



HAL
open science

Validation d'une méthode de déréplication de substances naturelles marines

Léonie Pellissier

► **To cite this version:**

Léonie Pellissier. Validation d'une méthode de déréplication de substances naturelles marines. Sciences pharmaceutiques. 2017. dumas-01500339

HAL Id: dumas-01500339

<https://dumas.ccsd.cnrs.fr/dumas-01500339>

Submitted on 3 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il n'a pas été réévalué depuis la date de soutenance.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact au SID de Grenoble :
bump-theses@univ-grenoble-alpes.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

<http://www.cfcopies.com/juridique/droit-auteur>

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

UNIVERSITÉ GRENOBLE ALPES
FACULTÉ DE PHARMACIE DE GRENOBLE

Année : 2017

N°

VALIDATION D'UNE METHODE DE DEREPLICATION DE PRODUITS NATURELS
MARINS

THÈSE
PRESENTÉE POUR L'OBTENTION DU TITRE DE DOCTEUR EN PHARMACIE
DIPLOME D'ÉTAT

Léonie PELLISSIER

[Données à caractère personnel]

THÈSE SOUTENUE PUBLIQUEMENT À LA FACULTÉ DE PHARMACIE DE
GRENOBLE

Le : 03/02/2017

DEVANT LE JURY COMPOSÉ DE

Président du jury :

Mme. Delphine ALDEBERT

Membres :

M. Ahcène BOUMENDJEL

M. Benjamin BOUCHERLE

M. Matthieu SANCHEZ

La Faculté de Pharmacie de Grenoble n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions sont considérées comme propres à leurs auteurs.

Doyen de la Faculté : **M. le Pr. Michel SEVE**

Vice-doyen et Directrice des Etudes : **Mme Christine DEMEILLIERS**

Année 2016-2017

ENSEIGNANTS A L'UFR DE PHARMACIE

STATUT	NOM	PRENOM	LABORATOIRE
MCU	ALDEBERT	Delphine	LAPM - UMR CNRS 5163
PU-PH	ALLENET	Benoit	THEMAS TIMC-IMAG UMR CNRS 5525
PU	BAKRI	Aziz	TIMC-IMAG CNRS UMR 5525
ATER	BARDET	Jean-Didier	
MCU	BATANDIER	Cécile	LBFA - INSERM U1055
MCU-PH	BEDOUCHE	Pierrick	THEMAS TIMC-IMAG UMR CNRS 5525
MCU	BELAIDI-CORSAT	Elise	HP2 - INSERM U1042
MAST	BELLET	Béatrice	-
ATER	BOUCHERLE	Benjamin	DPM
DCE	BOULADE	Marine	SyMMES
PU	BOUMENDJEL	Ahcène	DPM -UMR 5063 UJF CNRS
DCE	BOURDIER	Guillaume	HP2
MCU	BOURGOIN	Sandrine	IAB - CRI INSERM UJF U823
DCE	BOUVET	Raphaël	HP2
MCU	BRETON	Jean	L.C.I.B. - UMR E3 CEA UJF
MCU	BRIANCON-MARJOLLET	Anne	HP2 - INSERM U1042
DCE	BROCCO	Benjamin	ILL
MCU	BUDAYOVA SPANO	Monika	IBS - UMR 5075 CEA CNRS UJF
PU	BURMEISTER	Wim	UVHCI - UMI 3265 UJF EMBL CNRS
MCU-PH	BUSSER	Benoit	IAB - CRI INSERM UJF U823
Professeur émérite	CALOP	Jean	-
MCU	CAVAILLES	Pierre	LAPM - UMR 5163 CNRS UJF
AHU	CHANOINE	Sébastien	THEMAS TIMC-IMAG UMR CNRS 5525
MCU	CHOISNARD	Luc	DPM -UMR 5063 UJF CNRS
AHU	CHOVELON	Benoit	DPM -UMR 5063 UJF CNRS
PU-PH	CORNET	Murielle	THEREX - TIMC IMAG UMR 5525 CNRS UJF
DCE	COUCHET	Morgane	LBFA
PU-PH	DANEL	Vincent	SMUR SAMU
PU	DECOUT	Jean-Luc	DPM -UMR 5063 UJF CNRS
MCU	DELETRAZ- DELPORTE	Martine	Equipe SIS -EAM 4128 UCB
MCU	DEMEILLIERS	Christine	LBFA - INSERM U1055
PU	DROUET	Christian	AGIM - CNRS 3405

STATUT	NOM	PRENOM	LABORATOIRE
PU	DROUET	Emmanuel	UVHCI - UMI 3265 UJF-EMBL- CNRS
MCU	DURMORT - MEUNIER	Claire	I.B.S – UMR 5075 CEA UJF CNRS
PU-PH	FAURE	Patrice	HP2- INSERM U1042
PRCE	FITE	Andrée	-
AHU	GARNAUD	Cécile	THEREX – TIMC IMAG UMR 5525 NCRS UJF
PRAG	GAUCHARD	Pierre-Alexis	-
MCU-PH	GERMI	Raphaëlle	UVHCI, UMI 3265 UJF-EMBL- CNRS
MCU	GEZE	Annabelle	DPM –UMR 5063 UJF CNRS
MCU	GILLY	Catherine	DPM –UMR 5063 UJF CNRS
PU	GODIN-RIBUOT	Diane	HP2- INSERM U1042
PRCE	GOUBIER MATHYS	Laurence	-
Professeure émérite	GRILLOT	Renée	-
MCU	GROSSET	Catherine	DPM –UMR 5063 UJF CNRS
MCU	GUIEU	Valérie	DPM –UMR 5063 UJF CNRS
AHU	HENNEBIQUE	Aurélie	sous réserve de création de poste
MCU	HININGER-FAVIER	Isabelle	LBFA - Inserm U1055
MCU	JOYEUX-FAURE	Marie	HP2- INSERM U1042
MCU	KHALEF	Nawel	TIMC-IMAG CNRS UMR 5525
MCU	KRIVOBOK	Serge	LCBM, IRTSV CEA
DCE	LE	Cong Anh Khanh	CERMAV
PU	LENORMAND	Jean Luc	THEREX, TIMC-IMAG
DCE	MARILLIER	Mathieu	HP2
PU	MARTIN	Donald	TIMC-IMAG, UMR 5525 UJF CNRS
AHU	MAZET	Roseline	DPM –UMR 5063 UJF CNRS
MCU	MELO DE LIMA	Christelle	L.E.C.A – UMR CNRS 5553
PU	MOINARD	Christophe	LBFA - Inserm U1055
DCE	MONTEMAGNO	Christopher	LRB
ATER	MORAND	Jessica	HP2
PU-PH	MOSSUZ	Pascal	THEREX - TIMC-IMAG UMR 5525 CNRS
MCU	MOUHAMADOU	Bello	L.E.C.A – UMR CNRS 5553
DCE	MOULIN	Sophie	HP2
DCE	NADER	Serge	LCBM
DCE	NGUYEN	Kim-Anh	DPM
MCU	NICOLLE	Edwige	DPM –UMR 5063 UJF CNRS
MCU	OUKACINE	Farid	DPM –UMR 5063 UJF CNRS
MCU	PERES	Basile	DPM- UJF/CNRS UMR 5063
DCE	PERONNE	Lauralie	IAB
MCU	PEUCHMAUR	Marine	DPM –UMR 5063 UJF CNRS
PU	PEYRIN	Éric	DPM –UMR 5063 UJF CNRS
MCU	RACHIDI	Walid	L.C.I.B - UMR E3 CEA/UJF
MCU	RAVELET	Corinne	DPM –UMR 5063 UJF CNRS
PU	RIBUOT	Christophe	HP2- INSERM U1042

STATUT	NOM	PRENOM	LABORATOIRE
PAST	RIEU	Isabelle	-
Professeure émérite	ROUSSEL	Anne -Marie	-
PU-PH	SEVE	Michel	CR INSERM / UJF U823 Institut Albert Bonniot
MCU	SOUARD	Florence	DPM –UMR 5063 UJF CNRS
DCE	TAHER	Raleb	IBS
ATER	TAHMASEBI	Faezeh	TIMC-IM2AG
MCU	TARBOURIECH	Nicolas	UVHCI, UMR 3265 UJF-EMBL- CNRS
DCE	TODOROV	Zlatomir	BCI
PAST	TROUILLER	Patrice	-
DCE	VACHEZ	Yvan	CRI-GEN
MCU	VANHAVERBEKE	Cécile	DPM –UMR 5063 UJF CNRS
DCE	VERNET	Céline	CRI-IAB
DCE	VRAGNIAU	Charles	UVHCI
PU	WOUESSIDJEWE	Denis	DPM –UMR 5063 UJF CNRS

ATER : Attachés Temporaires d'Enseignement et de Recherches

BCI : Biologie du Cancer et de l'Infection

CHU : Centre Hospitalier Universitaire

CIB : Centre d'Innovation en Biologie

CRI : Centre de Recherche INSERM

CNRS : Centre National de Recherche Scientifique

DCE : Doctorants Contractuels Enseignement

DPM : Département de Pharmacochimie Moléculaire et de Cognition et Ontogenèse »

HP2 : Hypoxie Physiopathologie Respiratoire et Cardiovasculaire

IAB : Institut Albert Bonniot,

IBS : Institut de Biologie Structurale

JR : Jean Roget

LAPM : Laboratoire Adaptation et Pathogenèse des Microorganismes

LBFA : Laboratoire Bioénergétique Fondamentale et Appliquée

LCBM : Laboratoire Chimie et Biologie des Métaux

LCIB : Laboratoire de Chimie Inorganique et Biologie

LECA : Laboratoire d'Ecologie Alpine

LR : Laboratoire des Radio pharmaceutiques

MCU : Maitre de Conférences des Universités

MCU-PH : Maitre de Conférences des Universités et Praticiens Hospitaliers

PAST : Professeur Associé à Temps Partiel

PRAG : Professeur Agrégé

PRCE : Professeur certifié affecté dans l'enseignement

PU : Professeur des Universités

PU-PH : Professeur des Universités et Praticiens Hospitaliers

TIMC-IMAG : Laboratoire Technique de l'Imagerie, de la Modélisation

UMR: Unité Mixte de Recherche

UVHCI: Unit of Virus Host Cell Interactions

Remerciements

À ma famille qui me donne la liberté d'explorer et de vivre de merveilleuses aventures, en m'apportant toute l'affection et le soutien dont j'ai besoin pour les accomplir.

A Matthieu Sanchez, dont le travail de pharmacien m'inspire depuis que je suis toute petite.

A mes professeurs et copains de Grenoble pour l'environnement de travail et de vie très sympathiques qu'ils ont pu créer et qui étaient pour nous une, source d'inspiration pour que nous devenions de bons pharmaciens et chimistes! Merci surtout aux professeurs Boumendjel et Aldebert qui m'ont donné l'opportunité de construire ma propre voie dans le monde de la recherche!

A mes professeurs et camarades du master à Paris, les divers univers qu'ils nous ont fait découvrir, au sein de ce jardin botanique fascinant.

A mes amis en Ecosse, leur accueil chaleureux, leur compagnie lors de belles aventures dans la campagne et lors de mon initiation aux rudiments de la vie écossaise.

À Jioji, mon patient mentor, pour toutes les connaissances qu'il m'a appris, des astuces de laboratoire aux éléments essentiels de l'élucidation structurale. Au personnel du MBC, notamment Russel et Andrea, qui ont pris grand soin de nos analyses.

À mes collègues de laboratoire, notamment Alan, sa générosité et sa patience, et Nathalie, ses conseils, sa présence rassurante et ses conversations animées en français.

Table of Contents

Abstract	1
Abbreviations	2
Préambule	4
Introduction Générale	4
Chapter 1 General introduction	16
1.1 A growing health crisis	16
1.2 Nature as a source of new medicine	17
1.3 Marine biodiversity	18
1.4 Marine chemo diversity	19
1.5 Exploring the hidden potential of oceans for new drugs	26
1.6 Bioprospecting pipeline for marine drug discovery	27
1.7 The dereplication process	28
1.8 Aim of the thesis	31
Chapter 2 Validation of the PharmaSea dereplication method	32
2.1 Introduction: the challenges of the dereplication process	32
2.2 Method for the validation of the dereplication strategy	36
2.3 Results and discussion	37
2.3.1 Prioritization by LCMS.....	37
2.3.2 Prioritization by NMR.....	39
2.4 Conclusion	44
2.5 Case study n°1 : Tavarua 2	45
2.5.1 Introduction.....	45
2.5.2 Methods.....	46
2.5.3 Results.....	47
2.6 Case study n°2 : Balolo 3.14	52
2.6.1 Methods.....	52
2.6.2 Results.....	52
2.7 General discussion	55

Chapter 3 A different perspective: isolation of new neuroactive compounds from the marine bacteria <i>Leeuwenhoekiella sp.</i>	60
3.1 Introduction	60
3.2 Methods	62
3.2.1 Isolation, typing and phylogeny of strain in University of Tromso (UoT)	62
3.2.2 Original Bioassays of X0127.....	62
3.2.3 Bioassay-Guided Purification of X0127 extract.....	63
3.3 Ongoing studies	63
Chapter 4 Conclusion and outlook.....	64
Conclusion	65
References	67
Supplementary data	72

Index of figures

Figure 1. Processus de déréplication utilisé au Marine Biodiscovery Centre.....	11
Figure 2. Méthode utilisée pour la validation de la méthode de déréplication PharmaSea	13
Figure 3. The chemical structure of Sorafenib.....	17
Figure 4. a. Lamellarin D (alkaloid) b. Halichondrin B (terpene)	19
Figure 5. Jorumycin isolated from a Pacific nudibranch <i>Jorunna funebris</i>	20
Figure 6. Distribution of new natural products (n=5286) isolated from 2000 to 2009 according to their chemical families.....	21
Figure 7. Distribution of nitrogenous and non-nitrogenous compounds among sea species.....	23
Figure 8. Helminal and kainic acid	24
Figure 9. Structure of Bryostatin 1.....	25
Figure 10. A schematic presentation of the PharmaSea project	27
Figure 11. Bioprospecting pipeline for Marine Natural Products discovery	28
Figure 12. Structure elucidation workflow used at MBC	30
Figure 13. Dereplication workflow followed at the MBC	32
Figure 14. Ion chromatogram and ions extraction	33
Figure 15. Presentation of IntelliXtract software and data input	34
Figure 16. Method used for the validation of the dereplication method.....	36
Figure 17. Number of new masses extracted by IX in each sample	37
Figure 18. Percentage of new masses (m/z) compared to the total number of masses per extracts	38
Figure 19. Exact mass (m/z) compared to retention time (min), for each SPE fraction of Fiji samples.....	40
Figure 20. Exact mass (m/z) compared to retention time (min), for each SPE fraction of Fiji samples.....	41
Figure 21. ¹ H NMR spectra of the four fractions of Tavarua 2 after fractionation by C18-SPE	42
Figure 22. The cyclopeptide Patellamide D isolated from the Tunicate <i>Lissoclinum Patella</i> ...	46
Figure 23. The cyclopeptide Didemnid B isolated from the Caribbean Ascidian <i>Trididemnum solidum</i>	46
Figure 24. Structure of compound 1 and selected HMBC and COSY correlations.....	50
Figure 25. Structure of compound 2 isolated from Tavarua 2 SPE 100% MeOH.....	50
Figure 26. Compounds 3 and 4 isolated from BALOLO 3.....	54

Figure 27. Exact mass (m/z) compared to retention time (min), for Tavarua 2 with labelled compound 1	56
Figure 28. Palaemon macrodactylus and 2,3 indolinediol	62

Index of tables

Table 1. Pharmacodiversity of Marine organisms	20
Table 2. Examples of marine natural products currently under clinical trials	26
Table 3. Excel table of new masses extracted from IntelliXtract analysis.....	35
Table 4. ¹ H NMR based classification of Fiji samples	43
Table 5. ¹ H-NMR and ¹³ C-NMR data (600 MHz, CD ₃ OD) for 1 , δ _C in ppm, J in Hz	49
Table 6. ¹ H-NMR and ¹³ C-NMR data (400 MHz, CD ₃ OD) for 2 , δ _C in ppm, J in Hz	51
Table 7. ¹ H-NMR and ¹³ C-NMR data (400 MHz, CD ₃ OD) for 3 and 4 δ _C in ppm, J in Hz.....	53
Table 8. Summary of compounds isolated from prioritized samples.....	55

Index of supplementary data

Figure S1. Exact mass (m/z) compared to retention time (min), for few samples studied.....	73
Figure S2. ^1H NMR spectrum of some of the top ranked samples in CD_3OD at 400 MHz....	75
Figure S3. HPLC chromatogram of Tavarua 2 SPE 50% MeOH-water.....	93
Figure S4. UV/Vis spectra of compound 1 in MeOH.....	93
Figure S5. Search and matching of UV spectra of compound 1 (green) and an indole derivative (red) in ACD lab spectrus processor (ACD/Structure Elucidator, version 15.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2015).	94
Figure S6. LC-HRESIMS spectrum of the new compound 1.....	95
Figure S7. HRESIMS/MS spectrum of the new compound 1.....	96
Figure S8. Fragmentation pattern of compound 1.....	96
Figure S9. ^1H NMR spectrum of compound 1 in CD_3OD at 600 MHz.....	97
Figure S10. HSQC spectrum of compound 1 in CD_3OD at 600 MHz.....	97
Figure S11. COSY spectrum of compound 1 in CD_3OD at 600 MHz.....	98
Figure S12. HMBC spectrum of compound 1 in CD_3OD at 600 MHz.....	98
Figure S13. HPLC chromatogram of Tavarua 2 SPE 100% MeOH.....	99
Figure S14. UV/Vis spectra of compound 2 in MeOH.....	99
Figure S15. LC-HRESIMS spectrum of compound 2.....	100
Figure S16. ^1H NMR spectrum of compound 2 in CD_3OD at 600 MHz.....	101
Figure S17. HSQC spectrum of compound 2 in CD_3OD at 600 MHz.....	101
Figure S18. HMBC spectrum of compound 2 in CD_3OD at 600 MHz.....	102
Figure S19. COSY spectrum of compound 2 in CD_3OD at 600 MHz.....	102
Figure S20. HPLC chromatogram of Balolo 3 SPE 100% MeOH.....	103
Figure S21. UV/Vis spectra of compound 3 in MeOH.....	103
Figure S22. UV/Vis spectra of compound 4 in MeOH.....	103
Figure S23. LC-ESIMS spectrum of the compound 3.....	104
Figure S24. HRTOFMS spectrum of compound 3.....	104
Figure S25. ^1H NMR spectrum of compound 3 in CD_3OD at 600 MHz.....	105
Figure S26. ^{13}C NMR spectrum of compound 3 in CD_3OD at 600 MHz.....	105
Figure S27. HSQC spectrum of compound 3 in CD_3OD at 600 MHz.....	106
Figure S28. HMBC spectrum of compound 3 in CD_3OD at 600 MHz.....	106
Figure S29. COSY spectrum of compound 3 in CD_3OD at 600 MHz.....	107
Figure S 30 LC-ESIMS spectrum of compound 4.....	108

Figure S 31 ¹ H NMR spectrum of compound 4 in CD ₃ OD at 600 MHz.....	108
Figure S 32 HSQC spectrum of compound 4 in CD ₃ OD at 600 MHz.....	109
Figure S 33 HMBC spectrum of compound 4 in CD ₃ OD at 600 MHz.....	109
Figure S 34 COSY spectrum of compound 4 in CD ₃ OD at 600 MHz.....	110
Figure S 35 ¹ H NMR spectrum of compound 1 HPLC fraction H12 (top) and ¹ H NMR spectrum of compound 5 from HPLC fraction H11 (bottom) isolated from Tavarua 2 SPE 50% MeOH-water ,in CD ₃ OD at 600 MHz.....	111
Figure S 36 HRESIMS spectrum of Tavarua 2 SPE 50% MeOH (with fraction H11).....	112
Figure S 37 HRESIMS spectrum of Tavarua 2 SPE 50% MeOH (with fraction H10).....	112
Figure S 38 ¹ H NMR spectrum of the compound 6 isolated from Tavarua 2 SPE fraction 100% MeOH+TFA in CD ₃ OD at 600 MHz.....	113
Figure S 39 HPLC chromatogram of Tavarua 2 SPE 100% MeOH+TFA.....	114
Figure S 40 UV/Vis spectra of compound 6 in MeOH.....	114
Figure S 41 ¹ H NMR spectrum of the compound 6 isolated from Tavarua 2 SPE fraction 100% MeOH+TFA in CD ₃ OD at 600 MHz.....	115
Figure S 42 HSQC spectrum of compound 6 in CD ₃ OD at 600 MHz.....	115
Figure S 43 HMBC spectrum of compound 6 in CD ₃ OD at 600 MHz.....	116
Figure S 44 COSY spectrum of compound 6 in CD ₃ OD at 600 MHz.....	116

Abstract

Rising incidences of various cancers and an ever-growing range of infections caused by microorganisms are the main issues of modern medicine, co evolving with the respective drug targeting resistances. This “growing health crisis”, as defined by the World Health Organization in 2014, prompts researchers to find new resources of new possible drug leads. Despite the overwhelming number of terrestrial natural products derivatives approved by the Food and Drug Administration (FDA) as pharmaceuticals, industry has reduced its reliance on it, in part due to the costs related to high rate of rediscovery of known compounds. Therefore, academicians and private industry have turned their efforts toward exploring the marine world as new sources of bioactive molecules. In this context, we contributed to this effort in participating in a research project named PharmaSea. Scientists from this EU-funded project focus on the bio discovery, the development and commercialisation of new substances from marine organisms.

The key step to successful drug bio discovery from marine sources is the rapid identification of the known compounds before isolating them. Called dereplication, this approach is essential to assess chemical novelty of crude extracts and their fractions. The large amount of data sets and the lack of confirmatory data make this task arduous. To meet these challenges, different strategies of dereplication procedures involving LC-HRMS are currently emerging. In this regard, we investigated the PharmaSea dereplication method on 21 tunicates extracts. This method is based on a combination of both the ACDLabs IntelliXtract (IX) software with the MarinLit database from the Royal Society of Chemistry (RSC) to dereplicate LCMS data. Additionally, ^1H NMR spectra analysis was incorporated in the strategy to streamline the prioritization of samples for the discovery of new secondary metabolites. The extracts were ranked for prioritization using both methods. The most interesting sample according to this methodology was selected for further purification. This led to the isolation of seven compounds of which five were new as evidence of the success of the method.

Abbreviations

ACD	Advanced Chemistry Development
ACN	Acetonitrile
AMR	Anti-Microbial Resistance
CASE	Computer Assisted Structure Elucidation
COSY	Homonuclear Correlation Spectroscopy
DB	Database
DNA	DesoxyriboNucleic Acid
DNP	Dictionnary of Natural Products
ESI	Electro-Spray Ionisation
FDA	Food and Drug Administration
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High Pressure Liquid Chromatography
HRESIMS	High Resolution Electro Spray Ionisation Mass Spectrometry
HRMS	High Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Correlation
IX	IntelliXtract
LC	Liquid Chromatography
LTQ	Linear Trap Quadrupole
MBC	Marine Biodiscovery Center
MeOH	Methanol
MNP	Marine Natural Product
MS	Mass Spectrometry
NCI	National Cancer Institute
NMR	Nuclear Magnetic Resonance

NP	Natural Products
RNA	RiboNucleic Acid
RP	Reverse Phase
RSC	Royal Society of Chemistry
SPE	Solid Phase Extraction
TFA	Trifluoroacetic Acid
T_R	Retention Time
UV	UltraViolet
WHO	World Health Organization
δ	Chemical shift [ppm]
J	Coupling constant

Préambule

Ce travail de thèse est issu de mon projet de stage de recherche, faisant partie intégrante du diplôme de Master 2 intitulé «Molécules et Cibles Thérapeutiques», réalisé au Muséum National d'Histoire Naturelle de Paris. Ce stage s'est déroulé au sein du «Marine Biosdiscovery Center» à Aberdeen en Ecosse, de Janvier à Juillet 2016 et portait sur la recherche de substances bioactives dans des organismes marins.

Le contexte, les objectifs, les méthodes et les résultats de ce projet seront présentés succinctement en français. Ces mêmes items seront présentés et développés en anglais.

Introduction générale

Nous sommes actuellement dans un contexte de «crise de santé planétaire» déclarée par l'Organisation Mondiale de la Santé (OMS) en 2014. Cette crise est marquée par l'augmentation persistante des maladies infectieuses et cancéreuses, associée au développement d'une résistance aux traitements actuels. Suite à un récent rapport basé sur les données recueillies auprès de 114 pays, l'OMS a reconnu la résistance antimicrobienne (AMR) comme « une menace de plus en plus grave pour la santé publique mondiale capable d'affecter toute personne, de tout âge, dans toute région du monde ». Ainsi, face à l'augmentation du nombre de microbes devenant résistants, l'arsenal thérapeutique couramment utilisé devient presque inutile. Par exemple, la résistance aux Fluoroquinolones, antibiotiques des plus largement utilisés dans les infections des voies urinaires, s'est répandue dans de nombreuses régions du monde où ce traitement est maintenant inefficace chez plus de la moitié des patients (1). En outre, la résistance aux antibiotiques est inévitablement associée à une durée de traitement plus longue et des coûts de traitement plus importants. Conjointement, l'évolution de la résistance aux anticancéreux, constitue un obstacle majeur aux traitements par chimiothérapie. Une fois qu'un cancer a développé une résistance à une classe de médicaments, il est plus susceptible de développer une résistance à d'autres classes de molécules, impliquant divers mécanismes d'action, (2). Face, à ces challenges, la recherche dans le domaine du médicament est plus que jamais sollicitée pour découverte de nouvelles molécules médicamenteuses. Au cours des 30 dernières années, la

chimie combinatoire a été utilisée comme source de découverte de nouveaux candidats médicaments. Cependant, à ce jour, seulement un faible nombre de Nouvelles Entités Chimiques (NEC) résultant de cette approche a mis sur le marché (3).

Dans ce contexte, les molécules d'origine naturelle sont de nouveau devenues très courtoisies par la recherche académique et l'industrie du médicament.

Les végétaux ont une longue histoire dans le domaine de la santé. Jusqu'au 20^{ème} siècle, les extraits de végétaux, animaux, microbes et minéraux ont été les seules « thérapies » utilisées. Les études menées sur les fossiles ont pu établir que l'usage des plantes dans la médecine traditionnelle remonte jusqu'au paléolithique (il y a 60 000 ans) (4). Actuellement, selon les rapports de l'OMS, environ 65% de la population mondiale a intégré la médecine traditionnelle par les plantes comme moyen de traitements. Des plantes telles que la réglisse (*Glycyrrhiza glabra*), la Myrrhe (*Commiphora Myrrha*), et la capsule de pavot (*Papaver somniferum*) sont indiquées dans les premiers écrits de Mésopotamie en 2600 av JC et sont aujourd'hui utilisées comme plantes médicinales ou ingrédients de médicaments commercialisés. Le 20^{ème} siècle a ensuite révolutionné l'utilisation de ces traitements. En effet, l'idée que l'effet du médicament dans le corps humain soit médié le résultats d'interactions spécifiques entre le principe actif et des macromolécules biologiques a conduit les scientifiques à penser que des composés chimiques contenus dans les extraits, plutôt que certains « pouvoirs mystiques » seraient les facteurs de leur activité biologique. Ainsi, sur la base de la médecine traditionnelle, de nombreuses entités chimiques uniques ont pu être introduites dans la pratique moderne et utilisées par l'industrie pharmaceutique comme matière première dans l'élaboration de nombreuses thérapeutiques.

En raison de leur coévolution avec des cibles biologiques, les produits naturels (PN) sont en effet de bons points de départ pour la découverte de médicaments. Leur nouveauté chimique est supérieure à toute autre source : 40% des motifs structuraux du Dictionnaire des Produits Naturels (DNP, base de données publiée) sont absents de la chimie synthétique. L'étude de leur diversité chimique montre que les collections de composés naturels sont plus variées que ceux synthétiques, notamment ceux émanant de la chimie combinatoire. De plus, ces produits naturels communément appelés métabolites secondaires étant élaborés au sein de systèmes vivants, montrent plus de compatibilité biologique avec les êtres humains et semblent posséder une « drug-likeness » supérieure aux molécules synthétiques. Cela signifie qu'ils peuvent être absorbés, distribués et métabolisés par le corps humain. Ainsi, les coûts de développement de la

production de médicaments actifs par voie orale sont vraisemblablement plus faibles qu'avec les produits biotechnologiques ou avec la plupart des composés de synthèse (5).

L'étude de la bio activité en fonction de la source des composés a démontré que les PN sont utiles pour traiter environ 87% des pathologies humaines, entre autres comme antibiotiques, anti parasitaires, anti-cancéreuses, anticoagulantes ou encore immunosuppressives (6). Les PN et leurs dérivés composent près de 60% des nouveaux médicaments anticancéreux et 75% des nouveaux agents antibactériens (7). C'est ainsi que la morphine et la codéine isolées du pavot somnifère (*Papaverum Somniferum*) ont été développées comme médicaments et sont encore actuellement sur le marché en tant qu'analgésiques. Citons aussi l'hemisuccinate de carbenoxolone sodium, dérivé hémisynthétique de l'acide glycyrrhétique isolé de la réglisse, utilisé dans le traitement d'ulcères gastro duodénaux. L'exploitation du règne végétal a aussi conduit à la découverte d'important agents anticancéreux tels que la paclitaxel issu de *Taxus brevifolia* ou encore la camptothécine issue de *Camptotheca acuminata* (8).

Entre 1981 et 2002, environ 28% des 868 Nouvelles Entités Chimiques (NEC) étaient des produits naturels et 24% des molécules créées par synthèse ont été inspirées d'un pharmacophore naturel (9). Ces molécules ont représenté respectivement 40, 24 et 26% des 35 médicaments les plus délivrés au monde en 2000, 2001 et 2002 (6). D'après une étude américaine, 84 prescriptions médicales sur 150 contenaient des produits naturels ou dérivés. Ces produits étaient prescrits principalement comme anti allergiques, agents respiratoires, analgésiques, cardiovasculaires ou anti infectieux (10). Parmi les NEC mis sur le marché entre 2000 et 2005, nous pouvons citer trois exemples. L'apomorphine hydrochloride (Apokyne, Bertek, 2004) dérivé de la morphine issu du pavot (*Papaver somniferum*). C'est un agoniste des récepteurs dopaminergiques utilisé actuellement en sous cutané dans les stades de fluctuations de la maladie de Parkinson. Le tiotropium bromide est lui un dérivé de l'atropine issue la Belladonne (*Atropa Belladonna*). Il constitue un inhibiteur réversible non sélectif des récepteurs muscarinique préconisé dans le traitement de la Broncho Pneumopathie Chronique Obstructive (Spiriva Handihaler, Boehringer Ingelheim, 2004). Citons enfin l'Arteether (Artecef, Artecef BV, 2000) agent anti malarique, dérivé de l'artémisine isolée de l'Armoise annuelle (*Artemisia annua*). En 2010, les produits naturels comptent toujours pour quasiment la moitié des NEC, et notamment deux des sept nouveaux agents antitumoraux en dérivent : la vinflunine, dérivé alcaloïdiques issu de la Pervenche de Madagascar (*Catharanthus roseus L.*), et le cabazitaxel, dérivé de taxanes issu des plantes du genre *Taxus*.

L'attribution du prix Nobel 2016 de médecine ou physiologie à des scientifiques travaillant sur la valorisation des produits naturels est un argument fort en faveur de l'intérêt majeur des plantes comme sources de médicaments.

Le monde marin comme source de molécules bioactives

La contribution écrasante des produits naturels à l'expansion de la ressource pharmaceutique est alors clairement évidente, mais une grande partie de ce «trésor de petites molécules» de la nature reste à explorer. Cependant, cette ressource terrestre tant à s'épuiser et à devenir difficile d'accès. De plus, on observe un manque de rentabilité et d'innovation due à la "redécouverte" fréquente d'agents connus dans les PN végétaux (11). C'est ainsi que la recherche s'est progressivement tournée vers le milieu marin encore peu exploré pour trouver une nouvelle source de biodiversité et de nouveauté. La richesse et la diversité des produits naturels issus de l'environnement marin suggèrent en effet de grandes opportunités pour découvrir de nouveaux candidats médicaments.

Représentant environ 71% de la surface de la Terre, les océans sont fondamentaux pour notre planète. Le premier recensement de la vie marine (2000-2010) (12) et de récentes expéditions (ANDEEP Antarctic benthic deep-sea biodiversity project, 2007) ont confirmé un niveau de biodiversité particulièrement élevé, comprenant un large variété de taxons, bien supérieure à celle du milieu terrestre. Le nombre d'espèces marines connues est estimé à 250000 et extrapolé à au moins un million et le nombre d'espèces microbiennes à des centaines de millions. Vivant dans un milieu hostile et dilué, et ayant parfois très peu de moyens de transports, les organismes marins dépendent de messagers chimiques primordiaux (produits naturels) pour communiquer et se défendre. Ils produisent donc une panoplie de nouveaux métabolites secondaires (« écomones »), comprenant une variété de familles chimiques, d'activités pharmacologiques et des structures originales (13). Issu d'organismes procaryotes et eucaryotes et caractérisés par la présence d'éléments chimiques rares dans les PN (chlore, brome...), ces composés exhibent une chimie unique et des structures originales, différentes de celles observées dans les produits naturels terrestres (14). Leur spécificité pour les organismes procaryotes et leur faible toxicité pour les cellules eucaryotes en ont fait une source de composés d'intérêt thérapeutique, auxquels la recherche s'intéresse de plus en plus. Le premier agent marin autorisé sur le marché par la FDA en 1969 fut la Cytarabine, agent chimio thérapeutique dérivé d'une éponge marine, *Cryptotethya crypta* (14). A l'heure actuelle, environ 15 produits naturels marins sont en cours d'essais cliniques, un exemple étant le Ziconotide (Prialt®), issu du venin d'un escargot marin (*Conus magus*), approuvé par la FDA en 2004 dans le traitement des douleurs chroniques. Ces

découvertes ont donné aux scientifiques l'espoir de trouver dans les eaux marines les nouveaux modèles moléculaires et pharmacologiques nécessaires à l'élaboration de notre « armement thérapeutique ».

C'est ainsi que plusieurs laboratoires européens ont fondé le projet PharmaSea, les menant dans les milieux marins les plus extrêmes : zones les plus froides, les plus chaudes ou les plus profondes dans le but de récolter des échantillons dans ces milieux encore très peu explorés. Ce projet vise à rechercher de nouvelles molécules dans les organismes marins notamment les éponges ou bactéries, afin d'évaluer leur potentiel thérapeutique.

Introduction et objectifs de mes travaux

Le travail présenté dans cette thèse est issu de mes 6 mois de stage de master réalisé au sein du projet PharmaSea. Au cours de ce projet, nous nous sommes focalisés sur une étape clé de cette bioprospection: la dérégulation.

Actuellement, la bioprospection définie comme étant la recherche de nouvelles molécules d'intérêt biologique parmi les substances naturelles repose essentiellement sur des méthodes de fractionnement bio guidé, c'est-à-dire l'isolement guidé par l'activité biologique. Les objectifs sont donc d'isoler les composés bioactifs pour usage direct ou bien de produire par hémi synthèse des composés dérivés de leurs structures. La bioprospection suit un cheminement allant de la collecte des échantillons, identification, extraction, fractionnement de l'échantillon menant à l'isolement de molécules pures. Cela mène ensuite à la détermination de structures potentiellement nouvelles et originales et la mesure de l'activité biologique des molécules isolées. Les extraits sont souvent des mélanges complexes de composés. L'extraction liquide-liquide par l'utilisation de solvants non miscibles va permettre de se débarrasser des composés non désirés tels que les composés trop hydrophiles ou trop hydrophobes (acides gras, sucres, sels). Afin de réduire leur complexité, les extraits vont être ensuite pré fractionnés par différentes méthodes telles que l'Extraction en Phase Solide (SPE) ou la chromatographie liquide (LC), en utilisant diverses combinaisons d'éluant. La séparation finale des fractions résultantes va être ensuite réalisée par Chromatographie, telle que la Chromatographie Liquide à Haute Performance (HPLC), un outil des plus utilisés. L'HPLC permet une séparation des composés selon leur polarité par leur distribution entre une phase stationnaire (support) et une phase mobile (mélange de solvants de différentes polarités), qui vont être élués de la colonne selon leurs

caractéristiques physico-chimiques. L'HPLC va pouvoir ensuite être couplée à une analyse par spectrométrie de masse (MS) qui permet l'ionisation des molécules de l'extrait et leur séparation en phase gazeuse selon leur rapport masse/charge (m/z). Elle va permettre de détecter et mesurer la masse de molécules pures issues de l'extrait, et ainsi, aider à la caractérisation de leur structure. L'élucidation structurale est ensuite poursuivie par analyse en Résonance Magnétique Nucléaire (RMN), qui permet d'identifier les fonctions, le pharmacophore et les substituants présents de la molécule. Malheureusement, fréquemment, ces étapes longues et fastidieuses aboutissent à des composés connus.

Dans le domaine des produits naturels marins, on recense environ 30000 composés déjà identifiés et répertoriés dans la base de données des produits naturels marins, MarinLit (15). Dans le but d'accélérer le processus de découverte de nouveaux composés, il est très utile de scanner les extraits pour détecter la présence de composés déjà connus ou peu intéressants avant l'isolement, pour ne se concentrer que sur des composés ayant de nouvelles structure ou activité biologique. Ce processus clé est appelé la déréplication.

Plusieurs stratégies de déréplication ont été développées, impliquant généralement l'utilisation de recherche chimio taxonomiques dans les bases de données couplées à une séparation analytique et spectroscopique. La plupart sont basées sur des combinaisons de techniques. La Chromatographie Liquide (LC) couplée à la Spectrométrie de Masse (MS) ou à l'analyse en détection Ultraviolet (UV) (LC-MS ou LC-UV) a été la technologie fondamentale utilisée pour la déréplication (16). La LC, et notamment l'HPLC, est une technique des plus polyvalentes permettant une séparation efficaces des composés, et nécessitant une préparation simple des extraits. Les possibilités de séparation se sont améliorées avec aujourd'hui un grand choix de phases stationnaires, de la rapidité, résolution et sensibilité, applicables a une grande variété de composés et adaptables à plusieurs détecteurs spectroscopiques. La spectrométrie de masse, sensible et précise, permet de caractériser chaque composé individuellement par mesure résolutive de leur masse et leur formule moléculaire (17). La combinaison des temps de rétention obtenus par LC et des données de masse par MS est une méthode sensible, robuste et un contenu pertinent en informations. Additionnée aux informations chimio taxonomiques, cette technique permet une recherche précise et efficace des bases de données en réduisant le nombre de compositions élémentaires prédites et donc, le nombre de candidats possibles pour une molécule. La détection UV, réalisée en général en sortie de colonne HPLC, va fournir des informations sur les chromophores présents dans les molécules, répertoriée également dans les bases de données. Les données UV sont généralement utilisées comme second critère après une recherche LC-MS,

ou en complément d'informations. Les spectres UV des chromophores peuvent être en effet un moyen de discriminer des composés de même composition élémentaire, permettant d'exclure ou confirmer un candidat du profilage (18). Les bases de données, par leur qualité et quantité de données disponibles, autorisent une comparaison directe des données d'un extrait (*e.g.*, SciFinder, AntiMarine, ChemSpider). La base de données pour les produits naturels marins, MarinLit, contient des données bibliographiques actualisées sur les organismes marins et plus de 21000 composés, avec des références provenant de 1200 journaux/livres.

Au Marine Biodiscovery centre (MBC), le schéma de déréplication utilisé en routine est basé sur une méthode manuelle combinant LC-MS (**Figure 1**) et recherche dans les bases de données. Pour confirmer les données issues de la déréplication, l'élucidation complète de la structure de la molécule est réalisée par analyse en spectrométrie de masse haute résolution (HRMS) et RMN 1D (1 dimension) et 2D (2 dimensions). Ces données structurales sont comparées à des bases de données pour confirmer la nouveauté ou l'exclusion d'un composé.

Cette méthode manuelle est cependant rendue délicate par le grand nombre de composés, la complexité des extraits et l'ampleur des jeux de données à traiter qui augmente le temps consacré aux analyses, et qui peut parfois être associé à un manque de données significatives (chimio taxonomie, identification...).

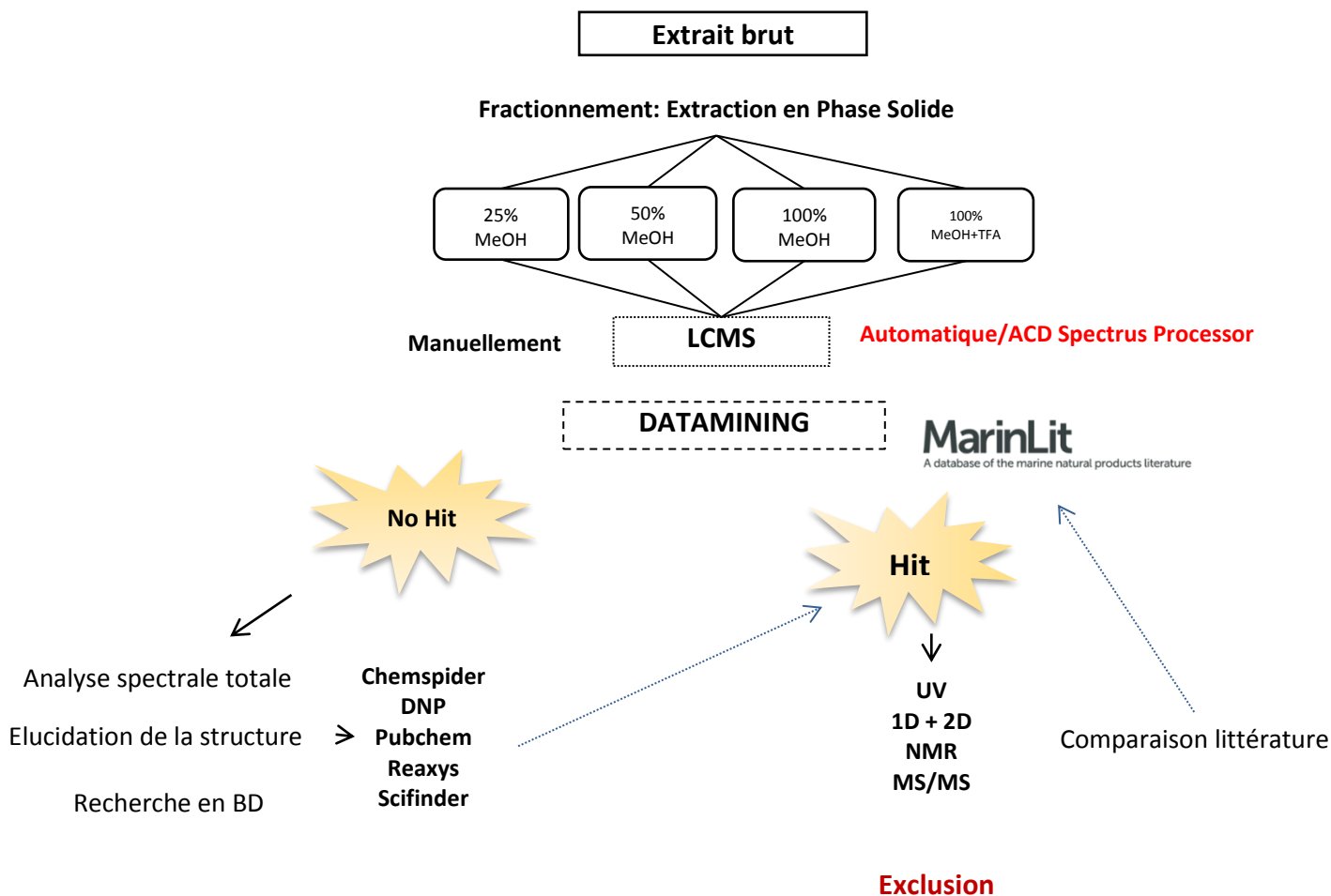


Figure 1. Processus de déréplication utilisé au Marine Biodiscovery Centre

Pour faire face à ces problèmes, de nombreuses stratégies sont élaborées, impliquant notamment des méthodes telles que la LCHRMS (Chromatographie Liquide couplée à la Spectrométrie de Masse Haute Résolution), ou la LC-MS/MS (Chromatographie Liquide couplée à la fragmentation de masse), qui permettent d'augmenter la précision des données de masse. Le couplage de la LC à la RMN est aussi expérimenté. La RMN permet de discriminer deux composés et de réduire l'ambiguïté liée à leur identification (16). Cependant le couplage LC-RMN reste encore limité et les bases de données RMN sont encore à enrichir.

