Irelands' Hidden Pharmacy: Marine Biodiscovery Applied to Deep-Sea Sponges and Corals

By

Sam Afoullouss

A thesis presented for the degree of

Doctor of Philosophy

National University of Ireland, Galway

2017-2022

Abstract

Natural products have provided over 60% of our medicines to date, but their search has focused on easy to access habitats, leaving the marine realm underexplored. The deep-sea, which is the world's largest biome and contains some of the most biodiverse and inaccessible ecosystems, could represent an untapped source of natural products, especially since harsh environmental gradients and competition for resources in deep-sea habitats can lead to evolutionary adaptations with the potential to yield novel internal chemistry.

A series of three expeditions to Ireland's continental slope and associated submarine canyon systems, collected corals and sponges from rich habitats. As part of a wider project whose aim was to isolate novel bioactive deep-sea metabolites with pharmaceutical potential, herein focus was applied to two species. *Characella pachastrelloides* was selected because of the interesting chemical profile of an initial extract. *Paragorgia arborea* was selected for its bioactivity in a screening programme.

Four novel glycolipopeptides which contain a rare sugar moiety and rare D-amino acids were extracted and elucidated from the deep-sea sponge *Characella pachastrelloides* collected in Whittard Canyon at 800 m depth. These four compounds represent two pairs of stereo enantiomers which differ in the length of the fatty acid chain. Enantiomers Characellide A and B showed potent anti-inflammatory activity in a bioassay measuring ROS production. Subsequent synthesis of Characellide A by international colleagues demonstrated that further analysis of the absolute configuration of the natural product is required.

It became apparent that optimisation of classical and feature-based data acquisition parameters was required to effectively apply molecular networking to fractions of the *Characella pachastrelloides* extract. A fractional factorial analysis applied to three extracts (an alga, a sea squirt and a zoanthid) that varied widely in their chemodiversity showed that four factors have the greatest effect on molecular network topology: concentration, liquid chromatography separation, the number of precursors per cycle, and collision energy. The relative importance of factors varied between featured-based and classical molecular networking and depends on the planned application of the network itself.

Using the optimised data acquisition parameters for molecular networking the full chemodiversity of *Characella pachastrelloides* was analysed. This led to the discovery and elucidation of two additional characellides, light sensitive poecillastrins (cytotoxic polyketides), and a novel methylated histidine (6-methyl hercynine) whose imidazole ring has a unique methylation pattern not previously seen in nature whereby the methylation occurs on a C. In addition, cyanocobalamine was identified revealing its occurrence in nature for the first time.

Bioactivity screening led to prioritisation of the bubble gum coral *Paragorgia arborea* for natural product isolation. New bicyclic diterpenoids were elucidated. These included two miolenols with a rare cyclobutanol ring and one new and five known xeniolides. Miolenol showed interesting conformational flexibility as evidenced by temperature-variable NMR and NOESY correlations. All diterpenoids showed mild anti-plasmodial activity and the known compound 9-deoxyxenolide showed strong activity against a drug resistant strain of malaria.

This discovery of new and novel compounds illustrates the untapped chemistry of the deep sea. The pharmaceutical potential shown by these metabolites further highlights the importance of this biome in future drug discovery efforts.

Declaration

I declare that the work presented in this thesis is original material of my own research unless otherwise stated. This work, in whole or in part, has not been submitted for consideration for any other degree in NUI Galway, or elsewhere. To the best of my knowledge and belief, all sources have been properly acknowledged.

Sam Afoullouss 16/05/2022

Table of Contents

List of Figures

Chapter 6.

List of Tables

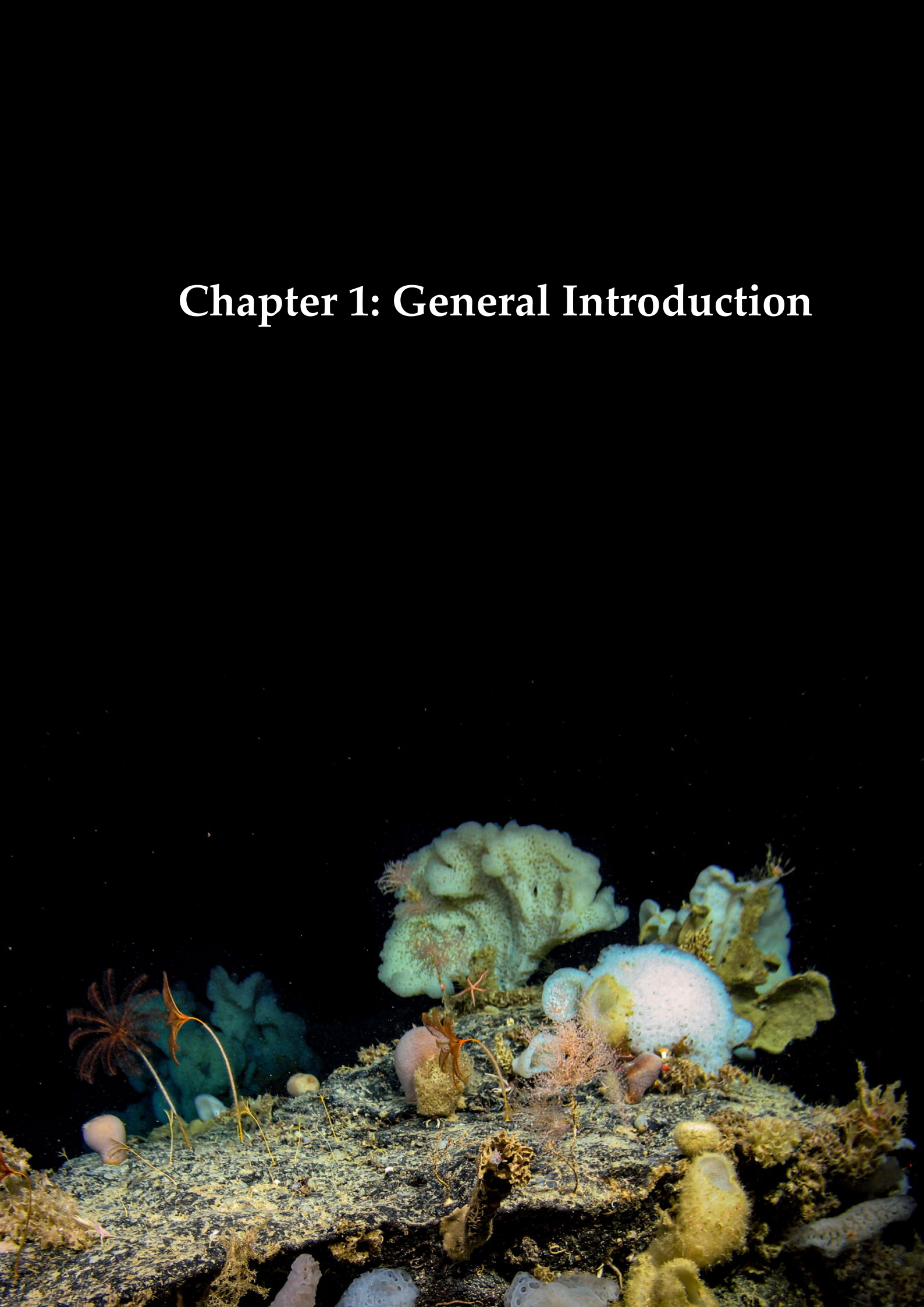
List of Abbreviations

ACN	Acetonitrile
ADC	Antibody-Drug Conjugates
amu	Atomic Mass Unit
br.	Broad
CE	Collision energy
CLMN	Classical molecular networking
COSY	1H-1H Correlation Spectroscopy
d	Doublet
Da	Daltons
DAD	Diode Array Detector
DCFH-DA	7',2'-DiChloroFluorescein Diacetate
DCM / CH_2CL_2	Dichloromethane
DFT	Density Functional Theory
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
DOE	Design of Experiments
EBSA	Ecologically or Biologically Significant marine Areas
EC_{50}	Half maximal effective concentration
ECD	Electronic Circular Dichroism
ELSD	Evaporative Light Scattering Detector
eV	Electron volts
FA	Formic Acid
FBMN	Featured-Based Molecular Networking
FDA	Food and Drug Administration
GC	Gas Chromatography
GCRMN	Global Coral Reef Monitoring Network

GI_{50}	50 % Growth inhibition
GNPS	Global Natural Products Social Molecular Networking
GPS	Global Positioning System
HMBC	Heteronuclear Multiply Bond Correlation
HOV	Human-Occupied vehicle
HPLC	High Performance Liquid Chromatography
HRESIMS	High-Resolution ElectroSpray Ionisation Mass Spectrometry
HSQC	Heteronuclear Single Quantum Correlation
HTS	High Throughput Screening
IC_{50}	Half maximal inhibitory concentration
IR	InfraRed
ISDB	*In silico* MS/MS database
IUCN	Centre of the International Union for Conservation of Nature and Natural Resources
KBA	Key Biodiversity Areas
l-FDAA	1--Fluoro-2,4-Dinitrophenyl-5-l-Alanine Amide
m	Multiplet
MeOH	Methanol
MNP	Marine natural products
MPA	Marine Protected Areas
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
MTT	3-(4,5-diMethylThiazol-2-yl)-2,5-diphenylTetrazolium bromide
mult.	Multiplicity
NAP	Network annotation propagation
NMR	Nuclear Magnetic Resonance
NO	Nitric Oxide
NOESY	Nuclear Overhauser Enhancement SpectroscopY
NP	Natural products
OFAT	One factor at a time
PCA	Principal Component Analysis
PM	Primary metabolites
PPC	Precursor per cycle
ppm	Parts per million
q	Quadruplet
qTOF	Quadrupole time of flight
ROS	Reactive Oxygen Species
ROVs	Remotely Operated Vehicles
RP-C_{18}	Reverse-phase silica gel column
RP-HPLC	Reverse Phase-High Performance Liquid Chromatography
RT	Retention time
s	Singlet
S.C.U.B.A	Self-contained under water breathing apparatus
SM	Secondary metabolites
SPE	Solid Phase Extraction
Std.E	Standardised effects
t	Triplet

TBHP	*tert*-butyl hydroperoxide
TFA	TriFluoroacetic acid
TNP	Terrestrial natural products
tR	Retention time
U/HPLC	Ultra/high pressure liquid chromatography
UN	United Nations
UNEP	United Nations Environmental Programme
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
VCD	Vibrational Circular Dichroism
VLC	Vacuum Liquid Chromatography
WoRMS	World Register of Marine Species
WPD	World Porifera Database
δ	Chemical Shift (ppm)

Chapter 1: General Introduction



Chapter 1. General Introduction

1.1 Marine Biodiscovery

1.1.1 Marine Biodiversity

Figure 1. Example of biodiversity found in marine ecosystems. This kelp forest contains a wide range of organisms including algae, sponges, corals, arthropods, and echinoderms. New Quay, Co. Clare, Ireland.

The marine biosphere contains some of the most biodiverse ecosystems on earth. Variations in environmental conditions such as water temperature, water current, dissolved carbon content, depth etc. gave rise to a broad range of ecosystems from kelp forests in temperate waters, to coral reefs in some shallow seas of the tropics[1]. These ecosystems contain a diversity of living organisms that is unmatched when compared to terrestrial ecosystems. As a result of life originating in our oceans before colonizing land, there are 34 phyla in the ocean, while 15 phyla are found on land. This results in ecosystems with large phylogenetic distances between species in marine habitats, when compared to terrestrial habitats[2].

1.1.2 Marine Natural Products and Drug Discovery

Marine organisms have proven to be a reservoir of structurally diverse natural products with potential therapeutic application[3]. Initial Interest in marine natural products was sparked by various toxins, such as tetrodotoxin, which were causing poisoning in coastal communities.[4-6] Chemical studies into these toxins highlighted unique chemical adaptations (e.g. bromination) that were previously unknown to occur in nature[7, 8]. Early knowledge of marine natural products paled in comparison to that of terrestrial organisms, particularly plants. This was caused by the difficulty in collecting marine organisms. Discovery of these new chemical features and scaffolds further encouraged searching for novel natural products with therapeutic potential from marine life forms[4, 7, 9]. 70% of marine natural product chemical scaffolds are unique to marine organisms[10]. Large intertidal and shallow water organisms such as algae, sponges and soft corals were the initial subjects of secondary metabolite investigation, due to the ease with which they could be collected. The introduction and wide adaption of Self-Contained Underwater Breathing Apparatus (SCUBA) revolutionised the early marine natural products field. This allowed collections of a diverse range of marine organisms down to depths of ~50 m.

Two phyla of marine invertebrates that were quickly identified as rich in original natural products were sponges (Phylum: Porifera) and corals (Phylum: Cnidaria), comprising 30.1% and 18.9% of 38,439 marine natural products isolated to date, respectively[11]. Multiple factors contribute to the rich diversity of natural products in sponges and corals [11].

Unlike any known terrestrial animals, sponges and corals in the adult phase are sessile. This inability to move led sponges and corals to build structures that allow them to filter-feed on food items

suspended in the water current (e.g., planktonic, bacteria, cell debris etc). This creates competition for space to get primary access to food in the water current, and/or light for symbiotic cyanobacteria. To prevent neighbouring organisms from over growing them (i.e. biofouling) and to deter predation, sponges and corals have been found to produce specialised metabolites which can inhibit potential colonisers[12-15]. These same metabolites are used in an offensive role, allowing sponges and corals to expand over neighbouring organisms. Many of these chemical defence metabolites inhibit cellular growth and functions that have been applied to therapeutic applications, including anti-tumour treatments.

Some shallow water coral and sponge species have developed symbiotic relationships with single-celled dinoflagellates and zooxanthellae, which can photosynthesise, providing the host organism with energy in the form of organic carbon products from photosynthesis. Sponges in particular host taxonomically diverse microbiomes, which in some species can comprise up to 35% of the sponge biomass. Recent studies have suggested that some species-specific microbial communities can metabolize nitrogen, sulphur, and phosphorus, and have the ability to fix carbon dioxide and biosynthesise B-vitamins, which are essential to life. The diversity of holobionts (e.g. bacteria, fungi, cyanobacteria) in combination with the different types of relationships (symbiotic, parasitic, etc.) effectively results in sponges and corals hosting a plethora of chemical architectures representing the diversity of biosynthetic gene cluster (BGC) present in the animals and microbes. Many natural products isolated from sponges and corals are combination products from different metabolic pathways and organisms[4].

To date, ~38,000 marine natural products have been reported[11]. Multiple comparative studies show a general trend of marine natural products possessing more potent bioactivities when compared to terrestrial natural products. In particular a large number of marine natural products show cytotoxic/antiproliferation activities[16]. This may be due to the role of these compounds in chemical defence[17, 18]. As of April 2022, there are 17 marine derived pharmaceuticals approved by the Federal Drugs Administration (FDA), with an additional 29 compounds in clinical trials[19].

Table 1. Approved marine derived pharmaceuticals with their corresponding source organism, chemical class, molecular target, and disease area of the therapeutic[19].

Compound Name	Trademark / Prescription	Year of FDA-Approval	Marine Organism	Chemical Class	Molecular Target	Disease Area
Cytarabine (Ara-C)	Cytosar-U®	-1969	Sponge	Nucleoside	DNA polymerase	Cancer: Leukemia
Vidarabine (Ara-A)	Arasena A®	-1976	Sponge	Nucleoside	Viral DNA polymerase	Antiviral: Herpes Simplex Virus
Ziconotide	Prialt®	-2004	Cone snail	Peptide	N-Type Ca channel	Pain: Severe Chronic Pain
Omega-3-acid ethyl esters *status is debatable at the moment	Lovaza®	-2004	Fish	Omega-3 fatty acids	Triglyceride-synthesizing enzymes	Hypertriglyceridemia
Eicosapenta enoic acid ethyl ester	Vascepa®	-2012	Fish	Omega-3 fatty acids	Triglyceride-synthesizing enzymes	Hypertriglyceridemia
Omega-3-carboxylic acid	Epanova®	-2014	Fish	Omega-3 fatty acids	Triglyceride-synthesizing enzymes	Hypertriglyceridemia
Eribulin Mesylate (E7389)	Halaven®	-2010	Sponge	Macrolide	Microtubles	Cancer: Metastatic Breast Cancer
Brentuximab vedotin (SGN-35)	Adcetris®	-2011	Mollusk/cyanobacterium	ADC (MMAE)	CD30 & microtubules	Cancer: Anaplastic large T-cell systemic malignant lymphoma, Hodgkin's disease
Trabectedin (ET-743)	Yondelis®	-2015Tunicate		Alkaloid	Minor groove of DNA	Cancer: Soft Tissue Sarcoma and Ovarian Cancer
Panobinostat	Farydak®	-2015	Sponge	Hydroxamic acid	Histone	Cancer: Multiple Myeloma
Plitidepsin**	Aplidin®	(2018) [Australia]	Tunicate	Depsipeptide	Deacetylase	Cancer: Multiple Myeloma, Leukemia, Lymphoma
Polatuzumab vedotin (DCDS-4501A)	Polivy™	-2019	Mollusk/cyanobacterium	ADC (MMAE)	eEF1A2	Cancer: Non-Hodgkin lymphoma, Chronic lymphocytic leukemia, Lymphoma, B-Cell lymphoma, Folicular
Enfortumab Vedotin-ejfv	PADCEV™	-2019	Mollusk/cyanobacterium	ADC (MMAE)	Nectin-4	Metastatic urothelial cancer
Lurbinectedin	Zepzelca™	(2020) * [Australia 2021]	Tunicate	Alkaloid	RNA Polymerase II	Cancer: Metastatic Small Cell Lung Cancer
Belantamab Mafodotin-blmf	Blenrep™	-2020	Mollusk/cyanobacterium	ADC (MMAF)	BCMA	Cancer: Relapsed/refractory multiple myelo
Disitamab Vedotin	Aidixi™	2021 (China)	Mollusk/cyanobacterium	ADC (MMAE)	HER2	Cancer: Urothelial Carcinoma, Advanced Cancer, Gastric Cancer, HER2 Overexpressi Gastric Carcinoma, Advanced Breast Cance Solid Tumors
Tisotumab vedotin-tftv	TIVDAK™	-2021	Mollusk/cyanobacterium	ADC (MMAE)	Microtubule	Metastatic cervical cancer

1.2 The Deep Sea

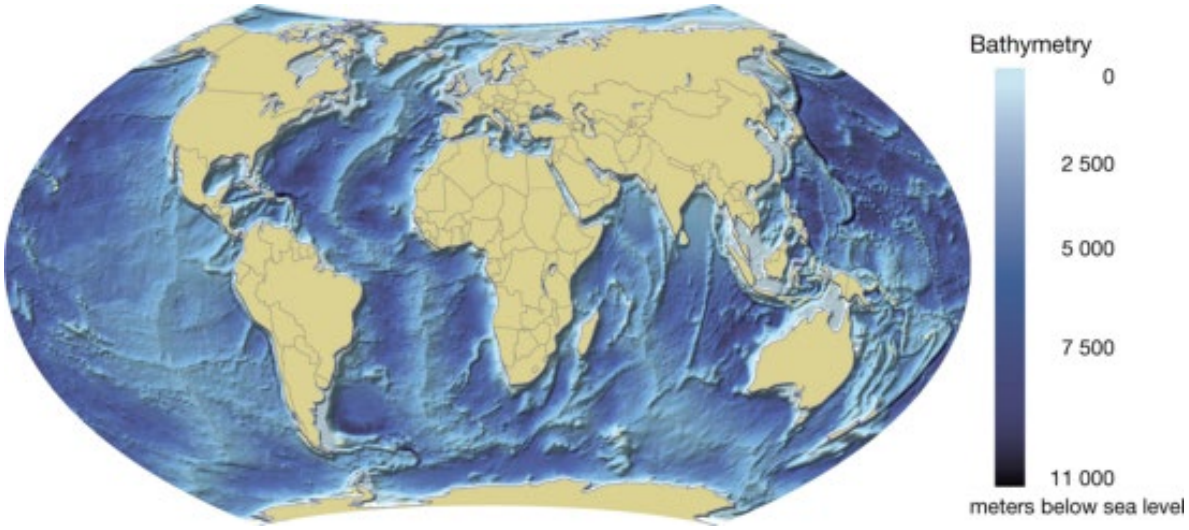


Figure 2. A global bathymetry map displaying the ocean depth and various submarine geographical features.[20]

The deep sea, ranging from 200 m to 11,034 m depth, is the largest biome on earth, encompassing 90% of the world's oceans[21]. The deep sea is often described as the last frontier on earth, with only ~5% of the deep sea being explored and less than 0.01% being sampled and studied in detail[22]. Initial studies into life in the deep sea used dredging to collect samples. In the 19th century, Edward Forbes concluded life could not exist past 600 m depths, due to the extreme environmental conditions [23]. This theory was tested with the first deep-sea expeditions on the H.M.S. Challenger and H.M.S. Porcupine, which discovered multiple forms of life at depth past 600 m, triggering deep-sea research, vastly improving our knowledge on these unique fauna and habitats[24]. Over the past 30 years, technological advances in robotics have enabled the development of Remotely Operated Vehicles (ROVs) with the ability to explore depths down to 3,000 meters[25, 26]. The improvement in technology has unveiled over 29 different deep-sea ecosystems[21, 27]. Some of these ecosystems are among the most biodiverse in the world, including coral reefs and sponge fields[28-30]. One of these unique habitats located around hydrothermal vents relies on chemosynthesis for energy and is one of the only ecosystems on earth that does not ultimately draw its energy from the sun[31].

The animals that inhabit these deep-sea ecosystems have had to adapt to extreme environmental conditions. As depth increases, two main changes to the environmental conditions occur. The first is the reduction of light, caused by the water column and its suspended particles absorbing the sun-light, resulting in permanent darkness past around 250 m depth. As there is no light for photosynthesis in the deep sea, organisms that inhabit these depth mainly depend on organic matter (cell debris, fecal matter), termed marine snow, that descend from the sun-lit waters above, as their main source of energy. This drives competition between the filter-feeding animals for space, in order to access food in the water currents. Although there is no oxygen production in the deep sea, cool, oxygen-rich water from the euphotic (sun-lit waters) are convected downwards at the poles to form deep water currents and water masses. In turn these water masses provide deep-sea animals with a sufficient oxygen supply. The most diverse ecosystems are located around geographic features (e.g. seamounts, continental shelfs) that effect the hydrographic conditions, which causes mixing of different water masses, containing both oxygen and source of energy in the form of organic matter. The second environmental condition to change with depth is the increase in hydrostatic pressure, 1 atm for every 10 meters depth. Increases in hydrostatic pressure have been shown to affect the biochemistry of cells, by altering the folding, conformation, and stability of proteins[32, 33].

These physical conditions also appear to effect speciation, with species diversity higher in deep-sea fauna, as compared to shallow water fauna[22, 24, 29, 30]. A habitat discovery curve, used to predict the number of undescribed species in a habitat, suggests there are undescribed deep-sea ecosystems that are yet to be discovered.

1.2.1 Submarine canyon systems

Submarine canyons systems, that cut into continental shelves, act as concentrators of the limited food resources. Water currents, driven by differences in water densities, draw food resources from the sun-lit shallow waters down to the benthic communities that inhabit these canyon systems. Submarine canyon systems are known to exist globally, but the technology required to map these features in high enough detail for in-depth studies has only been accessible over the past 20 years. These canyons are classified as essential habits for local fauna, both benthic and pelagic, and provide the rocky outcrops required for reefs to develop[27, 34].

Abiotic factors including depth, slope, sediment type, water current etc. give rise to multiple ecosystems within the marine canyon systems. These benthic communities which are centred around reef-forming sponges and corals, host a diverse range of organisms including echinoderms, polychaetes, tunicates, molluscs etc. Many of the animals that inhibit these reefs including black corals, bamboo corals, and glass sponges have a higher number of species in deep-sea habitats, highlighting the deep-sea's unique biodiversity. A recent study by Steffen et al. highlighted the change in prokaryotic communities and metabolomes of deep-sea sponges with depth and oceanographic conditions[35]. These findings indicated that studies on the same species from different locations may result in the discovery of new metabolites, further highlighting the deep seas undiscovered natural product potential.

A prominent ecosystem of the continental slope, between 600-900 m depth, is cold water coral (CWC) reef, comprising stony corals (Scleractinia) such as *Desmophyllum pertusum* [36]. CWC reefs are globally distributed and are also present on ridge systems and seamounts. These ecosystem-engineering corals create a habitat with specific niches for organisms such as soft corals and sponges. Initial studies suggest they are as biodiverse as their shallow water counterparts.

Watling et al. (2022), examined the global distribution of anthozoans group (20,846 records, 715 species), collected from depths between 700-3,000m. Species clustered into 25 bioregions, with the north Atlantic being the most diverse with 250 species. 160 of these were endemic to the north Atlantic with 62 species endemic to the northeast Atlantic (bioregion 12)[27].

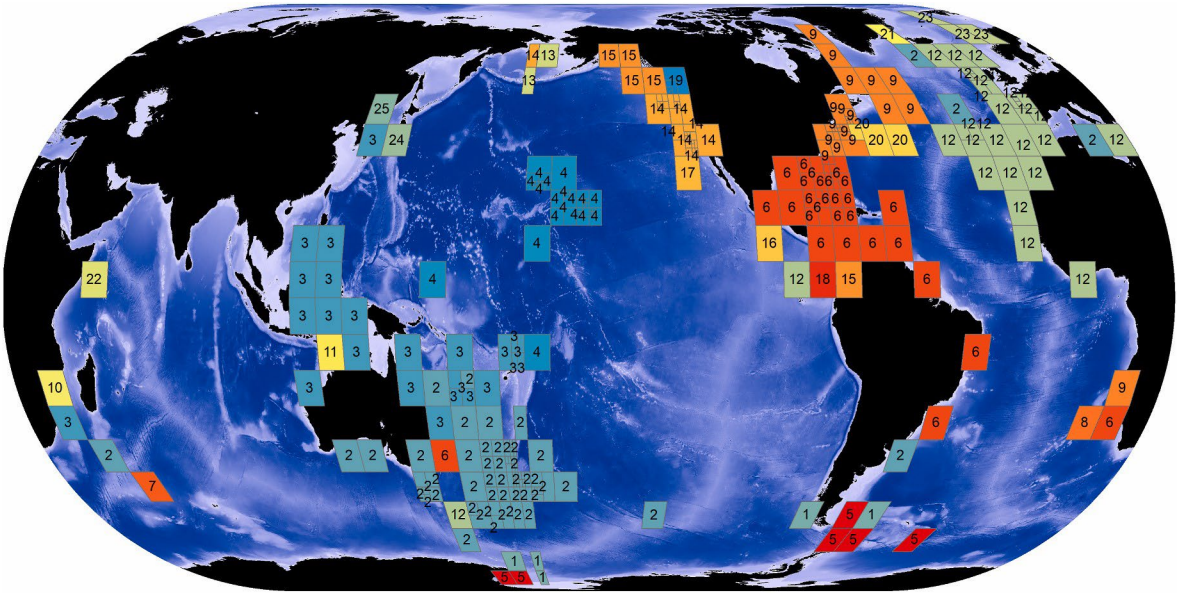


Figure 3. Bioregions generated by Infomap Bioregions network analysis. Different lower bathyal (700-3,000 m depth) bioregions are represented by variations in grid number and colour[27].

1.2.2 North-east Atlantic

The technological requirements to study deep-sea ecosystems is high. As a result, deep-sea research is limited to regions/countries where researchers have access to advanced research vessels, Remotely Operated Vehicles (ROVs) and high-resolution sonar mapping. The northeast Atlantic, specifically in the Irish exclusive economic zone (EEZ), is one of the most studied regions of deep sea in the world. High resolution mapping of the seafloor along the Celtic Continental margin was carried out by the INFOMAR project and is the first of its kind[37, 38]. These high-resolution maps are a vital requirement for more in-depth analysis by hydrology, oceanography, and ecology, to model water currents, species distribution etc. These maps combined with thousands of hours of ROV footage and sample collections have resulted in a high level of understanding with regards to the various ecosystems and the factors affecting them along the continental shelf. Whittard canyon system, which is located on the south-eastern edge of Ireland EEZ, is one the most studied deep-sea canyon systems, with over 800 publications to date. Depths range from 250 m at the top of the continental shelf to 3,500-4,000 meters at the abyssal plain. Whittard canyon is an active system, where sediment is being transported and distributed by different water bodies. This movement of sediment, combined with the organic matter from the rich phytoplankton blooms of the Atlantic, results in a high diversity of macrofauna within the canyon system[39]. This is most evident in the eastern branches of the canyon. The multitude of different sediment types, depths, currents etc aid in increasing the biodiversity found with the canyon systems[36, 40].

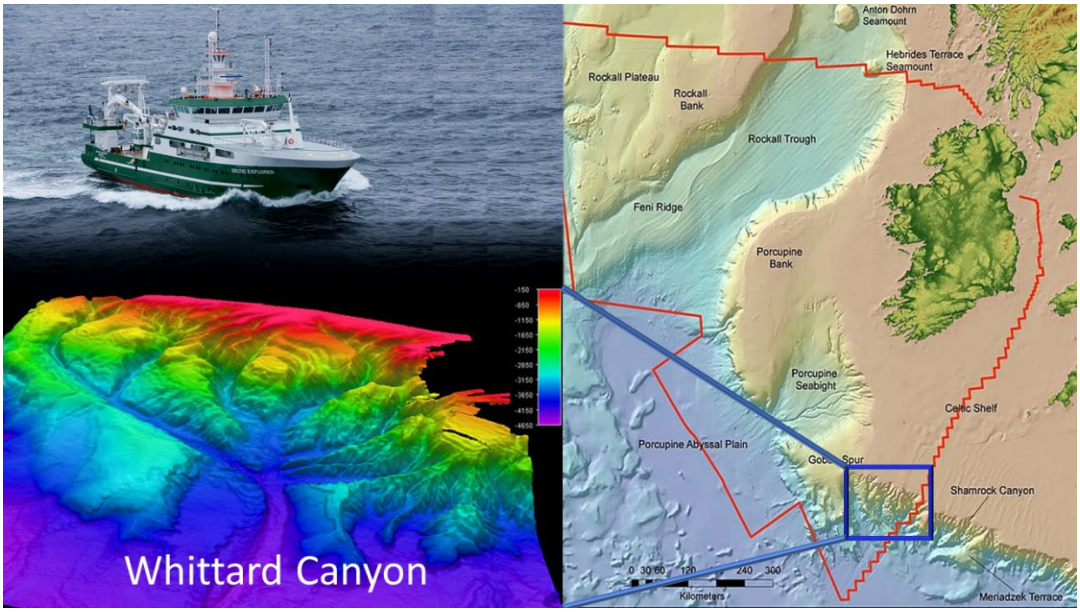


Figure 4. Image of the RV Celtic Explorer at sea (top left). The "True Map of Ireland" (right) indicating the EEZ boundry (red)of ireland, with a bathymetery maps of Whittard Canyon system (bottom left).

1.3 Deep-sea natural products

1.3.1 Metabolic adaptations

The harsh environmental conditions in the deep sea have forced the organisms living at these depths to adapt their metabolisms to high hydrostatic pressures, low oxygen and food availability, and low temperatures. Nevertheless, deep-sea organisms have a lower average metabolic rate when compared to their shallow water counterparts, which correlates strongly with food (energy) availability.

As rates of chemical reactions are governed by Le Chatelier's principles, the effect of pressure is more significant than that of temperature[41]. Pressure also plays a significant role in the conformation of molecules. This can affect the metabolic processes by changing the shape and size of binding sites in enzymes and carrier-protein interactions. Barret et al. identified the first pressure-regulated gene[42]. Future studies have proposed metabolic responses to increased pressure, including changes too membrane fatty acid synthesis[43, 44]. An *in-silico* study, by Wright et al., examining the effect of pressure on the polyketide synthase (PKS) pathway, found pressure significantly reduced the volume of the final metabolite[41]. It is hypothesised that the modifications to metabolic pathways required by piezotolerant organisms may result in novel secondary metabolites[41].

While the deep sea hosts some of the most species rich ecosystems in the world, less than 2% of marine natural products have been isolated from ecosystems depths greater than 1000 meters. The difficulty in collecting deep-sea specimens has left many benthic deep-sea species undescribed and so leaves an untapped potential of novel natural products. It is estimated that over 95% of deep-sea species are yet to be described, let alone undergo chemical analysis. The high number of single rare species, modified biosynthetic gene clusters, and complex microbial communities, indicates that deep-sea organisms may possess unique metabolites with novel chemical architectures[21, 30, 45, 46].

In the quest for finding natural products with therapeutic applications, novel bioactive structures possess great potential as drug leads. Novel structures, with unique chemical scaffolds, can bind to new biological targets, resulting in many being developed into first of class therapeutics[4]. This is

especially important in the search for novel antibiotics to treat drug resistant bacteria[47]. An analysis on ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profile of 179 deep-sea natural products found 2/3 occupied the Known Drug Space (KDS) and 40% possessed drug like properties[48]. Interest in deep-sea natural product chemistry has increased, with deep-sea natural product publications representing 0.7%-2.4% of MNP papers published annually in up to 2020, rising to 4.5% in 2021. The dramatic increase in research interest is mostly due to increased access to deep-sea ROVs and increased success cultivating deep-sea microbes.

The subsequent review of deep-sea sponge and coral natural products is a non-exhaustive list and contains representative examples of metabolites reported from specimens collected from depths deeper than 200 m.

1.3.2 Deep-sea Sponge Natural Products

Deep-sea sponges, to date, are the largest source of deep-sea natural products. The natural products isolated cover a broad range of molecular classes, including polyketides, terpenoids, polypeptides and modified nucleic acids. Halichondrin B and discodermolide are both examples of deep-sea sponge natural products that have been proven to be promising drug leads with anticancer activities[49, 50]. Two main sponge classes are prevalent in the deep-sea, glass sponges (class Hexactinellida) and demosponges (class Demospongiae).

1.3.2.1 Hexactinellida

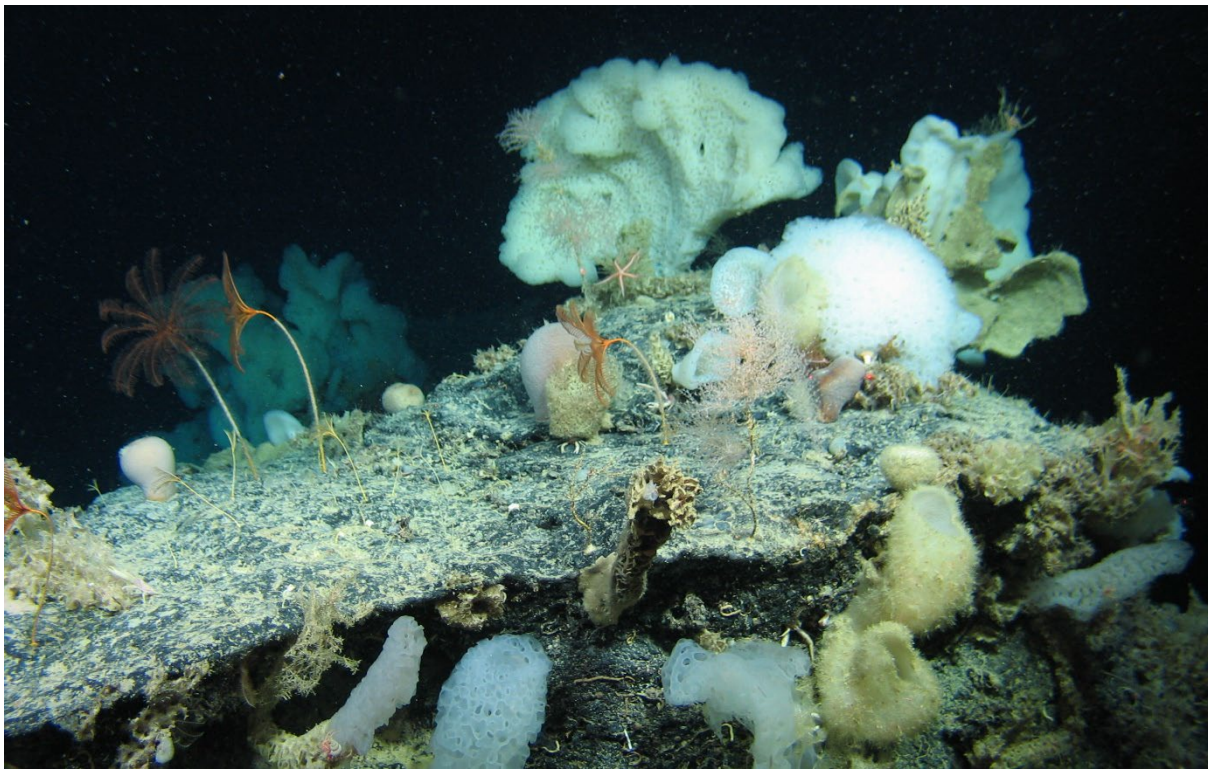


Figure 5. An assemblage of deep-sea glass sponges in the NE Atlantic.

Glass sponges are cold-water specialists and have a global distributed. Although some species of glass sponges can be found in the shallow waters of the Southern Ocean and Chilean fjords, they predominantly found in the deep sea. Comprised mostly of large, intertwined silica spicules, glass sponges possess strong physical defence against predation. Over 50 natural products have been isolated from glass sponges. Most are cell membrane derivatives including fatty acids, steroids and

cerebrosides. Collected at 505 m depth off the Kuril Islands, a specimen of *Aulosaccus* sp., yielded nine fatty acid derivatives (**1, 2**), with three featuring cyclopropane rings(**3**)[51]. Two new steroids (**4**, **5**) and 27 cerebrosides (**6**, **7**) were also isolated from the same specimen of *Aulosaccus sp*. by Santalova et al[52-54]. Glassponsine (**8**), a taurine derivative, was isolated from an Antarctic glass sponge, *Anoxycalyx joubini*, collected at a depth of 280 m[55]. Glassponsine showed anti-feed activity, indicating it may play a role as a defence metabolite[12].

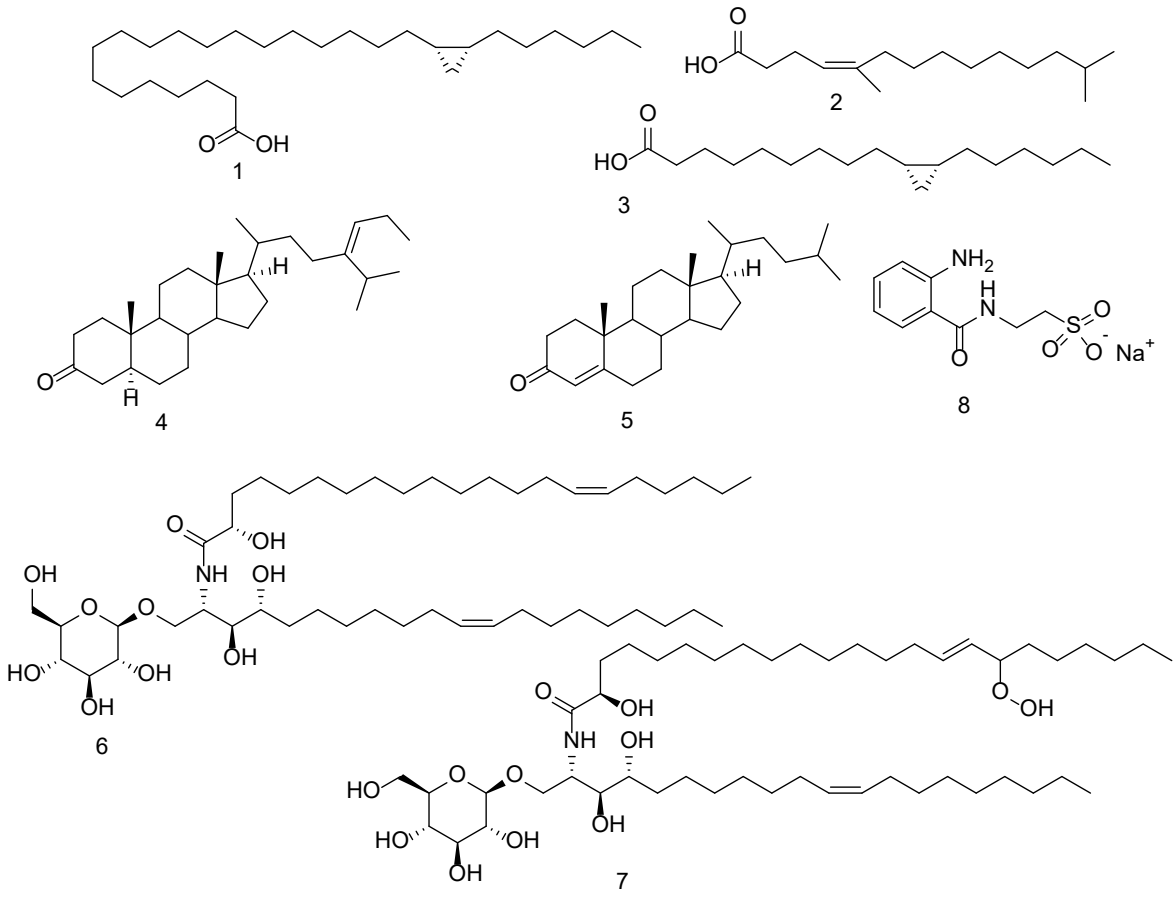


Figure 6. Chemcial strucutres of metabolites isolated from deep-sea glass sponges.

1.3.2.2 Demospongiae

Order Tetractinellida

A deep-sea *Geodia* sponge covered with an encrusting *Halichondria sp.* led to the isolation of frankliolides A-C (**9**, **10**), novel polyketide phosphodiesters with cytotoxic bioactivities[56]. A manned submersible off the coast of the Bahamas collected a *Dercitus sp.*, which contained an aromatic pentacyclic alkaloid. Dercitin (**11**) possesses an unprecedented thiazole-containing fused ring[57]. Dericitin has shown potent antitumor activity both *in vitro* and *in vivo*, as well as antiviral activity[58]. Continued studies on the deep-sea tetractinellid sponges are uncovering multiple bioactive aromatic alkaloids with fused thiazole rings. Studies of pachastrellid sponges uncovered dercitamide (**12**), dercitamine (**13**), cycolodercitin (**14**), and nordercitin (**15**) from *Stelletta sp.* Poecillastrin A (**16**) is a class of macrolide lactams first isolated from *Poecillastra sp.,* with antitumour activity[59]. Further studies isolated and characterised bioactive poecillastrins B-H (**17**), found in deep-sea sponges of the family Vulcanellidae [60-63]. The chemical investigation of a sponge of the genus *Erylus*, sampled at 500 m depth near New Caledonia, resulted in the characterization of two novel oligogalactosides, eryloside C and D(**18**, **19**)[64].

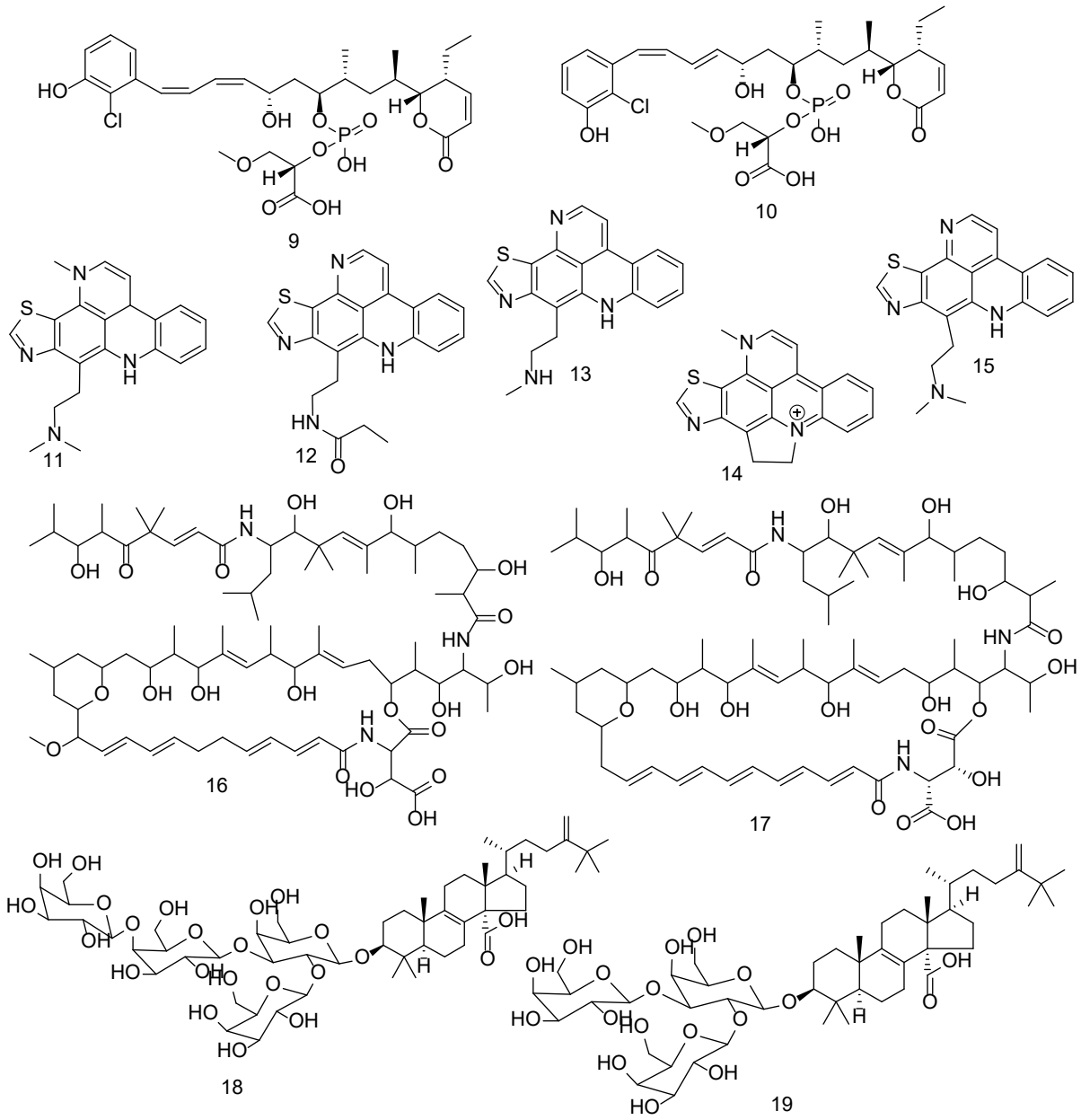


Figure 7. A variety of secondary metabolites isolated from deep-sea sponges of the order Tetractinellida.

The Genus *Geodia* has been shown to host a diverse array of secondary metabolites, thanks in part to its global distribution and the varying depth distribution of different species. *Geodia japonica*, collected in the south China sea, was shown to contain two new nortriterpenoids, geoditins A and B (**20**, **21**), and a new sterol (**22**)[65]. A *Geodia barretti* specimen, collected from the deep sea in the northeast Atlantic, lead to the discovery of the polypeptides, barretides A and B(**23**, **24**)[66]. Further investigation of the sponge's genome highlighted sequences for five more barretides C-G, which possess antifouling activity[67]. Evidence suggests the barretides are produced by sponge biosynthetic pathways rather than associated microbes. A *Geodia sp*. sampled from Australia, contained geodin A (**25**), a novel macrocyclic polyketide[68].

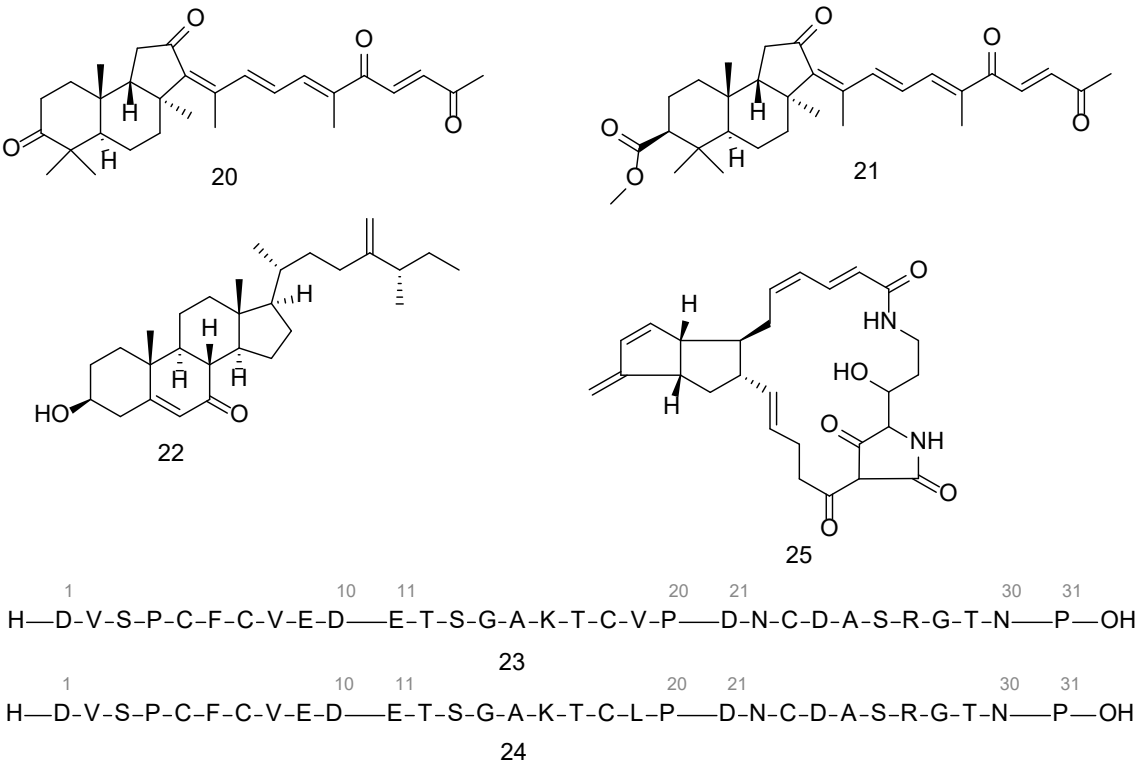


Figure 8. Chemical structures of metabolites isolated from deep-sea sponges of the order Tetractinellida.

Six new sesquiterpenes with antifungal and cytotoxic activity, the sollasins A-F (**26**-**31**), were isolated after a successful bioassay guided approach highlighted the potential of an extract of the deep-sea sponge *Poecillastra sollasi[69].* A specimen of *Dercitus* (*Stoeba) extensa* trawled at 260 m depth, led to the discovery of novel cytotoxic furansosesterterpens, shinsonefuran (**32**) and halisulfate 7 (**33**)[70]. Both metabolites have strong activities against HeLa cells.

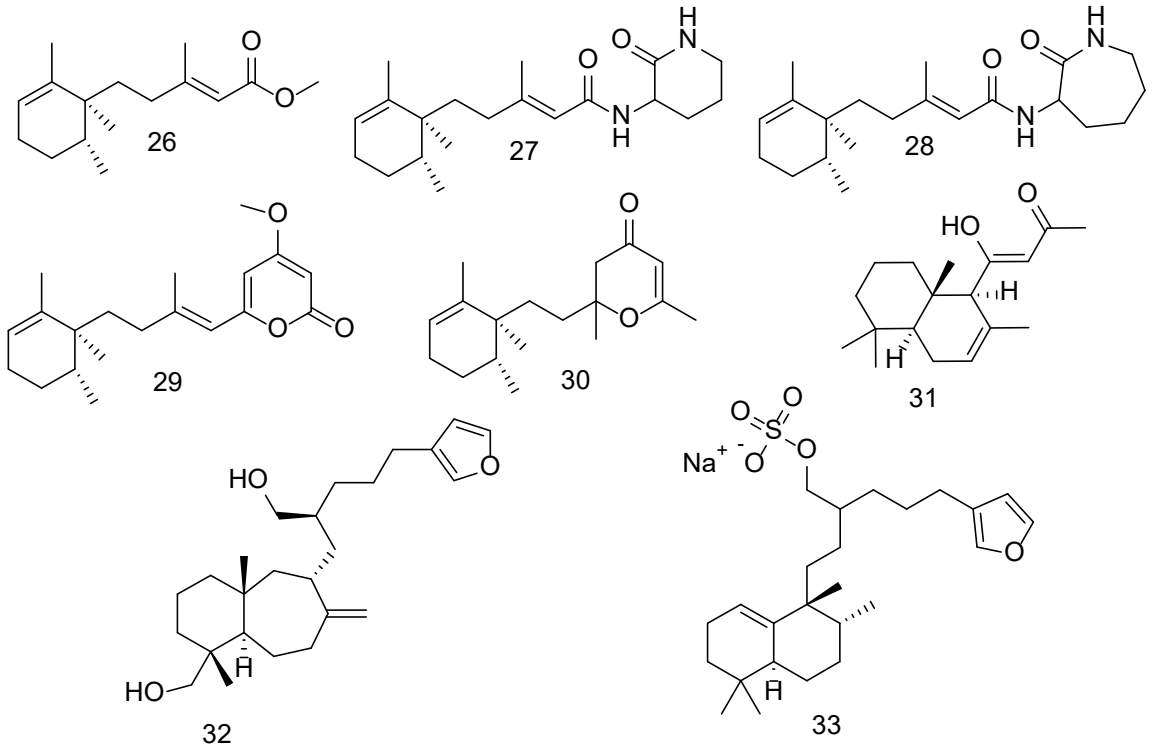


Figure 9. Bioactive metabolites isolated from deep-sea sponges of the order Tetractinellida.

Leiodermatolide A (**34**) is a novel macrolide with an unprecedented 16-membered skeleton discovered after bioassay guided isolation of a deep-sea sponge extract of *Leiodermatium sp.[71].* Analysis of leiodermatolide bioactivity reported IC_{50} values lower than 10 nM against PANC-1 carcinoma, human A549 lung adenocarcinoma, and DLD-1 colorectal carcinoma. Two more leiodermatolides B and C (**35**, **36**) were isolated from a deep-sea sponge of the same genus[72].

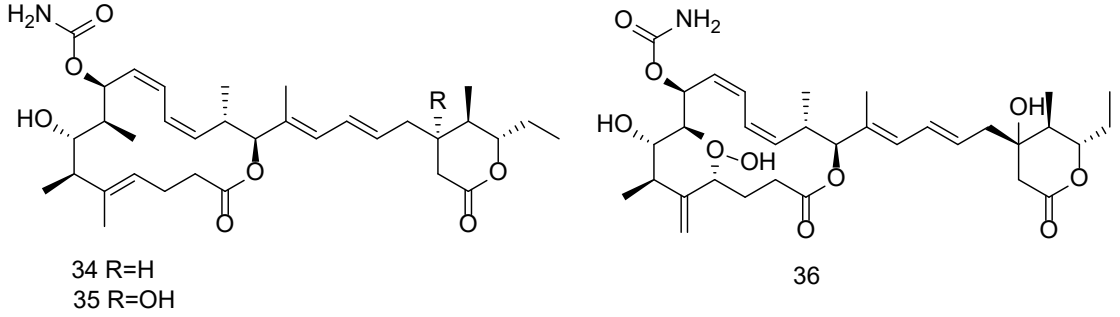


Figure 10. Chemical structures of the Leiodermatolides isolated from *Leiodermatium sp.*

A Caribbean lithistid sponge *Neoschrammeniella fulvodesmus* (Lévi & Lévi, 1983) *,* sampled from 300 m depth, revealed corallistin A (**37**), a porphyrin pigment with *in vitro* cytotoxicity to human KB cell lines[73]. A new polynitrogen metabolite, corallistine (**38**), was isolated from the same specimen[74]. A new depsipeptide, polydiscamide A (**39**), which displayed cytotoxic effects against A549 cancer cell line, was isolated from the deep-sea sponge *Discodermia sp[75].* A deep-sea sponge of the genius *Leiodermatium* collected at 229 m depth, resulted in the isolation of a new class of 19-membered macrolides, leiodolides A and B (**40**, **41**). Leiodolide A possessed strong cytotoxic activities against multiple cancer cell lines including OVCAR-3 ovarian cancer and HL-60 Leukaemia[76]. An interesting family of cyclic peptides, microsclerodermins, which incorporate modified amino acids, were first

isolated from the lithistid deep-sea sponge *Microscleroderma sp.[77]*. Microsclerodermin A (**42**) showed antifungal activity towards *C. albicans.*

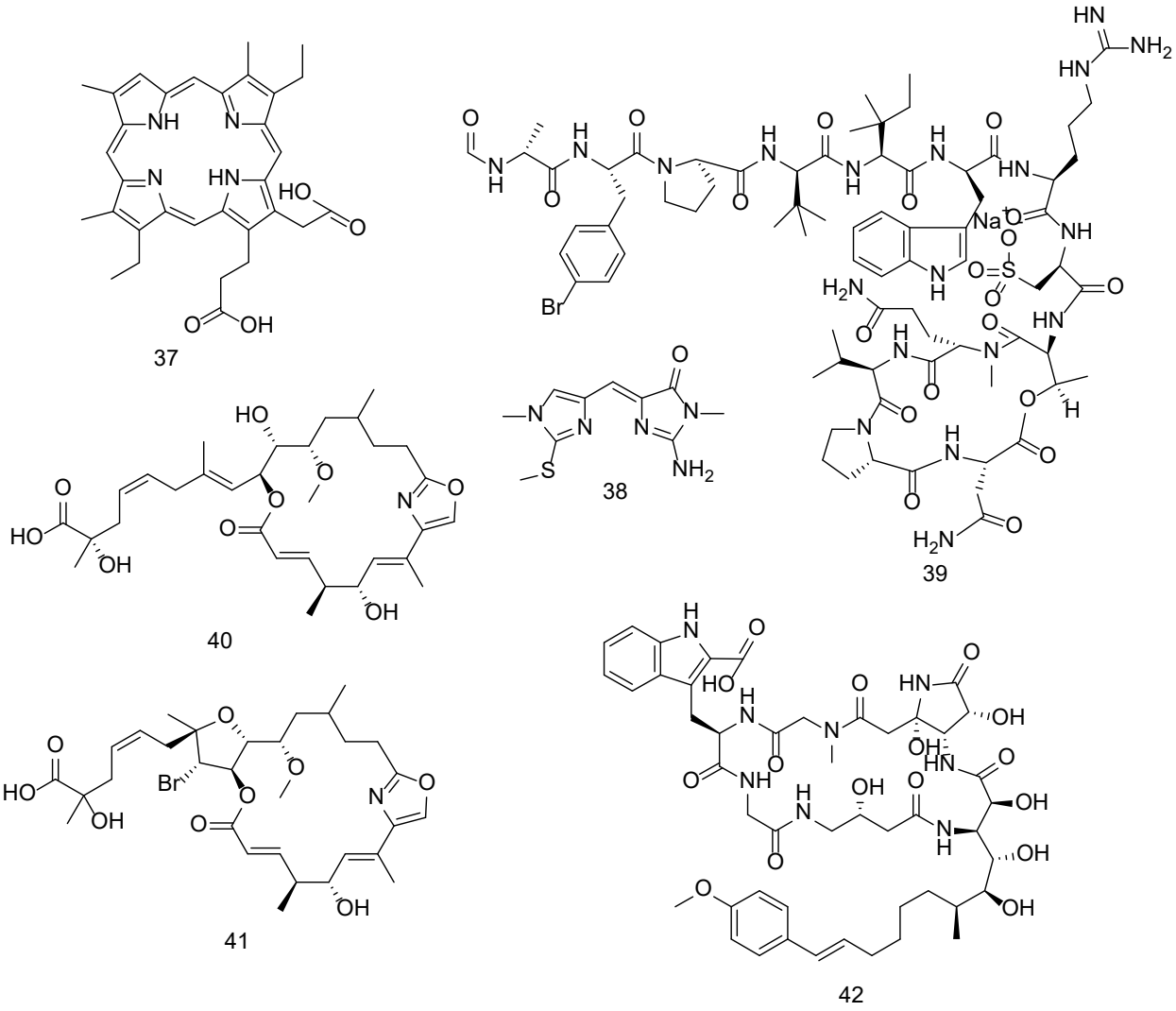
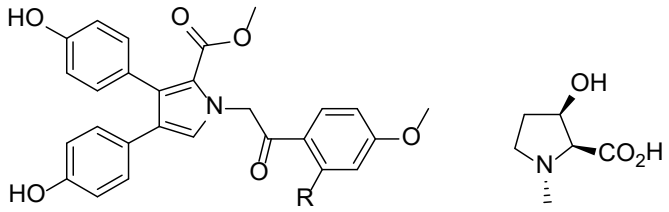


Figure 11. Natural products isolated from the deep-sea sponges of the order Tetractinellida.

Order Dendroceratida

Dredging at a depth of 1,000 m of Bass strait, Australia, collected a specimen of *Dendrilla cactos*. A chemical investigation uncovered two novel aromatic alkaloids, lamellarin O and P (**43**, **44**), and a new amino acid in high abundance, *cis*-3-hydroxyN-methyl-L-proline (**45**) [78, 79]. Lamellarin O and P have mild antimicrobial activity.



43 R=H

44 R=OH

45

Figure 12. Metabolites isolated from deep-sea sponges of the order Dendroceratida, including aromotic alkaloids, new amino acids.

Order Dictyoceratida

New sesquiterpene quinones and quinols (**46**, **47**), with mild antimicrobial actives were isolated from the sponge *Euryspongia sp.,* collected from ~200 m depth off the Great Australian Bight[80].The fatty acids heterofibrins (**48**-**50**) were isolated from a deep-sea specimen of *Spongia sp.* after bioassay-guided fractionation[81]. Despite their relatively simple chemical structure, they were shown to affect lipid droplet biogenesis in AML12 hepatocytes and affect fatty acid distribution in zebrafish embryos. Interesting meroterpenes with selective antimicrobial activity against Gram-positive bacteria, fascioquinols A-F, were discovered in a deep-sea sponge of the genus *Fasciospongia*. Fascioquinols A (**51**) and B (**52**) displayed IC_{50} values of 0.9-2.5 µM and 0.3-7.0 µM towards *S. aureus* and *B. subtilis,* respectively[82]. A sponge, *Sarcotragus fasciculatus* (Pallas, 1766) *,* collected from 500m depth from the Norfolk Ridge in New Caledonia revealed a hydroxylated hydroquinone sulphate (**53**), which inhibited HIV-Integrase activity[83]. A large family of novel bicyclic sesterterpenes, the luffarins A-C, (**54**-**56**) were isolated from the deep-sea sponge *Luffariella geometrica*, sampled at 350 m depth on the Great Australian Bight[84].

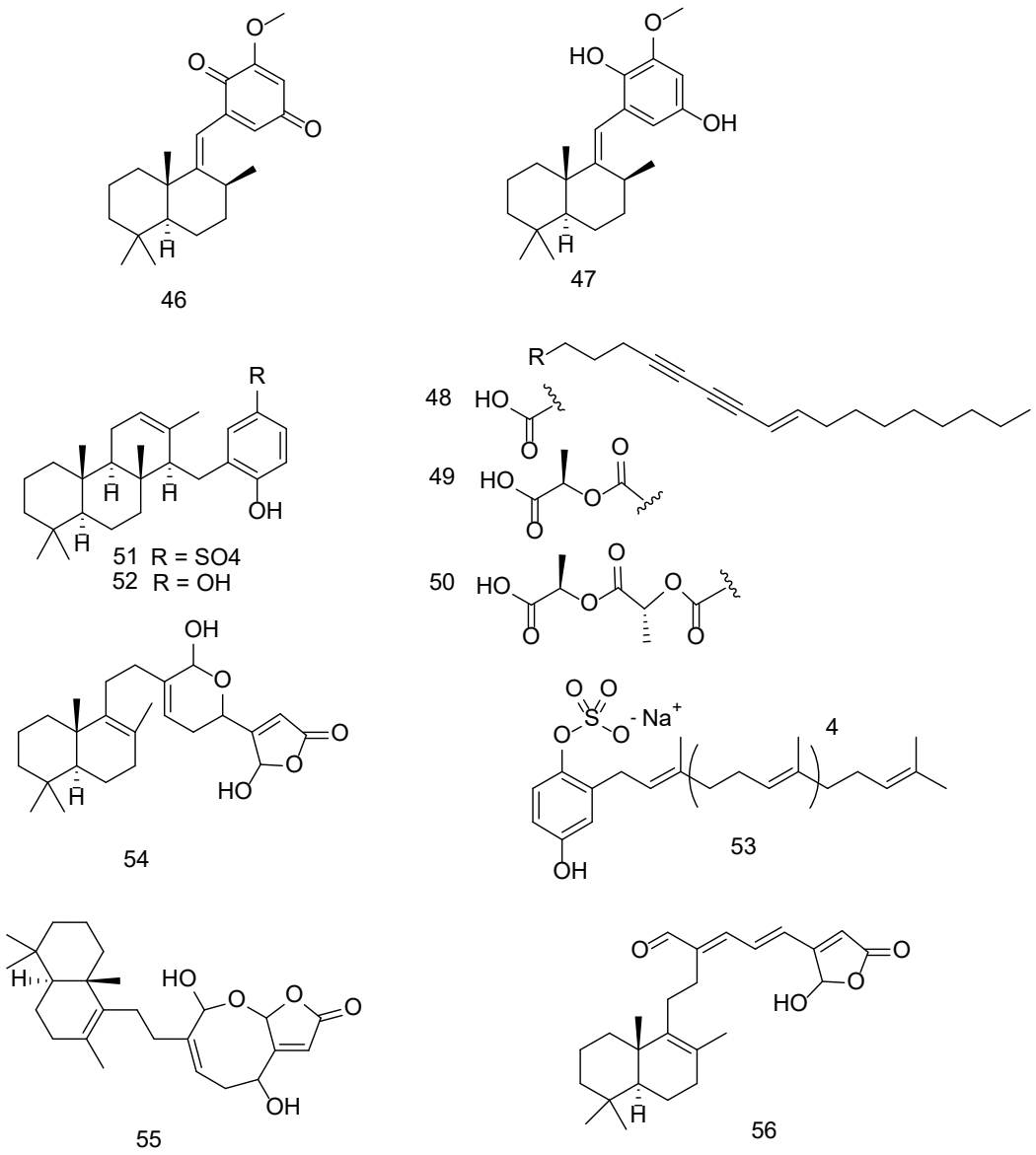


Figure 13. Metabolites isolated from deep-sea sponges of the order Dicytoceratida, including aromotic terpenes, modified fatty acids and polycyclic terpenes.

Order Axinellida

Sponges from the order Axinellida has been the source of numerous natural products of various chemical architectures, particularly terpenes and alkaloids. A deep-sea axinellid sponge, *Axinella sp*. hosted a diverse array of known bioactive secondary metabolites and three new massadine. 3-O-methyl massadine chloride (**57**) showed strong antibacterial activities against both gram-positive and gram-negative bacteria[85]. Their promising bioactivities and interesting structure has led to several attempts to synthesise the massadines[86]. A novel cytotoxic bisindole, Dragmacidin (**58**), was recovered from the sponge *Dragmacidon sp.* Dragmacidin, which comprises of two indole rings bridged by a piperazine ring, had excellent picomolar activities toward lung cancer (human A549 cells) tumour cell lines (MDAMB)[87]. Novel diterpenoids, reiswigins A and B (**59**, **60**), were isolated from the deep-sea sponge *Myrmekioderma gyroderma* (Alcolado, 1984), from 330 m depth. Enantiospecific synthesis was used to determine their absolute stereochemistry[88]. Both reiswigins show antiviral activity against murine hepatitis virus and HSV-1[89].

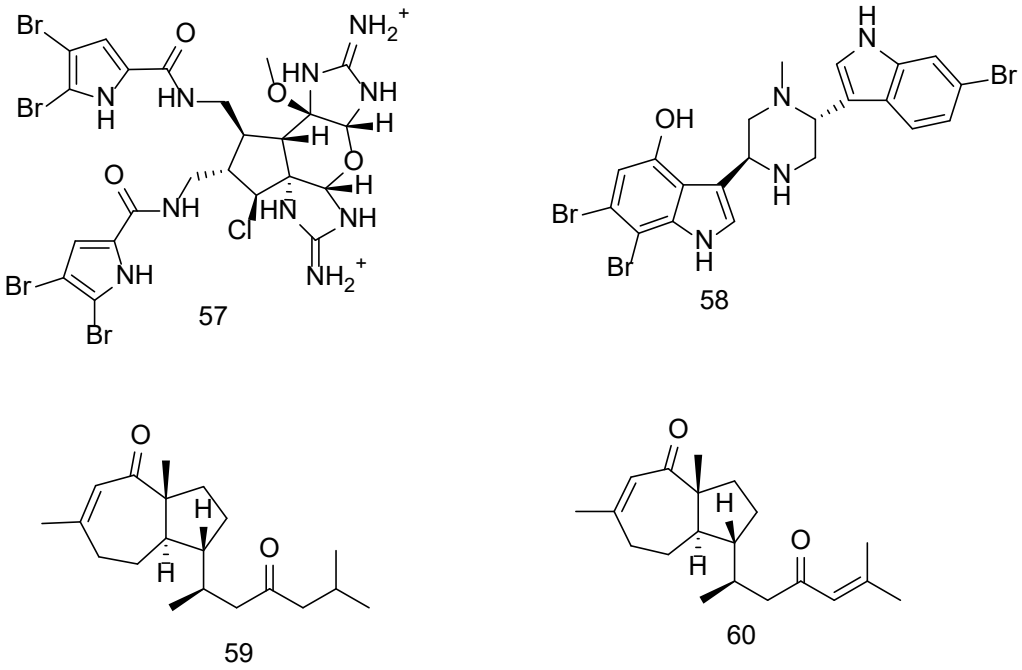


Figure 14. Secondary metabolites isolated from deep-sea sponges of the order Axinellida.

Order Suberitida

New sesquiterpenes, (+)-curcuphenol (**61**) and (+)-curcudiol (**62**), were discovered from the sponge *Epipolasis sp*.A. Suberitid sponges collect from 174-355 m depth, contained new bisindole alkaloids, topsentin (**63**) and bromotopsentin (**64**). Both alkaloids displayed antiviral (HSV-1 and corona A-59) and *in vivo* antitumor activity against P388 (T/C 137%, 150 mg/kg) and B16 melanoma (T/C 144%, 37.5 mg/kg)[90] .

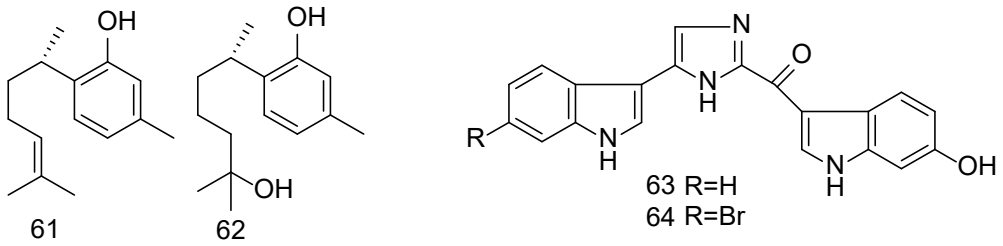


Figure 15. An array of bioactive metabolites isolated from deep-sea sponges of the order Suberitida.

Order Haplosclerida

Neopetrosiquinones A and B (**65**, **66**), sesquiterpene benzoquinones with in vitro anticancer activities were isolated from the deep-sea sponge *Neopetrosia proxima*. Neopetrosiquinone A displayed activity against DLD-1 human colorectal adenocarcinoma cell line and PANC-1 human pancreatic carcinoma cell line with IC_{50} values of 3.7 and 6.1 mM, respectively[91]. Neopetrosiquinone B showed activity against the same cell lines with IC_{50} values of 9.8 and 13.8 mM. A specimen of the sponge *Petrosia sp.* collected by ROV from 450 m depth, led to the discovery of durynes A-F[92, 93]. While their linear acetylene structures are of limited structural interest, durynes A (**67**) with nanomolar cytotoxicity IC_{50} against HeLa cells as low as 80 nM, is notable. This led to the development of a one-pot synthesis of duryne[94]. With the aim of finding natural products with antimalarial activity against *Plasmodium falciparum*, bioactivity screening of the deep-sea sponge extracts led to the discovery of four new meroterpenes alisiaquinones A-C (**68**-**70**) and alisiaquinol (**71**)[95]. These new compounds show µM

activity using the protein farnesyl transferase (PFTase) assay. Alisiaquinones A and C also exhibited strong inhibition against *Plasmodium falciparum,* with selectivity indices (SI) of 4 and 500, respectively.

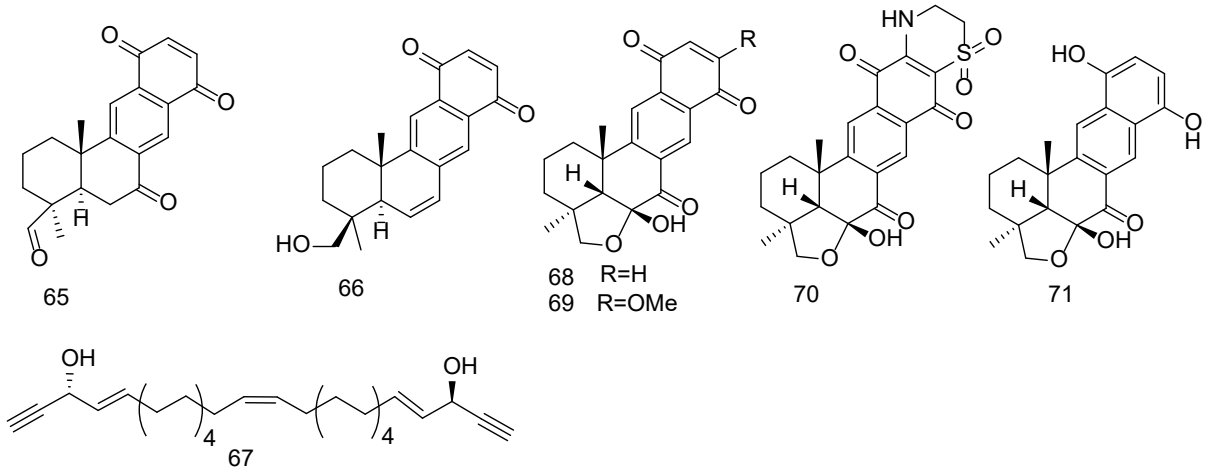


Figure 16. Natural products isolated from the deep-sea spongs of the Order Haplosclerida.

Order Poecilosclerida

A trawling expedition of the Great Australian Bight collected a sponge from the genus *Mycale* (*Arenochalina*) Lendenfeld, 1887 from 400 m depth. Chemical investigation into the sample highlighted a new family of alkaloids, mirabilins A–F, with mirabilins A and B (**72**, **73**) shown in Figure 19[96]. Multiple discorhabdins have been isolated from deep-sea poecilosclerid sponges. Two sperate species of sponge, *Latrunculia sp.* collected from Antarctic waters, and *Negombata sp.* sampled from Australian waters, contained the antimicrobial discorhabdin R (**74**)[97]. Cytotoxic analogues, with 5-S-methyls, discorhabdins S (**75**), were later isolated from a sponge of the genus *Batzella.* A new derivative of halichondrin B, isohomohalichondrin B (**76**), was isolated from a large-scale collection (~200 Kg) of the sponge *Lissodendoryx sp.,* from 100 m depth[98]. The same group also identified five new polyester macrolides. Screening in the National Cancer Institute highlighted isohomohalichondrin B as possessing selective cytotoxicity[99-101].

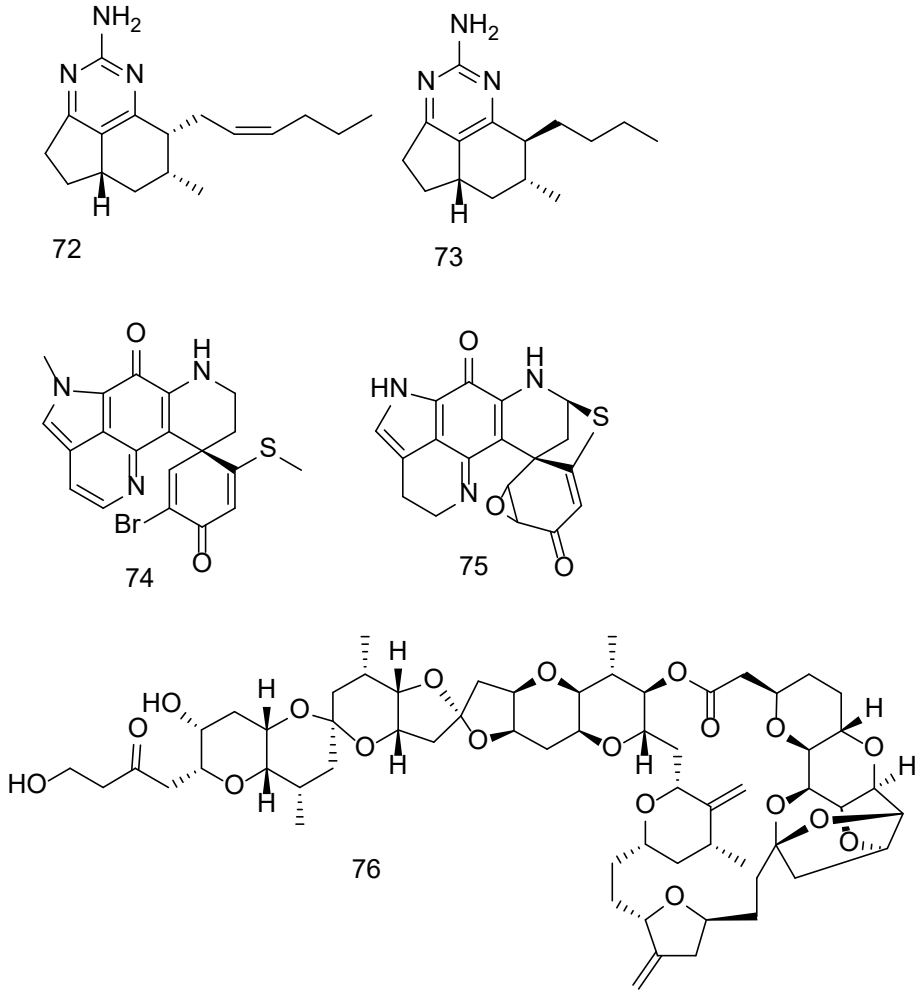


Figure 17. Secondary metabolites reported from deep-sea sponges of the order Poecilosclerida

Order Merliida

A sponge collected at 550 m depth, thought to be a new species of the genus *Hamacantha,* furnished new bisindole alkaloids hamacanthin A and B (**77**, **78**). Both metabolites showed strong antimicrobial activities against *C. albicans* (1.6; 6.2 µg/mL MIC)*, C. neoformans* (3.2; 6.2 µg/mL MIC)*,* and *Bacillus subtilis* (3.1; 1.6 µg/mL MIC), for hamacanthin A and B respectively.

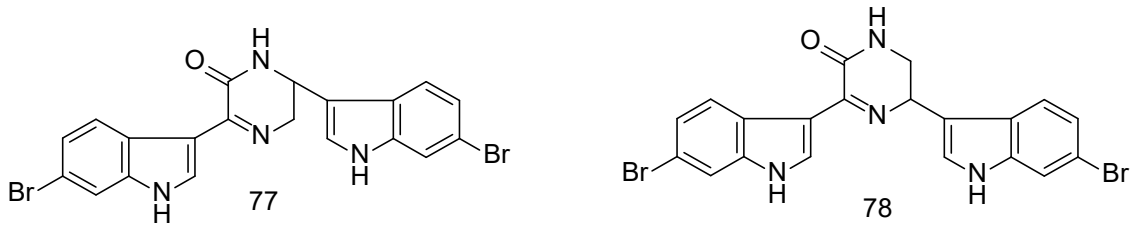


Figure 1 Bisindole alkaloids isolated from the deep-sea Merliida sponge.

Order Verongiida

A deep-sea sponge *Suberea* sp., collected off Guam yielded novel psammaplysins I and J (**79**, **80**), bromotyrosine derivatives with a rare bromotyramine moiety[102].

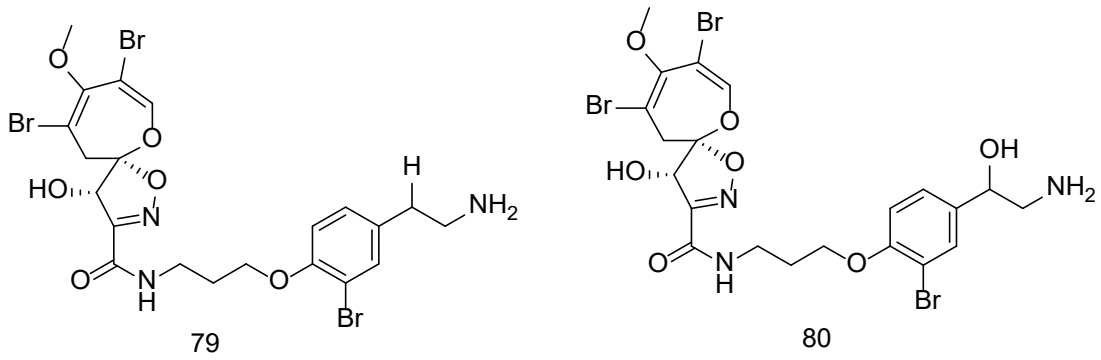


Figure 19. Bromotyrosine derivatives isolated from a deep-sea verongiid sponge

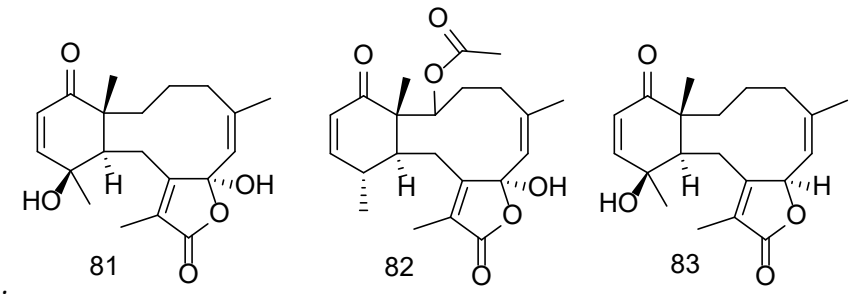
1.3.3 Deep-sea Cnidaria natural products

Of the marine macro-organisms, Cnidaria are the second largest source of new natural products, mainly comprising terpenoid derived natural products. The phylum Cnidaria comprises four classes, Anthozoa, Hydrozoa, Cubozoa, and Scyphoza. The class Anthozoa contains over 6,500 species and is the most species rich class of Cnidarians. Within the class Anthozoa, soft corals (Order Alcyonacea, Subclass Octocorallia) have proven to be the richest source of new metabolites.

Subclass Octocorallia (octocorals)

Order Pennatulacea (sea pens)

Sea pens are a diverse group of octocorals with over 200 species identified. Their distribution ranges in depth from the intertidal zone to 6,000 m.[103] Sea pens can be found in multiple deep-sea ecosystems including coral reefs and soft sediment benthic communities called sea pen fields. Shallow water sea pens have been shown to produce cembranes, diterpenes, and briaranes[11]. Only one report of deep-sea sea pen natural products was published. In 2019, Thomas et al., isolated three new briarane diterpenes, bathyptilone A-C (**81**-**83**), from the Antarctic deep-sea seapen, *Anthoptilum grandiflorum[104]*



.

Figure 20. Briarane diterpenes isolated from the deep-sea sea pen *Anthoptilum grandiflorum*.

Order Alcyonacea

The soft-coral *Echinogorgia pseudosassapo* revealed the presence of two new alkaloids, pseudozoanthozanthins III and IV (**84**, **85**), which showed mild activity against HSV-1 and RSV[105]. Two sesquiterpenes (**86**, **87**) were also isolated from the same specimen which showed antilarval activity towards *Balanus amphitrite* larvae[105]. Papakusterol (**88**), a cholesterol derivative containing a cyclopropyl ring was isolated from a mixture of gorgonians, collected at 300 m depth off the coast of Hawaii[106]. Novel tricyclic sesquiterpenes, paesslaerins A and B (**89**, **90**), were discovered from *Alcyonium paessleri*, collected from the south Georgia islands at a depth of 200 m[107].

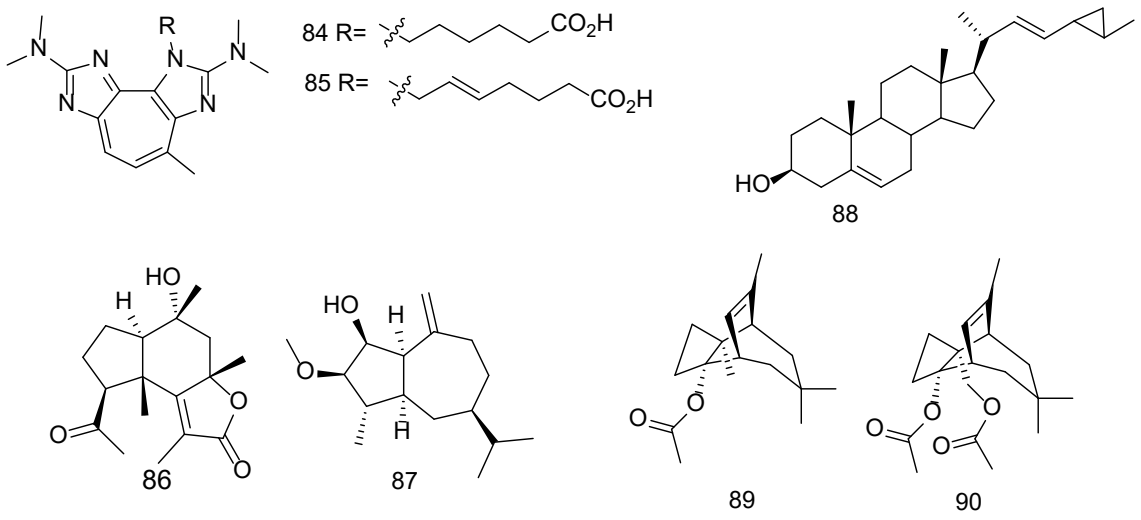


Figure 21. Natural products isolated from deep-sea soft corals.

A range of bicyclic diterpenoids (**91**-**94**), with chemical similarities to xenicins, were discovered from the deep-sea gorgonian *Corallium sp.,[108].* A chemical investigation of a similar deep-sea gorgonian, *Paragorgia arborea,* resulted in the isolation of a novel xeniolide, arboxeniolide-1 (**95**)[109]. Three cytotoxic linderazulenes (**96**, **97**), were recovered from a gorgonian *Paramuricea clavata* collected using a manned submersible at 350 m depth of the coast of Curacao[110]. A blue gorgonian, sampled at 350 m depth off Hawaii, lead to the discovery of three novel halogenated azulenes (**98**-**100**)[111]. The specimen also yielded N,N,-dimethylamino-3-guaiszulenylmethane (**101**)[112]. An interesting tricyclic aromatic compound, guaianolide (**102**), was isolated from *Placogorgia sp*., which has strong yellow pigmentation due to its conjugated bonds[113].

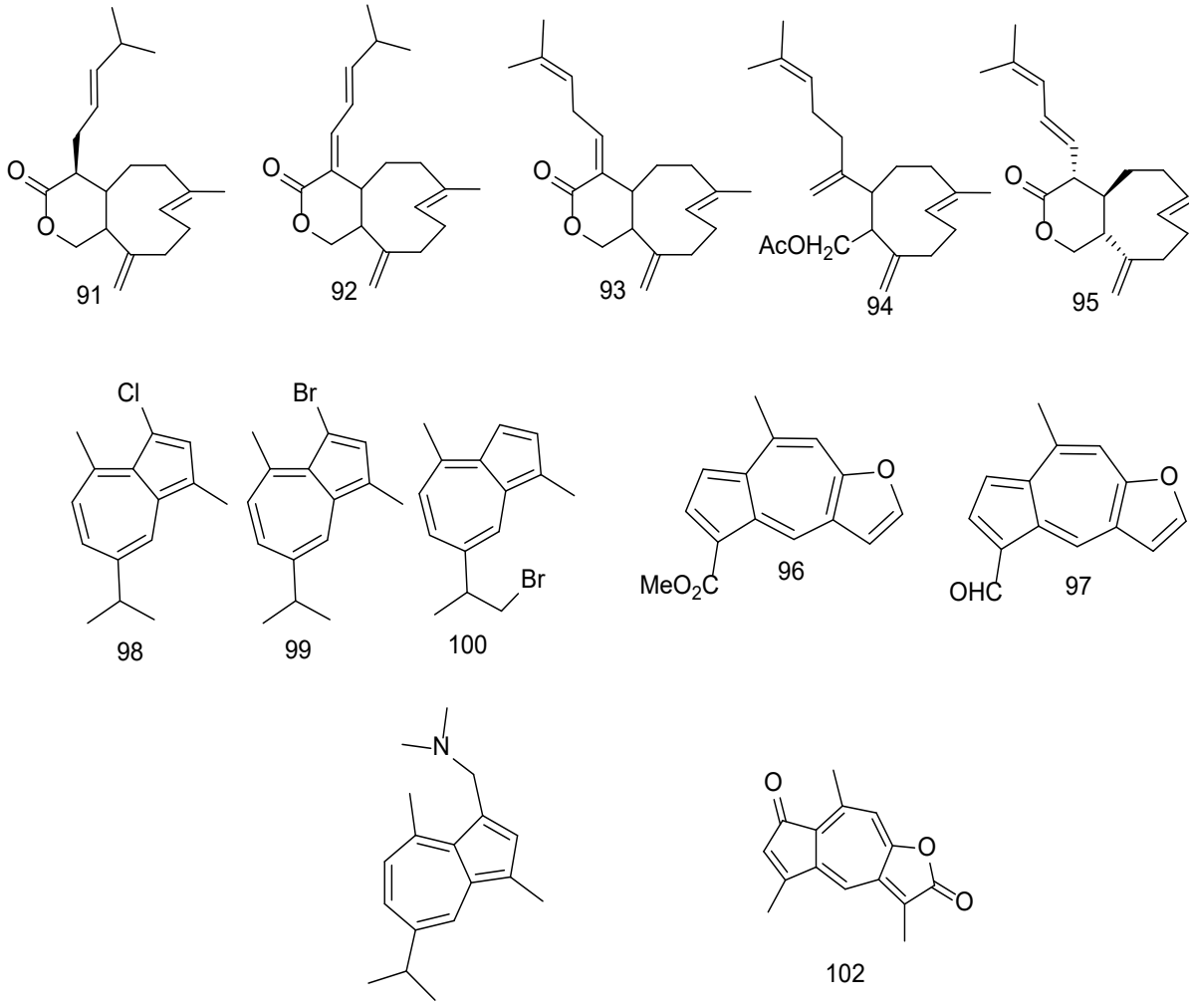


Figure 22. Terpenoids and aromatic metabolites isolated from deep-sea soft corals.

Bamboo corals (Family Keratoisididae)

Bamboo corals are ubiquitous in the deep-sea at both lower and upper bathyal depths. Although these octocorals are relatively common, no natural products to date have been isolated from this family.

Subclass: Hexacorallia

Order Scleractinia

Stony corals (Order Scleractinia) are the main reef-forming corals with over 800 species known[114]. While mainly found in shallow waters, in upper bathyal depths (200-700 m) cold-water coral reefs mainly comprisestony corals, such as *Desmophyllum pertusum[115].* Cold-water coral reefs have a global distribution and have been found as far north as Norway, to as far south as Antarctica. Four novel cholic-acid-type steroids (**103**-**105**) were the first deep-sea compounds isolated from pacific stony coral *Deltocyathus magnificus[116].* A β-glucosylceramide (**106**) with a structurally similar sphinogoid moiety and α-hydroxy fatty acid was isolated from a Mediterranean *Deltocyathus sp.,* and a novel (20R)-22E-cholesta-4-22-diene-3,6-dione(**107**)[117]. *Madrepora oculata* sampled from the southern Indian Ocean, revealed the presence of three novel 10-hydroxydocosapolyenoic acids (**108**-**110**), which also were isolated from a deep-sea coral in Norway[118].

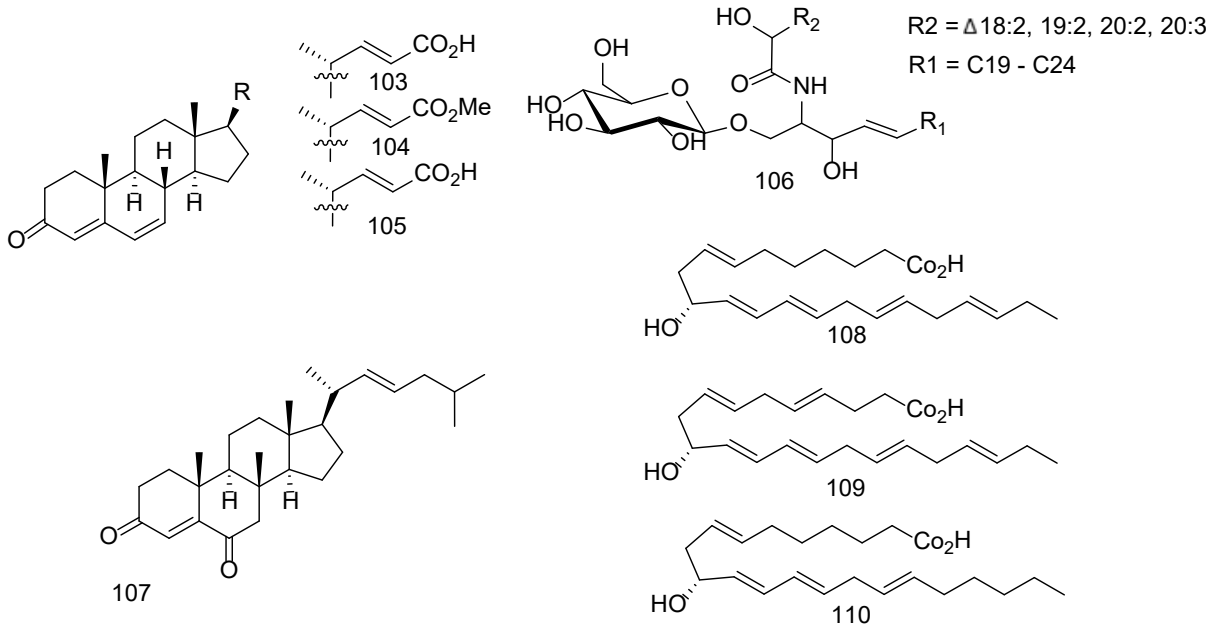


Figure 23. Metabolites isolated from deep-sea stony corals.

Order Zoantharia

Zoantharians have been shown to possess an abundance of bioactive secondary metabolites, especially alkaloids*[119]*. Deep-sea zoantharians remain underexplored with few natural products reported. Two fluorescent tetraazacyclopentazulene derivatives were isolated from *Savalia sp.* collected from 350 m depth of Hawaii, pseudozoanthoxanthin I and II (***111***, ***112***)*[120]*.

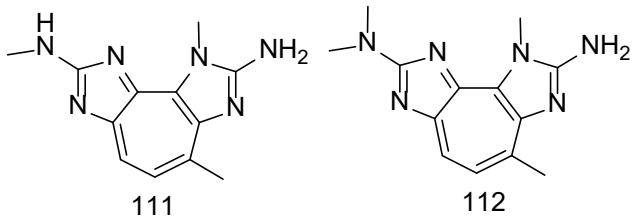


Figure 24. Azulene derivatives isolated from deep-sea zoantharians.

Order: Antipatharia

Black corals are widely distributed in the deep sea, in both the upper and lower bathyal zone. While not intensively studied due to their limited appearances in shallow waters, black corals of the genus *Antipathes* were shown to poses a new carbazole alkaloid, antipathine A (**113**), an antimicrobial sphingolipid (**114**), and a steroid (**115**)[121]. Multiple new sterols have been isolated from this genus including four new polyhydroxylated sterols (**116**, **117**) and five new polar sterols (**118**, **119**)[121-123].

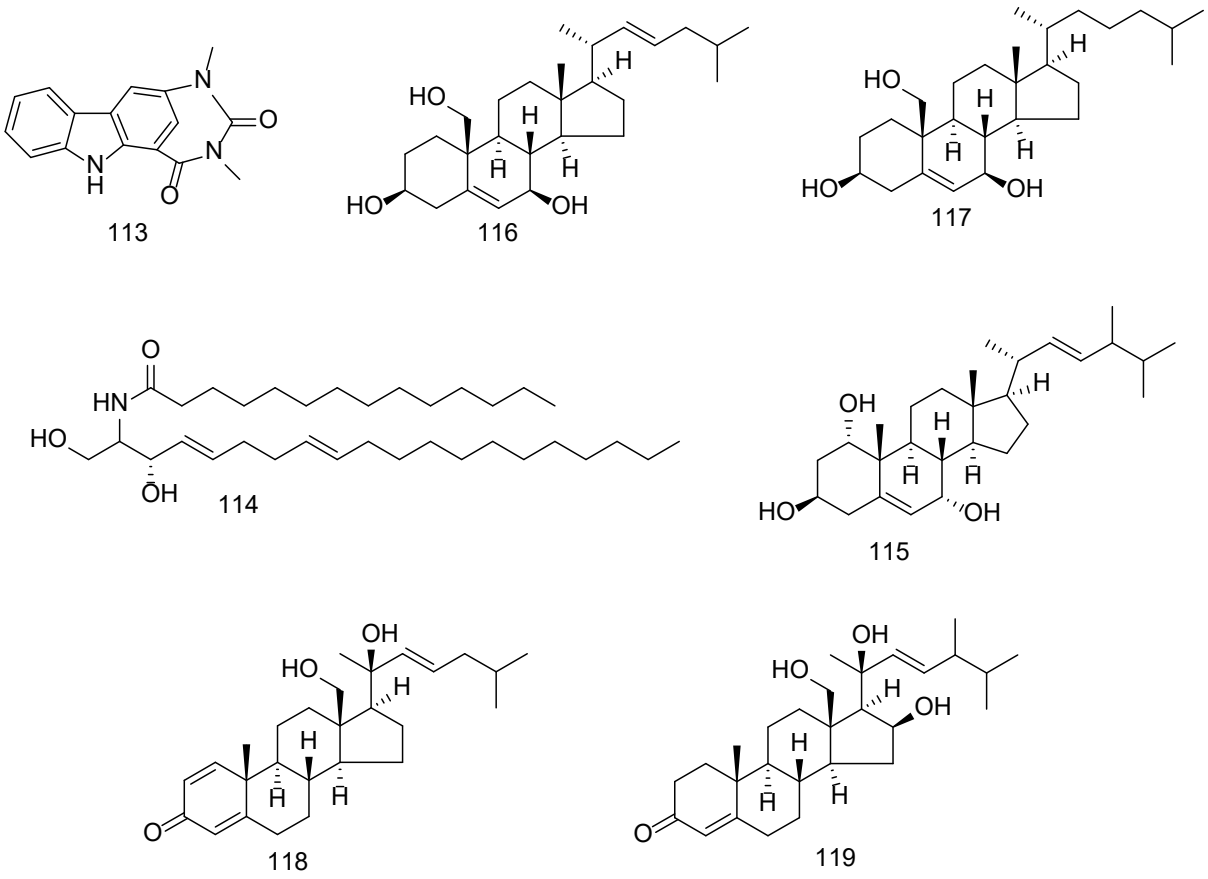


Figure 25. Metabolites and their chemical structures isolated from deep-sea black corals.

1.4 Techniques in natural products chemistry

1.4.1 General Workflow

For the purpose of drug discovery, the isolation and characterisation of natural products generally follow a multistep process of extraction of metabolites from biomass, isolation of metabolites, characterization of the metabolites chemical structure, followed by analysis of biological activities. Over the previous 50 years, researchers have faced many challenges in this process including the diversity of chemical structures in extracts, the low quantities of compounds and the difficulty faced when targeting new or novel metabolites. In this section I will discuss recent improvements to natural product methodology that researchers have developed and are developing, to overcome many of these challenges.

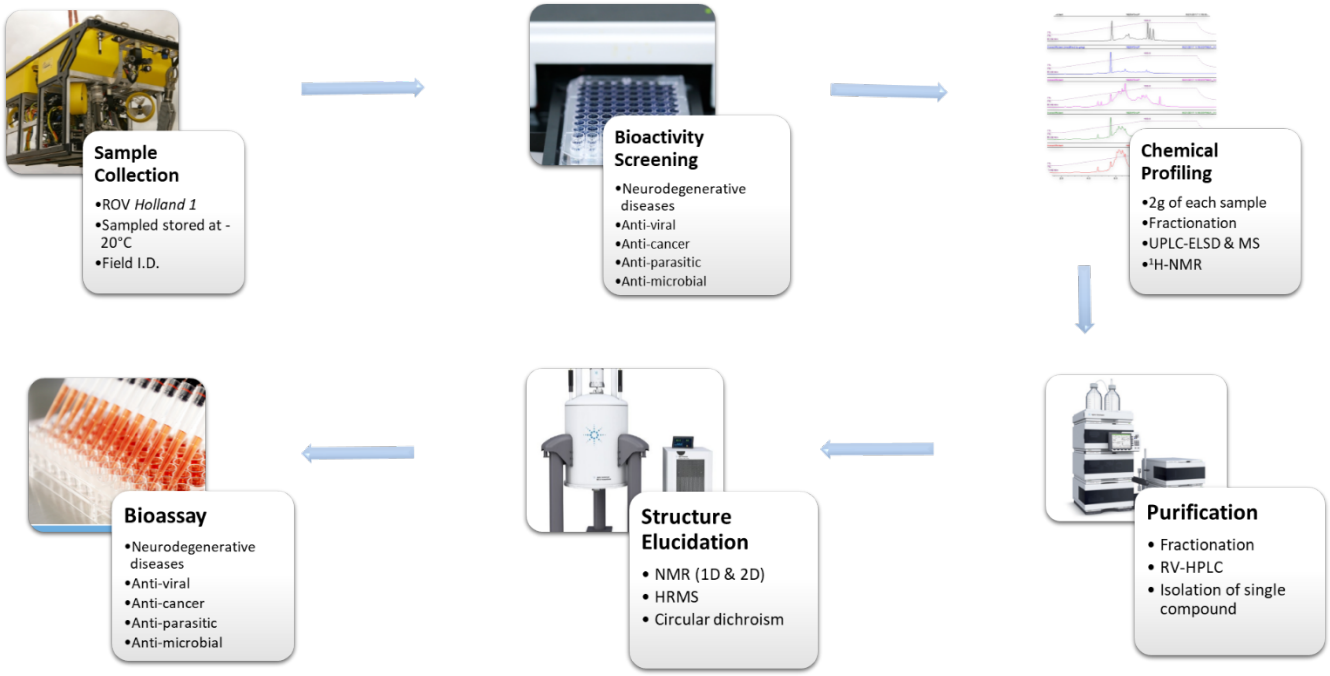


Figure 26. Deep-sea drug discovery workflow starting with sample collection, bioactivates screening, chemical profiling, purification, structure elucidation and biological activity assays.

The mixture of primary and secondary metabolites in biological extracts covers a broad diversity of chemical architectures and concentrations. Due to the complexity of the biosynthetic pathways of many organisms, many metabolites are analogues, varying slightly by stereochemistry, methylation, degrees of oxidation etc. This increases the difficulty in isolating natural products as analogues, especially in the cases of stereoisomers which can have identical retention times across various chromatography techniques. High pressure liquid chromatography (HPLC) has persisted as the most effective way to isolate natural products. Recent advances in reverse phase and normal phase columns, such as chiral columns, have provided researchers with a diverse toolbox to improve natural production separation and isolation.

The quantity of natural products isolated tends to be a limiting factor in terms of discovering and fully characterizing new metabolites. For some specimens, acquiring more biomasses is not an issue (e.g. microbial cultures), however acquiring enough/more biomass from specimens from remote locations can be.

The two main structural elucidation techniques for natural products are 1D and 2D nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry. Mass spectrometry is experimentally sensitive, allowing the analysis of natural products at concentrations in the ppm range, while NMR is information rich, but significantly less sensitive than other techniques. Thus acquiring a usable NMR spectrum can be a bottleneck in the discovery of natural products. Most 500-600 MHz NMRs require approximately 1 mg of pure compound to acquire an adequate data set for structure elucidation. Modification to NMRs including cryoprobes and higher-MHz magnets can dramatically increase their resolution and sensitivity. As many undescribed metabolites are present in minute concentrations as minor metabolites, the increased sensitivity of NMRs can massively increase the number of new compounds isolated.

While NMR can be used to determine the 2D structure of natural products and determine their relative configuration, determining the absolute configuration of natural products is often a challenge. Crystallography is considered the "gold standard" in terms of determining the absolute configuration. While a powerful technique, crystalizing natural products can require relatively large quantities of sample and can be challenging or next to impossible for some chemical classes. Two common techniques, Mosher's method for secondary alcohols and Marfey's for amino acids, are both destructive techniques using derivatisation[124, 125]. This can significantly reduce the quality of sample available for analysis of bioactivities. While NMR based techniques are useful for determining the relative configurations for molecules, the ability to determine the absolute configuration is limited.

Recent advances in computational chemistry and access to supercomputers have enabled the absolute configuration to be determined using circular dichroism. Comparing experimental spectra to the calculated spectra for the enantiomers allows for the non-destructive analysis of absolute stereochemistry. The accuracy of these techniques has been improved dramatically with higher-level algorithms for geometry optimization[126-129]. For natural products with one or more chromophores in their structure, electronic circular dichroism (ECD) is an effective technique which requires relatively small amounts of UV-active natural products. For compounds without chromophores, vibrational circular dichroism (VCD) is a powerful technique, producing spectra with a higher level of information than those produced by ECD. VCD spectrometers are less sensitive than ECD spectrometers, requiring significantly more of a sample to acquire usable spectrum but they may be the only option for determining the absolute configuration of compounds which cannot be crystalized and lack an appropriate chromophore.

Dereplication is the process of identifying known compounds in a sample, be they primary or secondary metabolites. As natural product isolation is labour, time and resource intensive, dereplication is a vital step for improving the efficiency of any natural product workflow. Dereplication is also useful in helping to prioritise samples for in-depth analysis as it can indicate the types of chemistry present in a specimen. The past ten years has seen dramatic improvements to advances in dereplication techniques and strategies. The key to these improvements is natural product databases which incorporate chemical properties (^{1}H-NMR shifts, MS^2 fragmentation spectra etc) and biological information (taxonomy of source species, gene barcoding etc). The incorporation of metadata into these databases can reduce the time required to complete the dereplication process. Recent attempts to improve the dereplication process have aimed to automate it as much as possible.

1.4.2 Molecular Networking (GNPS)

Molecular networking is a tool developed to help visualize the chemical diversity in a particular sample, by analysing XC-MS^2 data. The MS^2 spectra from each parent ion are combined to make a consensus spectrum. These consensus spectra are represented as a node. Molecules which have comparable chemical structures produce similar MS^2 spectra when undergoing collision induced fragmentation (CID) [130-132]. Algorithms used by GNPS, such as MsCluster, compare the MS^2 spectra for each node. Nodes whose spectra contain six fragments in common, or a cosine score >0.7, are connected by an edge. Nodes connected by edges tend to share similar chemical skeletons. Groups of nodes connected by edges are referred to as clusters and can be considered "molecular families"[133].

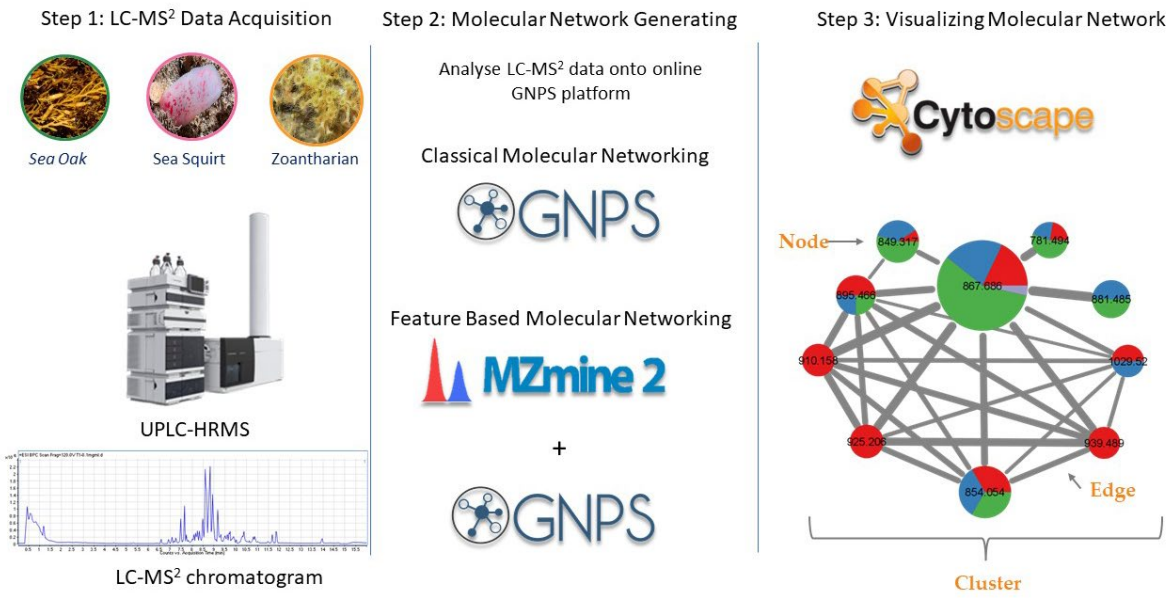


Figure 27. The workflow for generating GNPS-based molecular networking from acquiring LC-MS^2 data from a sample, through generating a network using the online GNPS platform, to visualization of the network using Cytoscape.[132]

Molecular networking was introduced to the natural products field though the online platform Global Natural Products Social Network (GNPS)[134]. This tool allows for a quick and effective way to produce molecular networks from LC-MS^2 data. It also hosts a user-generated open-access natural product library containing the MS^2 spectra for known compounds. This allows users to automatically carry out dereplication on their samples while simultaneously generating a molecular network with known metabolites "annotated" into the network[134]. The GNPS platform is continuously being updated with new techniques such as feature-based and ion-identity molecular networking, which increase the accuracy and reproducibility of this tool[135].

1.4.3 Small Molecule Accurate Recognition Technology (SMART)

Gerwick et al. introduced a tool to hypothesise the structure of small molecules from experimental ^{1}H-^{13}C HSQC spectra[136]. SMART generates these hypnotized structures from an artificial-intelligence tool trained against a library of HSQC spectra from over 100,000 natural products. SMART can be used on both pure compounds and mixtures, allowing it to be integrated into the early steps of a chemical investigation. As SMART is NMR-based, it can highlight unique chemical features, by identifying unique/uncommon chemical shifts. One major advantage of this AI based tool is the rapid improvements that can be expected from increased use and a broadening of its training data set. Users can place these hypothesised structures within the "Moliverse" to aid the visualization of the chemical space. This is a useful tool when targeting unique/novel metabolites and has great potential for metabolic studies and dereplication[137, 138].

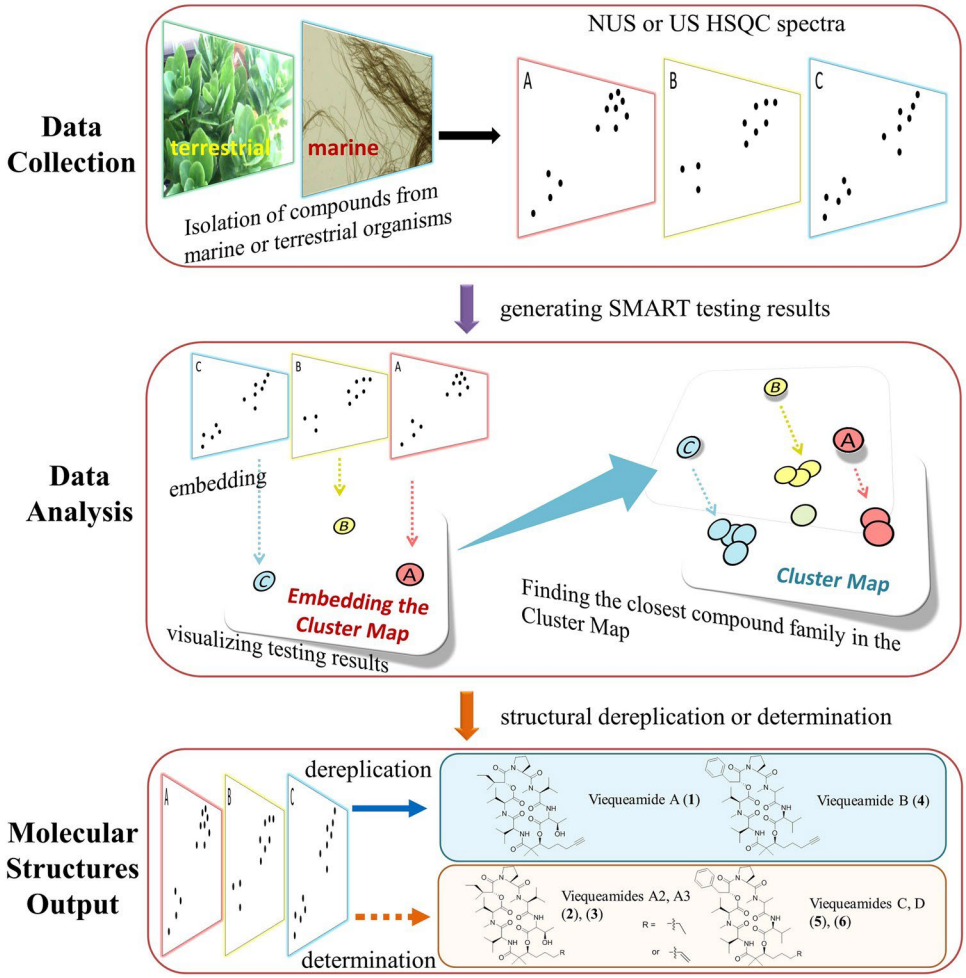


Figure 28. The Small Molecule Accurate Recognition Technology (SMART) workflow for hypothesising natural product chemical structures based on ^{1}H-^{13}C HSQC NMR spectra. HSQC spectra (NUS or US) acquired from a pure compound or mixture of metabolites are analysed and compared to clusters (families) of known compounds. Differences and similarities between the unknown and known spectra are used to predict the structures of the metabolites in the sample.[136]

1.5 Research Project

As part of multidisciplinary drug discovery project, a series of three expeditionary cruises, on the RV Celtic Explorer, collected deep-sea coral and sponges from the Irish Continental Margin. The two main locations for sample collection were Whittard canyon system and the Porcupine Bank. The ROV *Holland I* was used to collect over 800 specimens from the rich benthic ecosystems between 700 m - 3,000 m depth.

The aims of this research thesis are;

1. Isolate and fully characterise new natural products from deep-sea sponges and corals with pharmaceutical potential.
2. Integrate molecular networking into our drug discovery workflow as tool from dereplication and visualization of specimens chemodiversity.

1.6 References

1. Council, N. R., *Understanding marine biodiversity*. National Academies Press: 1995.
2. Chapter 13 Species diversity: land and sea. In *Developments in Palaeontology and Stratigraphy*, Briggs, J. C., Ed. Elsevier: 1995; Vol. 14, pp 371-389.
3. Liang, X.; Luo, D.; Luesch, H., Advances in exploring the therapeutic potential of marine natural products. *Pharmacological research* **2019,** 147, 104373.
4. Gerwick, W. H.; Moore, B. S., Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chemistry & biology* **2012,** 19, (1), 85-98.
5. Grindberg, R. V.; Shuman, C. F.; Sorrels, C. M.; Wingerd, J.; Gerwick, W. H., Neurotoxic alkaloids from cyanobacteria. In *Modern Alkaloids, Structure, Isolation, Synthesis and Biology*, Wiley-VCH Verlang GmbH & Co., Weinheim: 2008; pp 139-170.
6. Goto, T.; Kishi, Y.; Takahashi, S.; Hirata, Y., Tetrodotoxin. *Tetrahedron* **1965,** 21, (8), 2059-2088.
7. Molinski, T. F.; Dalisay, D. S.; Lievens, S. L.; Saludes, J. P., Drug development from marine natural products. *Nature reviews Drug discovery* **2009,** 8, (1), 69-85.
8. Suzuki, M.; Vairappan, C., Halogenated secondary metabolites from Japanese species of the red algal genus Laurencia (Rhodomelaceae, Ceramiales). *Current Topics in Phytochemistry, Volume 7* **2005**, 1-34.
9. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of natural products* **2020,** 83, (3), 770-803.
10. Kong, D.-X.; Jiang, Y.-Y.; Zhang, H.-Y., Marine natural products as sources of novel scaffolds: Achievement and concern. Elsevier Current Trends: 2010; Vol. 15, pp 884-886.
11. MarinLit.
12. Núñez-Pons, L.; Avila, C., Defensive Metabolites from Antarctic Invertebrates: Does Energetic Content Interfere with Feeding Repellence? *Mar. Drugs* **2014,** 12, (6).
13. Thoms, C.; Schupp, P. J.; Custódio, M.; Lôbo-Hajdu, G.; Hajdu, E.; Muricy, G., Chemical defense strategies in sponges: a review. *Porifera research: biodiversity, Innovation and sustainability* **2007,** 28, 627-637.
14. Dicke, M.; Takken, W., *Chemical ecology: from gene to ecosystem*. Springer Science & Business Media: 2006; Vol. 16.
15. Thoms, C.; Ebel, R.; Proksch, P., Activated chemical defense in Aplysina sponges revisited. *Journal of chemical ecology* **2006,** 32, (1), 97-123.
16. Hu, Y.; Chen, J.; Hu, G.; Yu, J.; Zhu, X.; Lin, Y.; Chen, S.; Yuan, J., Statistical research on the bioactivity of new marine natural products discovered during the 28 years from 1985 to 2012. *Mar. Drugs* **2015,** 13, (1), 202-221.
17. Haefner, B., Drugs from the deep: marine natural products as drug candidates. *Drug discovery today* **2003,** 8, (12), 536-544.
18. Hay, M. E., Marine chemical ecology: what's known and what's next? *Journal of Experimental Marine Biology and Ecology* **1996,** 200, (1-2), 103-134.
19. Marine Pharmacology. https://www.marinepharmacology.org (04/04/2022),
20. Skropeta, D.; Wei, L., Recent advances in deep-sea natural products. *Natural product reports* **2014,** 31, (8), 999-1025.
21. Ramirez-Llodra, E.; Brandt, A.; Danovaro, R.; De Mol, B.; Escobar, E.; German, C. R.; Levin, L. A.; Martinez Arbizu, P.; Menot, L.; Buhl-Mortensen, P., Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. *Biogeosciences* **2010,** 7, (9), 2851-2899.
22. Rex, M. A.; Etter, R. J., *Deep-sea biodiversity: pattern and scale*. Harvard University Press: 2010.
23. Anderson, T. R.; Rice, T., Deserts on the sea floor: Edward Forbes and his azoic hypothesis for a lifeless deep ocean. *Endeavour* **2006,** 30, (4), 131-137.
24. Murray, J.; Renard, A. F., *Report on deep-sea deposits based on the specimens collected during the voyage of HMS Challenger in the years 1872 to 1876*. HM Stationery Office: 1891.

25. Omerdic, E.; Toal, D.; Nolan, S.; Ahmad, H., ROV LATIS: next generation smart underwater vehicle. *Further advances in unmanned marine vehicles (IEE Control Eng. Series–Vol. 77)* **2012**, 9-44.
26. Bowen, A. D.; Jakuba, M. V.; Farr, N. E.; Ware, J.; Taylor, C.; Gomez-Ibanez, D.; Machado, C. R.; Pontbriand, C. In *An un-tethered rov for routine access and intervention in the deep sea*, 2013 oceans-san diego, 2013; IEEE: pp 1-7.
27. Watling, L.; Lapointe, A., Global biogeography of the lower bathyal (700–3000 m) as determined from the distributions of cnidarian anthozoans. *Deep Sea Research Part I: Oceanographic Research Papers* **2022**, 103703.
28. Van den Hove, S.; Moreau, V., *Deep-sea Biodiversity and Ecosystems: A Scoping Report on Their Socio-economy, Management and Governanace*. UNEP/Earthprint: 2007.
29. Stuart, C. T.; Rex, M.; Etter, R. J., Large-scale spatial and temporal patterns of deep-sea benthic species diversity. *Ecosystems of the World* **2003**, 295-312.
30. McClain, C. R., The commonness of rarity in a deep-sea taxon. *Oikos* **2021,** 130, (6), 863-878.
31. Baker, M. C.; Ramirez-Llodra, E. Z.; Tyler, P. A.; German, C. R.; Boetius, A.; Cordes, E. E.; Dubilier, N.; Fisher, C. R.; Levin, L. A.; Metaxas, A., Biogeography, ecology, and vulnerability of chemosynthetic ecosystems in the deep sea. *Life in the World's Oceans: Diversity, Distribution, and Abundance, edited by: McIntyre, A* **2010**, 161-183.
32. Chen, C. R.; Makhatadze, G. I., Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nature communications* **2017,** 8, (1), 1-9.
33. Silva, J. L.; Foguel, D.; Royer, C. A., Pressure provides new insights into protein folding, dynamics and structure. *Trends in biochemical sciences* **2001,** 26, (10), 612-618.
34. Fernandez-Arcaya, U.; Ramirez-Llodra, E.; Aguzzi, J.; Allcock, A. L.; Davies, J. S.; Dissanayake, A.; Harris, P.; Howell, K.; Huvenne, V. A.; Macmillan-Lawler, M., Ecological role of submarine canyons and need for canyon conservation: a review. *Frontiers in Marine Science* **2017,** 4, 5.
35. Steffen, K.; Indraningrat, A. A. G.; Erngren, I.; Haglöf, J.; Becking, L. E.; Smidt, H.; Yashayaev, I.; Kenchington, E.; Pettersson, C.; Cárdenas, P., Oceanographic setting influences the prokaryotic community and metabolome in deep-sea sponges. *Scientific reports* **2022,** 12, (1), 1-16.
36. Morris, K. J.; Tyler, P. A.; Masson, D. G.; Huvenne, V. I.; Rogers, A. D., Distribution of cold-water corals in the Whittard Canyon, NE Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography* **2013,** 92, 136-144.
37. Judge, M. T.; McGrath, F.; Cullen, S.; Verbruggen, K. In *INFOMAR, Ireland's National Seabed Mapping Programme; Sharing Valuable Insights*, AGU Fall Meeting Abstracts, 2017; pp OS14B-03.
38. Coveney, S., Integration of INFOMAR Bathymetric LiDAR with external onshore LiDAR datasets. **2009**.
39. Amaro, T.; Huvenne, V.; Allcock, A.; Aslam, T.; Davies, J.; Danovaro, R.; De Stigter, H.; Duineveld, G.; Gambi, C.; Gooday, A., The Whittard Canyon–A case study of submarine canyon processes. *Progress in Oceanography* **2016,** 146, 38-57.
40. Duineveld, G.; Lavaleye, M.; Berghuis, E.; De Wilde, P., Activity and composition of the benthic fauna in the Whittard Canyon and the adjacent continental slope (NE Atlantic). *Oceanologica Acta* **2001,** 24, (1), 69-83.
41. Wright, P. C.; Westacott, R. E.; Burja, A. M., Piezotolerance as a metabolic engineering tool for the biosynthesis of natural products. *Biomolecular engineering* **2003,** 20, (4-6), 325-331.
42. Bartlett, D., Pressure effects on in vivo microbial processes. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2002,** 1595, (1-2), 367-381.
43. Abe, F.; Horikoshi, K., The biotechnological potential of piezophiles. *Trends in biotechnology* **2001,** 19, (3), 102-108.

44. Metz, J. G.; Roessler, P.; Facciotti, D.; Levering, C.; Dittrich, F.; Lassner, M.; Valentine, R.; Lardizabal, K.; Domergue, F.; Yamada, A., Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science* **2001,** 293, (5528), 290-293.
45. Borchert, E.; Jackson, S. A.; O'Gara, F.; Dobson, A. D. W., Diversity of Natural Product Biosynthetic Genes in the Microbiome of the Deep Sea Sponges Inflatella pellicula, Poecillastra compressa, and Stelletta normani. *Frontiers in Microbiology* **2016,** 7.
46. Marchese, P.; Young, R.; O'Connell, E.; Afoullouss, S.; Baker, B. J.; Allcock, A. L.; Barry, F.; Murphy, J. M., Deep-sea coral garden invertebrates and their associated fungi are genetic resources for chronic disease drug discovery. *Mar. Drugs* **2021,** 19, (7), 390.
47. Tortorella, E.; Tedesco, P.; Palma Esposito, F.; January, G. G.; Fani, R.; Jaspars, M.; De Pascale, D., Antibiotics from deep-sea microorganisms: current discoveries and perspectives. *Mar. Drugs* **2018,** 16, (10), 355.
48. Pilkington, L. I., A Chemometric analysis of deep-sea natural products. *Molecules* **2019,** 24, (21), 3942.
49. Schreiber, S. L., The rise of molecular glues. *Cell* **2021,** 184, (1), 3-9.
50. Phillips, E.; Jones, R. L.; Huang, P.; Digklia, A., Efficacy of Eribulin in Soft Tissue Sarcomas. *Frontiers in Pharmacology* **2022,** 13, 869754-869754.
51. Santalova, E. A.; Denisenko, V. A., Fatty acids from a glass sponge Aulosaccus sp. occurrence of new cyclopropane-containing and methyl-branched acids. *Lipids* **2017,** 52, (1), 73-82.
52. Santalova, E. A.; Denisenko, V. A., Steroids from a Far-Eastern glass sponge[space]Aulosaccus[space]sp. *Nat. Prod. Commun.* **2019,** 14, (6), 1-8.
53. Santalova, E. A.; Denisenko, V. A.; Dmitrenok, P. S.; Drozdov, A. L.; Stonik, V. A., Cerebrosides from a Far-Eastern glass sponge Aulosaccus sp. *Lipids* **2015,** 50, (1), 57-69.
54. Santalova, E. A.; Denisenko, V. A.; Dmitrenok, P. S., Structural analysis of oxidized cerebrosides from the extract of deep-sea sponge Aulosaccus sp.: occurrence of amide-linked allylically oxygenated fatty acids. *Molecules* **2020,** 25, (24), 6047.
55. Carbone, M.; Núñez-Pons, L.; Ciavatta, M. L.; Castelluccio, F.; Avila, C.; Gavagnin, M., Occurrence of a taurine derivative in an Antarctic glass sponge. *Nat. Prod. Commun.* **2014,** 9, (4), 469-539.
56. Zhang, H.; Conte, M. M.; Capon, R. J., Franklinolides A-C from an Australian Marine Sponge Complex: Phosphodiesters Strongly Enhance Polyketide Cytotoxicity. *Angew. Chem. Int. Ed.* **2010,** 49, (51), 9904-9906.
57. Gunawardana, G. P.; Koehn, F. E.; Lee, A. Y.; Clardy, J.; He, H. Y.; Faulkner, D. J., Pyridoacridine alkaloids from Pachastrellidae sponges: struc. rev. of dercitin & related comp. & corr. with kuanoniamines. *J. Org. Chem.* **1992,** 57, 1523-1526.
58. Burres, N. S.; Sazesh, S.; Gunawardana, G. P.; Clement, J. J., Antitumor activity and nucleic acid binding properties of dercitin, a new acridine alkaloid isolated from a marine Dercitus species sponge. *Cancer Res.* **1989,** 49, 5267-5274.
59. Rashid, M. A.; Gustafson, K. R.; Crouch, R. C.; Groweiss, A.; Pannell, L. K.; Van, Q. N.; Boyd, M. R., Application of High-Field NMR and Cryogenic Probe Technologies in the Structural Elucidation of Poecillastrin A, a New Antitumor Macrolide Lactam from the Sp. *Org. Lett.* **2002,** 4, (19), 3293-3296.
60. Irie, R.; Hitora, Y.; Ise, Y.; Okada, S.; Takada, K.; Matsunaga, S., Poecillastrin E, F, and G, cytotoxic chondropsin-type macrolides from a marine sponge Poecillastra sp. *Tetrahedron* **2018,** 74, (13), 1430-1434.
61. Suo, R.; Takada, K.; Irie, R.; Watanabe, R.; Suzuki, T.; Ise, Y.; Ohtsuka, S.; Okada, S.; Matsunaga, S., Poecillastrin H, a chondropsin-type macrolide with a conjugated pentaene moiety, from a Characella sp. marine sponge. *J. Nat. Prod.* **2018,** 81, (5), 1295-1299.
62. Takada, K.; Choi, B. W.; Rashid, M. A.; Gamble, W. R.; Cardellina, J. H., Structural Assignment of Poecillastrins B and C, Macrolide Lactams from the Deep-Water Caribbean Sponge Poecillastra Species. *J. Nat. Prod.* **2007,** 70, (3), 428-431.

63. TAKEMOTO, D.; TAKEKAWA, Y.; SOEST, R. W. M. v.; FUSETANI, N.; MATSUNAGA, S., Poecillastrin D: A New Cytotoxin of the Chondropsin Class from Marine Sponge Jaspis serpentina. *Bioscience, Biotechnology, and Biochemistry* **2007,** 71, (11), 2697-2700.
64. Structure characterization by two-dimensional NMR spectroscopy, of two marine triterpene oligoglycosides from a Pacific sponge of the genus Erylus. *Tetrahedron* **1992,** 48, 491-498.
65. Zhang, W.; Che, C., Isomalabaricane-type nortriterpenoids and other constituents of the marine sponge Geodia japonica. *J. Nat. Prod.* **2001,** 64, 1489-1492.
66. Carstens, B. B.; Rosengren, K. J.; Gunasekera, S.; Schempp, S.; Bohlin, L.; Dahlstrom, M.; Clark, R. J.; Goransson, U., Isolation, characterization, and synthesis of the barrettides: disulfide-containing peptides from the marine sponge Geodia barretti. *Journal of Natural Products* **2015,** 78, (8), 1886-1893.
67. Steffen, K.; Laborde, Q.; Gunasekera, S.; Payne, C. D.; Rosengren, K. J.; Riesgo, A.; Goransson, U.; Cárdenas, P., Barrettides: A Peptide Family Specifically Produced by the Deep-Sea Sponge Geodia barretti. *Journal of natural products* **2021,** 84, (12), 3138-3146.
68. Capon, R. J.; Skene, C.; Lacey, E.; Gill, J. H.; Wadsworth, D.; Friedel, T., Geodin A magnesium salt: a novel nematocide from a southern Australian marine sponge, Geodia. *J. Nat. Prod.* **1999,** 62, 1256-1259.
69. Killday, K. B.; Longley, R.; McCarthy, P. J.; Pomponi, S. A.; Wright, A. E.; Neale, R. F.; Sills, M. A., Sesquiterpene-derived metabolites from the deep water marine sponge Poecillastra sollasi. *J. Nat. Prod.* **1993,** 56, 500-507.
70. Phuwapraisirisan, P.; Matsunaga, S., Shinsonefuran, a cytotoxic furanosesterterpene with a novel carbon skeleton, from the deep-sea sponge Stoeba extensa. *Tetrahedron Lett.* **2004,** 45, (10), 2125-2128.
71. Paterson, I.; Dalby, S. M.; Roberts, J. C.; Naylor, G. J.; Guzman, E. A.; Isbrucker, R.; Pitts, T. P.; Linley, P.; Divlianska, D.; Reed, J. K.; Wright, A. E., Leiodermatolide, a Potent Antimitotic Macrolide from the Marine Sponge Leiodermatium sp. *Angew. Chem. Int. Ed.* **2011,** 50, (14), 3219-3223.
72. Wright, A. E.; Roberts, J. C.; Guzmán, E. A.; Pitts, T. P.; Pomponi, S. A.; Reed, J. K., Analogues of the potent antitumor compound leiodermatolide from a deep-water sponge of the genus Leiodermatium. *J. Nat. Prod.* **2017,** 80, (3), 735-739.
73. 161. Corallistin A, a second example of a free porphyrin from a living organism. Isolation from the Demosponge Corallistes sp. of the Coral Sea and inhibition. *Helv. Chim. Acta* **1989,** 72, 1451-1454.
74. Debitus, C.; Cesario, M.; Guilhem, J.; Pascard, C.; Pais, M., Corallistine, a new polynitrogen compound from the marine sponge Corallistes fluvodesmus. *Tetrahedron Lett.* **1989,** 30, 1535-1538.
75. Gulavita, N. K.; Gunasekera, S. P.; Pomponi, S. A.; Robinson, E. V., Polydiscamide A: a new bioactive depsipeptide from the marine sponge Discodermia sp. *J. Org. Chem.* **1992,** 57, 1767-1772.
76. Sandler, J. S.; Colin, P. L.; Kelly, M.; Fenical, W., Cytotoxic macrolides from a new species of the deep-water marine sponge Leiodermatium. *J. Org. Chem.* **2006,** 71, (19), 7245-7251.
77. Bewley, C. A.; Debitus, C.; Faulkner, D. J., Microsclerodermins A and B. Antifungal cyclic peptides from the lithistid sponge Microscleroderma sp. *J. Am. Chem. Soc.* **1994,** 116, 7631-7636.
78. Urban, S.; Butler, M. S.; Capon, R. J., Lamellarins O and P: new aromatic metabolites from the Australian marine sponge Dendrilla cactos. *Aust. J. Chem.* **1994,** 47, 1919-1924.
79. Capon, R. J.; Ovenden, S. P.; Dargaville, T., cis-3-hydroxy-N-methyl-L-proline: a new amino acid from a southern Australian marine sponge, Dendrilla sp. *Australian journal of chemistry* **1998,** 51, (2), 169-170.

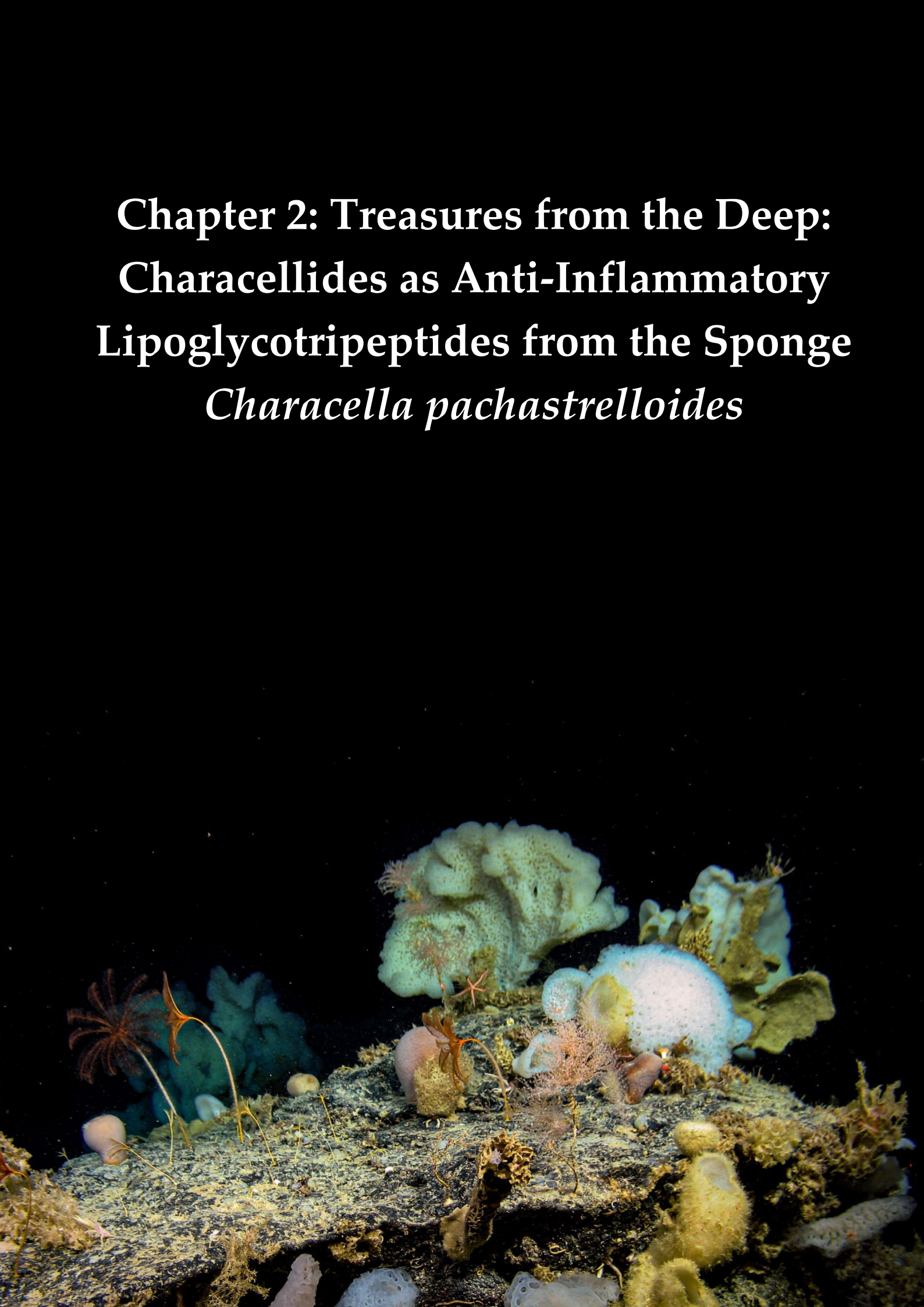
80. Urban, S.; Capon, R. J., Deoxyspongiaquinones: New sesquiterpene quinones and hydroquinones from a southern Australian marine sponge Euryspongia sp. *Australian journal of chemistry* **1996,** 49, (5), 611-615.
81. Salim, A. A.; Rae, J.; Fontaine, F.; Conte, M. M.; Khalil, Z.; Martin, S.; Parton, R. G.; Capon, R. J., Heterofibrins: inhibitors of lipid droplet formation from a deep-water southern Australian marine sponge, Spongia (Heterofibria) sp. *Org. Biomol. Chem.* **2010,** 8, (14), 3188-3194.
82. Zhang, H.; Khalil, Z. G.; Capon, R. J., Fascioquinols A-F: bioactive meroterpenes from a deep-water southern Australian marine sponge, Fasciospongia sp. *Tetrahedron* **2011,** 67, (14), 2591-2595.
83. Venkateswarlu, Y.; Reddy, M. V. R., Three new heptaprenylhydroquinone derivatives from the sponge Ircinia fasciculata. *J. Nat. Prod.* **1994,** 57, 1286-1289.
84. Butler, M. S.; Capon, R. J., The luffarins (A-Z), novel terpenes from an Australian marine sponge, Luffariella geometrica. *Aust. J. Chem.* **1992,** 45, 1705-1743.
85. Zhang, H.; Khalil, Z.; Conte, M. M.; Plisson, F.; Capon, R. J., A search for kinase inhibitors and antibacterial agents: bromopyrrolo-2-aminoimidazoles from a deep-water Great Australian Bight sponge Axinella sp. *Tetrahedron Lett.* **2012,** 53, (29), 3784-3787.
86. Sivappa, R.; Hernandez, N. M.; He, Y.; Lovely, C. J., Studies toward the total synthesis of axinellamine and massadine. *Organic Letters* **2007,** 9, (20), 3861-3864.
87. Kohmoto, S.; Kashman, Y.; McConnell, O. J.; Rinehart, K. L.; Wright, A.; Koehn, F., Dragmacidin, a cytotoxic bis(indole)alkaloid from deep water marine sponge, Dragmacidon sp. *J. Org. Chem.* **1988,** 53, 3116-3118.
88. Snider, B. B.; Yang, K., Stereo- and enantiospecific syntheses of (-)-reiswigins A and B. Assignment of absolute and relative configuration. *J. Org. Chem.* **1990,** 55, 4392-4399.
89. Kashman, Y.; Hirsch, S.; Koehn, F.; Cross, S., Reiswigins A & B, antiviral diterpenes from sponge Epipolasis reiswigi. *Tetrahedron Lett.* **1987,** 28, 5461-5464.
90. Tsujii, S.; Rinehart, K. L.; Gunasekera, S. P.; Kashman, Y.; Cross, S. S.; Lui, M. S.; Pomponi, S. A.; Diaz, M. C., Topsentin, bromotopsentin, and dihydroxybromotopsentin: antiviral and antitumor bis(indolyl)imidazoles from Caribbean deep-sea sponges of the family Halichon. *J. Org. Chem.* **1988,** 53, 5446-5453.
91. Winder, P. L.; Baker, H. L.; Linley, P.; Guzman, E. A.; Pomponi, S. A.; Diaz, M. C.; Reed, J. K.; Wright, A. E., Neopetrosiquinones A and B, sesquiterpene benzoquinones isolated from the deep-water sponge Neopetrosia cf. proxima. *Bioorg. Med. Chem.* **2011,** 19, (22), 6599-6603.
92. Hitora, Y.; Takada, K.; Okada, S.; Ise, Y.; Matsunaga, S., (-)-Duryne and Its Homologues, Cytotoxic Acetylenes from a Marine Sponge Petrosia sp. *J. Nat. Prod.* **2011,** 74, (5), 1262-1267.
93. Wright, A. E.; McConnell, O. J.; Kohmoto, S.; Lui, M. S.; Thompson, W.; Snader, K. M., Duryne, a new cytotoxic agent from the marine sponge Cribrochalina dura. *Tetrahedron Lett.* **1987,** 28, 1377-1380.
94. Gung, B. W.; Omollo, A. O., Total synthesis of (+)- and (-)-duryne: A potent anticancer agent from the marine sponge Cribrochalina dura. Establishment of the central double bond geometry and the absolute configuration of the. *J. Org. Chem.* **2008,** 73, (3), 1067-1070.
95. Desoubzdanne, D.; Marcourt, L.; Raux, R.; Chevalley, S.; Dorin, D.; Doerig, C.; Valentin, A.; Ausseil, F.; Debitus, C., Alisiaquinones and Alisiaquinol, Dual Inhibitors of Plasmodium falciparum Enzyme Targets from a New Caledonian Deep Water Sponge. *J. Nat. Prod.* **2008,** 71, (7), 1189-1192.
96. Barrow, R. A.; Murray, L. M.; Lim, T. K.; Capon, R. J., Mirabilins (A-F): new alkaloids from a southern Australian marine sponge, Arenochalina mirabilis. *Aust. J. Chem.* **1996,** 49, 767-773.
97. Ford, J.; Capon, R. J., Discorhabdin R: a new antibacterial pyrroloiminoquinone from two Latrunculiid marine sponges, Latrunculia sp. and Negombata sp. *J. Nat. Prod.* **2000,** 63, 1527-1528.