Pour automatiser la déréplication, des systèmes de traitement de données semi ou totalement automatiques ont été développés. Nous pouvons ainsi citer le logiciel proposé par Wang *et al.* (19), appelé MIDAS (Metabolite Identification via DatabAse Searching) qui permet de comparer en ligne, automatiquement et à haut débit des spectres de masse et de fragmentation expérimentaux à ceux théoriques d'une base de données de métabolites.

Une méthode fut conçue dans le cadre du projet PharmaSea qui associe l'utilisation du logiciel IntelliXtract (IX), fournit par ACD/labs (Advanced Chemistry Development company) et la base de données MarinLit de la « Royal Society of Chemistry ». Le logiciel ACD/Labs IX extrait tous les composants chromatographiques issus des données LC-MS et assigne les ions moléculaire pour chaque composant en soustrayant les artefacts et bruits de fond. Il permet ainsi d'obtenir des données de LC-MS (temps de rétention et masse m/z) dé-complexifiés. Des macros et algorithmes insérés dans ce logiciel vont permettre la lecture et comparaisons des données LC-MS d'un extrait avec les données des composés répertoriés dans MarinLit. Cette stratégie résulte en l'obtention d'un fichier contenant les jeux de données de composés déjà connus, et un contenant ceux des composés non répertoriés antérieurement.

Dans le cadre de notre étude consacrée aux organismes marins, nous avons utilisé et investigué cette méthode pour aider à l'identification de composés déjà connus et prioriser des échantillons paraissant plus intéressants en terme de nouveauté et originalité. Elle devrait en effet permettre une automatisation et accélération de la méthode manuelle « classique » utilisée au MBC. Cette investigation est détaillée dans le travail de thèse présenté ultérieurement.

Notre objectif principal était donc de valider cette technique de déréplication en l'appliquant à nos échantillons et utiliser les données obtenues pour sélectionner ceux contenant plus de nouveauté et diversité. De manière indépendante, une analyse par RMN du proton a été réalisée pour une sélection plus discriminante des échantillons. Le but est ensuite d'isoler les composés nouveaux et de caractériser leur structure pour affirmer ou infirmer leur nouveauté.

Les organismes étudiés sont des tuniciers, invertébrés marins vivant seuls ou en colonie dans les mers du monde entier à toute profondeur. La variété fascinante de ces organismes laisse entrevoir une multiplicité de nouvelles possibilités en matière de découvertes pharmaceutiques. Ils se sont en effet révélés source de composés possédant une large gamme d'activités biologiques et des structures originales et intéressantes, pour certaines actuellement en études cliniques. Ces échantillons ont été récoltés aux îles Fidji en 2009.

Matériel et Méthodes

21 Echantillons de tuniciers provenant des îles Fidji ont été extraits et fractionnés. Chaque fraction a été analysée par LCMS et dérépliquée par la méthode PharmaSea pour fournir une liste de masses connue et une liste de masses inconnues pour chacune d'entre elle. Pour valider la méthode, des études de cas ont été classifiées et sélectionnés après analyse des données LC-MS, ainsi que des spectres RMN, séparément (**Figure 2**). Les deux extraits priorités ont été purifiés par (HPLC), et les molécules pures entièrement caractérisées structurellement par Spectrométrie de Masse Haute Résolution et RMN 1D et 2D. Elles ont été ensuite comparées à des bases de données pour vérifier leur éventuelle nouveauté.

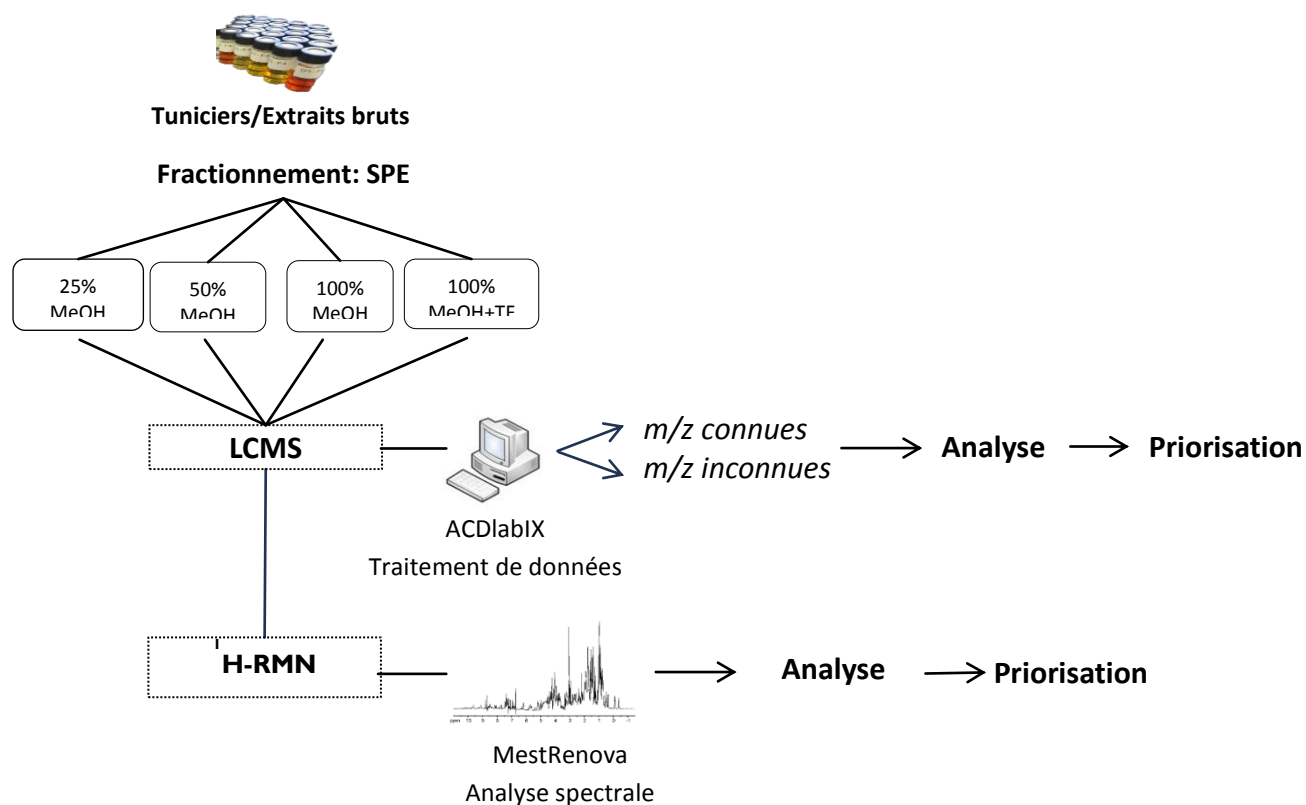


Figure 2. Méthode utilisée pour la validation de la méthode de déréplication PharmaSea

Résultats

Sur les 7 molécules isolées de la première étude de cas, un Tunicier de l'île de Tavarua, 5 apparaissent être nouvelles. Deux nouveaux composés, ont été isolés et entièrement caractérisés, se révélant être des dérivés de nucléosides non connus à ce jour et trois autres composés de masse nouvelle sont en cours de caractérisation. Un des 7 composés est le Drimenol, connu en milieu terrestre mais non répertorié dans Marinlit. Le 7^{ème} composé pur est la Psammapsyline, molécule connue pour être produite par une éponge. La deuxième étude de cas était un tunicier de l'île de Balolo. Deux composés de type éthers diphenyl polybromés ont été isolés: ce sont des molécules déjà connues pour être synthétisées par des bactéries marines et retrouvées dans des émissions anthropologiques.

Conclusion et perspectives

La méthode de déréplication testée au cours de ce stage a été conçue dans le but de hiérarchiser les échantillons à étudier en fonction de leur potentialité de fournir des molécules nouvelles et originales. Notre étude a permis de valider que cette méthode est fiable car nous avons en effet trouvé des molécules nouvelles dans les extraits privilégiés. Cette étude nous a permis de faire la lumière sur les avantages et inconvénients de chaque stratégie utilisée ainsi que la valeur ajoutée de leur combinaison.

La LCMS permet d'apporter des informations précises sur les formules chimiques des molécules, cependant, la variabilité de l'ionisation elle-même et de sa source peuvent entraîner de nombreux biais de mesure. La RMN est la méthode de choix fournissant d'importantes informations structurales. De plus, elle permet de réaliser une analyse quantitative des échantillons. Cependant, la sensibilité est moins bonne que la LCMS. Ceci nous amène à la conclusion que ces deux approches sont complémentaires facilitant le processus de déréplication, permettant d'accélérer l'isolement de nouvelles structures.

Cette étude a permis l'isolement de nouveaux composés structurellement intéressants, notamment des dérivés nucléosidiques ce qui nous replace dans le cadre pharmaceutique de recherche de nouvelles molécules bioactives. Ce type de molécules s'est avéré être prometteur en terme d'activité biologique, et nous prompte à poursuivre notre recherche par des tests antiviraux et anticancéreux sur ces nouveaux composés. Ces molécules vont ainsi être

répertoriées et permettre d'enrichir la collection de molécules déjà existantes dans les bases de données, notamment MarinLit.

Etudes complémentaires

En parallèle au travail résumé ci-dessus, un projet de fractionnement bio guidé fut également mis en place lors de ce stage, afin d'explorer une approche différente dans la bio prospection des substances naturelles. Son objectif était l'identification de nouveaux composés bioactifs (neuroactifs) à partir d'extraits de la bactérie marine *Leeuwenhoekiella sp.* Une première purification de l'extrait par extraction en phase solide fut réalisée à Aberdeen. Les fractions obtenues sont actuellement en cours de tests biologiques. Les résultats devraient nous permettre de sélectionner la fraction active qui sera purifiée afin d'obtenir des sous-fractions qui seront à leur tour envoyées aux essais biologiques. Nous espérons pouvoir ensuite isoler les composés purs responsables de l'activité biologique et élucider leurs structures.

Ci-dessous, nous présentons en anglais le détail de mes travaux réalisés au cours de mon stage de master.

Chapter 1 General introduction

1.1 A growing health crisis

Rising incidences of various cancers, and an ever-growing range of infections caused by microorganisms are the main issues of modern medicine, co evolving with the respective drug targeting resistances. Following on a new report based on data from 114 countries, the World Health Organization (WHO) has recognised Anti-Microbial Resistance (AMR) as a growing health crisis, calling it "an increasingly serious threat to global public health happening right now in every region of the world able to affect anyone, of any age, in any country". With the increasing number of microbes developing resistance to standard drugs, the previously trusted medical weapons are becoming almost useless. For instance, resistance to one of the most widely used antibiotic in urinary tract infections caused by *Escherichia coli*, fluoroquinolones, has spread in many parts of the world where this treatment is now ineffective in more than half of patients (1). Moreover, antibiotic resistance causes people sickness for longer time, increasing the costs of treatments and ever the number of deaths. As an example, every year, methicillin resistant *Staphylococcus aureus* is responsible for the deaths of about 25,000 people worldwide. Besides, the evolution of resistance against cancer targeted drugs represents a significant obstacle to the successful control of tumours since it abrogates the response to therapy. For certain drugs, a wide variety of resistance mechanisms have been reported (2). Once a cancer has developed resistance to one type of drug, it is more likely to develop resistance to other drugs, with various mechanisms, making treatment more difficult (20).

Given those facts, it is important to build up the best possible chemotherapeutic "armamentarium" and to use it first when the probability of resistance is lowest. Since 2003, there hasn't been a completely new antibiotic registered. During the last 30 years, combinatorial chemistry has been used as a discovery source in approximately 70% of cases. However, disappointing results were obtained, as to date, we can only find a small number of *de novo* chemical entity reported in the public domain resulting from this method and approved for drug use (3). This is the antitumour multikinase inhibitor from Bayer known as sorafenib (**Figure 3**). It was approved by the Food and Drug Administration (FDA) in 2005 for treatment of renal cell carcinoma and in 2007 for hepatocellular carcinoma. Hence, interest and inspiration have been therefore renewed in the discovery of new agents from Nature,

considered as a rich source of novel active secondary metabolites that may serve as the leads and scaffolds to elaborate efficacious drugs, for many indications.

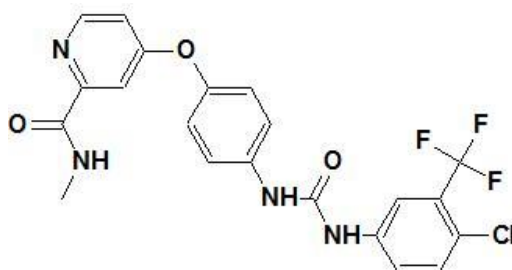


Figure 3. The chemical structure of Sorafenib

1.2 Nature as a source of new medicine

Because of their coevolution with biological targets, Natural Products (NP) are good starting points for drug discovery. Their chemical novelty is shown to be higher than any other source: 40% of the chemical scaffolds in the Dictionary of Natural Products (DNP, published database) are absent from synthetic chemistry. In 2010, approximately half of the new drugs released (New Chemical Entities) so far have been discovered or designed based on the chemical structures elucidated from natural products and half of the 20 best-selling medicines are related to NP (21). NP and derivatives have been deemed as 60% of new anticancer drugs and almost 75% of all new antibacterial agents. Within the seven antitumor agents approved in 2010, one, romidepsin, a histone deacetylase inhibitor, is a microbial natural product without any modification and still produced by fermentation. Then, four are derived from NP (vinflunine, cabazitaxel, and the totally synthetic halichondrin B-derived eribulin, tubulin-interactive agents) (7). Moreover, NP that are biologically active in assays are generally small molecules with drug-like properties. This means that they can be absorbed and metabolized by the body. Thus, the development costs of producing orally active medicines are likely to be much lower than with biotechnological products or with most compounds produced to date from combinatorial chemistry.

The overwhelming contribution of NP to the expansion of the drug resource is then clearly evident, but much of Nature's "treasure trove of small molecules" remains to be explored. However, due to decreasing rates of new compound discovery and to frequent "rediscovery"

of known agents, the research has now gradually drifted to marine environment and deep-sea derived microorganisms.

1.3 Marine biodiversity

The abundance and diversity of natural products from the marine environment with interesting biological activities lead to opportunities for discovering new drug leads. Due to the area (71% of the surface of the Earth) they represent and the ecosystem services they provide, Oceans are fundamental to our planet. They represent 1 370 million km³, that is 361 million km² for 3 800 m depth. Marine area is thus 3 times bigger than terrestrial one.

Greater biodiversity can be observed in marine environment compared to terrestrial one. The reason generally advanced is that marine ecosystem is more constant therefore species have evolved very little within the same phylum. Furthermore, the large variety of biotopes on the ground has resulted in a consequent adaptation of species belonging to the same order, family or genus. Recent sampling expeditions by the Antarctic benthic deep-sea biodiversity project (ANDEEP, 2007) and first Census of Marine life (2000-2010) (12) confirmed an extremely high level of biodiversity across a wide range of taxa. They estimate the number of known marine species to approximately 250,000 and extrapolated their findings to at least a million of marine species and tens or even hundreds of millions of microbial species. According to William Fencal (22), within the 34 phyla of life, 17 occur on land while 32 occur in the sea, all forms of life first appeared in the oceans being still there, from Archaea to mammals. On estimated 1.8 million species, only 13% are described nowadays, showing that marine biodiversity still need to be investigated.

1.4 Marine chemo diversity

Many marine invertebrates are sessile and soft bodied, and therefore must rely on primordial chemical messengers (natural products), arisen from their evolutionary history, to communicate, attract food, deter predators and; in having no means of transport nor, in general, any way to hide, most of them have genuine and highly evolved chemical weapons to defend themselves against various attacks. Hence, seaweeds, sponges, ascidians produce a diverse array of novel secondary metabolites, called “ecomones”. They are the vector of chemical communication between organisms and belong to all chemical family including terpenes, alkaloids and polyphenolics, and exhibiting different and sometimes opposing pharmacological activities and original structures. To date, more than 20000 secondary metabolites have been isolated from marine organisms as sponges, ascidians, algae, cnidarians and bryozoans (13).

Because they originate from prokaryotes and eukaryotes, some of these compounds differ fundamentally from terrestrial secondary metabolites in that they are characterized by the presence of chemical elements or structures that are seldom observed in the latter. Indeed, in addition to the classical elements which are carbon, oxygen, hydrogen and nitrogen, marine organisms will incorporate sulphur and halogens presents in seawater such as chlorine and bromine to produce halogenated derivatives. This is also true for less abundant elements such as phosphorus, iodine, silicon and arsenic. By consequence, marine organisms demonstrate a unique chemistry that lead to unique and original structures. For all of the chemical families represented among marine organisms, it is possible to identify totally original structures of compounds, without any known equivalent to date in terrestrial ones. Two examples showing the large diversity of NP from marine microorganisms are presented in **Figure 4** (23).

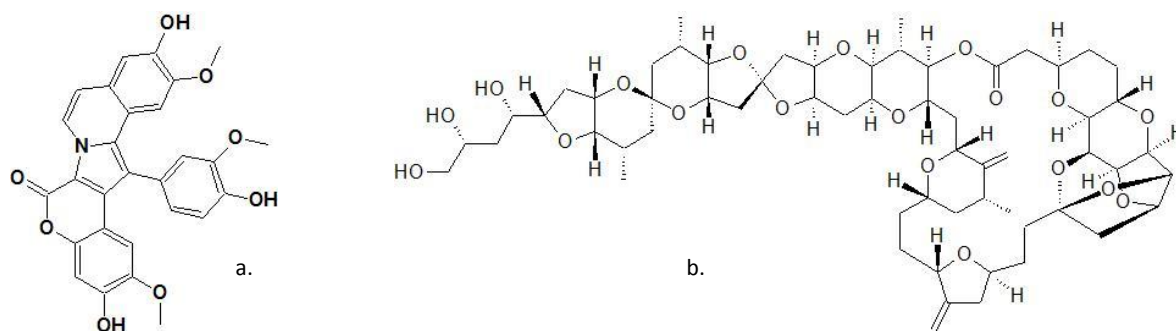


Figure 4. a. Lamellarin D (alkaloid) b. Halichondrin B (terpene)

Another example of very complex nitrogenous secondary metabolites isolated from certain classes of invertebrates such as nudibranchs is the jorumycine (**Figure 5**) (24).

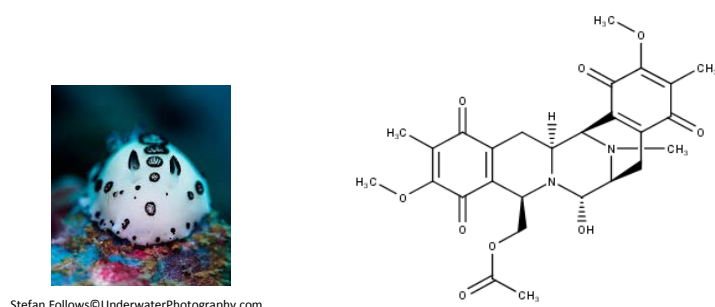


Figure 5. Jorumycine isolated from a Pacific nudibranch *Jorunna funebris*

Pharmacodiversity is logically as important as chemodiversity. Since the 1970s, it has become a self-evident fact that every possible biological and pharmacological activity can be found in marine organisms (**Table 1**) (14, 15).

Table 1. Pharmacodiversity of Marine organisms

Activity	Organism ^b Spo	Cni	RAI	BAI	Asc	Mol	Ech	GAI	Cya	Bac	Bry	Fun	Din	Ann	Hem	Sel	ArB	Nem
Anthelmintic	+	-	+	+	-	-	+	-	-	-	-	-	-	-	-	+	+	-
Antibacterial, Antibiotic	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
Antifungal	+	+	+	+	+	+	+	+	+	-	+	-	+	-	+	-	-	-
Antihistamine	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Antihypertensive	+	+	+	+	-	+	+	-	-	-	-	+	-	-	-	-	-	-
Anti-inflammatory	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-
Antimalarial	+	+	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-
Antimutagen	+	-	+	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-
Antiosteoporosis	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Antipsychotic (Alzheimer's disease)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Antispasmodic	+	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-
Antitumor, Anticancer	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	-	-
Antiviral	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-
Immunostimulant	+	+	-	+	+	-	-	-	-	-	+	-	-	-	-	-	-	-
Immunosuppressive	+	+	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-
Neurostimulant	+	+	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-
Neurosuppressive	+	+	+	-	+	+	+	-	-	+	-	-	-	-	-	-	-	-
Vasoconstrictor	+	+	+	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-
Vasodilator	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-
"Bioactivity index" ^c	90	79	74	68	58	58	53	37	32	21	21	21	16	11	11	11	5	5

A large panel of chemical families can be found as shown **Figure 6** (25).

Released in the water, those products are rapidly diluted and therefore must act quickly and be potent at low concentration (14). Because of the peculiarity of the marine environment, marine bioactive compounds usually have higher biological activity in a wide range (26). In the National Cancer Institute preclinical cytotoxicity screening, approximately 1% of the tested marine samples showed anti-tumor potentials *versus* 0.1% of the tested terrestrial samples (27). Furthermore, because they lack of an acquired memory type immune system, marine invertebrates rely entirely on an innate immune mechanism including cellular and humoral responses. To fight against attack by invading microorganisms that can be up to 10⁶ bacteria/mL and 10⁹ viruses/mL of seawater, they have developed an effective and robust use of this innate immune system including antimicrobial peptides production. Some show great specificity for prokaryotes alongside low toxicity for eukaryotic cells (28). Research on marine NP is then expected to increase again in the years to come, providing new and improved therapeutics for human illnesses.

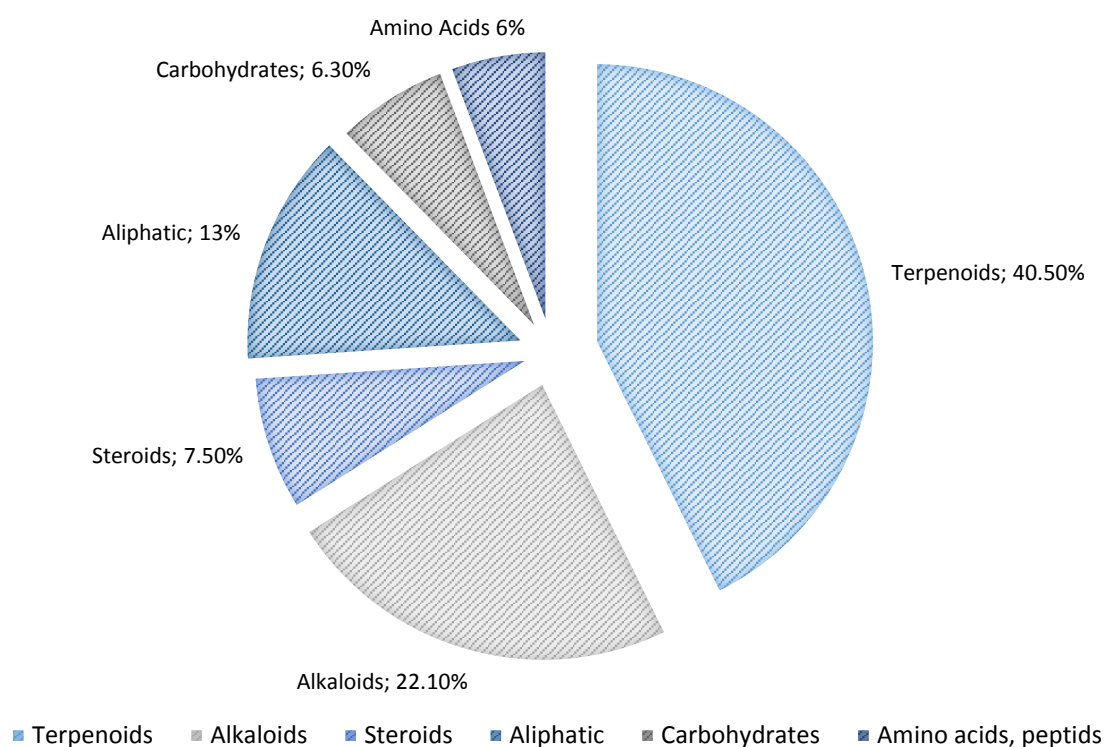


Figure 6. Distribution of new natural products (n=5286) isolated from 2000 to 2009 according to their chemical families

As biologically active substances often contain at least one nitrogen atom, it is interesting to look at the repartition of nitrogenous and non-nitrogenous secondary metabolites in marine world. It is thus essential to find out which organism produce it and the relative amounts of substances to know where we can focus our researches. It is clearly apparent that chances to find new nitrogenous metabolites are higher among tunicates, bryozoans, sponges than among cnidarians, echinoderms or molluscs (**Figure 7**) (14). The non-nitrogenous metabolites are usually derived from the mevalonate pathway (fatty acids, polyketides and derivatives), the two known terpenes biosynthetic pathway (mevalonate and desoxyxylulose) and the shikimate pathway. It is also known the existence of more complex pathways which are un-encountered among terrestrial organisms. The nitrogenous compounds are usually very complex and the mechanisms of their biosynthesis is sometimes far from being solved. They often derived from the coupling of two diverse metabolic pathways (29).

As marine habitat is considered as the first source of life on earth, evolution connects humans to marine organisms. Indeed, life appeared in the sea 3.8 billion years ago, while the first terrestrial species is only about 400 million years. The species diversification process began in the oceans long before that of the terrestrial environment. Our signalling pathways probably evolved from it and respond to their primordial 'exocrine' system, whence the profound effects on human cells and tissues exerted by those primordial messengers.

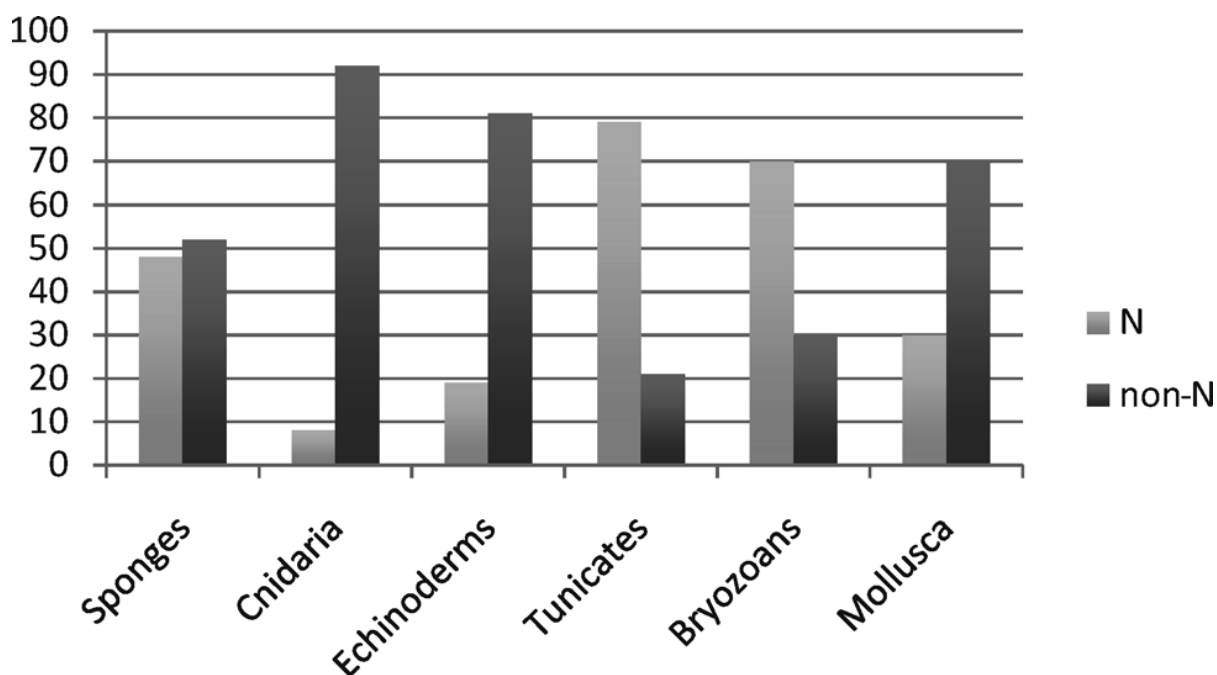


Figure 7. Distribution of nitrogenous and non-nitrogenous compounds among sea species

Though they have been used in Chinese medicine since ancient times, very few traditional pharmacopoeia and remedies referring to marine organisms is remaining. The Chinese *Pen Ts'ao*, oldest known pharmacopoeia (~2800 BC), comprises one entire chapter showing that Brown algae were used to treat stomach ulcers thanks to alginates, used as a gastric bandage, and also in goitres illness due to iodine deficiency. *Digenea simplex*, a red algae spread in Indian ocean and Mediterranean and red sea, was used for more than thousand years against nematodes in some parts of Asia (30). Considering that people from the western coast of Kyushu in Japan used the seaweed as a vermifuge, Murakami *et al.* first isolated the active substance known as kainic acid (31). The preparation of the extract was commercially available, marketed under the name Helminal (The merck Index, 1968) (Figure 8).



©Pharmakoteka

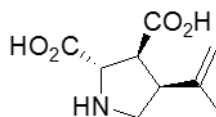


Figure 8. Helminal and kainic acid

Benkendorff reported that marine mollusks appear in traditional South African invertebrate medicine markets and are also used in Indian and Pacific Island remedies (32). But, the systematic study of marine therapeutic potent only started in the second half of the 20th century and immediately prompts a great deal of interest through the discovery of rare and active molecules (33). In 1948, the Professor Giuseppe Brotzu discovered a new generation of antibiotics, the cephalosporins, by cultivating a microscopic fungus from lagoon sediment, *Cephalosporium acremonium* (34). Almost at the same time, the American chemist Werner Bergmann isolated several molecules from a Caribbean sponge, *Cryptotethya crypta*. Those novel chemicals called “arabinosides”, spongothymidine and spongouridine showed strong anticancer and antiviral properties. Two synthetic analogs led to the development of Ara-C used against leukemia and Ara-A for treating viral infections (35).

At the present time, around 15 marine natural products are in various phases of clinical development, mainly in the oncology area. Cytarabine (Cytosar-U®) was first approved by the FDA in 1969 as a chemotherapeutic agent.

Those discoveries gave to many scientist hopes to find into the sea new models of molecules and new pharmaco-diversity. Little progress was made till the 70s: about only 1/10 of the molecules discovered in 50 years have resulted in marketed drugs. This can be attributed to the absent of documentation and pharmacopoeia for the marine environment, except for some local exceptions that we mentioned earlier. Moreover, the considerable chemo and biodiversity Lastly, according to MarinLit database, for over a more than 25-year period (1980 to 2009) a

new molecule has been extracted from a sponge every two days, increasing from 416 to 1548 entities observed is hard to study and inventory (15). The task is really vast and deep sea diving is often necessary to search for refine indices. Availability of the biomass and difficulties of harvesting complicate it as well. Nevertheless, thanks to the various systems now available and their improvement, the future possibilities are most significant. Indeed, modern techniques of harvesting, analysis and screening have reached a higher level of sophistication. Informatics chemotaxonomic system allows now to compare the pertinence of organisms. Despite a usual high financial need, results are finally there: a number of active molecules (candidates) have reached the stage of clinical trial in humans.

Few bioactive natural molecule from molluscs are undergoing pharmaceutical drug development : Ziconotide (Prialt®), from the venom of the cone sail, *Conus magnus* was the first marine drug to be approved for clinical use as a chronic pain treatment in 2004 followed by Trabectedin (Yondelis®) in 2007, both commercialised by European companies (14). The bryostatins (**Figure 9**), discovered in 1982 by Pettit et al. (36) in two species of bryozoans *Bugua neritina* and *Amathia convolute*. Despite a small amount of product available, it has now successfully entered in phase II clinical trial for breast cancer. Examples of marine natural products currently undergoing clinical trial are shown **Table 2** (14).

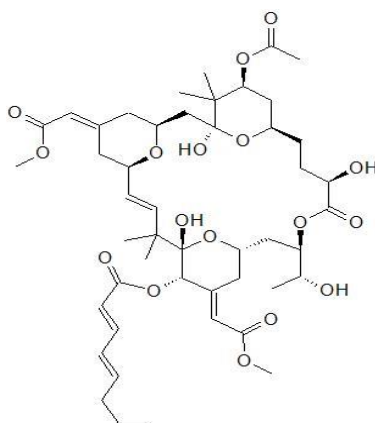


Figure 9. Structure of Bryostatin 1

Table 2. Examples of marine natural products currently under clinical trials

Denomination	Clinical phase trial (Laboratory)	Origin	Activity	Class
Ziconotide (Prialt®)	Phase III (Neurex)	Conus magnus (Mollusk)	Analgesic	Peptide
Dehydrodidemnin B (Aplidine®)	Phase II (PharmaMar)	<i>Trididemnum solidum</i> (Ascidian)	Anticancer	Cyclic depsipeptide
Bryostatin-1	Phase II (OrthoBiotech)	<i>Bugula neritina</i> (Bryozoan)	Anticancer	Macrolide
Dolastatin 10	Phase II (NCI)	<i>Dolabella auricularia</i> (mollusc)	Anticancer	Peptide
IPL-576	Phase II (Aventis)	<i>Petrosia contignata</i> (sponge)	Anti-inflammatory	Steroid
Manoalide	Phase II (Allergan)	<i>Luffariella variabilis</i> (sponge)	Anti-psoriasis	Terpene
DMXBA	Phase I/II (Taiho Pharm. Co.)	<i>Paranemertes peregrina</i>	Alzheimer's disease	Alkaloid
Isohomohalichondrin B	Phase I (PharmaMar)	<i>Lissodendoryx</i> sp. (sponge)	Anticancer	Macrolide polyether

1.5 Exploring the hidden potential of oceans for new drugs

Keeping this in mind, some European researchers have launched a collaborative project named PharmaSea to look for novel antibiotics focusing on the bio discovery and the development of potent new drug leads from marine organisms and microbes. Its primary goal is to collect samples all over the world, from the oceans, home to some of the hottest, deepest and coldest places on the planet. Choosing those extreme environments aims to tap novel diversity not seen before and to yield new species of marine microorganisms. Sediment samples are collected from ocean trenches - deep, narrow valleys in the sea floor, down to almost 11km depth thanks to fishing vessels to drop sampling equipment on a reel of cables to the trench bed. The organisms will be characterized and analysed by metagenomics to generate extract libraries and identify biosynthetic pathways. Those libraries will be subjected to high throughput screening and high content bioassay to find eventual lead compounds further analysed, structurally characterized and integrated in database platforms. Finally, evaluation of the potential of up scaling and optimisation will be undertaken to aim at future pre-clinical development (**Figure 10**).

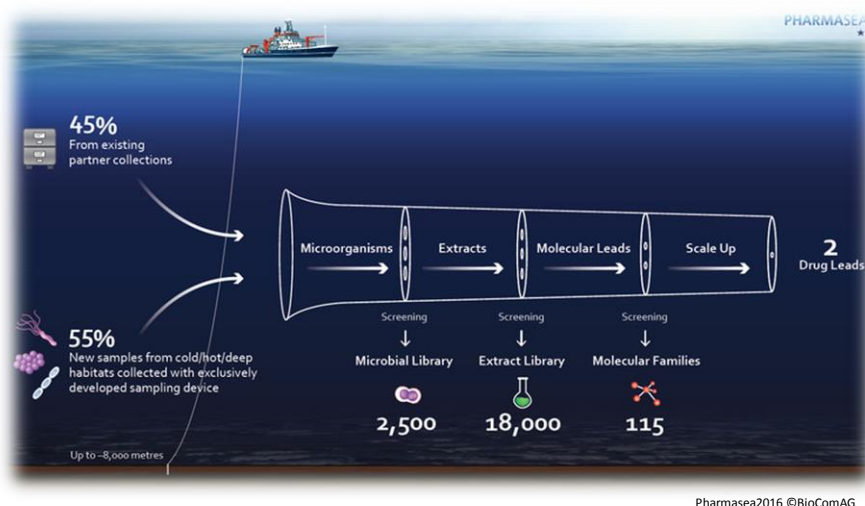


Figure 10. A schematic presentation of the PharmaSea project

1.6 Bioprospecting pipeline for marine drug discovery

Two schools of thought are currently established for the discovery of marine NP: “isolate and then test” (chemically guided) versus “test and then isolate” (biologically guided) and both have a historical proof of success. A fusion approach is thus commonly employed. One of the main aims of using the chemically driven approach is to find novel chemical structures of marine natural products, without any application in mind. Hence, extracts are screened for unusual and interesting patterns and then, biological properties of purified compounds are investigated. If one of the compounds presents any interesting activity, pharmacological research will continue, if not, the new structure can be published. The bio directed strategy where biological activity is hunt above all uses bioassay-guided extraction of crude extracts. Pharmacological tests are performed on all the extracts and the process is pursued only with those showing positive results. It has been the preferred method by modern marine natural product researchers and pharmaceutical industry (37). In this thesis we will approach each methods of bioprospecting.

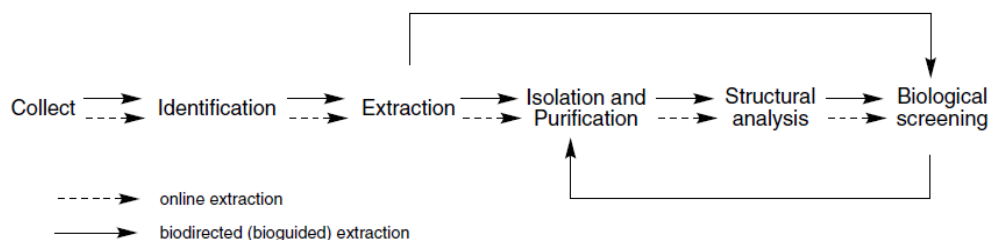


Figure 11. Bioprospecting pipeline for Marine Natural Products discovery

The bioprospecting pipeline will generally follow the same flow: partitioning, prefractionation, dereplication, isolation and structure elucidation (**Figure 11**). Crude samples are often mixtures of various substances not ideal for pharmaceutical objectives (oral administration...). Solvent partitioning separating solutes between two immiscible liquid phases may thus be needed to remove undesirable material as primary metabolites, highly hydrophobic or lipophilic compounds (fatty acids, sugars, salts...). The extract will be then pre-purified into several fractions to help reducing the complexity of the extract, enhance its rate and accelerates the separation and identification processes for further bioassays. Prefractionation can be processed thanks to various methods as Solid Phase Extraction (SPE), chromatography, using various compositions of eluent. Then final NP separation isolation can be achieved through High Pressure Liquid Chromatography (HPLC), one of the tool most commonly used since it has reproducibility, chromatic resolution, various choice of polar phases and embedded groups, and suitability with mass spectrometry.

1.7 The dereplication process

One of the main issues of any drug discovery approach is to avoid the isolation of previously characterized compounds. Many marine natural products (about 30000 entries in the MarinLit database (15)), are already identified secondary metabolites or fractions containing known compounds. Crude extracts must then be scanned for the presence of those compounds before initiation of their isolation, to focus on the discovery of structurally or biologically novel ones. This process is commonly known as “dereplication” and generally involves the use of analytical separation and spectroscopic methods combined with database and chemotaxonomic searching. High Resolution Mass Spectrometry is the core technology, used in combination with Liquid Chromatography (LC-MS/HRMS) and UV/vis analysis. This combination is fast,

robust, sensitive and provides high content in information. The high separation efficiency of HPLC combine with high resolved mass data permits characterisation of individual components in extracts. These properties make this strategy ideal for dereplication of extracts: elemental composition, retention times, and taxonomical data can be search in internal or commercial databases. The database (DB) provides straightforward comparison of data thanks to the huge amount of available compounds data (e.g., SciFinder, AntiMarine, ChemSpider). The Marine Natural Products DB MarinLit- contains up to date bibliographic data on marine organisms with the number of references from 1,200 journals/books and data for ~21,000 compounds. To aid the screening of the large number of NP currently available in DB, semi or fully automated systems covering data processing and compound searching has now been developed, such as the software proposed by Wang *et al.* (19). This vital dereplication process will be discuss in the second chapter of this thesis.

At the Marine Biodiscovery Centre (MBC), the structure of each compound will then be fully characterised using modern spectroscopic methods and manual interpretation using the usual workflow (**Figure 12**). HPLC will be coupled with a High Resolution Electrospray Ionization Mass Spectrometer (HRESIMS) in positive and negative ionization mode to measure unambiguously the accurate mass and determine molecular formula. Extended MS-MS analysis will be used to predict the structure of unknown compounds based on careful inspection of the obtained fragmentation patterns. Analysis of 1D NMR after HPLC purification will help to obtain the full structures including stereochemistry present in the elute and elucidate an accurate structure. Various 2D-NMR spectra such as Heteronuclear Single Quantum Coherence HSQC, Heteronuclear Multiple Bond Correlation HMBC, Homonuclear Correlation Spectroscopy ^1H - ^1H COSY, will be combined with UV and MS data. Structures will be confirmed by Computer Assisted Structure Elucidation (CASE) method.

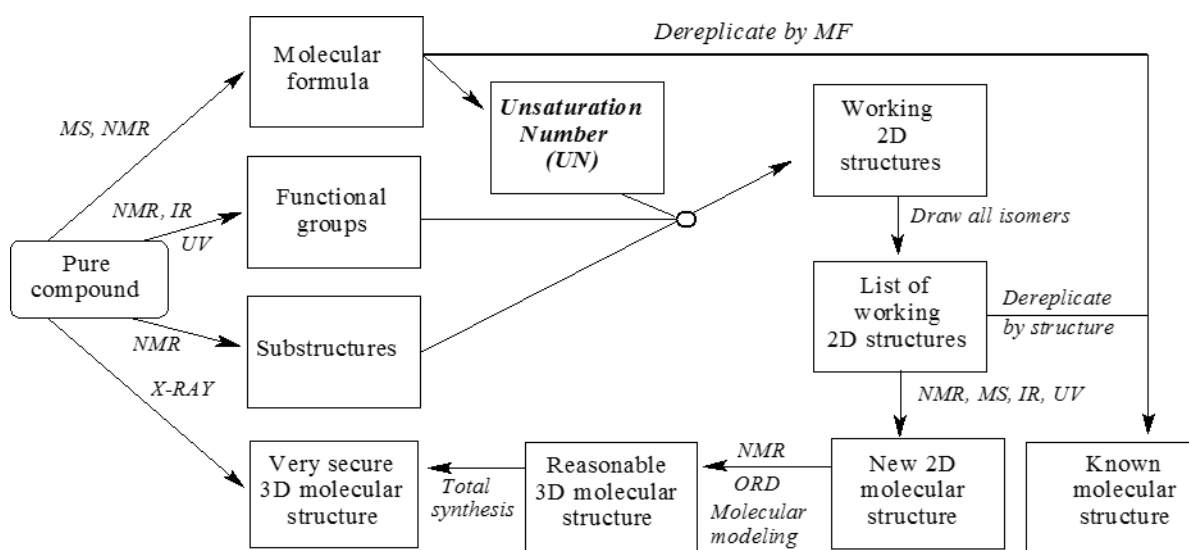


Figure 12. Structure elucidation workflow used at MBC

1.8 Aim of the thesis

The first aim of this thesis is to validate the PharmaSea compound dereplication method. The dereplication method used by the Pharmasea project was devised to help identify known compounds so that resources can be focused on samples that will lead to the isolation and identification of new ones.