98. Litaudon, M.; Hart, J. B.; Blunt, J. W.; Lake, R. J.; Munro, M. G. H., Isohomohalichondrin B, a new antitumor polyether macrolide from the New Zealand deep-water sponge Lissodendoryx sp. *Tetrahedron Lett.* **1994,** 35, 9435-9438.
99. Bergamaschi, D.; Ronzoni, S.; Taverna, S.; Faretta, M., Cell cycle perturbations and apoptosis induced by isohomohalichondrin B (IHB), a natural marine compound CA130:291176. *Br. J. Cancer* **1999,** 79, 267-277.
100. Garcia-Rocha, M.; Garcia-Gravalos, M. D.; Avila, J., Characterisation of antimitotic products from marine organisms that disorganise the microtuble network:ecteinascidin 743, isohomohalichondrin B and LL-15. *Br. J. Cancer* **1996,** 73, 875-883.
101. Litaudon, M.; Hickford, S. J. H.; Lill, R. E.; Lake, R. J.; Blunt, J. W.; Munro, M. H. G., Antitumor polyether macrolides: new and hemisynthetic halichondrins from the New Zealand deep-water sponge Lissodendoryx sp. *J. Org. Chem.* **1997,** 62, 1868-1871.
102. Wright, A. D.; Schupp, P. J.; Schror, J. P.; Engemann, A.; Rohde, S.; Kelman, D., Twilight Zone Sponges from Guam Yield Theonellin Isocyanate and Psammaplysins I and J. *J. Nat. Prod.* **2012,** 75, (3), 502-506.
103. Williams, G. C., The Global Diversity of Sea Pens (Cnidaria: Octocorallia: Pennatulacea). *PLOS ONE* **2011,** 6, (7), e22747.
104. Thomas, S. A. L.; Sanchez, A.; Kee, Y.; Wilson, N. G.; Baker, B. J., Bathyptilones: Terpenoids from an Antarctic Sea Pen, Anthoptilum grandiflorum (Verrill, 1879). *Mar. Drugs* **2019,** 17, (9), 513.
105. Gao, C. H.; Wang, Y. F.; Li, S.; Qian, P. Y.; Qi, S. H., Alkaloids and Sesquiterpenes from the South China Sea Gorgonian Echinogorgia pseudossapo. *Mar. Drugs* **2011,** 9, 2479-2487.
106. Bonini, C.; Kinnel, R. B.; Li, M.; Scheuer, P. J.; Djerassi, C., Minor and trace sterols in marine invertebrates 38: Isolation, structure elucidation and partial synthesis of papakusterol, a new biosynthetically unusual. *Tetrahedron Lett.* **1983,** 24, 277-280.
107. Paesslerins A and B: novel tricyclic sesquiterpenoids from the soft coral Alcyonium paessleri. *Org. Lett.* **2001,** 3, (10), 1415-1417.
108. Schwartz, R. E.; Scheuer, P. J.; Zabel, V.; Watson, W. H., The coraxeniolides, constituents of pink coral, Corallium sp. *Tetrahedron* **1981,** 37, 2725-2733.
109. Arboxeniolide-1, a new, naturally occurring xeniolide diterpenoid from the gorgonian Paragorgia arborea of the Crozet Is. (S. Indian Ocean). *Z. Naturforsch.* **1984,** 39, 1180-1183.
110. Imre, S.; Thomson, R. H.; Yalhi, B., Linderazulene; a new naturally occurring pigment from the gorgonian Paramuricea chamaeleon. *Experientia* **1981,** 37, 442-443.
111. Li, M. K. W.; Scheuer, P. J., Halogenated blue pigments of a deep sea gorgonian ; chloro- and bromoazulenes. *Tetrahedron Lett.* **1984,** 25, 587-590.
112. Li, M. K. W.; Scheuer, P. J., N;N-dimethylamino-3-guaiazulenylmethane from a deep sea gorgonian. *Tetrahedron Lett.* **1984,** 25, 4707-4708.
113. Li, M. K. W.; Scheuer, P. J., A guaianolide pigment from a deep sea Gorgonian. *Tetrahedron Lett.* **1984,** 25, 2109-2110.
114. Cairns, S. D., Deep-water corals: an overview with special reference to diversity and distribution of deep-water scleractinian corals. *Bulletin of marine Science* **2007,** 81, (3), 311-322.
115. Kiriakoulakis, K.; Fisher, E.; Wolff, G. A.; Freiwald, A.; Grehan, A.; Roberts, J. M., Lipids and nitrogen isotopes of two deep-water corals from the North-East Atlantic: initial results and implications for their nutrition. In *Cold-water corals and ecosystems*, Springer: 2005; pp 715-729.
116. Guerriero, A.; Dambrosio, M.; Zibrowius, H.; Pietra, F., Novel cholic-acid-type sterones of Deltocyathus magnificus, a deep-water scleractinian coral from the Loyalty Islands, SW Pacific. *Helv. Chim. Acta* **1996,** 79, 982-988.
117. Kontiza, I.; Abatis, D.; Malakate, K.; Vagias, C.; Roussis, V., 3-Keto steroids from the marine organisms Dendrophyllia cornigera and Cymodocea nodosa. *Steroids* **2006,** 71, (2), 177-181.

118. Mancini, I.; Guerriero, A.; Guella, G.; Bakken, T.; Zibrowius, H.; Pietra, F., Novel 10-hydroxydocosapolyenoic acids from deep-water scleractinian corals. *Helv. Chim. Acta* **1999,** 82, 677-684.
119. Guillen, P. O.; Jaramillo, K. B.; Genta-Jouve, G.; Thomas, O. P., Marine natural products from zoantharians: bioactivity, biosynthesis, systematics, and ecological roles. *Natural Product Reports* **2020,** 37, (4), 515-540.
120. Komoda, Y.; Shimizu, M.; Ishikawa, M., Structures of biologically active minor bases related to paragracine from Parazoanthus gracilis Lwowsky. *Chem. Pharm. Bull.* **1984,** 32, 3873-3879.
121. Qi, S. H.; Su, G. C.; Wang, Y. F.; Liu, Q. Y.; Gao, C. H., Alkaloids from the South China Sea Black Coral Antipathes dichotoma. *Chem. Pharm. Bull.* **2009,** 57, (1), 87-88.
122. Aiello, A.; Fattorusso, E.; Menna, M., Four new bioactive polyhydroxylated sterols from the black coral Antipathes subpinnata. *J. Nat. Prod.* **1992,** 55, 321-325.
123. Al-Lihaibi, S. S.; Ayyad, S. E. N.; Shaher, F.; Alarif, W. M., Antibacterial Sphingolipid and Steroids from the Black Coral Antipathes dichotoma. *Chem. Pharm. Bull.* **2010,** 58, (12), 1635-1638.
124. Ohtani, I.; Kusumi, T.; Kashman, Y.; Kakisawa, H., High-field FT NMR application of Mosher's method. The absolute configurations of marine terpenoids. *Journal of the American Chemical Society* **1991,** 113, (11), 4092-4096.
125. Bhushan, R.; Brückner, H., Marfey's reagent for chiral amino acid analysis: a review. *Amino acids* **2004,** 27, (3), 231-247.
126. Marcarino, M. O.; Cicetti, S.; Zanardi, M. M.; Sarotti, A. M., A critical review on the use of DP4+ in the structural elucidation of natural products: the good, the bad and the ugly. A practical guide. *Natural Product Reports* **2021**.
127. Sarotti, A. M., In silico reassignment of (+)-diplopyrone by NMR calculations: use of a DP4/J-DP4/DP4+/DIP tandem to revise both relative and absolute configuration. *The Journal of Organic Chemistry* **2020,** 85, (17), 11566-11570.
128. Orio, M.; Pantazis, D. A.; Neese, F., Density functional theory. *Photosynthesis research* **2009,** 102, (2), 443-453.
129. Huh, D. S.; Choe, S. J., Comparative DFT study for molecular geometries and spectra of methyl pheophorbides-a: test of M06-2X and two other functionals. *Journal of Porphyrins and Phthalocyanines* **2010,** 14, (07), 592-604.
130. Xu, R.; Lee, J.; Chen, L.; Zhu, J., Enhanced detection and annotation of small molecules in metabolomics using molecular-network-oriented parameter optimization. *Molecular Omics* **2021,** 17, (5), 665-676.
131. Olivon, F.; Roussi, F.; Litaudon, M.; Touboul, D., Optimized experimental workflow for tandem mass spectrometry molecular networking in metabolomics. *Analytical and bioanalytical chemistry* **2017,** 409, (24), 5767-5778.
132. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A., Molecular networking as a dereplication strategy. *Journal of natural products* **2013,** 76, (9), 1686-1699.
133. Perez De Souza, L.; Alseekh, S.; Brotman, Y.; Fernie, A. R., Network-based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation. *Expert Review of Proteomics* **2020,** 17, (4), 243-255.
134. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature biotechnology* **2016,** 34, (8), 828-837.
135. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M., Feature-based molecular networking in the GNPS analysis environment. *Nature methods* **2020,** 17, (9), 905-908.

136. Zhang, C.; Idelbayev, Y.; Roberts, N.; Tao, Y.; Nannapaneni, Y.; Duggan, B. M.; Min, J.; Lin, E. C.; Gerwick, E. C.; Cottrell, G. W., Small molecule accurate recognition technology (SMART) to enhance natural products research. *Scientific reports* **2017,** 7, (1), 1-17.
137. McAlpine, J. B.; Chen, S.-N.; Kutateladze, A.; MacMillan, J. B.; Appendino, G.; Barison, A.; Beniddir, M. A.; Biavatti, M. W.; Bluml, S.; Boufridi, A., The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Natural product reports* **2019,** 36, (1), 35-107.
138. Reher, R.; Kim, H. W.; Zhang, C.; Mao, H. H.; Wang, M.; Nothias, L.-F.; Caraballo-Rodriguez, A. M.; Glukhov, E.; Teke, B.; Leao, T., A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. *Journal of the American Chemical Society* **2020,** 142, (9), 4114-4120.

Chapter 2: Treasures from the Deep: Characellides as Anti-Inflammatory Lipoglycotripeptides from the Sponge *Characella pachastrelloides*



Chapter 2. Treasures from the Deep: Characellides as Anti-Inflammatory Lipoglycotripeptides from the Sponge *Characella pachastrelloides*

Published: 20th December 2018 *Organic Letters.* 2019, 21, 1, 246–251

DOI: 10.1021/acs.orglett.8b03684

2.1 Abstract

The chemical investigation of marine invertebrates from the deep Northeastern Atlantic revealed new lipoglycotripeptides named characellides isolated from the tetractinellid sponge Characella pachastrelloides. This new family of natural products features a central tripeptide linked to a rare sugar unit and a long alkyl chain ending with a 2,3- dimethyltetrahydropyran. The configurations of all 13 chiral centers were determined by extensive use of NMR data and circular dichroism spectra combined with calculations.

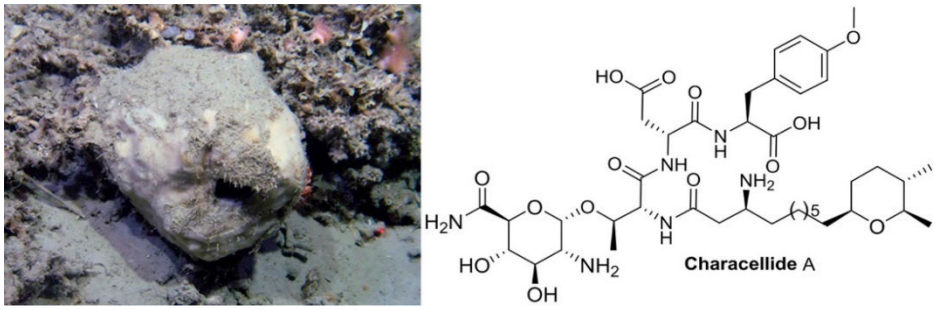


Table 1. Publication authors and their contributions.

Authors	Sample Identification	Isolation	Characterization	Bioactivity	Original Draft	Review and Editing	Resources	Supervision
Sam Afoullouss		X	X		X			
Kevin Calabro		X	X		X	X		
Gregory Genta-Jouve			X		X	X	X	
Sandra Gegunde				X	X			
Amparo Alfonso				X				
Robert Nesbitt		X						
Christine Morrow	X		X			X		
Eva Alonso				X				
Luis M. Botana				X	X	X	X	X
A. Louise Allcock						X	X	X
Olivier P. Thomas			X		X	X	X	X

2.2 Introduction

Among the new horizons for marine biodiscovery, the deep ocean has recently attracted much attention. Technological advances exemplified by the development of underwater remote operated vehicles (ROVs) have enabled in situ photography and careful collection of specimens from depths that were previously inaccessible, except by dredging. In recent years, natural product chemists have been turning to the deep sea where the physio-chemical conditions are extreme in comparison with other ecosystems with the aim of discovering new chemical structures[1]. A rapid review of the literature on deep-sea natural products reveals a particular focus on microorganisms like bacteria and fungi that have already led to new chemical architectures[2]. However, chemical studies on marine invertebrates are still scarce, mainly because of the difficulties in collection and identification of the samples, compounded by the scarcity of taxonomists, and the high percentage of novel species. Several worldwide programs have recently been funded to fill the gaps in our knowledge of the marine biodiversity present below 500 m. We undertook the first chemical screening of sponges collected from deep water off the southwest coast of Ireland. In order to make best use of the samples and within the context of sustainable exploitation, we did not choose a bioguided approach, but rather we aimed at a broad description of the metabolome of each collected specimen. A UPLC-DAD-MS/MS chemical screening prioritized the marine sponge Characella pachastrelloides as a rich source of chemical diversity[3]. Molecular networking showed a high level of novelty among the different clusters present. Interestingly, a large diversity of poecillastrins recently identified in a species of this genus were uncovered[4]. Nevertheless, the presence of other clusters of major metabolites with high molecular weights led us to undertake the first chemical study of this species. We report herein the isolation and structure elucidation of four new metabolites named characellides A–D (**1**–**4**) (Figure 1). Their unique structures include three moieties: a core tripeptide (*O*-Me-Tyr-Asp-Thr) and unprecedented long alkyl chains and sugar units linked on the terminal threonine. The chemical structures were first elucidated after extensive NMR spectroscopic analyses (1D and 2D NMR) and then confirmed by MS/MS analysis. We performed a full assessment of the configurations on **1** and tested their anti-inflammatory effects.

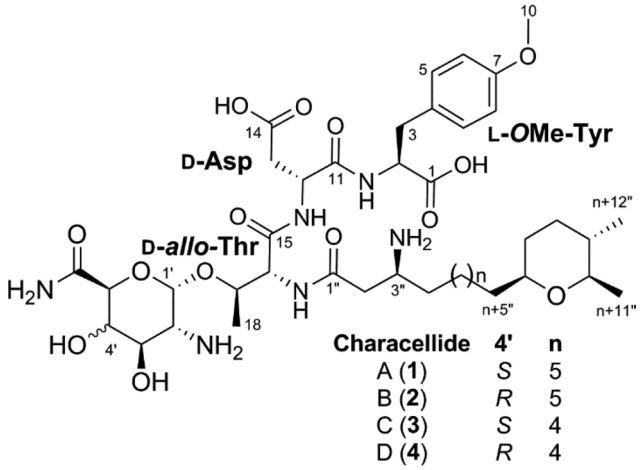


Figure 1. Chemical structures of characellides A-D (**1**-**4**).

2.3 Results and Discussion

The methanolic was selected for purification due to the presence of strong UV and ESLD signals after chemical profiling. Our first purification attempts led to the isolation of a mixture of two isomers difficult to separate. The use of an HILIC amide column finally allowed a straightforward separation of

the two epimers **1** and **2**. Compound **1** was isolated as a colorless oil, and its molecular formula was determined as $C_{41}H_{66}N_6O_{14}$ based on the ion peak at [M + H]+ 867.4726. The ^{1}H-NMR spectrum of **1** disclosed some characteristic signals of three moieties: a para-substituted phenolic part reminiscent of a tyrosine amino acid was evidenced by the signals at δ_H 7.13 (d, J = 8.4 Hz, 2H, H-5/9) and 6.83 (d, J = 8.4 Hz, 2H, H-6/8); oxygenated methines of a sugar unit with a characteristic signal at δ_H 5.32 (d, J = 3.7 Hz, 1H, H-1′) for the anomeric proton; and a series of aliphatic methylenes at δ_H 1.30 (m) corresponding to the lipophilic part of the molecule (Table 1) The identities of the three amino acids were deduced from interpretation of COSY, HSQC, and HMBC correlations. The methoxy group on the phenol of the tyrosine was inferred from the signals at δ_H 3.75 (s, 3H, H-10) and δ_C 55.7 (CH3, C-10) and the key H-10/C-7 HMBC correlation. At this point, we were not able to discriminate between an aspartic acid and an asparagine due to the uncertainty around the signal at δ_C 174.8 (C, C-14). The signals at δ_H 1.31 (d, J = 6.4 Hz, 3H, H-18) and 4.20 (quint, J = 6.4 Hz, 1H, H-17) were key features for the identification of the threonine. Additional H-2/C-11 and H-12/C-15 HMBC correlations allowed the connection between the three amino acids. A sugar residue was placed on the secondary alcohol of the threonine after interpretation of a key H-1′/C-17 HMBC correlation starting from the signal of the anomeric proton H-1′. The shielded value of the chemical shift corresponding to C-2′ at δ_C 55.3 (CH, C-2′) was consistent with a primary amine at this position, while the signal assigned to a carbonyl group at δ_C 174.1 (C, C-6′) left some ambiguity between a carboxylic acid or a primary amide at this position. A clear H-16/C-1′′ HMBC correlation revealed the presence of a carbonyl group substituted on the amino end of the threonine. The location of another primary amine at position C-3′′ was further evidenced through COSY and HSQC correlations with characteristic signals at δ_H 3.54 (m, 1H, H-3′′) and δ_C 50.1 (CH, C-3′′). This side chain extended through a set of methylene units, but the absence of a terminal methyl triplet prompted a thorough inspection of the lipophilic ending part. Two methyls associated with the signals at δ_H 1.15 (d, J = 6.2 Hz, 3H, H-16′′) and δ_H 0.83 (d, J = 5.9 Hz, 3H, H-17′′) were readily placed on vicinal carbons after 2D NMR inspection. Unexpectedly, a key H-15′′/C-11′′ HMBC correlation depicted a tetrahydropyran ring substituted at positions C14′′ and C-15′′ by the two methyls and at position C-11′′ by the long methylene chain. To the best of our knowledge, this is the first occurrence of such a terminal unit for a long alkyl chain, and this moiety raised the question of its metabolic origin. The molecular formula confirmed the presence of both one carboxylic acid and one primary amide placed at position C-14 or C-6′. The location of the amide at C-6′ and the carboxylic acid at C-14 were unequivocally deduced from two fragments in the HRMS/MS spectrum of **1**: the first one at m/z 175.0700 (Aα), related to the sugar unit containing two nitrogens, and the second one at *m/z* 311.1244 (Y_2) in agreement with an aspartic acid as the second amino acid of the peptide (see figure 2).

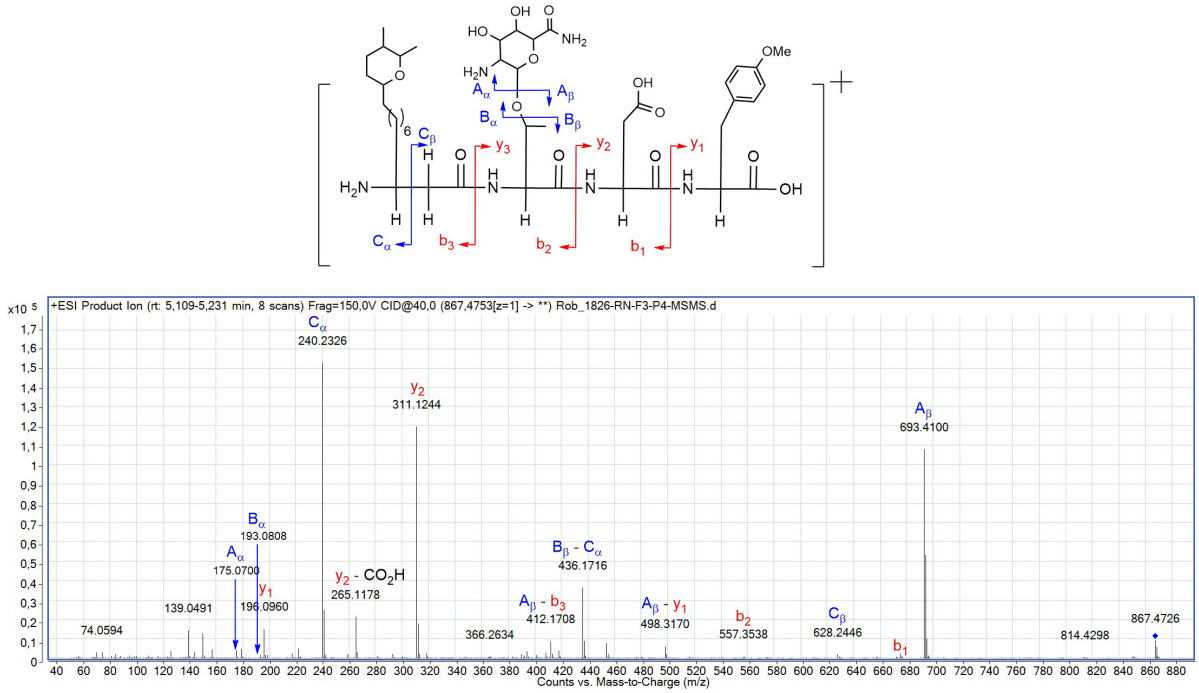


Figure 2. MS/MS analysis and fragmentation pattern of 1 and 2.

Table 2. ^{1}H (600 MHz) and ^{13}C (150 MHz) NMR Data for Compounds **1** in MeOH-d_4 and **2** in MeOH-d_4 and DMSO-d_6

		compd **1**		compd **2**			
residue		NMR solvent: MeOH-d_4		NMR solvent: MeOH-d_4		NMR solvent: DMSO-d_6	
	no.	δ_H, mult (*J*, Hz)	δ_C	δ_H, mult (*J*, Hz)	δ_C	δ_H, mult (*J*, Hz)	δ_C
O-Me-tyrosine	1		175.1		175.1		172.8
	2	4.56, dd (8.3, 5.2)	55.8	4.57, dd (8.5, 5.4)	55.7	4.38, td (8.0, 5.6)	53.8
	3	3.12, dd (14.0, 5.1) 2.93, dd (13.9, 8.5)	37.6	3.12, dd (14.0, 5.4) 2.93, dd (14.0, 8.5)	37.6	2.94, dd (13.9, 5.6) 2.81, dd (13.9, 8.2)	36.1
	4		130.2		130.2		129.2
	5/9	7.13, d (8.4)	131.3	7.15, d (8.6)	131.3	7.12, d (8.3)	130.2
	6/8	6.83, d (8.4)	114.9	6.84, d (8.8)	114.9	6.82, d (8.3)	113.6
	7		160.1		160.1		157.9
	10	3.75, s	55.7	3.76, s	55.7	3.71, s	55.0
	NH					8.30, d (7.9)	
aspartic acid	11		171.8		171.8		169.0
	12	4.68, dd (7.9, 4.5)	50.6	4.66, dd (8.0, 4.3)	50.6	4.53, td (7.3, 5.3)	48.7
	12	2.83, dd (16.1, 8.0) 2.79, dd (16.0, 4.4)	37.3	2.84, dd (16.1, 7.9) 2.77, dd (16.2, 4.2)	37.3	2.63, dd (16.2, 5.1) 2.58, dd (16.2, 7.0)	36.1
	14		174.8		174.8		172.6
	NH					8.35, d (7.5)	
threonine	15		171.4		171.5		167.9
	16	4.57, d (6.5)	58.1	4.53, d (7.2)	58.1	4.69, dd (8.5, 3.9)	55.6
	17	4.20, quint (6.4)	76.9	4.20, quint (6.5)	76.9	4.09, dq (6.2, 4.4)	76.1
	18	1.31, d (6.4)	18.2	1.27, d (6.4)	18.1	1.10, d (6.3)	16.3
	NH					8.60, d (8.4)	
sugar unit	1′	5.32, d (3.7)	97.4	5.37, d (3.8)	97.7	5.28, d (3.5)	95.9
	2′	3.22, dd (10.6, 3.7)	55.3	3.48, dd (10.8, 3.7)	52.1	3.32, m	50.3
	3′	3.79, dd (10.6, 8.9)	71.1	3.95, dd (10.8, 3.1)	68.0	3.69, (10.5, 2.8)	66.5
	4′	3.52, dd (9.8, 8.8)	73.6	4.24, dd (3.1, 1.4)	70.0	4.04, bs	68.3
	5′	4.05, d (9.9)	73.1	4.29, d (1.4)	73.1	4.02, d (1.7)	71.9
	6′		174.1		173.3		170.0
	NH_2					7.89, bs	
alkyl chain	1″		172.5		172.5		170.5
	2″	2.74, dd (16.5, 4.1)	37.7	2.72, dd (16.6, 4.1)	37.7	2.65, dd (16.1, 5.9)	37.4
	2″ a	2.58, dd (16.6, 8.7)	50.1	2.56, dd (16.5, 8.5)	50.1	2.50, m	48.0
	2″ b	3.54, m		3.54, m		3.38, m	
	3″						
	4″	1.67, m	33.9	1.67, m	33.9	1.51, m	32.3
	4″ a	1.62, m 1.41, m	26.2	1.62, m 1.41, m	26.2	1.47, m 1.32, m	24.3
	4″ b						
	5″a						
	5″b					1.27, m	
	6″	1.32, m	30.4	1.32, m	30.4		28.9

					1.22, m	
7″	1.33, m	30.4				
n+3″	1.30, m	30.7	1.30, m	30.7	1.23, m	28.8
n+4	1.33, m	26.7	1.33, m	26.7	1.32, m	25.1
″a *n*+4	1.48, m	37.4	1.48, m	37.4	1.23, m 1.37, m	35.9
″b *n*+5″a						
n+5″b	1.37, m		1.37, m		1.29, m	
n+6″	3.25, m	79.3	3.26, m	79.3	3.16, m	76.7
n+7	1.60, m	33.4	1.60, m	33.4	1.54, m	31.8
″a *n*+7	1.21, m 1.75, m	33.9	1.21, m 1.75, m	33.9	1.12, m 1.68, m	32.3
″b *n*+8″a						
n+8″b	1.20, m		1.20, m		1.10, m	
n+9″	1.19, m	38.7	1.19, m	38.7	1.11, m	36.9
n+10″	3.03, dq (8.5, 6.2)	81.1	3.04, dq (8.7, 6.2)	81.1	2.94, m	78.6
n+11″	1.15, d (6.2)	20.0	1.16, d (6.2)	20.0	1.06, d (6.1)	19.8
n+12″	0.83, d (5.9)	18.2	0.83, d (5.9)	18.2	0.77, d (5.2)	17.8
NH_2					7.89, bs	

The new characellide A (**1**) contains 13 chiral centers, and we decided to embark on the full configurational assignment of these centers using a set of different techniques. First, the relative configuration of the sugar unit was deduced from interpretation of the coupling constants of the different glycosidic protons. The coupling constant $^{3}J_{H1'\text{-}H2'}$ = 3.7 Hz was in accordance with an equatorial position for H-1′ and, therefore, an α substitution of the oxygen at C-1′, while all of the other substituents were placed in an equatorial position to respect values of about 9 Hz found for the other coupling constants of H-2′ to H-4′. The sugar unit of **1** therefore corresponds to a 2-amino-2-deoxyglucuronamide which has no precedent in the literature for a natural product. The relative configuration around the tetrahydropyran ring was then inferred from the ROESY spectrum where a H-15″/H-17″ rOe first placed the two methyls in a trans relative configuration and a second H11″/H-15″ rOe placed the methyl C-16″ in a cis relative configuration with C-10″.

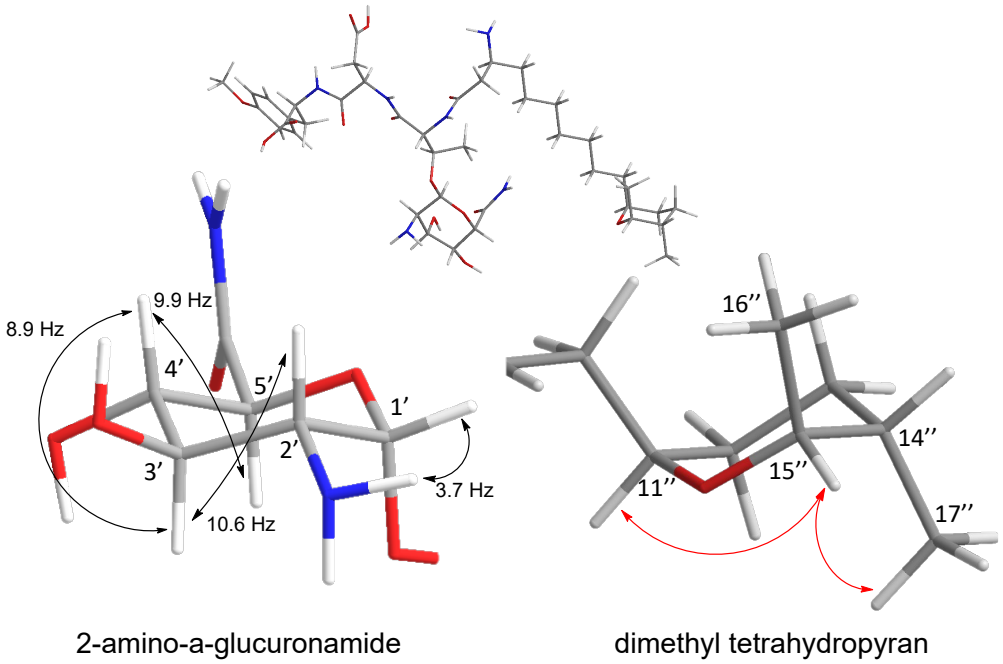


Figure 3. J-coupling constants (black arrows) of sugar unit (left) and nOe correlation (red arrows) of dimethyl tetrahydropyran ring (right) of characellide A

For the absolute configuration, we first decided to apply a Marfey's analysis on the three amino acids[5]. We were then able to assign the *L*-OMe tyrosine, *D*-aspartic acid, and *D-allo*-threonine among the four possible threonines. This core assignment set the stage for the determination of the absolute configurations of both the sugar unit and the primary amine on the long alkyl chain as chiral centers placed close to the amino acids. We decided to use a comparison between experimental and theoretical values of ^{13}C-NMR chemical shifts combined with the DP4 probability calculation. Our computing capacities allowed us to include the change in absolute configuration on the tetrahydropyran ring, even if this chiral part of the molecule was remote from the core tripeptide. We therefore ran a comparative analysis on eight possible stereoisomers: two for the sugar residue, two for the amino group at C-3'', and two for the tetrahydropyran ring respecting its relative configurations. The conformational analysis with an energy threshold of 1 kcal/mol gave 3, 2, 5, 3, 2, 1, 4, and 1 conformers for **1a**, **1b**, **1c**, **1d**, **1e**, **1f**, **1g**, and **1h**, respectively. After geometry optimization using DFT at the b3lyp/6-31g(d) level and verification of the presence of real energy minima for each conformer, NMR parameters were predicted using the GIAO method at the mpw1pw91/6-311+g(2d,p) level[6]. The DP4 probabilities were then calculated, and **1a** was identified as the most probable diastereoisomer with 80.4% confidence and **1e** the second most probable with 14.3% confidence (Figure 2). With an overall confidence of 95%, we could confirm the 3''S absolute configuration, but some ambiguity was left for the absolute configuration of the tetrahydropyran at 11''.

1a *2S,12R,16R,17R,1'S,3"S,11"S*
1b *2S,12R,16R,17R,1'S,3"R,11"S*
1c *2S,12R,16R,17R,1'R,3"S,11"S*
1d *2S,12R,16R,17R,1'R,3"R,11"S*
1e *2S,12R,16R,17R,1'S,3"S,11"R*
1f *2S,12R,16R,17R,1'S,3"R,11"R*
1g *2S,12R,16R,17R,1'R,3"S,11"R*
1h *2S,12R,16R,17R,1'R,3"R,11"R*

1a (80.4%)	**1e** (14.3%)	**1b**

Figure 4. DP4 probabilities for the eight diastereoisomers of **1**.

Although DP4 has been successfully applied to determine the relative configuration of flexible molecules[7], the length of the aliphatic chain could hamper the correct assignment of the absolute configuration for the tetrahydropyran moiety. While the absence of chromophore on the ring prevented the use of ECD calculations, we turned toward vibrational circular dichroism (VCD) calculations as the amount of sample was sufficient for compound **1**.

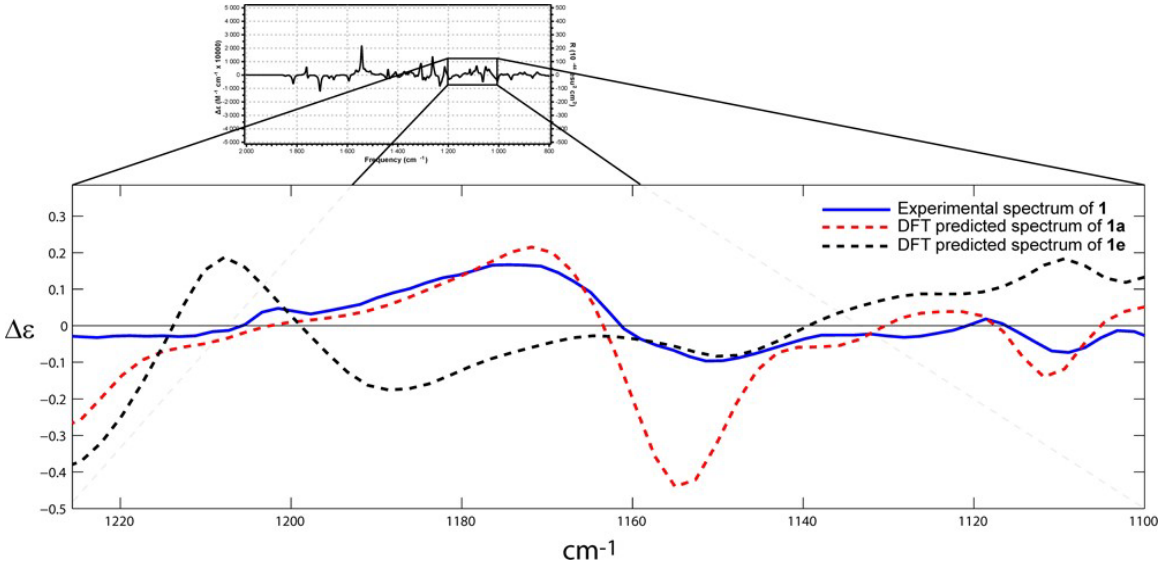


Figure 5. Comparison between experimental (blue) and theoretical VCD spectra (dotted lines) of the two diastereoisomers of 1 for the tetrahydropyran ring (red for **1a** and black for **1e**).

As shown in Figure 5, part of the experimental spectrum of **1** was in good agreement with the first proposed absolute configuration 1'S, 3''S, and 11''S (1a). Even if both analyses are in favour of an 11''S absolute configuration, this assignment still requires a definite confirmation through chemical synthesis. Compound **2** was isolated as an oil with an identical molecular formula as for **1**. The NMR spectra were very similar, and some differences appeared for the anomeric protons and other osidic signals. The analysis of the coupling constants of the osidic signals quickly revealed a galactose derivative for **2** with a characteristic signal at δ_H 4.24 (dd, J = 3.1; 1.4 Hz, 1H, H-4'). To inspect further the chemical diversity of this family of metabolites, we decided to look in more detail into the cluster associated with the new characellides **1** and **2** in the molecular network (see the Supporting Information). The analysis revealed two epimers **3** and **4** with masses in accordance with two nor analogues of **1** and **2**. We were able to isolate the two epimers as minor compounds in a mixture, but unfortunately, the low amount available did not allow a full separation on the HILIC column and we only report the ^{1}H-NMR spectrum of **3** and **4** in a mixture (see the SI). The mass fragmentation pattern of **3** and 4 was in perfect agreement with the NMR data of these compounds, therefore validating the use of the fragmentation patterns for the proposition of structures of other analogues in the cluster. Another interesting node at *m/z* 781.3978 drew our attention in the molecular network. Indeed, the analysis of the fragments of this ion suggested a terminal vinyl instead of the tetrahydropyran ring. This loss could therefore shed light on the biosynthetic events leading to the construction of the tetrahydropyran ring. Indeed, the dimethylated ring of characellides could stem from the cyclization of an isoprenyl unit on the terminal vinyl part of the latter molecule. This assumption would certainly require further confirmation through the isolation and structure determination of the corresponding metabolite (figure 6).

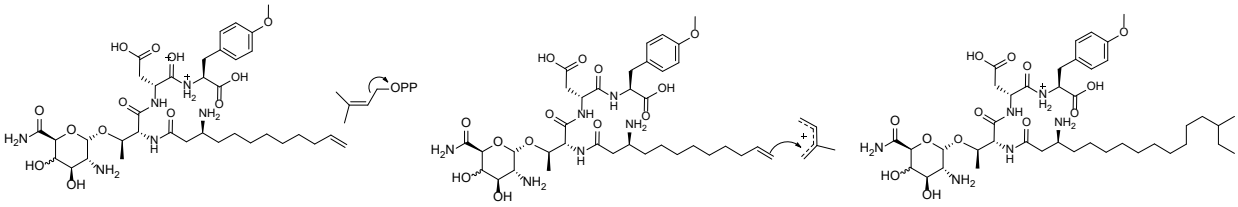


Figure 6. Proposed biosynthesis of dimethyl tetrahydropyran ring.

The amphiphilic nature of characellides prompted us to investigate their biological activities as anti-inflammatory agents, due to their similarity to sphingolipids. Only characellides A and B (**1** and **2**) were tested in the inflammatory cellular model microglia BV-2 cell line because the other two were isolated in a mixture. First, the effect of compounds over cell viability was checked. Microglia BV-2 cells were treated with different concentrations of compounds (0.001, 0.01, 0.1, 1, and 10 μM) for 24 h. Cell viability was determined by MTT assay. Compounds **1** and **2** were found to be noncytotoxic at all concentrations tested. Redox-mediated signaling has an important role in inflammation. In this context, oxidative damage results from an imbalance between the production of reactive oxygen species (ROS) and antioxidant cell defenses. Microglia-mediated inflammation is known to induce oxidative damage through ROS release. Therefore, the ability of these compounds to reduce ROS production in inflammatory conditions was tested. For this purpose, microglia BV-2 cells were stimulated with lipopolysaccharide (LPS) (500 ng/mL) for 24 h, and ROS generation was determined by fluorescence measurements with the dye CDFH-DA. First, cells were pretreated with different concentrations of compounds (0.001, 0.01, 0.1, 1, and 10 μM), and after 1 h, LPS was added. No significant variations were observed in the presence of compounds. However, when BV-2 cells were stimulated with LPS, intracellular ROS production was increased 30%, as shown in Figure 4. This increment was significantly attenuated ($p < 0.05$) in the presence of 10 μM of 1 (from 30 to 10%) and at concentrations down to 0.01 μM of 2 (from 30 to 15%). Therefore, a 50% reduction of intracellular ROS production on the microglia BV-2 cell line was observed after treatment with characellides. The increase in ROS production is related with variations in the mitochondrial membrane potential. To test whether the effect of characellides is mitochondrionmediated, the variations in mitochondrial membrane potential in these conditions were determined by using tetramethylrhodamine methyl ester (TMRM). No significant variations were observed in the presence of compounds. However, when cells were stimulated with LPS, a significant increase (20%, $p < 0.05$) was observed. When cells were preincubated with characellides, the effect of LPS was strongly inhibited and mitochondrial membrane potential values returned to control values. On the other hand, the relationship between inflammation and the release of ROS and NO is well known. Therefore, the effect of compounds over NO production was studied by measuring levels of nitrite (a marker of NO production) in LPS-stimulated BV-2 microglia cells. Cells were pretreated with characellides for 1 h prior to LPS (1 μg/mL) stimulation. LPS treatment significantly ($p < 0.001$) induced NO production compared with control. Cell pretreatment with compounds (0.1, 1, and 10 μM) decreased, but not significantly, nitrite production. In summary, these results suggest that characellides inhibit ROS production in LPS-stimulated microglia cells probably by a mechanism mediated by mitochondrion activity and not related to NO production, with the epimer **2** showing higher potency than **1**. Further experiments are necessary to clarify their mechanism of action.

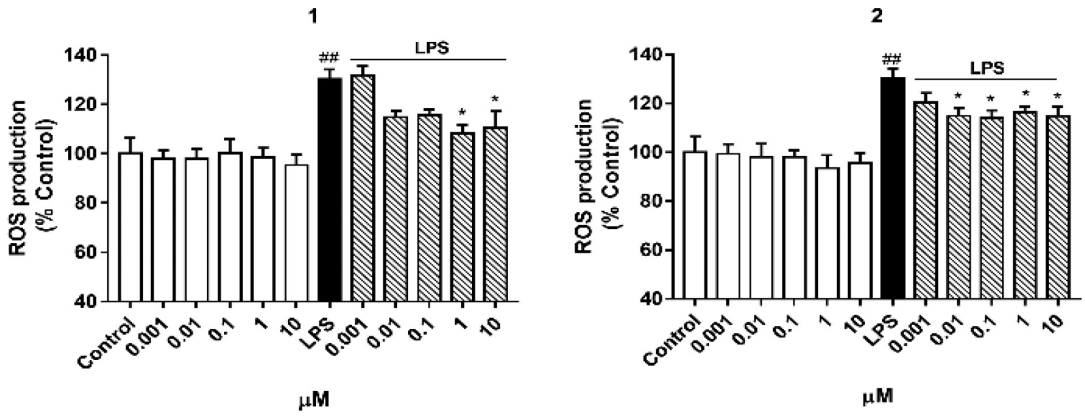


Figure 7. Effect of characellides A (1) and B (2) on intracellular ROS production in LPS-stimulated microglia BV-2 cells. Cells were pretreated with 1 and 2 at different concentrations (0.001, 0.01, 0.1, 1, and 10 µM) for 1 h, and then they were stimulated with LPS (500 ng/mL) for 24 h. ROS levels were measured with DCFH-DA. Dates are represented as a percentage of control cells, being the result of mean fluorescence intensity ± SEM of a minimum of N = 3 independent experiments performed in triplicate. The values are shown as the difference between cells treated with LPS alone versus cells treated with compounds in the presence of LPS by ANOVA statistical analysis followed by post hoc Dunnet's t-test. *$p < 0.05$, or cells treated with LPS versus control cells. $p < 0.01$.

2.4 Conclusions

Characellides represent a new family of natural products characterized by a combination of unique features including a hydrophilic glycosylic tripeptide linked to a long and liphophilic alkyl chain. Interestingly, citronamides are other tripeptides linked to an unusual sugar unit and a lipophilic chain, and they have been isolated from a sponge[8]. As always, the origin of characellides and citronamides might be the sponge, the bacteria, or the association of both. To the best of our knowledge, this is the first identification of a 2-amino-2- deoxy-6-carboxamide sugar residue in a small natural product, although it has already been described in lipopolysaccharides[9-11]. The lipophilic end also features a unique 2-alkyl-5,6- dimethyltetrahydropyran. The construction of HRMS/MS molecular networks applied to this valuable deep-sea sponge extract was useful to propose the structure of analogues of characellides including a possible biosynthetic intermediate[12]. Even if these compounds have been found to exhibit interesting anti-inflammatory activities, they belong to a new class of lipoglycopeptides, and some metabolites of this class like telavancin or oritavancin have been shown to possess strong antibacterial potential[13]. Work is ongoing to assess this potential. These results obtained on the first studied Irish deep-sea sponge are highly promising, and they highlight the huge potential of the deep marine biodiversity for biodiscovery.

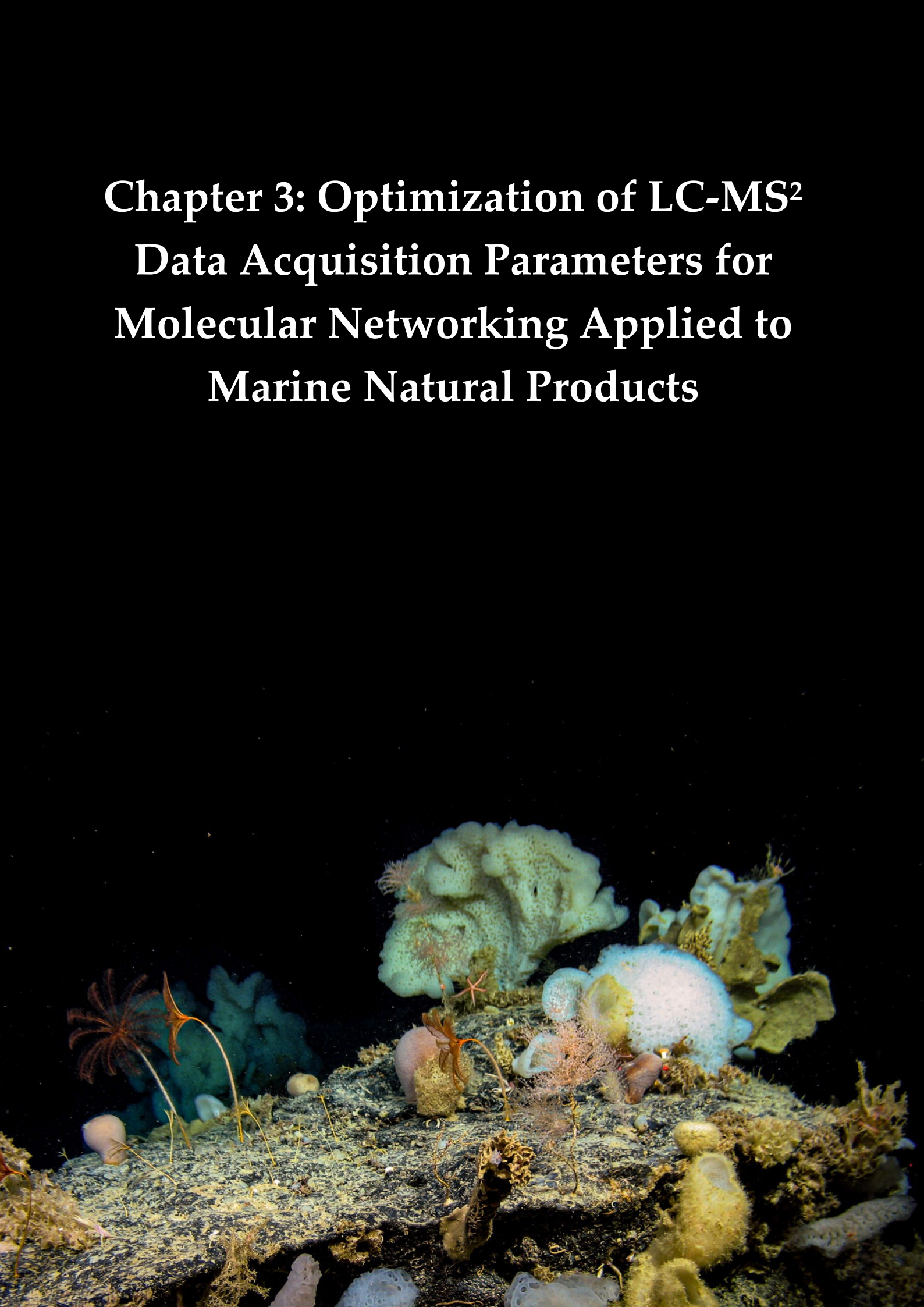
2.5 Supplementary Materials and Data Availability Statement

The Supporting Information, containing NMR and MS for compounds **1**-**4**, and ECD/ VCD spectra of **1**, available at DOI: 10.1021/acs.orglett.8b03684. Molecular networking data is available at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f0f277796394477e8d186099aa5ee1af.

2.6 References

1. Skropeta, D.; Wei, L., Recent advances in deep-sea natural products. *Natural product reports* **2014,** 31, (8), 999-1025.
2. Daletos, G.; Ebrahim, W.; Ancheeva, E.; El-Neketi, M.; Song, W.; Lin, W.; Proksch, P., Natural products from deep-sea-derived fungi a new source of novel bioactive compounds? *Current medicinal chemistry* **2018,** 25, (2), 186-207.
3. Carter, H., XXVII.—Descriptions and figures of deep-sea sponges and their spicules, from the Atlantic Ocean, dredged up on board HMS 'Porcupine,'chiefly in 1869 (concluded). *Journal of Natural History* **1876,** 18, (106), 307-324.
4. Suo, R.; Takada, K.; Irie, R.; Watanabe, R.; Suzuki, T.; Ise, Y.; Ohtsuka, S.; Okada, S.; Matsunaga, S., Poecillastrin H, a chondropsin-type macrolide with a conjugated pentaene moiety, from a Characella sp. marine sponge. *J. Nat. Prod.* **2018,** 81, (5), 1295-1299.
5. Bhushan, R.; Brückner, H., Marfey's reagent for chiral amino acid analysis: a review. *Amino acids* **2004,** 27, (3), 231-247.
6. Guillen, P. O.; Jaramillo, K. B.; Genta-Jouve, G.; Sinniger, F.; Rodriguez, J.; Thomas, O. P., Terrazoanthines, 2-aminoimidazole alkaloids from the Tropical Eastern Pacific zoantharian Terrazoanthus onoi. *Organic letters* **2017,** 19, (7), 1558-1561.
7. Cooper, J. K.; Li, K.; Aubé, J.; Coppage, D. A.; Konopelski, J. P., Application of the DP4 probability method to flexible cyclic peptides with multiple independent stereocenters: the true structure of cyclocinamide A. *Organic letters* **2018,** 20, (14), 4314-4317.
8. Carroll, A. R.; Duffy, S.; Avery, V. M., Citronamides A and B, tetrapeptides from the australian sponge Citronia astra. *Journal of natural products* **2009,** 72, (4), 764-768.
9. Kondakova, A. N.; Novototskaya-Vlasova, K. A.; Drutskaya, M. S.; Sof'ya, N. S.; Shcherbakova, V. A.; Shashkov, A. S.; Gilichinsky, D. A.; Nedospasov, S. A.; Knirel, Y. A., Structure of the O-polysaccharide chain of the lipopolysaccharide of Psychrobacter muricolla 2pST isolated from overcooled water brines within permafrost. *Carbohydrate research* **2012,** 349, 78-81.
10. Yildiz, F.; Fong, J.; Sadovskaya, I.; Grard, T.; Vinogradov, E., Structural characterization of the extracellular polysaccharide from Vibrio cholerae O1 El-Tor. *PLoS One* **2014,** 9, (1), e86751.
11. Zdorovenko, E. L.; Varbanets, L. D.; Shashkov, A. S.; Kiprianova, E. A.; Knirel, Y. A., Structure of the O-polysaccharide of the lipopolysaccharide of Pseudomonas chlororaphis subsp. aureofaciens UCM B-306. *Carbohydrate research* **2015,** 410, 47-50.
12. Otogo N'Nang, E.; Bernadat, G.; Mouray, E.; Kumulungui, B.; Grellier, P.; Poupon, E.; Champy, P.; Beniddir, M. A., Theionbrunonines A and B: dimeric vobasine alkaloids tethered by a thioether bridge from Mostuea brunonis. *Organic Letters* **2018,** 20, (20), 6596-6600.
13. Morata, L.; Mensa, J.; Soriano, A., Corrigendum to "New antibiotics against gram-positives: present and future indications"[Curr. Opin. Pharmacol. 24 (2015) 45–51]. *Current Opinion in Pharmacology* **2015**, (24), 147.

Chapter 3: Optimization of LC-MS^2 Data Acquisition Parameters for Molecular Networking Applied to Marine Natural Products



Chapter 3. Optimization of LC-MS^2 Data Acquisition Parameters for Molecular Networking Applied to Marine Natural Products

Published: 14th March 2022 *Metabolites* **2022**, *12*(3), 245;

DOI:/10.3390/metabo12030245

3.1 Abstract

Since the introduction of the online open-source GNPS, molecular networking has quickly become a widely applied tool in the field of natural products chemistry, with applications from dereplication, genome mining, metabolomics, and visualization of chemical space. Studies have shown that data dependent acquisition (DDA) parameters affect molecular network topology but are limited in the number of parameters studied. With an aim to optimize LC-MS^2 parameters for integrating GNPS-based molecular networking into our marine natural products workflow, a design of experiment (DOE) was used to screen the significance of the effect that eleven parameters have on both Classical Molecular Networking workflow (CLMN) and the new Feature-Based Molecular Networking workflow (FBMN). Our results indicate that four parameters (concentration, run duration, collision energy and number of precursors per cycle) are the most significant data acquisition parameters affecting the network topology. While concentration and the LC duration were found to be the two most important factors to optimize for CLMN, the number of precursors per cycle and collision energy were also very important factors to optimize for FBMN.

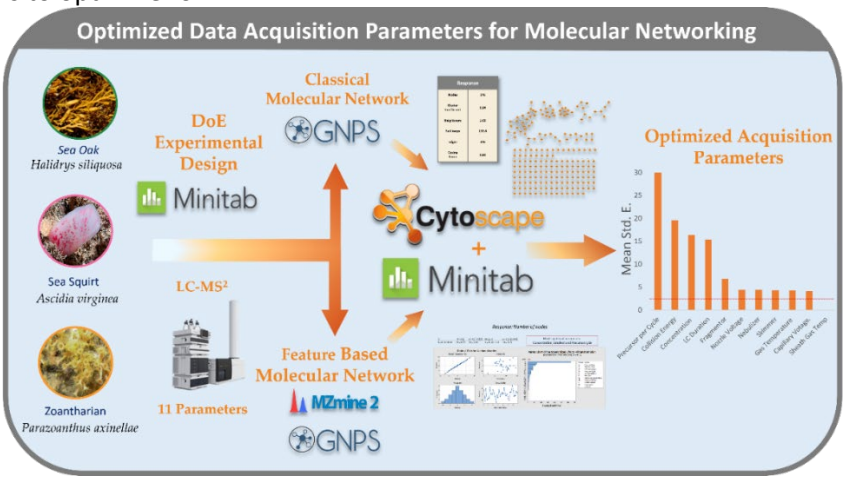


Table 1. Authors and their contributions

Authors	Conceptualization	Methodology	Software	Formal Analysis	Investigation	Data Curation	Resources	Original Draft	Review And Editing	Visualization	Funding Acquisition	Supervision	Administration project
Sam Afoullouss	x	x	x	x	x	x		x	x	x			
Agata Balsam		x	x	x	x	x							
A. Louise Allcock							x		x		x	x	x
Olivier P. Thomas		x					x		x		x	x	x

3.2 Introduction

Molecular networking is an informatics tool that allows visualization of non-targeted tandem mass spectrometer data (MS^2), to highlight structure similarities between metabolites of a complex mixture and help in the annotation of the detected metabolites [1]. The most common data acquisition technique for molecular networking uses data-dependent acquisition (DDA) [2]. DDA is an autonomous data acquisition mode and works by first taking an MS^1 scan and collecting the *m/z* and relative abundance of analytes. This is instantly followed by multiple MS^2 scans, targeting the major analytes selected from the MS^1 scan. Molecular networking is now broadly used in the field of natural products (NP) with the introduction of online molecular networking Global Natural Products Social (GNPS) platform, developed by Wang et al. [3] in 2016. GNPS has been applied to a wide range of applications including dereplication [4–7], metabolomics [8–11], and genome mining [12,13].

GNPS uses an algorithm to compare the similarities of fragmentation spectra (MS^2) in each dataset, generating a cosine score for each pair of MS^2 spectra. Working on the principle that structurally similar molecules will produce MS^2 spectra with fragment ions in common [14], the cosine scores aim to measure spectral similarity. Using the MS-Cluster algorithm [15], MS^2 spectra with identical parent ion masses are combined to produce a single consensus spectrum and are represented by a single node, characterized by its *m/z* and MS^2 fragment ion patterns. Nodes with similar MS^2 spectra, and therefore high cosine scores (usually set at more than 0.7), are connected via edges to form a cluster. Clusters can be considered "molecular families", as the metabolites should share key chemical features. When combined with automated searches of databases containing MS^2 spectra, the known compounds can quickly be annotated, turning molecular networking and GNPS into a powerful dereplication tool [16]. A key feature of GNPS is the ability for users to generate and share MS^2 spectra for identified compounds, which can be uploaded to open-access GNPS reference spectra libraries. Thus, the natural products community can contribute to the rapid growth of reference spectra libraries, increasing the range of natural products which can be annotated quickly.

GNPS contains two main workflows to create molecular networks: Classical Molecular Networking (CLMN) and Feature-Based Molecular Networking (FBMN) [17]. Both workflows use the same untargeted LC-MS^2 data. CLMN was the first tool to be introduced as a quick and effective way to visualize the chemical space of a sample, creating networks using only MS^2 spectra. The FBMN workflow advances on CLMN by using both MS^1 (e.g., isotopic pattern, retention time) and MS^2 data to create more reproducible and accurate molecular networks. In the FBMN workflow, LC-MS^2 data are processed using MZmine2 [18] or similar. This processing allows FBMN to be used for relative quantification and increases reproducibility including the ability to resolve isobaric isomers [17]. Both molecular networking workflows have been widely utilized in natural product related fields [3].

Recent works showed that molecular network topologies are affected by DDA parameters, including intensity threshold, collision energy, and exclusion after n scans [18–20]. Each of these studies employed a One Factor at a Time (OFAT) approach and investigated a limited number of parameters. Our preliminary investigation did likewise indicate that the four parameters tested (concentration, liquid chromatography duration, precursors per cycle, and collision energy) had a significant impact on CLMN [21]. We recognized that an OFAT approach, while simple to implement, was inefficient for testing a large range of parameters. This approach also limits the ability to observe interactions between parameters. We therefore used a fractional factorial design to evaluate the effect of multiple LC-ESI-MS^2 data acquisition parameters on the resulting molecular network. Fractional factorial design uses a subset of experimental runs required for a full factorial analysis. Two key advantages of using the fractional factorial design are: (i) it allows for the statistical significance of each parameter to be determined; (ii) it assesses the significance of interactions between parameters. To cover the diverse

range of families of marine natural products with distinct physico-chemical properties, we analysed extracts of three widely differing marine organisms present off the coasts of Ireland: the sea squirt *Ascidia virginea*, rich in small phenolic derivatives currently under chemical investigation, the zoantharian *Parazoanthus axinellae* known to contain a range of polar alkaloids, and the macroalga *Halidrys siliquosa*, rich in meroterpenoids (Figure 1). To the best of our knowledge this is the first study optimizing LC-MS² data acquisition parameters for FBMN.

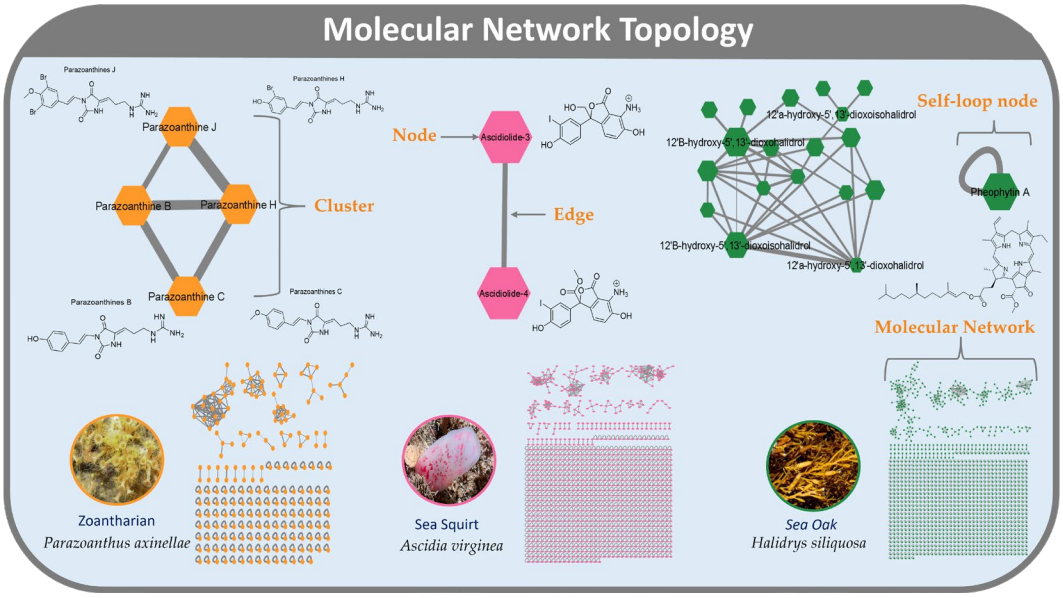


Figure 4 Annotated molecular networks of the studied samples, *Parazoanthus axinellae* (yellow, left), *Ascidia virginea* (pink, centre), and *Halidrys siliquosa* (green, right), showing the diverse range of metabolites. Elements of molecular networking topology are labelled.

3.3 Results and Discussion

Nomenclature

Molecular Networking		Response Modelling	
Self-loop Nodes	A self-loop is an edge that originates from and terminates the same node. Self-loop nodes tend to indicate a high noise/signal ratio.	**Factor**	Factors are the variables that are controlled during an experiment in order to determine their effect on the response variable. e.g. Concentration, Collision energy
Cluster Coefficient	A measure of how well nodes connect together in clusters. It indicates how molecular families are represented in the network.	**Response**	Responses are the characteristics of the MN that are measured to observe the factors effects. e.g. Number of nodes
Cosine Score	A measure between 0 and 1 of how similar MS² spectra for two compounds are. The more fragment peaks in common, the higher the cosine score.	**Standardize Effect**	Standardized effect is a unitless measure of a factors effect on a response. It allows for the comparison of a factors' effect on different responses.
Average Number of Neighbours	An indicator of how well molecular families are represented in the network.	**Predicted R^2**	Indicated the ability for a regression model to predict responses. Higher predicted R^2 values indicated a better quality model.
Average Number of Edges	An indicator of how well molecular families are represented in the network.	**P-value**	The probability of a null-hypothesis being true. Lower p-values indicate a more accurate predictability of the model.

Figure 2. Nomenclature used for molecular networking (left) and DoE response modelling (right).

3.3.1. Response Models

Of the 36 response models generated using fractional factorial analysis, the majority had high Predicted R^2 (R^2 Pred.) and *p*-values < 0.01 (Table S8) indicating a high goodness-of-fit. Residual plots

(Figures S4–S39) were also analyzed to evaluate the fit of the models. Deviation in the quality of response models was seen between the two molecular networking workflows and between the three samples. Key nomenclature for this study is summarized in Figure 2.

Models with poor fit (p-value > 0.05 or R^2 Pred. < 20) were not included in the response analysis. Analysis of CLMN significant factors excluded response models for clustering co-efficient, number of neighbors and two of three average cosine models, due to poor model fit. Response models for FBMN produced better quality models. *Parazoanthus axinellae* models were an outlier with response models for the number of nodes, neighbors and average cosine being excluded from further analysis due to poor fit with data.

3.3.2. Significant Factors and Significant Factor Interactions

When averaged across response models and sample, only two factors, fragmentor voltage and drying gas temperature did not have a significant effect on the measured responses in the CLMN workflow. While the effect of all other factors was significant (Figure 3), they varied in the size of their Standardized Effect (Std E.). Sample concentration (Std E. 23.1) had the greatest effect, followed by LC duration (Std E. 5.8), precursor per cycle (Std E. 5.5), collision energy (Std E. 5.0), sheath gas temperature (Std E. 4.7), skimmer voltage (Std E. 4.4), nozzle voltage (Std E. 3.5), nebulizer pressure (Std E. 2.9), and capillary voltage (Std E. 2.7).

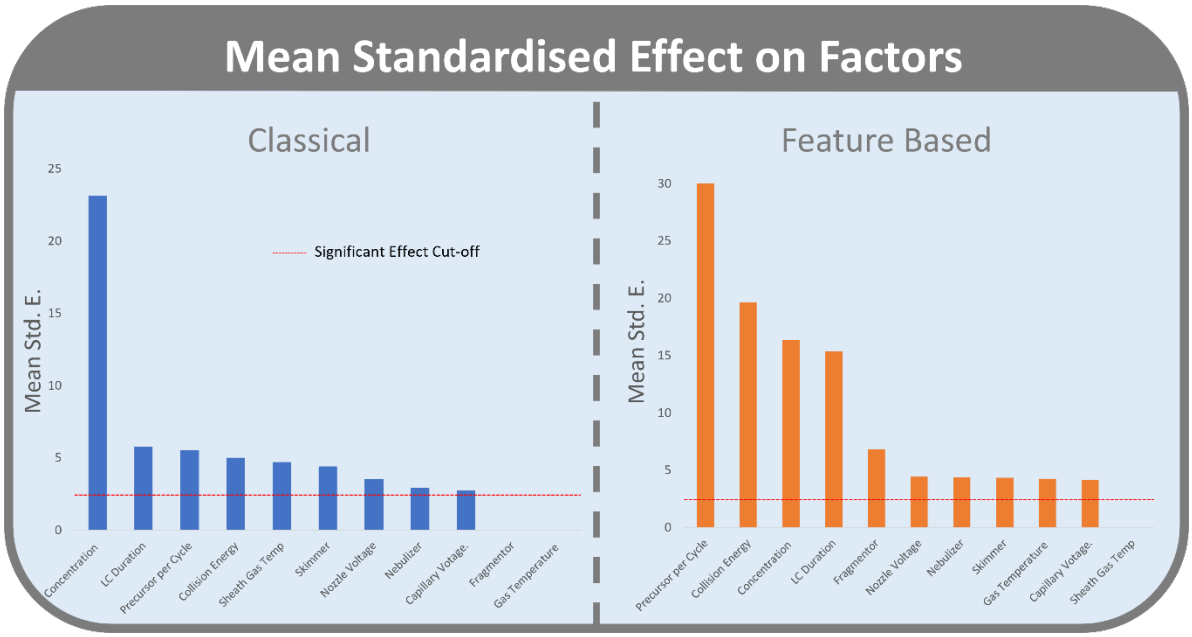


Figure 3. Significant effects of factors on CLMN (blue, left) and FBMN (orange, right), averaged across response models and samples. For included models, refer to Figure S11. Standardized Effects (Std E) of 2.27 or greater is considered significant (red dashed line).

In the FBMN workflow, all factors, except sheath gas temperature, had a significant effect on the measured responses. Precursor per cycle (Std E. 33.4) had the greatest standardized effect, followed by collision energy (Std E. 19.6), sample concentration (Std E. 16.4), LC duration (Std E. 15.4), fragmentor voltage (Std E. 6.9), nozzle voltage (Std E. 4.5), nebulizer pressure (Std E. 4.4), skimmer voltage (Std E. 4.3), drying gas temperature (Std E. 4.3), and capillary voltage (Std E. 4.2).

Four factors displayed a consistent statistically significant effect for both CLMN and FBMN workflows with high mean standardized effects: Sample concentration, number of Precursors per cycle, LC duration, and collision energy (Figure 1). These four factors are explored in more detail in Sections 2.2.1–2.2.4. Although fragmentor voltage, sheath gas temperature, skimmer voltage, and nozzle

voltage were shown to have a significant effect on molecular network topology, a lower mean standardized effect and inconsistent effect across the three samples show the lesser importance of optimizing these factors.

Interactions occur when the effect of a factor on a response is dependent on the level of another factor. For CLMN, there were nine significant interactions between factors effecting molecular network responses. These were interactions between concentration and eight other factors (collision energy, nebulizer pressure, skimmer voltage, nozzle voltage, drying gas temperature, precursors per cycle, LC duration, and fragmentor voltage), and between LC duration and precursors per cycle (Figure 4). That concentration was involved in eight of the nine significant factor interactions, further highlights the importance of this factor.

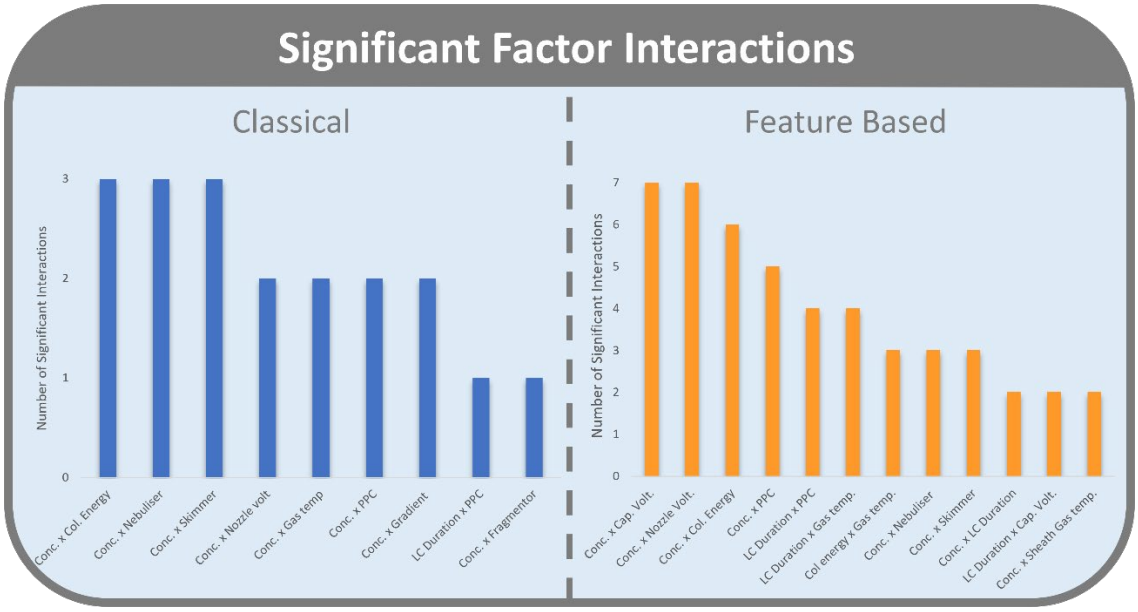
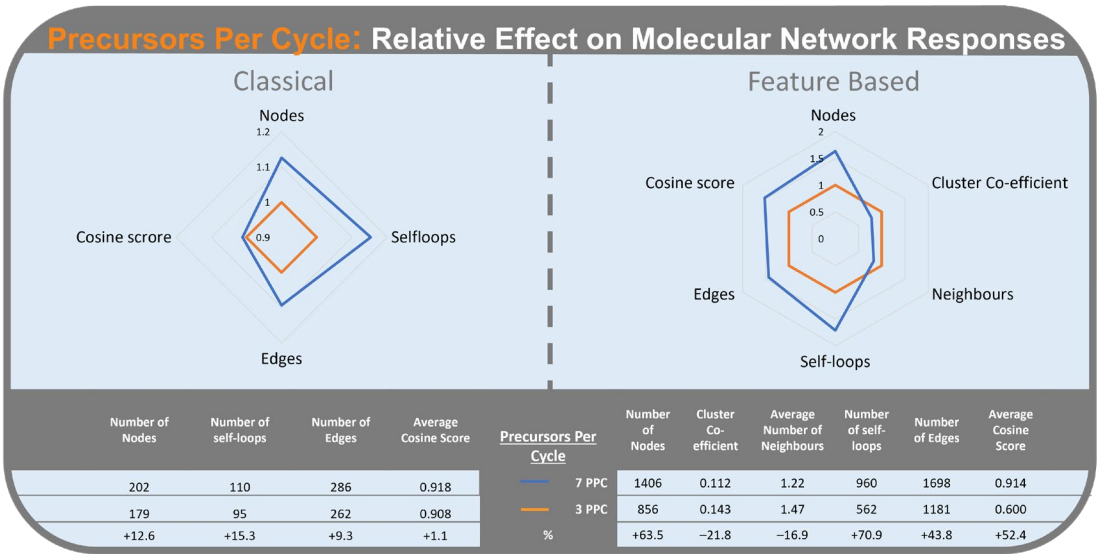


Figure 4. Number of significant factor interactions that affect responses for CLMN (blue, left) FBMN (orange, right).

A broader range of interactions was seen in the FBMN workflow, with 12 significant interactions. Nine of these significant interactions were seen between sample concentration and other factors (capillary voltage, nozzle voltage, collision energy, precursors per cycle, nebulizer pressure, skimmer voltage, LC duration, and sheath gas temperature). LC duration had further significant interactions with precursors per cycle, drying gas temperature, and capillary voltage.

3.3.2.1. Precursor Per Cycle

The number of precursors per cycle (PPC) had the third largest Standardized Effect in CLMN and was a significant factor for the four responses we considered for this workflow (Figure 3). Higher PPC (Figure 5) resulted in an increase in self-loop nodes (15.3%), number of nodes (12.6%), number of edges (9.3%), and average cosine score (1.1%).



Number of Nodes	Number of self-loops	Number of Edges	Average Cosine Score	Precursors Per Cycle	Number of Nodes	Cluster Co-efficient	Average Number of Neighbours	Number of self-loops	Number of Edges	Average Cosine Score
202	110	286	0.918	7 PPC	1406	0.112	1.22	960	1698	0.914
179	95	262	0.908	3 PPC	856	0.143	1.47	562	1181	0.600
+12.6	+15.3	+9.3	+1.1	%	+63.5	−21.8	−16.9	+70.9	+43.8	+52.4

Figure 5. Radar graph of the number of precursors per cycle effect (3 precursors per cycle in orange; 7 precursors per cycle in blue) on the relative change of responses, averaged across the three samples. Classical Molecular Networking (left) and Feature-Based Molecular Networking (right).

For FBMN, PPC had the highest Standardized Effect on molecular network topology and significantly affected all responses. Higher PPC increased the number of self-loop nodes (70.9%), number of nodes (63.5%), cosine score (52.4%), and number of edges (43.8%). Higher PPC also reduced the cluster co-efficient (−21.8%) and the average number of neighbours (−16.9%).

PPC consistently affected the number of nodes, number of self-loop nodes, and number of edges. More PPC results in more analytes being chosen for fragmentation per cycle, reducing competition between parent ions. As more analytes are being selected for MS^2 fragmentation, more nodes appear in the network. Increases in the number of edges with more PPC may result from more minor metabolites being selected for MS^2 fragmentation; these would have been outcompeted with a lower number of PPC.

In FBMN, higher PPC caused a decline in the number of neighbours and clustering co-efficient. As both values are calculated as an average for the whole network, this reflects the increase in the number of self-loop nodes. The number of clusters with just two nodes increases considerably at the higher level of PPC compared with the lower level. Since two node clusters are awarded a clustering co-efficient of 0, this translates to a decline in average clustering co-efficient for the network. This effect can be seen in the FBMN of *Ascidia virginea* under the condition of T25 (3 precursors per cycle) and T17 (7 precursors per cycle (Figure 6).

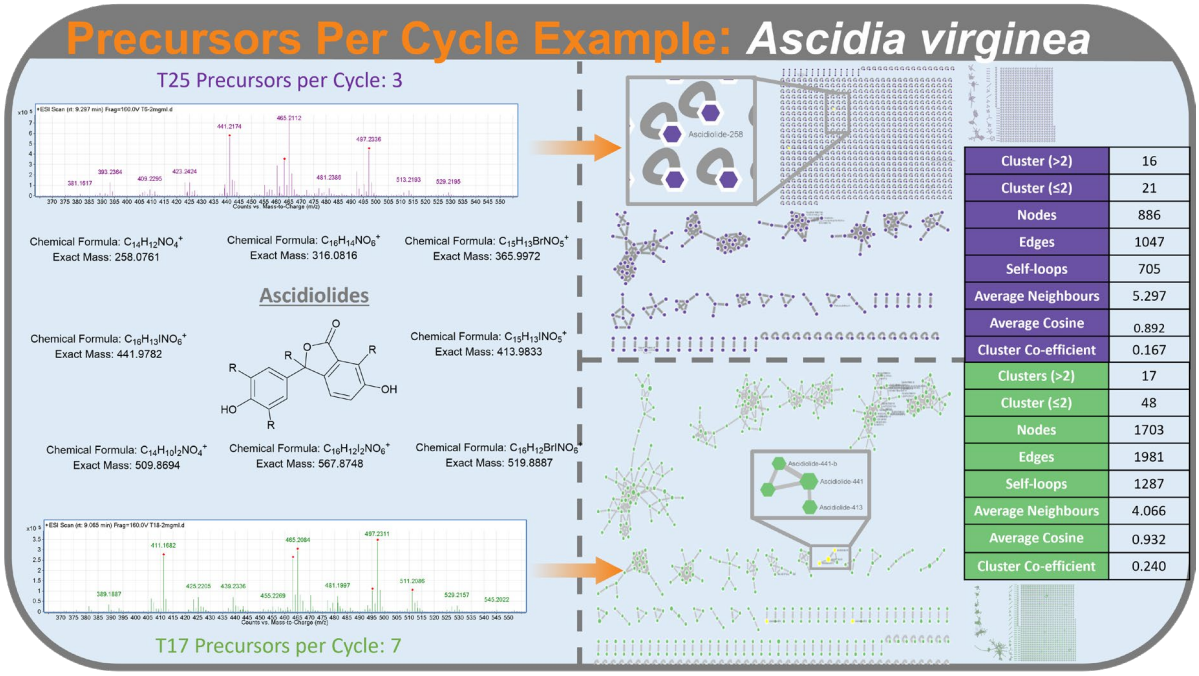
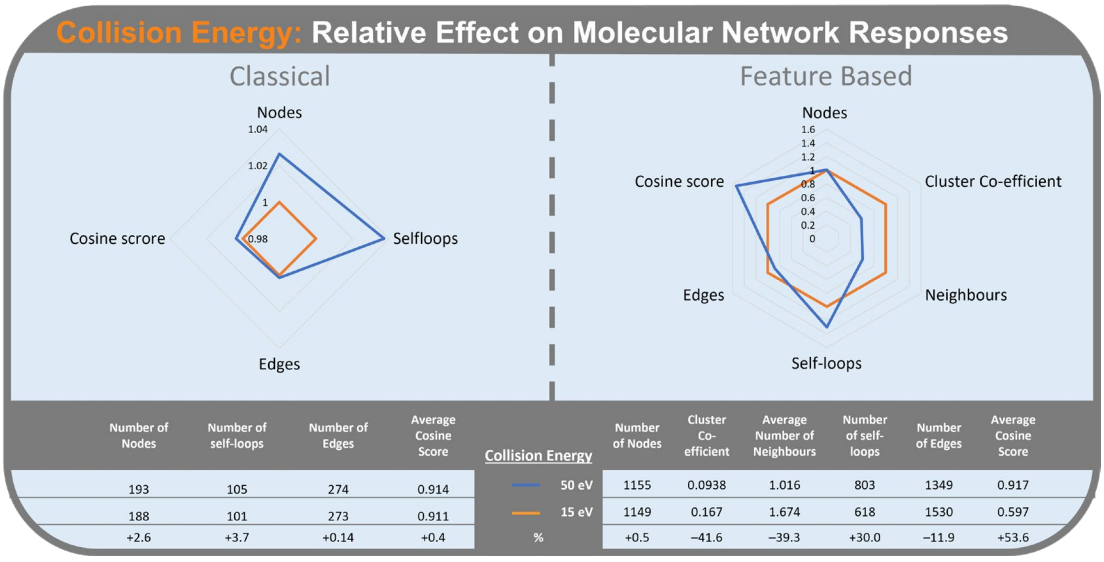


Figure 6. An example of the effects of PPC on molecular networking responses, of a Feature-Based Molecular Network for *Ascidia virginea*. MS spectra with ions selected form MS^2 fragmentation denoted with a red dot for precursor per cycle of 3 (top, purple, T25) and 7 (bottom, green, T17). Molecular networks produced from these parameters are displayed on the right, with tables displaying molecular network statistics. Eight ascidiolides were annotated with T17 molecular network (7 precursors per cycle) and three ascidiolides were annotated in T25 molecular network (3 precursors per cycle).

3.3.2.2. Collision Energy

In the CLMN workflow, collision energy (Figure 7) had the fourth largest mean Standardized Effect. Increasing collision energy from 15 eV to 50 eV increased number of self-loop nodes (3.7%), number of nodes (2.6%), average cosine score (0.4%), and number of edges (0.14%).



Number of Nodes	Number of self-loops	Number of Edges	Average Cosine Score	Collision Energy	Number of Nodes	Cluster Co-efficient	Average Number of Neighbours	Number of self-loops	Number of Edges	Average Cosine Score
193	105	274	0.914	50 eV	1155	0.0938	1.016	803	1349	0.917
188	101	273	0.911	15 eV	1149	0.167	1.674	618	1530	0.597
+2.6	+3.7	+0.14	+0.4	%	+0.5	−41.6	−39.3	+30.0	−11.9	+53.6

Figure 7. Radar graph of collision energy effect (15 eV in orange) (50 eV in blue) on the relative change of responses. Classical Molecular Networking (left) and Feature-Based Molecular Networking (right).

In the FBMN workflow, collision energy had the second highest mean Standardized Effect. Higher collision energy resulted in an increase in average cosine score (53.6%) and number of self-loop nodes (+30%). Higher collision energy also resulted in a decrease in cluster co-efficient (−41.6%), average

number of neighbours (−39.3%), and number of edges (−11.9%). Collision energy did not significantly affect the number of nodes in FBMN.

Increasing collision energy generates a higher number of fragments in MS^2 spectra for a given precursor ion. This provides more points of reference when comparing MS^2 spectra, leading to the increase in cosine score seen in both CLMN and FBMN. This increase in cosine score, combined with the decrease in number of edges in FBMN, indicates that while fewer similarities between metabolites are detected, the similarities (edges) are being detected more accurately (with higher cosine score). The decrease in cluster co-efficient and average number of neighbours appears to be a result of the combined effect of the increase in self-loop nodes (+30%) and decrease in number of edges (−11.9%). As the edges become more accurate due to more datapoints, matching between nodes is reduced, in turn reducing number of edges, neighbours and cluster size.

An example of collision energy effect on molecular network topology can be seen when comparing FBMN of *Parazoanthus axinellae* under parameters set for T6 (collision energy of 15 eV) and T32 (collision energy of 50 eV). The MS^2 spectra produced with a collision energy of 50 eV resulted in a higher number of fragments produced from parazoanthine E and a parazoanthine cluster with a lower number of edges and higher average cosine scores when compared to the same MS^2 spectra and cluster which used a collision energy of 15 eV (Figure 8).

As the effect of collision energy on fragmentation pattern is dependent on the parent ions chemical structure, the increase in the number of fragments observed for parazoanthine E was not seen with the meroterpenoids of the *Halidrys siliquosa,* or the ascidiolides detected in *Ascidia virginea.* This highlights the importance of optimizing the collision energy for the particular type of chemistries within the samples.

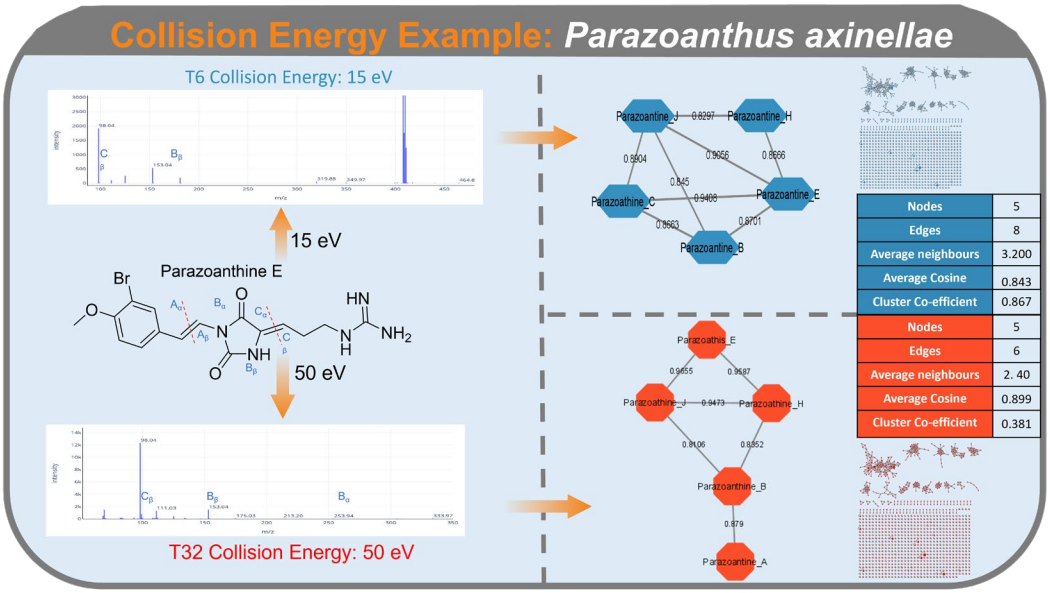
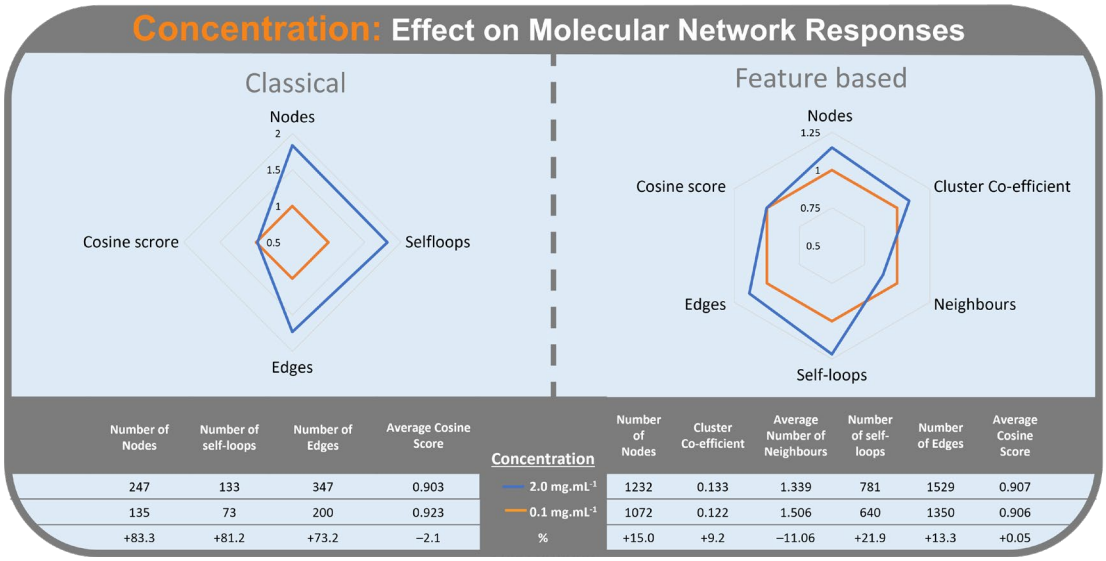


Figure 8. Fragmentation pattern and MS^2 spectra of parazoanthine E using a collision energy of 15 eV (top, blue, T6) and 50 eV (bottom, red, T32). Parazoanthine clusters with annotated nodes, resulting from a collision energy of 15 eV (top, blue, T6) and 50 eV (bottom, red, T32). Edges are labeled with cosine scores. Tables with cluster statistics under both conditions are displayed on the right.

3.3.2.3. Concentration

Concentration (Figure 9) had the highest Standardized Effect on CLMN responses. The highest concentration of 2.0 mg/mL increased the number of nodes (83.3%), number of self-loop node (81.2%)

and the number of edges (73.2%), when compared with the lowest concentration of 0.1 mg/mL. Average cosine score decreased (−2.1%) with the highest concentration.



Number of Nodes	Number of self-loops	Number of Edges	Average Cosine Score	Concentration	Number of Nodes	Cluster Co-efficient	Average Number of Neighbours	Number of self-loops	Number of Edges	Average Cosine Score
247	133	347	0.903	2.0 mg.mL⁻¹	1232	0.133	1.339	781	1529	0.907
135	73	200	0.923	0.1 mg.mL⁻¹	1072	0.122	1.506	640	1350	0.906
+83.3	+81.2	+73.2	−2.1	%	+15.0	+9.2	−11.06	+21.9	+13.3	+0.05

Figure 9. Radar graph of concentration (0.1 mg/mL in orange) (2.0 mg/mL in blue) on the relative change of responses. Classical Molecular Networking (left) and Feature-Based Molecular Networking (right).

For FBMN, concentration had the third highest Standardized Effect on molecular network responses. Increasing concentration (Figure 9) from 0.1 mg/mL to 2 mg/mL, increased the number of self-loop nodes (21.9%), number of nodes (15.0%), number of edges (13.3%), and cluster co-efficient (9.2%). The average number of neighbours decreased (−11.1%) at the highest concentration.

As expected, these results indicate that more analytes are detected as concentration increases. This result must be due to the detection of minor metabolites in the sample, which at low concentration are undetected in molecular networking as they do not pass the ion intensity threshold for MS^2 fragmentation. Increasing the sample concentration increases the ion intensity of these minor metabolites, allowing them to pass the threshold and be represented in the network. Increase in cluster coefficient and average number of edges in the FBMN workflow supports the hypothesis of more minor metabolites being detected with higher concentrations. The increase in self-loop nodes, seen in both CLMN and FBMN, may be a consequence of low intensity ions producing MS^2 spectra with a higher signal/noise ratio. These high noise spectra are difficult for the molecular networking algorithms to process, and result in an increase in self-loop nodes, as fragment matching is interrupted by increased noise. The variation in the effect of concentration on the number of self-loop nodes in CLMN versus FBMN is likely due to the more advanced data processing used in FBMN.

This hypothesis is supported when comparing the number of nodes in a low and high concentration molecular network. Comparison of equivalent clusters from the FBMN of the seaweed *Halidrys siliquosa* from run T5 (0.1 mg/mL) and run T17 (2.0 mg/mL), where only concentration differs, shows that at the highest concentration clusters in the network become more populated with low intensity metabolites being represented (Figure 10).

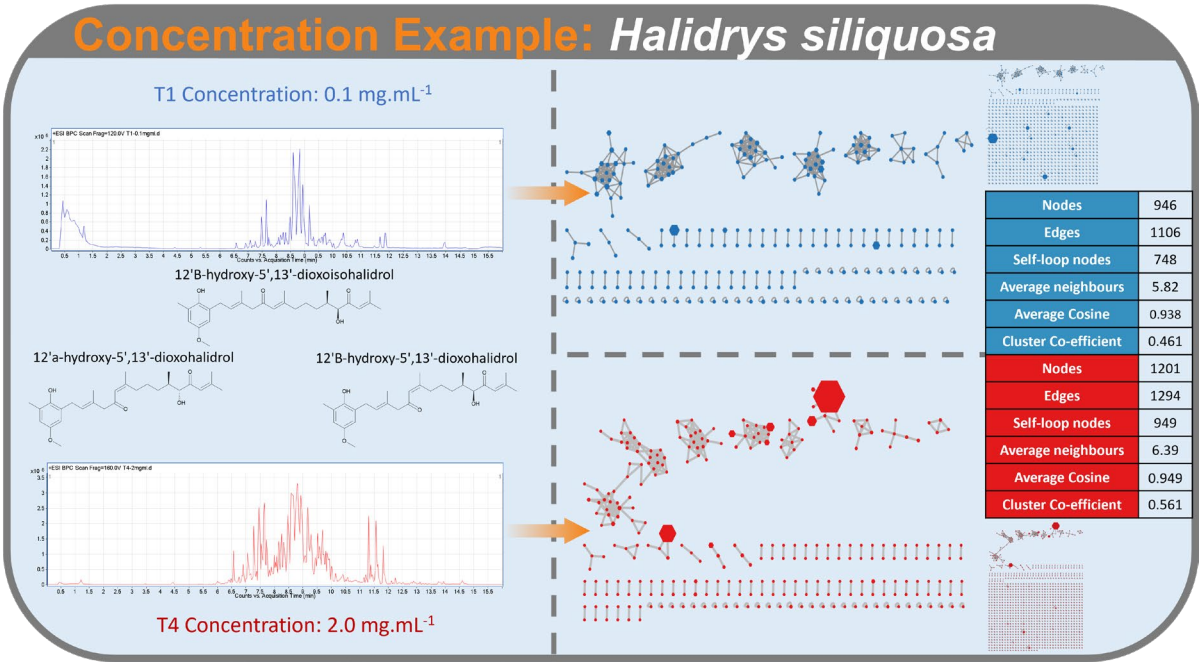


Figure 5 An example comparing the effect of low (0.1 mg/mL; left) and high (2.0 mg/mL; right) concentrations on Feature-Based Molecular Networking responses of *Halidrys siliquosa.*

3.3.2.4. LC Duration

LC duration had the second highest mean Standardized Effect on CLMN responses (Figure 3). Increased LC duration (Figure 11) resulted in an increase in the number of nodes (17.5%), number of self-loop nodes (17.4%), and number of neighbours (15.4%). LC duration did not have a significant effect on average cosine score (0.05%) in CLMN.

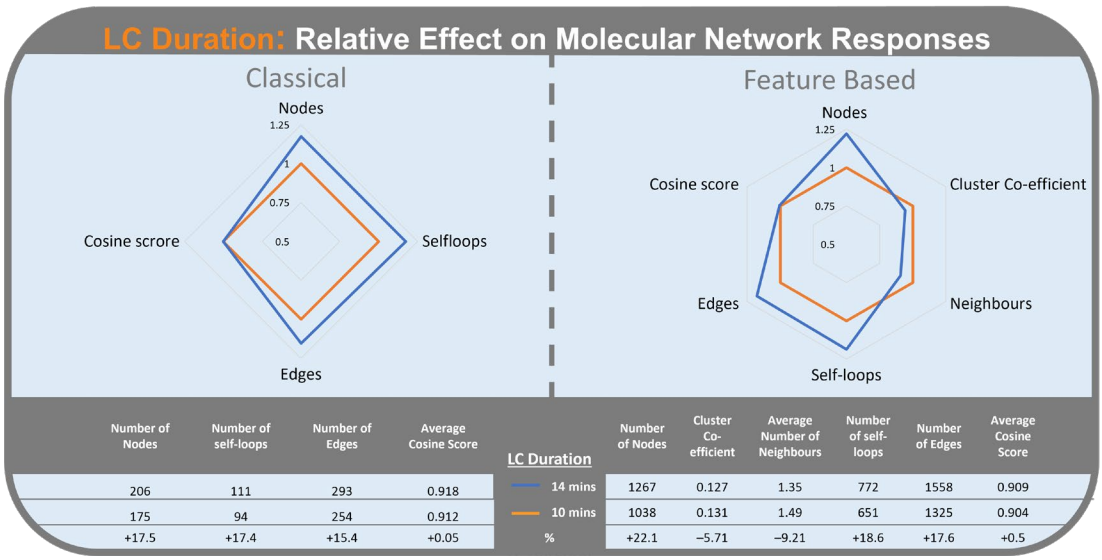


Figure 11. Radar graph of the effect of liquid chromatography duration (10 min in orange) (14 min in blue) on the relative change of responses of Classical Molecular Networking (left) and Feature-Based Molecular Networking (right).

In the FBMN workflow, LC duration had the fourth highest mean Standardized Effect (Std. E. 14.5), with the longer LC duration resulting in an increase of the number of nodes (22.1%), number of self-loop nodes (18.6%), number of edges (17.6%), and average cosine score (0.5%). Average number of neighbours (−9.21%) and cluster co-efficient (−5.7%) decreased with the longer LC duration (Figure 11).

An increase in LC duration improves separation between analytes (Figure 12), which reduces the number of analytes entering the mass spectrometer at any one time. As the number of parent ions that can undergo MS^2 fragmentation per MS scan is limited due to factors including number of precursors per cycle and exclusion time, we hypothesize that greater separation between analytes decreases the competition among parent ions for MS^2 fragmentation. This reduces competition, results in more analytes undergoing MS^2 fragmentation, corresponding to the increase observed in the number of nodes and edges in CLMN and FBMN. As the FBMN workflow utilizes retention time of analytes in the feature detection and alignment step, the increase in separation between analytes may benefit the detection of isomers/stereoisomers. This may contribute to the increase in number of nodes and edges.

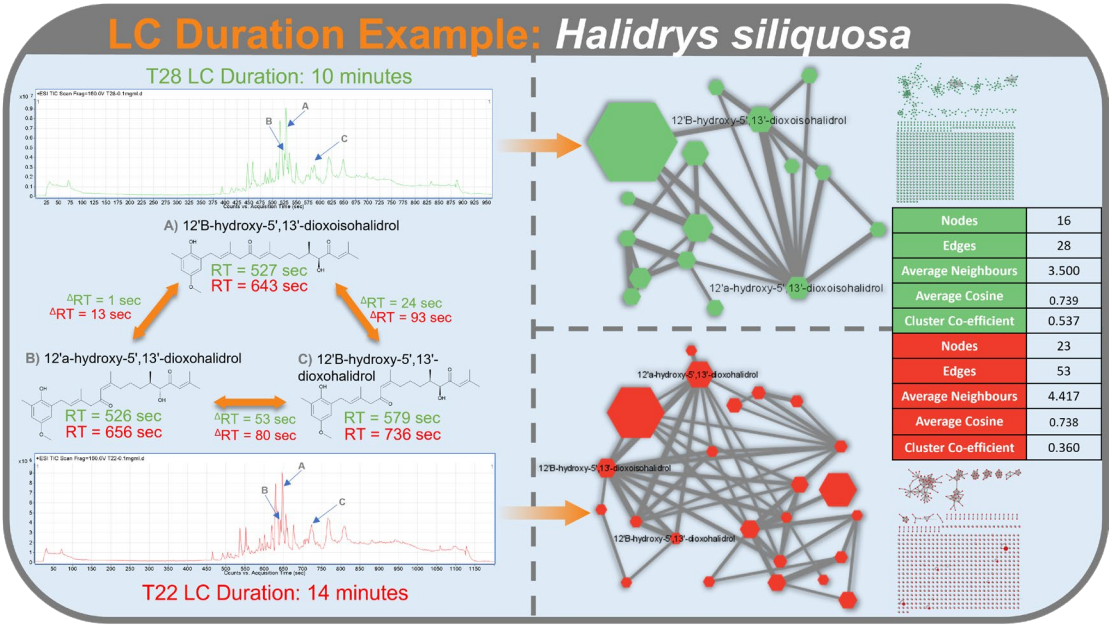


Figure 12. LC-MS^2 trace of *Halidrys siliquosa* using a LC duration of 10 min (top; green) and 14 min (bottom; red) with the retention times of the meroterpenoids under both conditions displayed. Cluster statistics for the meroterpenoid clusters for both LC-MS^2 runs are displayed on the right.

3.3.2.5. Comparison of CLMN and FBMN

According to the results obtained, data acquisition parameters had a larger and more consistent effect on FBMN than CLMN (Figure 3). For FBMN, the significant effects and their Standardized Effect values were consistent across all three samples, whereas for CLMN inconsistencies between the three samples were observed (Table S8). The higher rate of error associated with MSCluster, such as chimeric spectra represented by one node, could have translated into increased error with network statistics and therefore inaccuracies in the fractional factorial analysis. The increased reproducibility of FBMN, due to processing steps such as feature detection and alignment, resulted in more accurate measurements of the effects of parameters. The inclusion of retention times and isotope grouping also allow for more accurate networking, reducing error that would have been present in CLMN. This may be a contributing factor explaining why CLMN results are not as consistent as those for FBMN.

3.3.2.6. Optimization of Molecular Networking

Overall, the most important parameters to optimize for CLMN are sample concentration and gradient. The use of high concentrations where possible, compatible with the sensitivity of the mass spectrometer used, is recommended. High concentration paired with the longest practical LC duration had a desirable effect on the networks. A high collision energy also appeared beneficial. A higher rate

of fragmentation increased cosine scores indicating more accurate edges, improving the analysis of molecular families. For CLMN, the main goal is to visualize the whole chemical space, therefore a high PPC should be used to obtain a high number of nodes representing all the metabolites in the sample. As no in-depth statistical analysis is carried out with CLMN, lower quality data arising from higher PPC do not have as great an impact as they would on FBMN.

Sample concentration and gradient are also significant factors effecting FBMN topology. As the processing step in FBMN workflow can overcome problems associated with the liquid chromatography part of the analysis (e.g., overlapping peaks), the need to fully optimize these factors is reduced. The use of a higher concentration and longer gradient is still recommended. Increased liquid chromatography separation can improve FBMN's ability to resolve isomers as the difference in retention times is increased, especially for co-eluting isomers. Collision energy and precursor per cycle had the strongest effect on molecular network topology, and these should be the focus of optimization efforts, in line with the desired responses (e.g., high cosine score vs. low number of self-loop nodes). The use of a response surface Design of Experiment (DoE) could be used to optimize these two factors for the specific chemistry of the samples and aim of constructed molecular network.

As a result of using a screening DoE to generate our fractional factorial design, we could not optimize responses, as only two levels (high/low) were included for each factor (see methods). The 'response optimizer' method compares the generated worklist of acquisition parameters to the responses arising from each run and generates optimal data acquisition settings for a given set of responses. For a full optimization, a response surface DoE should be used to optimize the four significant factors (Concentration, Gradient, Collision Energy and Precursor per Cycle) chosen from the screening design.

3.4. Materials and Methods

The overall workflow of the main experiment is summarized in Figure 13.

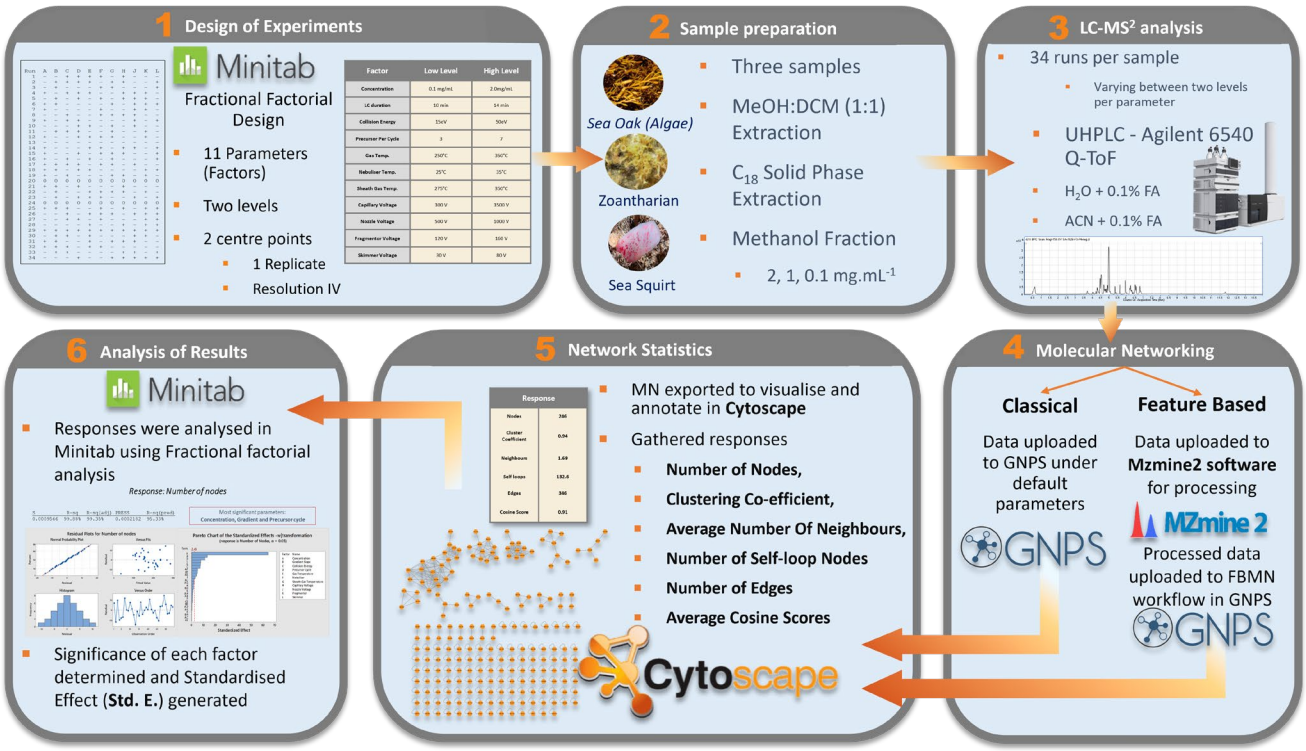


Figure 13. Workflow for evaluating the effect of data acquisition parameters on molecular network topology. Minitab fractional factorial design was used to design an experiment to determine whether a given parameter (factor) significantly affects molecular network topology. Samples were extracted with Methanol: Dichloromethane (1:1) and fractionated on a C18 SPE cartridge. A methanolic fraction of each sample was analysed 34 times on a UHPLC-Agilent 6540 q-TOF using unique parameter setting for each analysis. Each parameter had a high and low setting. Molecular networks were generated from untargeted LC-MS2 data using both the Classical and Feature-Based workflows offered by the GNPS online platform. Data pre-processing for the FBMN workflow was completed using MZmine 2. Molecular networks from both workflows were visualised in Cytoscape and responses were analysed in Minitab fractional factorial analysis to determine the Standardised Effect (Std. E.) of each parameter.

3.4.1. Sample Selection and Preparation

Three samples were selected for the main experiment: a sea squirt *Ascidia virginea*, which contained a range of terpene-derived quinone compounds; a zoantharian, *Parazoanthus axinellae*, containing a range of aromatic alkaloids, which were the most polar metabolites of the three samples; and the macroalga, *Halidrys siliquosa*, containing meroterpenoids, the least polar metabolites of the three samples (Figure 1).

For all three samples, 1.0 g of dried biomass was ground using a ball mill and extracted with MeOH/CH_2Cl_2 (1:1) under ultrasonification. The extract was fractionated using a 6cc RP-C18 Solid-phase extraction (SPE) cartridge into 4 fractions of decreasing polarity; 100% H_2O, 1:1 H_2O: MeOH, 100% MeOH, 1:1 MeOH: DCM using 15 mL of each solvent mixture. The fractions were dried and dissolved in DMSO at a concentration of 10 mg/mL. The 1:1 H_2O:MeOH and MeOH fractions were combined (1 mL of each) and diluted to obtain the concentrations used in the experimental design.

3.4.2. Experimental Design

Minitab® Statistical Software (Minitab, LLC, Sate College, PA, USA, (2019)) was used to design a resolution IV Fractional Factorial screening experiment, using the Design of Experiment (DoE) function. A single replicate was included, which is sufficient for a screening design. A single replicate design allows the significant factors to be discerned but does not allow for definitive conclusions on factor

effects, or for those factors to be optimized. However, since our aim was to discover which factors should be prioritized for optimization, this fractional factorial design (screening DoE) was adequate.

Eleven parameters (factors) were selected in the design: concentration, LC duration, collision energy, precursor per cycle, gas temperature, nebulizer voltage, sheath gas temperature, capillary voltage, nozzle voltage, fragmentor voltage and skimmer voltage, and were each tested at two levels, one low, one high (Table S1). These factors, levels, and their center points were determined based on mass spectrometry knowledge, literature review, GNPS recommendations and Agilent recommendations. The design included two experimental runs with all factors set at the centre point of the two levels, which increases the power of the design. Centre points can be used to determine whether the response surface is linear or curved. The final design had 34 runs per sample (see Table SI1) for the settings of each run.

To evaluate the characteristics of the generated molecular networks, six responses were chosen: number of nodes in the network, average number of neighbours, number of self-loop nodes, number of edges, cluster coefficient, and average cosine score. The number of nodes represents the number of metabolites represented by the network. Clustering co-efficient, average number of neighbours and number of edges are indicators of how molecular families are represented in the network. Average cosine scores are determined by the similarity of matches made between nodes in clusters. An increase in self-loop nodes is related to a decrease in clustering co-efficient and a decrease in number of edges, resulting from acquisition parameters that are sub-optimal. Alternatively, lowest intensity ions undergoing fragmentation produce spectra with higher noise ratios, and this noise can be represented by self-loop nodes.

3.4.3. Data Acquisition LC-MS²

High Resolution Electrospray Ionization Mass Spectrometry (HRESIMS) data were obtained from a Q-ToF Agilent 6540 in ESI (+) coupled to an Agilent 1290 Infinity II ultra-high performance liquid chromatography system (UHPLC), using a BEH C_{18} 2.1 × 75 mm 1.7 µm column (Acquity, Waters, Milford, CT, USA). Mobile phases of H_2O (A) + 0.1% FA and CH_3CN (B) + 0.1% FA were used with a flow set to 0.5 mL/min. An injection volume of 5 µL of sample was used for all LC-MS² experiments. The following gradient was applied: isocratic hold of **B** at 10% for 2 min followed by increase of **B** to 100% over a 10 or 14 min (LC duration specified for that run), then an isocratic hold of **B** at 100% for 4 min and a final decrease of **B** to 10% over 1 min. A 2-min post-run after each injection for equilibration. Other parameters for a standard LC-MS² experiment were set to low or high values as specified in the fractional factorial experimental design (Table S1).

3.4.4. File Conversion

LC-MS² data were converted from. d (Agilent data format) to .mgf using MS convert, part of the ProteoWizard software package [22].

3.4.5. Classical Based Molecular Networking

Molecular networks were created using the online workflow on the GNPS website (http://gnps.ucsd.edu) (accessed on 22/02/2022). The data were filtered by removing all MS² fragment ions within +/− 17 Da of the precursor m/z. MS² spectra were window filtered by choosing only the top six fragment ions in the +/− 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 2.0 Da and the MS² fragment ion tolerance to 0.5 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than six matched peaks. Further, edges between two nodes were kept in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the

molecular family size was below this threshold. 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.

3.4.6. Feature-Based Molecular Networking

LC-MS^2 data was pre-processed using MzMine2 features including feature detection, chromatogram builder, chromatogram deconvolution and isotopic peaks. Parameters used for processing can be seen in the Supplementary Materials. The processed data were uploaded to GNPS using the FBMN workflow, on the GNPS platform (https://gnps.ucsd.edu) (accessed on 22/02/2022).

3.4.7. Molecular Network Visualization and Network Analyses

Molecular networks were exported to Cytoscape software for visualization. The number of nodes, number of self-loop nodes, clustering-co-efficient and average number of neighbours were generating using the Network Analyser. Networks were treated as undirected. Number of edges and average cosine scores were generated from edge tables.

3.4.8. Design of Experiment Response Analysis

Responses (number of nodes in the network, average number of neighbours, number of self-loop nodes, number of edges, cluster coefficient, and average cosine score) were returned to Minitab for each sample, resulting in 36 response models (6 responses × 3 samples × 2 workflows). Two terms, all main effects (factor effects on responses e.g., concentration effect on cosine score) and 2-way interactions (two factors interacting to affect a response) were selected as model terms for analysing the factorial design, with no covariates. Two-sided confidence level for all intervals was set to 95%, to estimate the higher and lower values of the mean response. DoE response analysis generated Standardized Effects (Std. E.) for each factor and 2-way interaction effect on a response. Standardized effects incorporate standard deviations of observations and thus allow for the evaluation/comparison of the size of various factor effects that have different units on responses.

DoE response analysis in Minitab generates multiple outputs for analysing the effect of each factor/two-way interaction as well as the quality of each response model. Normal plots and half normal plots of standardized effects, and Pareto charts were used to identify and compare the relative magnitude of a factor's effect on a given response, as well as the statistical significance. A model summary was also generated for each response model, containing S, predicted R^2 (R^2 Pred.) and *p*-values which were used to determine the quality of the model. S is a measure, in terms of standard deviations, of how the data values differ from the fitted values and indicates how well the model describes the response. Lower S values indicate a model that better describes the response. Predicted R^2 is a measure of how well the model can predict a response; higher R^2 Pred. indicates better predicative ability. The *p*-value is the probability that the null hypothesis (i.e., factor has no effect on responses) is true. Residual plots were generated to detect issues with regression. S values in combination with residual plots were analysed to identify and excluded any model with poor fitting and/or biased data from response analysis. Models with poor fit (*p*-value > 0.05 or R^2 pred. < 20) were not included in the response analysis.

3.4.9. Visualization of Molecular Networking

Results from DoE response analysis were exported to Microsoft Excel. Factor effect on responses was averaged across the three samples for the CLMN and FBMN workflows. Bar charts were generated to summarize the Standardized Effect of each factor for both workflows. Four- and six-dimensional radar maps were used to illustrate PPC, collision energy, concentration, and LC durations effect on CLMN and FBMN responses respectively.

3.5. Conclusions

When applied to natural products, molecular networking is an incredibly versatile tool, that can be used to help answer a variety of questions. What makes a "good" molecular network depends on the purpose of the molecular network, and the most appropriate workflow, FBMN or CLMN, depends on the question. For a quick analysis of chemodiversity and dereplication, for example, to help prioritize samples for in-depth chemical analysis, CLMN would be preferable, due to the time efficient analysis of the workflow. For in depth analysis, such as metabolomic analyses or samples with known isomers, FBMN is preferable due to its ability to quantify metabolites, resolve isomers/stereoisomers and its increased reproducibility. FBMN could also be preferable when matching data against databases for dereplication as it holds more information allowing for more accurate comparison.

Once the appropriate workflow is selected, the next step is to determine the optimal responses. To determine whether there are analogues of a known bioactive compound in a sample, high number of nodes, edges, and cluster coefficient should be the responses selected for optimization. If the aim were to compare samples, where statistical accuracy and reproducibility are of high importance, a high cosine score and low number of self-loop nodes should be optimized.

Preliminary studies were necessary to recognize the need for optimization of molecular networking. The use of statistical tools and software, such as Minitab's screening DoE feature, is a time-saving and effective way to determine the significance of multiple factors and their interactions.

Mass spectrometry data acquisition parameters have a significant effect on the network topology and interpretation, with the most significant parameters shown to be concentration, LC duration, collision energy and number of precursors per cycle. When correctly used and interpreted, molecular networking can substantially speed up the dereplication of samples and provides a visual representation of sample components.

3.6 Supplementary Materials and Data Availability Statement

The following are available online at https://www.mdpi.com/2218-1989/12/3/245/s1?version=1647396998, LC-MS^2 chromatograms of samples (Figures S1–S3). DoE reports for three samples and two workflows (S4-S39. Data acquisition parameters (Table S1). Molecular networking responses (Tables S2–S7), Statistical analysis of response model (TableS8). The data presented in this study are openly available in GNPS-MassIVE Datasets at doi:10.25345/C5FN10W04.

3.7 Reference

1. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B.S.; Yang, J.Y.; Kersten, R.D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J.M. Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1743–E1752.

2. Guo, J.; Huan, T. Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics. *Anal. Chem.* **2020**, *92*, 8072–8080.

3. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.

4. Allard, P.-M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V.C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Anal. Chem.* **2016**, *88*, 3317–3323.

5. Gao, Y.L.; Wang, Y.J.; Chung, H.H.; Chen, K.C.; Shen, T.L.; Hsu, C.C. Molecular networking as a dereplication strategy for monitoring metabolites of natural product treated cancer cells. *Rapid Commun. Mass Spectrom.* **2020**, *34*, e8549.

6. Perez De Souza, L.; Alseekh, S.; Brotman, Y.; Fernie, A.R. Network-based strategies in metabolomics data analysis and interpretation: From molecular networking to biological interpretation. *Expert Rev. Proteom.* **2020**, *17*, 243–255.

7. Shishido, T.K.; Popin, R.V.; Jokela, J.; Wahlsten, M.; Fiore, M.F.; Fewer, D.P.; Herfindal, L.; Sivonen, K. Dereplication of natural products with antimicrobial and anticancer activity from Brazilian cyanobacteria. *Toxins* **2020**, *12*, 12.

8. Duncan, K.R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B.S.; Dorrestein, P.C. Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from Salinispora species. *Chem. Biol.* **2015**, *22*, 460–471.

9. Paulo, B.S.; Sigrist, R.; Angolini, C.F.; De Oliveira, L.G. New cyclodepsipeptide derivatives revealed by genome mining and molecular networking. *ChemistrySelect* **2019**, *4*, 7785–7790.

10. Raheem, D.J.; Tawfike, A.F.; Abdelmohsen, U.R.; Edrada-Ebel, R.; Fitzsimmons-Thoss, V. Application of metabolomics and molecular networking in investigating the chemical profile and antitrypanosomal activity of British bluebells (Hyacinthoides non-scripta). *Sci. Rep.* **2019**, *9*, 2547.

11. Silva, E.; da Graça, J.P.; Porto, C.; Martin do Prado, R.; Hoffmann-Campo, C.B.; Meyer, M.C.; de Oliveira Nunes, E.; Pilau, E.J. Unraveling Asian Soybean Rust metabolomics using mass spectrometry and Molecular Networking approach. *Sci. Rep.* **2020**, *10*, 138.

12. De Vijlder, T.; Valkenborg, D.; Lemière, F.; Romijn, E.P.; Laukens, K.; Cuyckens, F. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrom. Rev.* **2018**, *37*, 607–629.

13. Frank, A.M.; Monroe, M.E.; Shah, A.R.; Carver, J.J.; Bandeira, N.; Moore, R.J.; Anderson, G.A.; Smith, R.D.; Pevzner, P.A. Spectral archives: Extending spectral libraries to analyze both identified and unidentified spectra. *Nat. Methods* **2011**, *8*, 587–591. https://doi.org/10.1038/nmeth.1609.

14. Yang, J.Y.; Sanchez, L.M.; Rath, C.M.; Liu, X.; Boudreau, P.D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A. Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013**, *76*, 1686–1699.

15. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **2020**, *17*, 905–908.

16. Pluskal, T.; Castillo, S.; Villar-Briones, A. Oreši c M. MZmine: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, 395.

17. Xu, R.; Lee, J.; Chen, L.; Zhu, J. Enhanced detection and annotation of small molecules in metabolomics using molecular-network-oriented parameter optimization. *Mol. Omics* **2021**, *17*, 665–676.

18. Czitrom, V. One-factor-at-a-time versus designed experiments. *Am. Stat.* **1999**, *53*, 126–131.

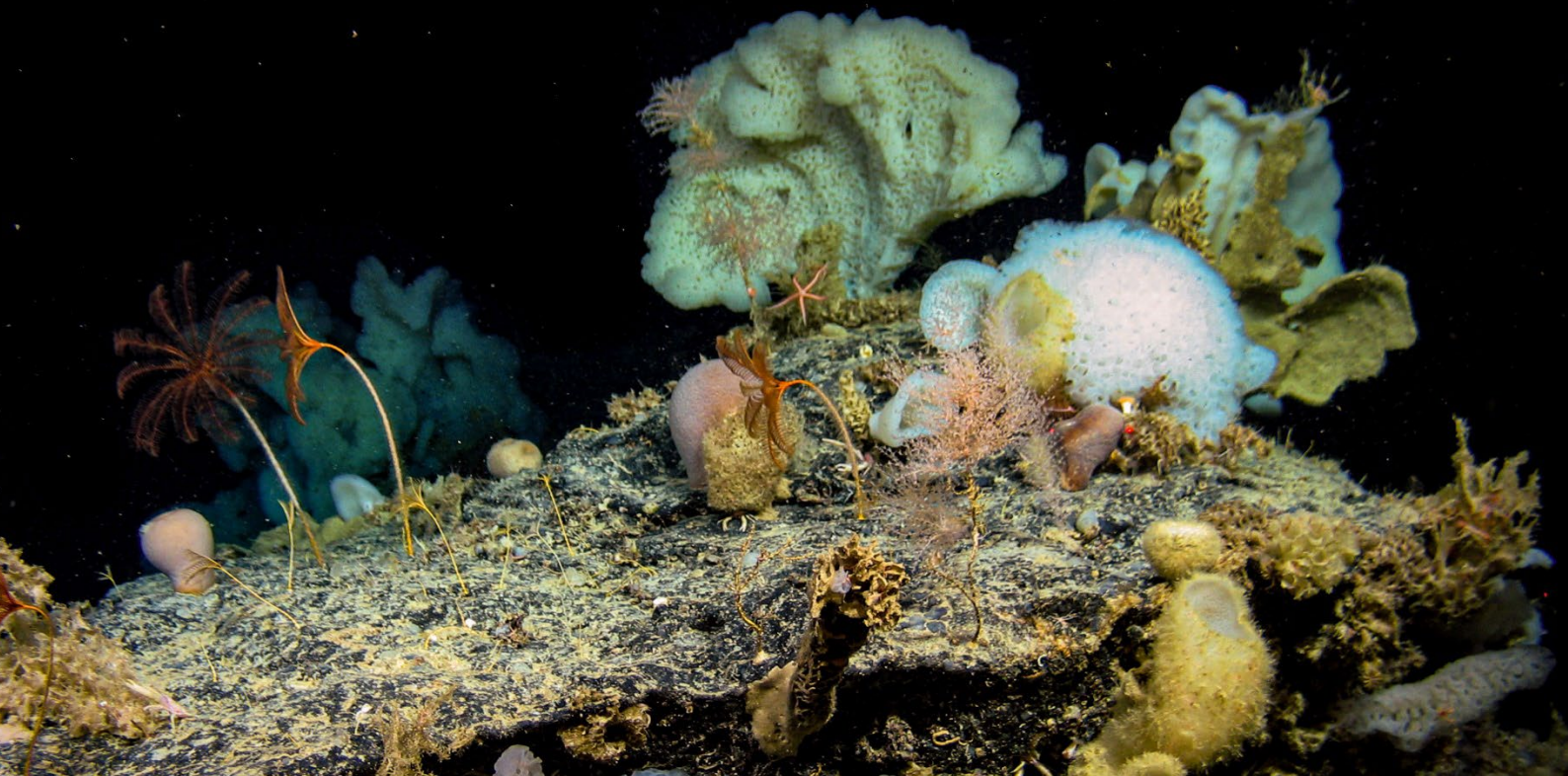
19. Hecht, E.S.; Oberg, A.L.; Muddiman, D.C. Optimizing mass spectrometry analyses: A tailored review on the utility of design of experiments. *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 767–785.

20. Olivon, F.; Roussi, F.; Litaudon, M.; Touboul, D. Optimized experimental workflow for tandem mass spectrometry molecular networking in metabolomics. *Anal. Bioanal. Chem.* **2017**, *409*, 5767–5778.

21. Balsam, A. *Optimization of Molecular Networking for Marine Natural Products Chemistry*; NUI Galway: Galway, *Ireland,* 2020.

22. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534–2536. https://doi.org/10.1093/bioinformatics/btn323.

Chapter 4. Unveiling the Chemical Diversity of the Deep-sea Sponge *Characella pachastrelloides*



Chapter 4. Unveiling the Chemical Diversity of the Deep-sea Sponge *Characella pachastrelloides*

Published: 5th January 2022 *Marine Drugs* **2022**, *20*(1), 52

DOI: 10.3390/md20010052

4.1 Abstract

Sponges are at the forefront of marine natural product research. In the deep sea, extreme conditions have driven secondary metabolite pathway evolution such that we might expect deep-sea sponges to yield a broad range of unique natural products. Here we investigate the chemodiversity of a deep-sea tetractinellid sponge, *Characella pachastrelloides*, collected from ~800 m depth in Irish waters. First, we analysed the MS/MS data obtained from fractions of this sponge on the GNPS public online platform to guide our exploration of its chemodiversity. Novel glycolipopeptides named characellides were previously isolated from the sponge and herein cyanocobalamin, a manufactured form of vitamin B_{12}, not previously found in nature, was isolated in large amount. We also identified several poecillastrins, a class of polyketide known to exhibit cytotoxicity, from the molecular network. Light sensitivity prevented the isolation and characterisation of these polyketides, but their presence was confirmed by characteristic NMR and MS signals. Finally, we isolated the new betaine 6-methylhercynine; which contains a unique methylation at C-2 of the imidazole ring. This compound showed potent cytotoxicity towards against HeLa (cervical cancer) cells.

Table 1: Authors and their contributions.

Authors	Conceptualization	Methodology	Software	Formal Analysis	Investigation	Bioassay	Data Curation	Resources	Original Draft	Review And Editing	Visualization	Funding Acquisition	Supervision	Project Administration
Sam Afoullouss	X	X	X	X			X		X		X			
Anthony R. Sanchez						X								
Laurence K. Jennings		X	X								X			
Younghoon Kee						X						X	X	X
A. Louise Allcock	X				X			X		X		X	X	X
Olivier P. Thomas	X	X			X			X		X		X	X	X

4.2 Introduction

Deep-sea habitats at 200 m – 2,000 m on the continental margin may be highly diverse, especially in submarine canyon systems, where hydrography concentrates food resources [1]. Such habitats are rich in sponges and corals [2], taxa whose holobionts represent the most promising sources of bioactivity [3]. Sponges alone have yielded nearly 50% of the marine natural products between 1990 and 2009 [4], only recently being surpassed by microbes [5]. Several factors promote the evolution of novel chemical architectures in deep-sea sponges: sessile organisms may evolve secondary metabolites that act as chemical defenses against biofouling and predation [6], while extreme environmental conditions can drive adaptations in biochemical pathways [3, 7]. However, there are known difficulties in collecting deep-sea samples, and of the ~9,620 compounds isolated from sponges, only ~290 were isolated from specimens collected from below 200 m [8].

In our quest for bioactive compounds from Irish deep-sea sponges, we focused on a tetractinellid sponge *Characella pachastrelloides,* collected in Whittard Canyon, one of the largest submarine canyon systems in the NE Atlantic. We previously identified and described four unique glycolipopeptides named characellides with anti-inflammatory properties from this sponge [9, 10]. The order Tetractinellida is known for its rich diversity of bioactive compounds. To date, the natural product chemistry of few deep-water tetractinellids has been investigated, but those few have yielded a variety of natural products including bisindole alkaloids [11], cytotoxic peptide lactones [12], macrolides [13, 14] and other polyketides [15], steroidal saponins [16], and cytotoxic linear acetylenes [17], suggesting this group has potential for yielding new bioactive metabolites.

Where species are chemodiverse, as seen in tetractinellids, molecular networking can aid in distinguishing and identifying the different families of natural products present. Molecular networking compares MS/MS spectra of metabolites with known or calculated MS/MS spectra and provides a visual representation of compound similarity, clustering structurally similar natural products together [18]. Molecular networking can aid in dereplication but can also help identifying multiple closely related compounds and analogs occurring in a complex mixture, allowing the components of extracts and fractions to be more easily discerned. We applied molecular networking to our MS/MS data and revealed clusters of diverse families of natural products that we explored in depth. Herein, we report on the presence of characellide analogs 1-4, polyketide poecillastrins 5 and 6, cyanocobalamin 7, and a new histidine derivative, 6-methylhercynine (8).

4.3 Results and Discussion

4.3.1 Molecular Networking

To inspect the chemical diversity of the deep-sea sponge *Characella pachastrelloides*, we first built molecular networks through the online Global Natural Products Social (GNPS) platform [18]. The extract prepared from *C. pachastrelloides* using a mixture of solvents CH_2Cl_2/MeOH (1:1) was fractionated using C_{18} Solid Phase Extractions (SPE) into five fractions of decreasing polarity from H_2O to MeOH and then CH_2Cl_2. All, but the first aqueous fraction, were analyzed by UHPLC-HRMS/MS on a Q-ToF instrument. MS/MS data from each analysed fraction were processed and combined to produce one molecular network for *Characella pachastrelloides*. Molecular networks were generated using both the *classical molecular networking* (CLMN) and the *feature based molecular networking* (FBMN) workflows. CLMN showed the presence of three large clusters (>7 nodes) and eighteen small clusters (2-7 nodes). 29 nodes were annotated against GNPS libraries but were comprised of compounds of low interest including. fatty acid derivatives, compounds synthetic origin, and dipeptides (see SI). FBMN was preprocessed using MZmine2 and produced a network with three large clusters and seven small clusters, with only six nodes annotated against GNPS library spectra. These hits include the same fatty acid derivatives annotated in the CLMN.

Dereplication was carried out manually, comparing predicted molecular formulae against marine natural product libraries such as MarinLit. Manual annotation was added by *node MS2 peaks visualizer* tool available on the online network analyzer, to find nodes sharing fragment peaks. The CLMN was visualized using Cytoscape (Figure 1).

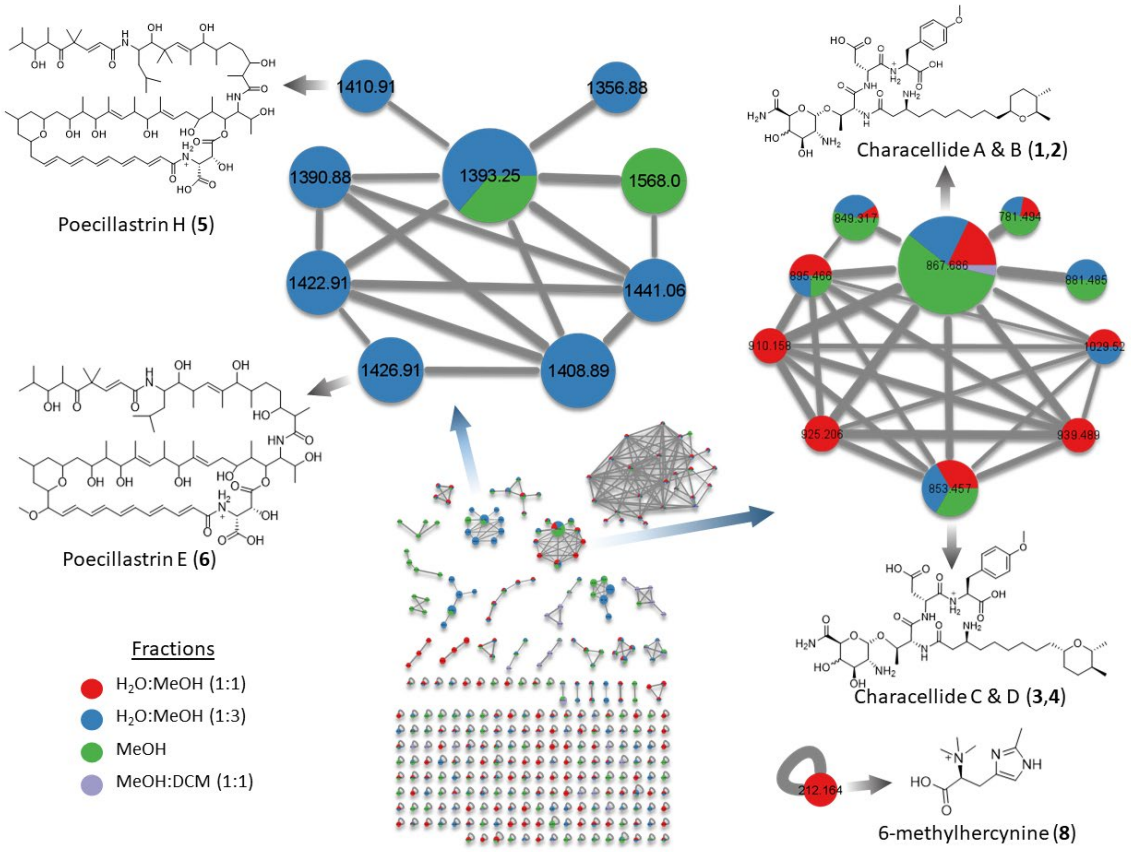


Figure 1. Molecular network of Characella pachastrelloides fractions, with annotated metabolites. Pie charts indicate metabolites distribution in fractions (sum precursor ion intensity). Size of node is relative to precursor ion intensity. Edge width increases with higher cosine score.

4.3.2 Glycolipopeptides: characellides

A first cluster contained 10 nodes with molecular masses between *m/z* ~853 and 1029 including the four previously described novel glycolipopeptides, characellide A-D (1-4) [9]. Further analysis of this cluster in FBMN, reveals the presence of two nodes, 854 and 868 not present in the CLMN. Interrogation of MS/MS spectra of both nodes, suggest there are analogs of characellide A and C, replacing a sugar unit containing two nitrogen m/z 175.07 (A_α) for sugar unit containing one nitrogen *m/z 176.06*(A_α) (figure S1).

Characellide B has since been synthesized [19], showing that the configurations of this compound differ from those of the proposed structure of the natural product. Four isomers synthesized with various $_L$ and $_D$ peptides were not in agreement with NMR data of the isolated natural product, leaving the configuration between the peptidic chain and the sugar unit the most likely source of the differences observed. An in-depth analysis of the MS/MS spectra between annotated and unannotated nodes allowed us to hypothesize the structure of other characellides C and D (3-4) in the cluster. Isolation of the new characellides was attempted but due to low quantities and the difficulty of sample collections, these attempts were unsuccessful.

4.3.3 Polyketides: poecillastrins

A second cluster of the molecular network, was only detected in CLMN, contained analogues of chondropsin macrolide lactams such as poecillastrin H (**5**) that was recently isolated from a deep-sea sponge**,** Characella sp. [20]. From this annotated node, we also identified the presence of poecillastrin E (**6**), which has been isolated from a deep-sea tetractinellid sponge [13]. The presence of poecillastrin H was confirmed by characteristic ^{1}H-NMR signals at δ_H 7.23 and doublet at δ_H 6.94 matching those found in the literature. Additional characteristic UV absorption at 370 nm was observed. Chondropsins are known to have potent cytotoxicity caused by selective inhibition of V-ATPases [21]. The presence

of unidentified nodes, with a high ratio of edges per node, within this cluster suggested the presence of undiscovered analogs. Retrieving the nodes' retention times and molecular weights from the molecular networking allowed us to carry out a targeted purification of these macrolides. Due to their extreme photosensitivity, the chemical structure of theses analogs could not be determined. These compounds may be only stable in the absence of light as in their natural environment.

4.3.4 Cyanocobalamine

Compound **7** was isolated (4.4 mg) as a pink amorphous solid. A molecular formula of $C_{63}H_{88}CoN_{14}O_{14}P$ was revealed using HRESIMS with m/z 1,355.5719 $[M+H]^+$ (Figure 2). Cyanocobalamin was identified in the literature with this exact molecular formula (CN-B_{12}). This was confirmed by comparing HSQC spectra of compound **7** and a cyanocobalamin standard by NMR.

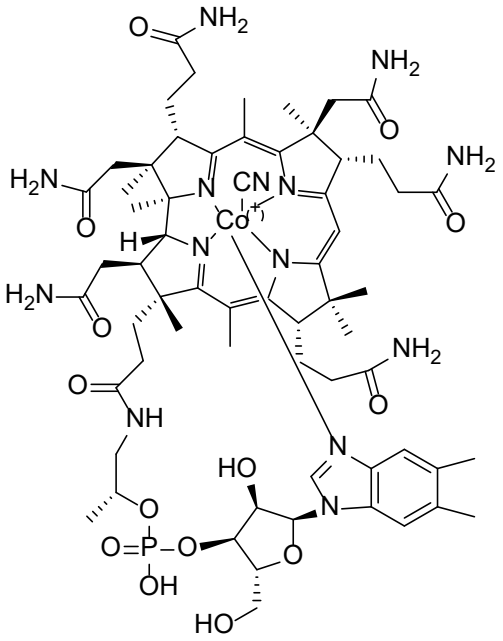
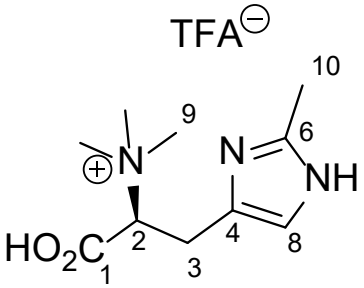


Figure 2. Structure of cyanocobalamin (**7**).