1. Application of the PharmaSea dereplication process
2. Prioritization of samples by LCMS
3. Prioritization of samples by $^1\text{H-NMR}$
4. Isolation of new compounds
5. Elucidation of chemical structures
6. Acquiring skills in processing LC-MS and LCMS/MS data for the identification of compounds.
7. Acquiring skills in use of 1D and 2D NMR techniques for the characterization of compounds.

Chapter 2 Validation of the PharmaSea dereplication method

2.1 Introduction: challenges of the dereplication process

In the chase for new compounds and original structures we principally use at the MBC an aggressive manual dereplication method combining LC-HRMS-MS/MS and NMR based structure elucidation to compare it with the MarinLit database (**Figure 13**).

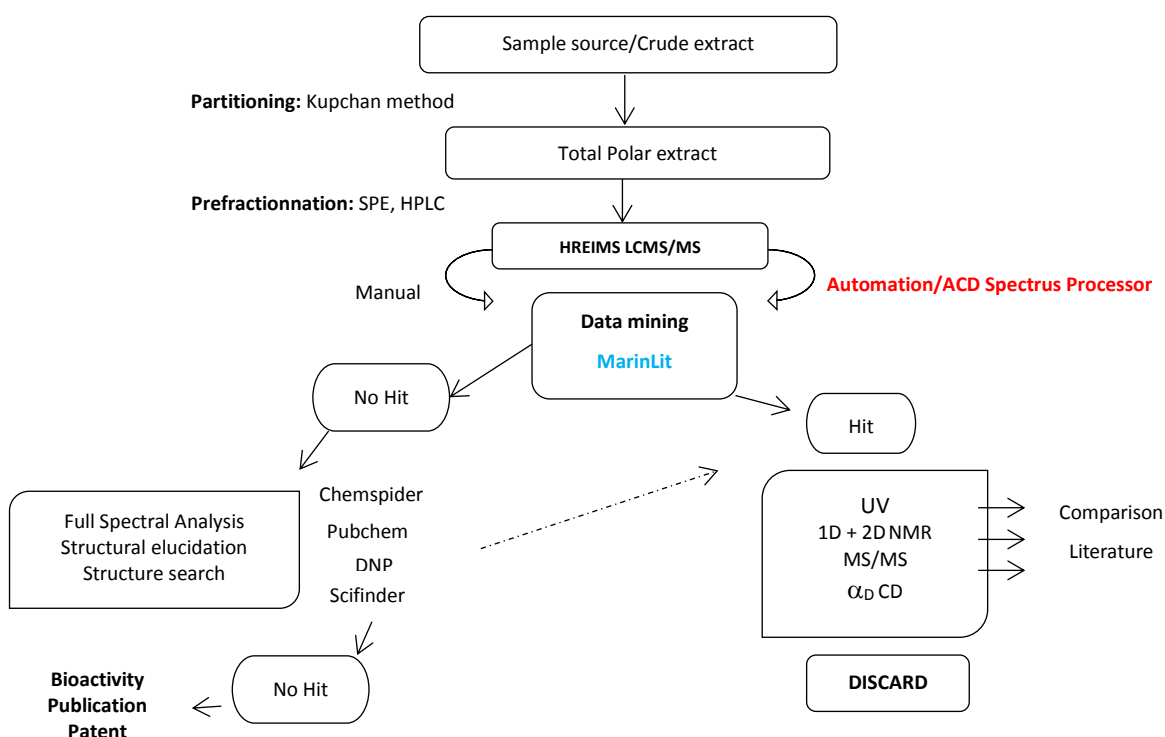


Figure 13. Dereplication workflow followed at the MBC

Unfortunately these processes have a few challenges. It is not suitable for large data sets, labour intensive and slow due to possible long run times. A lack of confirmatory data and confused compound identification can be experienced due to certain ion adducts or interfering minor ions (16). Furthermore, some marine organisms having uncertain taxonomy, the current database like MarinLit does not contain all marine natural compounds making dereplication difficult and time consuming.

The need for an easier and faster way to process dereplication, a new method was elaborated as part as the PharmaSea project. This method devised in conjunction with Advanced Chemistry Development (ACD) Labs and the Royal Society of Chemistry (RSC) uses a combination of both the ACDLabs IntelliXtract (IX) software with the MarinLit database from RSC. This software extracts all the chromatographic components from the LC-HRESIMS data sets and assigns the molecular ion for each one (**Figure 14**). LCMS data were inputted into the software and processed using algorithms to yield LC-HRESIMS data sets containing deconvoluted masses, retention times and ion intensities. It also subtracts masses and information found in a media and a blank used as a reference (**Figure 15**). Then, macros written within the ACDLabs IX were able to read and match masses in the MarinLit database. This gave rise to a set of data that was not found in the MarinLit database. These data sets contain masses of compounds (m/z) and their LC retention times that are potentially new and form the basis of the sample prioritization process (**Table 3**).

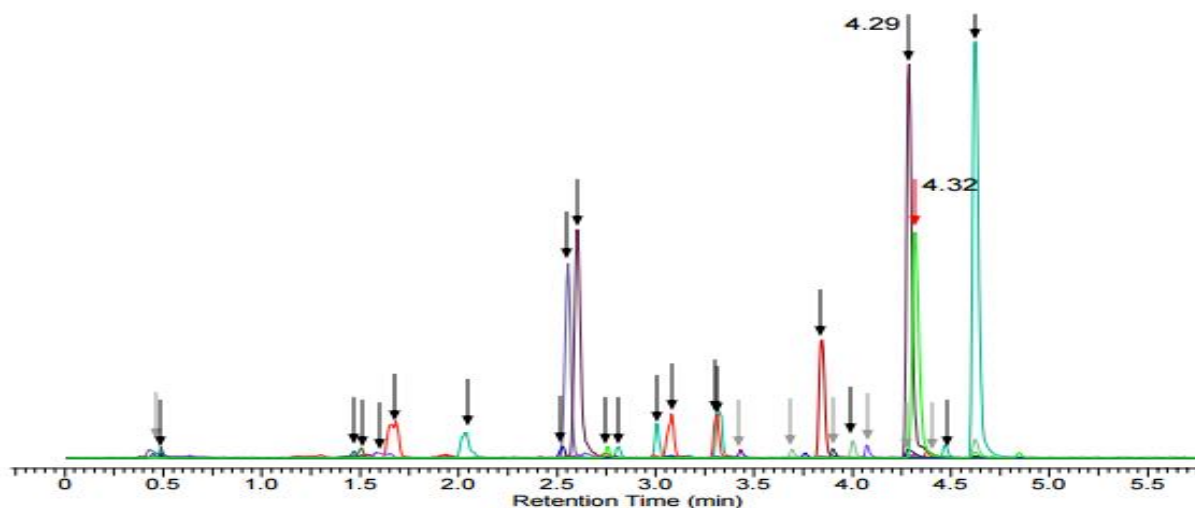


Figure 14. Ion chromatogram and ions extraction

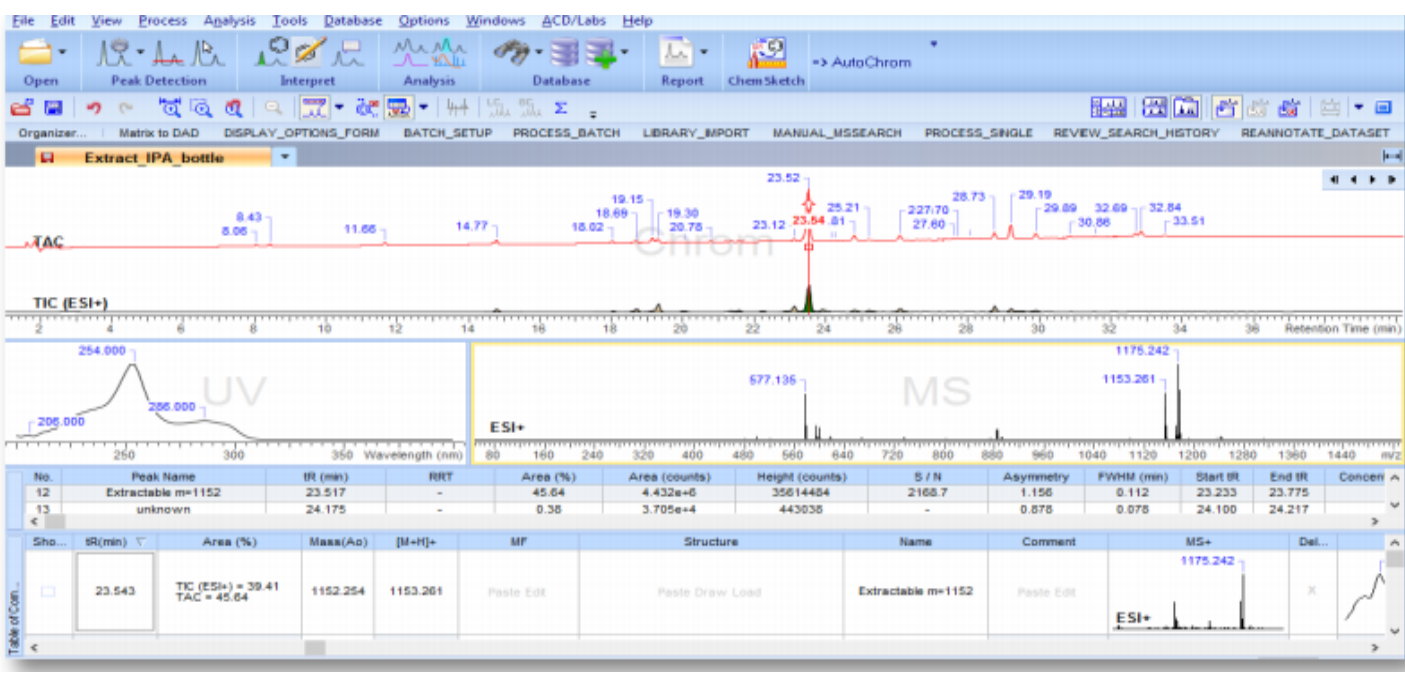
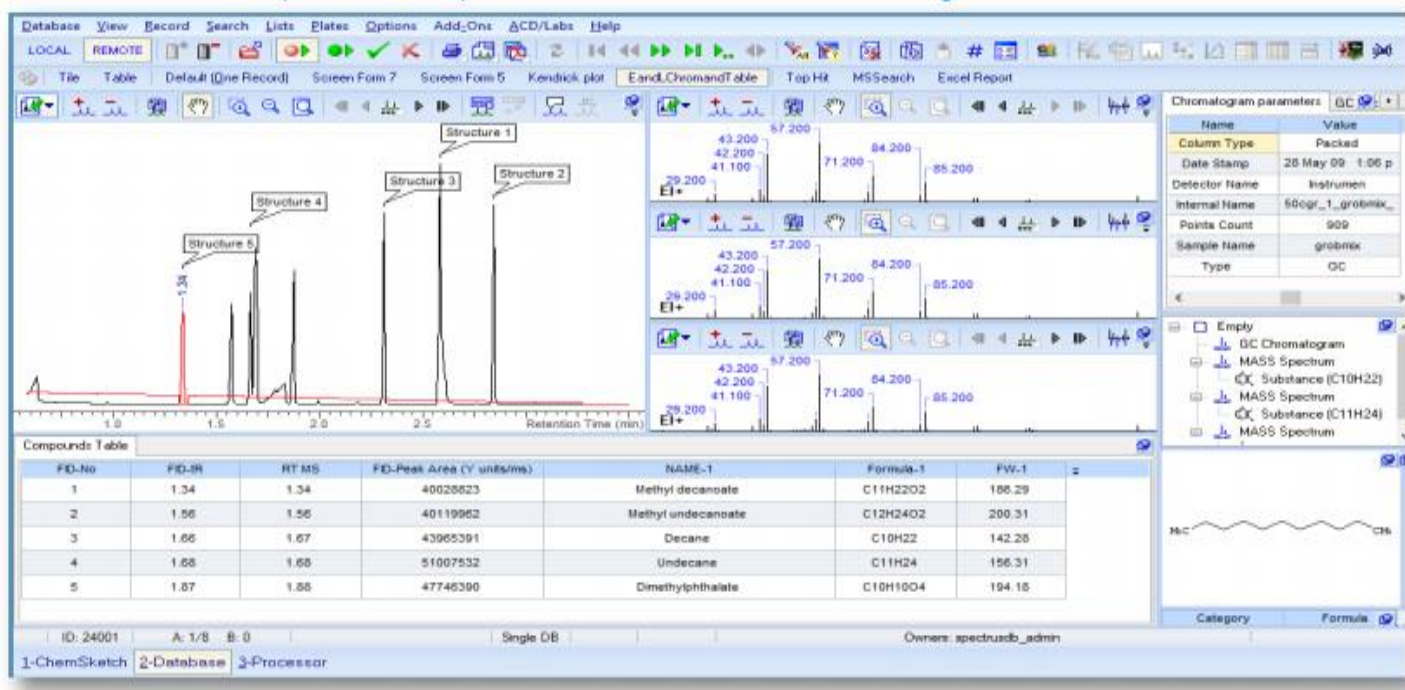


Figure 15. Presentation of IntelliXtract software and data input

Table 3. Excel table of new masses extracted from IntelliXtract analysis

Index	Notation	Adjusted PTM	PTM	RT
1	[M+H] ⁺ U ⁺ u	828.41669	829.42451	10.619258
2	[M+H] ⁺ U ⁺ u	408.153912	409.161732	11.969934
3	[M+H] ⁺ U ⁺ u	505.351807	506.359627	12.784425
4	[M+H] ⁺ U ⁺ u	880.496949	881.504769	12.920259
5	[M+H] ⁺ U ⁺ u	664.360546	665.368366	13.087925
6	[M+Na] ⁺ U ⁺ u	664.359532	687.349302	13.087925
7	[M+H] ⁺ U ⁺ u	895.463462	896.471282	14.612092
8	[M+H] ⁺ U ⁺ u	896.478139	897.485959	14.741259
9	[M+H] ⁺ U ⁺ u	896.478139	897.485959	15.475092
10	[M+H] ⁺ U ⁺ u	818.420952	819.428772	15.606758
11	[M+H] ⁺ U ⁺ u	902.325592	903.333412	15.901758
12	[M+H] ⁺ U ⁺ u	768.385774	769.393594	16.602924
13	[M+H] ⁺ U ⁺ u	738.425378	739.433198	17.107258
14	[M+H] ⁺ U ⁺ u	978.521588	979.529408	17.410091
15	[M+H] ⁺ U ⁺ u	754.400427	755.408247	17.643425
16	[M+H] ⁺ U ⁺ u	949.564321	950.572141	17.778767
17	[M+H] ⁺ U ⁺ u	980.58595	981.59377	17.97677
18	[M+H] ⁺ U ⁺ u	981.59673	982.60455	18.011425
19	[M+H] ⁺ U ⁺ u	732.424034	733.431854	18.070267
20	[M+Na] ⁺ U ⁺ u	732.419321	755.409091	18.097258
21	[M+H] ⁺ U ⁺ u	479.218195	480.226015	18.500425
22	[M+H] ⁺ U ⁺ u	1007.543307	1008.551127	18.602425
23	[M+H] ⁺ U ⁺ u	1007.543307	1008.551127	18.874758
24	[M+H] ⁺ U ⁺ u	980.585954	981.593774	19.077759
25	[M+H] ⁺ U ⁺ u	798.425901	799.433721	19.34326
26	[M+H] ⁺ U ⁺ u	1028.540479	1029.548299	19.750099
27	[M+Na] ⁺ U ⁺ u	964.54211	987.53188	20.076437
28	[M+H] ⁺ U ⁺ u	964.54452	965.55234	20.117758
29	[M+H] ⁺ U ⁺ u	980.515066	981.522886	20.159094
30	[M+H] ⁺ U ⁺ u	942.557916	943.565736	20.193768
31	[M+Na] ⁺ U ⁺ u	942.560192	965.549962	20.228434

2.2 Method for the validation of the dereplication strategy

To validate the method, a collection of 21 marine macro-organisms collected from the Fiji Islands in 2009 was used. Samples were extracted with methanol followed by dichloromethane, dried, and then fractionated by C₁₈ cartridge solid phase extraction (SPE) using successively 100% water, a mixture of 25% methanol in water, a mixture of 50% methanol in water, 100% methanol and a mixture of 100% methanol with 0.05% TFA as solvents to afford four fractions. Each of the fractions was dried under nitrogen and aliquots of each extract submitted for LCMS analysis (**Figure 16**).

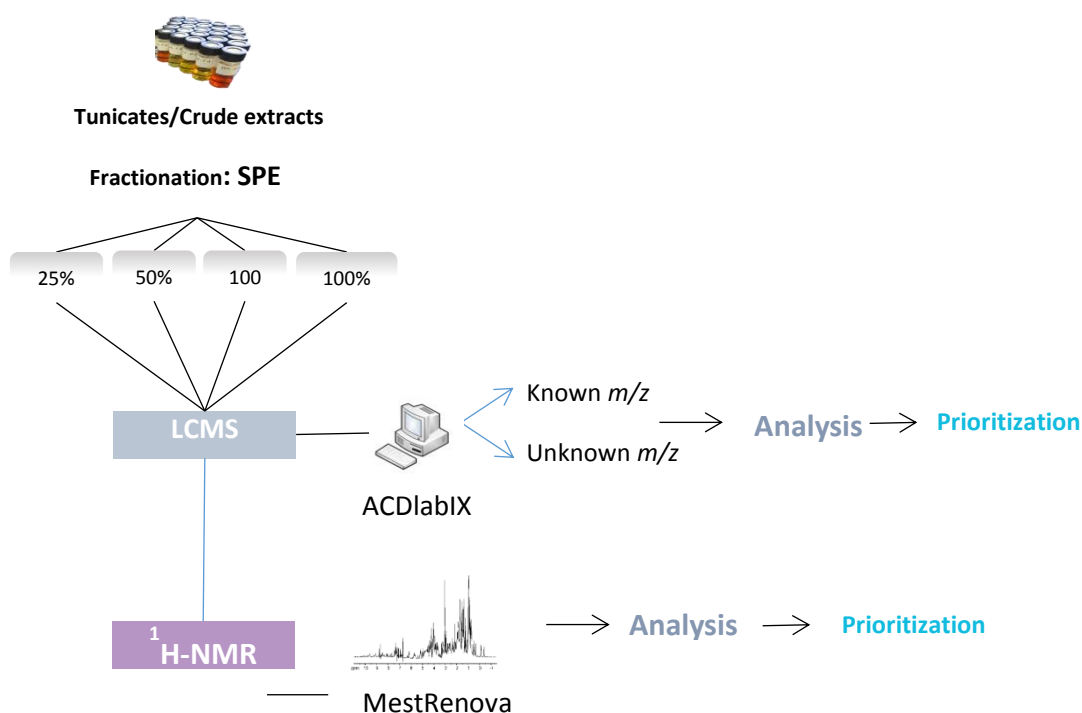


Figure 16. Method used for the validation of the dereplication method

LCMS analysis was carried out using a Linear Trap Quadrupole (LTQ) Orbitrap system (ThermoScientific, Waltham, MA, USA) coupled to a 1290 Infinity HPLC system (Agilent, Santa Clara, CA, USA). The following conditions were used: C₁₈ Reverse Phase (RP), column C₁₈ HL ACE 250x10mm 5 μ m separation using 0.1% formic acid in water and 0.1% formic acid in acetonitrile (CAN) gradient, UltraViolet (UV) (200-400nm) and MS positive resolution

(30000), MS/MS resolution (7500), capillary voltage 24 V, capillary temperature 350°C, gas flow rate 20 arbitrary units, spray voltage 4.50 kV, mass range 100-2000 amu. Each fraction was also measured by nuclear magnetic resonance (NMR) using an ID progidy cryoprobe in methanol-*d*4 as solvent, at 64 number of scans.

2.3 Results and discussion

2.3.1 Prioritization by LCMS

The total number of unknown masses (after dereplication) in each SPE fraction of each sample was plotted. The resulting chart allows the comparison of fractions in the same sample, additionally, it permits to compare each samples, as shown in **Figure 17**. This helped to identify samples and fractions that are rich in new masses or potentially new compounds. Assuming similar ionisation of compounds in the MS, in each fraction the results show that some samples are richer in new masses than others. As an example, we can see that the 100% methanol fractions of Lau1 and Tavarva 2 exhibits a higher amount of new compounds compared to the same fraction in Buago 2 or Eden 6. In addition, most new masses can be found in general in either 100% methanol-water or 100% methanol and Trifluoroacetic acid (TFA) extracted fractions. We can see this tendency in samples like Lau 1, Pin 5, Wall 1, Tavarva 2.

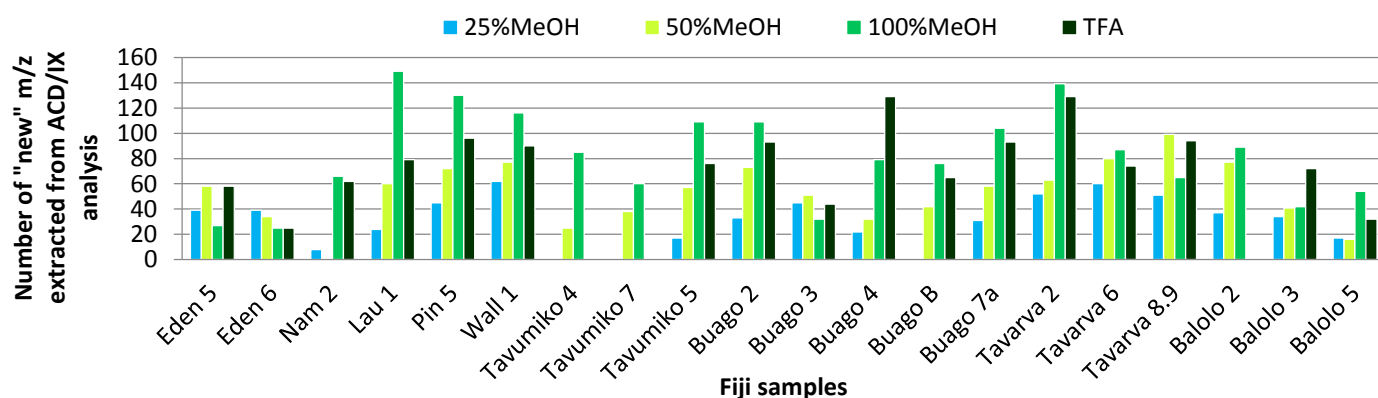


Figure 17. Number of new masses extracted by IX in each sample

Figure 18 shows a plot of the percentage of new masses out of the total masses identified by the processing software, assessed as novelty. The results show insignificant differences between each sample suggesting that each sample used in this work has a similar chance of yielding new compounds, except for some extracts as Eden 6.

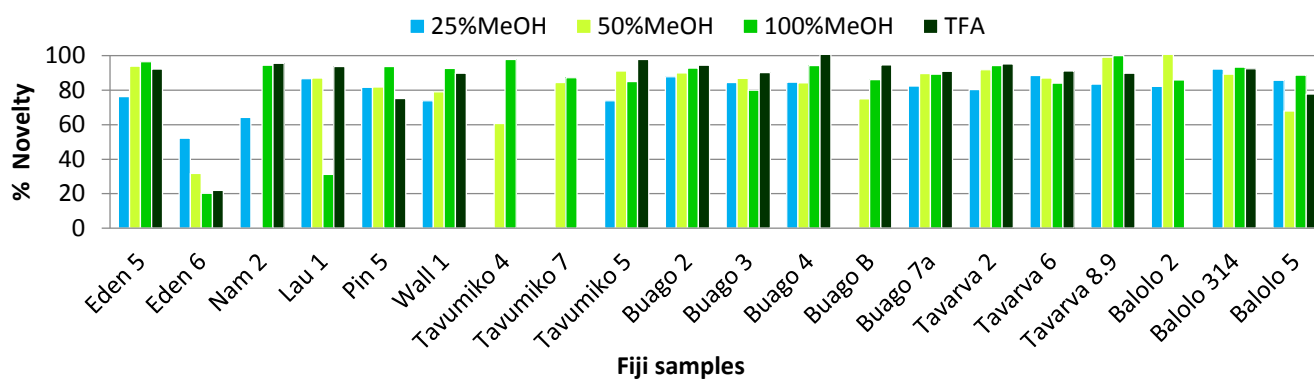


Figure 18. Percentage of new masses (m/z) compared to the total number of masses per extracts

To have an overview on the diversity and the complexity of each extract, the masses m/z from LCMS were then plotted against the retention time (T_R). The results for six of the samples are shown in **Figure 19** and **Figure 20**. The results show the spread of the new compounds as they are eluted by the reversed-phase column (Liquid Chromatography) providing information about structural diversity and polarity. The number of masses around a given polarity scale is important as it reflects their potential drug-like characteristics. Indeed, the likelihood that a molecule will have useful therapeutic activity (sometimes called ‘drug-likeness’) is predicted from its structural properties. A medicinal drug should have a suitable balance of solubility in polar and non-polar solvents. It should be sufficiently soluble in water so that it can be carried around the body in the bloodstream; but also in non-polar solvents so that it can pass through cell membranes (phospholipid bilayers) into cells (38). A medium polarity is thus generally targeted in those kind of research and having a first overview of polarity can be pertinent for the prioritization of extracts. The extent of different T_R provides an overview of the diversity, as for example two isomers may have the same mass but different T_R , due to their atoms and conformations. The complexity may be reflected by the clusterization of various masses in the same T_R area, revealing sometimes similar structures, which T_R is affected by slight conformation changes, e.g. Pin5.

The results in general show the expected correlation of compounds with percentage methanol used for fractionation. However, there are some samples like Tavarua 2 showing indications of compounds that do not correlate well to this trend suggesting a higher level of structural complexity.

Based on the number of new masses per sample, the top 8 ranked samples were: Tavarua 2, Tavarua 8.9, Lau1, Buago 4, Pin5, Wall1, Tavumiko 2 and Buago 2. Tavarua 2 also shows a higher level of structural complexity in addition to the presence of a higher number of masses with medium polarity (SPE 50% and 100% MeOH).

2.3.2 Prioritization by NMR

All NMR data were processed using the MestReNova software (39) and analysed visually for the presence or absence of interesting characteristics. Indeed, NMR spectra allows to assess chemical shifts (nature of the chemical environment in which a particular nucleus is located), spin-spin coupling (number and nature of nearby nuclei: connectivity information) and peak intensity (concentration of protons), thus provides information about compound features as well as the compound richness of the sample. Clues about compounds of interest such as key substructures and structural class can be obtained from a preliminary interpretation of a ^1H NMR spectrum details, e.g. chemical shifts, integrals and coupling patterns, protons attached to heteroatoms (OH, NH groups). Classes of compounds such as peptides, terpenoides, can be easily recognised as well as specific moieties: multiple bonds acetals, aromatic systems. NMR will also help to distinguish the structure of isomers, or two compounds having the same mass.

The NMR spectra of the four fractions of Tavarua 2 (**Figure 21**) show that each fraction contains different compounds and structural diversity. Each fraction was then analysed and ranked based on NMR profiles, signal intensity, number of compounds and possible fragment types. Sample classification of the top eight ranked samples are shown in **Table 4**. The proton NMR spectra of some of the top eight ranked samples are shown in the supplementary data (**Figure S2**).

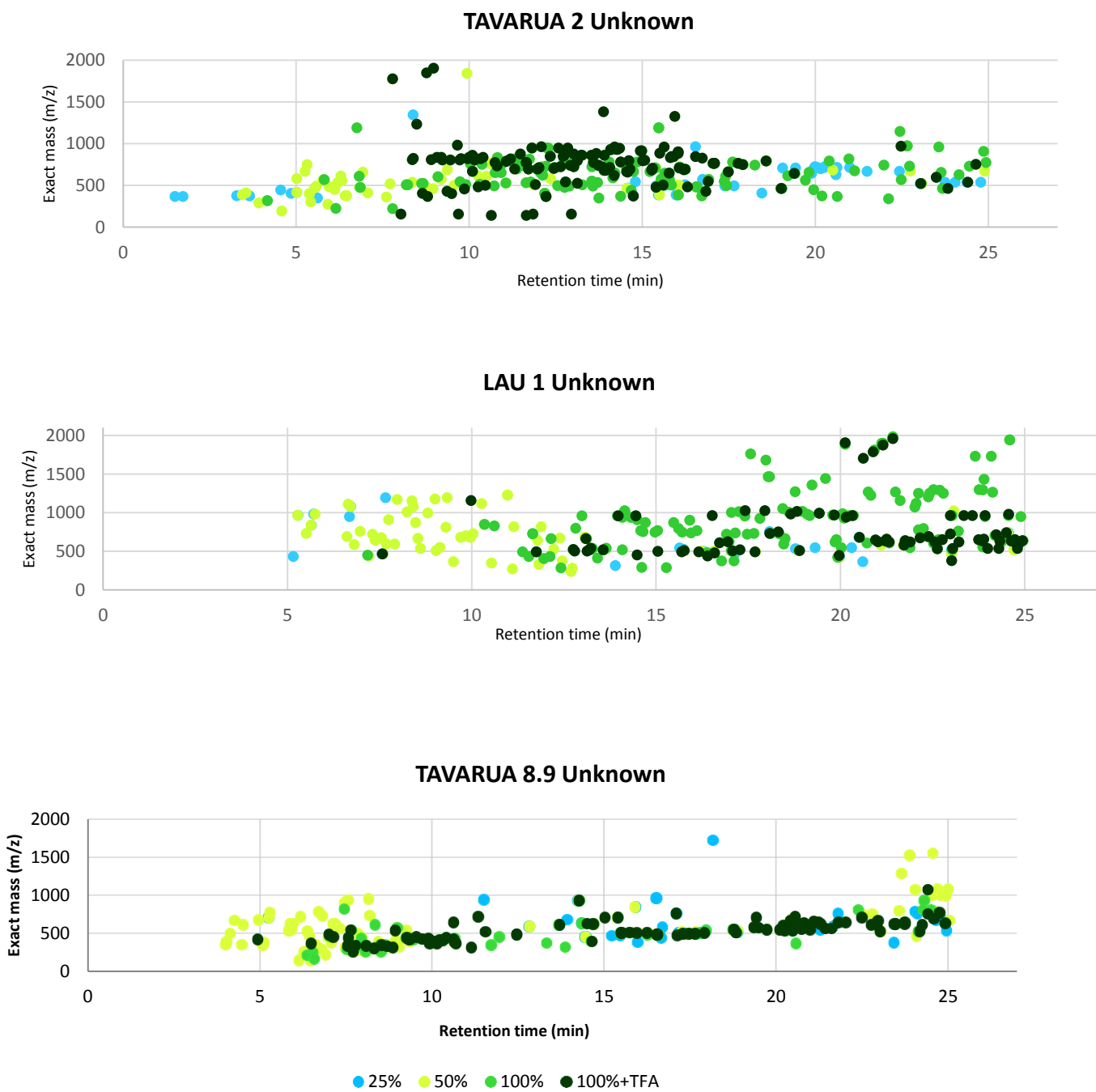


Figure 19. Exact mass (m/z) compared to retention time (min), for each SPE fraction of Fiji samples

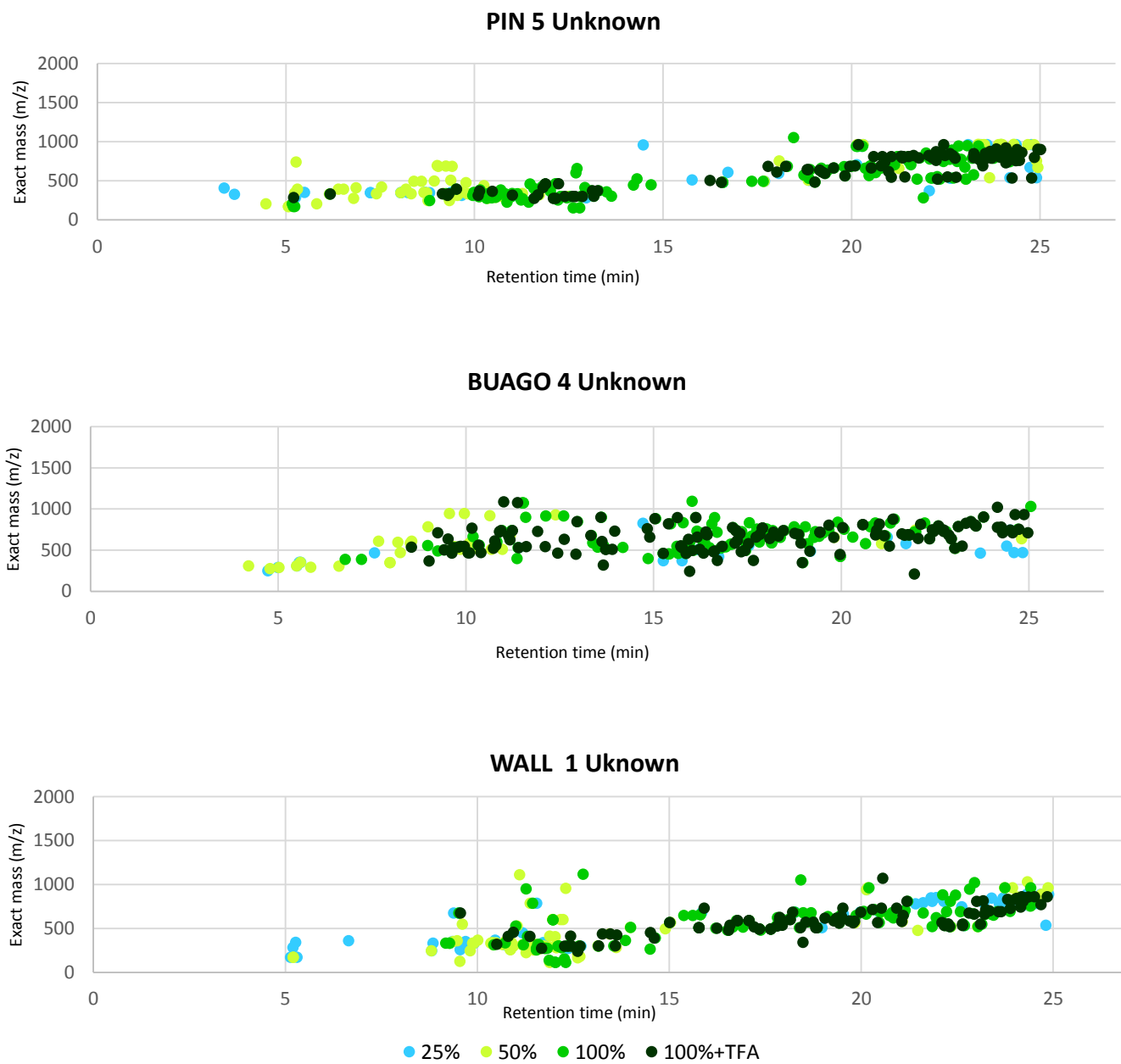


Figure 20. Exact mass (m/z) compared to retention time (min), for each SPE fraction of Fiji samples

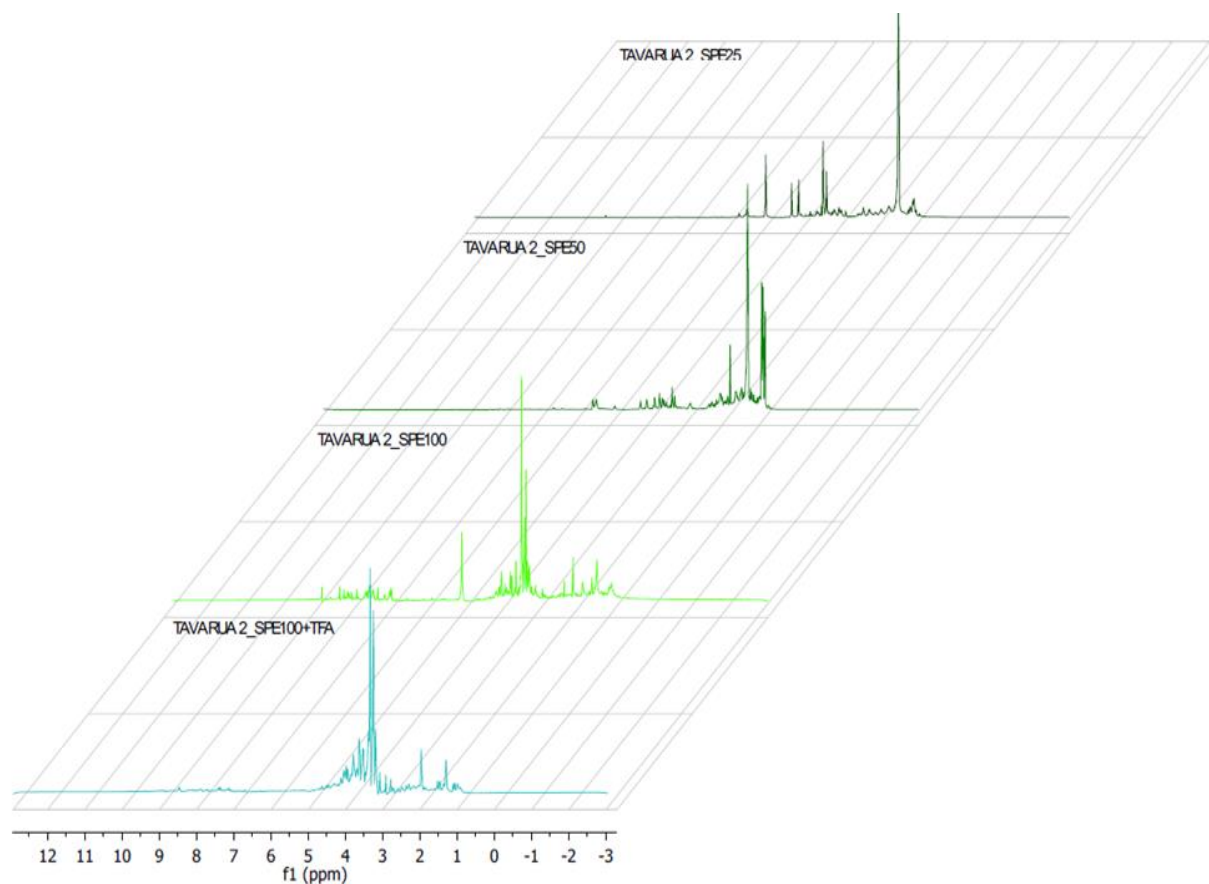


Figure 21. ^1H NMR spectra of the four fractions of Tavarua 2 after fractionation by C18-SPE

The top eight ranked samples were prioritized and classified based on NMR characteristics (**Table 4**), stronger NMR signal intensity and higher compound content. The presence of various chemical shifts and proton resonance area revealed their molecular diversity and their richness in compound substructures suggested their potential complexity. As an example, Tavarua 2 SPE50 % spectra exhibited strong signal intensity, extent proton resonance area from 0 to 9 ppm, and possible interesting structural moiety as benzene ring, or alkaloids resonances (**Figure 21**), thus was ranked as number one.

Table 4. ¹H NMR based classification of Fiji samples

Rank	Sample	SPE fraction	Nb of compounds	Signal intensity	6-9 ppm	5-6 ppm	3-5 ppm	2-3 ppm	0-2 ppm	Possible structure components
1	TAVARUA 2	25% MeOH	4	Strong	Yes	No	Yes	Yes	Yes	Aromatics, Aldehyde, N-group
		50% MeOH	5	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, Aldehyde, N-group
		100% MeOH	4	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, Alkene, N-group, O-group
		100% MeOH+TFA	3	Strong	No	Yes	Yes	Yes	Yes	Lipids, Alkaloid, O-groups
2	TAVARUA 8.9	25% MeOH	2	Weak	Yes	No	Yes	Yes	Yes	N-groups, O-groups
		50% MeOH	3	Medium	Yes	No	Yes	Yes	Yes	Aromatics,
		100% MeOH	5	Strong	Yes	No	Yes	Yes	Yes	Aromatics, N-group, O-groups
		100% MeOH+TFA	4	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, N-group, O-groups
3	PIN 5	25% MeOH	4	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, N-group, O-groups, lipids
		50% MeOH	4	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, N-groups, O-groups
		100% MeOH	3	Strong	Yes	Yes	Yes	Yes	Yes	Aromatic, N-groups, O-groups, lipids
		100% MeOH+TFA	3	Strong	No	Yes	Yes	Yes	Yes	Alkene, O-groups, lipids
4	EDEN 4	25% MeOH	2	Medium	Yes	Yes	Yes	Yes	Yes	Aromatic, N-groups, O-groups
		50% MeOH	2	Weak	No	No	Yes	No	Yes	?
		100% MeOH	3	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, N-groups, O-groups, Alkene
		100% MeOH+TFA	3	Strong	Yes	Yes	Yes	Yes	Yes	Aromatics, O-groups, N-group
5	WALL 1	25% MeOH	3	Medium	Yes	Yes	Yes	Yes	Yes	Lipids, aromatics, O-groups
		50% MeOH	3	Strong	Yes	Yes	Yes	Yes	Yes	Alkene, Aromatics, O-groups,
		100% MeOH	3	Strong	No	Yes	Yes	Yes	Yes	Alkene, O-groups
		100% MeOH+TFA	2	Medium	No	Yes	Yes	Yes	Yes	Alkene, O-groups
6	NAM 1	25% MeOH	2	Medium	No	No	Yes	Yes	Yes	N-groups, O-groups,
		50% MeOH	1	Weak	Yes	No	Yes	Yes	Yes	Aromatics, O-group, N-groups,
		100% MeOH	3	Medium	No	Yes	Yes	Yes	Yes	O-groups, Alkene
		100% MeOH+TFA	3	Medium	No	Yes	Yes	Yes	Yes	O-groups, Alkene
7	BALOLO 3	25% MeOH	2	Medium	No	No	Yes	Yes	Yes	?
		50% MeOH	1	Weak	No	No	Yes	Yes	Yes	N-group, O-groups
		100% MeOH	2	Strong	Yes	No	Yes	No	Yes	Aromatics, O-groups
		100% MeOH+TFA	2	Medium	Yes	No	Yes	Yes	Yes	Aromatics, O-groups
8	LAU 1	25% MeOH	2	Weak	No	Yes	Yes	No	Yes	N-group, O-groups, Alkene
		50% MeOH	2	Weak	No	Yes	Yes	Yes	Yes	O-groups, Alkene
		100% MeOH	4	Strong	Yes	Yes	Yes	Yes	Yes	Alkenes, N-groups, O-groups, lipids
		100% MeOH+TFA	2	Medium	No	No	Yes	Yes	Yes	N-group, O-groups

2.4 Conclusion

The results showed that the sample Tavarua 2 was ranked first followed by Tavarua 8.9 in either method. Three other samples were in the top 5 showing the complementary effects of both methods when used together. The main advantage of using LCMS is its high sensitive enabling measurements of data at very low concentrations. However, since the technique is based on the chemical ionisation process, compounds with very high ionisation potential can dominate the spectrum. This does not necessarily mean that they are present in significant quantities in the sample. It has been shown that access to 1D ^1H NMR data NMR can be very complementary and has been used in this case to provide additional data at the initial steps of dereplication. In being a less sensitive method than MS, it helps to identify samples containing compounds that can be isolated in good yields for further NMR structural analysis. It adds a quick discriminative approach of dereplication as just counting number and type of features chemical shift is enough to reduce the DB matching without the need to interpret them.

2.5 Case Study 1: Isolation of compounds from an unknown ascidian (Tavarua 2) from Fiji

2.5.1 Introduction

Tunicates (Urochordata) are among the most evolved invertebrates of marine environment, present in all the world's seas, at all depths, solitary or as a colony. They are characterized by the possession of a thick tunic composed of cellulose ("Tunicin") surrounding their barrel-shaped body, protecting it from predators and helping the fixation of their substrates (40). Chemical studies of Tunicates have mainly focused on the class Ascidiaceae since they furnish a potent source of active and structurally interesting and original compounds. According to Marinlit database, over a period of 40 years, about 1080 compounds have been isolated from ascidians and 50 new structures elucidated annually. Those molecules contain a lot of nitrogenous compounds: cyclic peptides and alkaloids and several one contains elements specific of seawater – sulphur and bromine. Most of them possess potent bioactivities and some are even currently involved in preclinical trials. Among them, Didemnidae has been the most chemically prolific ascidian family, responsible for at least 35% of known ascidian compounds (41). Didemnidae colonies can fuse and divide really fast, sometimes making difficult to collect sufficient material for drug discovery. But, it is interesting to note that if often genetically identical, metabolites biosynthetic pathway can be slightly different and so, chemical composition can vary, creating a combinatorial library of molecules, with unique mechanisms of action against diverse targets. Some are potent to ~100 pM against cancer-relevant cell lines (42). An example is the cyclic octapeptide, patellamide D (**Figure 22**), isolated from the marine tunicate *Lissoclinum patella*, acting as a selective antagonist of multidrug resistance in human leukemic cell line (41). The Didemnins, potent antiviral and antitumor cyclic peptides (**Figure 23**), were first isolated from the Caribbean ascidian *Trididemnum solidum*, which hosts the cyanobacterial symbiont *Synechocystis trididemni*. Didemnin B continues to be a main focus in clinical investigation, being the first marine natural product currently in clinical trials as an anti-cancer agent (43). Recent comprehensive studies that propose identified specific symbiotic bacteria living in ascidians and responsible for producing complex "ascidian" molecules aim at protecting the host. Indeed, recent publications find a gene cluster responsible for Patellamide production in the prochloron cyanobacteria and demonstrated its activity (44). Another example are Clavanins, α -helical antimicrobial peptide from the hemocytes of the ascidian *Styela clavela* that showed broadly activity against gram-positive and –negative bacteria, including MRSA.

and a flow of 2mL/min. HPLC chromatograms (*Figure S3*) extracted from 200 to 400 nm presented 12 different peaks separated, dried and submitted to NMR. One of the major peaks was submitted to HRESIMS and full structure elucidation to yield compound **1** (1.2 mg) (*Figure 24*).

The 100% SPE fraction was purified by a reversed phase C₁₈ HPLC using a mixture of methanol, water and TFA (0.005%) as an eluent with a gradient from 50-100% MeOH over 25 minutes and a flow of 2mL/min. HPLC chromatograms (*Figure S13*) extracted from 200 to 400 nm presented 14 different peaks separated, dried and submitted to NMR. One of the major peaks was submitted to HRESIMS and full structure elucidation to yield compound **2** (9.0 mg).

2.5.3 Results

The molecular formula of compound **1** was established as (C₂₆H₂₄N₃O₉) (522.1508 [M+H]⁺ Δ 0.47 ppm) by HRESIMS and suggested 17 degrees of unsaturation (*Figure S6; S7*). The correct structure was determined by analysing ¹H NMR (*Figure S9*), edited HSQC (*Figure S10*), HMBC (*Figure S12*) and COSY NMR spectra (*Figure S11*). Dereplication of this compound based on 1D, 2D NMR and LC-MS data in MarinLit indicated that it is an unknown compound. The mass fragmentation of the molecular ion displayed 3 fragments corresponding to C₂₁H₁₈NO₇ (396.108 [M+H]⁺), C₁₄H₁₂NO₄ (258.08 [M+H]⁺), C₉H₆NO₂ (160.04 [M+H]⁺) (*Figure S8*). In the ¹H and ¹³C 1D NMR spectra, proton signals at δ_H 7.41 (H-13) and 1.44 (CH₃-14) and carbon resonances at δ_C 150.88 (C-16), 164.74 (C-15), 110.67 (C-14), 135.43 (C-13), and 10.55 (CH₃-14) evidenced the presence of a thymine moiety in the molecule. HMBC correlation between the methyl group at δ_H 1.44 ppm, the carbon C-14 at δ_C 110.67, the carbon C-13 at δ_C 135.43, the carbon C-15 at δ_C 164.74 and between the carbon C-16 at δ_C 150.88 and the proton H-13 at δ_H 7.41 ppm and the absence of COSY correlation between this proton H-13 and the methyl group at δ_H 1.44 ppm confirmed the position of the methyl at the position 14. Proton signals at δ_H 6.31 (1H, *dd*, *J* = 8.7, 5.7 Hz, H-12), 2.50 *dd* (1H, *m*, H-11a), 2.64 (1H, *dd*, *J* = 14.2, 5.7 Hz -11b), 5.73 (1H, *d*, *J* = 6.6 Hz, H-10), 3.90 (1H, *m*, H-17), 4.52 (1H, overlap, H-18a), and 4.56 (1H, *m*, H-18b) and carbon resonances at δ_C 84.74 (CH, C-12), 37.10 (CH₂, C-11), 75.06 (CH, C-10), 83.07 (CH, C-17), and 62.33 (CH₂, C-18) supported the presence of a pentose moiety in the molecule. The HMBC spectrum showed correlations of H-10 and H-11 with C-12. In addition, the following correlations were observed in the ¹H-¹H COSY spectrum: H-12/H₂-11, H₂-11/H-10, H-10/H-17. HMBC long-range correlations from C-12 to H-13 and C-6 clearly located the 2'-desoxyribose moiety to N-1 of the 3-acetyl-5-methylthymine unit,

forming together a thymidine unit. Moreover, there were strong COSY correlations between the anomeric proton H-10 and the adjacent protons of the sugar unit H-11 and H-17. This part of the structure was confirmed by comparison with the literature data (35), and accounted for 5 degrees of unsaturation. The ^1H , ^{13}C -NMR data revealed the presence of 2 esters (δ_{C} 166.88, 164.45) accounting for other 2 degrees of unsaturation. The presence of 6 methines (δ_{C} , 104.65, 112.52, 112.33, 131.72, 114.79) and 6 quaternary carbons (δ_{C} 105.83, 126.78, 152.62, 131.40, 120.07, 162.53) led to two aromatic rings raising the number of unsaturation to 15. The COSY NMR showed the presence of two spin systems in the range of δ 6.60-7.90. The first system is a 1,4 disubstituted hydroxybenzoate ring. COSY correlations H-22 to H-21 together with the coupling constant $J=8.37$ Hz of the doublets confirmed their *ortho* coupling. Chemical shift of the quaternary carbon C-23 (δ_{C} 161.53) and its HMBC correlation to H-21 and H-22 indicated the position in 23 of a hydroxyl group, in *meta* position. Connections from H-21 to H-20/H-22 (COSY), and to the ester carbon C-19 (HMBC) confirmed the position of the ester link in 20, *para* to the hydroxyl group. These data confirmed the presence of the hydroxybenzoate system. The second spin system is part of an indole nucleus whose presence was clearly indicated by the ^1H and ^{13}C aromatic signals (Table 5) as well as the typical UV/vis spectra of this moiety compared with literature (45) (Figure S5). Within the 1,2,4 substituted benzene, *ortho* coupling H-6/H-7 was confirmed by COSY correlations together with the coupling constant of the doublet $J=8.80$ Hz. Meta coupling H-6/H-4 was confirmed by the coupling constant of the doublet $J=2.49$ Hz. HMBC correlations from the quaternary phenyl carbon C-5 (δ_{C} 152.62) to H-4/6/7, H-4 being a singlet confirmed the position of the phenyl group in 5. The remaining portion implied a fused 3 substituted pyrrole ring containing one olefinic carbon C-1 (δ_{C} 132.18) whose corresponding singlet proton H-1 at δ 8.00 ppm is correlated by HMBC to the quaternary carbons C-2 and C-3. The absence of correlation between the ester carbon C-9 (δ_{C} 164.45) and C-1, and the chemical shift of C-2 (δ_{C} 105.83) indicated that the pyrrole is substituted by the ester group in position 2. This portion accounted for the two last degrees of unsaturation of the molecule. HMBC correlation between C-9 and C-10 confirmed the link between the indole and the thymidine unit.

Table 5. ^1H -NMR and ^{13}C -NMR data (600 MHz, CD_3OD) for **1**, δ_{C} in ppm, J in Hz

Position	δ_{H}	δ_{C}	COSY ^1H - ^1H	HMBC C \rightarrow H
1	7.99 m	132.2		
2		105.8		H-1
3		126.8		H-1
4	7.52 d (2.2)	104.7		
5		152.6		H-4,H-6,H-7
6	6.80 dd (8.7,2.4)	112.5	H-7	
7	7.31 d (8.7)	112.3	H-6	
8		131.4		
9		164.5		H-10
10	5.73 d (6.6)	75.1	H-11, H-17	
11	a. 2.50 m b. 2.64 dd (14.2, 5.7)	37.1	H-10, H-12	
12	6.31 dd (8.7, 5.7)	84.7	H-11	H-10, H-11,H-13
13	7.41 m	135.4		CH3-14
14		110.7		CH3-14, H-13
15		164.7		H-13, CH3-14
16		150.9		H-13
17	3.90 m	83.1	H-10	
18	a. 4.52 overlap b. 4.56 m	62.3		
19		166.9		H-21
20		120.1	H-22	
21	7.97 d (8.8)	131.7	H-20, H-22	
22	6.87 d (8.8)	114.8	H-21	
23		161.5		H-21, H-22
24		127.1		
25		131.7		
CH₃-14	1.44 s	10.6		

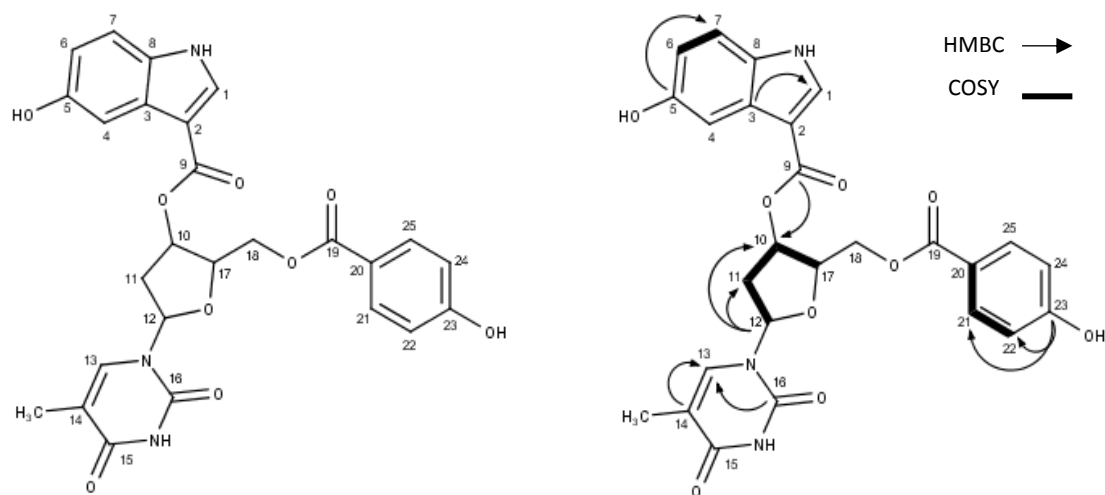


Figure 24. Structure of compound **1** and selected HMBC and COSY correlations

The molecular formula of compound **2** was established as $C_{15}H_{26}O$ ($223.2057 [M+H]^+ \Delta 0.3$ ppm) by HRESIMS and suggested 3 degrees of unsaturation (**Figure S15**). The correct structure was determined by analysing 1H , ^{13}C , edited HSQC, HMBC and COSY NMR spectra (**Figures S16; S17; S18; S19**) that led to (2,5,5,8a-tetramethyl-1,4,4a,5,6,7,8,8a-octahydronaphthalen-1-yl) methanol (**Figure 25**). Dereplication of this compound based on 1D, 2D NMR and LC-MS data in MarinLit indicated that it is an unknown compound. However, dereplication in other natural products DB indicated that it is a drimane sesquiterpenoid alcohol, drimenol, notably isolated from plants and fungi (46). The structure was confirmed by comparison with literature (**Table 6**).

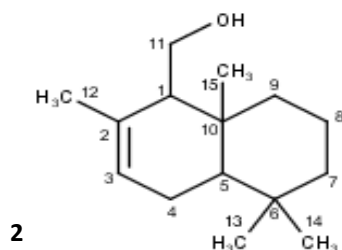


Figure 25. Structure of compound **2** isolated from Tavarua 2 SPE 100% MeOH

Table 6. ^1H -NMR and ^{13}C -NMR data (400 MHz, CD_3OD) for **2**, δ_{C} in ppm, J in Hz

Position	δ_{H}	δ_{H} Litt ^a	δ_{C}	δ_{C} Litt ^a	COSY ^1H - ^1H	HMBC C \rightarrow H
1	1.73 broad s	1.84 broad s	56.90	59.9	H-11a/b	H-3,H-5,H-9,H-11a/b,H-12
2			133.9	136.6		H-1,H-4,H-9, H-11a/b, H-12
3	5.35 broad m	5.45 m	122.4	125.3	H-4a/b	H-1,H-4,H-5,H-12
4	a. 1.82 broad m	1.85 m	23.2	26.2	H-4b, H-3	H-3,H-5,H-13
	b. 1.88 broad m	1.92 m			H-4b, H-5	
5	1.12 overlap	1.20 dd (12.1,4.6)	49.9	52.9	H-4a	H-3,H-7,H-9,H-13,H-14,H-15
6			36	35.4		H-5,H-7,H-13,H-14
7	a. 1.33 broad m	1.46 dquint (13.8,3.5)	41.9	44.9	H-7b, H-8a	H-8,H-9,H-13,H-14
	b. 1.12 m	1.21 ddd (13.4,13.4,3.5)			H-7a, H-8a/b,	
8	a. 1.50 m	1.60 qd (13.7, 3.2)	18.4	21.4	H-8b, H-9a/b	H-7,H-8,H-14
	b. 1.3 broad m	1.42 m			H-8a,H-9/b, H-7b	
9	a 1.91 broad m	1.96 m	39.2	42.2	H-9b, H-8a/b,	H-7,H-8,H-15
	b 1.03 m	1.12 dd (13.4,13.3,3.7)			H-9a,H-8a/b	
10			35.6	38.5		H-8,H-9,H-13,H-14,H-15
11	a 3.47 dd (11.1, 2.9)	3.55 dd (11.1,6.6)	59.8	2.8	H-11b, H-1	
	b 3.71 dd (10.6, 6.8)	3.79 dd (11.1, 3.)			H-11a, H-1	
12	1.66 broad m	1.74 m	20.8	23.9		H-3
13	0.82 s	0.91 s	20.9	23.8		H-4,H-5,H-7,H-8,H-14,
14	0.78 s	0.87 s	32.4	33.3		H-7,H-8,H-13
15	0.72 s	0.82 s	13.4	16.4		H-5,H-7,H-9

a.Data from literature recorded in CD_3OD (46)

2.6 Case Study 2: Isolation of compounds from an unknown ascidian (BALOLO 3) from Fiji

2.6.1 Method

The sample was originally collected from the Balolo island of Tavararua in Fiji (S17.10.750 E177.10.302) from the Didemnidae family, found on outcrops, at 12m depth. This sample was chosen only based on interesting NMR features, and for its expected ease for purification, as this was part of the training. The 100% SPE fraction was purified by a reversed phase C18 HPLC using a mixture of methanol, water and TFA (0.05%) as an eluent with a gradient from 10-100% MeOH over 13 minutes and a flow of 2mL/min. HPLC chromatograms (*Figure S20*) extracted from 200 to 400 nm presented 5 different peaks separated, dried and submitted to NMR. Two of the major peaks were submitted to HRESIMS and full structure elucidation was determined to yield compounds named **3** (42 mg) and **4** (30.2 mg).

2.6.2 Results

The molecular formula of compound **3** was established as $(C_{12}H_6Br_4O_2)$ (496.701 [M-H]⁻ Δ 0.76 ppm) by HRESIMS negative mode and suggested 8 degrees of unsaturation (*Figure S23*). This compound did not ionize at all in ESI+, so determination of the molecular mass required negative mode data. The high resolution mass spectrum (HRTOFMS) showed the unusual isotopic pattern (496.701, 498.699, 500.697, 502.695, 504.693) consistent with the presence of 4 bromines (*Figure S24*). The correct structure was determined by analysing ¹H, ¹³C, edited HSQC, HMBC and COSY NMR spectra (*Figures S25; S26; S27; S28; S29*) that led to 2,5-dibromo-3-(3,5-dibromo-2-hydroxyphenoxy)phenol (**Figure 26.3**). Dereplication of this compound based on 1D, 2D NMR and LC-MS data in MarineLit indicated that it is a known polybrominated diphenyl ether, isolated for the first time from a Callyspongiidae near Australia (47).

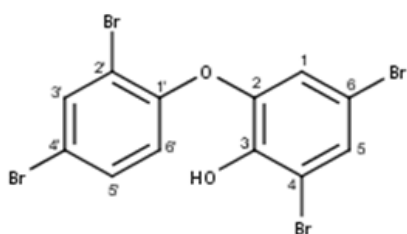
The molecular formula of compound **4** was established as $C_{13}H_8Br_4O_3$ (526.71 [M-H]⁻ Δ 0.10 ppm) by HRESIMS negative mode and suggested 8 degrees of unsaturation. The high resolution mass spectrum showed the unusual isotopic pattern (527.3, 529.17, 531.03, 532.99, 534.91) consistent with the presence of 4 bromines (*Figure S30*). The correct structure was determined by analysing ¹H (*Figure S31*), ¹³C, edited HSQC, HMBC and COSY NMR spectra

(Figures S32; S33; S34) that led to 3,5-dibromo-2-(3,5-dibromo-2-methoxyphenoxy)phenol (Figure 26.4). Dereplication of this compound based on 1D, 2D NMR and LC-MS data in MarineLit indicated that it is a known polybrominated oxydiphenol, ever isolated in a Fijian sponge *Dysidea herbacea* (48).

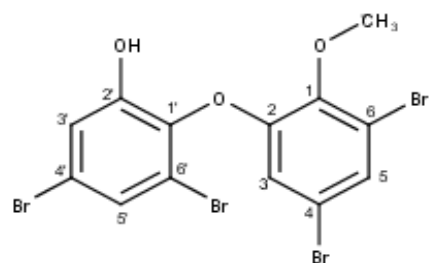
Table 7. ^1H -NMR and ^{13}C -NMR data (400 MHz, CD_3OD) for **3** and **4** δ_{C} in ppm, J in Hz

3						4					
P	δ_{H}	δ_{H} litt ^a	δ_{C}	COSY ^1H - ^1H	HMBC C \rightarrow H	δ_{H}	δ_{H} litt ^b	δ_{C}	δ_{C} litt ^a	COSY ^1H - ^1H	HMBC C \rightarrow H
1	6.69 d (2.3)	7.30 d (2.2)	120.1		H-5			145.5	145.0		H-3,H-5,H-7
1'			152.5		H-3',H-5',H-6'			138.3	137.7		H-5',H-3'
2			145.3		H-5,H-1			151.9	151.1		H-3,H-5
2'			115.0		H-3',H-6'			151.8	151.8		H-3',H-5'
3			144.3		H-5,H-1	6.4 d (2.3)	6.53 d (2.0)	116.5	116.2	H-5	H-5
3'	7.73 d (2.4)	7.79 d (2.0)	135.9		H-5',H-6'	7.07 d (2.3)	7.30 d (2.0)	119.4	119.7	H-5'	H-5'
4			111.8		H-5			118	118.5		H-5
4'			117.0		H-3',H-5',H-6'			119	118.9		H-5',H-3'
5	7.26 d (2.3)	7.41 s	130.2		H-1	7.29 d (2.3)	7.10 d (2.0)	128.2	128.0	H-3	H-3
5'	7.36 dd (8.7,2.4)	7.29 dd (8.5,2.0)	131.9	H-6'	H-3'	7.25 d (2.3)	7.25 d (2.0)	125.8	125.2	H-3'	H-3'
6			110.3		H-5, H-1			116	116.0		H-5
6'	6.80 d (8.6)	6.4 d (8.5)	121.3	H-5'				118.4	118.0		H-5',H-3'
7						3.35 s	3.96 s	60.5	60		

a.Data from literature recorded in CDCl_3 (47) b.Data from literature recorded in CD_3OD (48)



3



4

Figure 26. Compounds **3** and **4** isolated from BALOLO 3

2.7 General Discussion

The dereplication method was designed to prioritize samples that can have higher chances of producing new compounds. Our aim was to validate if the method works or not. **Table 8** shows compounds that have been isolated from our prioritized samples. Out of the seven compounds isolated from the top ranked sample, Tavarua 2, five potentially new compounds have been isolated. Structural studies are ongoing on 5 other major HPLC pure fractions. Notably, structural elucidation is ongoing on compound 6 isolated from the 100% MeOH+TFA fraction: HPLC and NMR analysis can be found *Figures S38 to S44*. According to analyse of early LCMS data and NMR spectra (*Figure S35*), the compound isolated from fraction H11 of the 50%SPE extract seems to be structurally related to compound 1. We could indeed confirm that it is the demethylated related structure. This highlight the structural diversity and complexity prioritization was based on. The new compound **1** isolated from Tavarua 2 can notably be identified in the extracted LCMS unknown masses and labelled **Figure 27**.

Table 8. Summary of compounds isolated from prioritized samples

Sample	HPLC fraction	Mass (m/z)	Status
BALOLO 3-SPE 100%MeOH	H2	497.7101	Known
	H3	527.7207	Known
TAVARUA 2- SPE 50% MeOH	H7	457.0921	New
	H10	475.1743	New
	H11	507.1322	New
	H12	521.1434	New
TAVARUA 2- SPE 100% MeOH	H4	744.8271	Known (Psammapsylin)
	H12	222.1978	Not in Marinlit DB
TAVARUA 2- SPE 100% MeOH+TFA	H6	789.5671	New

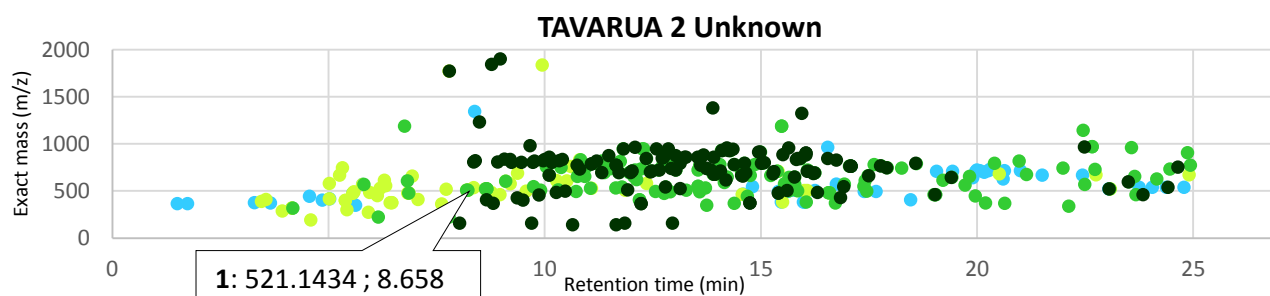


Figure 27. Exact mass (m/z) compared to retention time (min), for Tavarua 2 with labelled compound 1

Compound 2 isolated from the 100% SPE fraction appears to be a new compound in the Marinelit DB, even though it is a known terrestrial natural product. Interestingly, exact mass search in MarineLit led to its isomer, Laurobtusol, isolated from a red algae (49). Hence the usefulness of NMR and its input in the structure elucidation to discover new complexity and diversity in molecules should be highlighted. The second case study, Balolo 3 led to isolation of two polybrominated diphenyl ethers already known to be produced by the cyanobacterium (*Oscillatoria spongeliae*) associated to some sponges as *Dysidea sp.*. This sample was prioritized only on NMR analysis, but if we have a back look at the potential of new compounds (Figure 17), LCMS prioritization would have rank it to the bottom list, as it exhibited poor novelty. This demonstrates the value added by LCMS to NMR dereplication. LCMS is useful since it is sensitive; it provides accuracy in the determination of molecular formula and good resolution. But this case study also points out some of the limitations of the method. Initial LCMS was run in high resolution positive mode only, yet, we realize after isolation that that compounds 3 and 4 didn't ionize in positive mode, leading us to run LCMS in negative mode. We thus checked the processed mass data of the initial SPE extract and noticed that, indeed, this mass was missing in the ESI+ files which may explain the lack of processed data on some samples, and bias in the novelty prediction. According to Kristian F. *et al.* (50), if ESI+ helps to detect 93% of the ions, 12% still ionise poorly. This stresses out the versatile impact of the ionisation method and the compound dependant sensitivity of LCMS. Some may ionize better than compound of interest, avoiding their detection or leading to incorrect mass assignment from coeluting impurities or to ion suppression. In the key step that is dereplication, it is thus primordial to multiply ion-source setting, and applying a large ion range to keep all chances to a reliable detection of all the possible masses of the extract. Negative mode analysis would be a great improvement to consider for our method. As we saw measuring novelty, the software

extracted a high number of new masses that is most of the time not reflected after purification. It suggested that some compounds were not detectable or not sufficient to be isolated by HPLC purification for further analysis. Hence the significant advance the intensity add-on would bring to the method. UV Intensity is calculated by peak area but is not yet included in the calculated extract files. NMR is an interesting complementary approach since, as we said, it provides a quantitative information. Indeed, the intensity of the proton signal is proportional to the molar concentration of the metabolites. Added to its lower sensitivity than LCMS, it will help to pick out extracts with higher purification worthiness. In particular, it is a simple method: routine NMR doesn't require specific sample preparation and acquisition can be done in 10 minutes per sample (few tens of scans), and standard protocols can be used helping the reproducibility and the search in DB. This is however depending on the sensitivity of the instrument and the quantity of product available but, nowadays, new capabilities for exploration of sub milligram natural product samples through low-volume tube and capillary probes. However, contrary to the MS process, metabolite signals are not de-convoluted when mixture are profiled, making them sometimes superimposed. Moreover, many of them occur as multiplets because of spin-spin coupling, lowering the resolution. In the end, NMR still a useful tool to add to MS dereplication, considering the few NMR database that have emerged allowing dereplication, as ACD/HNMR DB (51).

Finally, the first case study provides the validation of our dereplication method. In complementation with the second case study, it demonstrates the usefulness of the strategy that allies accuracy and resolving power of LCMS to a more structural discriminative approach which is NMR. We showed that both method add value to one another and help to increase the level of confidence in the prioritization of natural products extracts.

This work allowed us to isolate few interesting molecules, candidates for further studies. Hydroxylated OH-PBDEs (**3**) and methoxylated MeOH-PBDEs (**4**) are metabolites of polybrominated diphenyl ethers (PBDEs). OH-PBDEs were first identified in the plasma of Baltic salmon and later in Baltic red algae and blue mussel while MeOH- PBDEs have been identified in a variety of fish and marine mammals, and form another class of bioaccumulated compounds (52). Interestingly, polybrominated diphenyl ethers (PBDEs) are synthetic compounds used as additives to retard fire and flames used since the 1960s. They have been detected in coastal and estuarine environments and accumulate in the air, soil, humans, marine

organisms and have emerged as pollutants of concern because of the detection of increasing levels in human fluids and tissues (53). If *meta*- or *para*-substituted OH-PBDEs are usually considered to be metabolites from anthropogenic origin, some studies have shown that OH-PBDEs and MeO-PBDEs with hydroxyl group in an *ortho* position had biogenic origin. Possible precursors are polybrominated phenols and anisoles which can be formed in living organisms by enzymatic bromination. Moreover, high levels and geographic distribution of these compounds in the Baltic Sea differ from what would be expected as a consequence of anthropogenic emissions, suggesting natural formation (54). Unson et al. showed that production of those compounds in *Dysidea sp.* is caused by the cyanobacterium (*Oscillatoria spongelliae*) and not by the sponge itself, acting as a defensive animal–microbe symbiosis, providing antimicrobial activity and prevention of fish-feeding (55). As we said, compound **2** was not found of MarinLit, however, searching other database as Chemspider, we found that this compound is known as *Drimenol*, a sesquiterpenoid alcohol first isolated from the bark of *Drimys winterii* tree. Recent studies suggest that *Drimenol* is the first skeletal pieces synthesized from Farnesyl Diphosphate and serve as the central precursor of drimane sesquiterpenoids (46). Those compounds are recognized as bioactive metabolites of plants, marine animals and fungi. They have aroused attention of chemists and pharmacists as they display a wide range of biological activities: antibacterial, antifungal, cytotoxic, antifeedant. Fungal drimanes have been recorded in various fungi, markedly; strains isolated from marine sponges, e.g the sponge-derived fungus *Aspergillus ustus*.

This strategy enables to focus on samples with greater potential to yield new and original molecules, accessible to any NP discovery laboratory, sparing resources, time and energy. This is exemplified by this project, with the discovery of a new nucleoside derived, 3-substituted indole (compound **1**). A number of 3-substituted indoles have been isolated from ascidians, exhibiting a wide range of complexity. Examples are 6-bromotryptamine and its derivatives isolated from *Didemnum candidum*. These alkaloids were shown to be strong inhibitors of several kinases, in particular protein kinase C (PKC); other activities include induction of cell cycle arrest and apoptosis, and the reversal of multidrug resistance in some cancer cell lines as well as inhibition of platelet aggregation and smooth muscle contraction. A good example of this potential is the Staurosporine isolated from ascidians (Polycitoridae) and also various actinomycetes. “Its derivatives have indeed shown great antitumour potential sometimes in the submicromolar range against several tumour cell lines, as 7-hydroxystaurosporine showing IC₅₀ values in the nM ranges, up to 14 times more cytotoxic than staurosporine” (56). This

molecule has been involved in clinical trial phase II in USA at the National Cancer Institute NCI (57) alone and its association with Topotecan in relapsed small cell lung cancer (58). Microorganisms and marine organisms are thus capable of producing different nucleosides with unusual structures and biological properties demonstrating a huge potential for therapeutic assays. Compound **1** features a pyrimidine nucleoside moiety i.e. a pyrimidine base (here a thymine nucleobase) linked to a 5-carbon sugar. This pyrimidine appeared to be an important pharmacophore interacting with the synthesis and function of nucleic acid. Phosphorylation of nucleosides on the primary hydroxyl group of the sugar moiety forms nucleotides, the structure elements of Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA). Both nucleosides and nucleotides are essential constituents of living cells found in several vital biological processes. Therefore scientists were interested in studying the naturally occurring nucleosides and their biochemical properties but also the effects of synthetic nucleosides on living organisms. According to literature, pyrimidine derivatives exhibit diverse ranges of biological activities. For instance, antitumour activities have been reported to involve various mechanisms, which prompts researchers to think they interact with various targets (receptors, enzymes) (59). Discoveries of these remarkable biological activities in marine nucleosides have promoted a great amount of research work on the synthesis of various analogues of these nucleosides and the further evaluation of their biological activities. An example is Spongothymidine obtained from the sponge *Cryptotethia crypta*, effective against HSV-1, HSV-2 (Herpes Simplex Viruses) and *Varicella zoster* virus (VSV). Inhibition of HSV was selective and orally effective (60). This prompts us to pursue its full structural investigation as well as its submission to antiviral and antitumor bioassays. Thereafter, the tunicate will also be phylogenetically characterised to further enlarged studies.

Bioassays will include measure of toxicity on human cells, initially *in vitro* because toxicity is a main criteria. Generally, along the drug discovery process, molecules that demonstrated *in vitro* activity and no toxicity for human cells will go forward to the *in vivo* animal testing phase and then maybe on the clinical trial in human within in a hospital setting. Those tests represent the first filter for molecules on their way towards the status of a new drug.

This molecule will be added to the database after full structure analysis.

Chapter 3 A different perspective: isolation of new neuroactive compounds from the marine bacteria *Leeuwenhoekiella sp.*

3.1 Introduction

Many sessile animals live in symbiosis with photosynthetic microorganisms, together called zooxanthellae, which provide their host with the metabolites they need and which they cannot synthesize *de novo*; this type of relationship is very frequent in coral reefs. The true origin of a substance may be either the animal, the symbiont, or even the symbiotic relation itself, that's why chemical studies will involve microbiological ones as well (61). For instance, embryos of the shrimp *Palaemon macrodactylus* are constantly covered by a strain of the bacterium *Alteromonas sp.* Removal of the bacterium results in > 90% mortality of the embryos because the bacterium produces 2,3-indolinedione (**Figure 28**) which chemically defends the embryos against attack by a pathogenic fungus (62). These symbiotic communities change with geography, which can contribute to the production of more different secondary metabolites. According to the literature, among more than 31600 microbial products discovered since 1900, about 20200 possess bioactivity: 35% were produced from filamentous fungi, 48% from actinomycetes and 17% from other (63). Some marine microbes are actually considered to be the true producers of some natural products previously assumed to come from their invertebrate host, seemingly relevant for the pharmaceutical and biotechnology industries. While the terrestrial microbial environment is very well studied so far, these microbes were largely unrecognized in the extreme world's oceans until recently, huge reservoir of bacteria such as Actinobacteria, Cyanobacteria, Myxobacteria and extremophiles bacteria. By virtue of their extreme environment where they live life (physical and chemical conditions...) and selective pressures to which they have adapted, they are considered apt to produce a variety of molecules with unique and new structural features. For instance, rapamycin (Sirolimus®), a well-known immunosuppressant, was found from the mycelium of *Streptomyces hygroscopicus* isolated from soil collected in Easter Island. "The Antarctic environment, as well as having incredibly low temperatures, possesses other diverse traits that may have helped to shape the unique way in which Antarctic bacteria have evolved" (64). This extreme environment contains hypersalinity that exists in sea ice brine channels, a lack of free water due to freezing temperatures, as well as low nutrient availability. This leads to opportunities for important and varied biotechnological exploitation.

Here we contributed to the biodiscovery pipeline aiming at the identification of novel bioactive compounds, from the bacteria *Leeuwenhoekiella sp. (Flavobacteria)*. Bacteria were tested for their neuroactivity potential and a bioassay-guided purification was performed leading to the isolation of molecules that were then structurally characterized.

The genus *Leeuwenhoekiella* is a heterotrophic, strictly aerobic, yellow-pigmented and gliding marine bacteria. It belongs to the family of *Flavobacteriaceae*, in the phylum *bacteroidetes*. The genus comprises three species *L. aequorea*, *L. marinoflava* and *L. blandensis*. Strains has been isolated from Antarctic sea water and one strain has been found in a sea urchin (65). Notably, *Bacteroidetes* are adapted to the synthesis and degradation of polymeric material such as proteins and polysaccharides, thus are involved in the degradation of organic matter during algal bloom. For instance, in the 1990s, it has been reported that *flavobacterium uliginosum* is a producer of an antitumor polysaccharide, named Marinactan. It is interesting to note that marine natural products are a rich source of potent, selective, and structurally novel intriguing compounds that can alter mammalian neurological activity. Classical example include N-type calcium channel blocker ziconotide (Prialt®), a cone snail venom peptide, recently approve by FDA for treatment of chronic pain. Few toxic neuroactive compounds are synthesized by bacteria associated with sponges that seems to be provided with neuronal-like receptor, such as the glutamate receptor. Literature analysis indicates that *Antarcticum vesiculatum* and *Psychroserpens burtonensis* isolated from the sponge *Halichondria panacea*, synthesize agonist of the NMDA receptor, important therapeutic targets for many Central Nervous System disorders including stroke, hypoxia, ischemia (66). Hence, searching for active compounds in marine microorganisms against those receptors seemed to be intelligible to discover new anti-epileptic drugs, one of the most common CNS disorders affecting about 50 million people worldwide (67). The potential of PharmaSea bacteria extracts as anti-epileptic drugs was further explored in a PTZ-induced epilepsy model in zebrafish.

The main aim here was to isolate and elucidate the structure of compounds found responsible for biological activity in extracts from *Leeuwenhoekiella sp.*, through bioassay guided purification of X0127 extract and reisolation the active compounds for full determination of biological activities.



Figure 28. *Palaemon macrodactylus* and 2,3 indolinediol

3.2 Method

3.2.1 Isolation, typing and phylogeny of strain at the University of Tromsø (UoT)

Bacteria strains were isolated from sediments on PYG minimal medium. The phylogenetic affiliation of bacterial isolates was performed through the 16S rRNA genes amplification and analysis: 16S rRNA genes were amplified by PCR and the nucleotide sequence of the amplicons determined. Each sequence was used as a query in a BLAST search to retrieve the most similar ones. Sequences were then aligned using the program ClustalW and the alignment revealed that strains are affiliated to the genus *Leeuwenhoekiella* sp. (*Flavobacteriaceae*).

3.2.2 Original Bioassays of X0127

X0127 strain was grown at Sealife pharma. The culture broth was extracted with ethyl acetate, and the extract separated by flash chromatography. Fractions were submitted to the PhotoMotor Response (PMR) bioassay in Leuven (Belgium). Two fractions demonstrated significant inhibitory activity, with a Maximum Tolerable Concentration $> 200\mu\text{g/mL}$ (68). PMR test is used to identify novel neuroactive chemicals and their targets, providing a robust and reproducible assay of nervous system function in zebrafish. The extract was tested against Pentylentetrazole PTZ induced activity. PTZ is a non-competitive antagonist of the GABA receptor complex that will induce measurable seizure-related behaviors in zebrafish, reversed by anti-epileptic drugs, creating a useful model system to identify anti-epileptic pharmaceuticals. Extracts were sent to Aberdeen to perform purification on HPLC. Bioactivity was confirmed on 4 fractions, however, yield of purified compounds being insufficient, a scale

up was undertaken by Sealife Pharma where the bacteria was regrown and sent to Aberdeen to be purified.

3.2.3 Bioassay-Guided Purification of X0127 extract

In Aberdeen, the crude extract (1g) was fractionated with an SPE C18 Cartridge. Elution was performed stepwise with an increasing MeOH concentration. The four eluted fractions were collected, dried and dissolved in DMSO to perform PMR bioassay in Leuven. The fraction eluted at 100% MeOH was shown to be the most active fraction and was separated by HPLC. Among HPLC fractions tested, 5 fractions showed high efficiency at tested concentration 4X Maximum Tolerated Concentration (MTC) (47.2 $\mu\text{g/ml}$), while 2 of them showed the most potent inhibitory activity at 12 $\mu\text{g/mL}$, with no toxicity at 4 times minimal tolerable concentration. Unfortunately, the observed bioactivity was due to minor HPLC peaks. Indeed, the major compounds isolated (found out to be peptide-like) were not the active ones. Thus, a new scale up was necessary do undertake the all bioassay guided purification on more sample. Here we contributed to reproduce the purification of another two batches of his bacteria to hunt for the molecule responsible of the bioactivity. The extracts were purified by SPE and sent to Leuven to perform bioassays.

3.3 Ongoing studies

This study is currently ongoing and awaiting bioassays results in Aberdeen. These results should allow to go further in the bio guided purification of those samples. The most active SPE fraction will then be purified by reversed phase C18 HPLC, and the resulting fractions sent back to bioassays in Leuven. The resulting positive HPLC fractions will then be purified again if needed, the corresponding compounds isolated and characterised by CASE method. We hope to yield enough active compounds to be isolated from the initial sample, as we aimed in this project.

Finally, this work gives an overview of the biologically driven strategy frequently used to hunt for new bioactive natural molecules.

Chapter 4 Conclusion and outlook

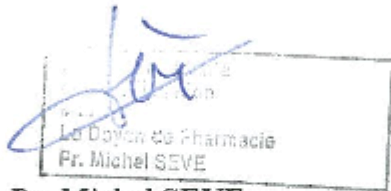
The proposed PharmaSea dereplication workflow is universal and based on a combination of online source tools allowing availability at the lab scale to any research group, institution, or company working on drug discovery from Marine Natural Products (MNPs). We hope that through our work, we were able to show that their efficient dereplication requires access to both state-of-the-art analytical/spectroscopic instrumentation and to suitable databases. The two complementary case studies described in this work give proof of concept to validate this method and highlight the interest of using an association of LCMS with NMR to efficiently prioritize natural products before purification efforts. We see that it is important to keep in mind that LCMS is a versatile method through the ionization and that several sources of ionization should be considered. Association with NMR increases the confidence in the choice of samples more pertinent to study in adding a proper structural informative value. Various aspect of this approach can be improved in the future to enhance the score ranking of candidates. The analysis of data has to be taken with caution and cannot replace the help and automation added by multiple software and sources of information. Ultimately, to attain its highest performance, the MNP dereplication toolbox should be employed cleverly. To assure this, the sagacity and experience of the MNP chemist still play an invaluable role. Nevertheless, this strategy encourages dereplication efforts by preventing researchers from working on extracts or compounds that have already been extensively studied for structure or activity. It expedites dereplication, evaluation, isolation and characterization in just one process. Samples with better potential to yield new and original molecules can be discovered with greater ease and less required resource, time and energy. This is exemplified by the novel compounds found in Tavarua's tunicate which will further go on taxonomic and bioactivity studies. Based on this prioritization method, the chase for novelty and complexity in new NP will be continued with full structural elucidation of any following isolation, to contribute to enlarge the panel of needed new drug leads and new structures to inspire pharmaco-chemists. Hence, this is the beginning of the pharmaceuticals discovery ventures. Finally, while the pharmaceutical industry has questioned the yields from traditional, terrestrial based drug discovery programs, the new molecules being observed from marine organisms indicate that this resource is unique, massive and virtually unexplored as a component of the complex drug discovery process.

Conclusion

Le processus de déréplication utilisé dans le projet PharmaSea est universel et fondé sur une association d'outils « Open source » permettant son accessibilité aux groupes de recherche, institutions ou entreprises travaillant dans la recherche de médicaments dérivés de produits naturels marins (MNP). Notre expérience montre que l'accès à un équipement analytique et spectroscopique de pointe ainsi qu'à des bases de données appropriées est primordial pour une déréplication efficace. Les deux études de cas présentées dans ce travail établissent une preuve de concept de cette méthode et mettent en évidence l'intérêt d'utiliser une association LCMS/RMN pour prioriser rentablement les extraits avant purification. Il est important de garder à l'esprit que, de par l'ionisation, la LCMS est une méthode versatile, nécessitant l'utilisation de plusieurs sources ionisantes. Par sa valeur informative structurale, la combinaison avec la RMN accroît le niveau de confiance dans le choix des échantillons les plus pertinents à étudier. Divers aspects de cette approche peuvent être améliorés à l'avenir pour une hiérarchisation rationnelle. L'analyse des données doit être prise avec prudence; l'aide et de l'automatisation apportées par l'usage de plusieurs logiciels sources d'information sont à prendre en considération. En fin de compte, la sagesse et l'expérience du chimiste MNP jouent aussi encore un rôle irremplaçable dans l'utilisation efficace du processus de déréplication. Néanmoins, cette stratégie permet de réunir déréplication, évaluation, isolement et caractérisation en un seul processus. Les échantillons ayant un meilleur potentiel de molécules nouvelles et originales peuvent être découverts plus aisément épargnant ressource, temps et énergie aux chercheurs. Ceci est illustré par les nouveaux composés trouvés dans le tunicier de Tavarua qui va ensuite être étudié taxonomiquement et pour son activité biologique. Sur la base de cette stratégie, la quête de nouveauté et complexité dans les substances naturelles sera perpétuée grâce à élucidation de la structure complète de toute nouvelle molécule isolée. Ceci pour contribuer à élargir le panel de nouveaux agents thérapeutiques et de nouvelle chimie pour inspirer les pharmaco-chimistes. Nous sommes ici donc à l'origine du processus de recherche et développement de molécules pharmaceutiques. Enfin, alors que l'industrie pharmaceutique a mis en doute la potentialité de « Drug discovery » à partir de ressources traditionnelles terrestres, les nouvelles molécules observées à partir d'organismes marins indiquent que cette ressource est unique, massive et presque inexplorée en tant que composante du processus de « réarmement » thérapeutique.