Cobalamin or vitamin B_{12} derivatives are members of the corrinoids, characterized by their cobalt-coordinated corrin ring. There are multiple vitamers of B_{12}, with axial ligands varying from hydroxyl, methyl, cyano or 5'-deoxyadenosyl, with cyanocobalamin being the most stable. Vitamin B_{12} plays an intrinsic role in metabolic processes as a co-enzyme, making it vitally important in the medical field. It is produced by bacteria and archaea and its biosynthesis is well studied [22]. From a review of the literature, this appears to be the first time cyanocobalamin has been isolated from a natural source. Mass spectrometry analysis of the extract was carried out to determine whether any other forms of Vitamin B_{12} were present, with cobalamin found to be present in minute amounts. As it is possible for cyanocobalamin to be produced as a byproduct from others forms of B_{12}, due to the affinity of those forms for cyanide, the absence of other vitamers indicated that the cyanocobalamin was not an artefact. This leaves two possibilities for the origins of the cyanocobalamin. First, it could be produced by prokaryotes living within the sponge, potentially via a new biosynthetic pathway. This may be a result of the extreme environmental conditions of the deep sea (pressure, temperature etc) encouraging the production of the more stable cyanocobalamin over the usually produced hydroxocobalamin. Since biosynthesis of Vitamin B_{12} is limited to bacteria and archaea, the presence of cyanocobalamin, indicates that deep-sea sponges, just like their shallow-water counterparts, host a productive microbial community. Second, cyanocobalamin could be accumulated within the sponge, aided by its filter feeding nature. Interestingly, in the marine world, the major producer of Vitamin B_{12} is Thaumarchaeota [23].

4.3.5 Betaine: 6-methylhercynine

Compound **8** was isolated as a white amorphous powder. Its molecular formula was determined by HRESIMS to be $C_{10}H_{17}N_3O_2$ with *m/z* 212.1394 for $[M+H]^+$ (Figure 3). ^{1}H-NMR data revealed the presence of four methyl singlets at δ_H 3.30 (s, 9H, H-9), 2.71 (s, 3H, H-10), one aromatic methine at δ_H 7.27 (s, 1H, H-8), and an ABX system at δ_H 3.97 (dd, 1H, *J* = 12.0, 3.5 Hz, H-2), 3.44 (dd, 1H, *J* = 14.0, 3.5 Hz, H-3a) and 3.37 (dd, 1H, *J* = 14.0, 12.0 Hz, H-3b) (Table 2). HSQC and HMBC spectra of **8** indicated resonances associated with three non-protonated carbons: a carboxylic acid at δ_C 170.2 (C-1) and two aromatic carbons at δ_C 144.6 (C-6) and 128.3 (C-4). The only spin coupled system (SCS) was assigned through H-2/H-3a and H-3b COSY correlations. The H-3/C-1 and H-9/C-2 HMBC correlations established the presence of an *N*-trimethylated amino acid derivative. The connection between the aromatic ring and the SCS was evidenced by key H-3a and 3b/C-4 and C-8 HMBC correlations. The chemical shifts of the aromatic signals at C-4, C-6 and C-8 suggested the presence of an imidazole ring but one aromatic proton was missing in the ^{1}H NMR spectrum. The fourth methyl at C-10 was located on the aromatic ring at C-6 due to a H-10/C-6 HMBC correlation. H-10 had only one correlation in HMBC confirming the presence of a substituted imidazole ring.



6-methylhercynine (**8**)

Figure 6. Structure of the new histidine derivative 6-methylhercynine (**8**).

Table 2. NMR spectroscopic data for compound **8** in D_2O (500 MHz for ^{1}H-NMR data and 125 MHz for ^{13}C-NMR).

No.	6-methylhercynine (8) δ_H, mult. (J in Hz)	δ_C
1	-	169.5
2	3.98, dd (12.0, 3.5)	76.5
3	3.46, dd (14.0, 3.5) 3.36, dd (14.0, 12.0)	22.6
4	-	127.7
6	-	144.0
8	7.29, s	118.8
9	3.32, s	52.2
10	2.73, s	15.3

The absolute configuration at C-2 was determined by comparison between experimental and calculated electronic circular dichroism (ECD) spectra. After a conformational analysis and geometry optimization, the ECD spectra of the two possible enantiomers of **8** were calculated using time-dependent density functional theory (TDDFT). The electronic transition and rotational strength calculations were conducted at the B3LYP/6-11G(d,p) level with 50 transition states. The charge state of the molecule was predicted to be +1 due to the use of acidic conditions during RP-HPLC purification. However, the neutral zwitterion and doubly charged ion were also calculated for comparison with experimental data. Spectra were Boltzmann weighted based on a free-energy distribution and corrected with the UV data. In the calculated spectra of the (2*S*) enantiomer for all charged states a positive Cotton effect at 215 nm that matches the experimental spectra of **8** was observed allowing the assignment of (*S*)-6-methylhercynine. Moreover, the experimental spectra closely matches previously published experimental spectra of L-histidine and *S*-hercynine, further supporting our TDDFT assignment [24, 25].

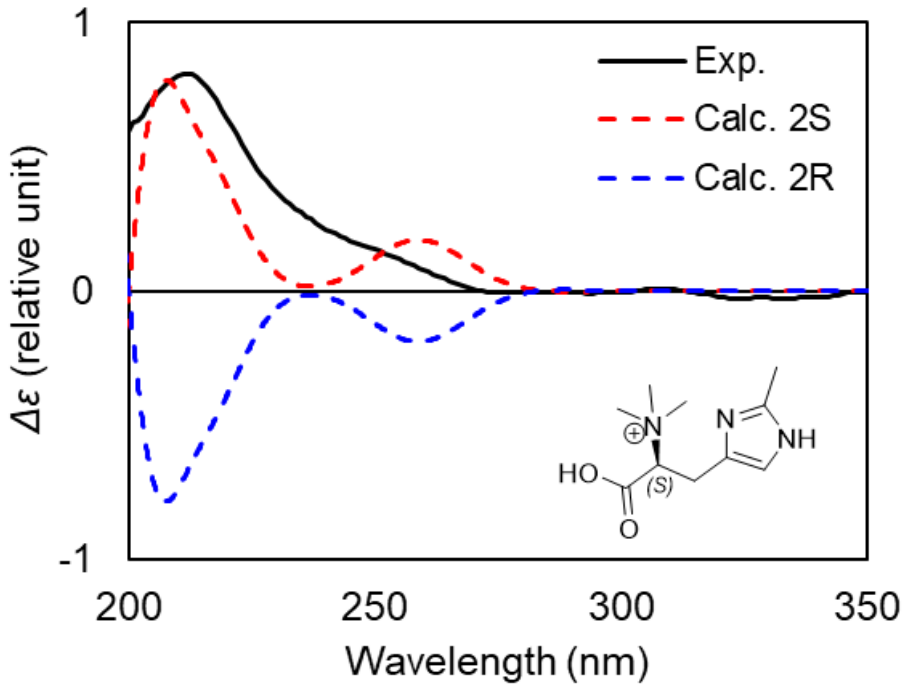
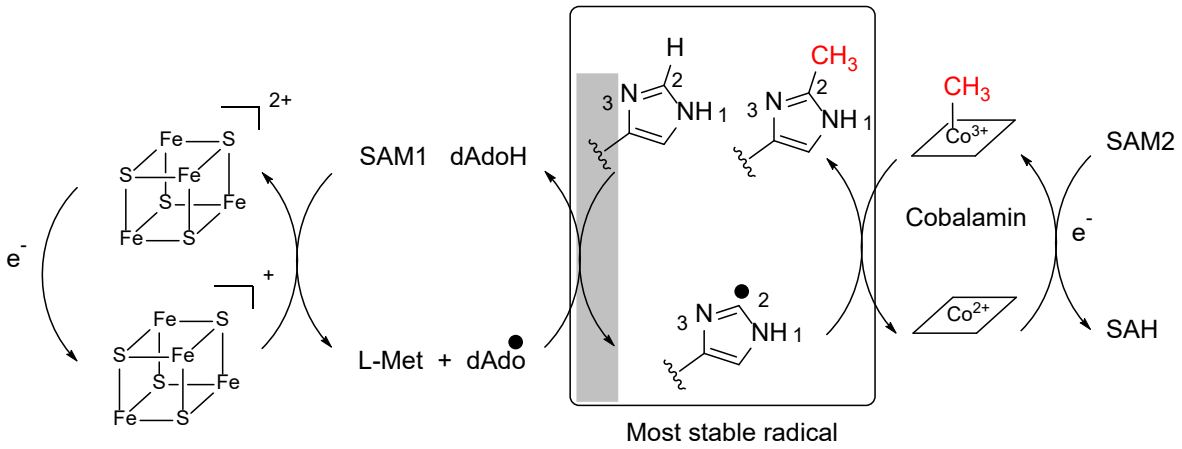


Figure 4. Comparison between the experimental and calculated ECD spectra for two enantiomers of compound **8.**

This new betaine is a methylated derivative at C-6 of hercynine and, to the best of our knowledge, it is the first natural product where an imidazole is methylated at this position. Indeed, positions *N*-1 and *N*-3 are the most reactive nucleophilic centers for an imidazole and position C-2 is rarely substituted. Searching in the literature, we could only identify synthetic compounds with a methyl at C-2 of an imidazole ring. They were obtained by chemical synthesis using radicalar conditions with

silver salts [26, 27]. Hercynine, the betaine corresponding to histidine, is also described as an intermediate in the biosynthesis of the "longevity vitamin" ergothioneine which possesses a sulfur at position C-2 of the imidazole [28]. Here again, the mechanism was shown to be radicalar. Radical-mediated enzymatic methylation has been well studied and involves radical SAM methyltransferases (RSMTs) of three different classes [29]. Interestingly, class B methyltransferases use cobalamin as a conduit of the methyl group to the substrate [30]. From the sponge, we were able to isolate good amounts of cobalamin and cyanocobalamin which is in accordance with the presence of this class of enzyme in the holobiont. As shown in Scheme 1, two molecules of *S*-Adenosyl Methionine would therefore be necessary to methylate the most stable radical at position C-2.



Scheme 1. Proposed metabolic pathway for **8.**

Compound **8** was assessed for cytotoxicity, using clonogenic assays, against HeLa (cervical cancer) cells and showed a significant bioactivity with a LC_{50} of 3.5 μM.

4.4 Materials and Methods

4.4.1 General Experimental Procedures

Optical rotations were measured on a UNIPOL 1000 Polarimeter. UV and ECD measurements were obtained on a Chirascan (Applied Photophysics) spectrophotometer. NMR experiments were measured on a 600 MHz equipped with a cryoprobe (Varian) or 500 MHz (Agilent). Chemical shifts (δ in ppm) are referenced to trace methanol (δ_H 3.34, δ_C 49.5) for NMRs in D_2O and residual proton and carbon signals (δ_H 3.31, δ_C 49.0) for NMRs in MeOH-D_4. High Resolution Electrospray Ionization Mass Spectrometry (HRESIMS) data were obtained from a Q-ToF Agilent 6540 in ESI(+). Preparative HPLC was preformed using a Jasco PU-2087 Plus equipped with a UV-Vis detector UV 2075 Plus and then by Agilent 1260 analytical HPLC series equipped with a DAD detector.

4.4.2 Biological Material

The specimen was collected from a depth of 809 m (48.6509° N, 10.4846° W) by the remotely operated vehicles Holland 1 during the CE16006 cruise of the RV Celtic Explorer. The sponge appeared as a white barrel sponge. In-situ pictures were taken to aid identification. All epibionts were removed and a small section was stored in 96% ethanol as a voucher specimen. The rest of the biomass was lyophilized and stored at -20 °C.

4.4.3 Extraction and Isolation

The sponge (330 g) was ground using a ball mill and extracted with a mixture MeOH-CH_2Cl_2 (1:1) and ultrasonification. The crude extract (20.6 g) was fractioned using RP-C_{18} Vacuum Liquid Chromatography (VLC) into five fractions of decreasing polarity from F1(H_2O), F2(H_2O: MeOH (1:1)), F3 (H_2O: MeOH (1:3)), F4 (MeOH) and F5 (MeOH:CH_2Cl_2 (1:1)).

The water-methanol fraction (942.2 mg) was separated using RP-HPLC on a semi-preparative T3 column, 250 mm × 19 mm, 5 µm (Xselect, Waters, Milford, CT, USA), using a mobile phase of water (A) and methanol (B), both with 0.1% TFA and a flow rate of 5 mL/min. The gradient method was developed with a 33 min run time: isocratic at 8% B for 5 min, a liner gradient for 18 mins to 65% B, isocratic at 65% for 5 mins and returning to 8% B for 5 minutes. Compound 7 (RT 13.0 min, 2.87 mg) was collected at sufficient purity and compound 8 was collected as a mixture (52 mg). Repurification was carried out on a Waters 2695 HPLC using a semi-preparative reversed-phase amide column (Waters analytical BEH column 5 µm 10 mm x 250 mm) with 5 $mL.min^{-1}$ flow rate and injections ranging from 10 µL 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, 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. Compound 8 (RT 8.5 min; 2.10 mg) was obtained in enough purity for structural elucidation using NMR.

Cyanocobalamin (7): Pink amorphous solid; UV (DAD) λ_{max} 360 nm & 270 nm; HRESIMS (+) m/z 1355.5719 $[M+H]^+$ (calcd. For $C_{63}H_{89}CoN_{14}O_{14}P$, 1355.5752,Δ -2.1 ppm)

6-methylhercynine (8): White amorphous solid; $[\alpha]_{20}^{D}$ +42°; UV (DAD) λ_{max} 254 nm; ^{1}H-NMR and ^{13}C NMR, see Table 1; HRESIMS (+) m/z 212.1394 [M+H]+ (calcd. For $C_{10}H_{18}N_3O_2$, 212.1394, Δ 0.0 ppm)

4.4.4 Computational Methods

A conformational analysis of **8** was performed using a Monte Carlo Minimum method (MCMM) and the molecular mechanics OPLS3 force field with an energy cut off of 5 kcal/mol in Schrodinger MacroModel [31]. These conformers were then optimized using DFT, at the M06-2X/6-31G(d,p) level in Gaussian 16 , with the zero-point energy, electronic transition, and rational strength also calculated [32]. Following this, the ECD spectra for each conformer were calculated in Gaussian 16 at the B3LYP/6-311G(d,p) level. All DFT calculations were performed using a polarizable continuum solvation model [33]. The final ECD spectra were extracted, Boltzmann weighted and corrected by alignment with the UV spectra using the freely available software SpecDis 1.7 (version 1.71, SpecDis, Berlin, Germany) [34].

4.4.5 Molecular Network

LC-MS/MS data was acquired in Data dependent acquisition (DDA) mode on a HRESIMS- Q-ToF Agilent 6540 in ESI(+), using a C_{18} column (Xselect, Waters, Milford, CT, USA) with a mobile phase of water (A) and acetonitrile (B), both with 0.1% formic acid (FA) and a flow rate of 0.5 mL/min. The gradient method was developed with a 18 min run time: isocratic at 10% B for 2 min, a liner gradient for 10 mins to 100% B, isocratic at 100% for 3 mins and returning to 10% B over 1 minutes and remaining isocratic at 10% B for 2 minutes. Parameters, conditions and spectra, used in a LC-MS/MS data acquisition are available in table S1.

Molecular networks were created using the Feature-Based Molecular Networking (FBMN) workflow (on the GNPS platform (https://gnps.ucsd.edu). The raw mass spectrometry data was converted to .mzXML using Proteowizard (version 3.18212). MS data was subsequently processed in MZmine2 using its MS peak detection, chromatogram builder, chromatogram deconvolution, isotopic grouping, and feature alignment tools. The resulting spectra were manually validated to ensure all spectra were processed correctly. The feature quantification table and MS/MS spectral summary were exported to GNPS feature based workflow for analysis.

Data filtering was carried out by removing MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. Mass tolerance for the precursor ion and the MS/MS fragments was set at 0.05 Da. The molecular networks were then created where edges appeared in the network if more than 6 peaks were matched and awarded a cosine of above 0.7. Furthermore, edges between two nodes were only kept if each of the nodes appeared in the others top 10 most similar nodes. The maximin size for a molecular family was set to 100 (the lowest scoring edges were removed from clusters until size of the molecular family fell below the set threshold). The uploaded spectra were compared against GNPS spectral libraries. GNPS libraries were filtered using the same conditions as

the input data. Matches between input data and library spectra were shown if they had a score of above 0.7 and at least 6 matched peaks. After file conversion using Proteowizard (version 3.18212), a molecular network was generated using the GNPS platform. The molecular network was visualized via Cytoscape (version 3.9.1).

Feature-based molecular network:
https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=18fcc78f96fd40f189b99a46fedf77b3
Classical molecular network:
https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5785a70c1741419793a5898079ebf64d

4.4.6 Biological Activities

Clonogenic survival assay

HeLa cells were seeded into 96-well plates at a concentration of 200 cells/mL in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Bovine serum. Drugs were added to the medium 24 hours after seeding at the indicated concentrations, and the cells were allowed to grow for 10 days. The cells were fixed with a 10% methanol, 10% Acetic acid solution (in water) for 15 min at room temperature and stained with 1% crystal violet (in methanol) for 5 minutes. Excess dye was removed with water and the plates were allowed to dry at RT overnight. Cells were de-stained with Sorenson's buffer (0.1 M sodium citrate, 50% ethanol). The colorimetric intensity of each solution was quantified using Gen5 software on a Synergy 2 (BioTek, Winooksi, VT) plate reader (OD at 595nm).

4.5 Conclusions

Molecular networking proved to be a useful tool to explore the chemodiversity of the deep-sea tetractinellid sponge *Characella pachastrelloides*. First, the molecular network allowed us to discern characellide analogs present in too small quantities, and highly light sensitive poecillastrins. The discovery of a uniquely methylated derivative of hercynine, which proved active against a HeLa cell line, alongside cyanocobalamin, the most common synthetic form of vitamin B_{12}, raised interesting questions about its biosynthetic origin. The chemodiversity of *Characella pachastrelloides* demonstrates the potential of deep-sea sponges, particularly tetractinellids, in biodiscovery. It is likely that some of the compounds, for example the cyanocobalamin, were produced by the sponge microbiome. Given the difficulties of culturing deep-sea microbes, collecting macrofauna can be an effective method of sampling deep-sea microbial diversity.

4.6 Supplementary Materials and Data Availability Statement

The following are available online at https://www.mdpi.com/article/ 10.3390/md20010052/s1, Table S1: Conditions and parameters used from HRESIMS analysis on Characella pachastrelloides using a UPLC-Q-ToF system for DDA LC-MSMS analysis. Figures S1–S17: HRESIMS NMR and UV spectra of 5, 7 and 8. Feature-based molecular network: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=18fcc78f96fd40f189b99a46fedf77b3 accessed on 21 December 2021. Classical molecular network: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5785a70c1741419793a5898079ebf64d accessed on 21 December 2021.

4.7 References

1. Johnson, M. P.; White, M.; Wilson, A.; Würzberg, L.; Schwabe, E.; Folch, H.; Allcock, A. L., A Vertical Wall Dominated by Acesta excavata and Neopycnodonte zibrowii, Part of an Undersampled Group of Deep-Sea Habitats. *PLOS One* **2013,** 8, (11), e79917.
2. Fernandez-Arcaya, U.; Ramirez-Llodra, E.; Aguzzi, J.; Allcock, A. L.; Davies, J. S.; Dissanayake, A.; Harris, P.; Howell, K.; Huvenne, V. A. I.; Macmillan-Lawler, M.; Martín, J.; Menot, L.; Nizinski, M.; Puig, P.; Rowden, A. A.; Sanchez, F.; Van den Beld, I. M. J., Ecological Role of Submarine Canyons and Need for Canyon Conservation: A Review. *Front. Mar. Sci.* **2017,** 4, 5.

3. Pilkington, L. I., A Chemometric Analysis of Deep-Sea Natural Products. *Molecules* **2019,** 24, (21).
4. Leal, M. C.; Puga, J.; Serôdio, J.; Gomes, N. C. M.; Calado, R., Trends in the Discovery of New Marine Natural Products from Invertebrates over the Last Two Decades – Where and What Are We Bioprospecting? *PLOS One* **2012,** 7, (1), e30580.
5. Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R., Marine natural products. *Nat. Prod. Rep.* **2021,** 38, (2), 362-413.
6. Genta-Jouve, G.; Thomas, O. P., Chapter four - Sponge Chemical Diversity: From Biosynthetic Pathways to Ecological Roles. In *Advances in Marine Biology*, Becerro, M. A.; Uriz, M. J.; Maldonado, M.; Turon, X., Eds. Academic Press: 2012; Vol. 62, pp 183-230.
7. Wright, P. C.; Westacott, R. E.; Burja, A. M., Piezotolerance as a metabolic engineering tool for the biosynthesis of natural products. *Biomol. Eng* **2003,** 20, (4), 325-331.
8. Marinlit. Royal Society of Chemistry: consulted 12/07/2019.
9. Afoullouss, S.; Calabro, K.; Genta-Jouve, G.; Gegunde, S.; Alfonso, A.; Nesbitt, R.; Morrow, C.; Alonso, E.; Botana, L. M.; Allcock, A. L.; Thomas, O. P., Treasures from the Deep: Characellides as Anti-Inflammatory Lipoglycotripeptides from the Sponge Characella pachastrelloides. *Org. Lett.* **2019,** 21, (1), 246-251.
10. Marcella, S.; Afoullouss, S.; Thomas, O. P.; Allcock, A. L.; Murphy, P. V.; Loffredo, S., Immunomodulatory properties of characellide A on human peripheral blood mononuclear cells. *Inflammopharmacology* **2021,** 29, (4), 1201-1210.
11. Ragini, K.; Piggott, A. M.; Karuso, P., Bisindole Alkaloids from a New Zealand Deep-Sea Marine Sponge Lamellomorpha strongylata. *Mar. Drugs* **2019,** 17, (12).
12. Nakamukai, S.; Takada, K.; Furihata, K.; Ise, Y.; Okada, S.; Morii, Y.; Yamawaki, N.; Takatani, T.; Arakawa, O.; Gustafson, K. R.; Matsunaga, S., Stellatolide H, a cytotoxic peptide lactone from a deep-sea sponge Discodermia sp. *Tetrahedron Lett.* **2018,** 59, (26), 2532-2536.
13. Irie, R.; Hitora, Y.; Ise, Y.; Okada, S.; Takada, K.; Matsunaga, S., Poecillastrin E, F, and G, cytotoxic chondropsin-type macrolides from a marine sponge Poecillastra sp. *Tetrahedron* **2018,** 74, (13), 1430-1434.
14. Wright, A. E.; Roberts, J. C.; Guzmán, E. A.; Pitts, T. P.; Pomponi, S. A.; Reed, J. K., Analogues of the Potent Antitumor Compound Leiodermatolide from a Deep-Water Sponge of the Genus Leiodermatium. *J. Nat. Prod.* **2017,** 80, (3), 735-739.
15. Dumdei, E. J.; Blunt, J. W.; Munro, M. H. G.; Pannell, L. K., Isolation of Calyculins, Calyculinamides, and Swinholide H from the New Zealand Deep-Water Marine Sponge Lamellomorpha strongylata. *J. Org. Chem.* **1997,** 62, (8), 2636-2639.
16. Calabro, K.; Kalahroodi, E. L.; Rodrigues, D.; Díaz, C.; Cruz, M. d. l.; Cautain, B.; Laville, R.; Reyes, F.; Pérez, T.; Soussi, B.; Thomas, O. P., Poecillastrosides, Steroidal Saponins from the Mediterranean Deep-Sea Sponge Poecillastra compressa (Bowerbank, 1866). *Mar. Drugs* **2017,** 15, (7).
17. Takanashi, E.; Takada, K.; Hashimoto, M.; Itoh, Y.; Ise, Y.; Ohtsuka, S.; Okada, S.; Matsunaga, S., Cytotoxic linear acetylenes from a marine sponge Pleroma sp. *Tetrahedron* **2015,** 71, (51), 9564-9570.
18. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.;

Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016,** 34, (8), 828-837.

19. Wang, Y.; Wang, Z.; Wang, Z.; Liu, X.; Jiang, Y.; Jiao, X.; Xie, P., Total Synthesis of the Proposed Structure of Characellide B. *Org. Lett.* **2021,** 23, (9), 3680-3684.
20. Suo, R.; Takada, K.; Irie, R.; Watanabe, R.; Suzuki, T.; Ise, Y.; Ohtsuka, S.; Okada, S.; Matsunaga, S., Poecillastrin H, a Chondropsin-Type Macrolide with a Conjugated Pentaene Moiety, from a Characella sp. Marine Sponge. *J. Nat. Prod.* **2018,** 81, (5), 1295-1299.
21. Bowman, E. J.; Gustafson, K. R.; Bowman, B. J.; Boyd, M. R., Identification of a New Chondropsin Class of Antitumor Compound That Selectively Inhibits V-ATPases*. *J. Biol. Chem.* **2003,** 278, (45), 44147-44152.
22. Fang, H.; Kang, J.; Zhang, D., Microbial production of vitamin B12: a review and future perspectives. *Microb. Cell. Fact.* **2017,** 16, (1), 15.
23. Doxey, A. C.; Kurtz, D. A.; Lynch, M. D. J.; Sauder, L. A.; Neufeld, J. D., Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J.* **2015,** 9, (2), 461-471.
24. Pulte, A.; Wagner, S.; Kogler, H.; Spiteller, P., Pelianthinarubins A and B, Red Pyrroloquinoline Alkaloids from the Fruiting Bodies of the Mushroom Mycena pelianthina. *J. Nat. Prod.* **2016,** 79, (4), 873-878.
25. Liu, D.; Tan, H.-Q.; Chen, W.-L.; Li, Y.-G.; Wang, E.-B., Resolution of chiral polyoxoanion [P2Mo18O62]6− with histidine. *CrystEngComm* **2010,** 12, (7), 2044-2046.
26. Rienzo, M.; Lummis, Sarah C. R.; Dougherty, Dennis A., Structural Requirements in the Transmembrane Domain of GLIC Revealed by Incorporation of Noncanonical Histidine Analogs. *Chem. Biol.* **2014,** 21, (12), 1700-1706.
27. Jain, R.; Cohen, L. A.; El-Kadi, N. A.; King, M. M., Regiospecific alkylation of histidine and histamine at C-2. *Tetrahedron* **1997,** 53, (7), 2365-2370.
28. Stampfli, A. R.; Blankenfeldt, W.; Seebeck, F. P., Structural basis of ergothioneine biosynthesis. *Curr. Opin. Struct. Biol.* **2020,** 65, 1-8.
29. Zhang, Q.; van der Donk, W. A.; Liu, W., Radical-Mediated Enzymatic Methylation: A Tale of Two SAMS. *Acc. Chem. Res.* **2012,** 45, (4), 555-564.
30. Zhou, S.; Alkhalaf, L. M.; de los Santos, E. L. C.; Challis, G. L., Mechanistic insights into class B radical-S-adenosylmethionine methylases: ubiquitous tailoring enzymes in natural product biosynthesis. *Curr. Opin. Chem. Biol.* **2016,** 35, 73-79.
31. Willoughby, P. H.; Jansma, M. J.; Hoye, T. R., A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts. *Nat. Protoc.* **2014,** 9, (3), 643-660.
32. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.;

Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

33. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** 105, (8), 2999-3094.
34. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G., SpecDis: Quantifying the Comparison of Calculated and Experimental Electronic Circular Dichroism Spectra. *Chirality* **2013,** 25, (4), 243-249.

Chapter 5. Bioactive Diterpenoids Isolated from Deep-Sea Soft Coral *Paragorgia arborea*

Chapter 5. Bioactive Diterpenoids Isolated from Deep-Sea Soft Coral *Paragorgia arborea*

5.1 Abstract

Octocorals are a major source of marine bioactive metabolites, and deep-sea octocorals remain an untapped resource. Here, we studied the chemical diversity of the bubble gum coral, *Paragorgia arborea* (Cnidaria: Anthozoa: Alcyonacea: Scleraxonia: Paragorgiidae), collected at 1,500 m depth from a submarine canyon in Irish waters. We first used an antiplasmodial bioassay and NMR guided fractionation to prioritize extracts. The purification of a targeted extract yielded three new and five known diterpenes. The new compounds comprised epoxycoraxeniolide A, as well as two new miolenols with a cyclobutanol ring, one of which showed a high energy barrier of inter-conversion between two conformers. The five known diterpenes were from four different families. The structures of the metabolites were elucidated via a combination of mass spectrometry and 1 and 2D NMR and the absolute stereochemistry assigned via electronic circular dichroism. Several metabolites displayed mild activity against the malaria parasite, Plasmodium falciparum, and miolenol was cytotoxic. Overall, the high diversity of metabolites emphasizes the potential of deep-sea corals as a source for new natural products.

Table 1. Authors and contributions

Authors	Sample Identification	Conceptualization	Methodology	Formal Analysis	Software	Investigation	Data Curation	Visualization	Original Draft	Review and Editing	Resources	Supervision	Funding acquisition
Sam Afoullouss			X	X	X	X	X	X	X				
Ryan M. Young			X			X							
Laurence K. Jennings			X	X	X								
Jason Doyle						X							
Deb Livorsi				X		X			X				
John H. Adams										X	X		
Bill J. Baker		X		X						X		X	X
A. Louise Allcock	X	X								X	X	X	X
Olivier P. Thomas			X							X	X	X	

5.2 Introduction

Submarine canyons host unique and complex ecosystems including vibrant deep-sea reef systems[1]. Advances in deep-sea technologies, namely remotely operated vehicles (ROVs), has allowed researching of these under-explored reefs, revealing a rich benthic biodiversity[2, 3]. The extreme physico-chemical conditions including high pressure, and low temperatures, together with intense competition for food and space [4, 5] ,drive metabolic adaptations, and deep-sea microbiome diversity [6, 7]. The fauna which inhabits these reefs are thus proving to be a rich source of secondary metabolites with novel chemical scaffolds and bioactivity[8, 9], such as the novel lipoglycotripeptide characellides [10].

With an aim to discover new bioactive metabolites from Irish submarine canyon deep-sea coral gardens, we used ROV *Holland I* to sample at depths to 2,500 m in the Whittard Canyon. We collected a diverse range of octocorals due to their known abundance of bioactive metabolites, which frequently include terpenes and their derivatives, for example, furanocembranoid diterpenes isolated from deep-sea primnoid octocorals [11], briarane diterpenes from an Antarctic sea pen [12], and antiplasmodial cembrane-type diterpenoids, caucanolides, from a gorgonian [13].

A bioactivity screening process, combined with NMR-based prioritization, led our attention to the bubble gum coral, *Paragorgia arborea* (Alcyonacea: Scleraxonia: Paragorgiidae), the organic extract of which showed antiplasmodial activity and cytotoxicity. Previous chemical studies into specimens of the genus *Paragorgia* reported steroid thioesters [14] and xenicane class diterpenes [15-17]. The latter class of compounds appeared to be present in our organic extract, as indicated by characteristic ^{1}H NMR exomethylene and methyl signals. An in-depth chemical investigation resulted in the isolation of i) three new xenicane diterpenoids: two xeniaphyllanes with an unusual bicyclo[7.2.0]undecane named miolenol[1] (**1**) and epoxymiolenol (**2**), and a new xeniolide, epoxycoraxeniolide A (**3**), ii) and five known xeniolides, acalycixeniolide F (**4**)[18] , coraxeniolide A (**5**) and C (**6**) [19], acalycigorgin E (**7**)[20], and 9-deoxyxeniolide A (**8**)[21]. Several of these metabolites retained mild activity in antiplasmodial screening and miolenol displayed cytotoxicity against J774 cell lines.

[1]Names for new compounds **1** and **2** derived from the Irish Gaelic word *míol mór,* inspired by the compounds' resemblance to a whale.

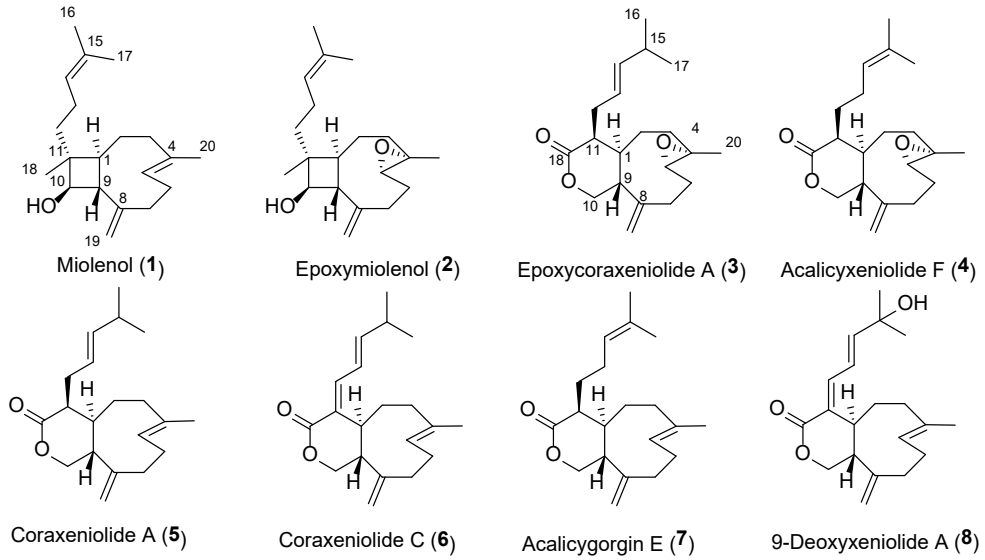


Figure 1. Structures and numbering of new metabolites miolenols (**1**) and epoxymiolenol (**2**), and epoxycoraxeniolide A (**3**), as well as known cembranes acalycixeniolide F (**4**) coraxeniolide A (**5**), coraxeniolide C (**6**), acalycigorgin E (**7**), and 9 deoxyxeniolide A (**8**).

5.3. Results and Discussion

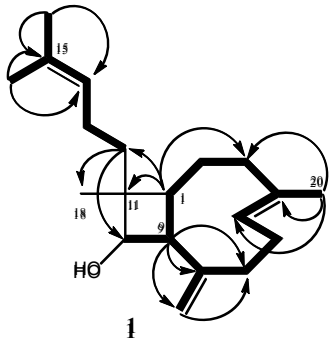
Miolenol (**1**) was isolated from the methanolic fraction, as a colourless oil. Its molecular formula was determined as $C_{20}H_{32}O$, based on the (+)-HRESIMS analysis ion peak at *m/z* 289.2536 (Δ 3.6 ppm), calculated for $[M + H]^+$ 289.2526, with five degrees of unsaturation. Examination of the HSQC spectrum in $CDCl_3$ showed the presence of four methyl groups (δ_H 1.69, δ_C 25.9, H_3-16; δ_H 1.62, δ_C 17.8, H_3-17; δ_H 0.95, δ_C 13.2, H_3-18; δ_H 1.56, δ_C 16.7, H_3-20), a cyclic and oxygenated methylene (δ_H 4.90, 4.87, δ_C 114.2, H_2-19), and olefinic signals at δ_C 134.8 (C-4), δ_C 125.0 (δ_H 5.27, H-5), δ_C 124.8 (δ_H 5.12, H-14), δ_C 131.7 (C-15), and δ_C 150.6 (C-8). These double bonds account for three degrees of unsaturation, indicating the remaining two degrees of unsaturation may stem from 2 cycles. The presence of a secondary alcohol at δ_H 3.52 (δ_C 75.8, H-10), was supported by a COSY correlation between H-10 and exchangeable proton at 1.53. Five methylene groups were observed from HSQC correlations (δ_H 1.64, 1.39, δ_C 29.0, H_2-2; δ_H 2.10, 1.90, δ_C 39.7, H_2-3; δ_H 2.35, 2.06, δ_C 29.9, H_2-6; δ_H 2.17, 2.05, δ_C 33.7, H_2-7; δ_H 1.42, δ_C 42.7, H_2-12; δ_H 2.00, 1.96, δ_C 23.2, H_2-13). Two bridge-head methines, with characteristic upfield proton and carbon (δ_H 1.38, δ_C 43.1, H-1; δ_H 2.19, δ_C 60.2, H-9) and a quaternary carbon at δ_C 43.4 (C-11) accounted for the remaining carbons and *C*-attached protons.

Table 1. ^{1}H NMR (600MHz) and ^{13}C NMR (150 MHz) spectroscopic data for miolenol (**1**), epoxymiolenol (**2**), and epoxycoraxeniolide A (**3**).

	Miolenol **(1)**				Epoxymiolenol **(2)**		Epoxycoraxeniolide A **(3)**	
	Chloroform-CDCl₃				*Chloroform-CDCl₃*		*Chloroform-CDCl₃*	
	Conformer A		*Conformer B*					
Pos	δ_C, Type	δ_H (Mult., *J* in Hz)	δ_C, Type	δ_H (Mult., *J* in Hz)	δ_C, Type	δ_H (Mult., *J* in Hz)	δ_C, Type	δ_H (Mult., *J* in Hz)
1	43.1, CH	1.38 (m)	47.7, CH	1.03	40.4, CH	1.47 (m)	43.9, CH	2.16 (m)
2	29.0, CH_2	1.64, 1.39 (m)	31.7, CH_2	1.69	27.2, CH_2	1.80, 1.42 (m)	28.2, CH_2	1.81, 1.09 (dt, 14.4, 4.08)
3	39.7, CH_2	2.10, 1.90 (m)	35.0, CH_2	2.51, 1.57	38.5, CH_2	2.10, 0.98 (m)	39.7, CH_2	2.24, 0.95 (m)
4	134.8, C		n.o.		59.6, C		59.2, C	
5	125.0, CH	5.27 (dd, 11.9, 5.24)	130.2, CH	5.35	64.3, CH	2.83 (dt, 10)	61.9, CH	2.88 (dd, 5.38, 4.85)
6	29.9, CH_2	2.35, 2.06 (m)	30.0, CH_2	2.46, 2.13	30.7, CH_2	2.27, 1.32 (m)	25.2, CH_2	2.28, 1.47 (m)
7	33.7, CH_2	2.17, 2.05 (m)	40.8, CH_2	2.48, 1.95	28.5, CH_2	2.32, 2.17 (m)	33.7, CH_2	2.56, 2.12 (m)
8	150.6, C		n.o.		148.4, C		150.8, C	
9	60.2, CH	2.19 (m)	60.1, CH	2.06	59.4, CH	2.44 (t)	48.5, CH	2.37 (m)
10	75.8, CH	3.52 (d, 8.34)	79.1, CH	3.65	75.1, CH	3.52 (t, 6.50)	70.9, CH_2	4.28, 3.96 (dd 12.0, 7.37)
11	43.4, C		n.o.		44.6, C		43.0, CH	2.84 (dt, 8.06, 6.16)
12	42.7, CH_2	1.42 (m)	43.1, CH	1.39	42.3, CH_2	1.41 (m)	29.9, CH_2	2.57, 2.16 (m)
13	23.2, CH_2	2.00, 1.96 (m)	23.2, CH_2	2.00, 1.96	23.0, CH_2	2.01, 1.96 (m)	122.9, CH_2	5.29 (m)
14	124.8, CH	5.12 (t, 6.01)	124.8, CH	5.12	124.9, CH	5.10 (t, 7.16)	141.2, CH_2	5.47 (dd, 8.67, 6.82)
15	131.7, C		n.o.		131.6, C		31.3, CH	2.25 (m)
16	25.9, CH_3	1.69 (s)	25.9, CH_3	1.69 (s)	25.7, CH_3	1.69 (s)	22.7, CH_3	0.95 (d, 2.01)
17	17.8, CH_3	1.62 (s)	17.9, CH_3	1.62 (s)	17.7, CH_3	1.62 (s)	22.7, CH_3	0.95 (d, 2.01)
18	13.2, CH_3	0.95 (s)	12.8, CH_3	0.93	12.5, CH_3	0.98 (s)	174.7, C	
19	114.2, CH_2	4.90, 4.87 (s)	111.24,	4.90	114.5, CH_2	4.99, 4.89 (s)	113.9, CH_3	5.17, 5.06 (s)
20	16.7, CH_3	1.56 (s)	22.1, CH_3	1.56	17.2, CH_3	1.15 (s)	16.3, CH_2	1.37 (s)

Interpretation of HMBC spectrum from miolenol includes correlations (Figure 2) from H_3-16 and H_3-17 to C-15, revealing the termination of a trisubstituted olefin, C-15/C-14. Allylic coupling observed in COSY correlations demonstrate H-16/17 coupling to H-14, which was further correlated with a methylene group H_2-13, and H_2-13 to a second methylene group at H_2-12. HMBC correlations from H_2-12 to a quaternary carbon (C-11), a methyl group (H_3-18), and a secondary alcohol (C-10), were key in establishing a linear termination. COSY correlations connect H-10 to a methine (H-9). The presence of a rare cyclobutanol ring was confirmed by a key COSY correlation between H-9 and a methine (H-1),

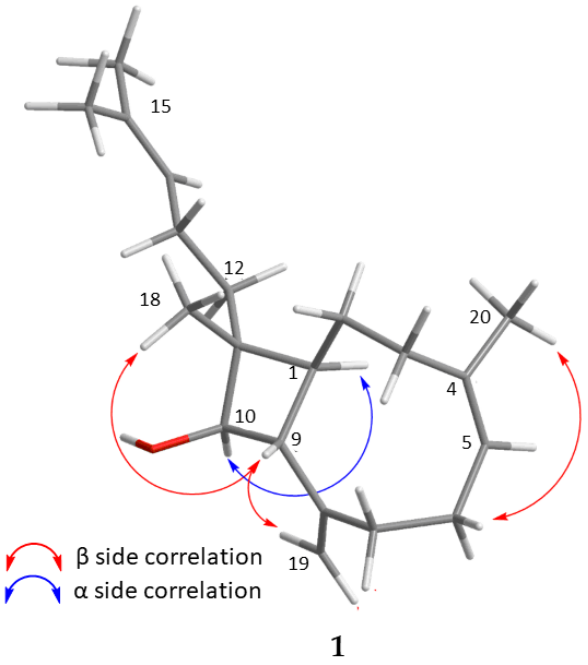
and HMBC correlations of H-1 to C-12, C-11 and C-18. Allylic coupling between H-9 and H_2-19, and HMBC signals from H-9 to C-8 and C-19, placed an exocyclic methylene group adjacent to the cyclobutanol ring. HMBC correlations from H-19 to methylenes C-7, combined with allylic COSY coupling between H_2-19 and H-7, confirmed the order of the methylenes. COSY correlations between H_2-6 and a triplet olefinic proton signal at δ_H 5.27 (H-5), established the location of the final olefinic bond. Key HMBC correlations of H-5 to a signal at δ_C 16.7 (C-20), show the presence of a methyl substituent on the double bond. HMBC correlations from H_3-20 to non-protonated olefinic carbon C-4, as well as C-5 and a methylene δ_C 39.7 (C-3) verified the trisubstituted nature of the C-4/C-5 olefinic bond. $^{2,3}J_{CH}$ HMBC correlations of H-1 to methylene carbons δ_C 29.0 (C-2) and C-3 establish a cyclononene ring, resulting in a bicyclic structure for miolenol.



1

Figure 2. Key COSY (▬) and HMBC (⌒) correlations establishing the planar structures of miolenol (**1**).

Stereochemical features of miolenol were investigated by 2D NOE spectroscopy (NOESY). The C-4/C-5 olefin was established as *E* based on the NOESY correlation between H-6 and H_3-20 (Figure 3). A relative configuration for the cyclobutanol ring of miolenol A was assigned based on a correlation of H-9/H_3-18, requiring H-9 and H_3-18 to occupy the same face of miolenol, while the H-1/H-10 correlation require H-1 and H-10 on the opposite face.



1

Figure 3. Key NOESY (⌒) correlations establishing the relative configurations in miolenol (**1**).

The absolute configuration of miolenol was determined by comparison of the experimental and calculated electronic circular dichroism (ECD) spectra. Due to the flexibility of the molecule the minimum energy conformer search was constrained using experimental NOESY correlations, resulting in 42 conformers with similar conformations as that obtained from the crystal structure of coraxeniolide [18]. Time-dependent density functional theory (TDDFT) at the M06-2X/def2-TZVP//B3LYP/6-311G(2d,p) level of theory was used to calculate ECD spectra for both enantiomers. The calculated ECD spectrum for (1*S*,9*R*,10*R*,11*R*)-miolenol matched the experimental spectrum (Figure 4). Interestingly, this stereo chemistry is opposite to that reported for most molecules of the xeniaphyllane family[22]. However, the absolute configuration of these coral-derived natural products has not been studied and most xeniaphyllanes, having been reported over 40 years ago, were assumed to have the same configuration as the plant-derived β-caryophyllene. Due to the lack of stereochemical studies on this family and to allow a more reliable assignment, the ECD of the optimised conformers of **1** were calculated again using another functional/basis set combination (ωB97XD/TZVP). These data matched the previous calculations further supporting our assignment. While our configuration does not match other xeniaphyllanes, more recently a number of other soft coral terpenes, not classified as xeniaphyllanes but containing the same rings, have also reported this configuration. The synthesis of the similar diterpene antheliolide A by Mushti et al. allowed for the stereochemical revision of the four-membered ring to the same absolute configuration as reported here [23]. Furthermore, the configuration of coral-derived caryophyllene-type sesquiterpenes isolated by Ahmed et al. was opposite to those based on optical rotation data from plant studies [24]. This suggests that the stereochemistry of a number of earlier reported xeniaphyllanes may be incorrectly assigned based on similar plant terpenes and likely needs reinvestigating [22].

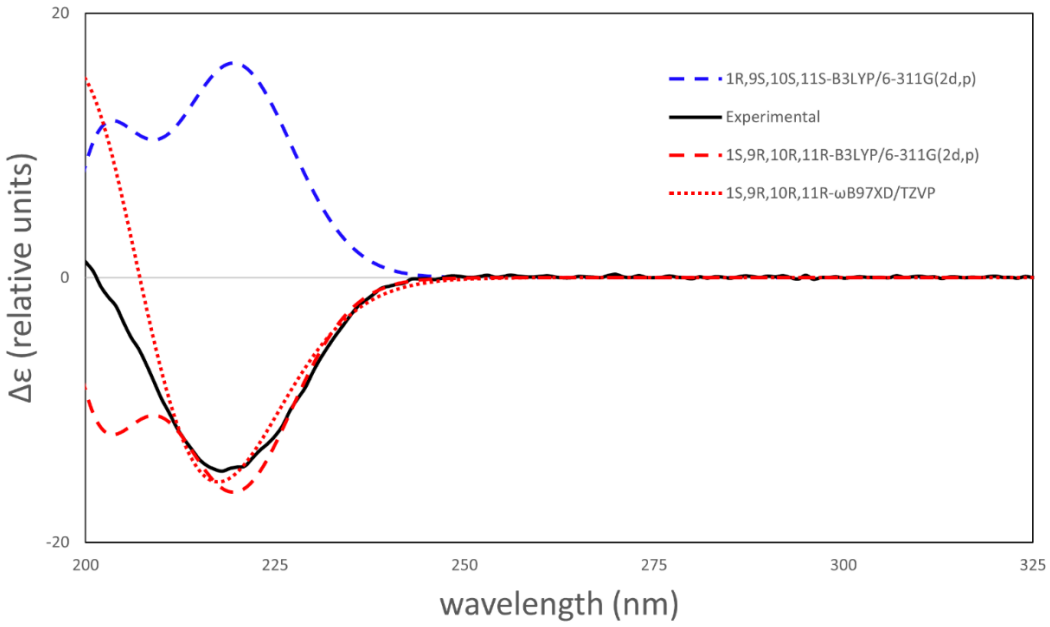


Figure 47. Calculated and experimental ECD spectra for miolenol (**1**).

Initial examination of the ^{1}H NMR spectrum of miolenol showed the presence of what appeared to be doubling of signals, possibly due to a stereoisomer, in a 3:1 ratio. Exchange correlations observed in the ROESY spectrum (Figure S10)**,** suggested interconversion between two conformers of miolenol A were responsible for the doubling of signals, similar to those reported by Urda et al.,[25]. Conformational interconversion was confirmed by a variable temperature ^{1}H NMR experiment (Figure 5/S12), obtained in deuterated pyridine at 10K increments from 288.15 K to 338.15 K. These experiments demonstrate proton signals **1a** H-16 (δ_H 1.61) and **1b** H-16 (δ_H 1.63) at 288 K merging at

δ_H 1.66 by 338 K. In-depth literature review of bicyclo[7.2.0]undecane molecular dynamics found the work of Hubner et al. [26]showing the formation of three interconverting conformers for (-)-β-caryophyllene, αα, βα, and ββ, with conformers αα and βα interconverting rapidly, producing only one set of NMR signals at room temperature (Figure 5). We hypothesize a similar phenomenon is occurring in miolenol A. NOE correlations from the major conformers H-9/H-7 and H-9/H-19, indicated it is rapidly converting between αα and βα, which can only be observed in low temperature NMR spectra. NOE correlations of the minor conformer indicate it is in ββ conformation. No evidence of αβ conformer for miolenol was observed in NOESY experimental spectra. This is in line with the finds of Hubner et al.[26]. Comparisons between major and minor conformer NMR spectra and absences of exchange correlations in NOESY spectra of epoxymiolenol (*vide infra*), indicate the trisubstituted double bond at $\Delta^{4,5}$ is key for the conformational flexibility observed in **1**.

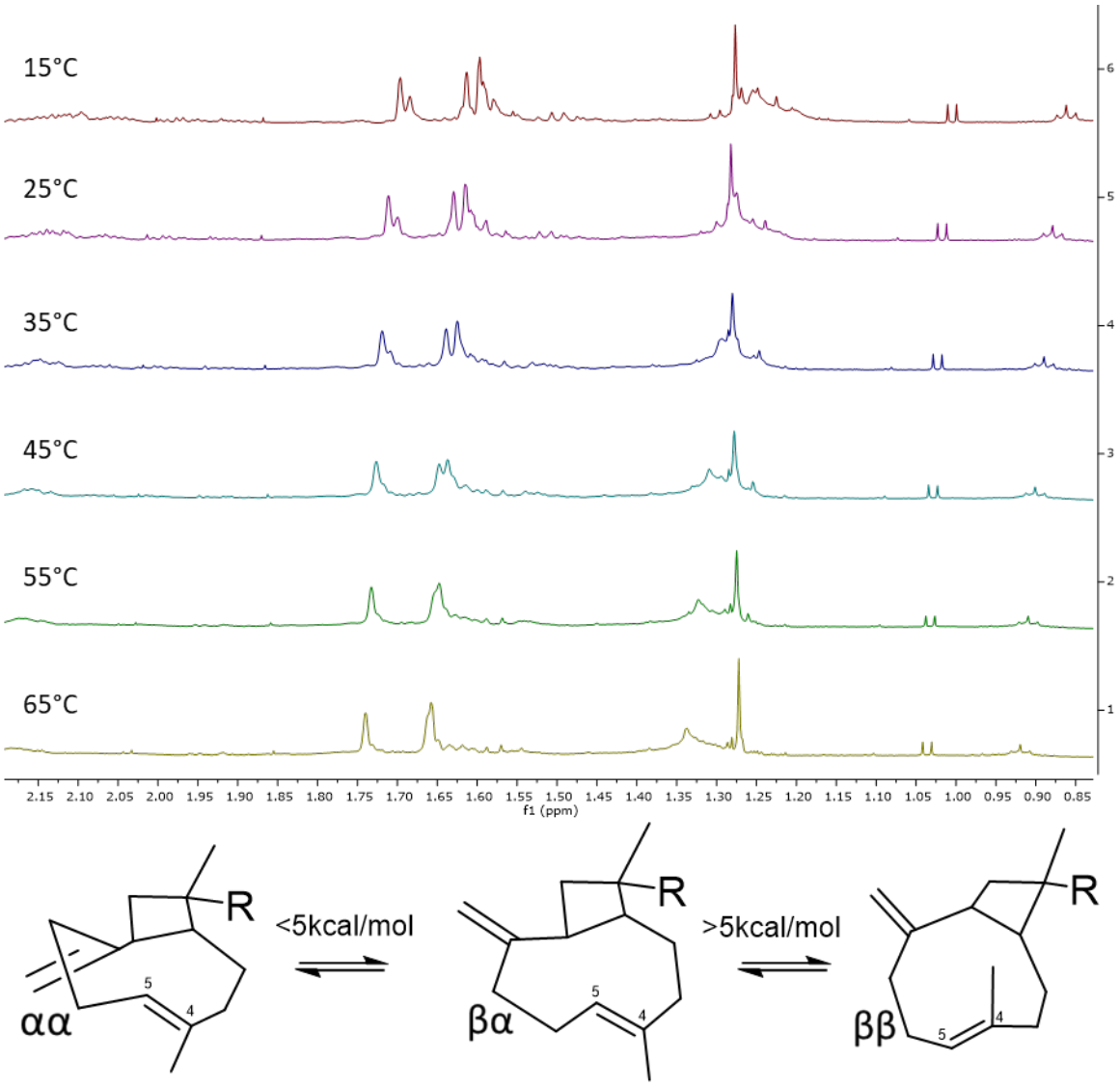
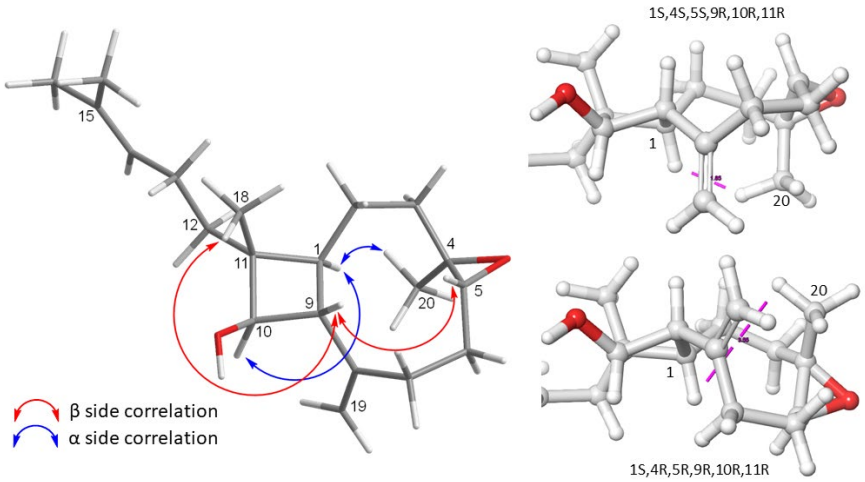


Figure 5. Variable temperature (15-65°C) ^{1}H-NMR in deuterated pyridine and geometries of three conformers of miolenol (**1**) with energy barriers of interconversion indicated.

Epoxymiolenol (**2**) was isolated as a colorless oil. Its molecular formula was determined as $C_{20}H_{32}O_2$, based on the (+)-HRESIMS analysis ion peak at *m/z* 305.2472, calculated for $[M + H]^+$ 305.2475. Comparison of 1D and 2D NMR spectra of miolenol and epoxymiolenol, revealed epoxymiolenol is also a xeniaphyllane, differing from miolenol, by an additional degree of oxidation, supported by the loss

of two olefinic carbon resonances at C-4 and C-5, and the gain of two oxygenated carbons δ_C 59.6 (C-4), δ_C 64.3 (C-5), consistent with epoxidation. A trisubstituted epoxide at C-4/C-5 was confirmed by a deshielded proton signal at δ_H 2.83 (H-5) and HMBC correlations of H-5 to C-6, and H_3-20 to C-3, C-4 and C-5.



2

Figure 6. Key NOESY ( ) correlations establishing the relative configurations in epoxymiolenol (**2**). Conformational analysis of diastereomers of epoxymiolenol.

Once the planar structure of epoxymiolenol was established, the relative configuration of the cyclobutanol ring was deduced from NOESY experimentation. Similar to miolenol, key NOE correlations from H-10 to H-1, H-1 to H-20 established their presence on the opposite face from H-9 and H_3-18. The relative configuration of the epoxide group was deduced from NOESY correlation indicating H-5 is on the opposite face from H_3-20. This was confirmed by analyzing the distances between H-1 and H-20, for the lowest energy conformers of epoxymiolenol-4S,5S and epoxymiolenol-4R,5R (Figure 6). The absolute configuration of epoxymiolenol was determined by experimental and computational ECD, using the same methodology as above, matching the (1*S*,4*S*,5*S*,9*R*,10*R*,11*R*)-epoxymiolenol ECD spectrum to the experimental ECD spectrum (Figure 7).

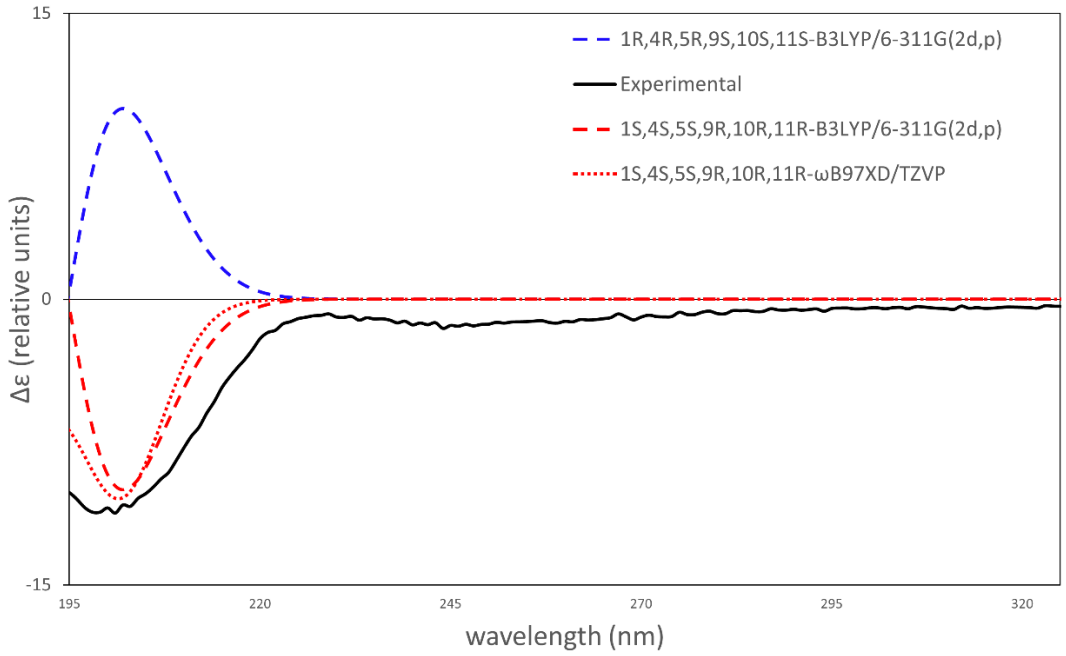
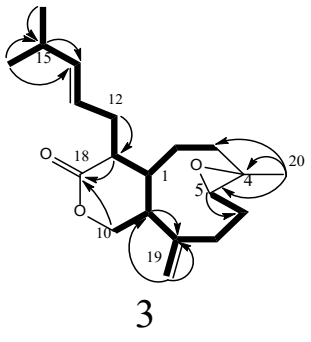


Figure 7. Calculated and experimental ECD spectra for epoxymiolenol (**2**).

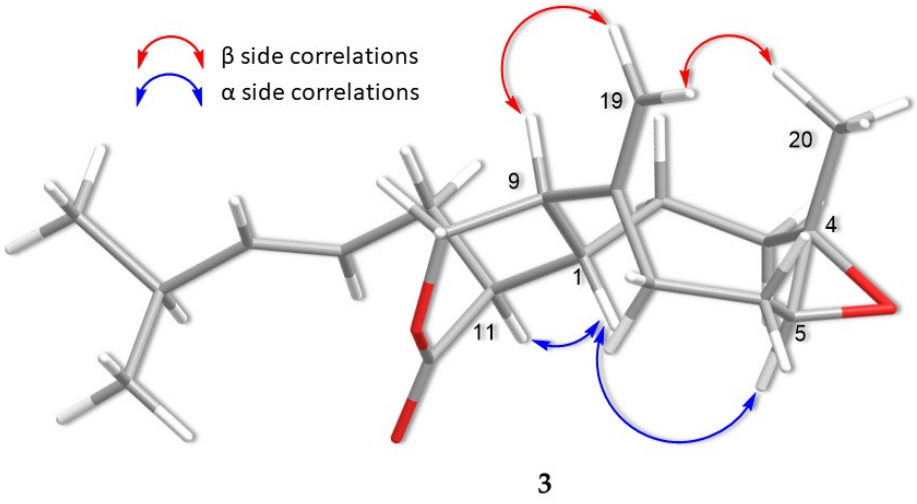
Epoxycoraxeniolide A (**3**) was isolated as a pale-yellow oil. Its molecular formula was determined by (+)-HRESIMS as $C_{20}H_{30}O_3$ (*m/z* 319.2269, calculated for $[M + H]^+$ 319.2268), demonstrating six degrees of unsaturation. Interpretation of ^{1}H NMR data revealed the presence of an isopropyl group (δ_H 0.95; δ_C 22.7, H_3-16/17), a singlet methyl group (δ_H 1.37; δ_C 16.3, H_3-20), and an exocyclic olefinic group (δ_H 5.17, 5.06; δ_C 113.9, H_2-19) respectively. An additional olefin was deduced from two signals at δ_H 5.47 (δ_C 141.2, H-14) and δ_H 5.29 (δ_C 122.9, H-13). Comparison of HSQC spectra of epoxymiolenol and epoxycoraxeniolide A indicated the shared presence of a C-4/C-5 epoxide-bearing cyclononane ring suggested by a ^{1}H-NMR signals at δ_H 2.83 (δ_C 64.3, H-5) and δ_H 2.88, (δ_C 61.9 ,H-5) in epoxymiolenol and epoxycoraxeniolide A respectively, and an exomethylene indicated by correlation at δ_H 4.99, 4.89 (δ_C 114.5, H_2-19) and δ_H 5.17, 5.06 (δ_C 113.9, H_2-19), respectively. Analysis of ^{13}C NMR signal at δ_C 174.7 (C-18) indicated the presence of a carboxylic acid or ester, accounting for two oxygens. Careful examination of gradient COSY, gradient HSQC and HMBC experiments, in comparison with miolenol and epoxymiolenol, indicated a bicyclic diterpene of the xenicane class. COSY correlations connect the isopropyl methyl groups (H_3-16/17) and olefinic methines (H-13/14) to a methylene δ_H 2.57, 2.16 (δ_C 29.9, H_2-12), as well as H_2-12 to a methine δ_H 2.84 (δ_C 43.0, H-11). Key HMBC correlations from H-11 to C-18, C-1, and C-12, and from deshielded methylene protons at δ_H 4.28, 3.96 (δ_C 70.9, H_2-10) to C-18, C-1, and C-9 confirm the presence of a lactone ring with an ester carbonyl at C-18. COSY correlations indicate a spin coupled system from H-11/H-1/H-2/H-3.



3

Figure 8. Key COSY (▬) and HMBC (⌒) correlations establishing the planar structure of epoxycoraxeniolide A (**3**).

Analysis of COSY spectra from epoxycoraxeniolide A indicated the presence of two additional proton-proton spin coupled systems (Figure 8). HMBC correlations from H-5 and H_2-3 to a quaternary carbon δ_C 59.2 (C-7) and correlations from H-9 to C-1, C-2 and C-11, establish a cyclononane ring. HMBC correlations from H-11 and H_2-10 to C-12 the established a lactone group. HMBC correlations from H_3-20 to C-3, C-4, and C-5 confirm the trisubstituted nature of the C-4/C-5 epoxide. The relative configuration of epoxycoraxeniolide A was deduced from NOE correlations showing H-1 and H-11 were on the opposite face as H-9. Key NOE correlations of H-1/H-5 and H-9/H-20 indicate H-20 and H-5 were on opposite faces (Figure 9). The absolute configuration of (1*S*,4*S*,5*S*,9*R*,11*S*) for epoxycoraxeniolide A was determined by comparison of experimental and calculated ECD spectra using the same method used for miolenol and epoxymiolenol (Figure 10).



3

Figure 9. Key NOESY correlations on the alpha side (blue) and beta side (red) of epoxycoraxeniolide A (**3**).

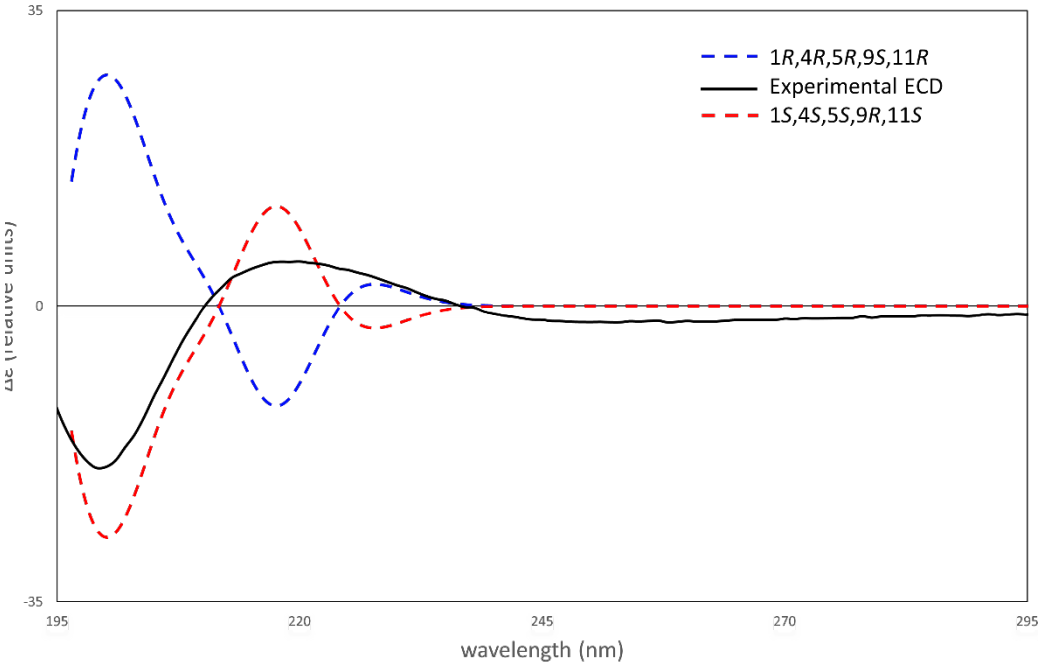


Figure 10. Calculated and experimental ECD spectra for epoxycoraxeniolide A (**3**).

Diterpenoids **1**-**8** were screened for antiplasmodial activity against *Plasmodium falciparum* strains NF54 (sensitive), Dd2 (Resistant) at concentrations of 5 µg/mL. Cytotoxicity was assessed against J774 macrophage cell lines, at concentrations of 10 µg/mL. Results from both strains indicate modest antiplasmodial activity for all metabolites (Figure 11). 9-Deoxyxeniolide A (**8**) showed >50% activity (57%) against the Dd2 strain in blood stage. Only miolenol (**1**) showed cytotoxicity, with 65% inhibition against J744 macrophages.

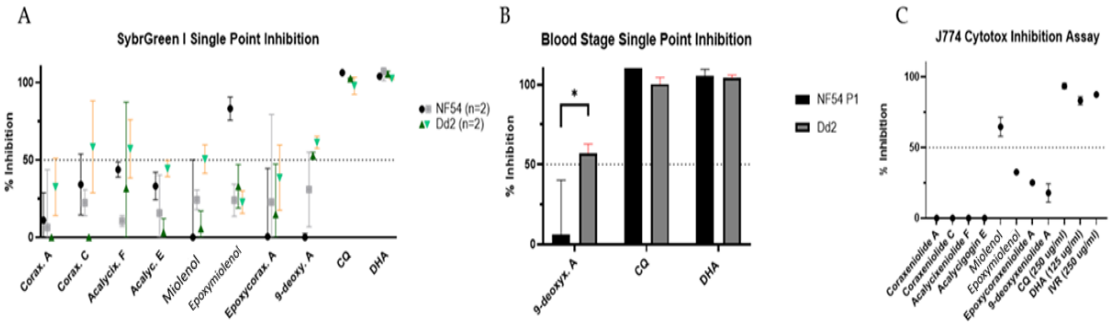


Figure 11. Compounds evaluated for inhibition against *Plasmodium falciparum* strains NF54 (sensitive), Dd2 (Resistant) and J774 macrophage toxicity. (a) While compound activity was variable when compared to control compounds, chloroquine (CQ) and dihydroartemisinin (DHA) at the same concentration, single point analysis of all compounds showed no significant inhibition to NF54. Error bars represent means with SD for technical replicates (n=2) with experimental replicates shown by independent points (n=2). (b) Two-way ANOVA analysis of the combined data between NF54 and Dd2 to the 9-deoxyxeniolide A (**8**) showed a significant response (P=0.0113). (c) Single point compound toxicity was evaluated against J774 macrophages at the 10 µg/mL concentration. Miolenol (**1**) showed >50% activity (64.7%); epoxymiolenol (**2**) (32.5%), epoxycoraxeniolide A (**3**) (25.2%) and 9-deoxyxeniolide A (**8**) (17.9%) had <50% activity when compared at the same concentration.

5.4. Materials and Methods

5.4.1. General Procedures

Optical rotations were measured on a UNIPOL 1000 Polarimeter. UV and ECD data were obtained on ChirascanTM V100 (Applied Photophysics, Leatherhead, UK). NMR experiments were measured on a 600/150 MHz Varian spectrometer equipped with a cryoprobe. Chemical shifts (δ in ppm) are referenced using residual $CHCl_3$ (δ_H 7.26 for ^1H and δ_C 77.0 for ^{13}C) and pyridine (δ_H 8.74, 7.58, 7.22 and δ_C 150.35, 135.91, 123.87). High Resolution Electrospray Ionization Mass Spectrometry (HRESIMS) data were obtained from a Q-ToF Agilent 6540 in positive mode. Preparative HPLC was preformed using a Jasco PU-2087 Plus equipped with a UV 2075 Plus detector and by an Agilent 1260 analytical HPLC series equipped with a DAD (model #G4212B) detector.