VU ET PERMIS D'IMPRIMER
Grenoble, le :

LE DOYEN



Le Doyen de Pharmacie
Pr. Michel SEVE

Pr. Michel SEVE

LE PRESIDENT DE LA THESE



Mme Delphine ALDEBERT

References

1. Antimicrobial resistance: global report on surveillance. World Health Organization Library Cataloguing-in-Publication Data: World Health Organization WHO; 2014.
2. Foo J, Michor F. Evolution of acquired resistance to anti-cancer therapy. *Journal of theoretical biology*. 2014;355:10-20.
3. Newman DJC, Gordon M. Natural Products as Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of natural products*. 2012;75(3):311-35.
4. Fabricant DS, Farnsworth NR. The Value of Plants Used in Traditional Medicine for Drug Discovery. *Environmental Health Perspectives*. 2001;109:69-75.
5. Feher M, Schmidt JM. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *Journal of Chemical Information and Computer Sciences*. 2003;43(1):218-27.
6. Chin Y-W, Balunas MJ, Chai HB, Kinghorn AD. Drug discovery from natural sources. *The AAPS Journal*. 2006;8(2):E239-E53.
7. Leal MC, Puga J, Serôdio J, Gomes NCM, 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.
8. Guenard D, Gueritte-Voegelein F, Dubois J, Potier P. Structure-activity relationships of Taxol and Taxotere analogues. *Journal of the National Cancer Institute Monographs*. 1993(15):79-82.
9. Newman DJ, Cragg GM, Snader KM. Natural Products as Sources of New Drugs over the Period 1981–2002. *Journal of Natural Products*. 2003;66(7):1022-37.
10. Grifo F ND, Fairfield A , Bhattacharya B , Grunehoff J The origins of prescription drugs. *Biodiversity and Human Health*.131- 63
11. Butler MS. The Role of Natural Product Chemistry in Drug Discovery. *Journal of Natural Products*. 2004;67(12):2141-53.
12. Costello MJ, Coll M, Danovaro R, Halpin P, Ojaveer H, Miloslavich P. A Census of Marine Biodiversity Knowledge, Resources, and Future Challenges. *PLOS ONE*. 2010;5(8):e12110.
13. El Gamal AA. Biological importance of marine algae. *Saudi Pharmaceutical Journal*. 2010;18(1):1-25.

14. Kornprobst J-M. Encyclopedia of Marine Natural Products-2nd greatly enlarged edition. s.l. . In: Kornprobst J-M, editor. Encyclopedia of Marine Natural Products: John Wiley and Sons; 2014. .
15. MarinLit. A database of the marine natural products literature. [Internet]. Royal Society of Chemistry. 2016. Available from: <http://pubs.rsc.org/marinlit/>.
16. Lang G, Mayhudin NA, Mitova MI, Sun L, van der Sar S, Blunt JW, et al. Evolving Trends in the Dereplication of Natural Product Extracts: New Methodology for Rapid, Small-Scale Investigation of Natural Product Extracts. *Journal of Natural Products*. 2008;71(9):1595-9.
17. Pérez-Victoria I, Martín J, Reyes F. Combined LC/UV/MS and NMR Strategies for the Dereplication of Marine Natural Products. *Planta Med*. 2016;82(09/10):857-71.
18. Wolfender J-L, Marti G, Thomas A, Bertrand S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A*. 2015;1382:136-64.
19. Wang Y, Kora G, Bowen BP, Pan C. MIDAS: A Database-Searching Algorithm for Metabolite Identification in Metabolomics. *Analytical chemistry*. 2014;86(19):9496-503.
20. Gottesman MM, Fojo T, Bates SE. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nature reviews Cancer*. 2002;2(1):48-58.
21. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Last 25 Years. *Journal of Natural Products*. 2007;70(3):461-77.
22. Fenical W. Marine pharmaceuticals: Past, present, and future. *Oceanography* 2006;19(2):110–9.
23. Imbri D, Tauber J, Opatz T. Synthetic Approaches to the Lamellarins—A Comprehensive Review. *Marine Drugs*. 2014;12(12):6142.
24. Fontana A, Cavaliere P, Wahidulla S, Naik CG, Cimino G. A New Antitumor Isoquinoline Alkaloid from the Marine Nudibranch *Jorunna funebris*. *Tetrahedron*. 2000;56(37):7305-8.
25. Leal MC MC, Brandão CA, Puga J, Calado R Bioprospecting of Marine Invertebrates for New Natural Products — A Chemical and Zoogeographical Perspective. *Molecules*. 2012;17(8):9842-54.
26. Hu Y, Chen J, Hu G, Yu J, Zhu X, Lin Y, et al. Statistical Research on the Bioactivity of New Marine Natural Products Discovered during the 28 Years from 1985 to 2012. *Marine Drugs*. 2015;13(1):202-21.

27. Montaser R, Luesch H. Marine natural products: a new wave of drugs? Future medicinal chemistry. 2011;3(12):1475-89.
28. Tincu JA, Taylor SW. Antimicrobial Peptides from Marine Invertebrates. Antimicrobial Agents and Chemotherapy. 2004;48(10):3645-54.
29. Blunt JW, Copp BR, Hu W-P, Munro MHG, Northcote PT, Prinsep MR. Marine natural products. Natural product reports. 2007;24(1):31-86.
30. Rabin Sen Gupta, Desa E. The Indian Ocean: A Perspective: A.A. Balkema; 2001.
31. Fusetani N. Marine Toxins: An Overview. In: Fusetani N, Kem W, editors. Marine Toxins as Research Tools. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 1-44.
32. Benkendorff K. Molluscan biological and chemical diversity: secondary metabolites and medicinal resources produced by marine molluscs. Biological Reviews. 2010;85(4):757-75.
33. Chapman V. Seaweeds in pharmaceuticals and medicine: a review. Marine Algae in Pharmaceutical Science. 1979 1:139-47.
34. Bo G. Giuseppe Brotzu and the discovery of cephalosporins. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2000;6 Suppl 3:6-9.
35. Sarfaraj H SF, Saba A, Mohd S. Indian Marine natural products: A lead for Anti-cancer. J Geomarine Sci. 2012;41(1):27-39.
36. Pettit GR HC, Doubek DL, Herald DL, Arnold E, Clardy J. Isolation and structure of bryostatin 1. J Am Chem Soc. 1982;104(24):6846-8.
37. Natural Products Isolation. 3 ed: Humana Press; 2012.
38. Sarah H. Breaking the rules. Chemistry world. June 2008.
39. Mestrelab Research- Chemistry Software Solution. s.l.: Mestrelab Research, S.L. Terms; 2016 [Available from: <http://mestrelab.com/>].
40. Tunicata. Ascidiacea World Database [Internet]. 2016. Available from: <http://www.marinespecies.org/aphia.php?p=taxdetails&id=146420>.
41. Schmidt EW, Donia MS, McIntosh JA, Fricke WF, Ravel J. Origin and Variation of Tunicate Secondary Metabolites. Journal of Natural Products. 2012;75(2):295-304.
42. Zheng L-H, Wang Y-J, Sheng J, Wang F, Zheng Y, Lin X-K, et al. Antitumor Peptides from Marine Organisms. Marine Drugs. 2011;9(10):1840-59.
43. Chun HG, Davies B, Hoth D, Suffness M, Plowman J, Flora K, et al. Didemnin B. The first marine compound entering clinical trials as an antineoplastic agent. Investigational new drugs. 1986;4(3):279-84.

44. Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J, et al. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol.* 2006;2(12):729-35.
45. Sundberg R. *The Chemistry of Indoles*: Elsevier Science; 2012.
46. Kwon M, Cochrane SA, Vederas JC, Ro D-K. Molecular cloning and characterization of drimenol synthase from valerian plant (*Valeriana officinalis*). *FEBS Letters.* 2014;588(24):4597-603.
47. Capon R, Ghisalberti EL, Jefferies PR, Skelton BW, White AH. Structural studies of halogenated diphenyl ethers from a marine sponge. *Journal of the Chemical Society, Perkin Transactions 1.* 1981(0):2464-7.
48. Norton RS CK, Wells RJ Polybrominated oxydiphenol derivatives from the sponge *Dysidea herbacea*. Structural determination. *Tetrahedron.* 1981;37(13):2341-9.
49. Caccamese S AV, Neri P, Foti M. . The structure of laurobtusol, a new rearranged sesquiterpenoid from the mediterranean red alga *laurencia obtusa*. *Tetrahedron.* 1991;47(48):10101-8.
50. Nielsen KF, Månsson M, Rank C, Frisvad JC, Larsen TO. Dereplication of Microbial Natural Products by LC-DAD-TOFMS. *Journal of Natural Products.* 2011;74(11):2338-48.
51. Mose A WP, Hayward S. . Dereplication of Natural Products by NMR: A Three-stage approach: Advanced Chemistry Development, Inc., Toronto, On, Canada,; 2011 [Available from: <http://www.acdlabs.com>].
52. Siddiqi MA, Laessig RH, Reed KD. Polybrominated diphenyl ethers (PBDEs): new pollutants-old diseases. *Clinical medicine & research.* 2003;1(4):281-90.
53. Bergman Å AM, Fäldt E, Jakobsson K. Hydroxylated PBDE metabolites in human blood. *Organohalogen Compounds.* 2006;68:635–8.
54. W. Gribble G. The diversity of naturally occurring organobromine compounds. *Chemical Society Reviews.* 1999;28(5):335-46.
55. Unson MD, Faulkner DJ. Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia.* 1993;49(4):349-53.
56. Jimenez PC, Wilke DV, Ferreira EG, Takeara R, de Moraes MO, da Silveira ER, et al. Structure elucidation and anticancer activity of 7-oxostaurosporine derivatives from the Brazilian endemic tunicate *Eudistoma vannamei*. *Mar Drugs.* 2012;10(5):1092-102.
57. UCN-01 (7-Hydroxystaurosporine) to Treat Relapsed T-Cell Lymphomas *ClinicalTrials.gov.* Bethesda (MD): National Library of Medicine (US): National Cancer

Institute (NCI); 2014 [Available from:

<https://clinicaltrials.gov/ct2/show/study/NCT00082017>.

58. 7-hydroxystaurosporine and Topotecan Hydrochloride in Treating Patients With Relapsed or Progressed Small Cell Lung Cancer ClinicalTrials.gov.Bethesda (MD): National Library of Medicine (US): National Cancer Institute (NCI); 2015 [Available from: <https://clinicaltrials.gov/ct2/show/study/NCT00098956>.

59. Kaur R, Kaur P, Sharma S, Singh G, Mehndiratta S, Bedi PM, et al. Anti-cancer pyrimidines in diverse scaffolds: a review of patent literature. Recent patents on anti-cancer drug discovery. 2015;10(1):23-71.

60. Huang RM, Chen YN, Zeng Z, Gao CH, Su X, Peng Y. Marine nucleosides: structure, bioactivity, synthesis and biosynthesis. Mar Drugs. 2014;12(12):5817-38.

61. Florez LV, Biedermann PH, Engl T, Kaltenpoth M. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. Natural product reports. 2015;32(7):904-36.

62. Gil-Turnes MS, Hay ME, Fenical W. Symbiotic marine bacteria chemically defend crustacean embryos from a pathogenic fungus. Science. 1989;246(4926):116-8.

63. Carrano L, Marinelli F. The relevance of chemical dereplication in microbial natural product screening. Journal of Applied Bioanalysis. 2015;1(2):55-67.

64. Jaspars M, Challis G. Microbiology: a talented genus. Nature. 2014;506(7486):38-9.

65. Nedashovskaya Ol VM, Kim SB, Zhukova NV, Han JH, Mikhailov VV. *Leeuwenhoekella palythoae* sp. nov., a new member of the family Flavobacteriaceae. International journal of systematic and evolutionary microbiology 2009;59(12):3074-7.

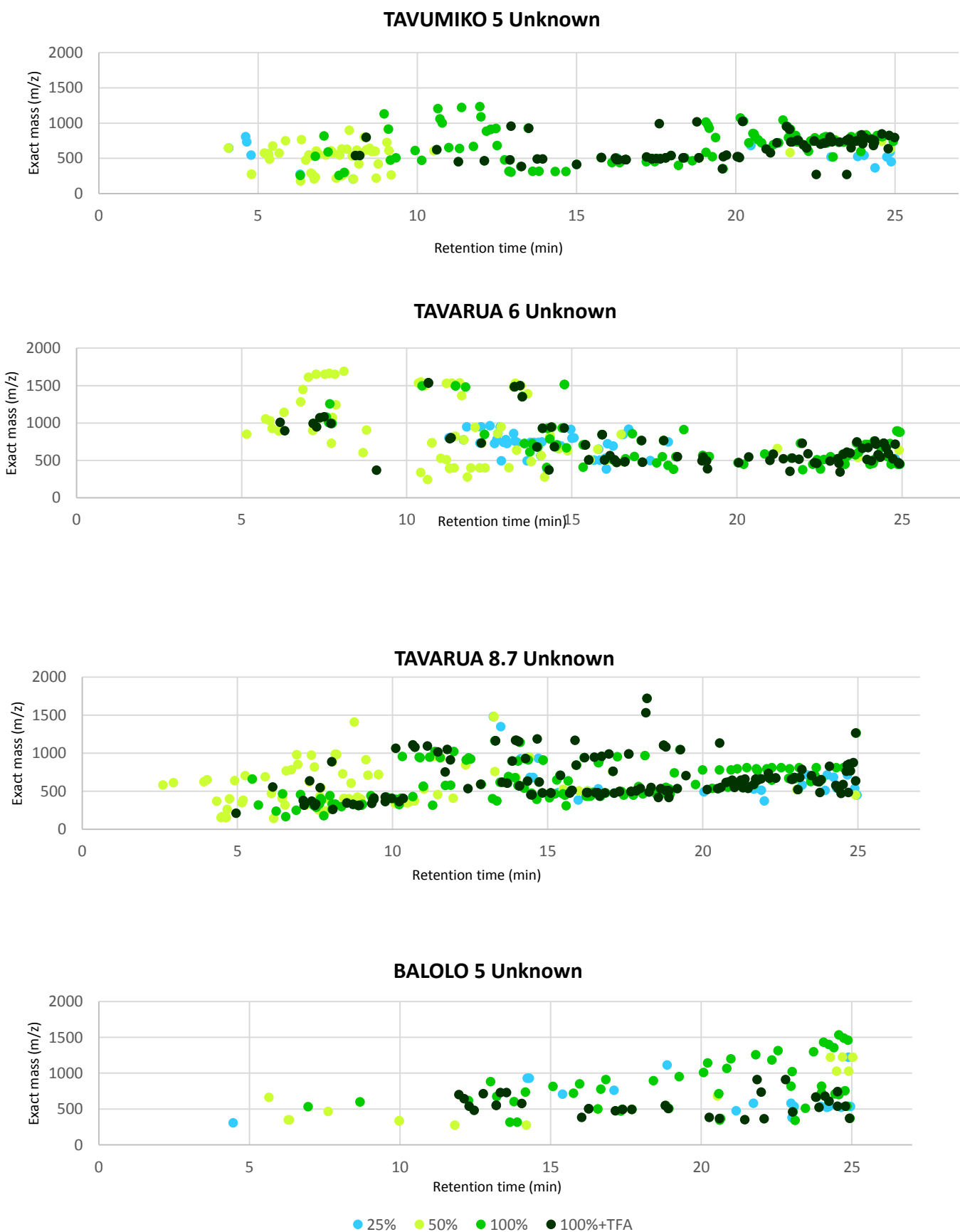
66. Perovic S, Wichels A, Schutt C, Gerdtts G, Pahler S, Steffen R, et al. Neuroactive compounds produced by bacteria from the marine sponge *Halichondria panicea*: activation of the neuronal NMDA receptor. Environmental toxicology and pharmacology. 1998;6(2):125-33.

67. Megiddo I CA, Chisholm D, Dua T, Nandi A, Laxminarayan R WHO Epilepsy: World Health Organization WHO; 2016 [Available from: <http://www.who.int/mediacentre/factsheets/fs999/en/>.

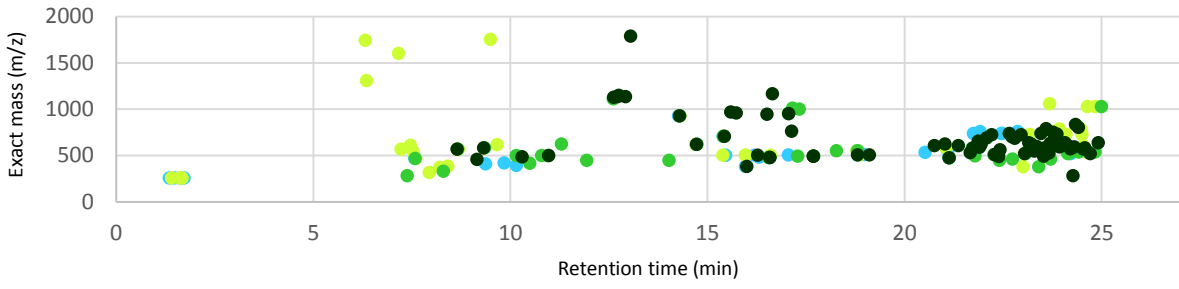
68. Long S-M, Liang F-Y, Wu Q, Lu X-L, Yao X-L, Li S-C, et al. Identification of Marine Neuroactive Molecules in Behaviour-Based Screens in the Larval Zebrafish. Marine Drugs. 2014;12(6):3307-22.

Supplementary data

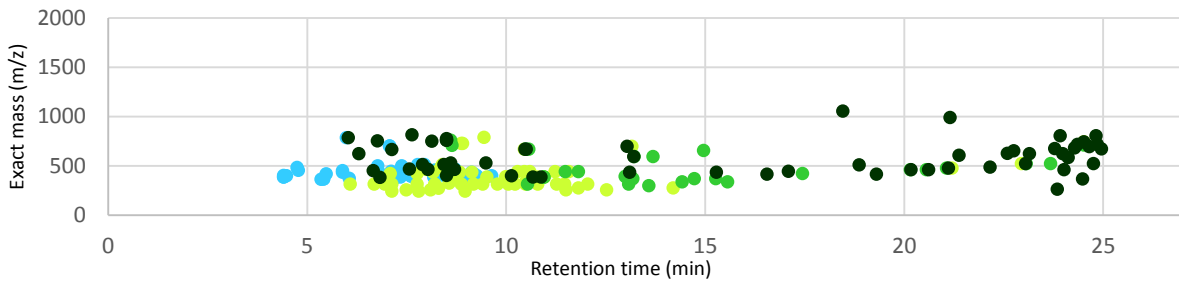
Figure S1. Exact mass (m/z) compared to retention time (min), for few samples studied



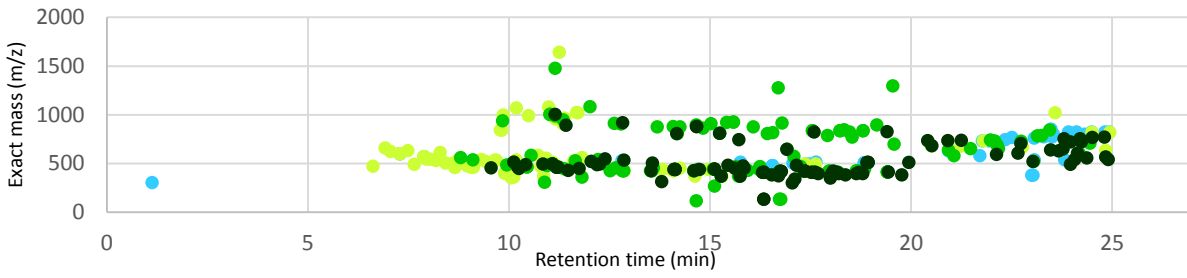
BALOLO 3 Unknown



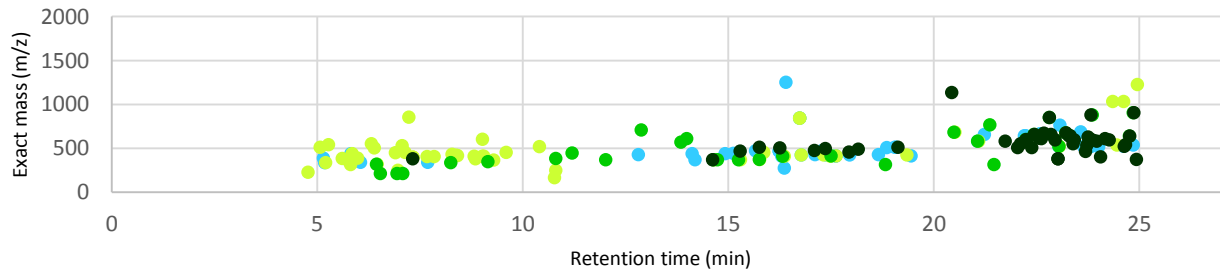
EDEN 5 Unknown



BUAGO 2 Unknown

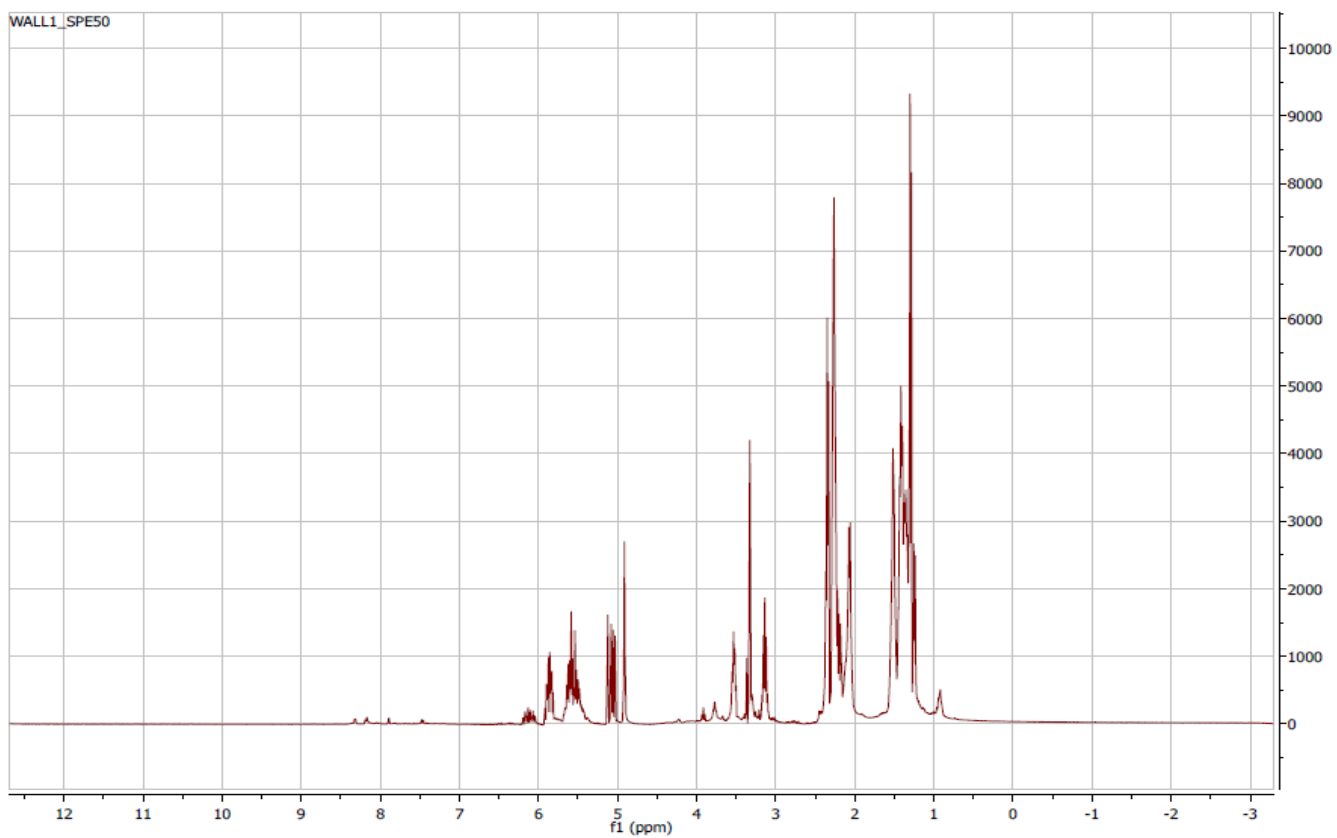
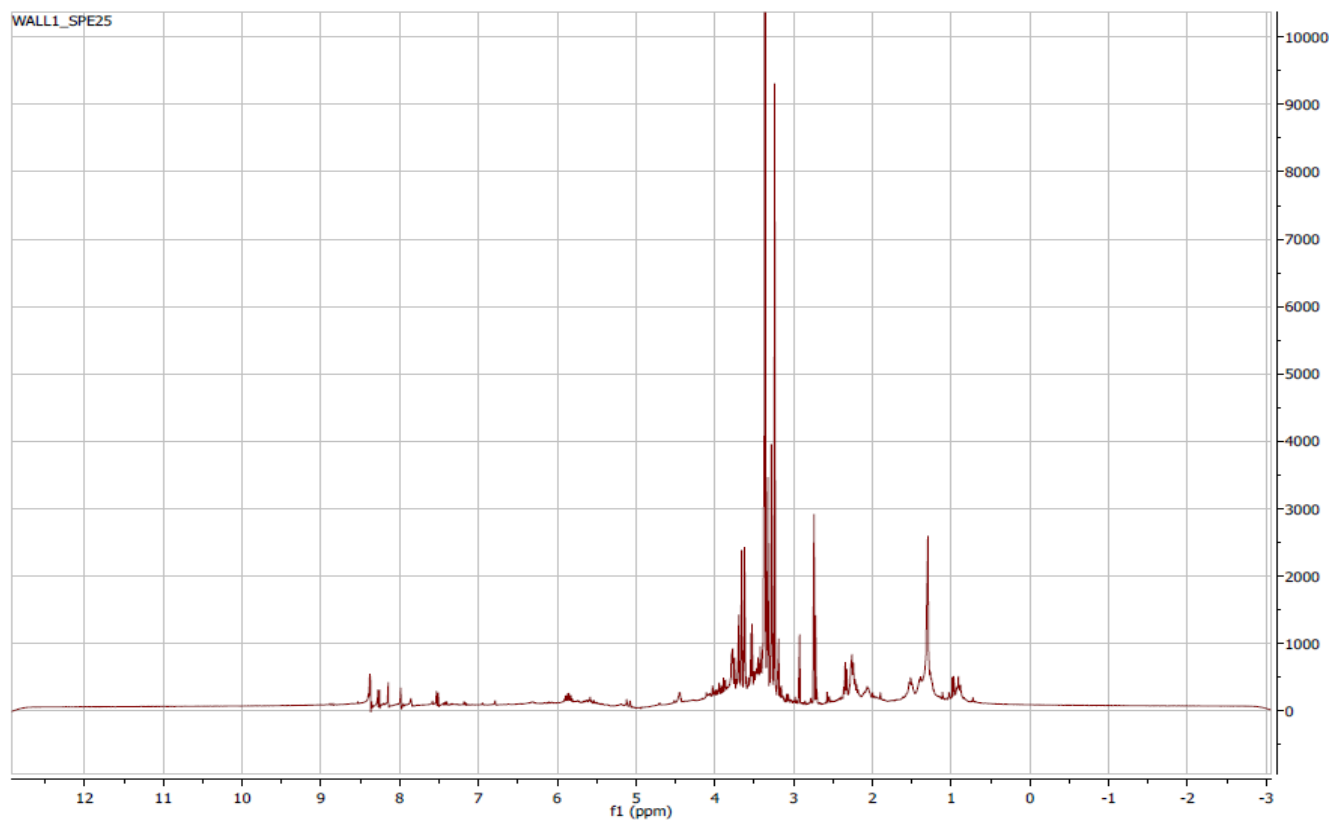


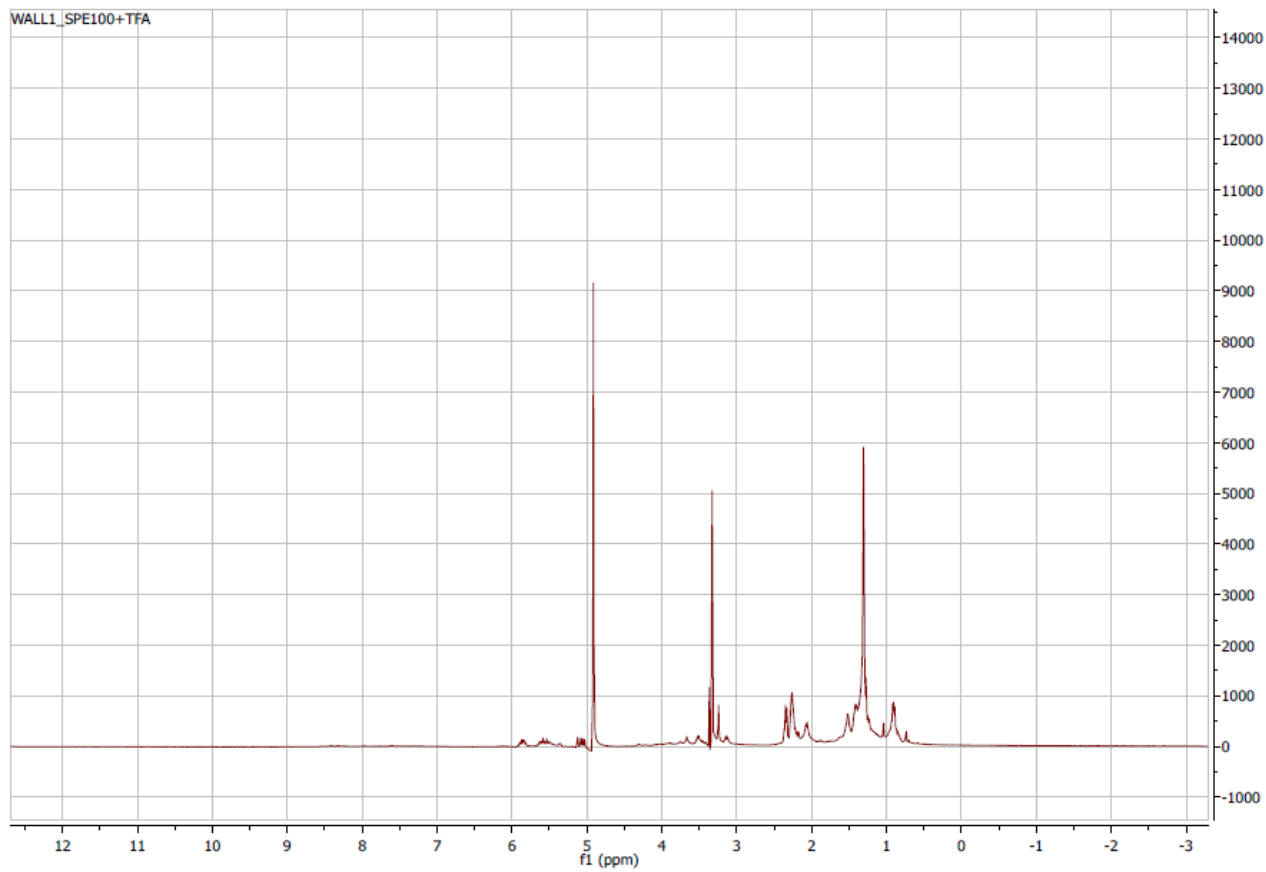
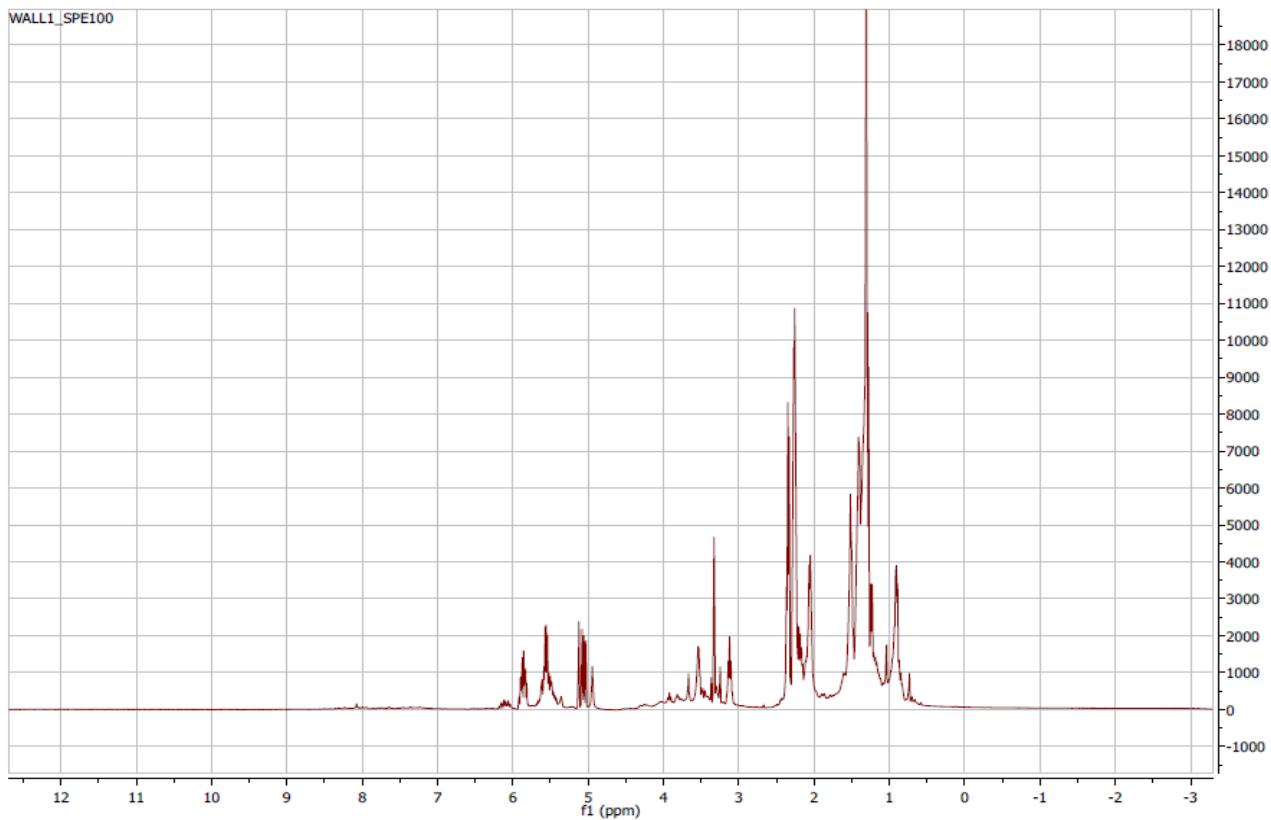
BUAGO 3 Unknown

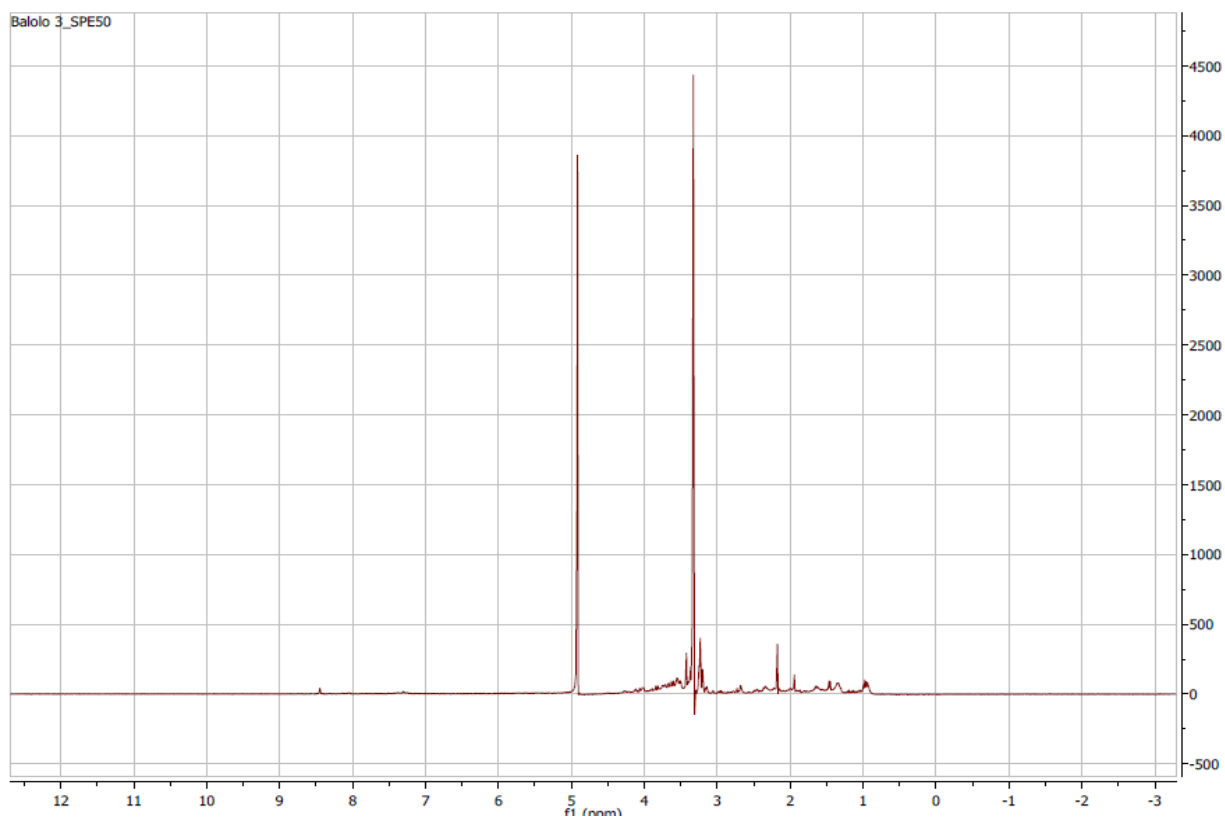
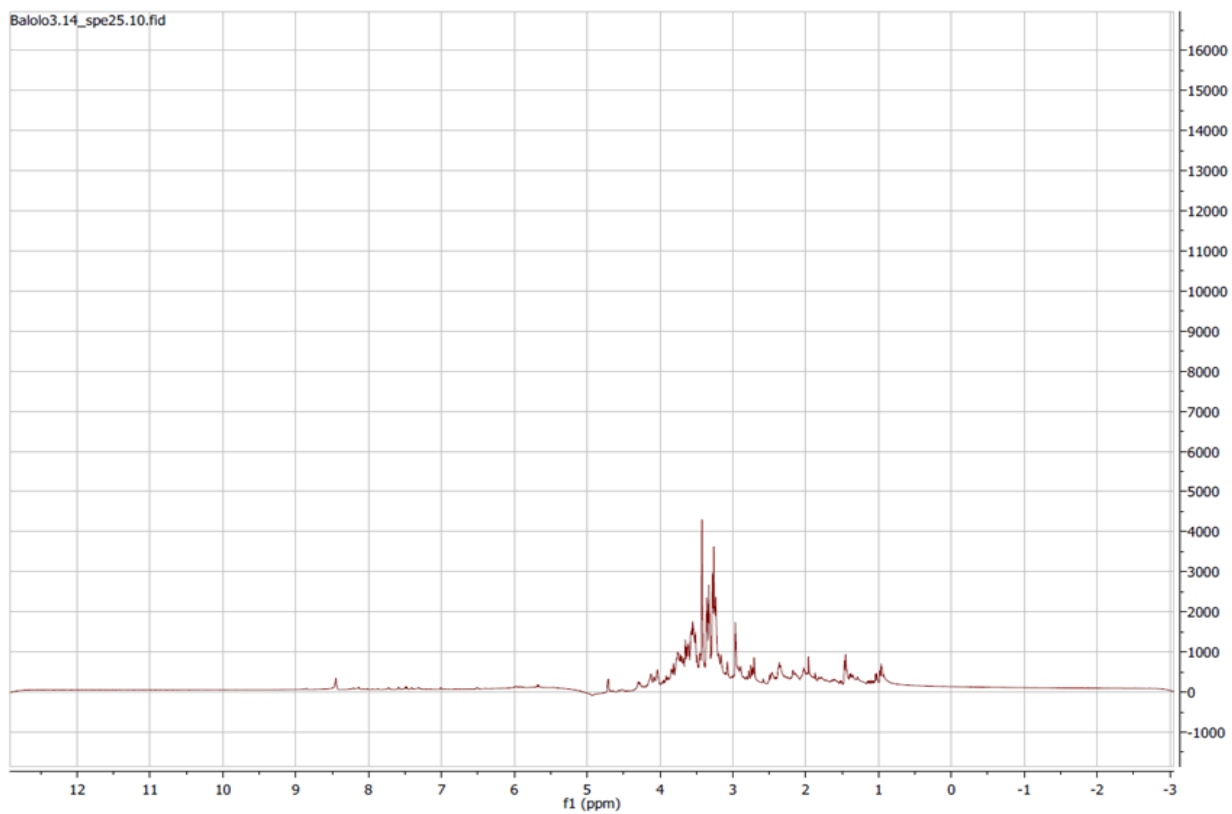


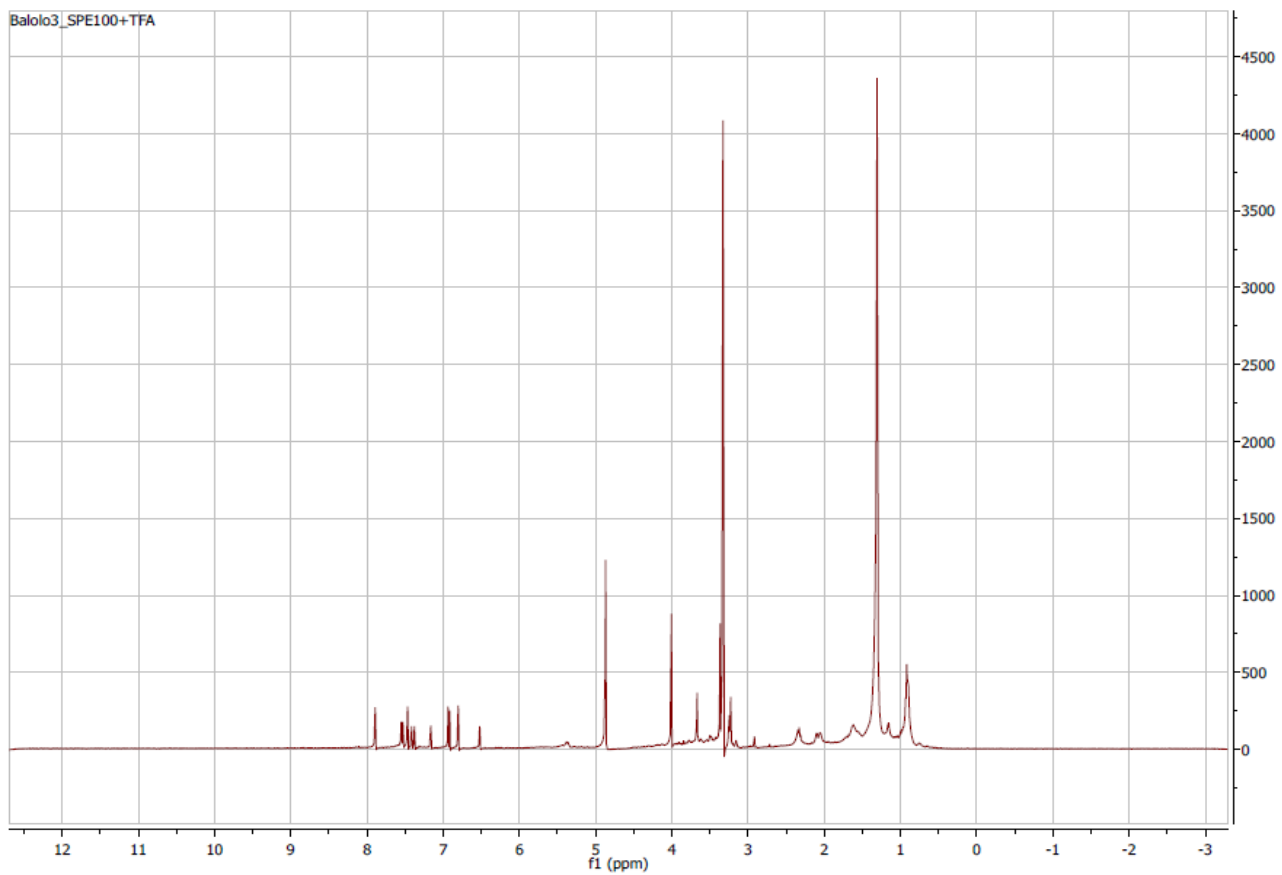
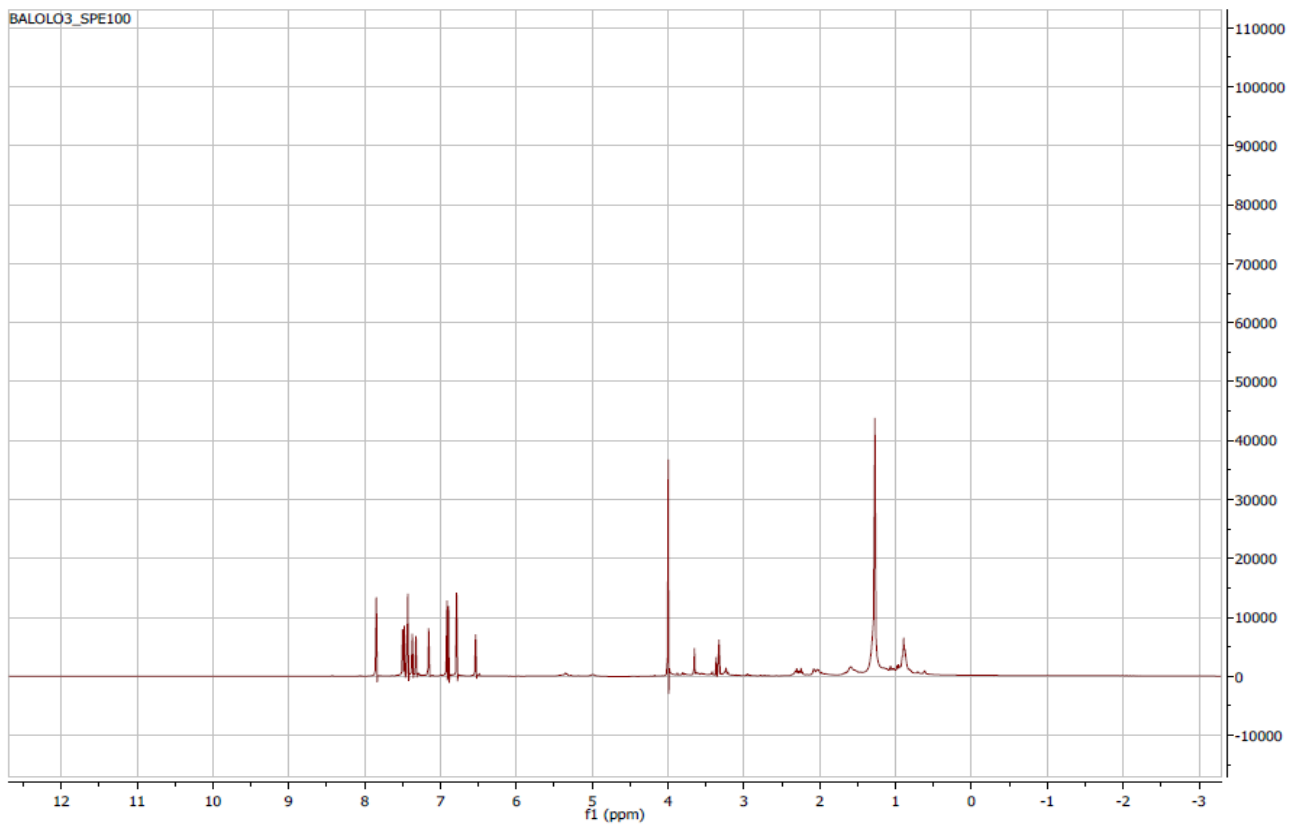
● 25% ● 50% ● 100% ● 100%+TFA

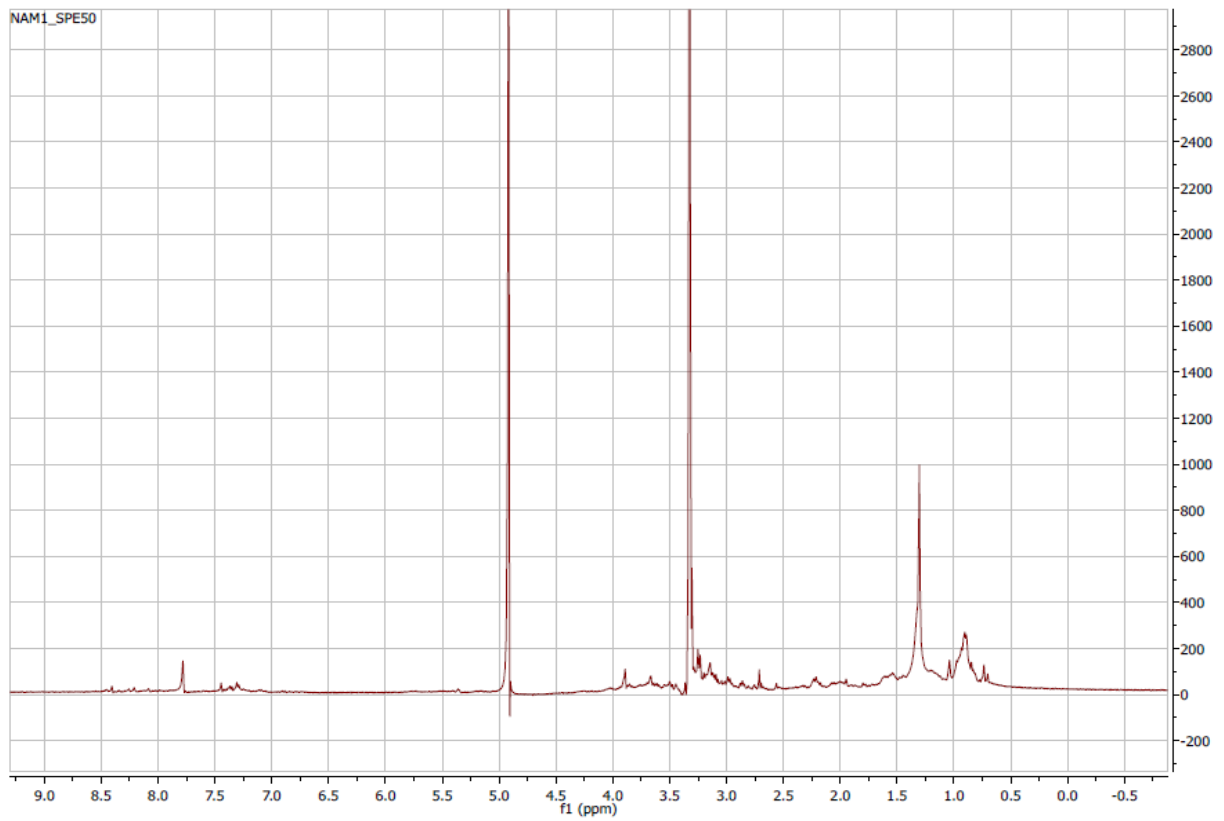
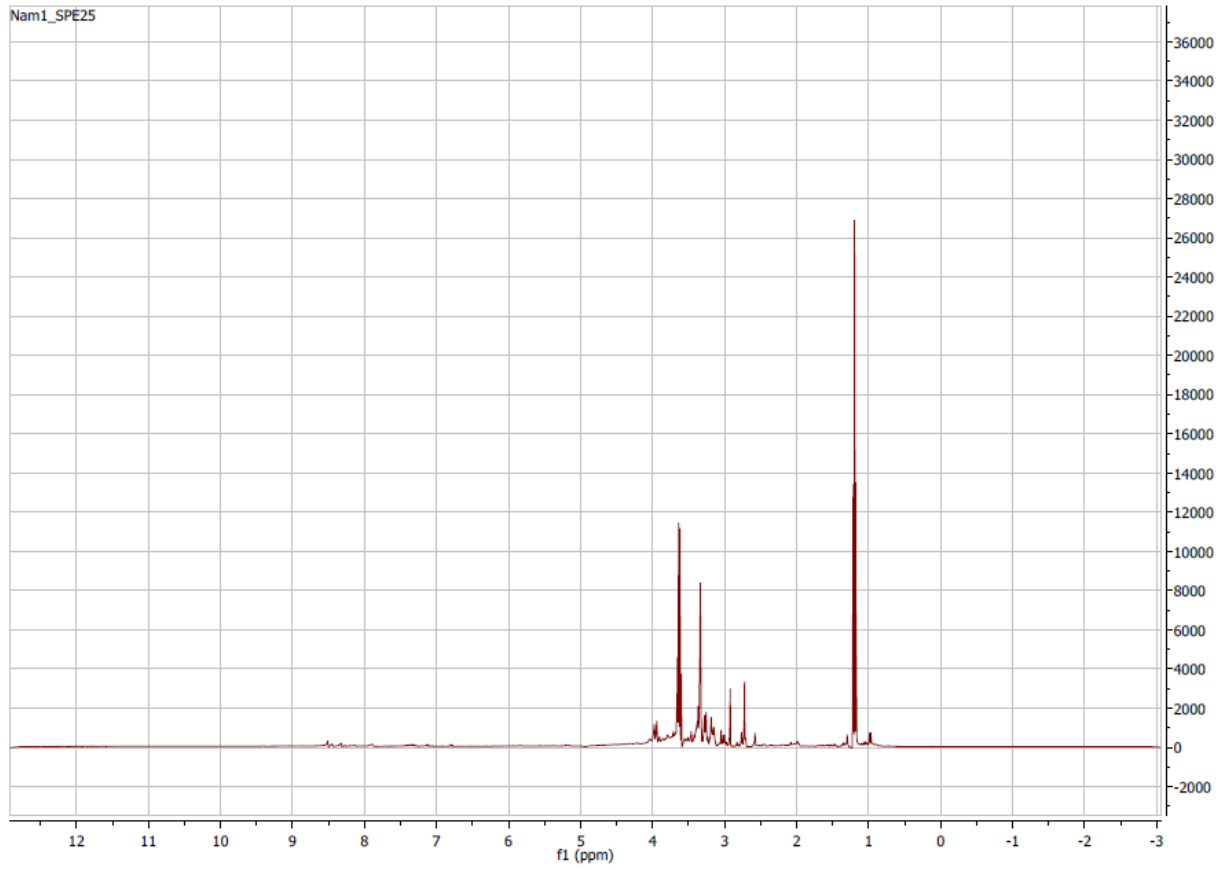
Figure S2. ^1H NMR spectrum of some of the top ranked samples in CD_3OD at 400 MHz

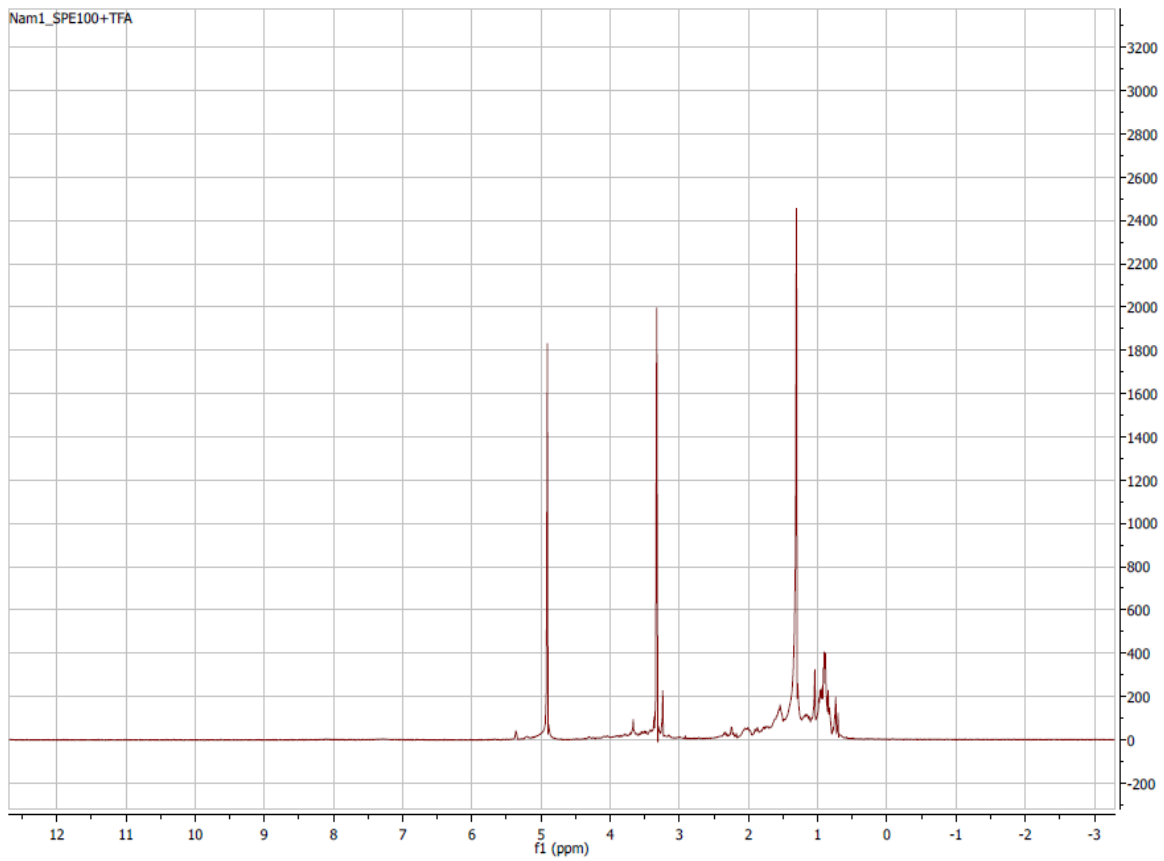
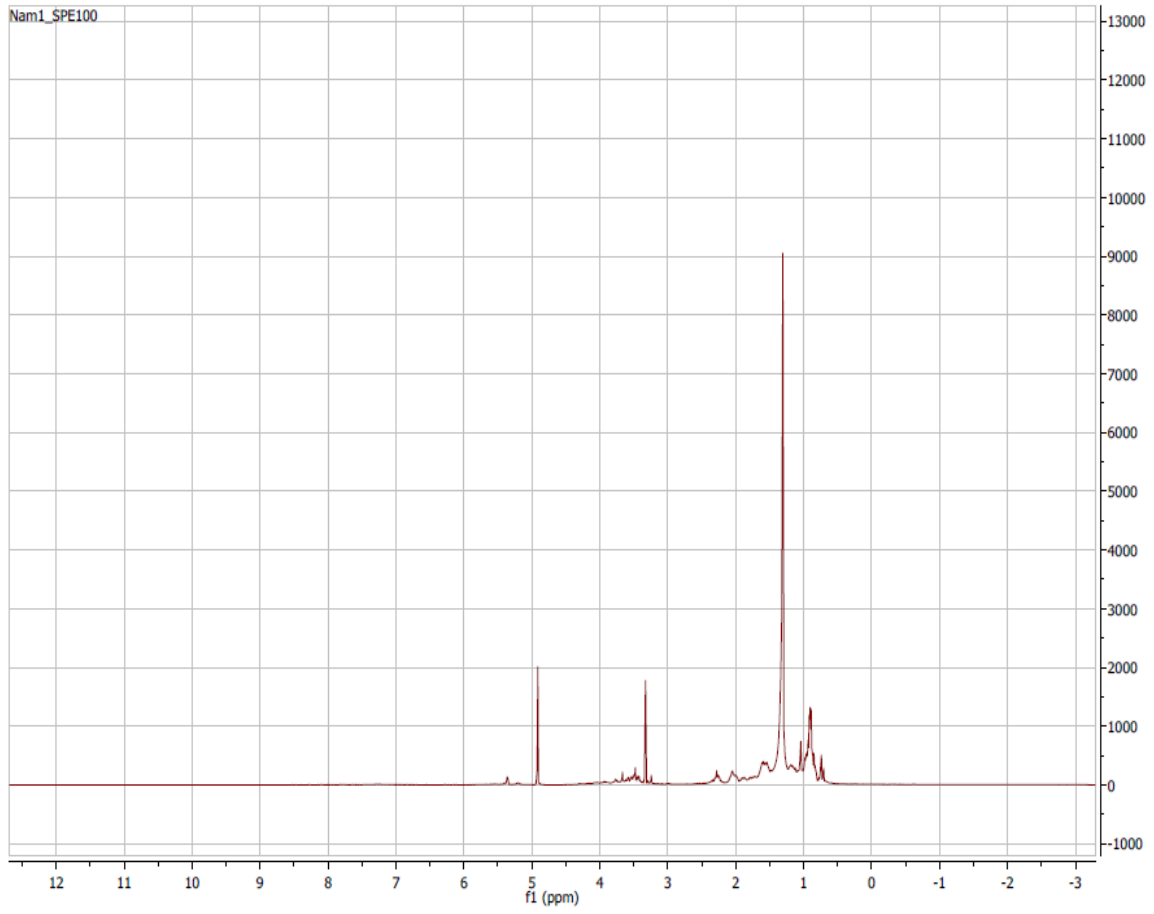


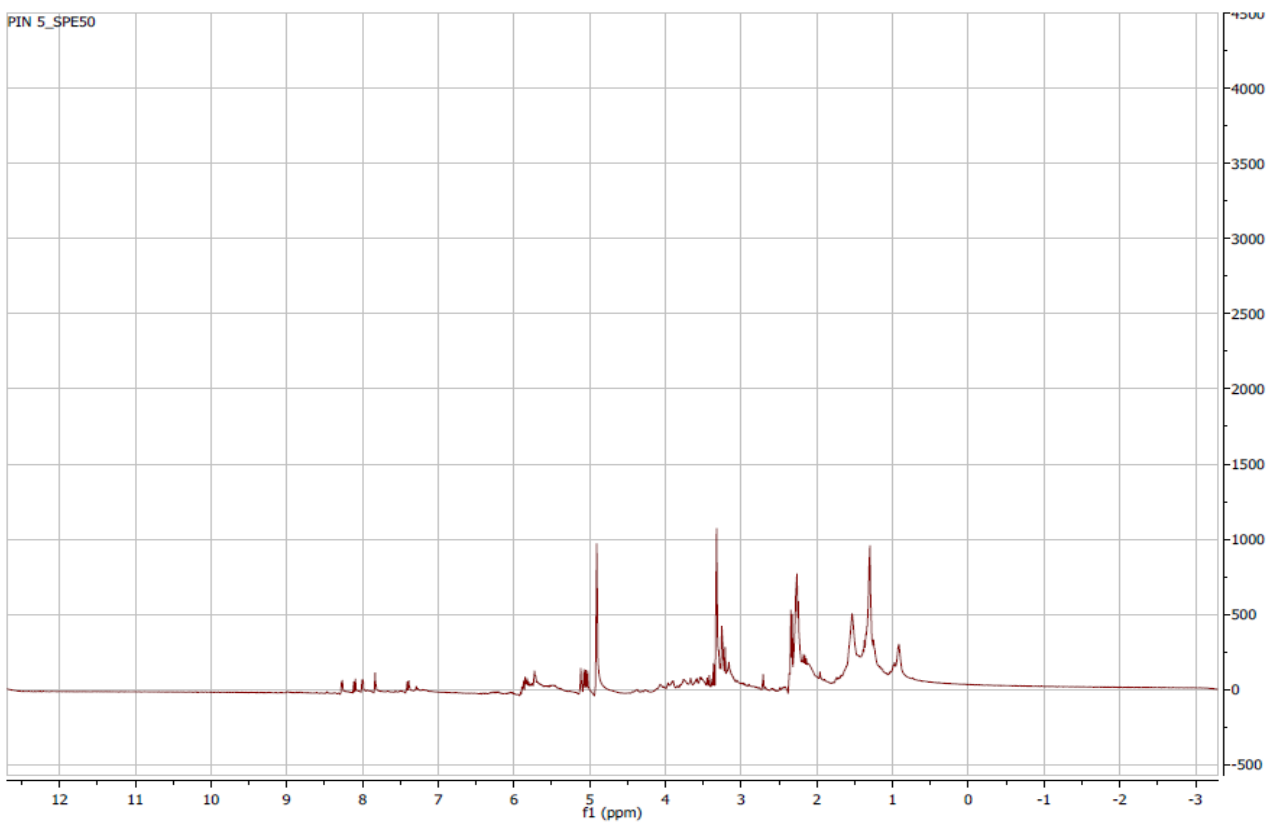
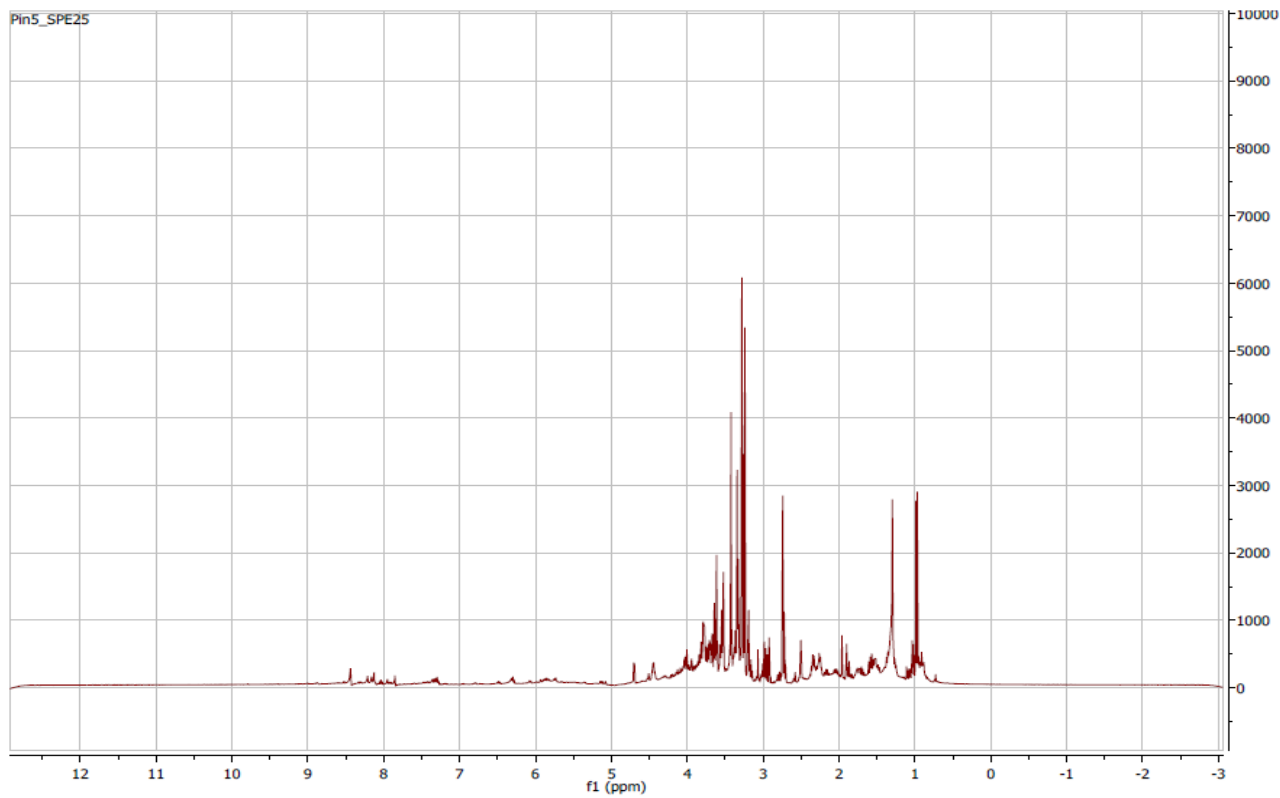


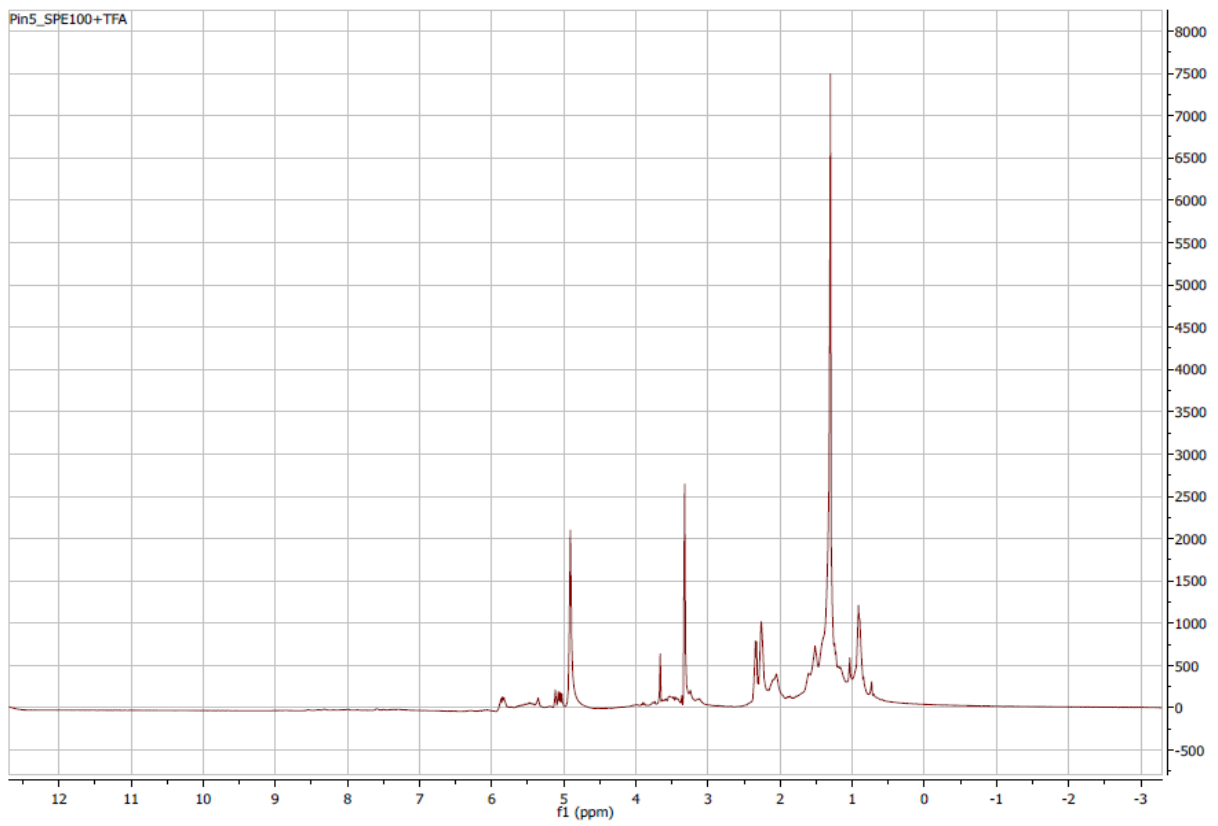
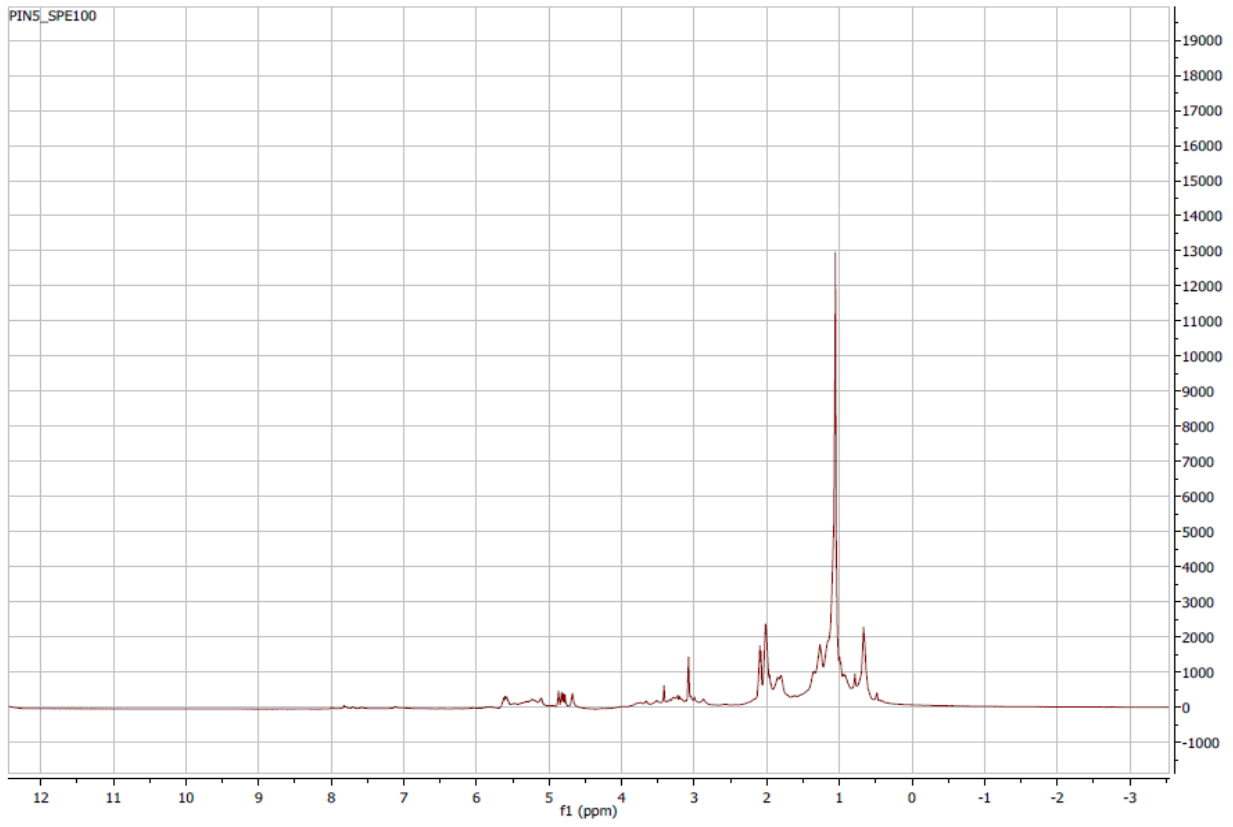


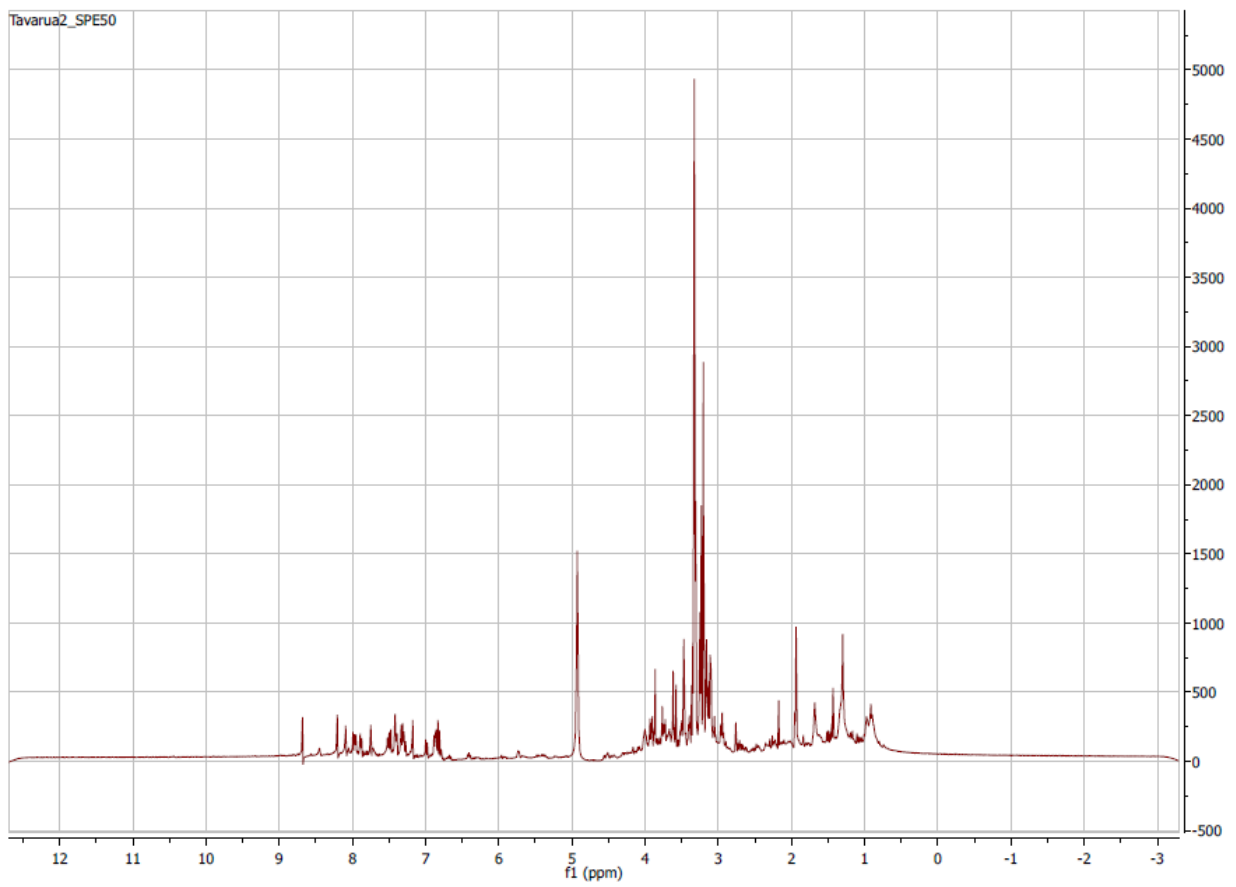
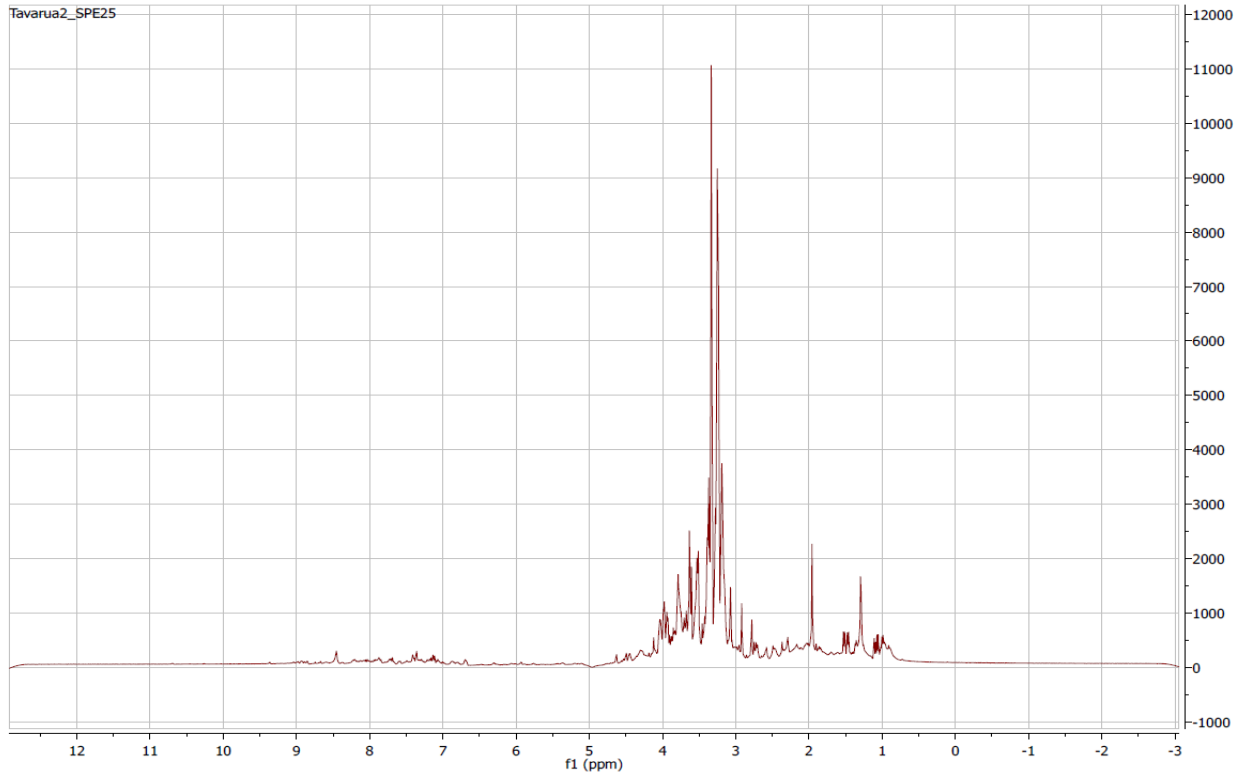


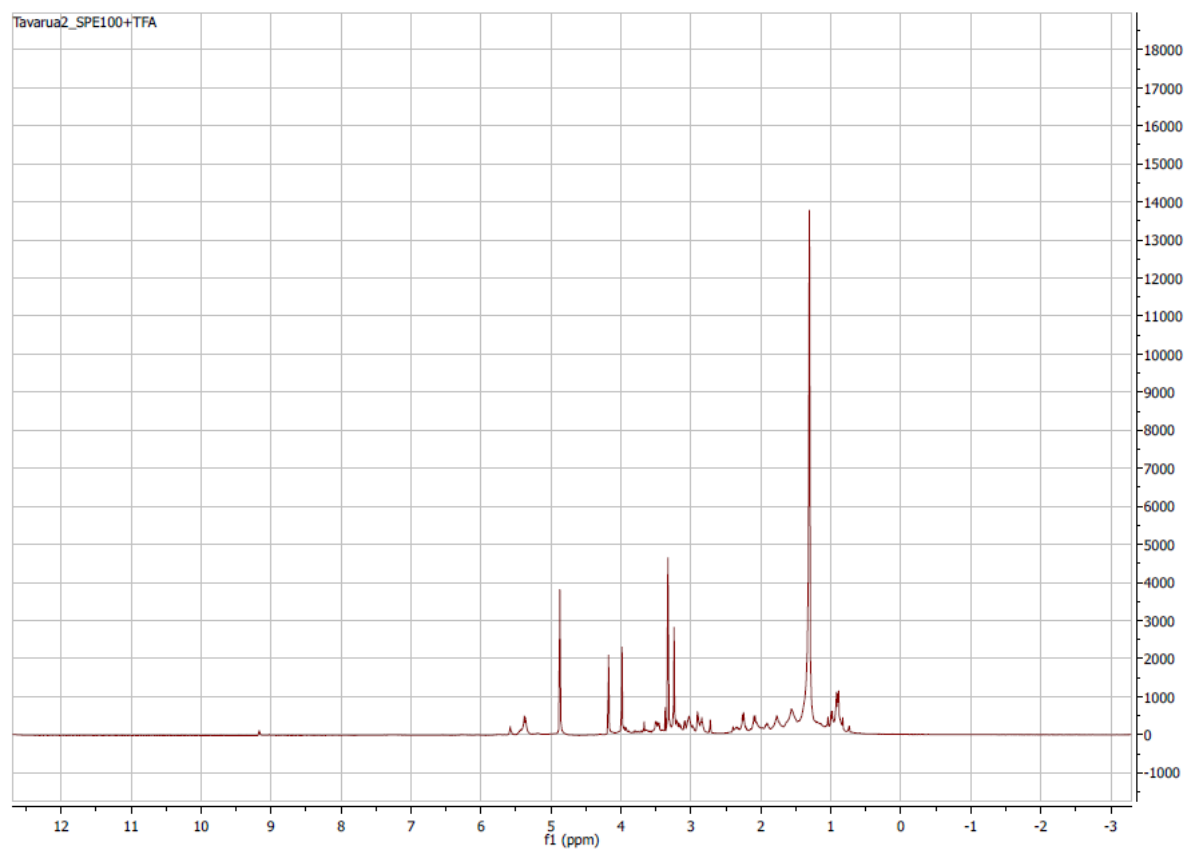
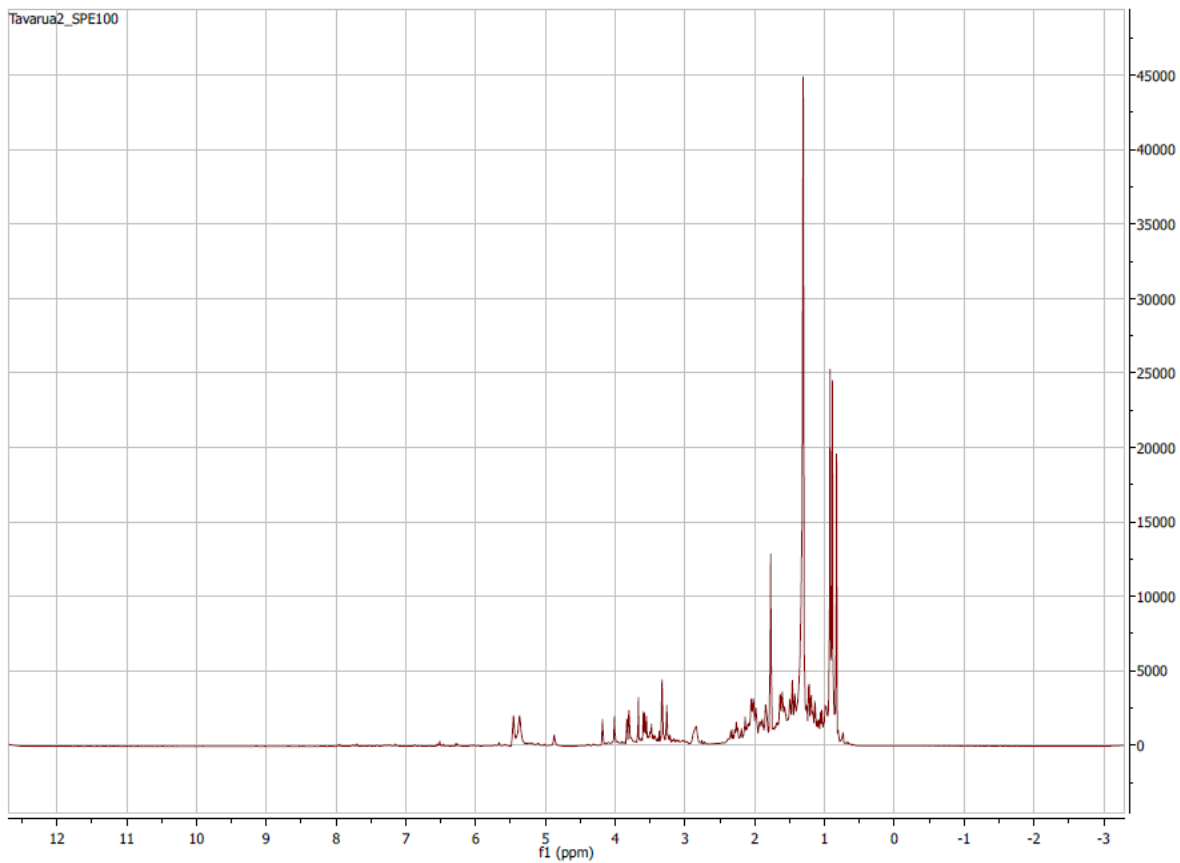