5.4.2. Collection and Identification of Paragorgia arborea

Two specimens were collected close together using the remotely operated vehicle *Holland I* during expedition CE1708 of RV *Celtic Explorer* at approximately 1500 m depth and at 48.830 °N, 11.048 °W. All epibionts were removed from the surface of the corals and a small amount was retained in 96% ethanol as a voucher specimen for DNA analysis and morphological identification. The rest of the sample was lyophilized and stored at -20 °C. Based on colony morphology and DNA sequencing of the octocoral mitochondrial gene MutS (GenBank Accession No. OM240806), the specimen was identified as *Paragorgia arborea.*

5.4.3. Extraction and Isolation of metabolites

Biomass totaling 758 g was lyophilized and extracted with 1 L dichloromethane utilizing a Soxhlet, resulting in 7.2 g of extract. Fractionation of the extract was achieved using reversed phase C_{18} vacuum

liquid chromatography, giving seven fractions (Figure S2). The methanolic fraction (675 mg) was purified by RP-HPLC-DAD on a semipreparative phenylhexyl column 5 µm, 10 × 250 mm (Xselect, Waters). Mobile phase consisted of A (H_2O) and B (CH_3CN), at a flow rate of 5 mL/min. The method was developed on 60 min acquisition time: isocratic 30% B for 4 min, then linear gradient to 70% B at 45 min, raised to 100% B at 47 min, held at 100% B for 54 min, back to 30% B in 1 min, and held at that percentage of B for 5 min. Initial purification, using acidified solvents (0.1% tetrafluoroacetic acid), appeared to cause degradation of metabolites after drying, indicated by ^{1}H NMR experiments (Figure S3), promoting the use of non-acidic solvent or MeOH for subsequent purifications. Sub-fraction 10 (t_R 34 min, 21 mg) containing miolenol (**1**) underwent further purification, conducted on an analytical T3 column 4.6 x 250 mm 5 µm (Xselect, Waters), using isocratic condition at 85% B for 31 min, yielding miolenol A (t_R 16 min, 0.66 mg, 9.2x10^{-5}% w/w). Sub-fraction 3 (t_R 22 min, 18 mg) containing epoxymiolenol (**2**), was repurified on an analytical C_{18} column 4.6 x 250 mm 5 µm (Xselect, Waters), yielding miolenol B (t_R 7 min, 0.74 mg, 9.7x10^{-4}% w/w). Sub-fraction 5 (t_R 24 min, 14 mg) holding epoxycoraxeniolide A (**3**) underwent repurification using T3 column 4.6 x 250 mm 5 µm (Xselect, Waters) with isocratic condition at 61%, at 1 mL/min, to yield epoxycoraxeniolide A (t_R 12 min, 0.83 mg, 1.2 x 10^{-4}% w/w). Sub-fraction 9 (t_R 33 min, 30 mg) underwent repurification using T3 column 4.6 x 250 mm 5 µm (Xselect, Waters) with isocratic condition at 82%, at 1 mL/min, to yield known xeniolides acalycinxeniolide F (**4**)(t_R 8 min, 0.26 mg, 3.4 x 10^{-5}% w/w), coraxeniolide C (**5**) (t_R 22 min, 0.73 mg, 9.6 x 10^{-4} w/w), acalycigorgin E (**6**) (t_R 23 min, 0.34 mg, 4.5 x 10^{-5}% w/w), and coraxeniolide A (**7**) (t_R 25 min, 0.42 mg, 5.8 x 10^{-5}% w/w). Sub-fraction 2 (t_R 20 min, 36.3 mg) underwent repurification using C_{18} column 4.6 x 250 mm 5 µm (Xselect, Waters) with isocratic condition at 54%, at 1 mL/min, to yield known xeniolide 9-deoxyxeniolide A (**8**) (t_R 13 min, 0.35 mg, 4.7 x 10^{-5}% w/w).

Miolenol (**1**): Colorless oil; $[\alpha]^{25}_D$ +19° (*c* 0.12, ACN); UV (ACN) λ_{max} 210 nm; ^{1}H NMR (600 MHz) and ^{13}C NMR (150 MHz), Table 1; (+)-HRESIMS analysis an ion peak at *m/z* 289.2536, calcd for $[M + H]^+$ 289.2526

Epoxymiolenol (**2**): Colorless oil; $[\alpha]^{25}_D$ +23° (*c* 0.12, ACN); UV (ACN) λ_{max} 210 nm; ^{1}H NMR (600 MHz) and ^{13}C NMR (150 MHz), Table 1; (+)-HRESIMS analysis an ion peak at *m/z* 305.2472, calcd for $[M+H^+]$ 305.2475

Epoxycoraxeniolide A (**3**): Colorless oil; $[\alpha]^{25}_D$ +49.5° (*c* 0.12, ACN); UV (ACN) λ_{max} 220 nm; ^{1}H NMR (600 MHz) and ^{13}C NMR (150 MHz), Table 1; (+)-HRESIMS as $C_{20}H_{30}O_3$ (*m/z* 319.2269, calcd for $[M + H]^+$ 319.2268)

5.4.4. Computational Methods

A conformational analysis of miolenol and epoxymiolenol and epoxycoraxeniolide A was performed using a Monte Carlo Minimum method (MCMM) and the molecular mechanics OPLS3 force field with an energy cut-off at 5 kcal/mol in Schrodinger MacroModel [27]. The conformational analysis was constrained using key NOESY correlation to limit the number of conformers. These conformers were then optimized, and frequency calculations were performed using DFT, at the M06-2X/def2-TZVP level in Gaussian 16 [28]. Gaussian 16 was then used to calculate ECD spectra for each conformer at the B3LYP/6-311G(2d,p) level. All DFT calculations were performed using a polarizable continuum solvation model[29]. The final ECD spectra were extracted, Boltzmann weighted and corrected by alignment with the UV spectra using the freely available software SpecDis 1.7 (version 1.71, SpecDis, Berlin, Germany) [30].

5.4.5. Bioassay

5.4.5.1. Antimalaria activity

Antimalaria activity was assessed with the adaptation of the sensitivity assay of Desjardins et al. [29] using SybrGreen fluorescence as an assessment of parasite growth. Control and sample compounds were prepared in 100% dimethylsulfoxide (DMSO) at 5 mg/mL. Compounds were further serially diluted 1:3 in DMSO. Assays were performed in 384-well microtiter plates; each plate contained 40 μL of parasite culture (0.5% parasitemia, 2.0 % hematocrit) of either NF54 wild type (WT) or Dd2 resistant (chloroquine, pyrimethamine and mefloquine resistant) and 40 nL of drug dispensed using the V&P Scientific pin tool mounted to a liquid handling robot. Control wells were run with 40 nL of 5 mg/ml of a control compound or 100% DMSO and 40 μL of cell suspension. Each compound was tested with two technical replicates per concentration, and parasite growth was compared to controls incubated at 37 °C under normal culture conditions (5%O_2, 5% CO_2, 95% N). Plates were frozen and thawed before adding a final 2x concentration of SybrGreen in lysis buffer and read on the CLARIOstar fluorescence plate reader (Ex/Em 484-15/528-15) following a 45-minute incubation at room temperature (RT). Single point analysis was determined after transforming the relative fluorescence response to percent inhibition using the equation (100*(1-(n-min)/(max-min))). Compounds were determined to be active if inhibition exceeded >70%.

5.4.5.2. Cytotoxic activity

Each compound was tested for cytotoxicity using mammalian J774A.1 cell lines (ATCC TIB-67™) in complete media; RPMI medium with phenol red containing l-glutamine and then supplemented with 10% fetal bovine serum (CM). Cells were seeded at 5x10^5 cell/mL and plated in a 96-well format with 100 µL/well fresh media on day 0 and were incubated overnight at 37 °C, 5% CO_2 for adherence. Following 24-hour incubation, spent media was removed and 100 µL of fresh media with test compounds were serially diluted 1:2 from a starting concentration of 10 ug/mL was added to the cells and incubated for an additional 68 hours before adding in 20 µL CellTiter 96® AQueous One Solution Cell Proliferation Assay reagent (Promega). Cells were incubated for an additional 4 hours before reading absorbance (490 nm) on the CLARIOstar plate reader. Single point analysis was determined after transforming the response to percent inhibition using the equation (100*(1-(n-min)/(max-min))). Compounds were determined to be active if inhibition exceeded >70%.

5.5 Supplementary Materials and Data Availability Statement

The following supporting information is available in the appendix. Figures S1-S47: In situ photograph of *Paragorgia arborea*, solvent scheme, and ^{1}H NMR, ^{13}C NMR, HRESIMS, gCOSY, gHSQCAD, gHMBCAD, ROESY, and NOESY spectra (as appropriate) of compounds **1** to **8**.

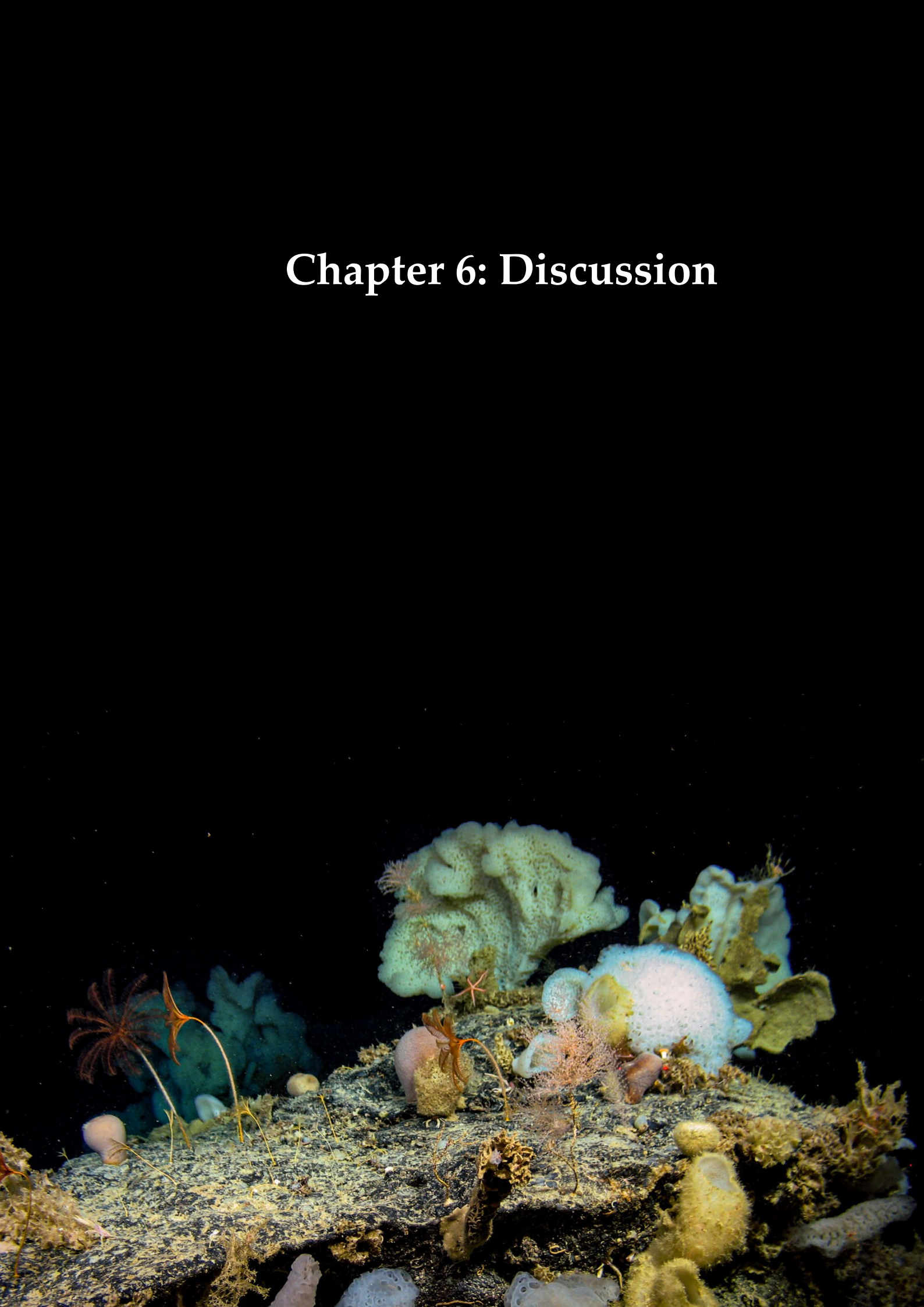
5.6 References

1. Johnson, M. P.; White, M.; Wilson, A.; Würzberg, L.; Schwabe, E.; Folch, H.; Allcock, A. L., A vertical wall dominated by Acesta excavata and Neopycnodonte zibrowii, part of an undersampled group of deep-sea habitats. *PloS One* **2013,** 8, (11), e79917.
2. Amaro, T.; Huvenne, V.; Allcock, A.; Aslam, T.; Davies, J.; Danovaro, R.; De Stigter, H.; Duineveld, G.; Gambi, C.; Gooday, A., The Whittard Canyon–A case study of submarine canyon processes. *Progress in Oceanography* **2016,** 146, 38-57.
3. Fernandez-Arcaya, U.; Ramirez-Llodra, E.; Aguzzi, J.; Allcock, A. L.; Davies, J. S.; Dissanayake, A.; Harris, P.; Howell, K.; Huvenne, V. A.; Macmillan-Lawler, M., Ecological role of submarine canyons and need for canyon conservation: a review. *Frontiers in Marine Science* **2017,** 4, 5.

4. Buhl-Mortensen, L.; Vanreusel, A.; Gooday, A. J.; Levin, L. A.; Priede, I. G.; Buhl-Mortensen, P.; Gheerardyn, H.; King, N. J.; Raes, M., Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. *Marine Ecology* **2010,** 31, (1), 21-50.
5. Beazley, L. I.; Kenchington, E. L.; Murillo, F. J.; Sacau, M. d. M., Deep-sea sponge grounds enhance diversity and abundance of epibenthic megafauna in the Northwest Atlantic. *ICES Journal of Marine Science* **2013,** 70, (7), 1471-1490.
6. Sogin, M. L.; Morrison, H. G.; Huber, J. A.; Mark Welch, D.; Huse, S. M.; Neal, P. R.; Arrieta, J. M.; Herndl, G. J., Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* **2006,** 103, (32), 12115-12120.
7. Xu, D. B.; Han, L. N.; Li, C. H.; Cao, Q.; Zhu, D. L.; Barrett, N. H.; Harmody, D.; Chen, J.; Zhu, H. N.; McCarthy, P. J.; Sun, X. M.; Wang, G. J., Bioprospecting Deep-Sea Actinobacteria for Novel Anti-infective Natural Products. *Front. Microbiol.* **2018,** 9, 9.
8. Skropeta, D.; Wei, L., Recent advances in deep-sea natural products. *Natural product reports* **2014,** 31, (8), 999-1025.
9. Skropeta, D., Deep-sea natural products. *Natural Product Reports* **2008,** 25, (6), 1131-1166.
10. Afoullouss, S.; Calabro, K.; Genta-Jouve, G.; Gegunde, S.; Alfonso, A.; Nesbitt, R.; Morrow, C.; Alonso, E.; Botana, L. M.; Allcock, A. L.; Thomas, O. P., Treasures from the Deep: Characellides as Anti-Inflammatory Lipoglycotripeptides from the Sponge Characella pachastrelloides. *Org. Lett.* **2019,** 21, (1), 246-251.
11. Thomas, S. A.; Von Salm, J. L.; Clark, S.; Ferlita, S.; Nemani, P.; Azhari, A.; Rice, C. A.; Wilson, N. G.; Kyle, D. E.; Baker, B. J., Keikipukalides, furanocembrane diterpenes from the Antarctic deep sea octocoral Plumarella delicatissima. *Journal of natural products* **2018,** 81, (1), 117-123.
12. Thomas, S. A. L.; Sanchez, A.; Kee, Y.; Wilson, N. G.; Baker, B. J., Bathyptilones: Terpenoids from an Antarctic Sea Pen, Anthoptilum grandiflorum (Verrill, 1879). *Mar. Drugs* **2019,** 17, (9), 513.
13. Ospina, C. A.; Rodríguez, A. D.; Sánchez, J. A.; Ortega-Barria, E.; Capson, T. L.; Mayer, A. M., Caucanolides A− F, Unusual Antiplasmodial Constituents from a Colombian Collection of the Gorgonian Coral Pseudopterogorgia b ipinnata. *Journal of natural products* **2005,** 68, (10), 1519-1526.
14. Poza, J. J.; Fernandez, R.; Reyes, F.; Rodriguez, J.; Jimenez, C., Isolation, biological significance, synthesis, and cytotoxic evaluation of new natural parathiosteroids a− c and analogues from the soft coral Paragorgia sp. *The Journal of organic chemistry* **2008,** 73, (20), 7978-7984.
15. Arboxeniolide-1, a new, naturally occurring xeniolide diterpenoid from the gorgonian Paragorgia arborea of the Crozet Is. (S. Indian Ocean). *Z. Naturforsch.* **1984,** 39, 1180-1183.
16. Stonik, V.; Makar'eva, T.; Dmitrenok, A., New diterpenoid of the xeniane series from the gorgonianParagorgia arborea. *Chemistry of Natural Compounds* **1990,** 26, (1), 103-104.
17. Bright-Diaz, L. M.; Strychar, K. B.; Shirley, T. C., Compounds from deep-sea bubblegum corals, Paragorgia arborea, elicit anti-predation behavior in fish. *Open Marine Biology Journal* **2011,** 5, 58-67.
18. Schwartz, R. E.; Scheuer, P. J.; Zabel, V.; Watson, W. H., The coraxeniolides, constituents of pink coral, Corallium sp. *Tetrahedron* **1981,** 37, 2725-2733.
19. Ochi, M.; Kataoka, K.; Tatsukawa, A.; Kotsuki, H.; Shibata, K., Biologically active xenicane diterpenoids from the gorgonian Acalycigorgia sp. *Heterocycles* **1994,** 38, (1), 151-158.
20. Rho, J.-R.; Lee, H.-S.; Seo, Y.; Cho, K. W.; Shin, J., New Xenicane Diterpenoids from the Gorgonian Acalycigorgia i nermis. *Journal of natural products* **2000,** 63, (2), 254-257.
21. Vervoor, H. C.; Fenical, W., Antibacterial diterpenoids from an undescribed soft-coral of the genus Xenia. *Natural Product Letters* **1995,** 6, (1), 49-55.
22. Ng, S.-Y.; Phan, C.-S.; Ishii, T.; Kamada, T.; Hamada, T.; Vairappan, C. S., Terpenoids from Marine Soft Coral of the Genus Xenia in 1977 to 2019. *Molecules* **2020,** 25, (22), 5386.
23. Mushti, C. S.; Kim, J.-H.; Corey, E. J., Total Synthesis of Antheliolide A. *Journal of the American Chemical Society* **2006,** 128, (43), 14050-14052.

24. Ahmed, A. F.; Su, J.-H.; Shiue, R.-T.; Pan, X.-J.; Dai, C.-F.; Kuo, Y.-H.; Sheu, J.-H., New β-Caryophyllene-Derived Terpenoids from the Soft Coral Sinularia nanolobata. *Journal of Natural Products* **2004,** 67, (4), 592-597.
25. Urda, C.; Fernández, R.; Pérez, M.; Rodríguez, J.; Jiménez, C.; Cuevas, C., Protoxenicins A and B, cytotoxic long-chain acylated xenicanes from the soft coral Protodendron repens. *Journal of Natural Products* **2017,** 80, (3), 713-719.
26. Hübner, M.; Rissom, B.; Fitjer, L., Conformation and Dynamics of (−)-β-Caryophyllene. *Helvetica chimica acta* **1997,** 80, (6), 1972-1982.
27. Willoughby, P. H.; Jansma, M. J.; Hoye, T. R., A guide to small-molecule structure assignment through computation of (1 H and 13 C) NMR chemical shifts. *Nature protocols* **2014,** 9, (3), 643-660.
28. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
29. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum mechanical continuum solvation models. *Chemical reviews* **2005,** 105, (8), 2999-3094.
30. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G., SpecDis: Quantifying the comparison of calculated and experimental electronic circular dichroism spectra. *Chirality* **2013,** 25, (4), 243-249.

Chapter 6: Discussion



6.Discussion

6.1. Deep-sea Sponge Natural Products

The potential of deep-sea sponges to provide a unique chemistry with pharmaceutical potential, is greatly dependent on the class of sponges. The two most prevalent classes, Hexactinellida and Demospongiae differ greatly in the diversity and quantity of secondary metabolites produces.

Although there is large diversity of glass sponge species in deep-sea ecosystems, their secondary metabolites are not diverse. Chemical profiling and bioactivity screening within the wider project that funded this thesis failed to highlight any glass sponges with interesting chemodiversity. The vast majority of metabolites present in their extracts were lipophilic cerebroside and desaturated fatty acid derivatives. The low abundance of interesting secondary metabolites is not unexpected. The large spicules of glass sponges, and high silica content of their biomass, provides a physical defence to deter predation. By having an effective physical defence, the need to produce a chemical defence is reduced. This physical defence also decreases the demand on the sponges limited chemical/ energy resources in these nutrient low environments. The low quantity of cellular tissues presents in glass sponges also limits the amount of extract yielded from a given specimen. Combined with the difficulty of collecting deep-sea hexactinellids, which are often very fragile and difficult to secure with robotic grips, little promise is shown for most glass sponges being a source of novel chemistry.

The exception to this trend, is the large number of deep-sea glass sponges which appear to have symbiotic relationships with zoanthids. Over three deep-sea sampling cruises, multiple specimens of glass sponges were collected with various species of zoanthids embedded into their structures: the chemistry of the zoanthids could prove to be more interesting.

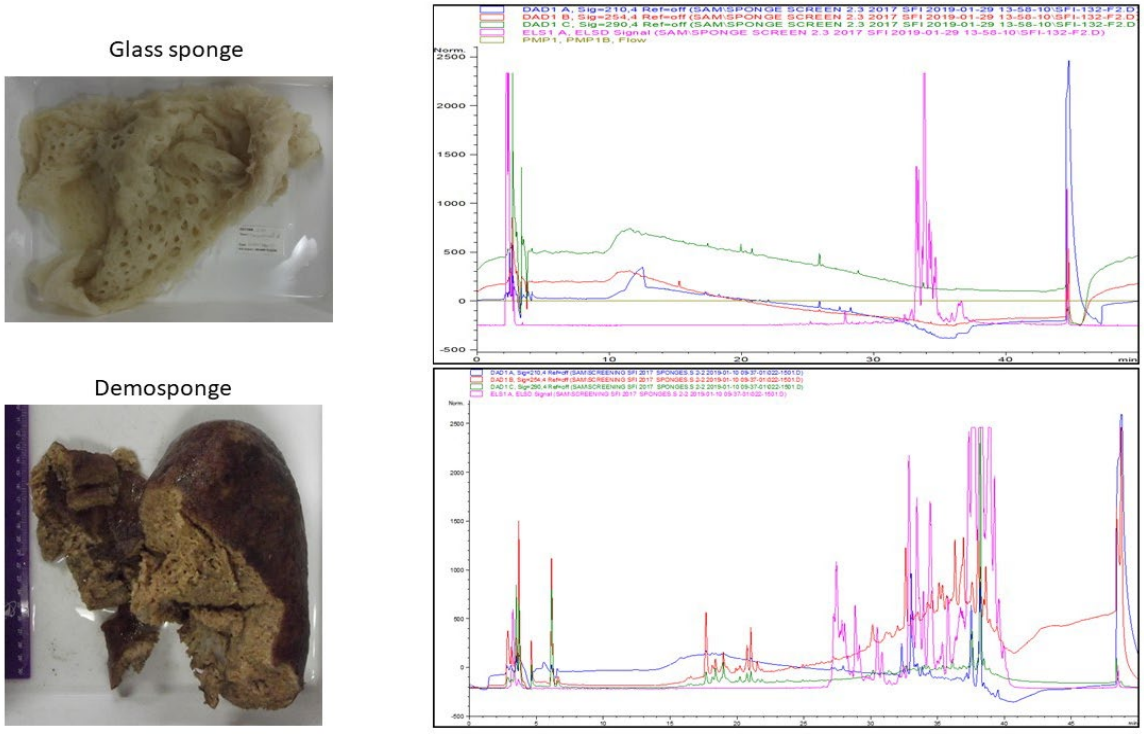


Figure 1. Examples of chemical profiling UHPLC-DAD-ELSD chromatograms of the methanolic fraction of a glass sponge (top) and a demosponge (bottom).

In contrast to glass sponges, deep-sea demosponges have been proven to host novel chemical structures, with many possessing the desired physiochemical properties and bioactivities to be developed into drug leads (figure 1). Of the demosponges, deep-sea sponges of the order Tetractinellida have been shown as rich source of novel secondary metabolites. The natural products isolated from these sponges include both compounds synthesised directly by the sponges, such as the barretides, and those of a microbial origin, for example polyketides such as the poecillastrins. This enhances the diversity of chemical classes isolated from tetractinellid sponges. A study into the microbiome composition of deep-sea sponge species found Actinobacteria, known to have large biosynthetic capacity, to be relatively abundant in tetractinellid sponges yet absent in other demosponges and glass sponges (figure 2)[1]. These complex species specific microbiomes have been shown to vary by depth and oceanographic conditions, resulting in changes to the metabolomic profiles of the sponges[2]. These findings highlight the potential for each deep-sea ecosystem to contain its own unique chemistry, dependent on the specific environmental condition.

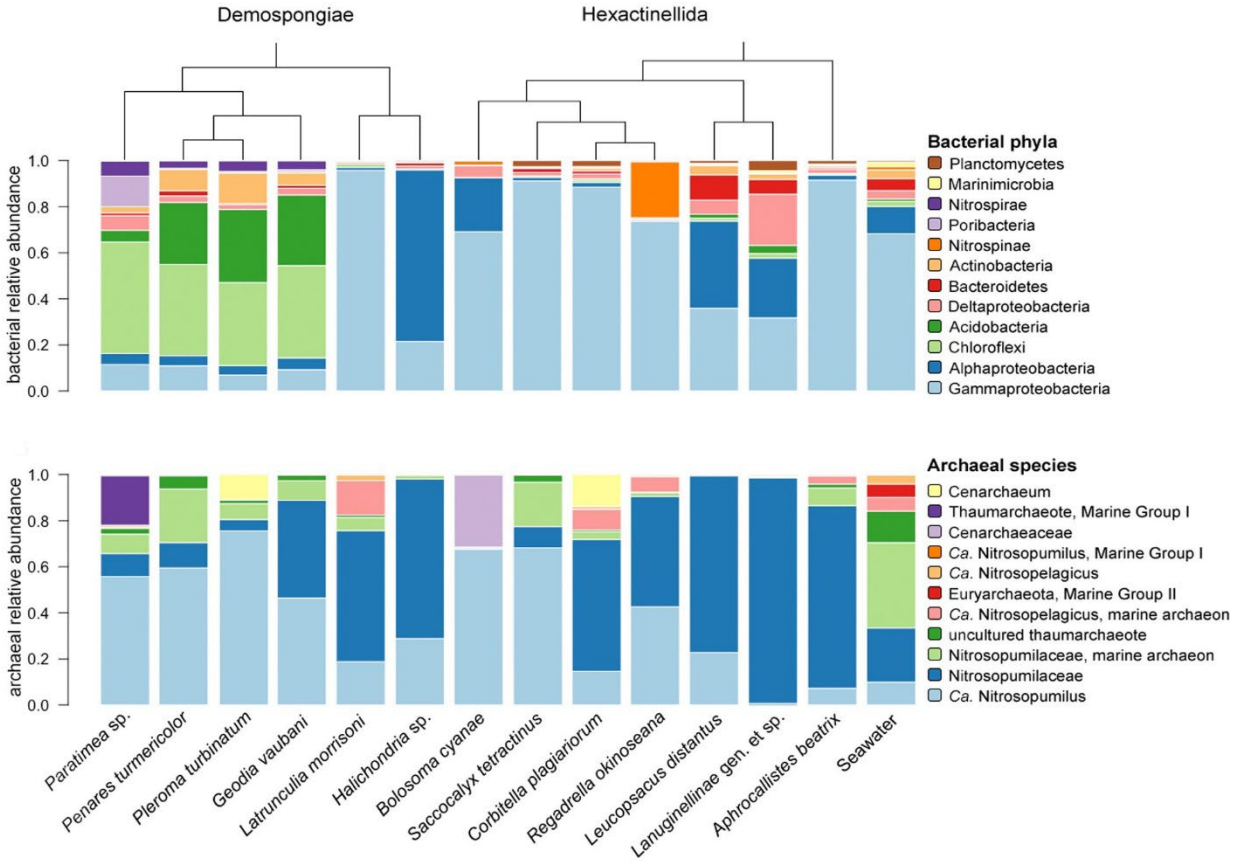


Figure 8. Relative abundance of the most abundant bacterial phyla(top) and archaeal species (bottom) from 13 sponges species (demosponges: left, glass sponges: right) and sea water.[1]

In our chemical screening of deep-sea sponges, multiple tetractinellid sponges showed promising chemical profiles, including the sponge *Characella pachastrelloides.* An in-depth chemical analysis of the chemodiversity of the sponge *C. pachastrelloides* led to the discovery of the novel glycolipopeptides, characellides A-D (Chapter 2). The characellides contained a rare 2-amino-2-deoxy-6-carboxamide sugar unit and an alkyl chain terminating with a dimethyl tetrahydropyran ring. A tripeptide moiety which connected the sugar moiety to the lipid moiety interestingly contained *D*-aspartic acid, and *D-allo*-threonine. The presence of *D* amino acids in nature are rare but they have been known to be produced by deep-sea microbes[3, 4]. This suggests a microbial

biosynthetic origin for the characellides. The poecillastrins present in these sponges along with the cyanocobalamin are known to be produced by bacteria also.

The synthesis of the reported structure of characellide B showed differences in NMR signals when compared to the natural products NMR shifts[5]. The differences of in both ^{1}H-NMR and ^{13}C-NMR signals for the aspartic acid at 12,13-position, and threonine at 16,17-position, were present in MeOH-d_4 and DMSO-d_4. These differences indicated the true absolute configuration of the characellides may differ from the absolute configuration in the reported structure. Amino acid stereoisomers of characellide B were synthesised with various combinations of *D* and *L* amino acids. None of the stereoisomers matched the chemical shift of the natural product. The strong experimental determination of the amino acid stereochemistry using the Marfey analysis combined with the results of the synthesised stereoisomers, indicated the true absolute configuration of the characellides either varies in the assignment of the sugar unit or the amine at 9-position. Although the stereochemistry for the dimethyl-THP in the proposed structure may be incorrect, the long alkyl chain separating the ring from the rest of the molecule would be unlikely to affect the chemical shifts of the amino acids. The assigned relative configuration of the sugar unit and amine were determined by comparison with experimental and theoretical ^{13}C-NMR chemical shift values using DP4 calculations. The absolute configuration was determined by comparing the experimental VCD spectra to the theoretical spectra of the two possible stereoisomers of the characellides. The assignment of the relative configuration which relied on computational methods is the most likely source of the error in the reported configuration.

The reliability of the assignment of the relative configuration would be improved by employing the new DP5 algorithm and increasing the conformer search window to a wider level of energy, from 1 Kcal/mol to 5 kcal/mol[6]. Experimental ROESY correlation can be used to inform restrictions in conformation searches to improve the accuracy of the optimised geometries, similar to the technique employed for the miolenols (Chapter 5). This approach should increase the accuracy of the theoretical VCD spectra. The high number of stereocentres and resulting possible stereoisomers of the characellides means determination of the absolute configuration is computationally resource intensive. This highlights the downside of this approach, but an issue which is increasing surmountable with easier access to supercomputers. To determine the absolute configuration experimentally, hydrolysis of the acetal bond to the sugar unit would allow for a comparison of UPLC retention times of the sugar unit with those of standards for the *L* and *D* forms of the sugar, when using an appropriate chiral column.

Analysis of the chemical space occupied by secondary metabolites present in the sponge *C. pachastrelloides* and secondary metabolites isolated from other deep-sea sponges, appears to show a trend of relatively “simple” modified primary metabolites (e.g. 6-methyl hercynine) or products of basic biosynthetic pathways (e.g. poecillastrins, barretins). The high degree of deep-sea sponge natural products with novel structures is aided by both the rarity of these modifications and the combination of these modified primary metabolites to form novel secondary metabolites (e.g. characellides,). An example of these rare modifications is seen in the interesting methylation of 6-methylhercyinene. The methylation at C-6 of hercynine, between nucleophilic *N-5* and *N-7* of the imidazole ring, indicates that the most common methylation pathway, involving radical SAM enzymes, was not the source of the methylation. This is the first natural product where an imidazole is methylated at this position, highlighting the rare biosynthetic pathways found in the deep-sea organisms. The presence of rare modification of primary metabolites is also exhibited in the amino sugar moiety and the presence of D amino acids seen in the characellides.

As deep sea reefs can be a low energy/nutrient environment, the ability to add simple modifications to primary metabolites, could be of benefit to deep-sea organisms, as it requires fewer chemical resources than more complex secondary metabolites. The rarity of the "simple" modification may also contribute to the strong biological activity of these natural products, as their similarity to primary metabolites may cause them to interfere with cell primary metabolic pathways.

The unique physical-chemical conditions in the deep sea (high pressure, low temperatures, and the absence of light) may allow for the production of natural products which would not be stable in standard laboratory conditions. The change of environmental conditions has been shown to affect the metabolome of sponge natural products, including oxidations and hydrolysis, when specimens are exposed to the air. While no studies have examined whether the change in environmental conditions degrades deep-sea sponge natural products, multiple deep-sea sponge natural products have been reported to be unstable. The poecillastrins, polyketides isolated from multiple deep-sea tetractinellid sponges have been shown to degrade under sun-light exposure and under acidic conditions[7]. Molecular networking could be a useful tool to examine whether degradation products of known metabolites are present in a sample, and to measure changes in the chemistry of a sample over time. This may indicate whether metabolites undergo chemical degradation, a factor that would increase the difficulty of isolation and characterisation of deep-sea natural products.

Many of the natural products isolated from deep-sea sponges are most likely produced by microbial biosynthetic pathways[8, 9]. The complexity of sponge microbiomes, containing undiscovered species of bacteria, archaea and fungi, and the extreme difficulty in culturing and isolating these deep-sea specific microbes has limited efforts to examine their biosynthetic potential. Although adapted to survive the high pressures encounter in the deep sea, a small portion (>0.1 %) of bacteria and fungi from deep-sea sponges have been cultured under standard laboratory condition, many of which are new strains. The past five years has seen a dramatic increase in the number for new compounds isolated from cultured deep-sea microbes, utilizing recent advances in our understand of epigenetics and biosynthetic gene clusters[10-12]. Multiple bioactive fungal extracts have been identified by Marchese et al., which strains were isolated from Irish deep-sea invertebrates[13, 14]. As the field of natural products is increasingly focusing on microbial natural products for novel chemical structures, deep-sea sponges act as a catalogue of possible secondary metabolites produced by deep-sea microbes.

6.2 Deep-sea corals

The diversity and species richness of deep-sea corals is impressive, especially those found in deep-sea reef communities. This diversity is observed in metabolomic richness of different species of coral, and even the same species collected from different depth and locations[15]. Many of the deep-sea coral natural products, both from *Paragorgia* (chapter 5) and from reviewing the literature, highlight an abundance and rich array of terpenic natural products[16]. This is somewhat to be expected as 90% of terpene marine natural products have been isolated from Cnidaria[17, 18]. The sessile natural of corals leads to the production of a chemical defence against predation and fouling[19]. This is a similar ecological niche to that filled by plants in terrestrial ecosystems. Interestingly, although vastly different forms of life, both corals and plants have been shown to produce the same secondary metabolites with anti-predation properties, including caffeine and prostaglandins[20, 21].

Reviewing deep-sea coral natural products in the literature and those isolated from *Paragorgia sp*., shows few novel chemical features distinguishing shallow water and deep-water coral metabolites. Unlike like deep-sea sponges, many of the natural products isolated from corals, in particular terpenoids, are products of the coral's biosynthetic pathways, although they do contain complex microbial assemblages adding to their metabolite diversity. An interesting correlation was noticed by Berrue et al., between the class of diterpenes and the phylogenetic groups of gorgonians they were isolated from[16]. This may be a contributing factor to similarity in metabolites between deep-sea coral and their shallow water counterparts, as well as the lower number of natural products isolated from deep-sea corals compared to sponges. The results from bioactivity screening of the wider funded project showed coral extracts were very bioactive (~10%), particularly in terms of cytotoxicity and antiplasmodial activity. The compounds responsible for these bioactivities are likely to play a role in the chemical defence of deep-sea corals.

Deep-sea soft corals have been shown to be a rich source of new bioactive terpenic natural products. Our work on *Paragorgia sp.* revealed the presence of eight bicyclic diterpenoids, comprised of an unusual bicyclo[7.2.0]undecane and either a new cyclobutanol ring, for the miolenols, or a lactone ring for the six xeniolides. The large number of xeniolides and other analogues we isolated from *Paragorgia sp.* and those in the literature highlight the diversity of metabolites that can be produced by the same precursor, by varying oxidations, reductions, and dehydrations to produce new metabolites.

Due to the high number of stereocentres, analysis of experimental and computational ECD spectra was well suited for determination of the absolute configuration. Compounds with a high degree of conformational flexibility, like miolenol, can be more difficult as the large number of possible conformers is computationally intensive. Using NOESY correlations to limit conformational flexibility allows for more accurate conformer searches, reducing the computational resource demand significantly. To further increase the reliability of the calculated ECD spectra, the employment of two different levels of theory is highly recommended. Our analysis of the absolute configuration of the xeniaphyllanes suggests that the early reported stereochemistry of the xeniaphyllanes is incorrect, as it was based on the absolute configuration of a similar plant terpene. It is of interest that a plant from the Himalayas and a deep-sea coral have evolved to produce similar chemical structures (stereoisomers) in vastly different ecosystems.

Most coral natural products have been derived from octocorals. However, black corals, which are hexacorals prevalent in deep-sea canyon systems, showed high levels of bioactivity (11%) in the extract screening process employed in the wider project. Of the black coral bioactive extracts, 35% showed cytotoxicity against cervical cancer and 41% displayed activity against tuberculosis. Attempts to isolate the metabolites which possess these bioactivities from the black coral *Bathypathes sp.* were unsuccessful, but NMR analysis indicated a large diversity of steroidal compounds.

6.3 Molecular Networking

Molecular networking, particularly the GNPS platform, has quickly grown to being the most used dereplication technique in natural products chemistry. A key pillar of this platform is the user-generated library of MS^2 spectra. This allows the platform to be applied to a diverse range of samples and has quickly grown a compound library of MS^2 spectra for the purposes of annotation.

GNPS-based molecular networking has been used in metabolomics, BGC research, chemical ecology, and toxicology, just to name a few[22-26].

Although widely adopted, multiple areas where improvement in the method could be made, have arisen. The presentation of nodes can show inaccuracies, such as some nodes being represented multiple times, inaccurate m/z and quantification data which relies on ion intensity[27]. The inability for classical molecular networking to resolve isobaric isomers was also noted[27]. Many research groups have been limited in their application of molecular networking due to poor coverage of certain chemotypes, for example, marine natural products. Multiple steps have been taken to improve the reproducibility and utilization to molecular networking, in terms of both improving input data quality and in the data analysis parameters.

Feature-based molecular networking (FBMN) was introduced to improve the reproducibility and accuracy of molecular networking[27]. This was achieved by incorporating a data-pre-processing step using third party programs such as MZmine3 and MS-DIAL, allowing for the integration of MS^1 and MS^2 data for generating networks[28, 29]. This pre-processing step, allows for the use of a metabolite's retention time, allowing for the separation of isomers in the networks, as well as improved accuracy in quantification of metabolites. Another key feature of many data pre-processing workflows is the ability the incorporate MS^2 spectra of the same metabolites with isotopic mass differences. Isotopic grouping is especially important from marine natural products as there is a high degree of halogenation, particularly for alkaloids[30-34]. Many of the data preprocesssing software packages allow for increased control of parameters. This provides users with a high degree of control and the ability to optimise molecular networking to their requirements. An example of one of these parameters, minimum intensity threshold, can be selected independently for MS^1 and MS^2, which can dramatically affect molecular network topology[35]. This allows users to generate a network that only includes major metabolites or incorporates more minor metabolites to aid in searching for new analogues, or biosynthetic precursors, as an example.

Within the software approach to improving molecular networking, continuous updates, and addition of new tools to mass spectrometry data processing packages, have improved the versatility and usability of molecular networking. An improved chromatogram builder, ADAP Module, was recently developed, is designed to reduce false positive and negatives in the resulting networks[36]. A major update to the open-source data processing package MZmine3 was published, with improvements to the user-friendly graphical interface and added visualization tools[30]. As molecular networking can be applied to a broad range of fields, a larger portion of GNPS-MN users will have only basic knowledge, if any, of mass spectrometry. For that reason, improvements to user-friendliness of the graphical interface and visualization tools are vital for improving quality of molecular networks and their uses. This same emphasis on a user-friendly approach, can be seen with the GNPS annotation platform. To add annotations to the GNPS library, previous interactions with the platform required users to extract a specific MS^2 spectra and upload it to the platform. Although not a difficult task, the multistep process required a steep learning curve. The most recent versions of the GNPS annotation workflow have dramatically improved the usability of this process, with automatic MS^2 spectrum selection when annotating a node. As the GNPS spectral library relies on user generated spectral libraries, the ease at which users can add to the library is vital for covering all chemotypes. Increased accessibility will also improve one of the main issues faced by GNPS users, the lack of spectral library coverage for certain classes of compounds.

The importance of the effect of LC-MS^2 data acquisition parameters on molecular networking appears to have been overlooked. Studies have shown that data-dependent acquisition (DDA) parameters affect molecular network topology but are limited in the number of parameters studied[37, 38]. Barbier et al., showed the energy selected for collision-induced dissociation affected MS^2 spectral quality[38]. The effect of intensity threshold, number of fragmented ions, and scan exclusion were examined by Olivon et al., and were shown to affect classical molecular networking topology[39]. Olivon et al,. introduced the use of six molecular network characteristics that can be used to analyse the quality of a molecular network. These six characteristics can be optimised to produce a network that will best fit the purpose of the molecular network. The effect of these parameters was confirmed by Xu et al,. who examined the parameters influence on Classical Molecular Networking (CLMN) topology in both positive and negative ionisation modes[40]. Although these studies showed molecular networks are affected by selected acquisition parameters, the effect of multiple other parameters such as liquid chromatography were not examined. This limitation is most likely due to the practical difficulties in running a large multiparameter study using the traditional one-factor-at-a-time (OFAT) approach.

Our study into optimizing data acquisition parameters (Chapter 3) built upon the knowledge gained from previous studies, including the how to quantify molecular network topology changes and an indication of which MS data-acquisition parameters to measure. To improve on previous works, we aimed to see the effects of a broad range of data-acquisition parameters. To overcome the practical issues with testing eleven parameters, a Fractional Factorial Design of Experiment (DoE) experimental strategy was followed. Not only was the number of test conditions reduced, but this DoE also allows for the statistical analysis of the significance of a parameter, and the significance of interactions between parameters. As different chemical classes require different MS^2 acquisition parameters from the best quality MS^2 spectra, our study incorporated three diverse marine extracts.

Our analysis of factor significance confirmed the influence of collision energies on MS^2 spectra and molecular network quality, as found in previous studies. This highlights parameters that affect the parent ion fragmentation process as key in optimizing molecular networks. Our research also found sample concentration, chromatographic separation, and precursors per cycle, as significant parameters to be optimized. These three parameters point to a new area of focus for molecular network optimization, which is competition for precursor ion selection. Reducing competition between precursors to be fragmented improved molecular network accuracy and reproducibility. This can be achieved by increasing sample concentration and chromatographic separation or by increasing the number of precursors per cycle.

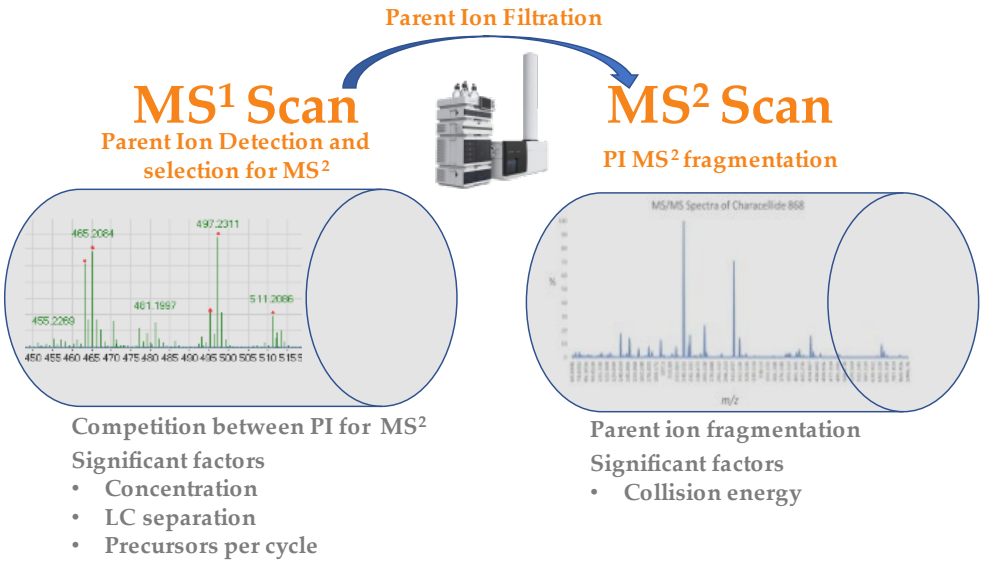


Figure 3. Two processes of MS^2 data acquisition that effect molecular networking topology. MS^1 scan when parent ions are detected and selected from MS^2 fragmentation (left). This process is affected by sample concentration, LC separation and the number of precursors per cycle. MS^2 scan (right), when the parent ion is fragmented, and is significantly affected by collision energy.

The DoE experimental strategy was shown to be a time-efficient technique for identifying significant parameters and may prove to be a good template for further optimization studies, both on data-acquisition and data-processing parameters. As different chemical structures respond to LC-MS^2 conditions differently, when carrying out an in-depth analysis, optimizations of parameters should be carried out on a per sample basis. Further studies to optimize both the data acquisition and processing parameters, could massively improve the versatility of molecular networking as a tool for metabolomics, drug discovery and toxicology.

The increase in quality of MS^2 spectra seen after LC-MS^2 parameter optimization can improve the accuracy of annotations. This remains a key area of study, to improve the accuracy and validity of annotation. Integration of the SMART NMR platform could help verify annotations by highlighting unique HSQC signals that would correspond to the annotated compound. New workflows like the Feature Based-ion identity molecular network (IIMN) show great potential for increasing the ability to annotate molecular networks with a high degree of accuracy and validity[41]. IIMN can also be used to identity ion-ligand complexes and improves MS^2 spectra quality by connecting and collapsing different ion species resulting from the same metabolite[41].

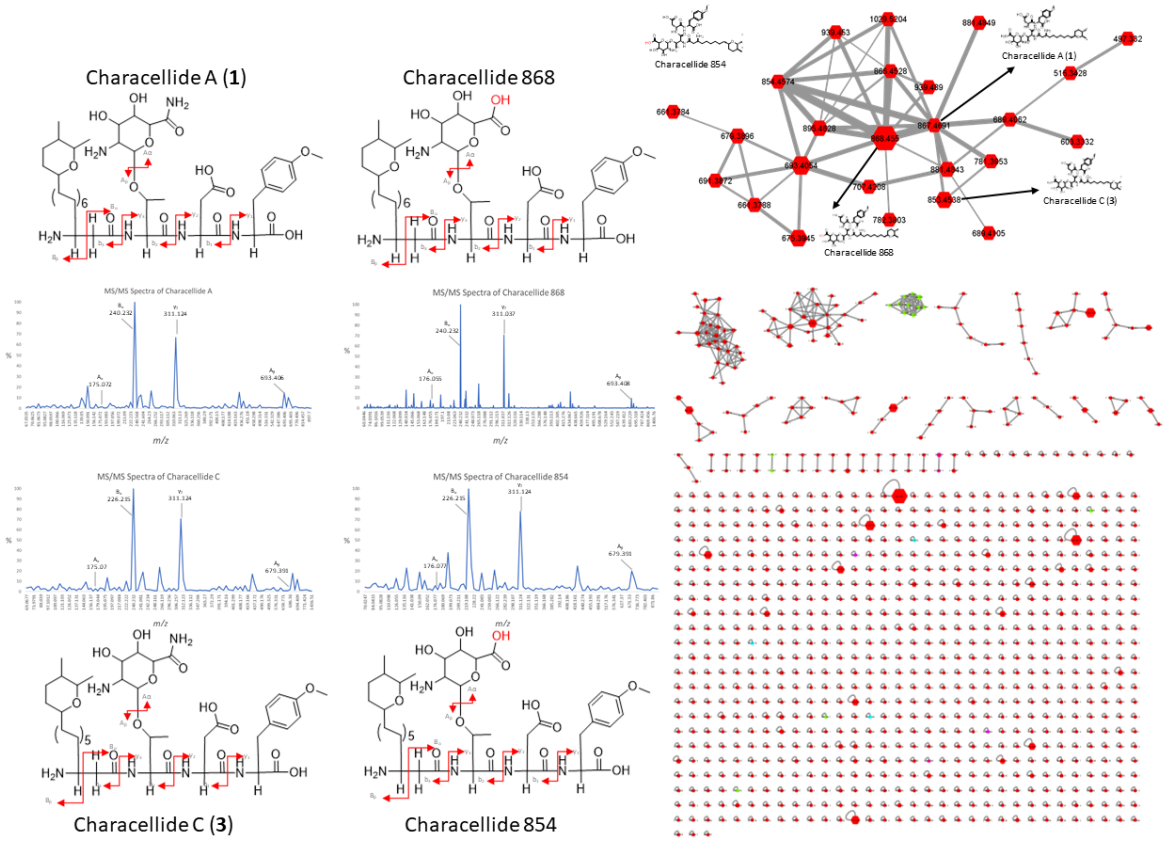


Figure 4. Example of analogue discovery using molecular networking. A cluster containing the characellides is highlighted on the right. MS^2 spectra of characellides are displayed (left) with their corresponding fragmentation pattern.

GNPS based classical molecular networking was shown to be a useful tool for detection and isolation of the characellides (Chapter 2). Analysis of the fragmentation patterns of nodes in the characellides cluster allowed the structure of eight additional characellide analogues to be hypothesised and the targeted isolation of characellides C and D. In depth examination of characellide nodes and their corresponding MS^2 spectra, revealed inaccuracies within the network including duplication of nodes, inaccurate mass data and inability to separate stereoenantiomers, characellides A and B, and characellides C and D. Implementation of the feature based molecular networking workflow, significantly improved the quality of the *Characella pachastrelloides* molecular network (Figure 26). The accuracy of nodes m/z and corresponding MS^2 spectra was greatly improved. The CLMN nodes corresponding to characellides A/B and C/D were releveled by FBMN to each represent two separate characellides, with sugar units containing one or two amides, evidenced by their MS^2 spectra. The higher quality output from Feature-Based Molecular Networking improves versatility in understanding the chemical space occupied by an organism's natural products.

The rapid improvement and iterations seen with the GNPS-based molecular networking platform highlights its potential in natural products chemistry. The ability to optimize molecular networks to provide easy visualization of chemical space aids in the versatility of this tool.

6.4 Deep-sea Natural products and their importance in the pharmacological space.

Since the early 2000s, untreatable microbial infections due to antibiotic resistant bacteria, has risen dramatically. In 2019, a global study found 4.95 million deaths resulted from antibiotic resistant microbial infections, surpassing the number of deaths caused by HIV/AIDS and malaria combined[42]. Over this same period of time the number of new antibiotics entering the market has reduced significantly, with only one new mechanistic class of antibiotic discovered in the past 10 years[43]. The cause of the rise of antibiotic resistant bacteria is clear: misuse of antibiotics (overprescribing, patients not finishing prescribed courses) and an inevitability for antibiotic resistance to spread given enough time. As a result, the useful lifespan of an antibiotic has reduced [44]. This reduced useful lifespan, limits the potential to recover the cost associated with bringing a new drug to market. The higher financial risk associated with antibiotic drug development has discouraged many pharmaceutical companies from investing in bringing new antibiotics to market.

As the historically rapid development of vaccines against SARS-COVD-19 shows, a high level of demand, and government investment into drug discovery, can play a key role in the rapid development of new medicines by the pharmaceutical industry. With illness and deaths due to multiple drug resistant bacteria increasing annually, and the number of effective treatments decreasing, we may soon reach a tipping point, where demand for new antibiotics drives similar government investment.

Natural products have always been a rich source of antibiotics, with over half of antibiotics being natural products or synthetic analogy of natural products. This may be result from lifeforms without complex immune systems requiring a chemical defence against microbial infections.

Recent advances in high throughput co-culturing of microbial communities, can encourage and enhance the production of antimicrobial compounds with drug lead potential. This could help address the “supply issue”, by focusing drug discovery effects on culturable bacteria, and might reduce the costs associated with manufacturing and production of antibiotics, decreasing the financial risks associated with antibiotic research and development.

6.5 References

1. Steinert, G.; Busch, K.; Bayer, K.; Kodami, S.; Arbizu, P. M.; Kelly, M.; Mills, S.; Erpenbeck, D.; Dohrmann, M.; Wörheide, G., Compositional and quantitative insights into bacterial and archaeal communities of South Pacific deep-sea sponges (Demospongiae and Hexactinellida). *Frontiers in microbiology* **2020,** 11, 716.
2. Steffen, K.; Indraningrat, A. A. G.; Erngren, I.; Haglöf, J.; Becking, L. E.; Smidt, H.; Yashayaev, I.; Kenchington, E.; Pettersson, C.; Cárdenas, P., Oceanographic setting influences the prokaryotic community and metabolome in deep-sea sponges. *Scientific reports* **2022,** 12, (1), 1-16.
3. Radkov, A. D.; Moe, L. A., Bacterial synthesis of D-amino acids. *Applied microbiology and biotechnology* **2014,** 98, (12), 5363-5374.
4. Kubota, T.; Kobayashi, T.; Nunoura, T.; Maruyama, F.; Deguchi, S., Enantioselective utilization of D-amino acids by deep-sea microorganisms. *Frontiers in microbiology* **2016,** 7, 511.
5. Wang, Y.; Wang, Z.; Wang, Z.; Liu, X.; Jiang, Y.; Jiao, X.; Xie, P., Total Synthesis of the Proposed Structure of Characellide B. *Organic Letters* **2021,** 23, (9), 3680-3684.
6. Howarth, A.; Goodman, J. M., The DP5 probability, quantification and visualisation of structural uncertainty in single molecules. *Chemical Science* **2022,** 13, (12), 3507-3518.

7. Suo, R.; Takada, K.; Irie, R.; Watanabe, R.; Suzuki, T.; Ise, Y.; Ohtsuka, S.; Okada, S.; Matsunaga, S., Poecillastrin H, a chondropsin-type macrolide with a conjugated pentaene moiety, from a Characella sp. marine sponge. *J. Nat. Prod.* **2018,** 81, (5), 1295-1299.
8. Saide, A.; Lauritano, C.; Ianora, A., A Treasure of Bioactive Compounds from the Deep Sea. *Biomedicines* **2021,** 9, (11), 1556.
9. Esposito, R.; Federico, S.; Bertolino, M.; Zupo, V.; Costantini, M., Marine Demospongiae: A Challenging Treasure of Bioactive Compounds. *Mar. Drugs* **2022,** 20, (4), 244.
10. Anh, C. V.; Yoon, Y. D.; Kang, J. S.; Lee, H.-S.; Heo, C.-S.; Shin, H. J., Nitrogen-Containing Secondary Metabolites from a Deep-Sea Fungus Aspergillus unguis and Their Anti-Inflammatory Activity. *Mar. Drugs* **2022,** 20, (3), 217.
11. Chen, J.; Chen, J.; Wang, S.; Bao, X.; Li, S.; Wei, B.; Zhang, H.; Wang, H., Amycolachromones A–F, Isolated from a Streptomycin-Resistant Strain of the Deep-Sea Marine Actinomycete Amycolatopsis sp. WP1. *Mar. Drugs* **2022,** 20, (3), 162.
12. Wang, J.; Li, T.; Wang, P.; Ding, W.; Xu, J., Tanzawaic Acids from a Deep-Sea Derived Penicillium Species. *Journal of Natural Products* **2022**.
13. Marchese, P.; Mahajan, N.; O'Connell, E.; Fearnhead, H.; Tuohy, M.; Krawczyk, J.; Thomas, O. P.; Barry, F.; Murphy, M. J., A Novel High-Throughput Screening Platform Identifies Itaconate Derivatives from Marine Penicillium antarcticum as Inhibitors of Mesenchymal Stem Cell Differentiation. *Mar. Drugs* **2020,** 18, (4), 192.
14. Marchese, P.; Young, R.; O'Connell, E.; Afoullouss, S.; Baker, B. J.; Allcock, A. L.; Barry, F.; Murphy, J. M., Deep-sea coral garden invertebrates and their associated fungi are genetic resources for chronic disease drug discovery. *Mar. Drugs* **2021,** 19, (7), 390.
15. Vohsen, S. A.; Fisher, C. R.; Baums, I. B., Metabolomic richness and fingerprints of deep-sea coral species and populations. *Metabolomics* **2019,** 15, (3), 1-13.
16. Berrue, F.; Kerr, R. G., Diterpenes from gorgonian corals. *Natural Product Reports* **2009,** 26, (5), 681-710.
17. Sang, V. T.; Dat, T. T. H.; Vinh, L. B.; Cuong, L. C. V.; Oanh, P. T. T.; Ha, H.; Kim, Y. H.; Anh, H. L. T.; Yang, S. Y., Coral and coral-associated microorganisms: A prolific source of potential bioactive natural products. *Mar. Drugs* **2019,** 17, (8), 468.
18. Kornprobst, J. M., Cnidaria and Ctenophora–1. *Encyclopedia of marine natural products* **2014**, 1087-1208.
19. Bright-Diaz, L. M.; Strychar, K. B.; Shirley, T. C., Compounds from deep-sea bubblegum corals, Paragorgia arborea, elicit anti-predation behavior in fish. *Open Marine Biology Journal* **2011,** 5, 58-67.
20. Imre, S.; Öztunç, A.; Çelik, T.; Wagner, H., Isolation of caffeine from the gorgonian Paramuricea chamaeleon. *Journal of natural products* **1987,** 50, (6), 1187-1187.
21. Bayer, F. M.; Weinheimer, A. J. *Prostaglandins from Plexaura homomalla: Ecology, utilization and conservation of a major medical marine resource. A symposium*; University of Miami: 1974.
22. Patiño, A. D.; Montoya-Giraldo, M.; Quintero, M.; López-Parra, L. L.; Blandón, L. M.; Gómez-León, J., Dereplication of antimicrobial biosurfactants from marine bacteria using molecular networking. *Scientific reports* **2021,** 11, (1), 1-12.
23. Silva, E.; da Graça, J. P.; Porto, C.; Martin do Prado, R.; Hoffmann-Campo, C. B.; Meyer, M. C.; de Oliveira Nunes, E.; Pilau, E. J., Unraveling Asian Soybean Rust metabolomics using mass spectrometry and Molecular Networking approach. *Scientific reports* **2020,** 10, (1), 1-11.
24. Paulo, B. S.; Sigrist, R.; Angolini, C. F.; De Oliveira, L. G., New cyclodepsipeptide derivatives revealed by genome mining and molecular networking. *ChemistrySelect* **2019,** 4, (27), 7785-7790.
25. Duncan, K. R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B. S.; Dorrestein, P. C., Molecular networking and pattern-based

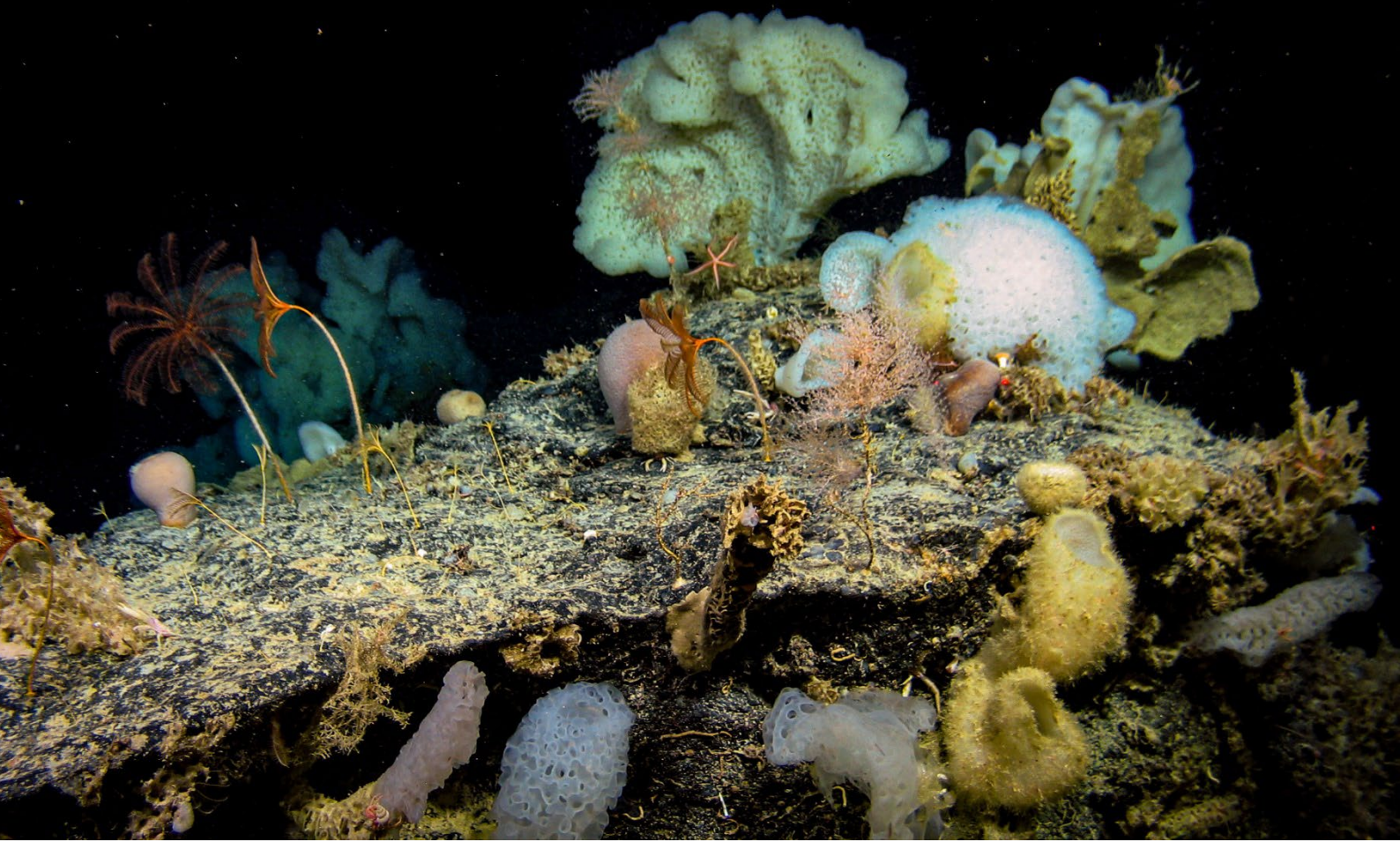
genome mining improves discovery of biosynthetic gene clusters and their products from Salinispora species. *Chemistry & biology* **2015,** 22, (4), 460-471.

26. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M., Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences* **2012,** 109, (26), E1743-E1752.
27. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M., Feature-based molecular networking in the GNPS analysis environment. *Nature methods* **2020,** 17, (9), 905-908.
28. Pluskal, T.; Castillo, S.; Villar-Briones, A., Oreši c M.(2010). MZmine.
29. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods* **2015,** 12, (6), 523-526.
30. Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D., MZmine 2 data-preprocessing to enhance molecular networking reliability. *Analytical chemistry* **2017,** 89, (15), 7836-7840.
31. Zhang, H.; Khalil, Z.; Conte, M. M.; Plisson, F.; Capon, R. J., A search for kinase inhibitors and antibacterial agents: bromopyrrolo-2-aminoimidazoles from a deep-water Great Australian Bight sponge Axinella sp. *Tetrahedron Lett.* **2012,** 53, (29), 3784-3787.
32. Butler, A.; Carter-Franklin, J. N., The role of vanadium bromoperoxidase in the biosynthesis of halogenated marine natural products. *Natural Product Reports* **2004,** 21, (1), 180-188.
33. Tsujii, S.; Rinehart, K. L.; Gunasekera, S. P.; Kashman, Y.; Cross, S. S.; Lui, M. S.; Pomponi, S. A.; Diaz, M. C., Topsentin, bromotopsentin, and dihydroxybromotopsentin: antiviral and antitumor bis(indolyl)imidazoles from Caribbean deep-sea sponges of the family Halichon. *J. Org. Chem.* **1988,** 53, 5446-5453.
34. Li, M. K. W.; Scheuer, P. J., Halogenated blue pigments of a deep sea gorgonian ; chloro- and bromoazulenes. *Tetrahedron Lett.* **1984,** 25, 587-590.
35. Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F., Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature protocols* **2020,** 15, (6), 1954-1991.
36. Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X., One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: new algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks. *Analytical chemistry* **2017,** 89, (17), 8696-8703.
37. Guo, J.; Huan, T., Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics. *Analytical Chemistry* **2020,** 92, (12), 8072-8080.
38. Barbier Saint Hilaire, P.; Rousseau, K.; Seyer, A.; Dechaumet, S.; Damont, A.; Junot, C.; Fenaille, F., Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics. *Metabolites* **2020,** 10, (4), 158.
39. Olivon, F.; Roussi, F.; Litaudon, M.; Touboul, D., Optimized experimental workflow for tandem mass spectrometry molecular networking in metabolomics. *Analytical and bioanalytical chemistry* **2017,** 409, (24), 5767-5778.
40. Xu, R.; Lee, J.; Chen, L.; Zhu, J., Enhanced detection and annotation of small molecules in metabolomics using molecular-network-oriented parameter optimization. *Molecular Omics* **2021,** 17, (5), 665-676.
41. Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K., Ion identity molecular networking for mass spectrometry-

based metabolomics in the GNPS environment. *Nature communications* **2021,** 12, (1), 1-12.

42. Laxminarayan, R., The overlooked pandemic of antimicrobial resistance. *The Lancet* **2022**.
43. Mitcheltree, M. J.; Pisipati, A.; Syroegin, E. A.; Silvestre, K. J.; Klepacki, D.; Mason, J. D.; Terwilliger, D. W.; Testolin, G.; Pote, A. R.; Wu, K. J. Y.; Ladley, R. P.; Chatman, K.; Mankin, A. S.; Polikanov, Y. S.; Myers, A. G., A synthetic antibiotic class overcoming bacterial multidrug resistance. *Nature* **2021,** 599, (7885), 507-512.
44. Dowling, A.; O'Dwyer, J.; Adley, C., Antibiotics: mode of action and mechanisms of resistance. *Antimicrobial research: Novel bioknowledge and educational programs* **2017,** 1, 536-545.

Conclusions



Conclusions

The deep-sea sponges present in the NE-Atlantic cover a broad range of taxonomic classes and orders. The two most prevalent classes, Hexactinellida and Demospongiae, vary greatly with regards to the presences of secondary metabolites. The low diversity and abundance of metabolites seen in glass sponges is most likely due to their strong physical defence in the form of spicules, and low microbial diversity. As a result, the majority of natural products isolated from glass sponges are derivatives of cell membrane components, with little pharmaceutical potential.

Demosponges, on the other hand, have been shown to possess an impressively large chemodiversity, including polypeptides, alkaloids, terpenes, polyketides, bisindoles, new amino acids, and fatty acid derivatives. The chemodiversity seen within demosponges appears to be due to two main factors. The first is the large diversity of bacterial taxa, archaeal species, and fungal strains, which are present in the microbiome of demosponges. Multiple studies have identified deep-sea microbes with unique biosynthetic gene clusters, adapted to the harsh environmental conditions found in the deep sea. The second is the high level of competition in resource-limited deep-sea ecosystems between sponges and other animals and within the complex microbiomes of individual sponges. This forces deep-sea sponges, and associated microbes, to produce a diverse range secondary metabolites for chemical defence, communication, and to aid symbiotic relationships.

A large number of these metabolites contain novel chemical scaffolds, as well as possessing potent biological activities. The deep-sea tetractinellid sponge *Characellida pachiosteroides* is a perfect case-study, highlighting the range of chemodiversity possible in the deep sea. The characellides, with their unique sugar moiety, and rare D-amino acids and tetrahydropyran ring are of great interest. Their high number of stereocentres presents an intriguing challenge, for both assigning their absolute configuration and as a target for synthesis. Further analysis is required to confirm or reassign the absolute configuration of the characellides. The proposed characellide analogues, detected through molecular networking, are also of interest as synthetic targets and future isolation efforts. Similarly, 6-methylhercyinine, poses interesting questions, in terms of its biosynthesis and the unique methylation of the imidazole ring. This compound may also be a challenge for synthetic chemists to replicate. The presence of poecillastrin H and cyanocobalamin in this sponge indicated the vital role the microbial community plays in deep-sea sponge chemodiversity.

Both the characellides and 6-methylhercyinine possess potent and desirable biological activities. The nanomolar anti-inflammatory activity reducing ROS production of characellide A, and additional study into the ability of characellides to modulate immune responses, could lead to its development as a drug lead candidate[1, 2]. The synthesis of characellide B[3] further increases its pharmaceutical potential as acquiring more of the natural product from deep-sea specimens for further biological analysis would be logistically difficult and cost prohibitive.