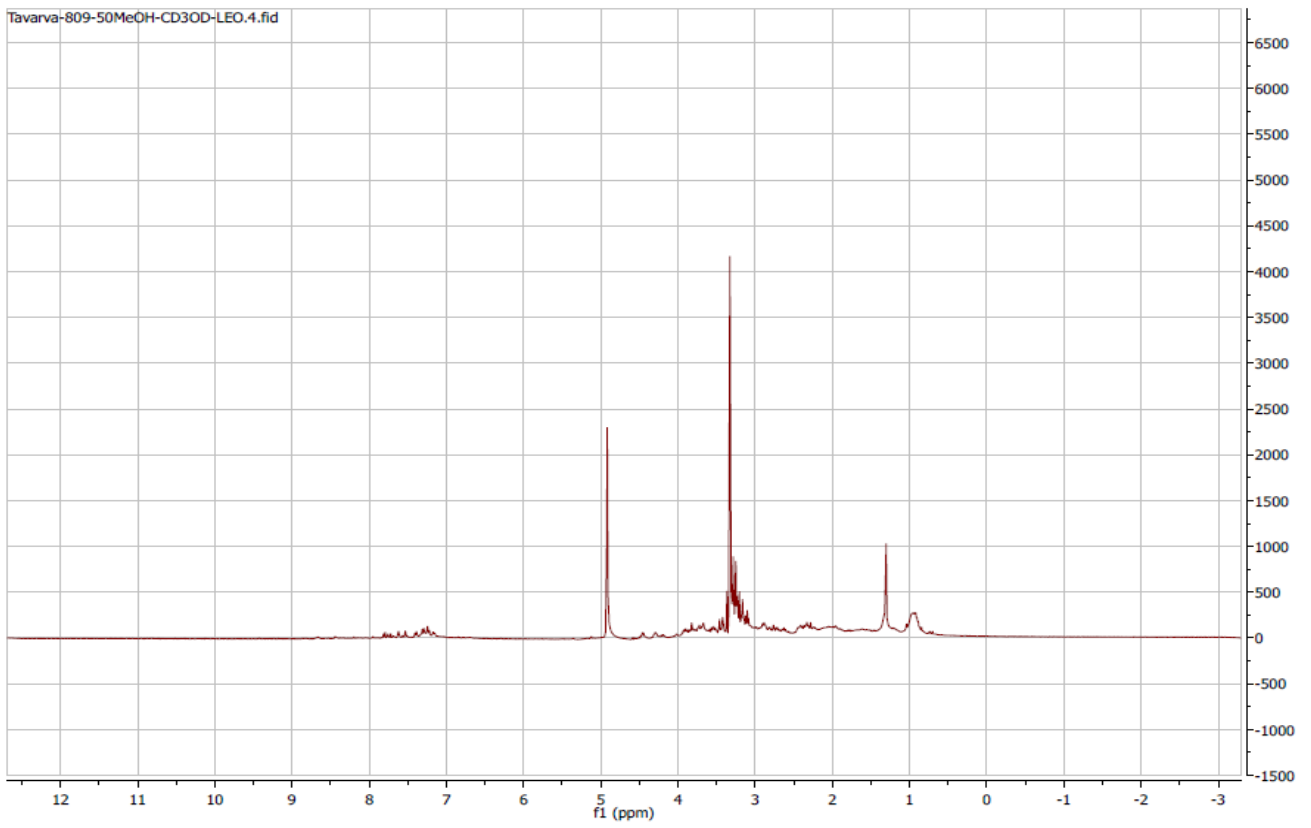
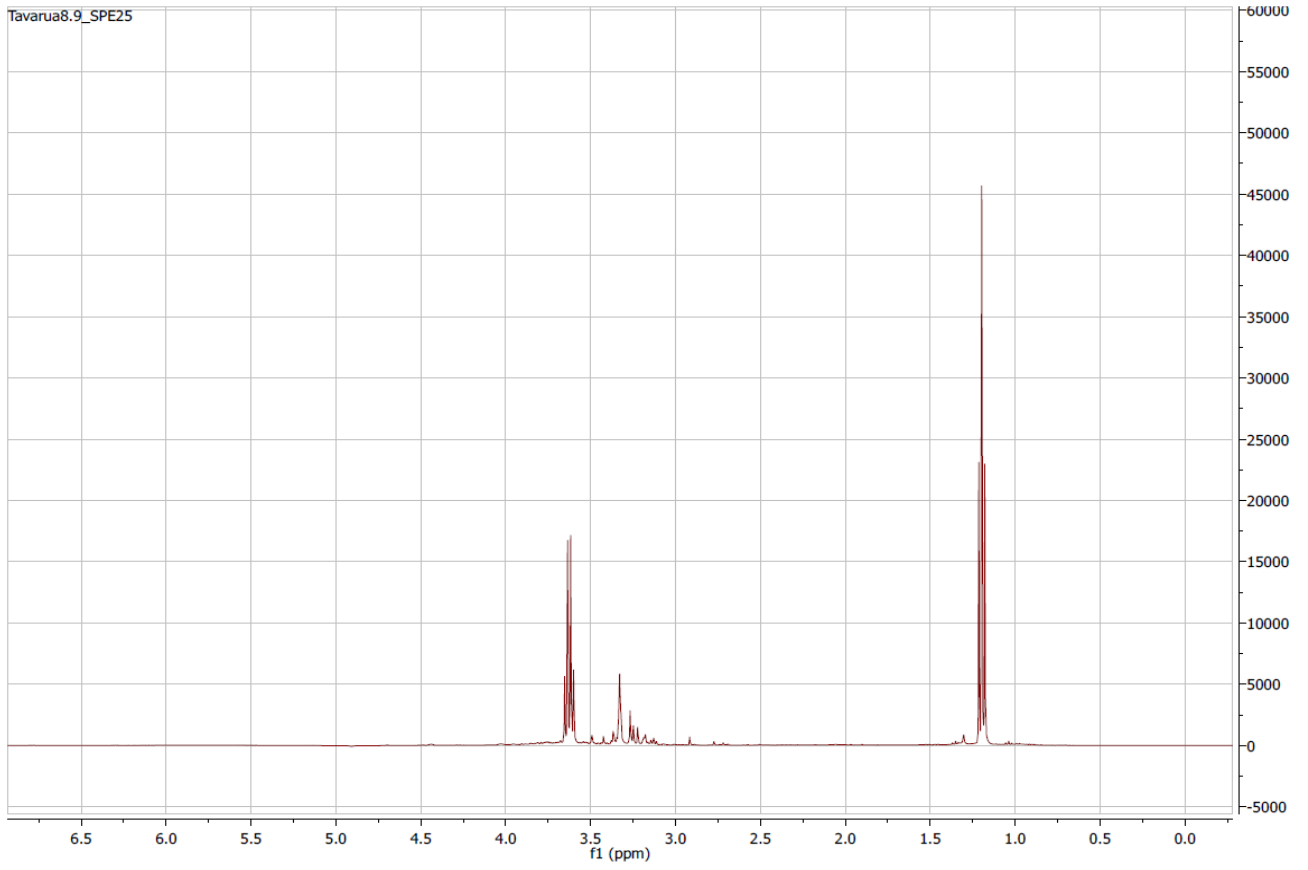


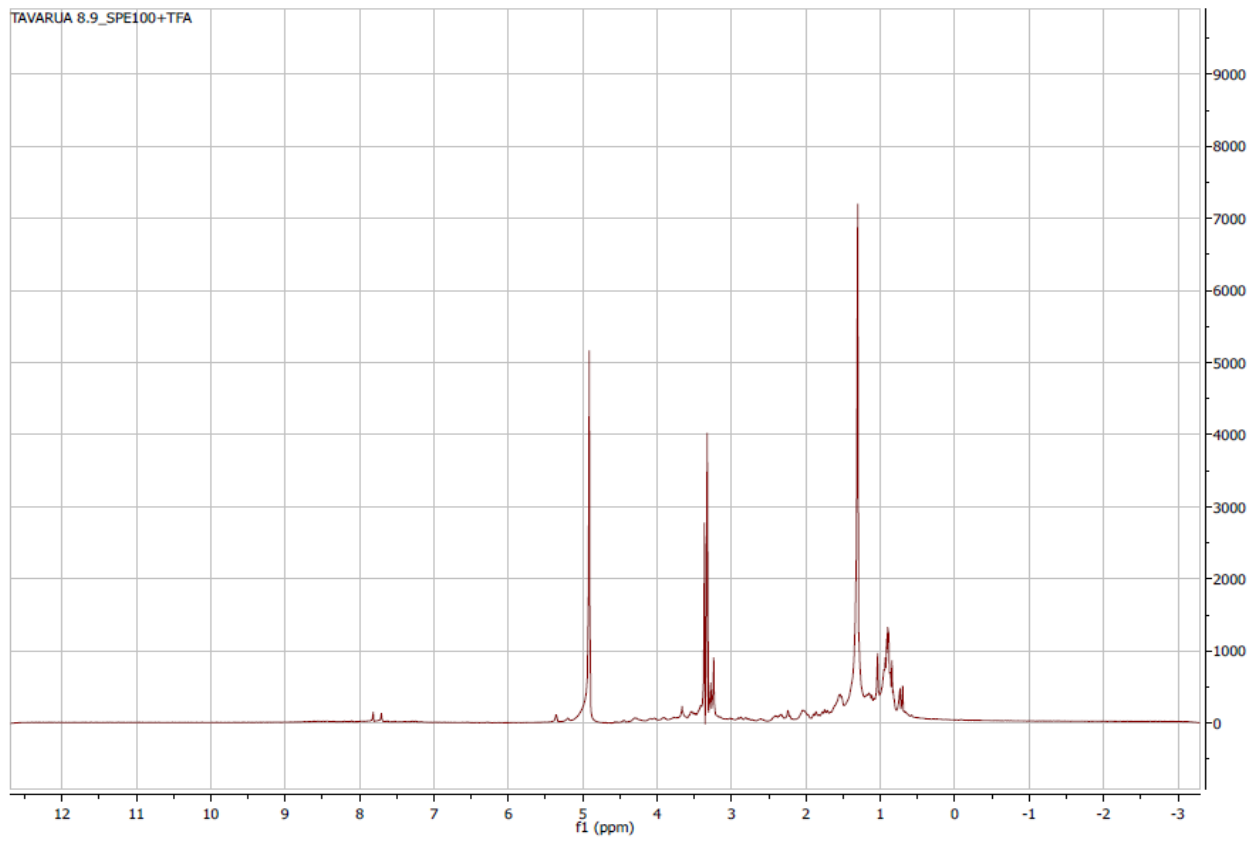
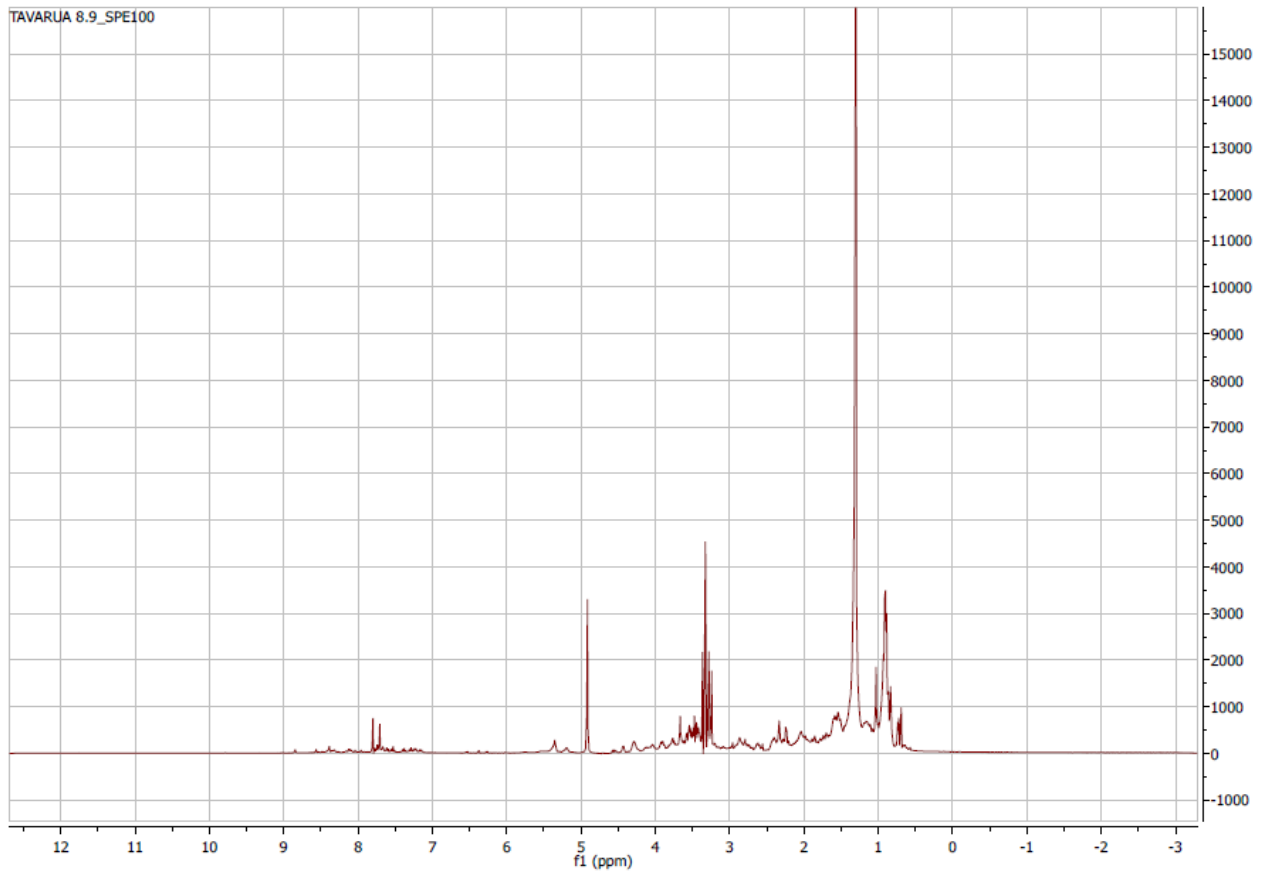


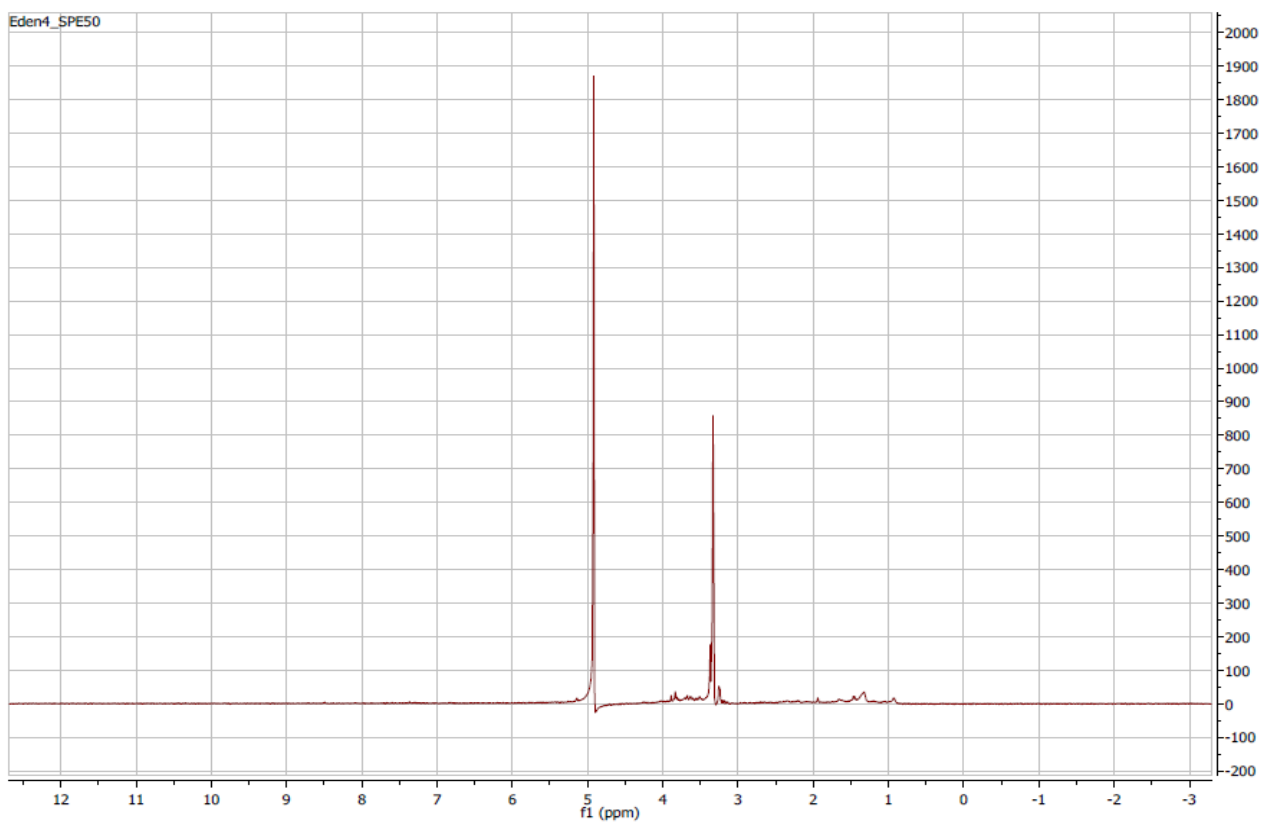
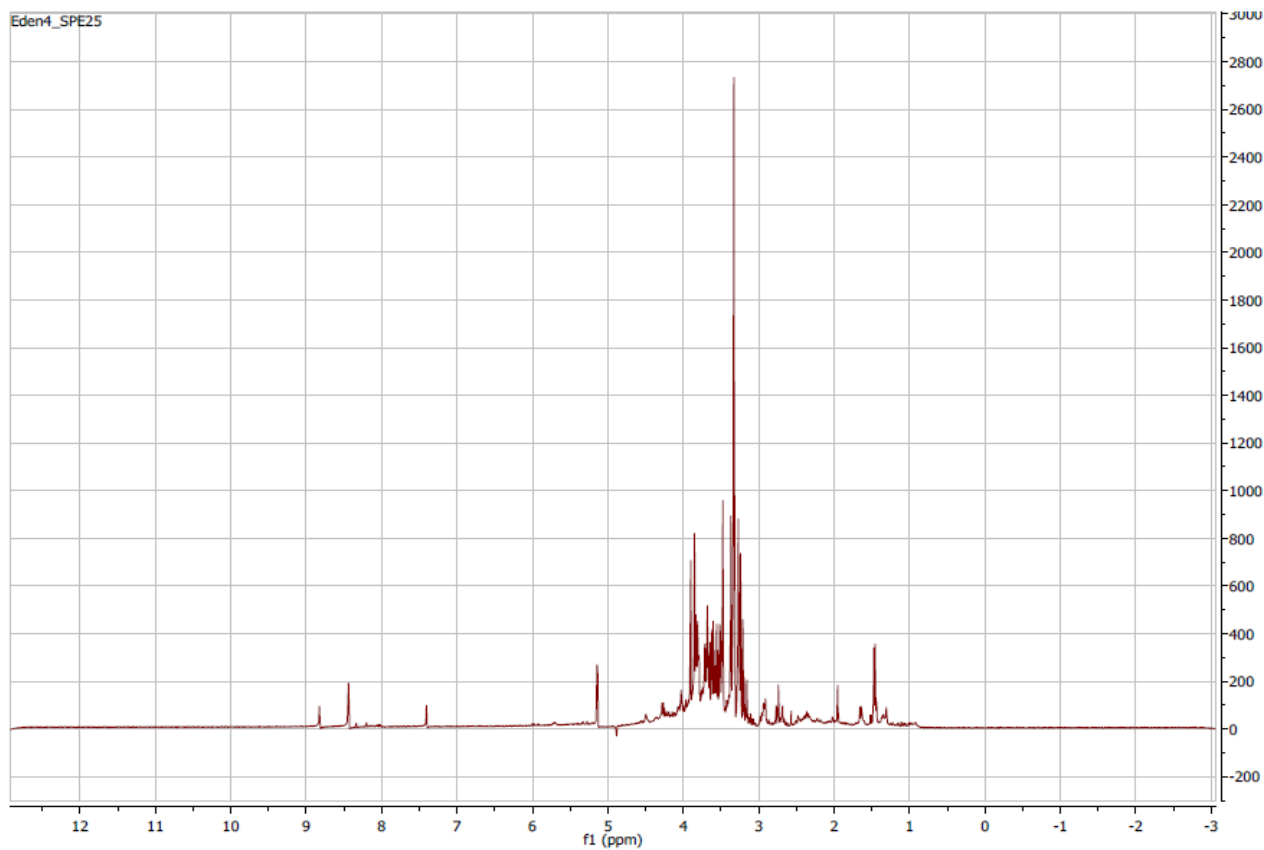


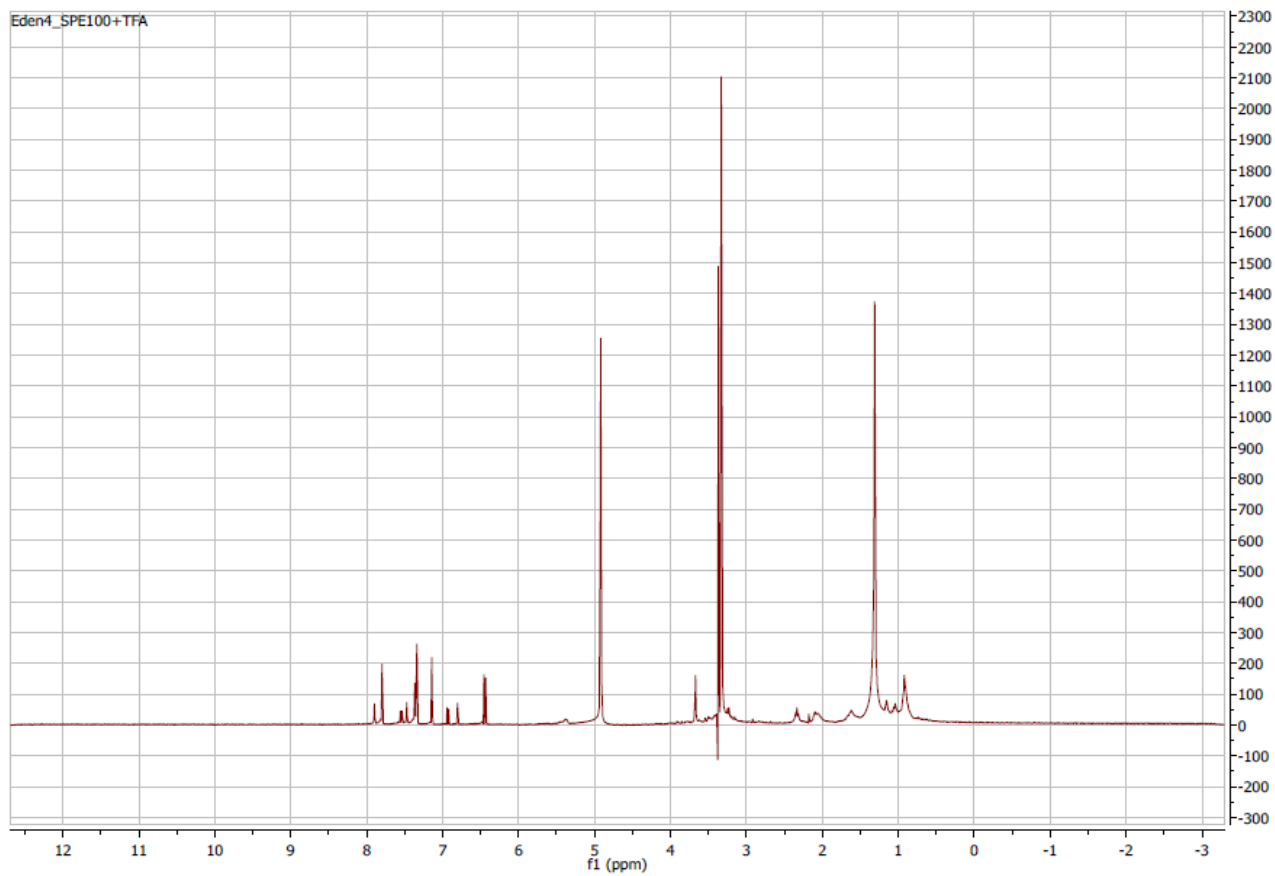
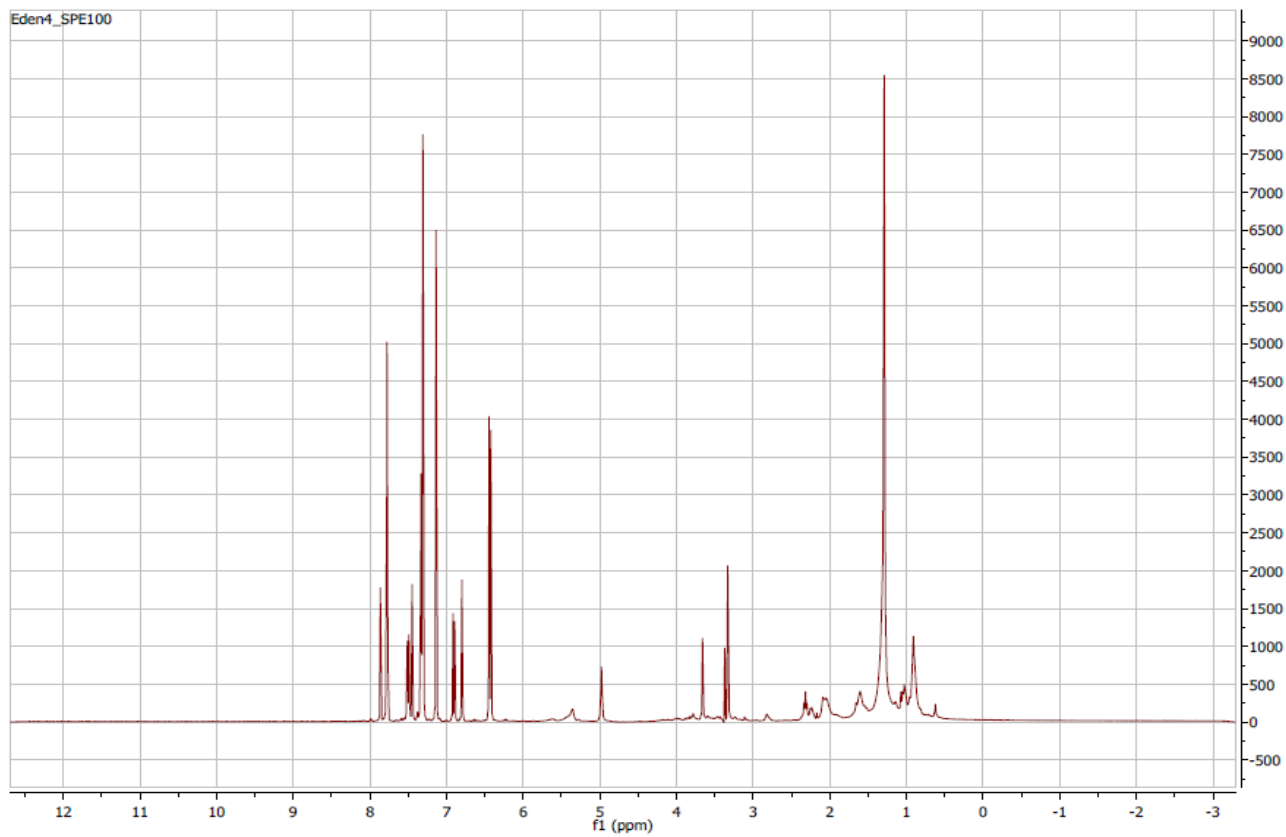


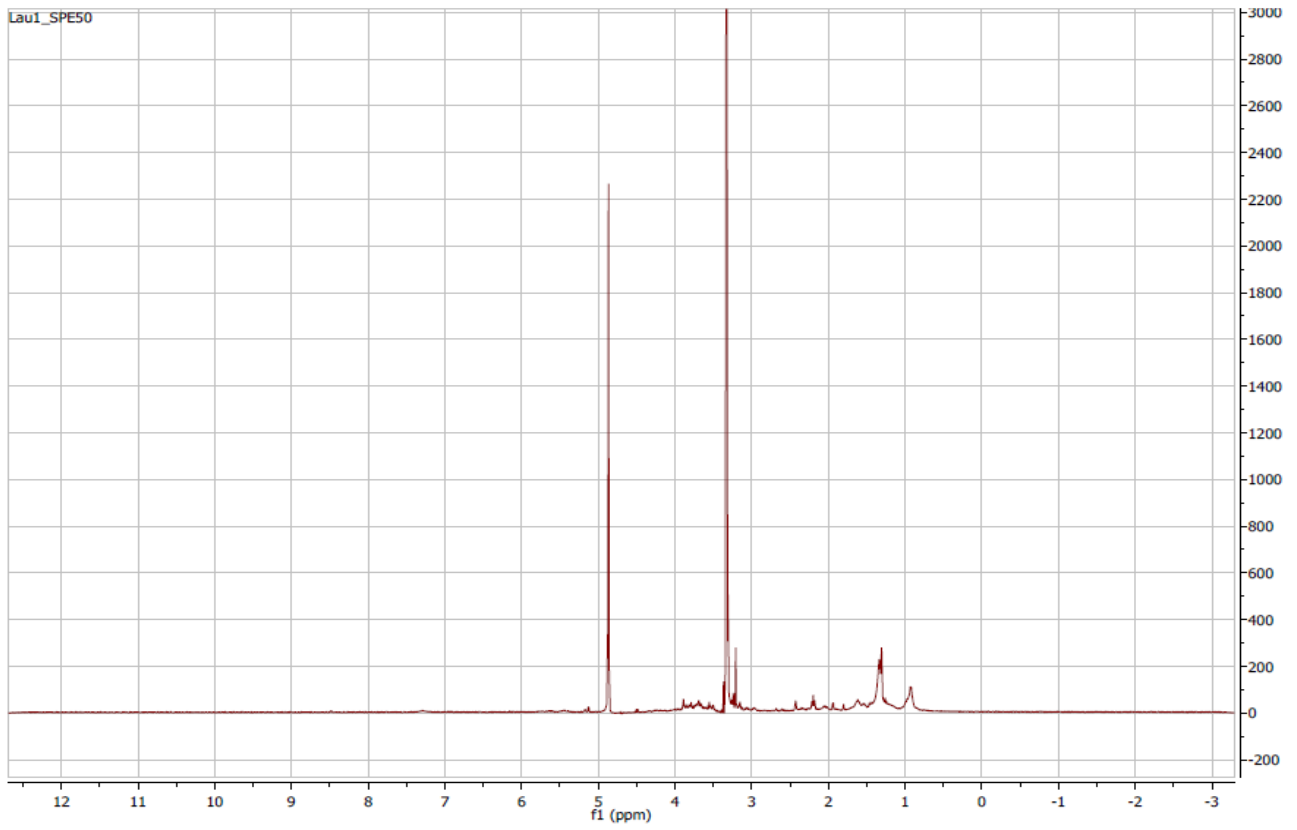
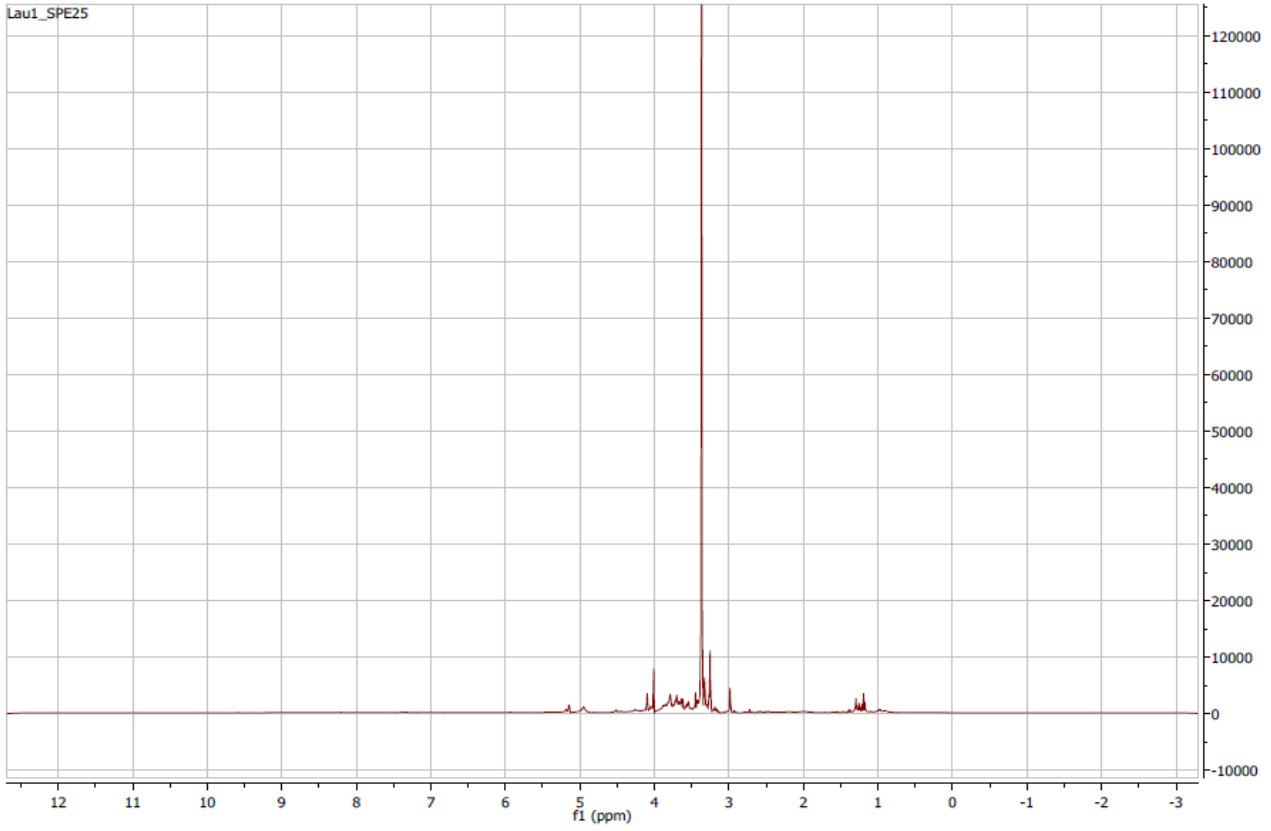


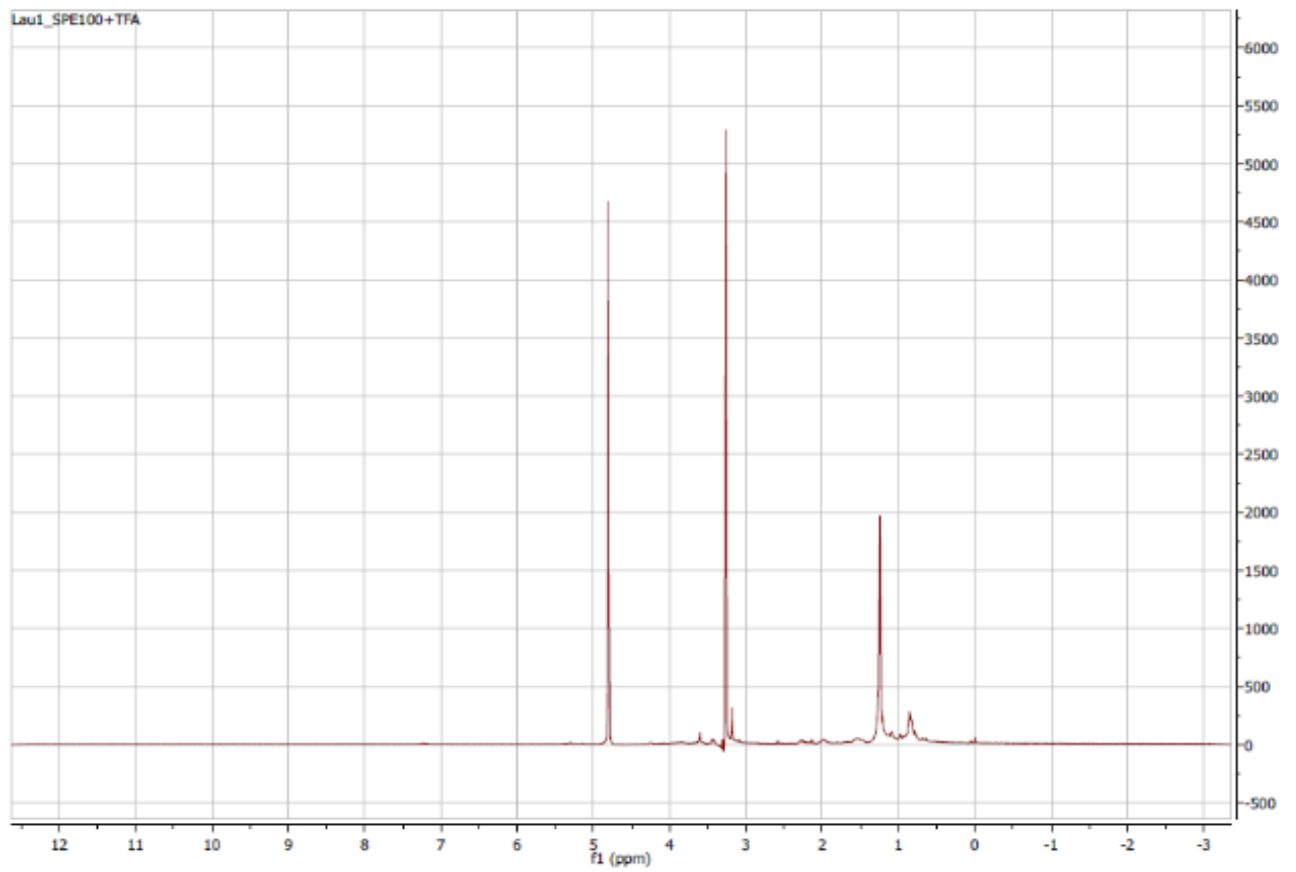
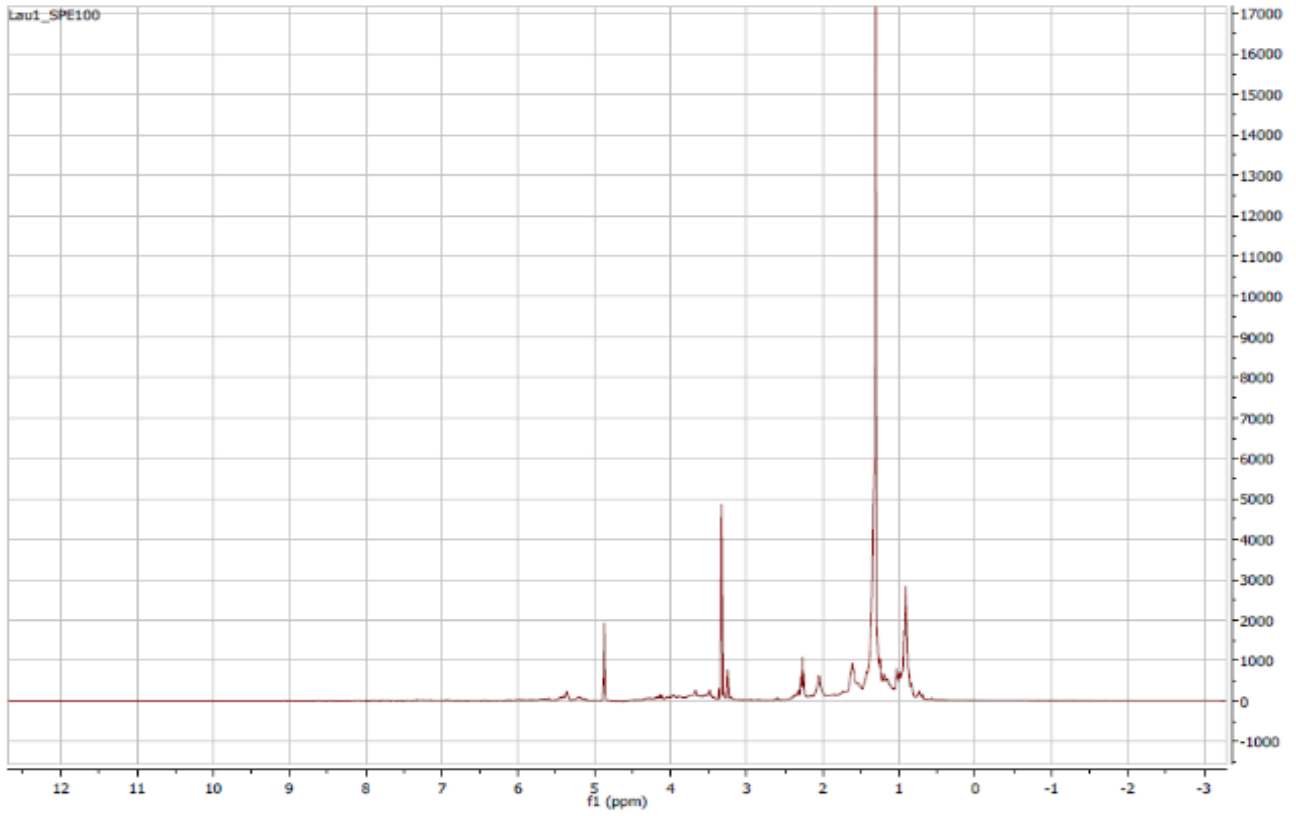