Deep-sea corals are both rich in species and showed strong activities when extracts were tested against a wide range of bioassays, indicating they are a good potential source of marine natural products with pharmaceutical potential. Soft corals in particular, appear to have the most promise, containing a large range of terpenoid derived metabolites. Our chemical investigation into the gorgonian *Paragorgia arborea* is a showcase of soft corals terpene richness, yielding eight diterpenoids of the xenicane class, three of which were new structures. Miolenol and epoxymiolenol both contain a unique cyclobutanol ring. The high degree of flexibility in the cyclononane ring leading to interesting conformational stability in miolenol. All metabolites isolated from the *Paragorgia arborea* showed mild antiplasmodial activities against both NF54 and Dd2 strains of *Plasmodium falciparum.* 9-deoxyxeniolide A displayed good activity (>50%) against the resistant Dd2 strain, which

could be of interest for further studies. Miolenol A possessed cytotoxicity against J744 macrophages, with inhibitions of ~ 65% when tested at 10 µL/mL.

Molecular networking has become an indispensable tool in the search for natural products. The ability to examine the chemical space occupied by organism's metabolome and rapidly determine the known metabolites though dereplication has led to it wide adoption in the natural products chemistry field. Since its introduction in 2012, GNPS based molecular networking has seen a wide range of improvements to its workflow, increasing it usability and accuracy. While most of the effort to optimize and improve molecular networking were focused on the software approach (Feature-Based Molecular Networking[4]), the quality and effected of MS^2 data-acquisition was often overlooked. Our study into the effect of LC-MS^2 parameters highlighted the importance of selecting appropriate concentration, LC separation, precursors per cycle, and collision energy for the desired information from molecular networking. The use of molecular networking to help predict structures of analogues characellides C/D and carry out a targeted isolation, reinforces its usefulness in the drug discovery process.

The chemical investigation into the pharmaceutical potential of deep-sea sponges and corals, has unveiled a large range of new and novel metabolites, spanning multiple chemical classes. The presence of rare and/or undiscovered chemical features as well as strong bioactivaties, highlight the deep-sea as a promising playground in the search for new drugs.

References

1. Marcella, S.; Afoullouss, S.; Thomas, O. P.; Allcock, A. L.; Murphy, P. V.; Loffredo, S., Immunomodulatory properties of characellide A on human peripheral blood mononuclear cells. *Inflammopharmacology* **2021,** 29, (4), 1201-1210.
2. Afoullouss, S.; Calabro, K.; Genta-Jouve, G.; Gegunde, S.; Alfonso, A.; Nesbitt, R.; Morrow, C.; Alonso, E.; Botana, L. M.; Allcock, A. L.; Thomas, O. P., Treasures from the Deep: Characellides as Anti-Inflammatory Lipoglycotripeptides from the Sponge Characella pachastrelloides. *Organic Letters* **2019,** 21, (1), 246-251.
3. Wang, Y.; Wang, Z.; Wang, Z.; Liu, X.; Jiang, Y.; Jiao, X.; Xie, P., Total Synthesis of the Proposed Structure of Characellide B. *Organic Letters* **2021,** 23, (9), 3680-3684.
4. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M., Feature-based molecular networking in the GNPS analysis environment. *Nature methods* **2020,** 17, (9), 905-908.

Appendix

Appendix: Chapter 5 Supplementary Information

Table of figures

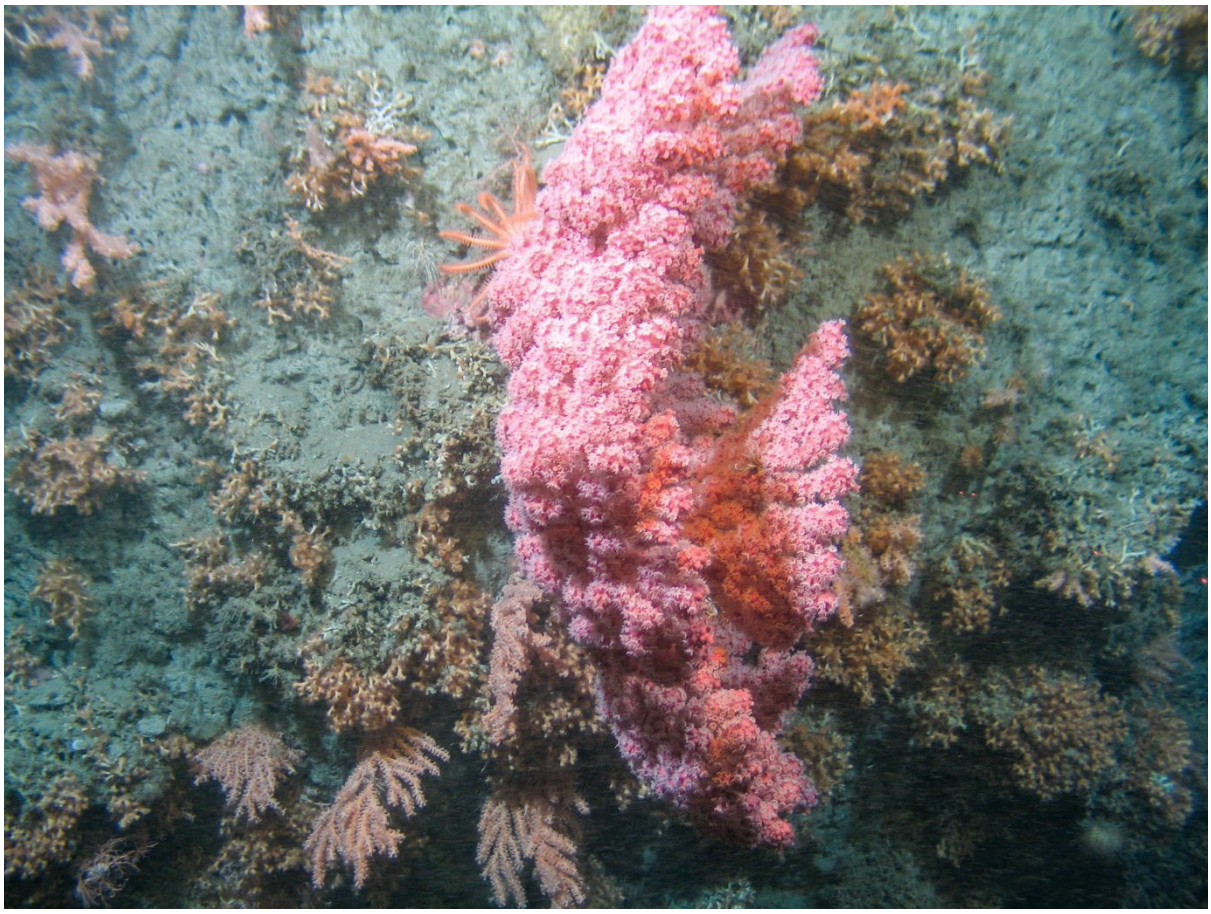


Figure S1: In-situ photograph of *Paragorgia arborea* taken by the ROV *Holland 1*, at a depth of 1,500 m in Whittard canyon.

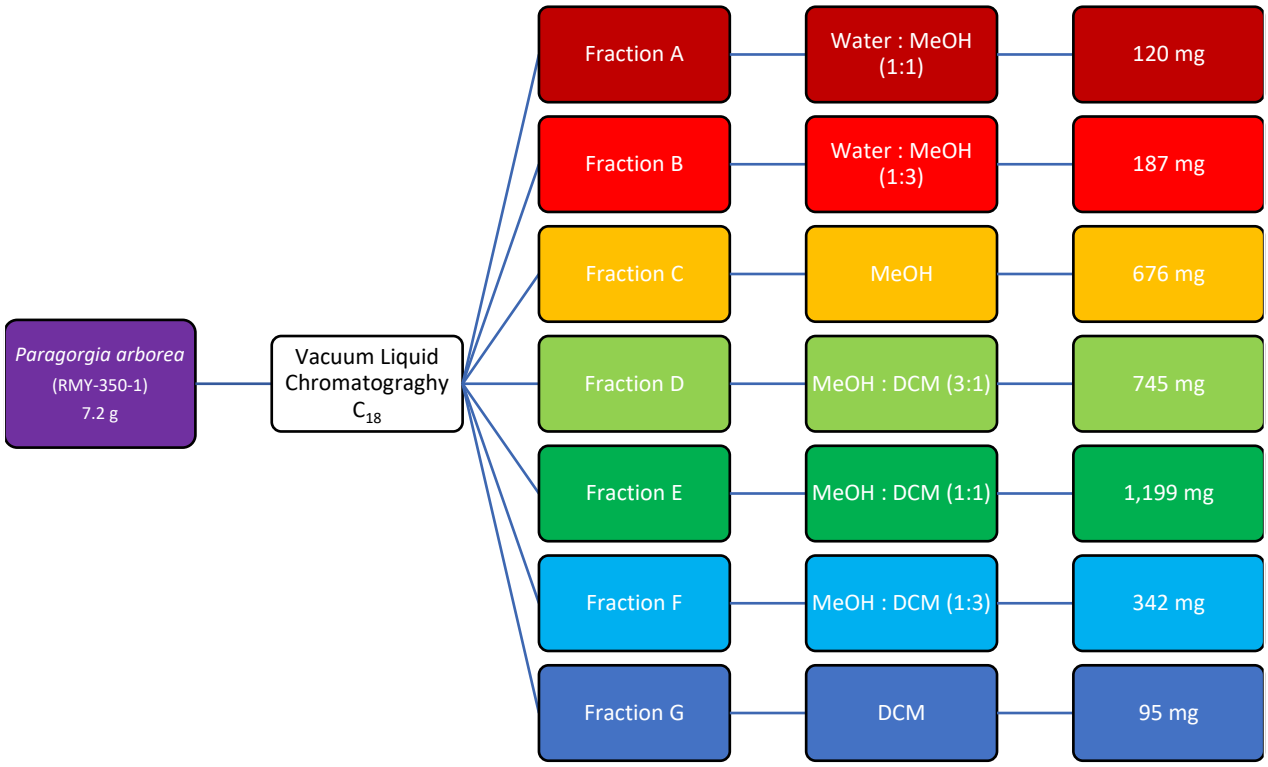


Figure S2: Solvent used for reversed phase C_{18} vacuum liquid chromatography fractionation and *Paragorgia arborea* extract and the mass of each fraction in milligrams.

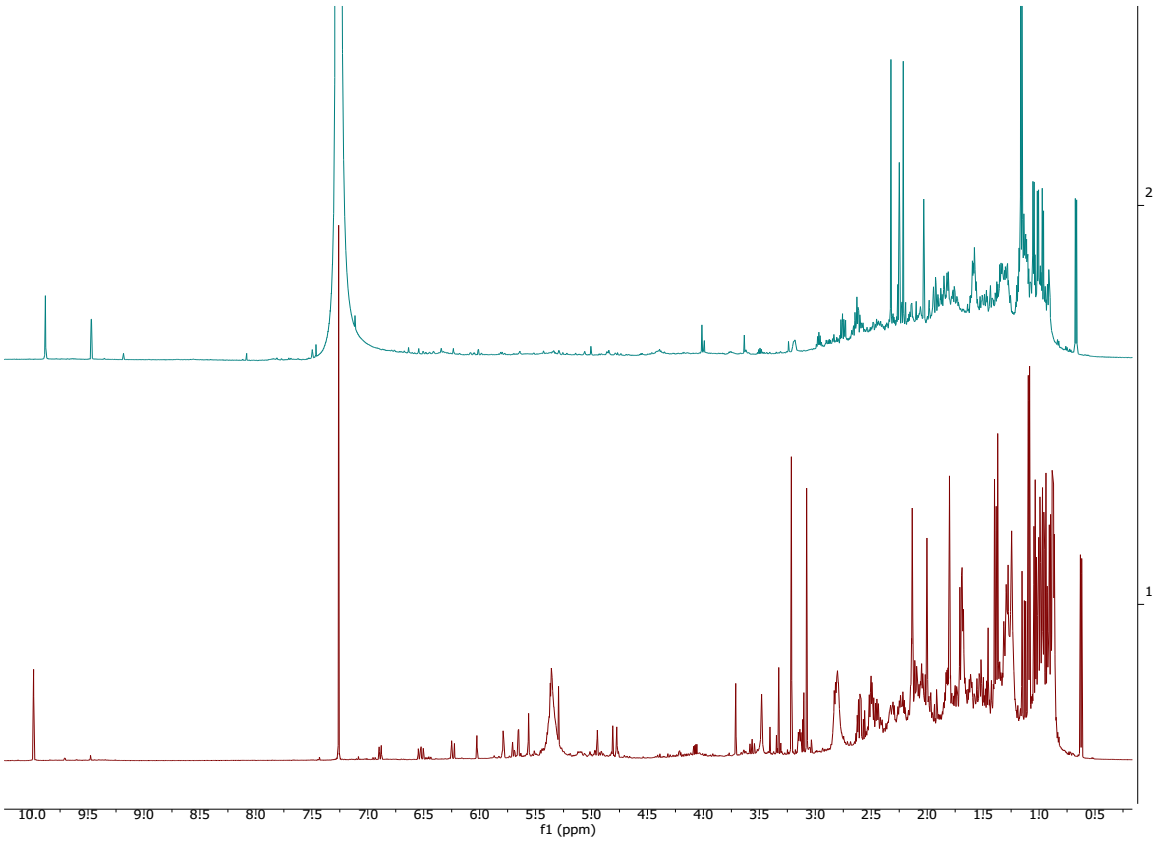


Figure S3: ^{1}H NMR in $CDCl_3$ of subfraction 2 after purification with acidified solvent (top) and non-acidified solvents (bottom) with the non-acidified sample show more signals in the olefinic region.

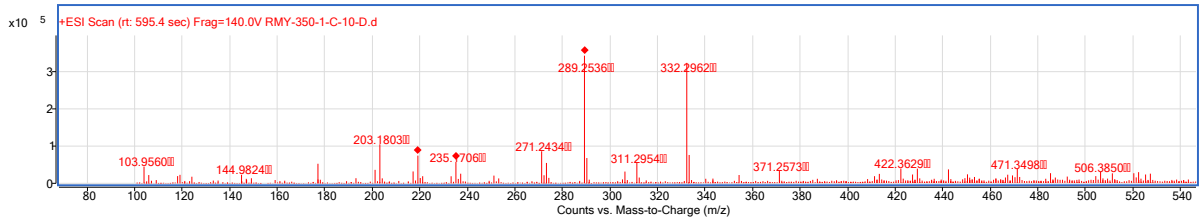


Figure S4: HRESIMS spectrum of miolenol (**1**).

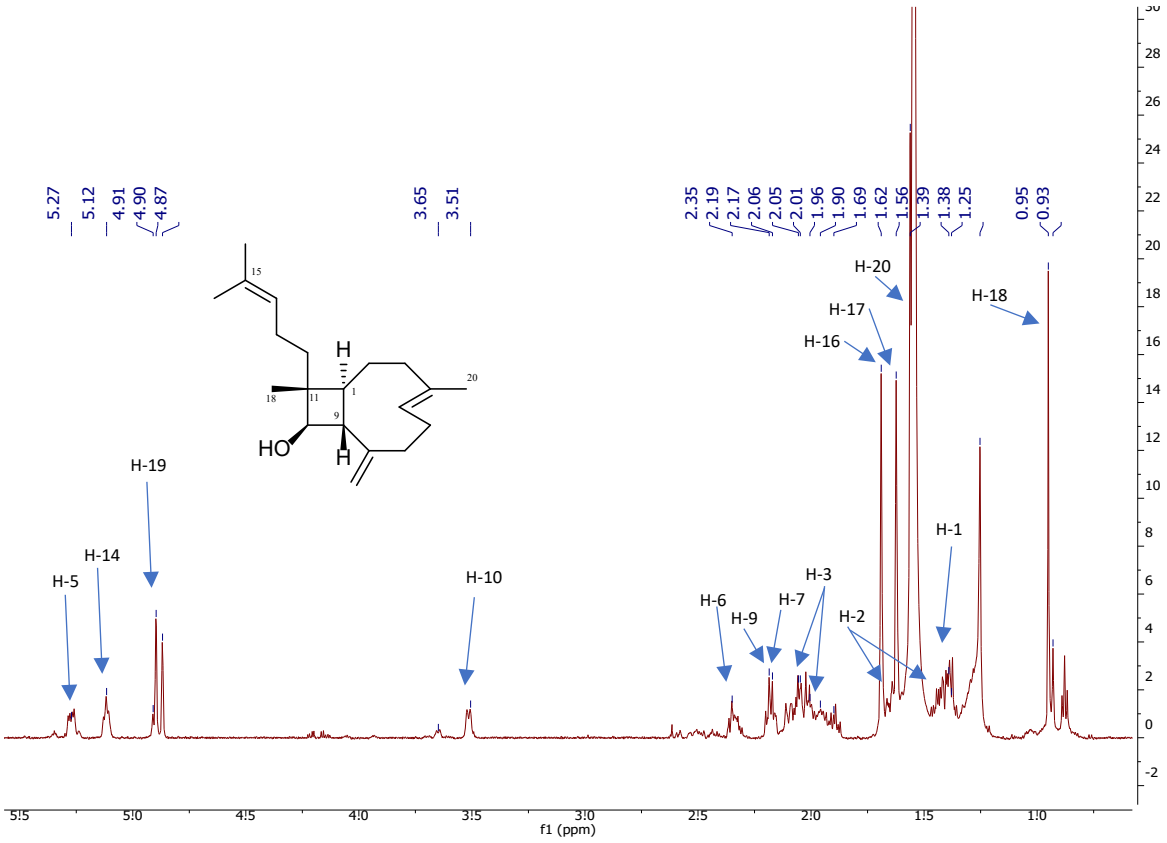


Figure S5: ^{1}H NMR spectrum of miolenol (**1**) in $CDCl_3$ (600 MHz).

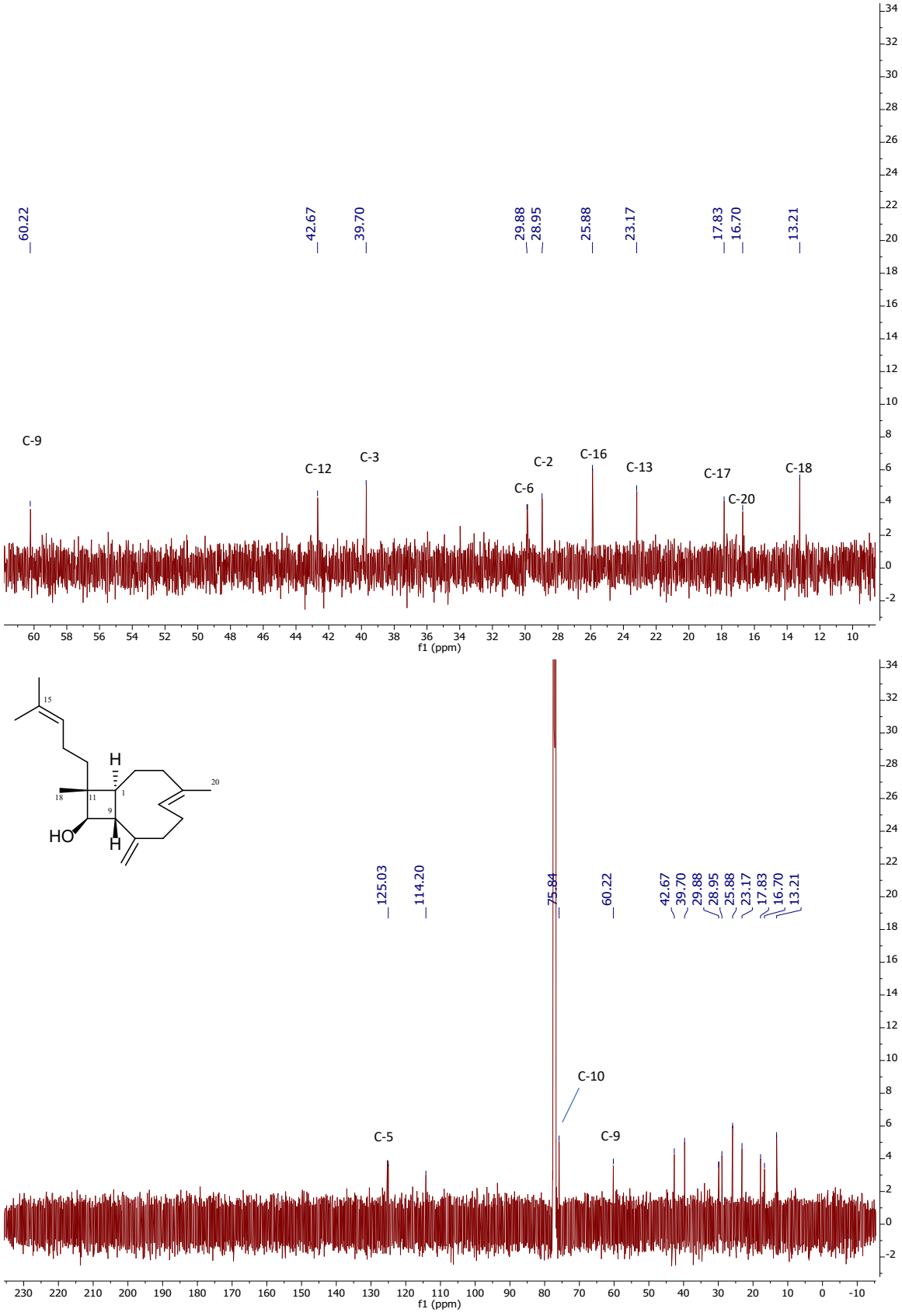


Figure S6:^{13}C NMR spectrum of miolenol (**1**) in $CDCl_3$ (150 MHz).

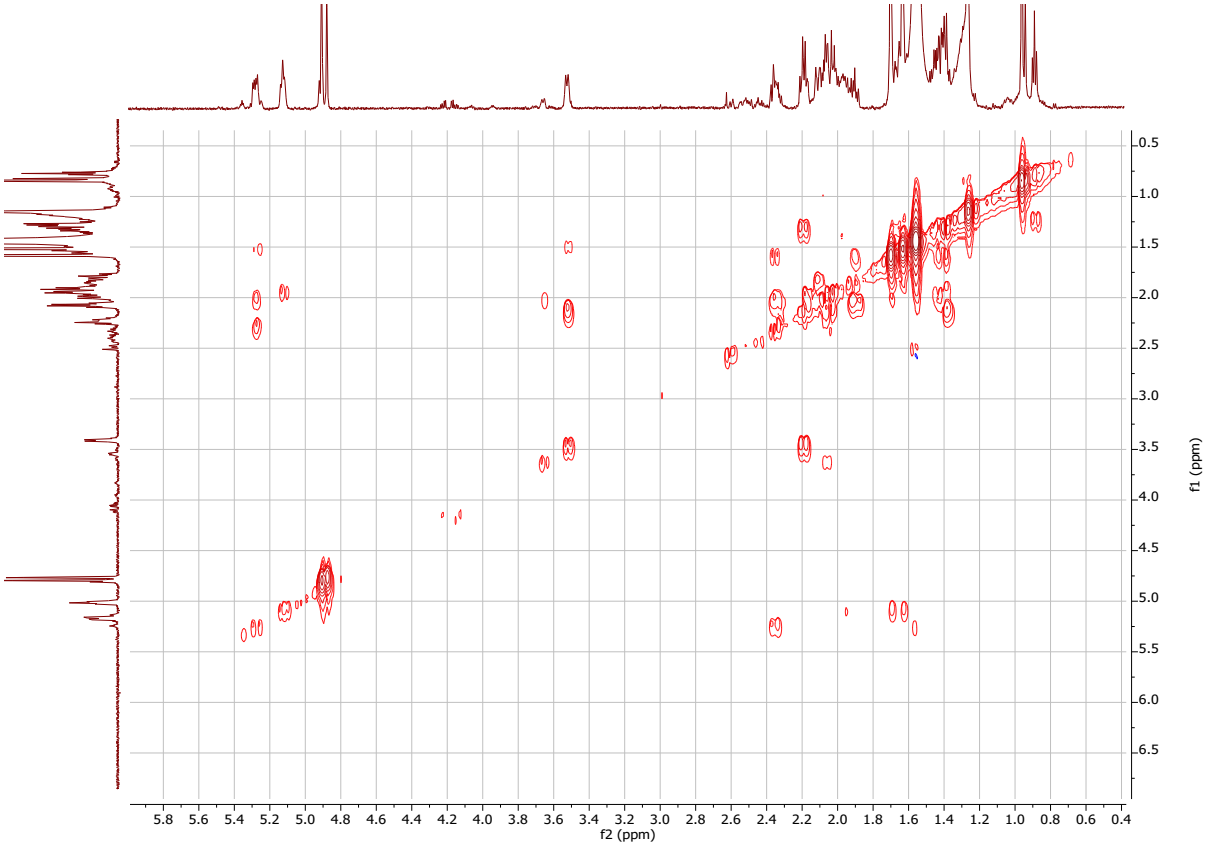


Figure S7: gCOSY spectrum of miolenol (**1**) in $CDCl_3$ (600 MHz).

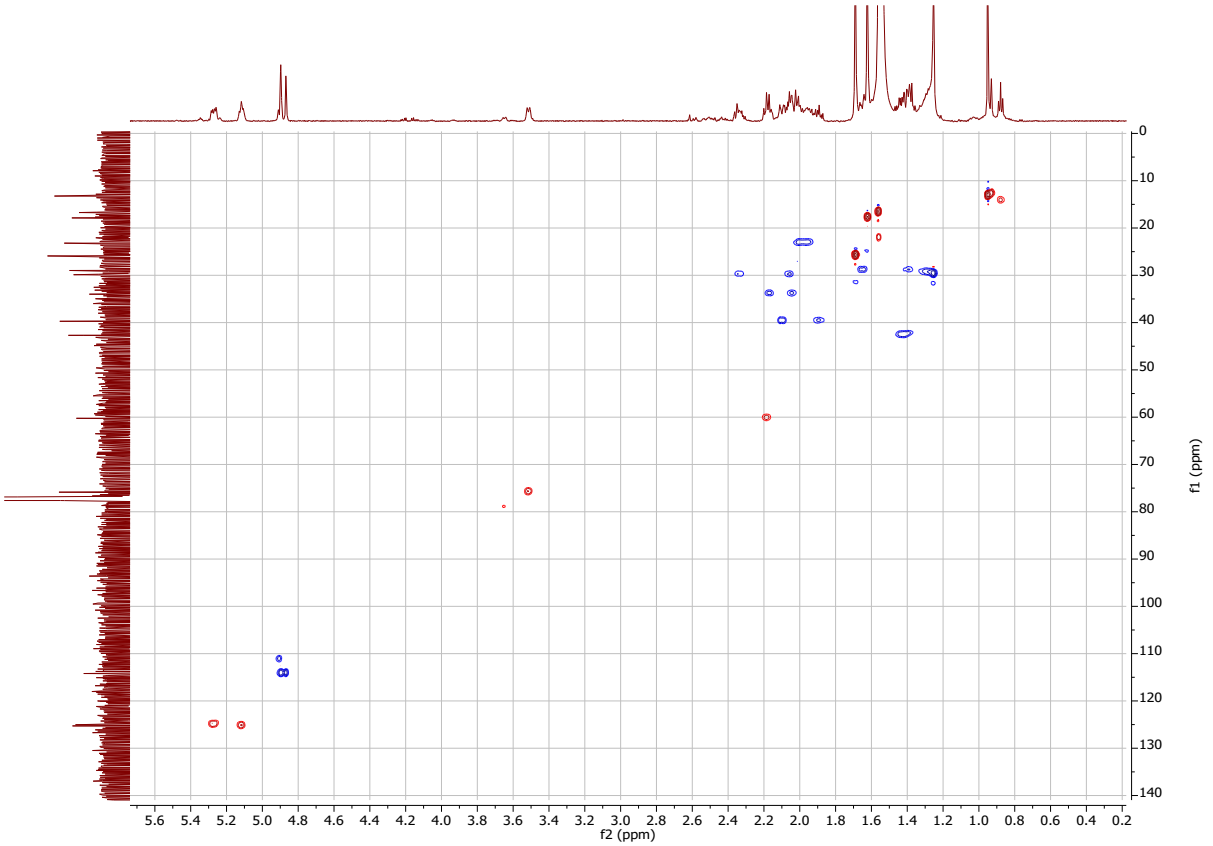


Figure S8: gHSQCAD spectrum of miolenol (**1**) in $CDCl_3$ (600 MHz).

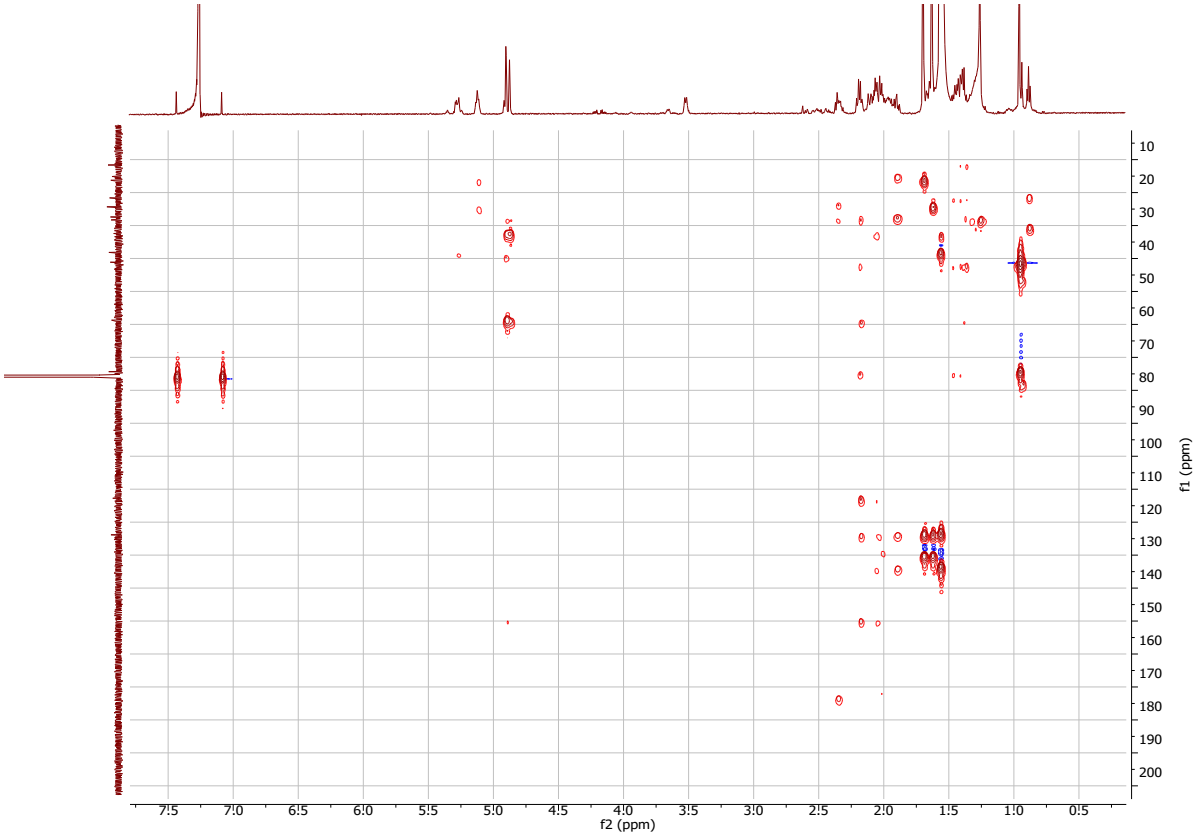


Figure S9: gHMBCAD spectrum of miolenol (**1**) in $CDCl_3$ (600 MHz).

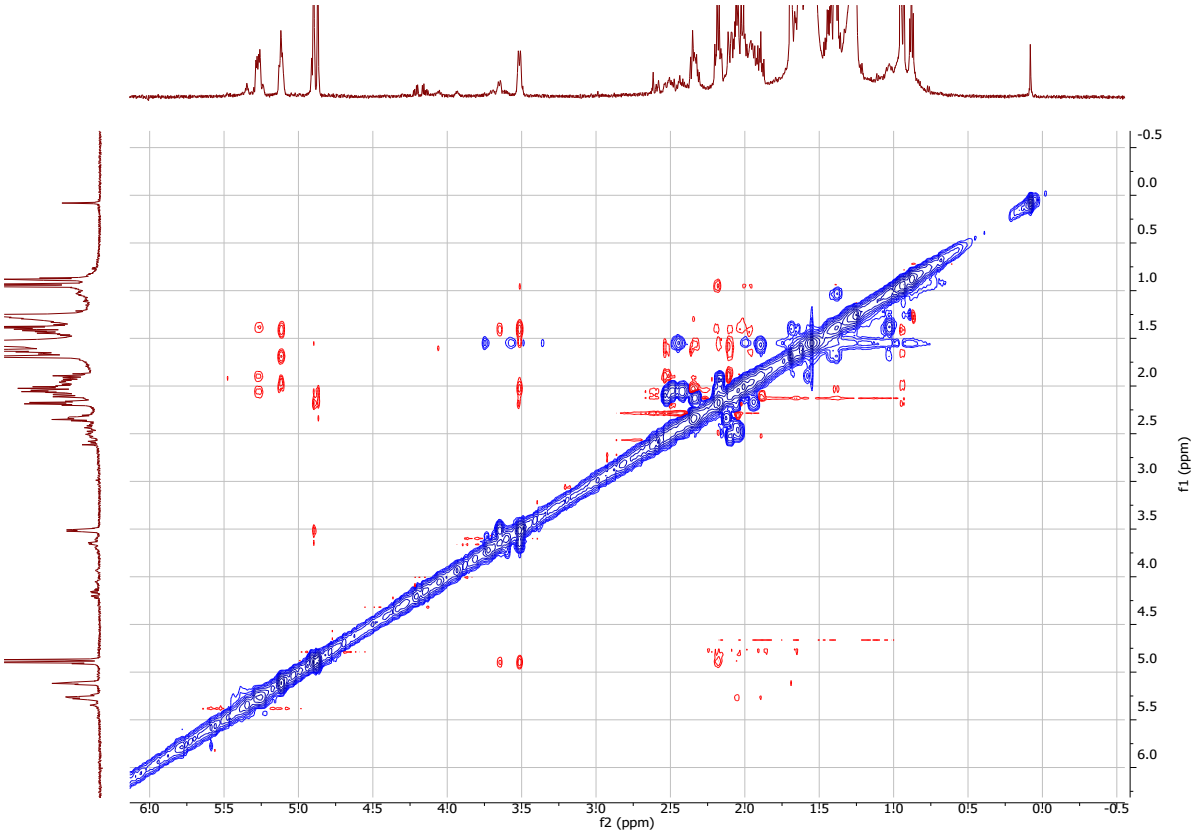


Figure S10: ROESY spectrum of miolenol (**1**) in $CDCl_3$ (600 MHz).

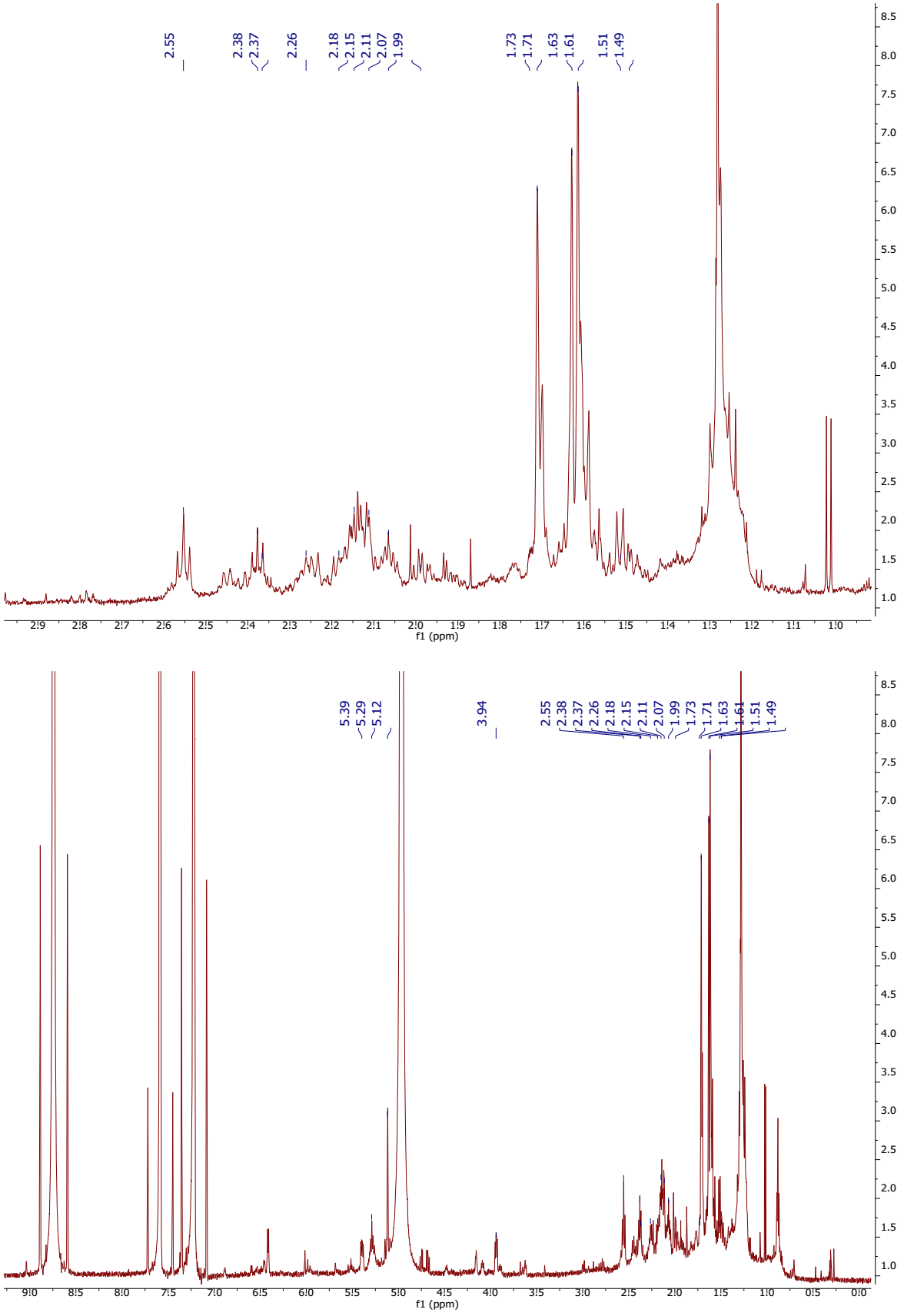


Figure S11: ^{1}H NMR spectrum of miolenol (**1**) in C_5D_5N (600 MHz).

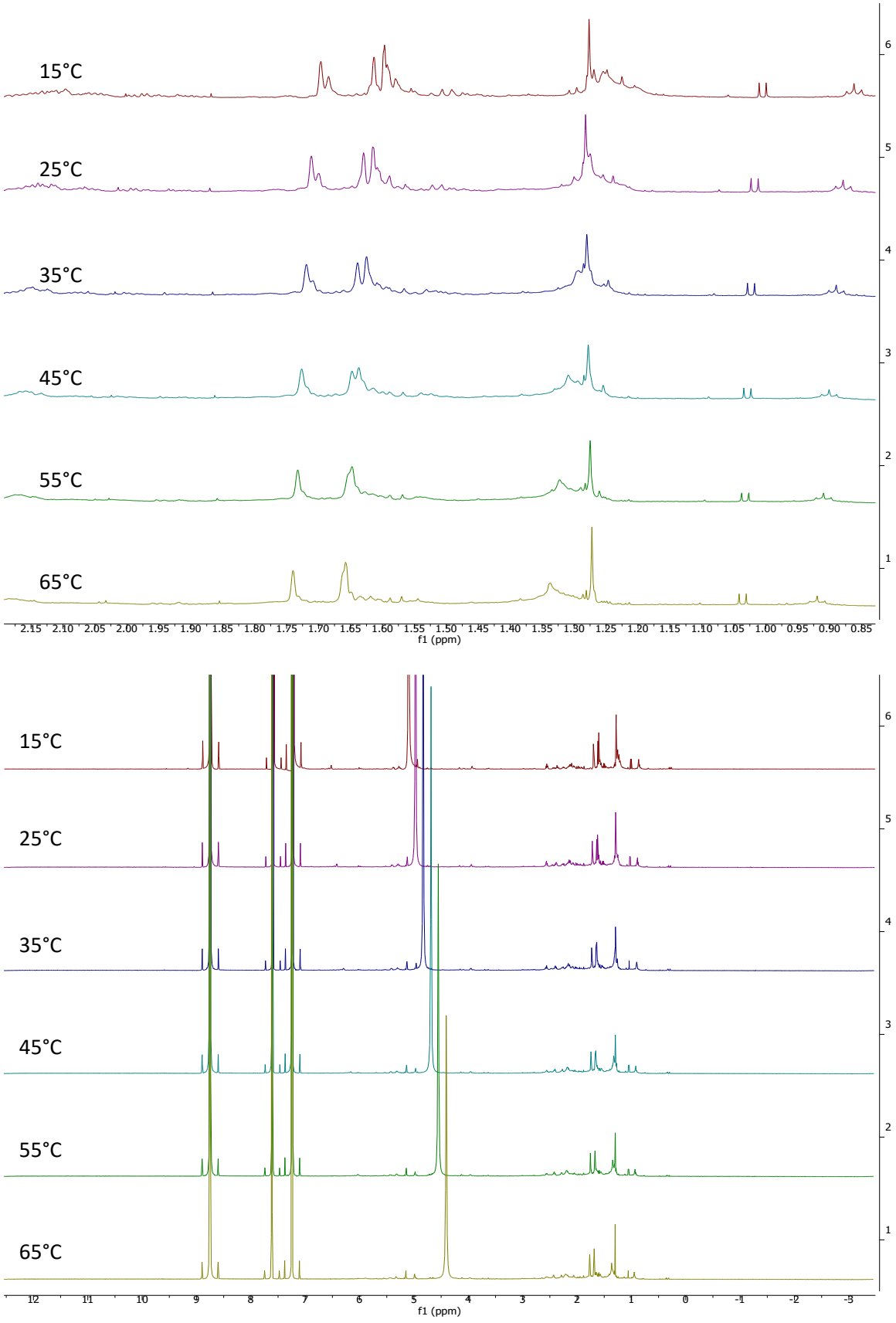


Figure S12: Variable temperature ^{1}H NMR spectrum of miolenol (**1**) in C_5D_5N (600 MHz).

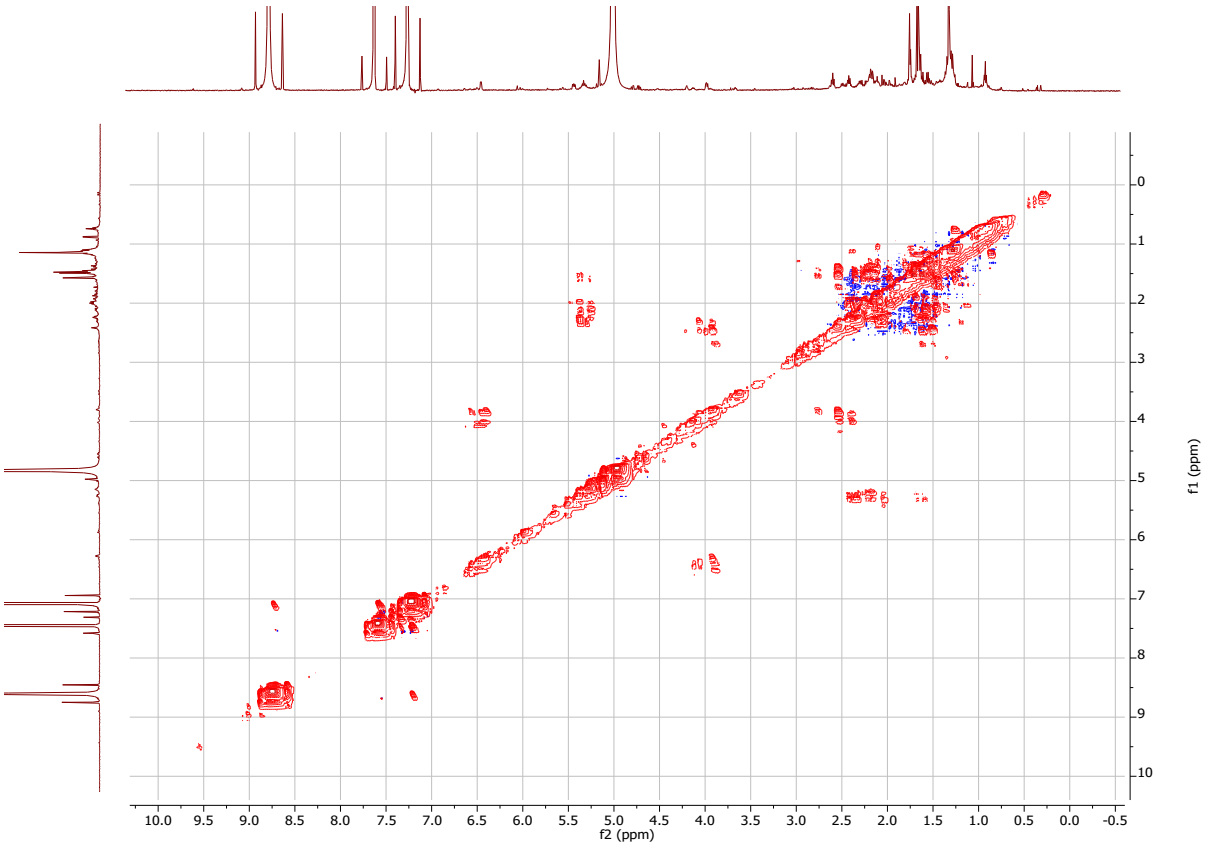


Figure S13: gCOSY spectrum of miolenol (**1**) in C_5D_5N (600 MHz).

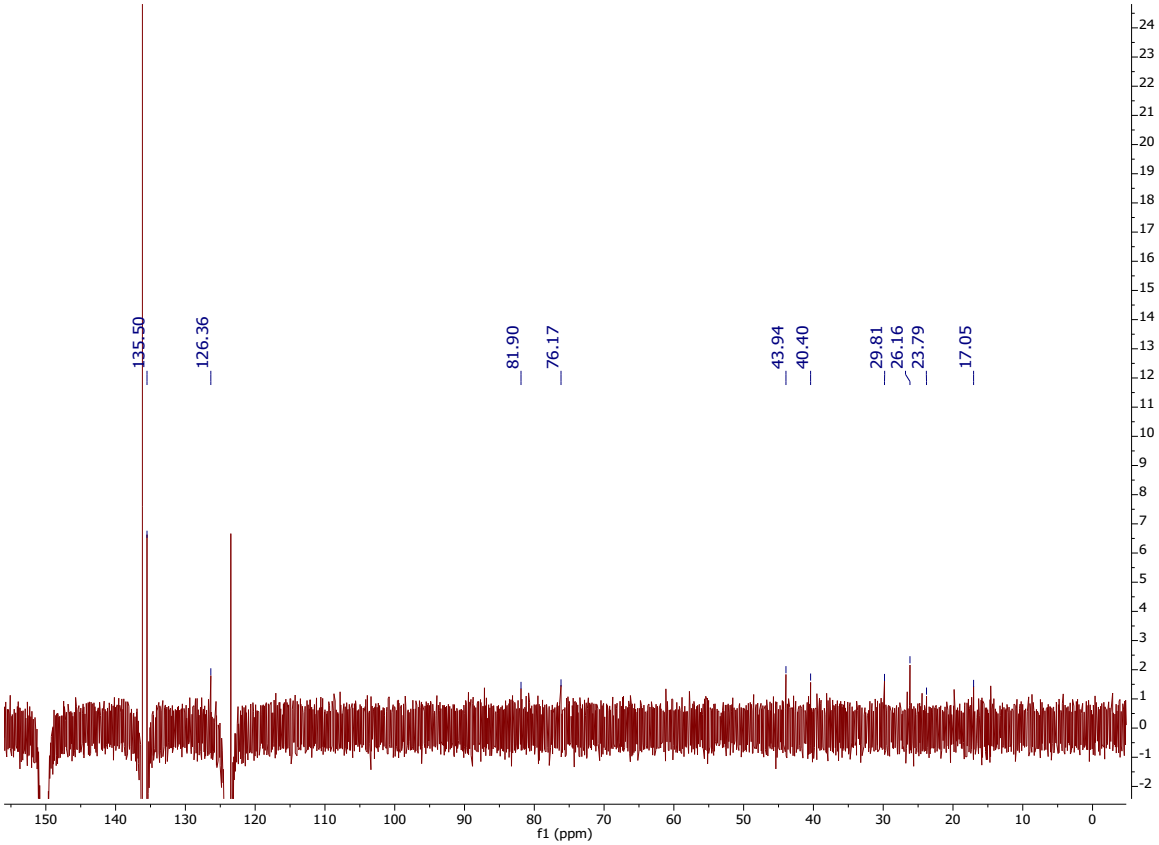


Figure S14: ^{13}C NMR spectrum of miolenol (**1**) in C_5D_5N (150 MHz).

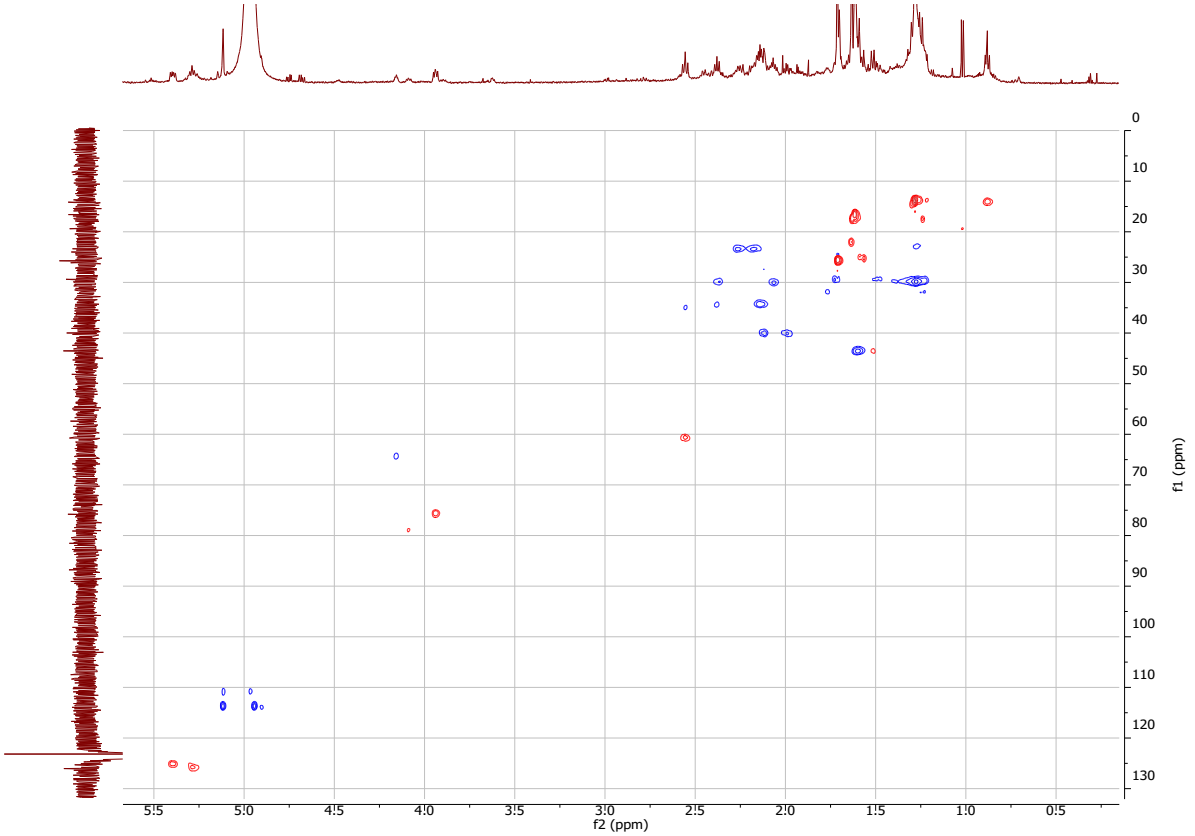


Figure S15:gHSQCAD spectrum of miolenol (**1**) in C_5D_5N (600 MHz).

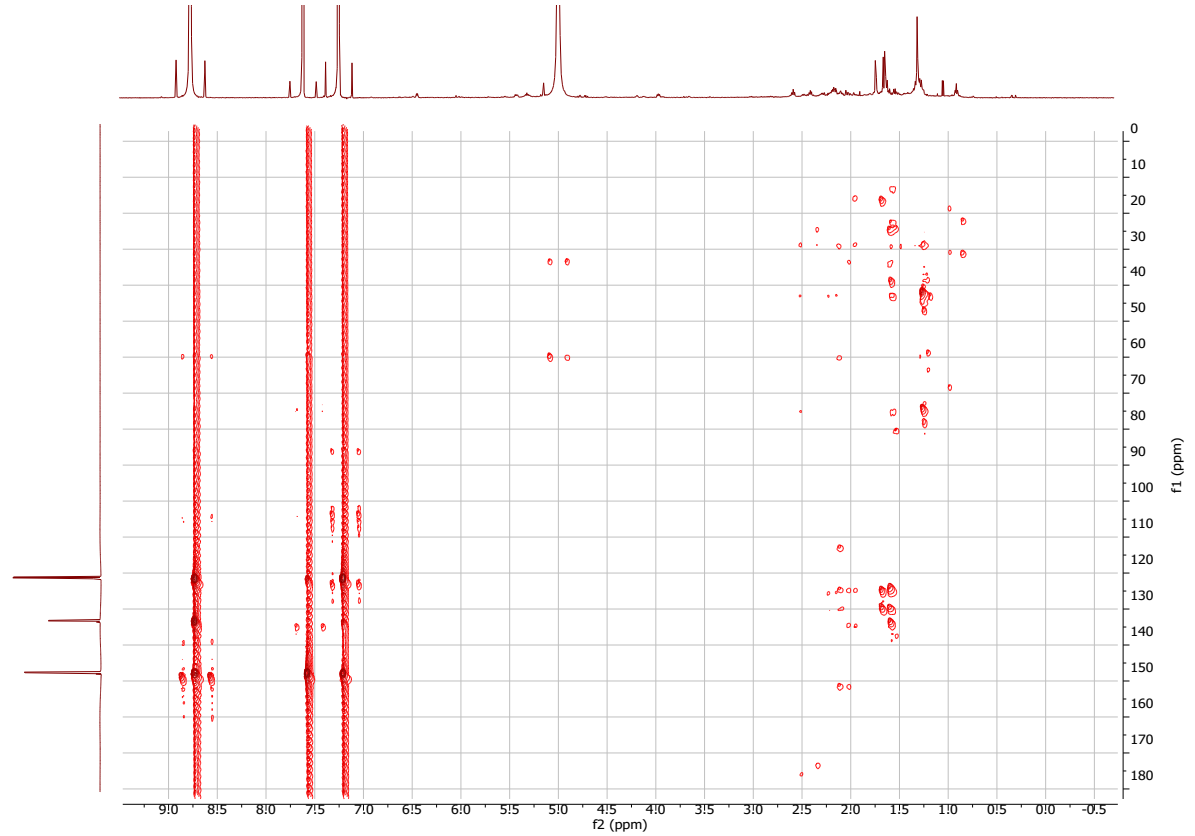


Figure S16: gHMBCAD spectrum of miolenol (**1**) in C_5D_5N (600 MHz).

Miolenol **(1)**		
	Pyridine- CD_5N	
Pos	δ_C, Type	δ_H (Mult., *J* in Hz)
1	43.8, CH	1.51 (m)
2	29.5, CH_2	1.73, 1.49 (m)
3	40.1, CH_2	2.11, 1.99 (m)
4	135.3, C	
5	125.3, CH	5.39 (dd, 11.0, 6.0)
6	30.1, CH_2	2.37, 2.07 (m)
7	34.4, CH_2	2.38, 2.15 (m)
8	152.5, C	
9	60.8, CH	2.55, (m)
10	75.9, CH	3.94 (d, 7.00)
11	44.1, C	
12	43.6, CH_2	1.60 (m)
13	23.5, CH_2	2.26, 2.18 (m)
14	126.0, CH	5.29 (t, 7.69)
15	131.1, C	
16	22.3, CH_3	1.63 (s)
17	25.9, CH_3	1.71 (s)
18	16.8, CH_3	0.95 (s)
19	113.7, CH_2	5.12, 4.94 (s)
20	16.8, CH_3	1.61 (s)

Table S1: ^{1}H NMR (600MHz) and ^{13}C NMR (150 MHz) spectroscopic data for miolenol (**1**) in deuterated pyridine,

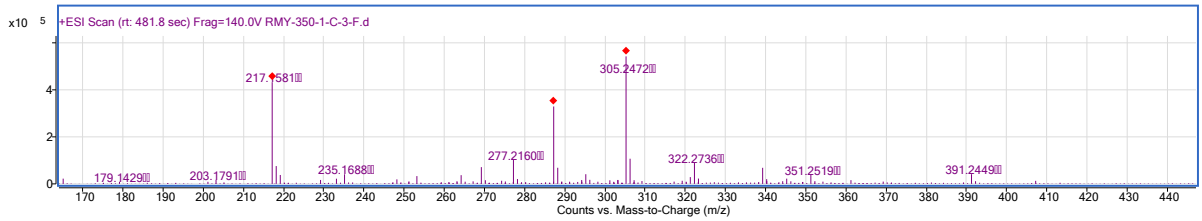


Figure S17:HRESIMS spectrum of epoxymiolenol (**2**).

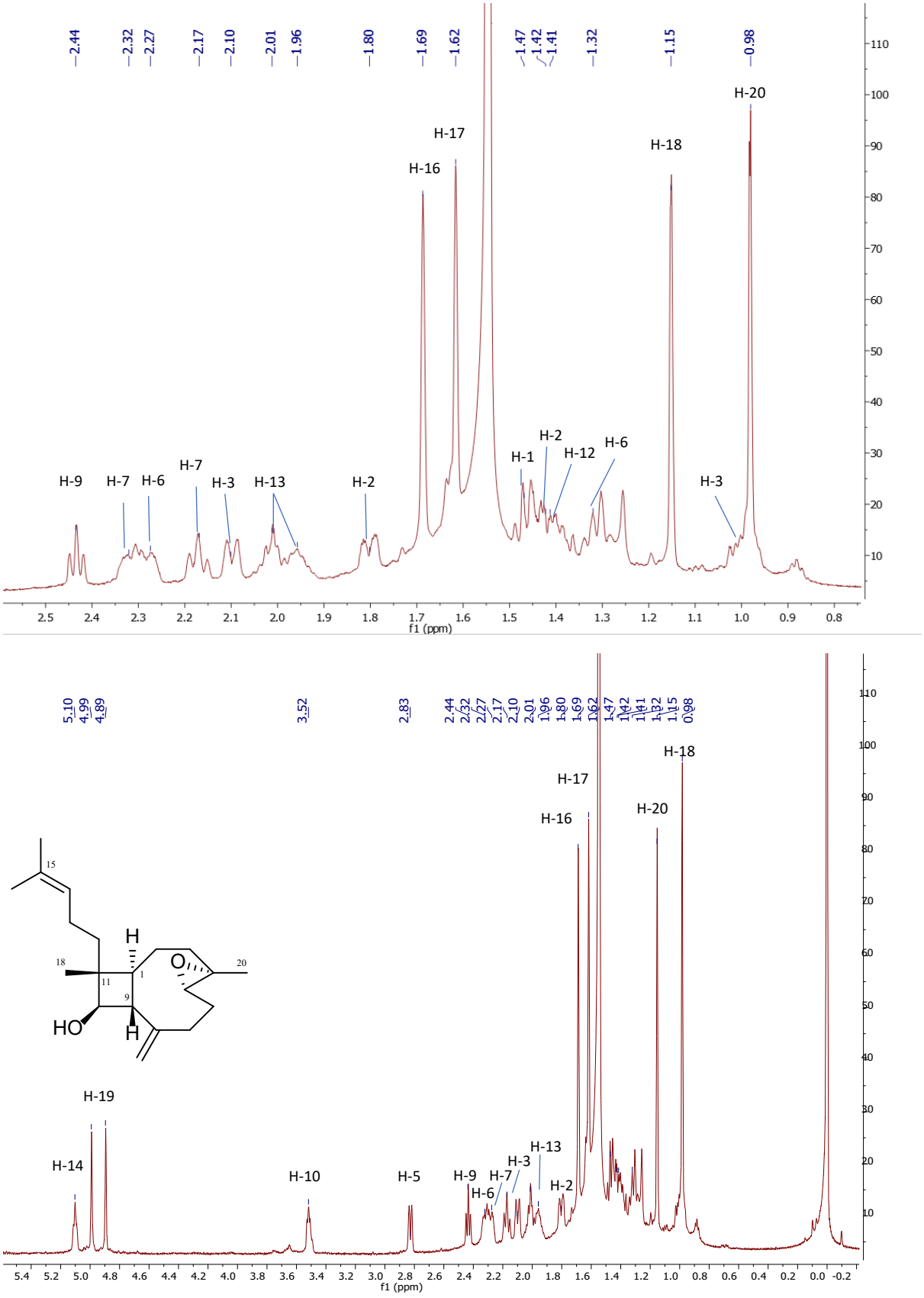


Figure S18: ^{1}H NMR spectrum of epoxymiolenol (**2**) in $CDCl_3$ (600 MHz).

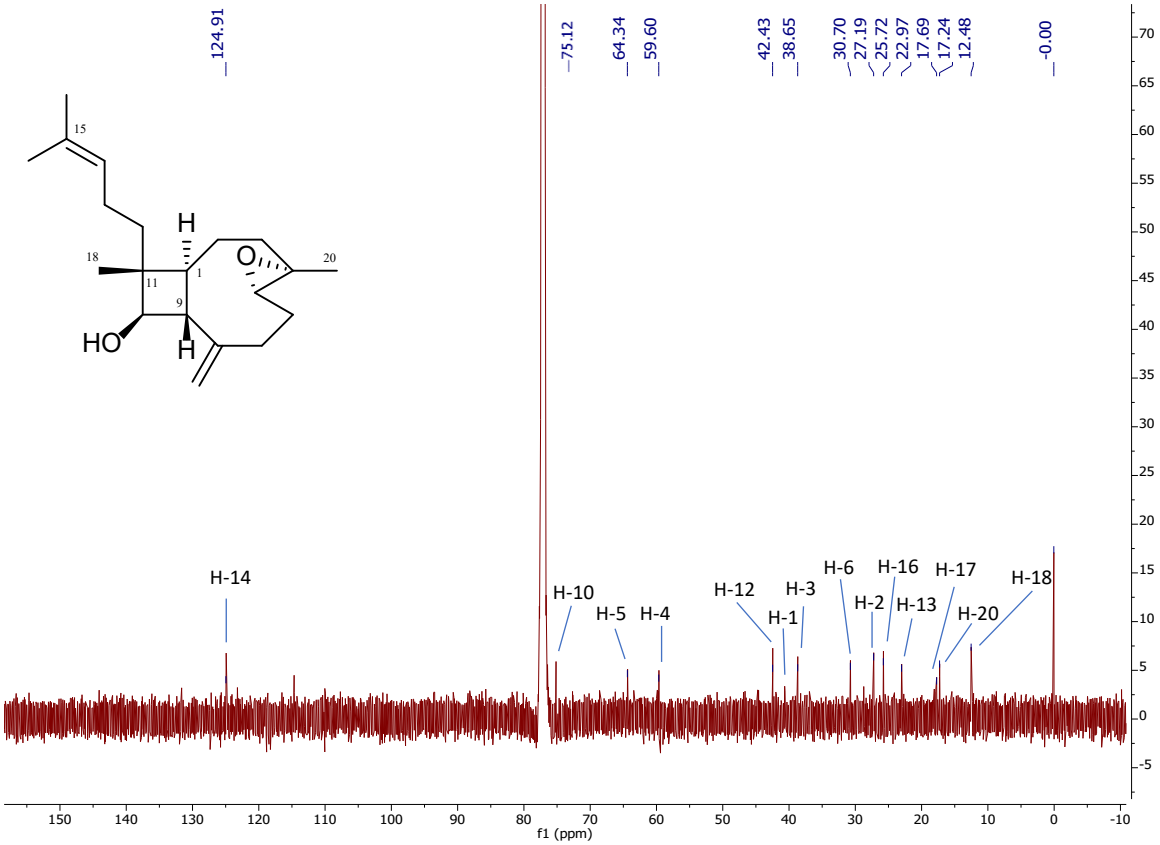


Figure S19: ^{13}C NMR spectrum of epoxymiolenol (**2**) in $CDCl_3$ (150 MHz).

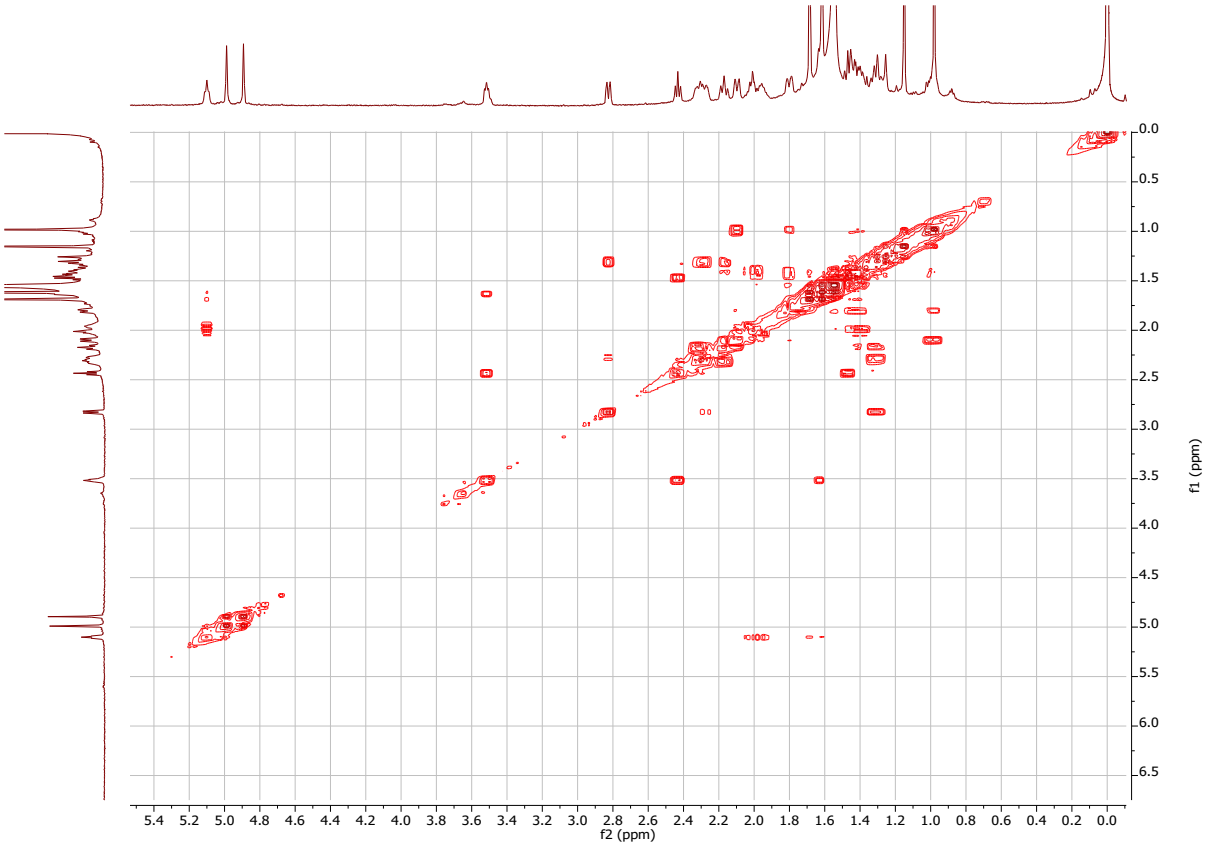


Figure S20: gCOSY NMR spectrum of epoxymiolenol (**2**) in $CDCl_3$ (600 MHz).

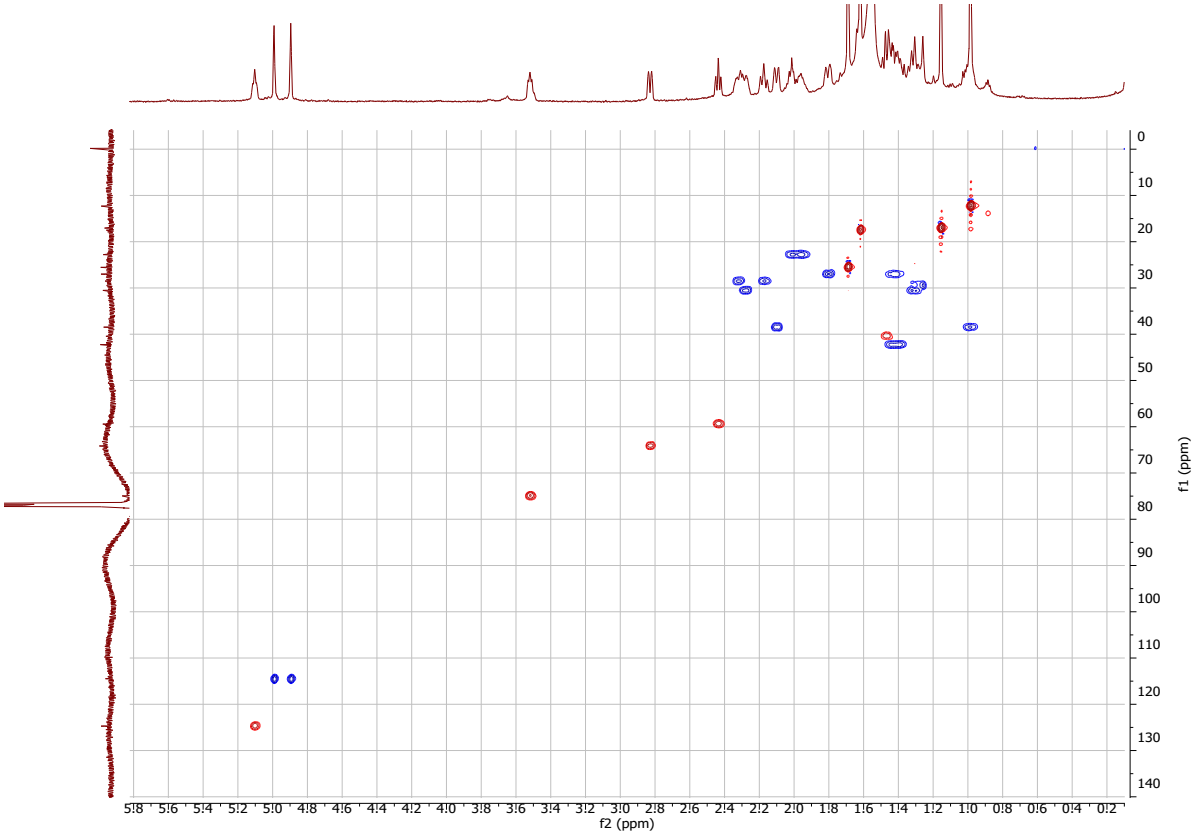


Figure S21: gHSQCAD NMR spectrum of epoxymiolenol (**2**) in $CDCl_3$ (600 MHz).

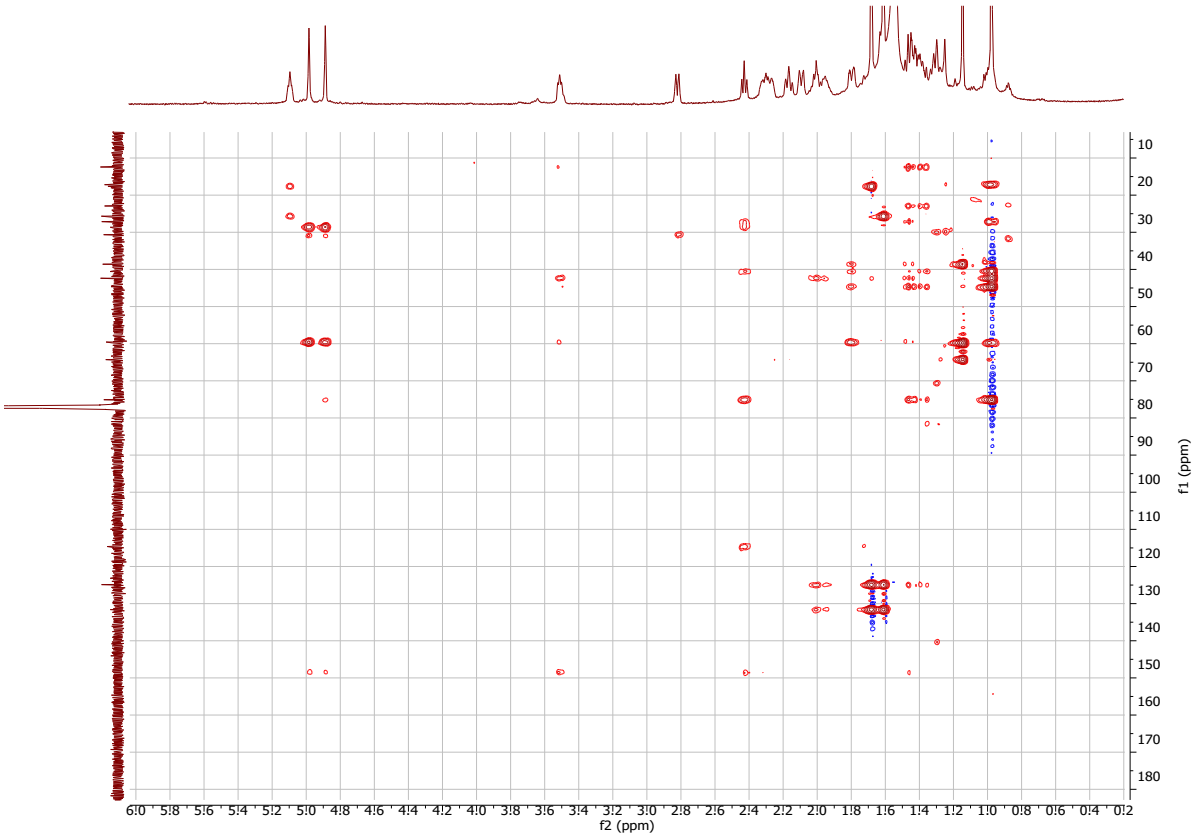


Figure S22: gHMBCAD NMR spectrum of epoxymiolenol (**2**) in $CDCl_3$ (600 MHz).

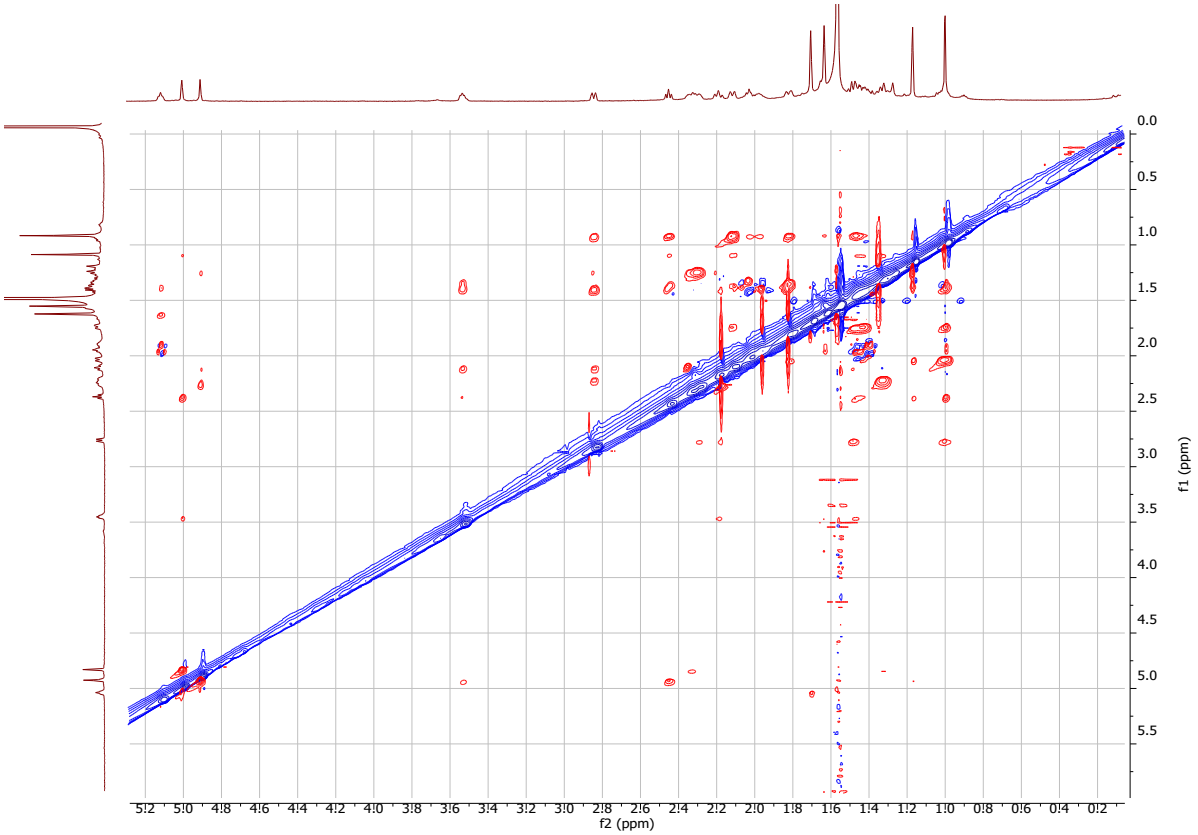


Figure S23: ROESY NMR spectrum of epoxymiolenol (**2**) in $CDCl_3$ (600 MHz).

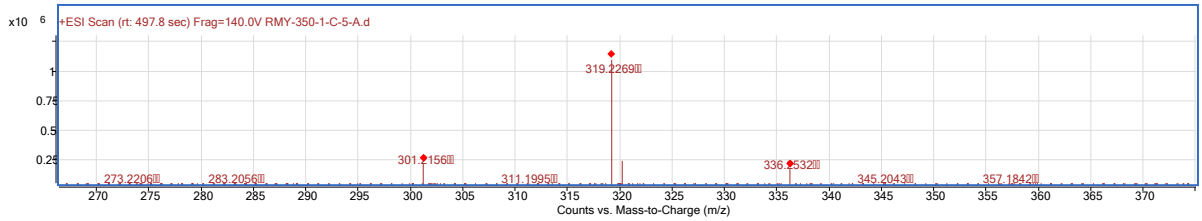


Figure S24: HRESIMS spectrum of epoxycoraxeniolide A (**3**).

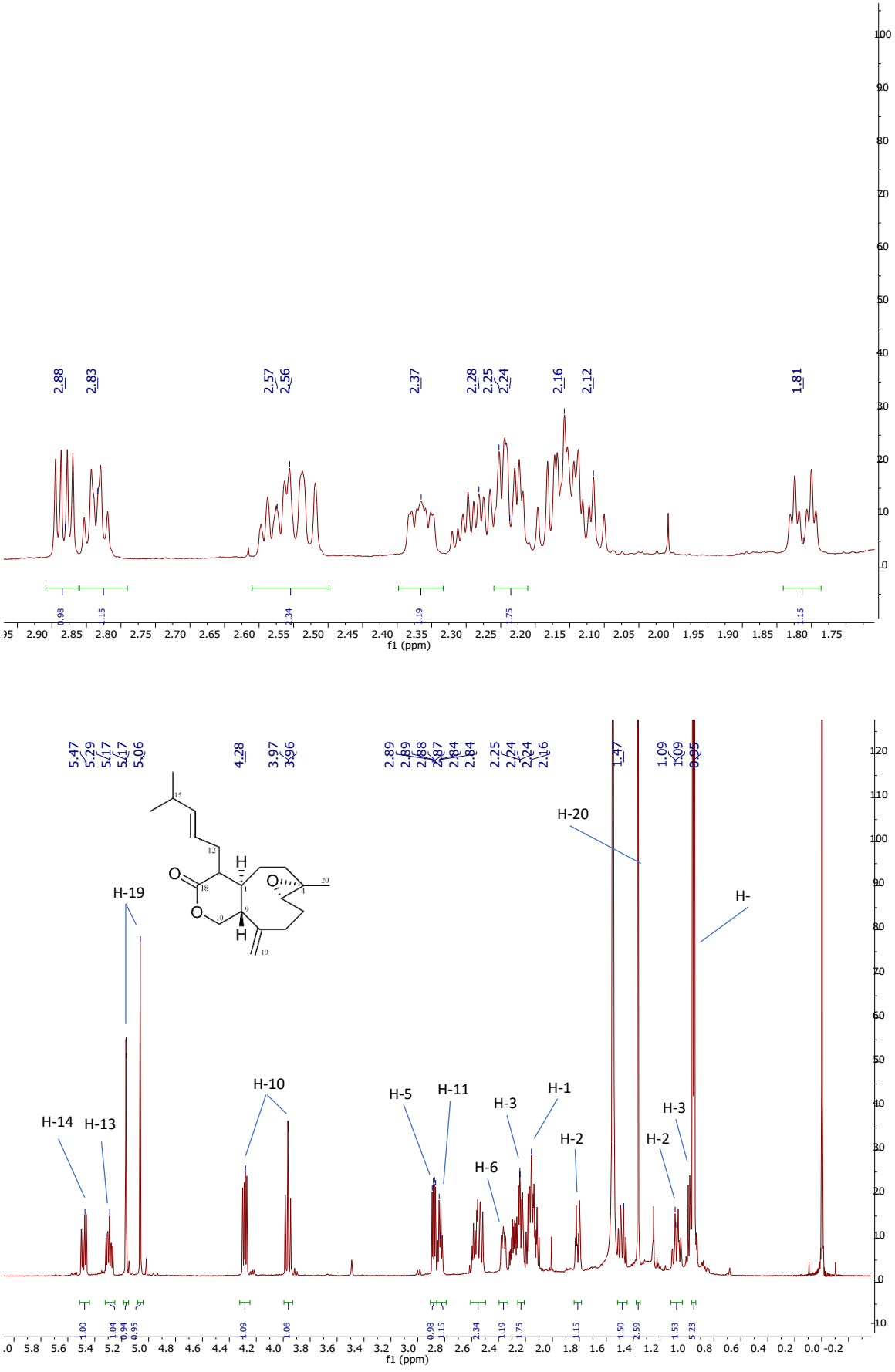


Figure S25: ^{1}H NMR spectrum of epoxycoraxeniolide A (**3**) in $CDCl_3$ (600 MHz).

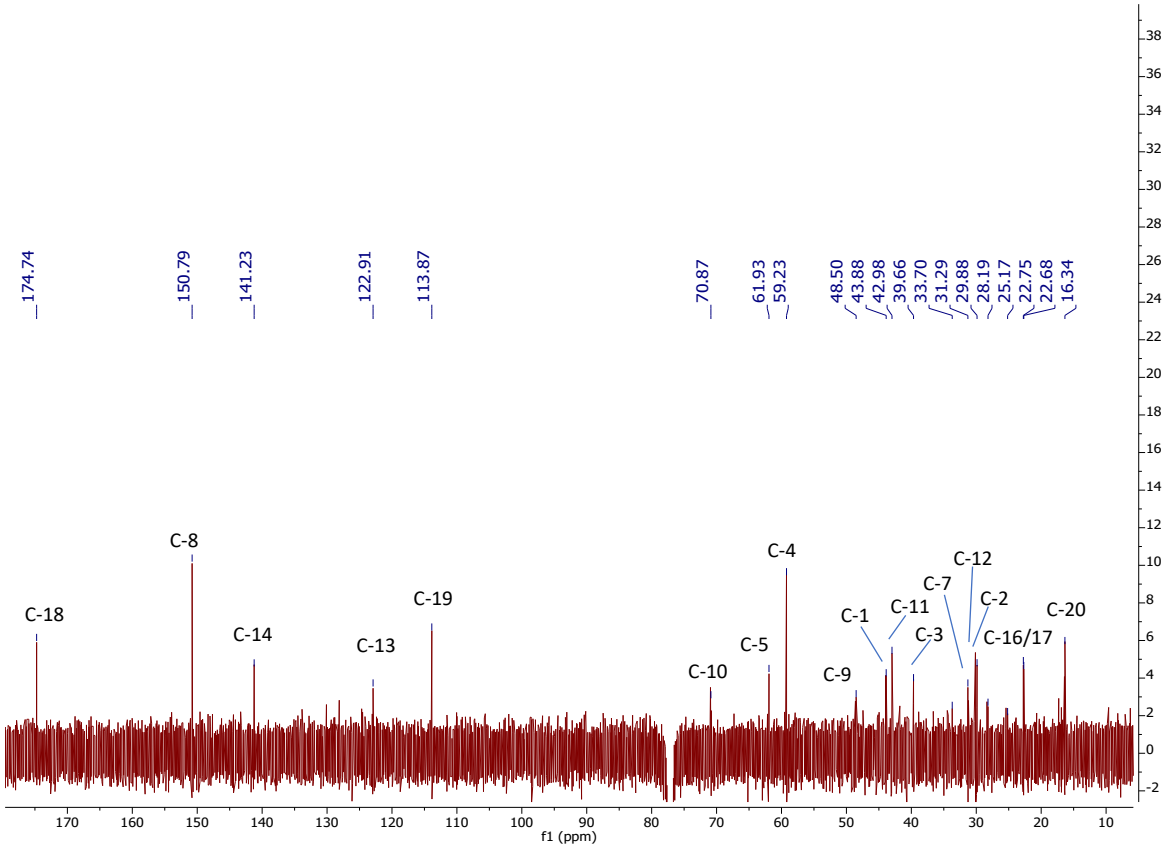


Figure S26: ^{13}C- NMR spectrum of epoxycoraxeniolide A (**3**) in $CDCl_3$ (150 MHz).

Figure S27: NOESY NMR spectrum of epoxycoraxeniolide A (**3**) in $CDCl_3$ (600 MHz).

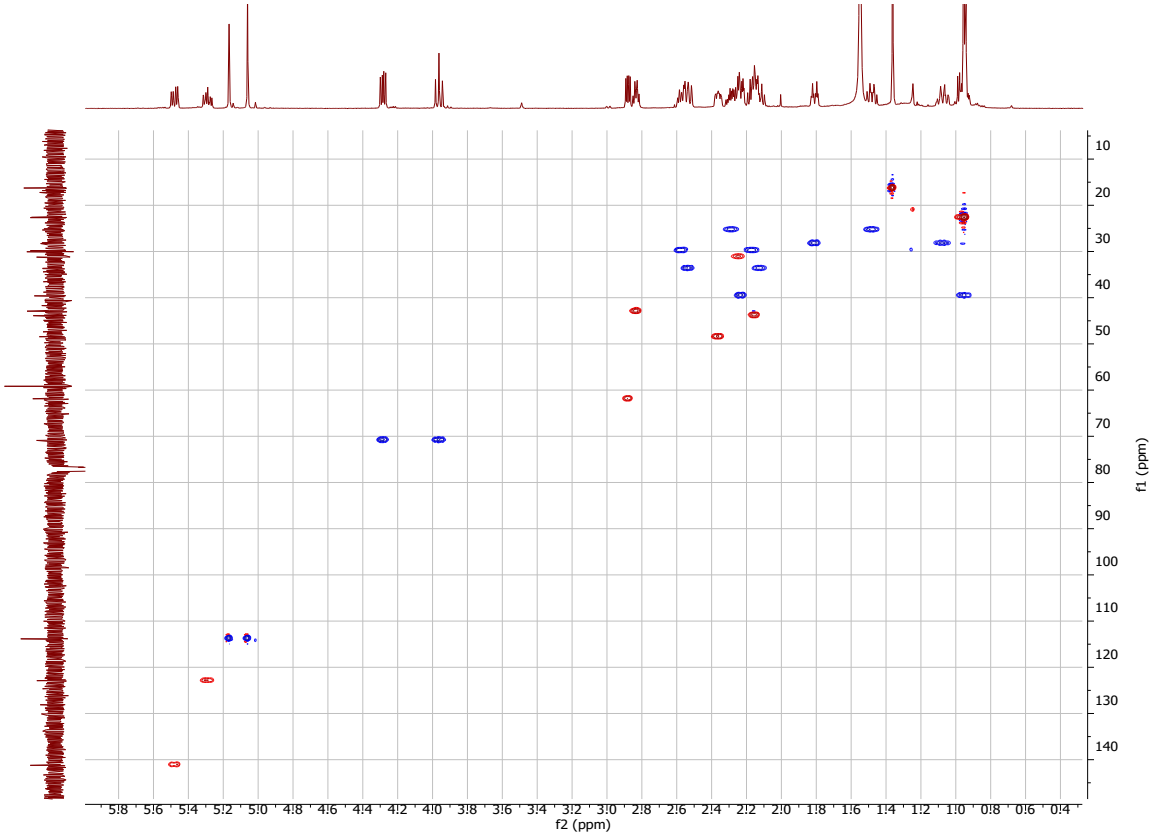


Figure S28: gHSQCAD NMR spectrum of epoxycoraxeniolide A (**3**) in $CDCl_3$ (600 MHz).

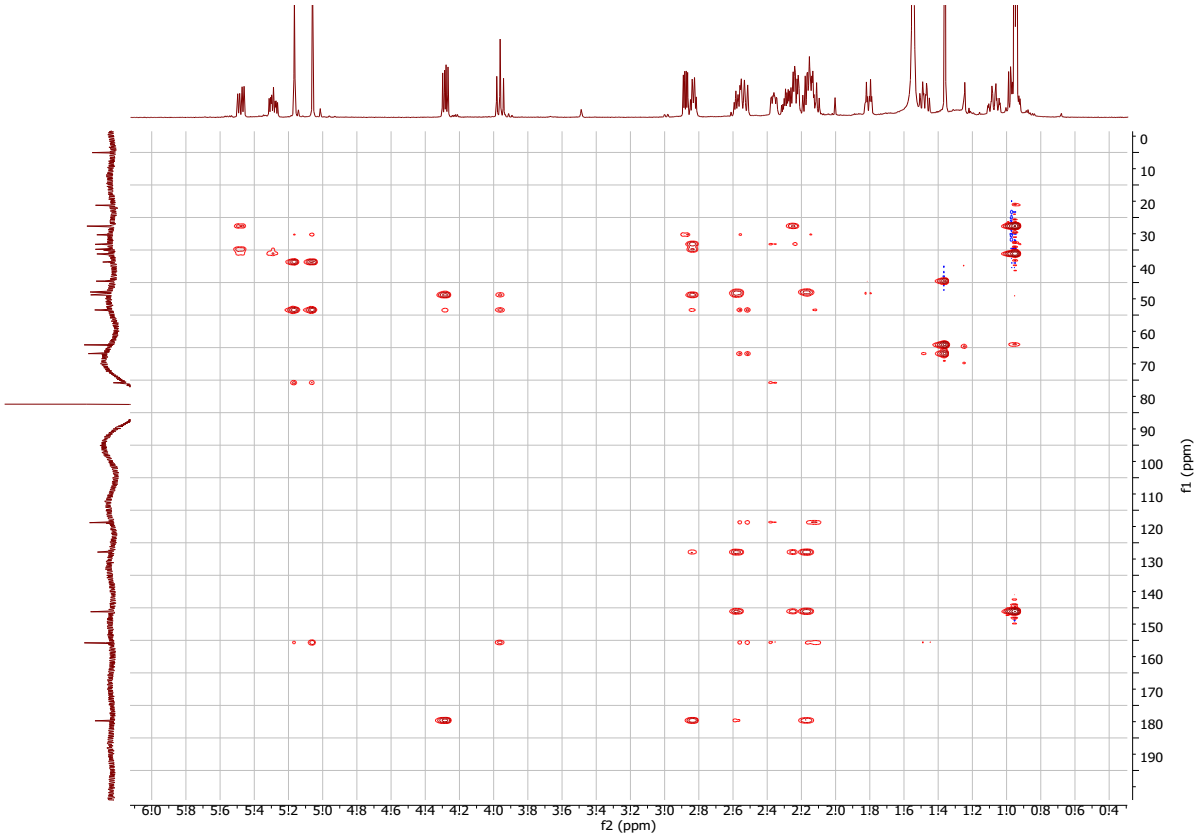


Figure S29: gHMBCAD NMR spectrum of epoxycoraxeniolide A (**3**) in $CDCl_3$ (600 MHz).

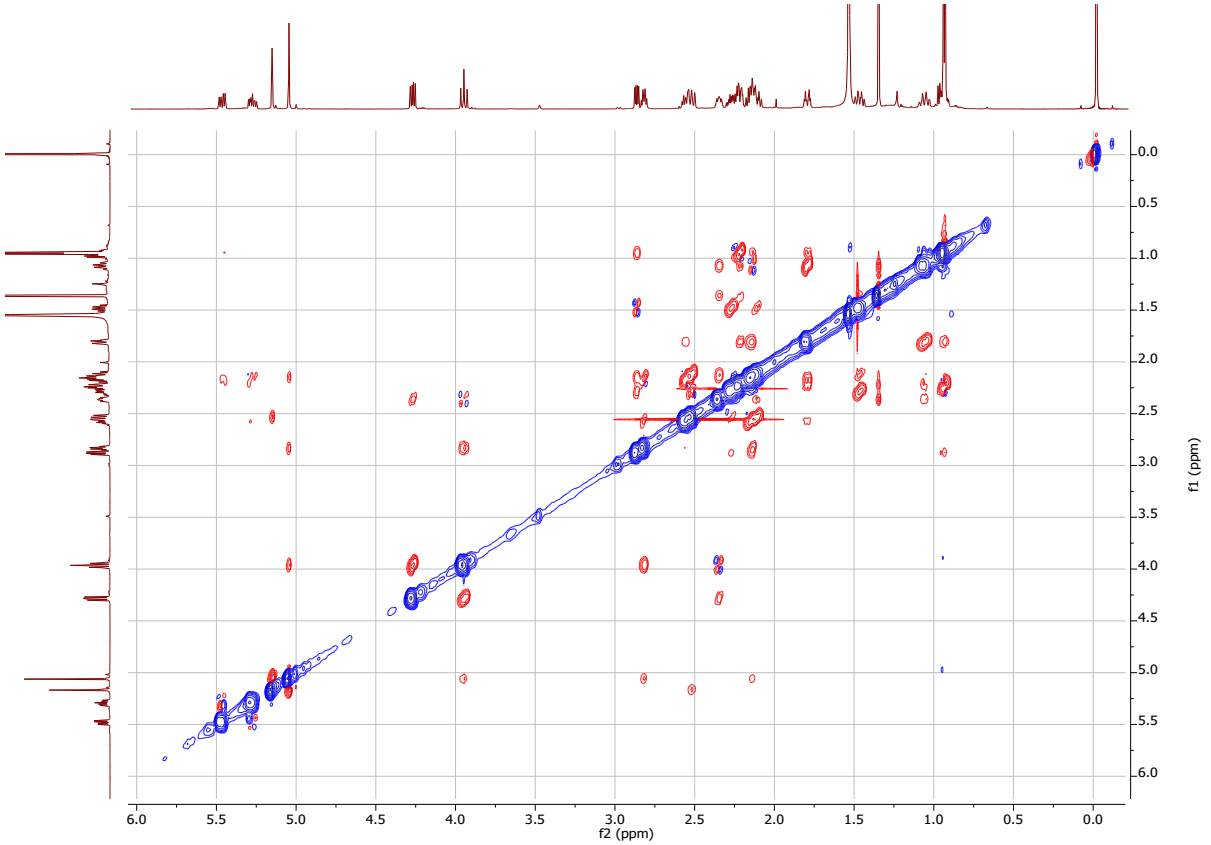


Figure S30: ROESY NMR spectrum of epoxycoraxeniolide A (**3**) in $CDCl_3$ (600 MHz).

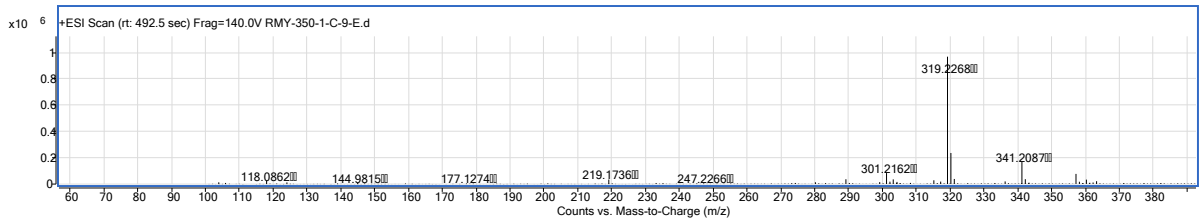


Figure S31: HRESIMS spectrum of acalycixeniolide F (**4**).

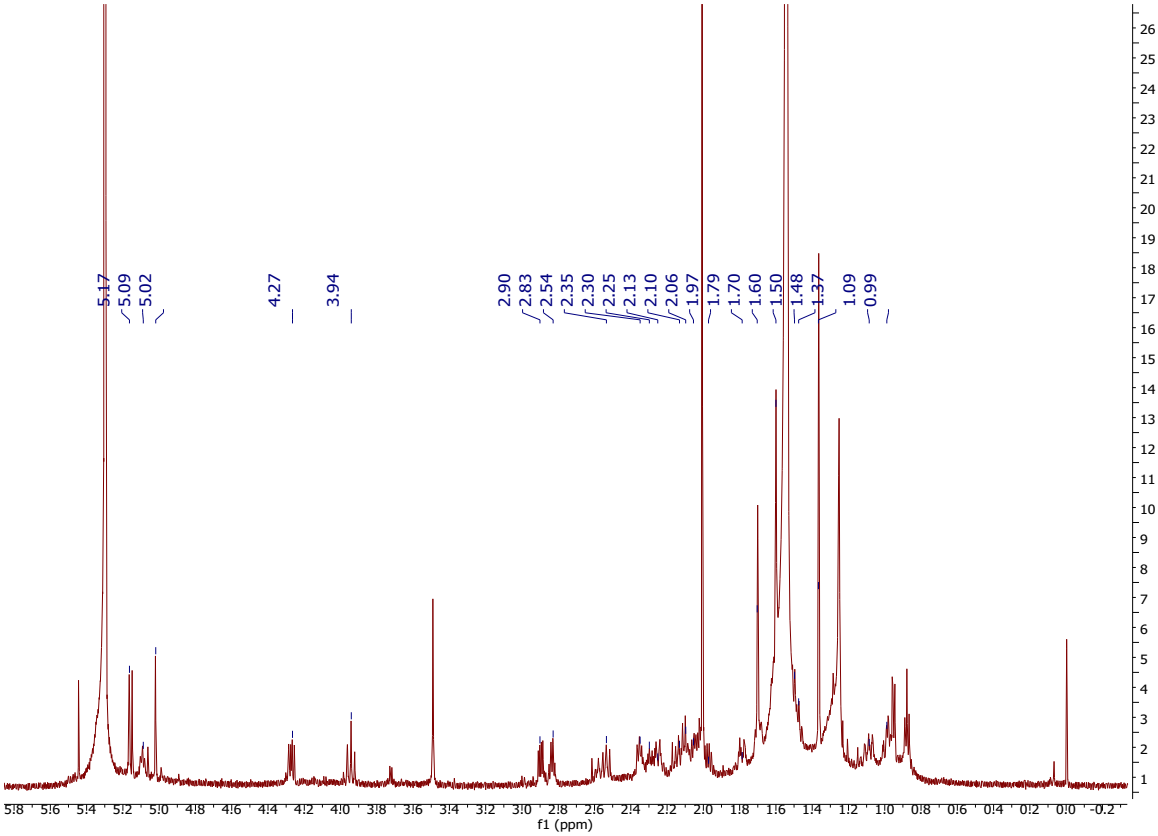


Figure S32: ^{1}H NMR spectrum of acalycixeniolide F (**4**) in $CDCl_3$ (600 MHz).

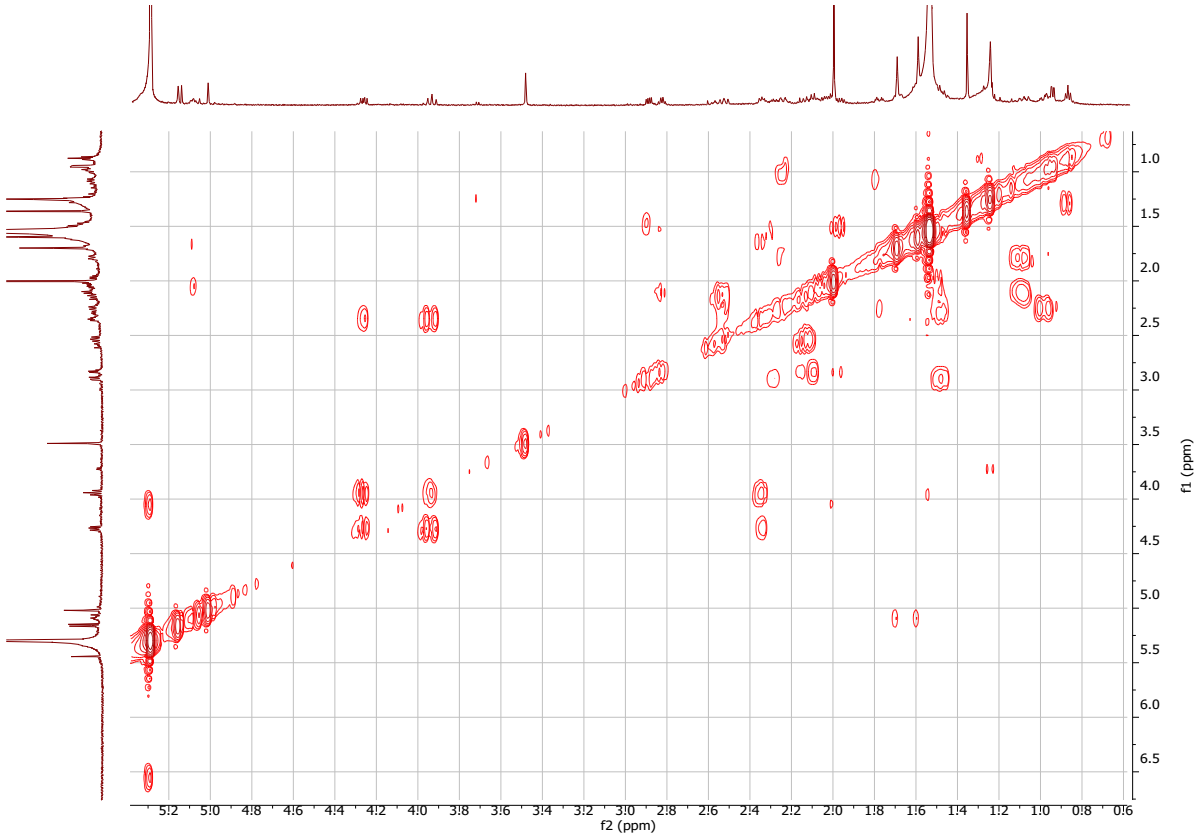


Figure S33: gCOSY NMR spectrum of acalycixeniolide F (**4**) in $CDCl_3$ (600 MHz).

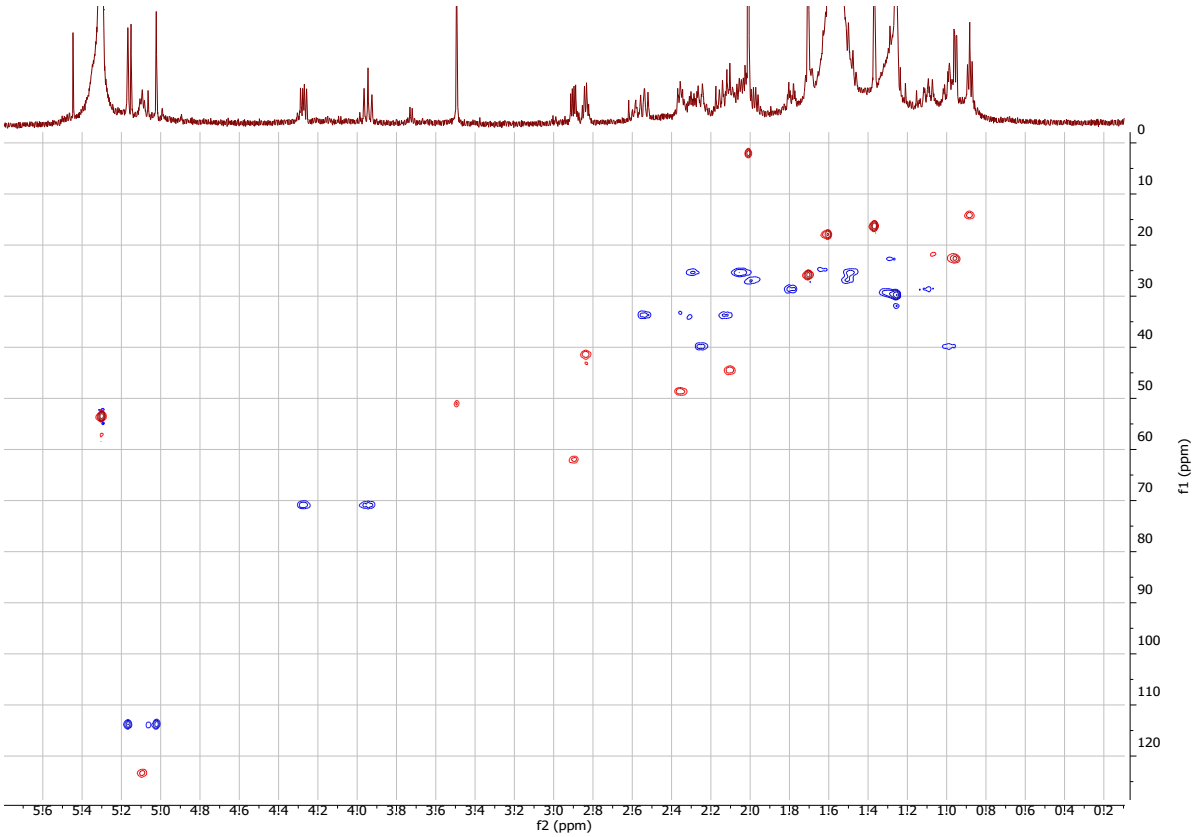


Figure S34: gHSQCAD NMR spectrum of acalycixeniolide F (**4**) in $CDCl_3$ (600 MHz).

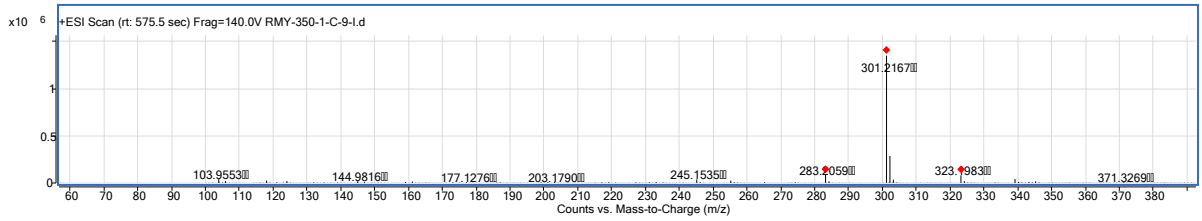


Figure S35: HRESIMS spectrum of coraxeniolide C (**5**).

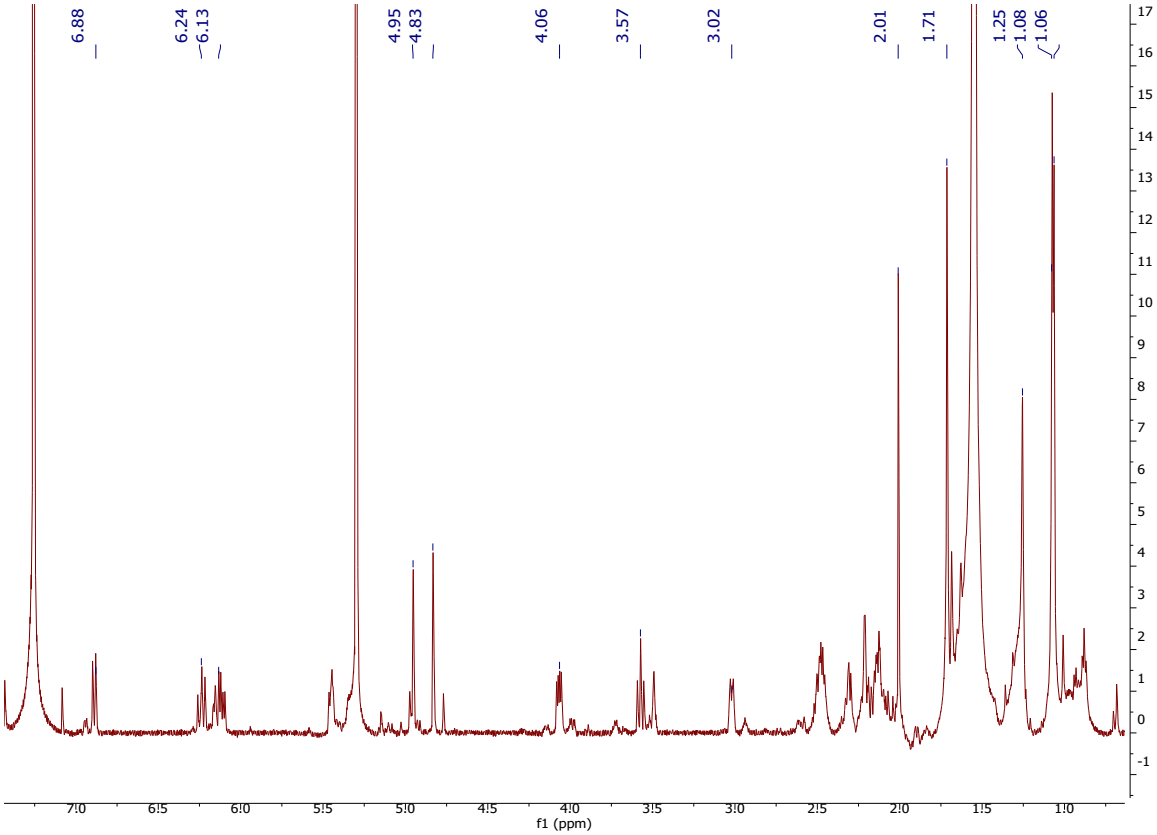


Figure S36: ^{1}H NMR spectrum coraxeniolide C (**5**) of in $CDCl_3$ (600 MHz).

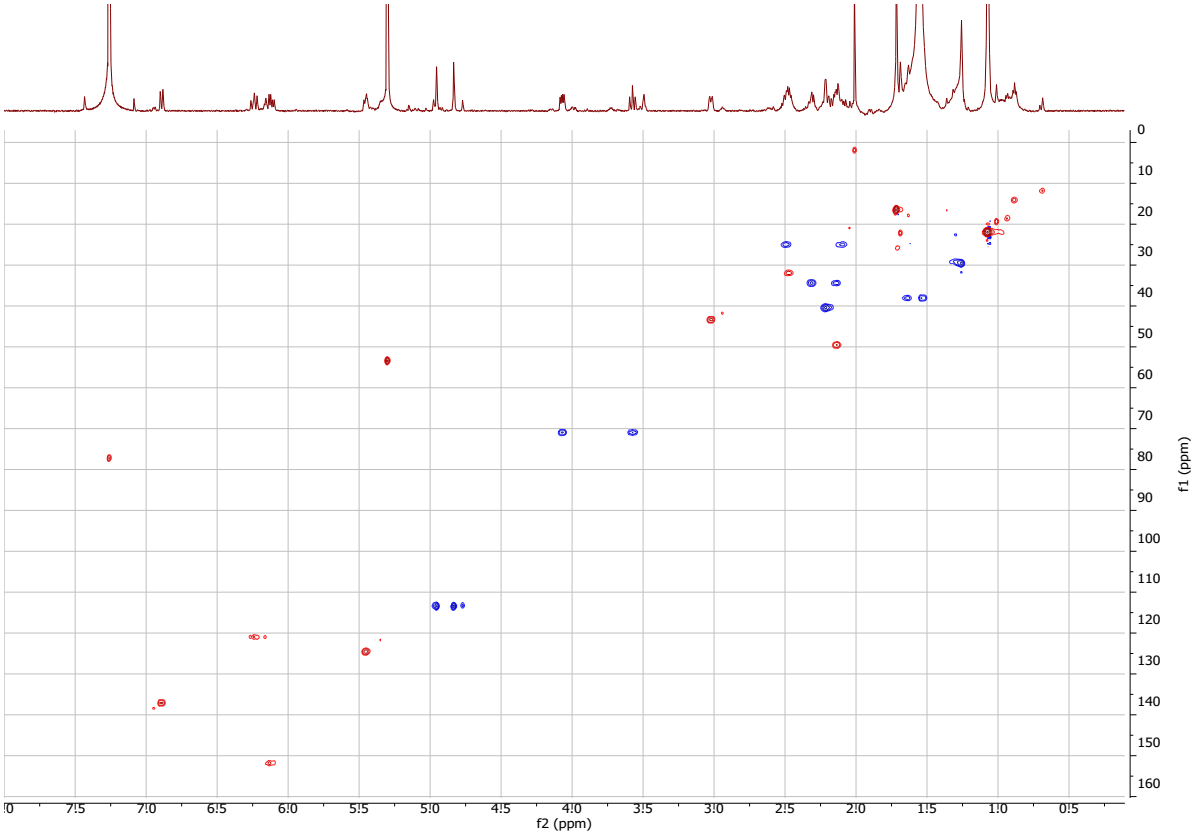


Figure S37: gHSQCAD NMR spectrum coraxeniolide C (**5**) of in $CDCl_3$ (600 MHz).

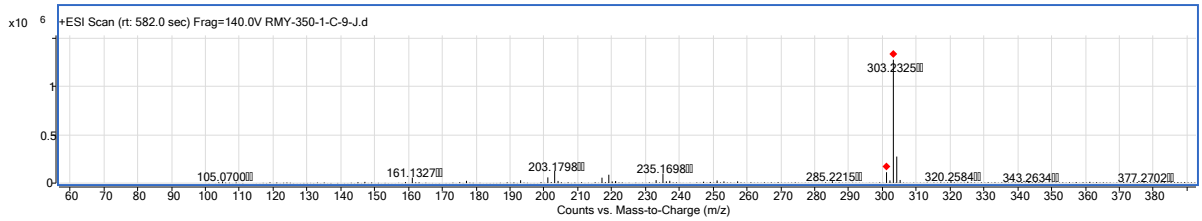


Figure S38: HRESIMS spectrum acalycigorgin E (**6**).

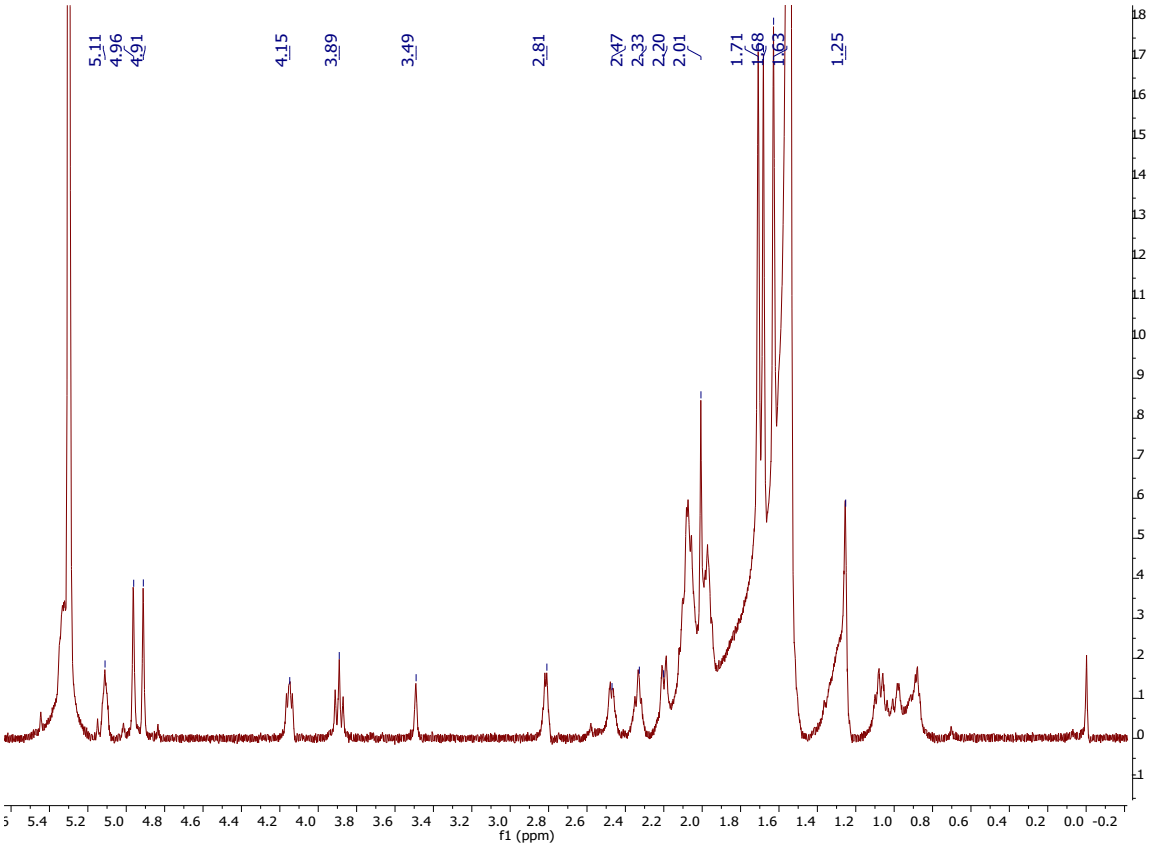


Figure S39:^{1}H NMR spectrum acalycigorgin E (**6**) of in $CDCl_3$ (600 MHz).

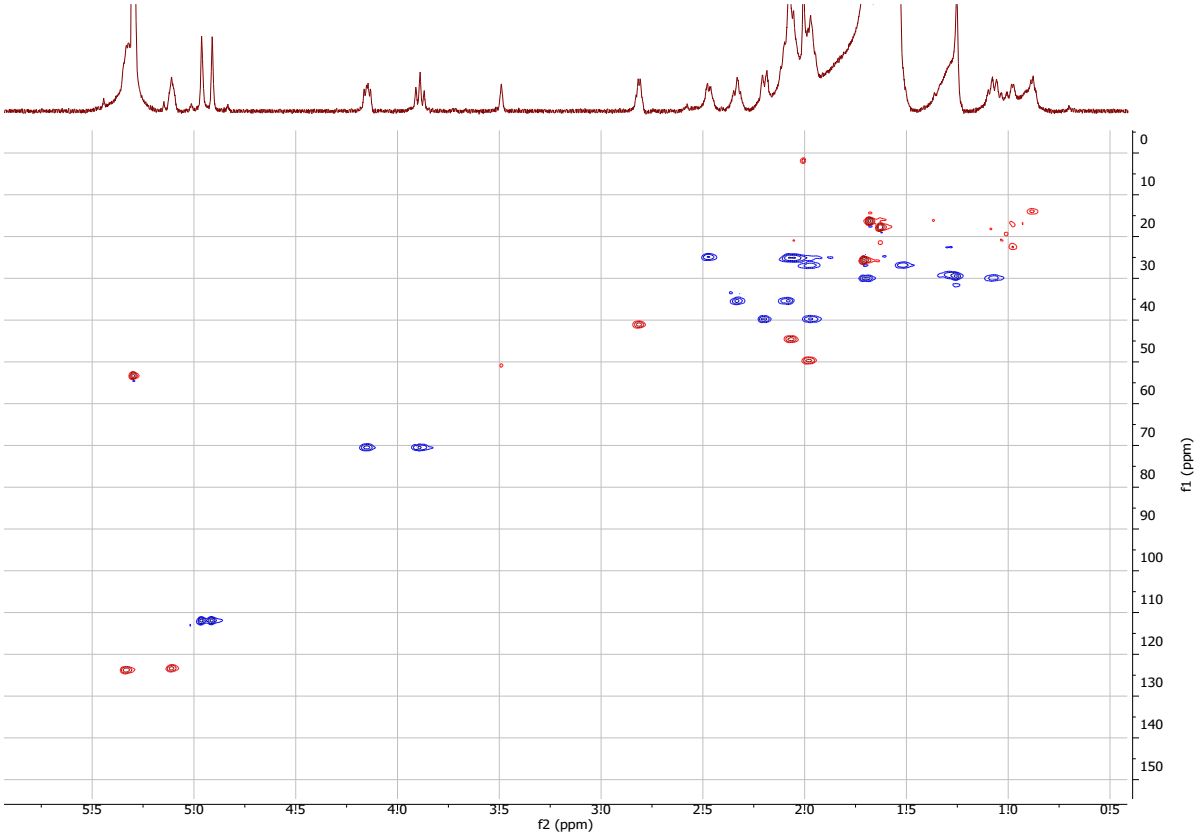


Figure S40: gHSQCAD NMR spectrum acalycigorgin E (**6**) of in $CDCl_3$ (600 MHz).

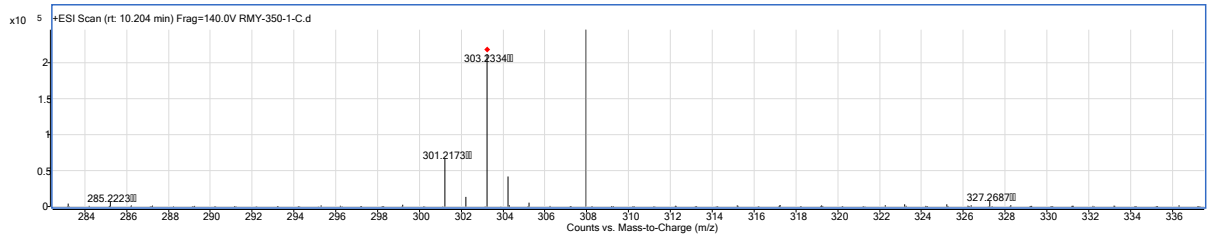


Figure S41: HRESIMS spectrum of coraxeniolide A (**7**)

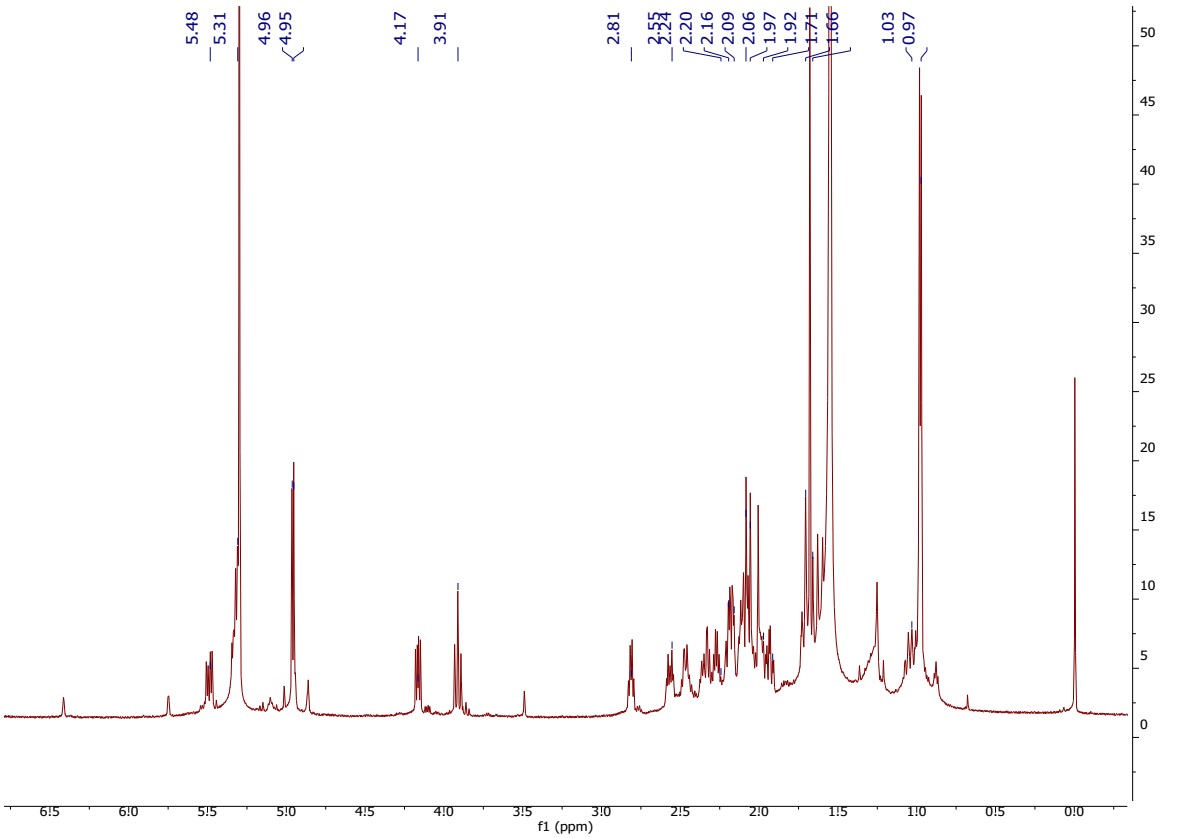


Figure S42: ^{1}H NMR spectrum coraxeniolide A (**7**) of in $CDCl_3$ (600 MHz).

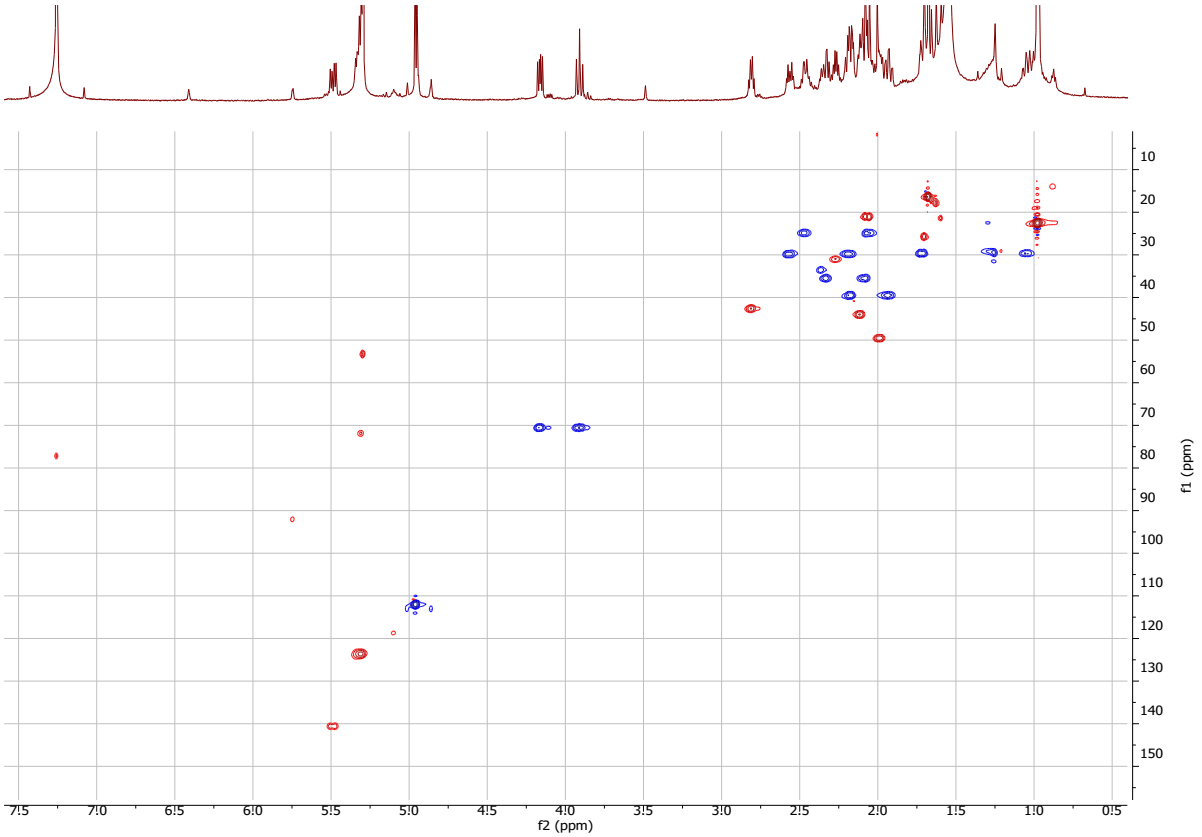


Figure S43: gHSQCAD spectrum coraxeniolide A (**7**) of in $CDCl_3$ (600 MHz).

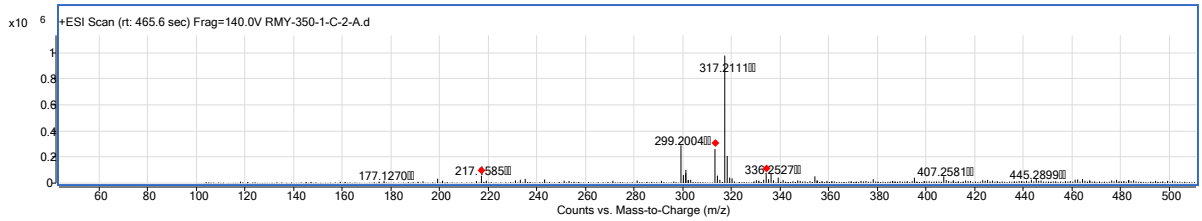


Figure S44:HRESIMS spectrum of 9-deoxyxeniolide A (**8**).

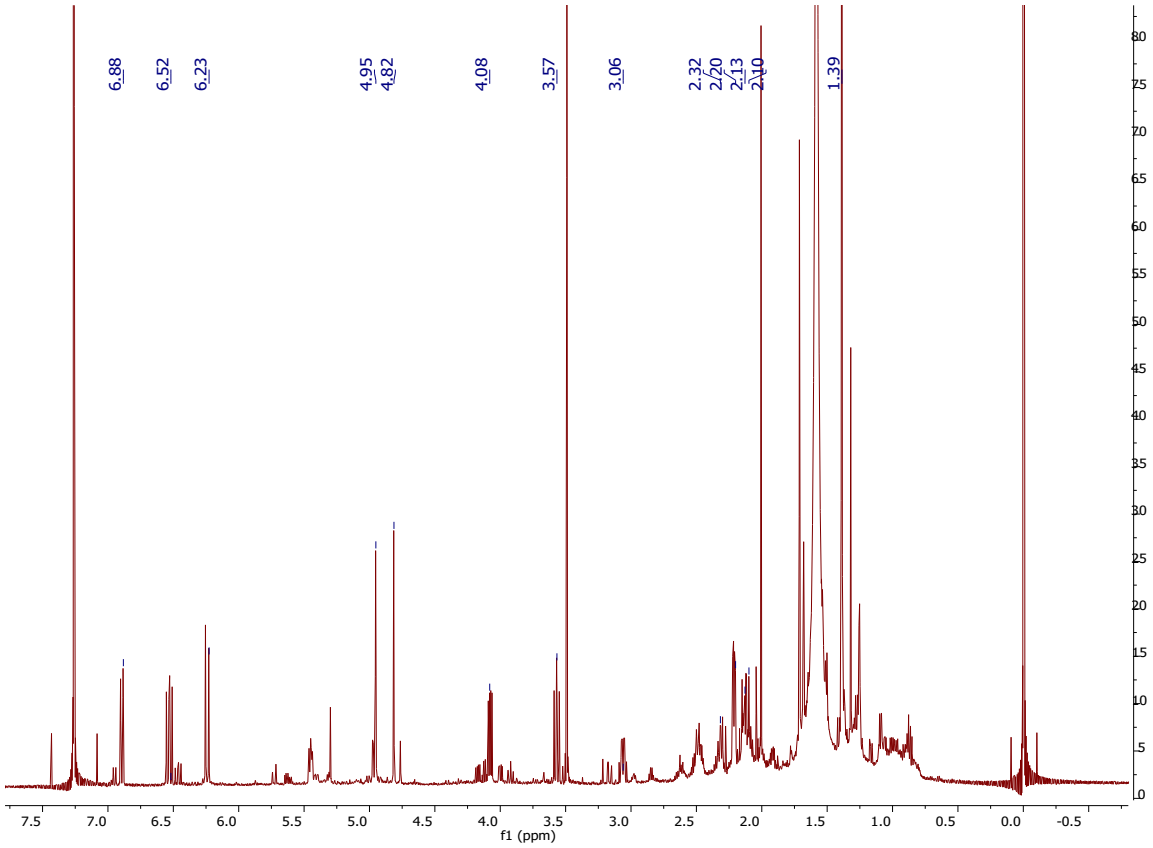


Figure S45: ^{1}H NMR spectrum 9-deoxyxeniolide A (**8**) of in $CDCl_3$ (600 MHz).

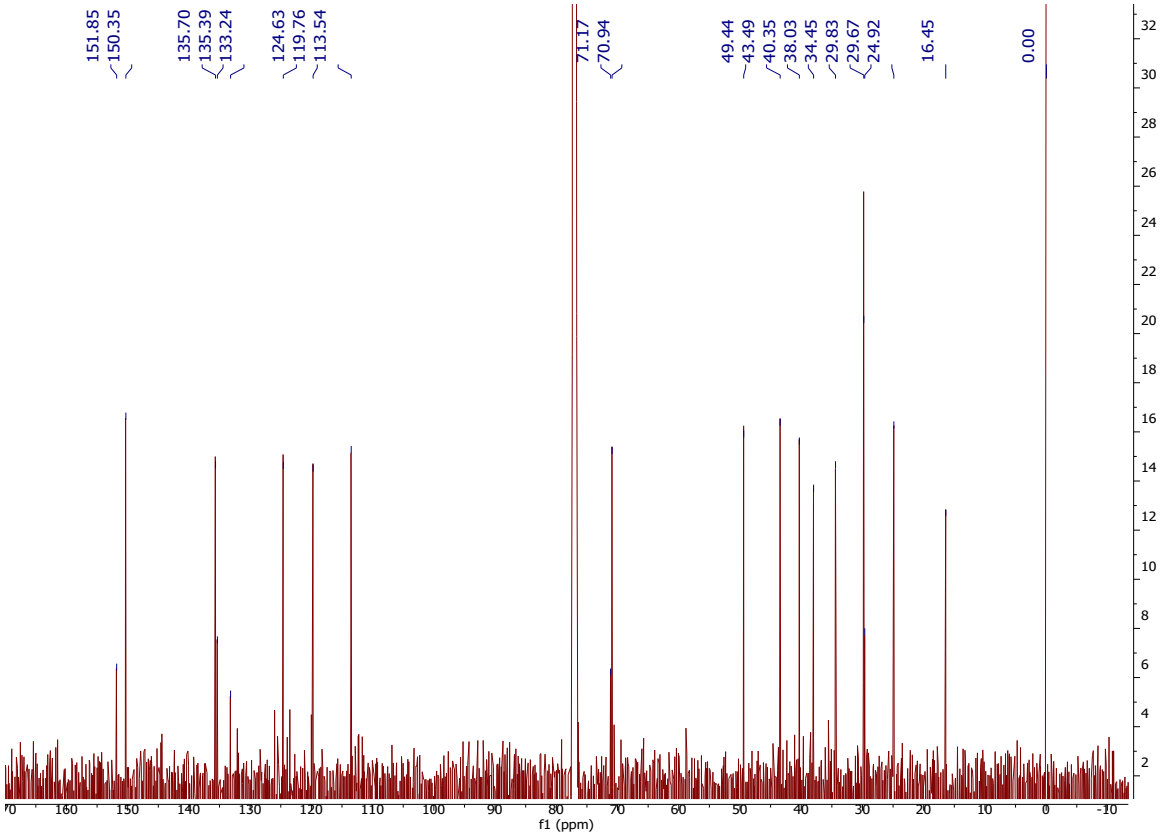


Figure S46: ^{13}C NMR spectrum 9-deoxyxeniolide A (**8**) of in $CDCl_3$ (150 MHz).

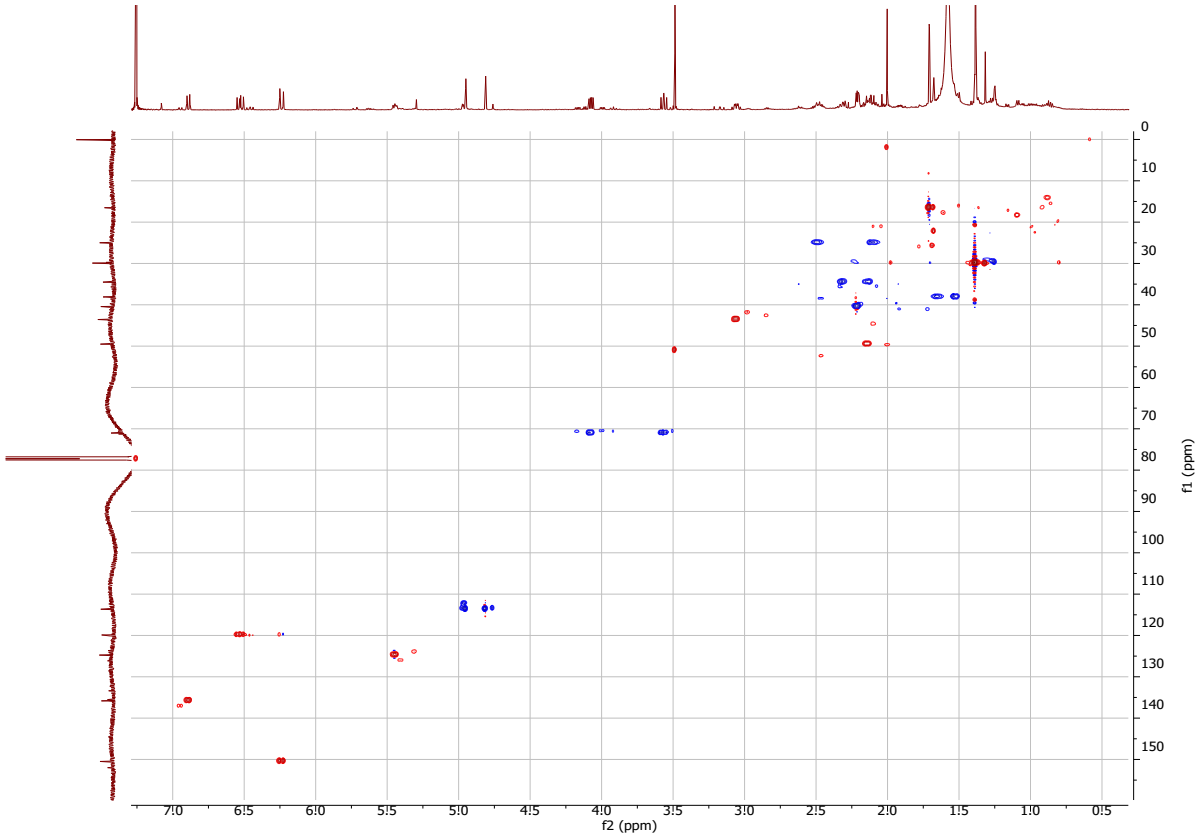


Figure S47: gHSQCAD NMR spectrum 9-deoxyxeniolide A (**8**) of in $CDCl_3$ (600 MHz).