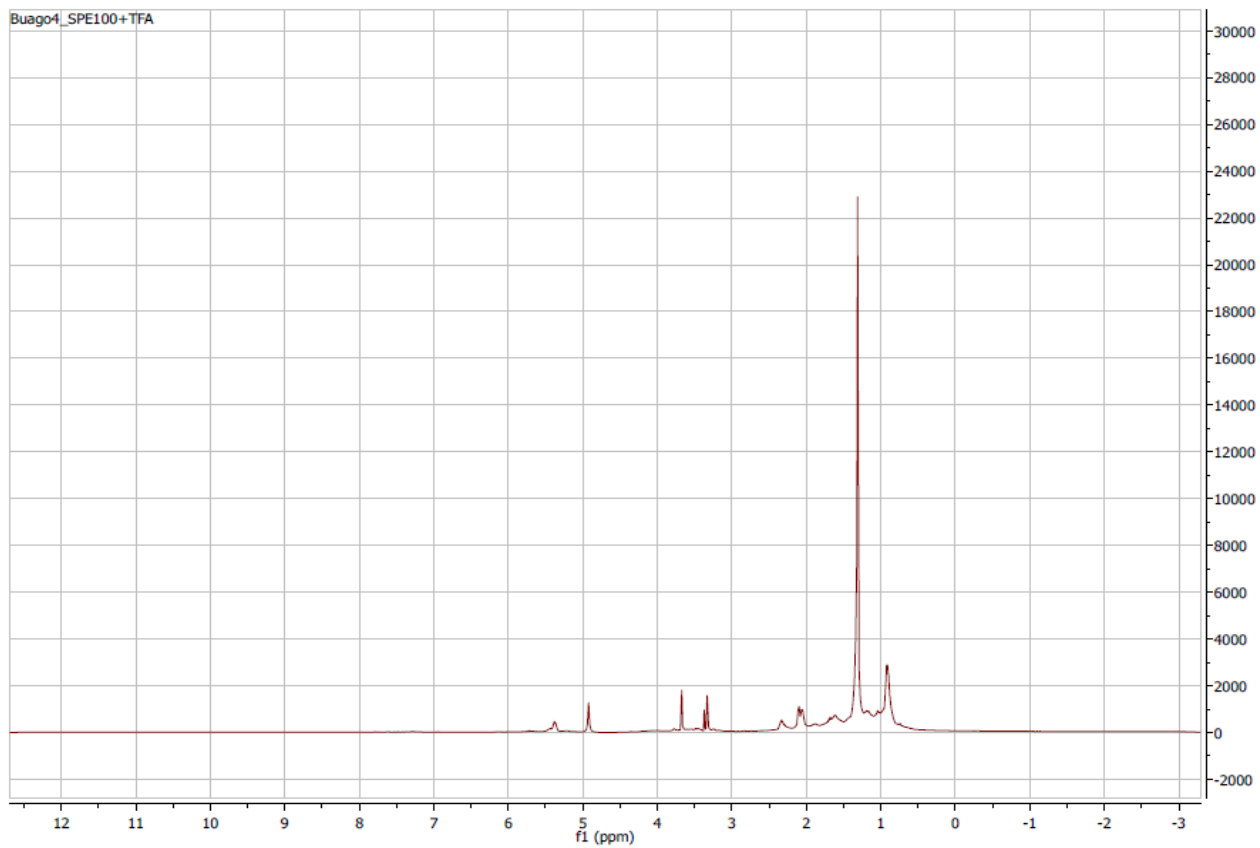
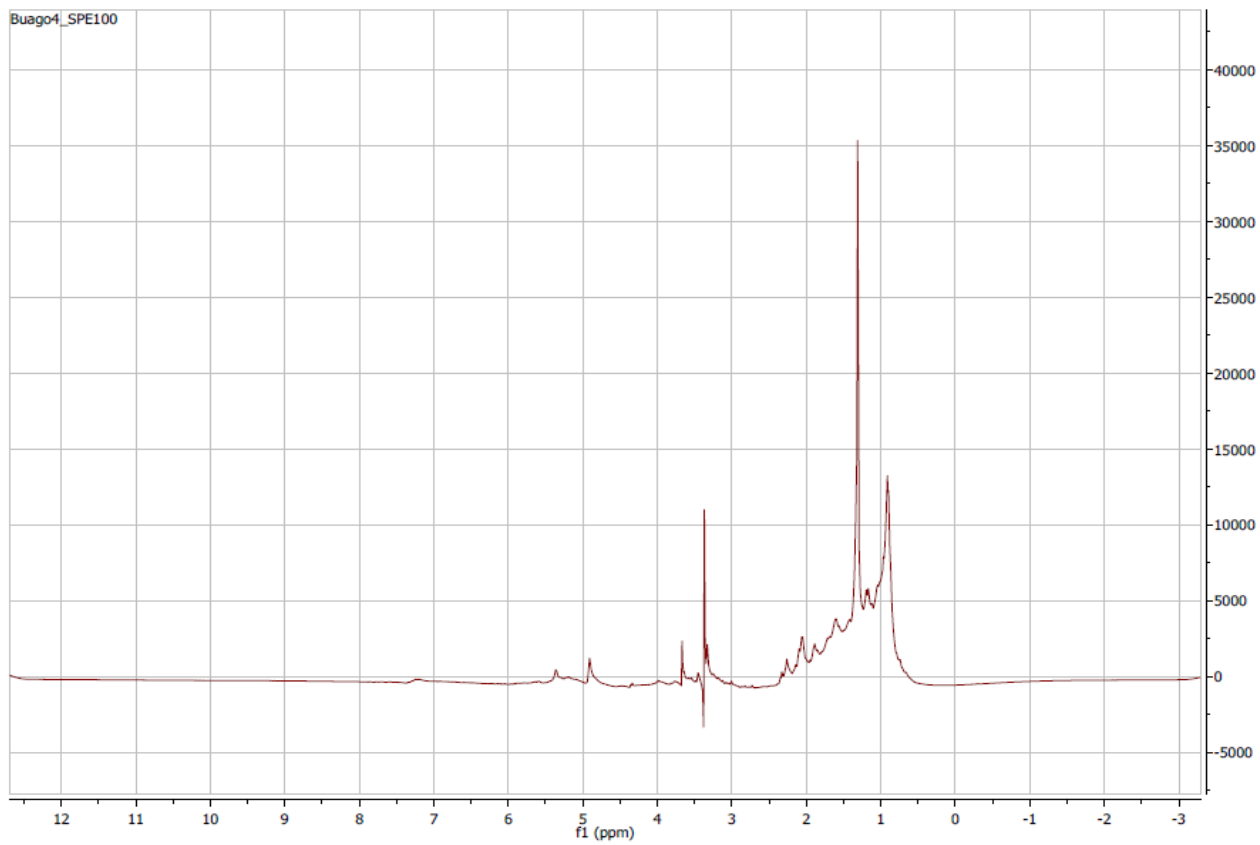


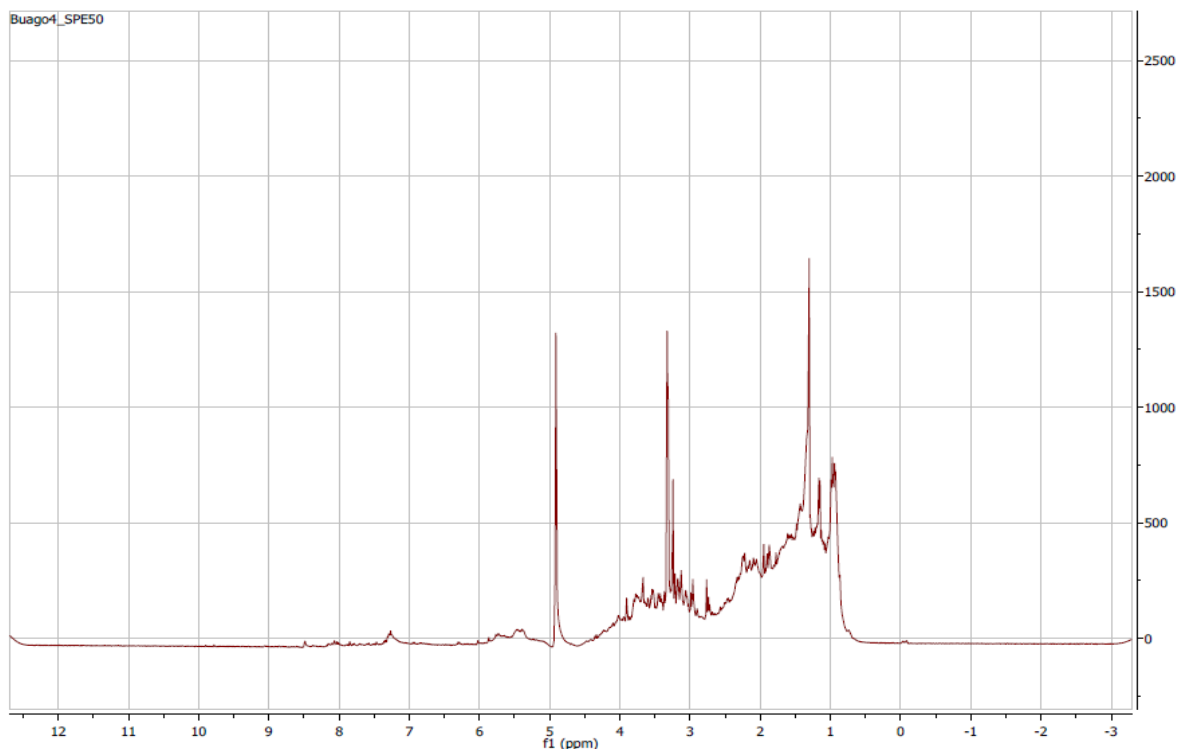












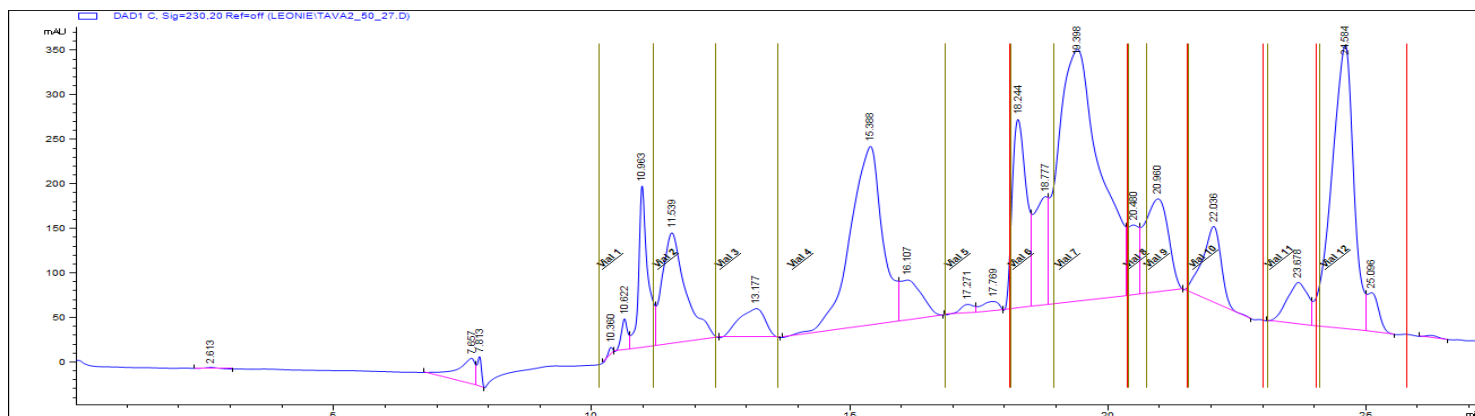


Figure S3. HPLC chromatogram of Tavarua 2 SPE 50% MeOH-water

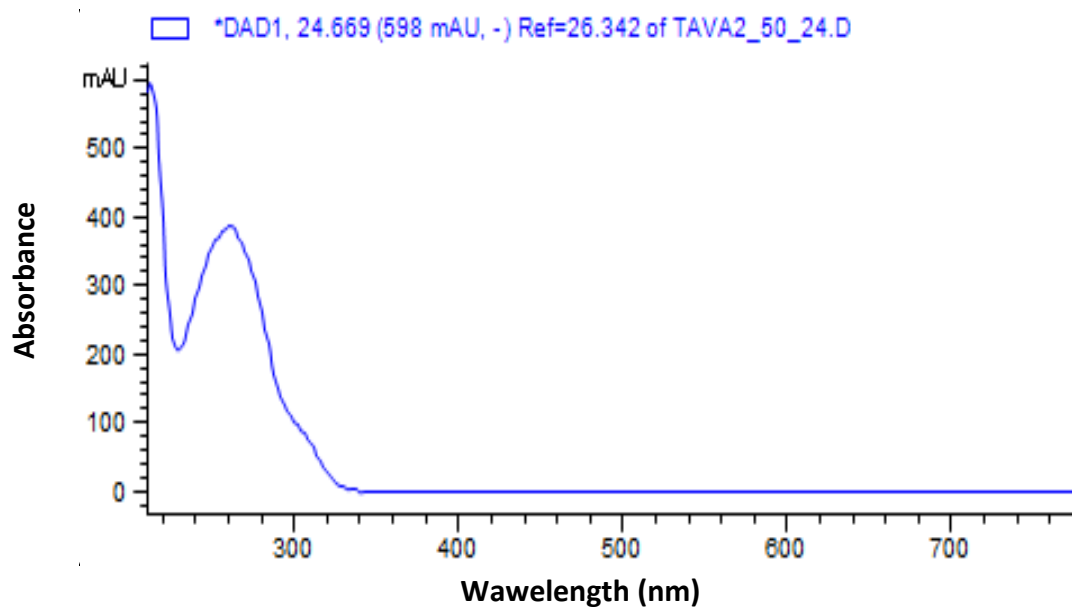


Figure S4. UV/Vis spectra of compound 1 in MeOH

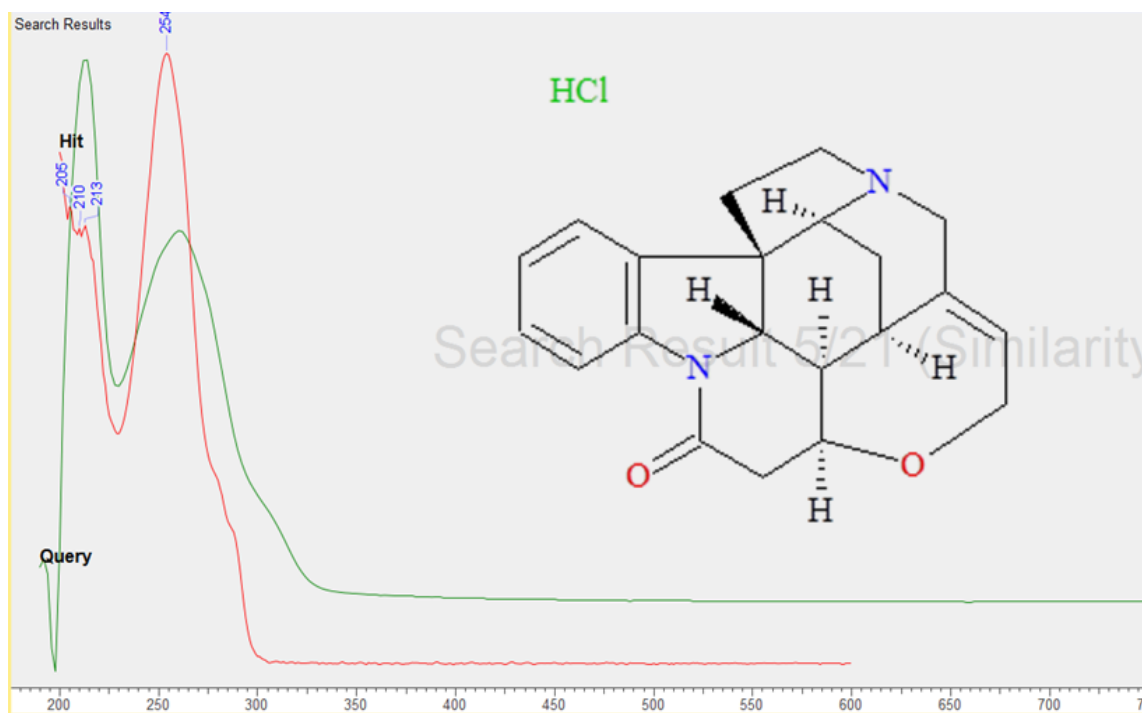
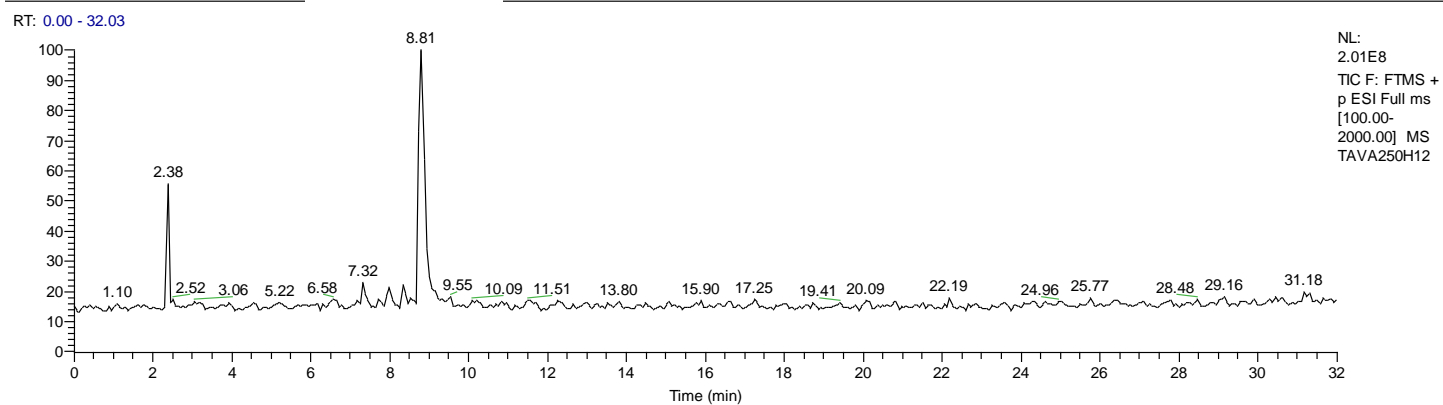


Figure S5. Search and matching of UV spectra of compound 1 (green) and an indole derivative (red) in ACD lab spectrus processor (ACD/Structure Elucidator, version 15.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2015)



TAVA250H12 #385-396 RT: 8.67-8.87 AV: 4 NL: 1.79E7
F: FTMS + p ESI Full ms [100.00-2000.00]

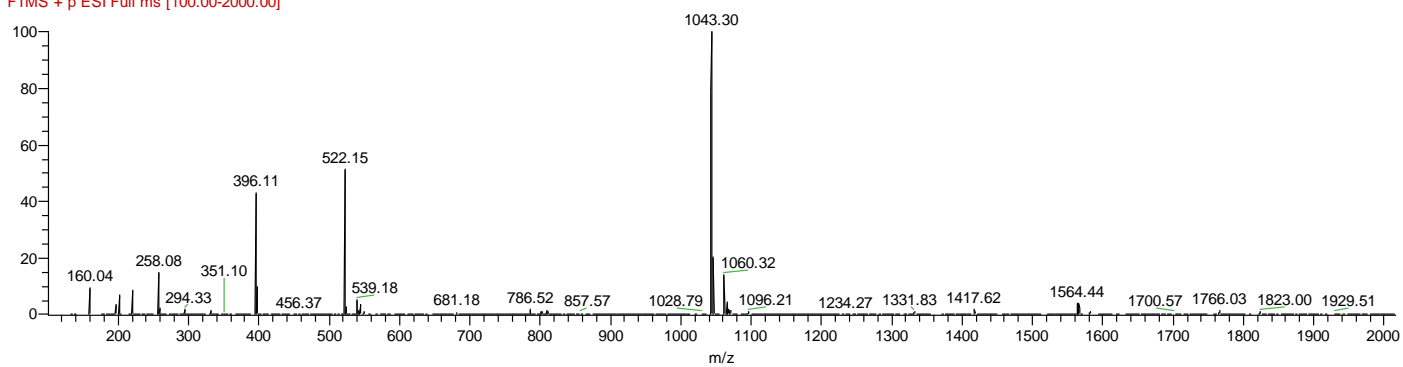


Figure S6. LC-HRESIMS spectrum of the new compound 1

KS94 #489 RT: 8.22 AV: 1 NL: 7.00E5
 F: FTMS + p ESI Full ms [100.00-2000.00]

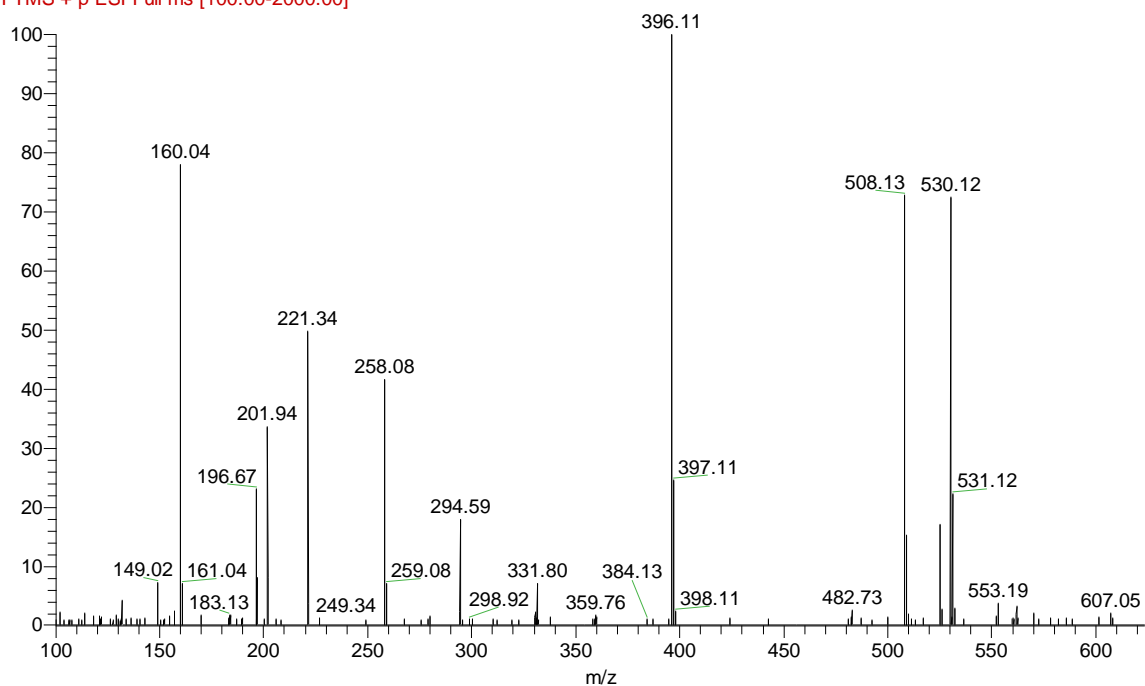


Figure S7. HRESIMS/MS spectrum of the new compound 1

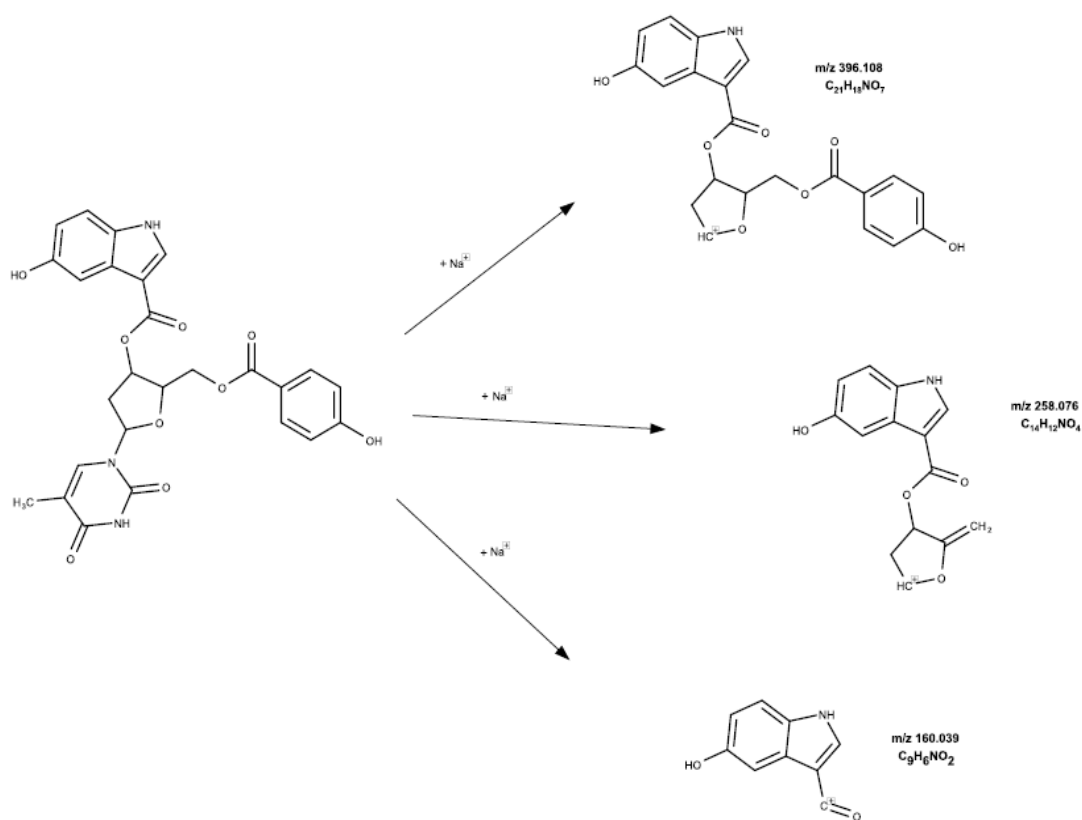


Figure S8. Fragmentation pattern of compound 1

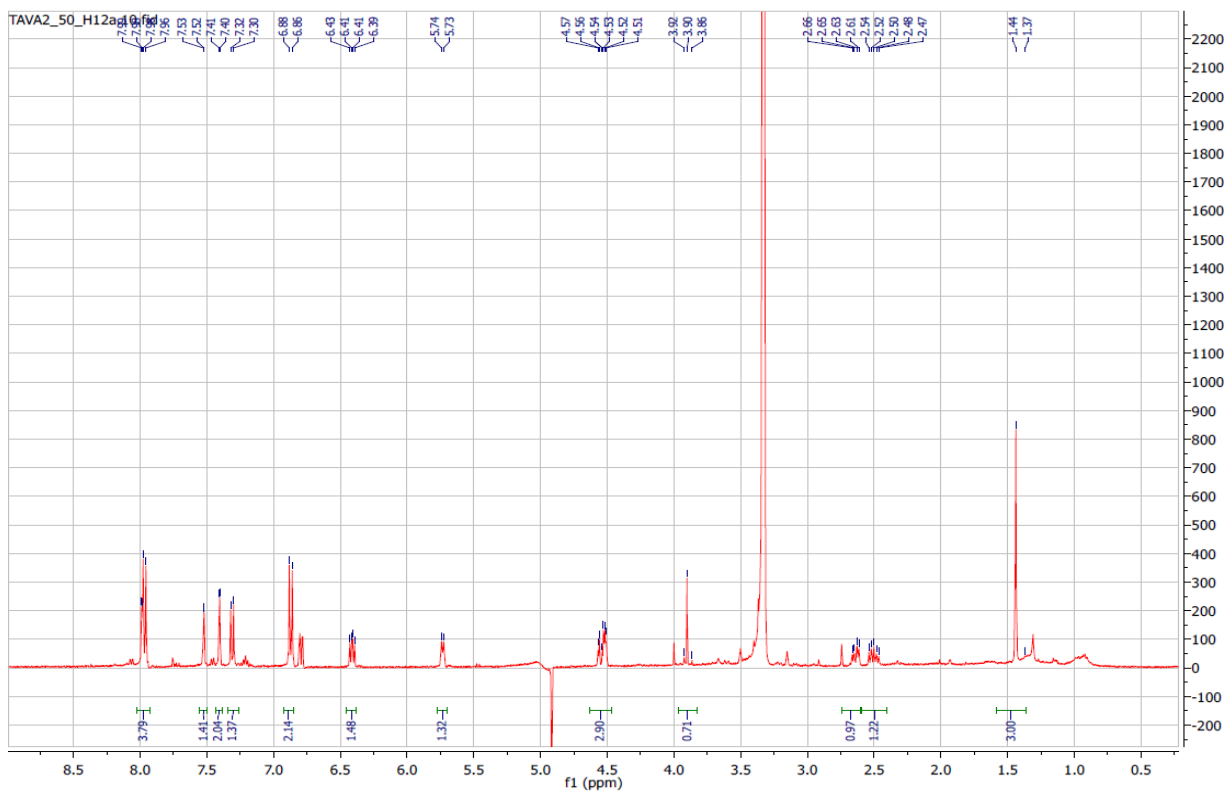


Figure S9. ^1H NMR spectrum of compound **1** in CD_3OD at 600 MHz

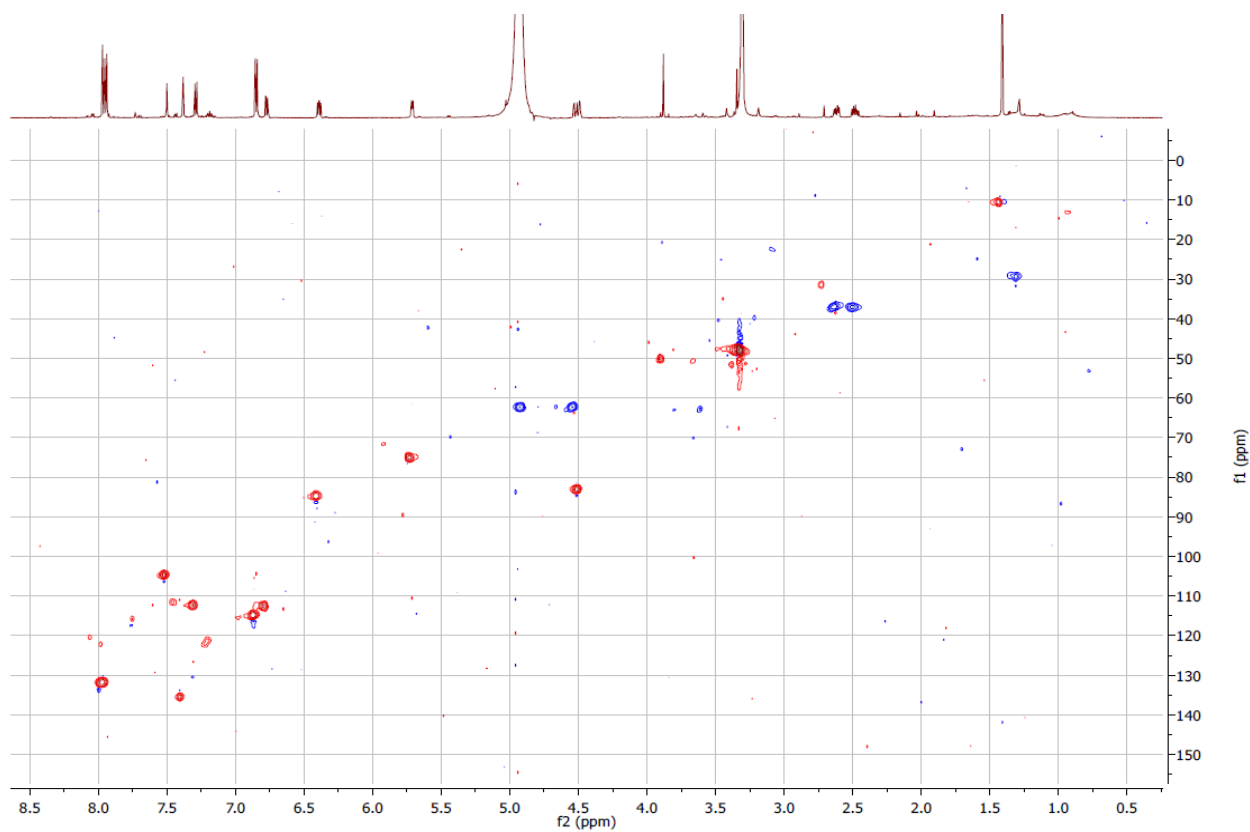


Figure S10. HSQC spectrum of compound **1** in CD_3OD at 600 MHz

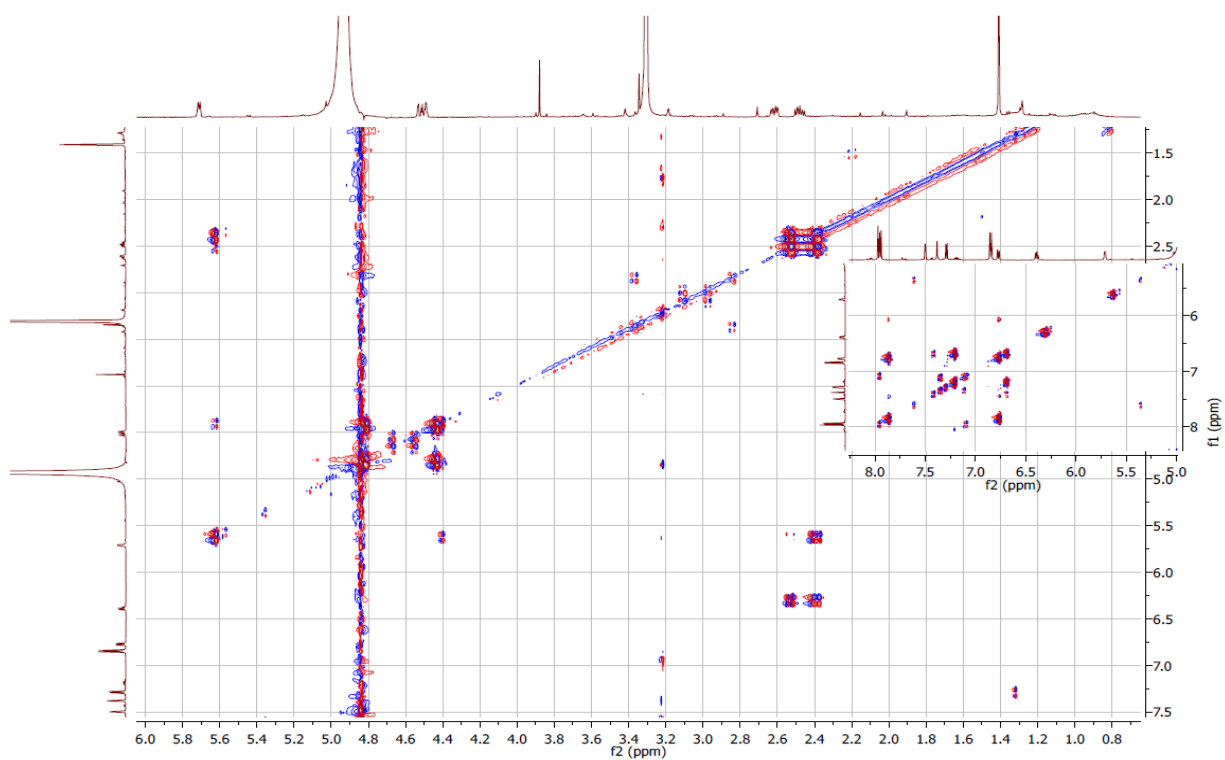


Figure S11. COSY spectrum of compound **1** in CD₃OD at 600 MHz

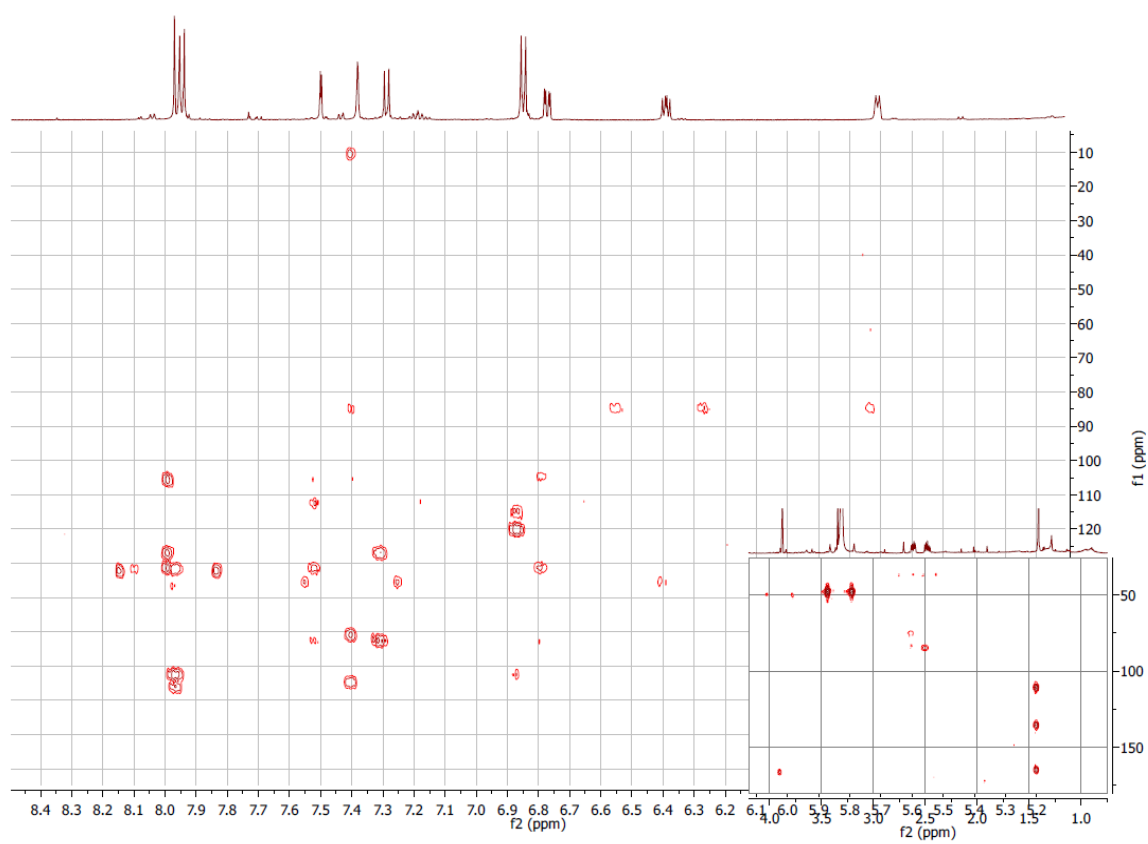


Figure S12. HMBC spectrum of compound **1** in CD₃OD at 600 MHz

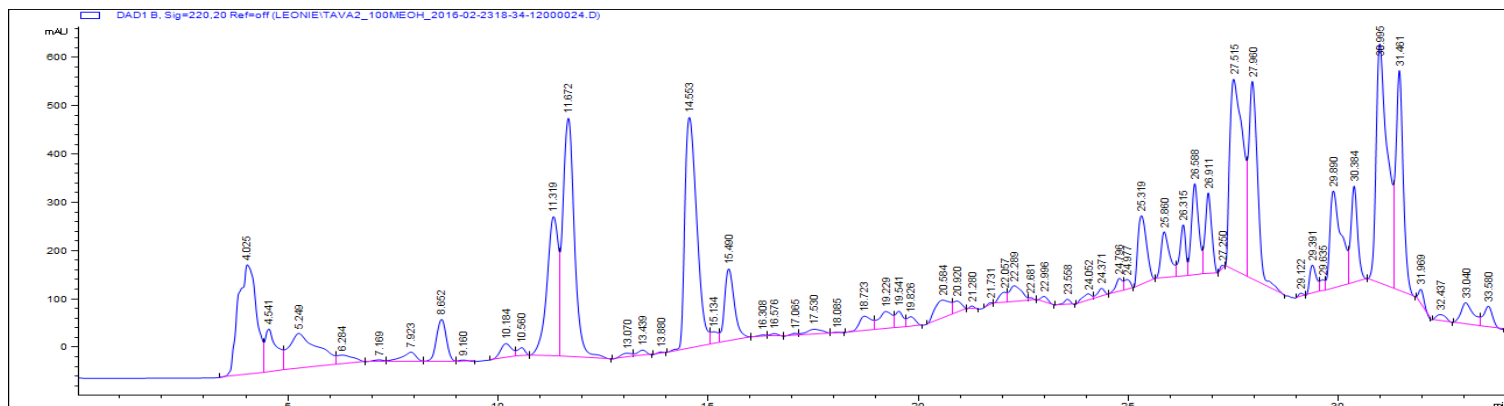


Figure S13. HPLC chromatogram of Tavarua 2 SPE 100% MeOH

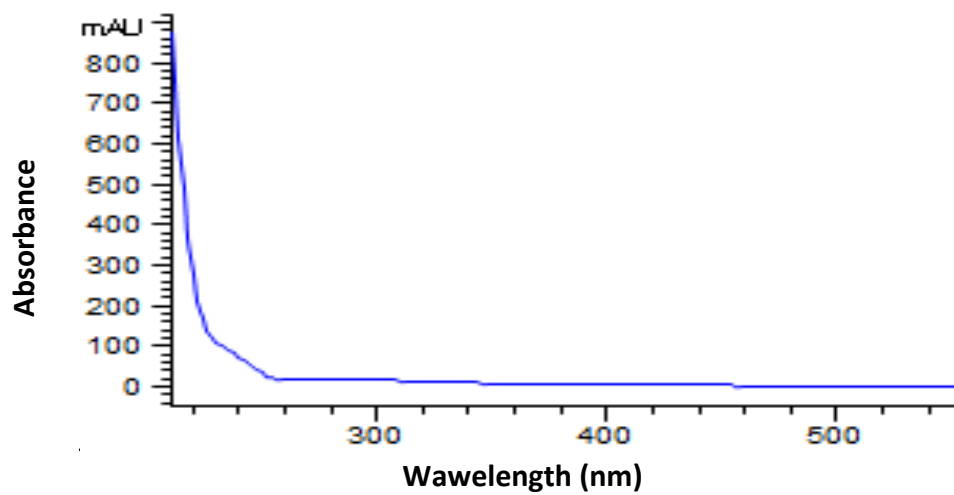


Figure S14. UV/Vis spectra of compound 2 in MeOH

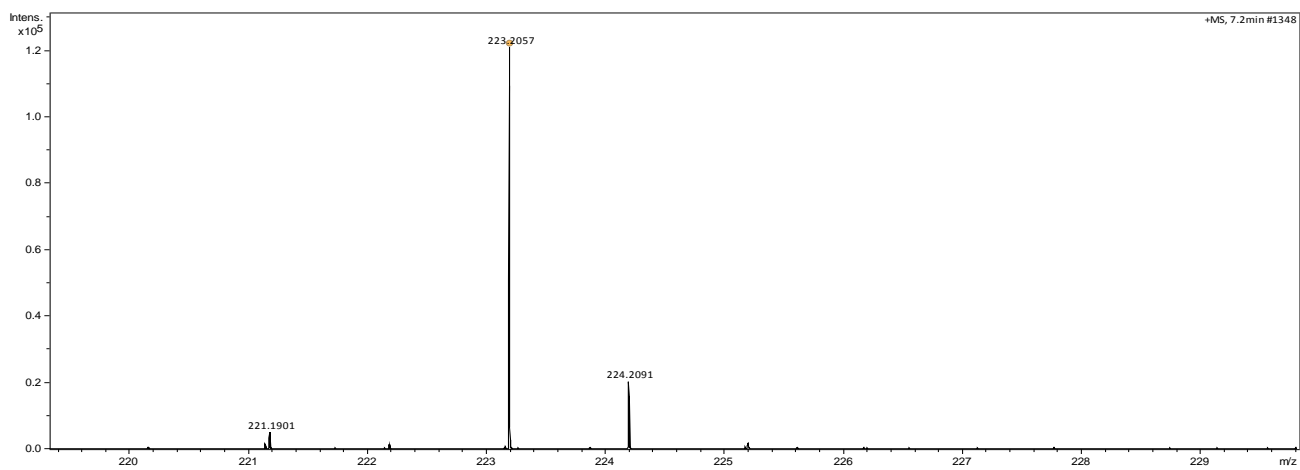
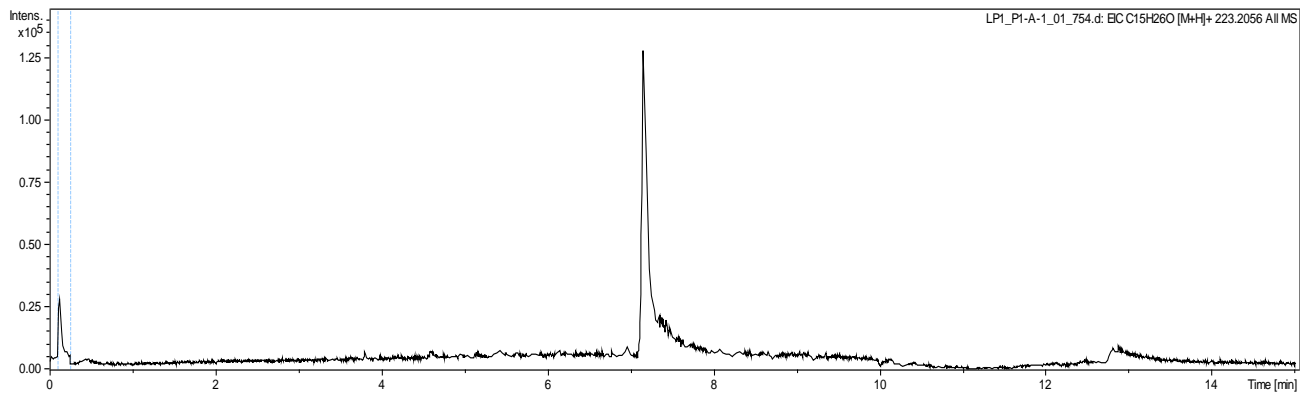


Figure S15. LC-HRESIMS spectrum of compound **2**

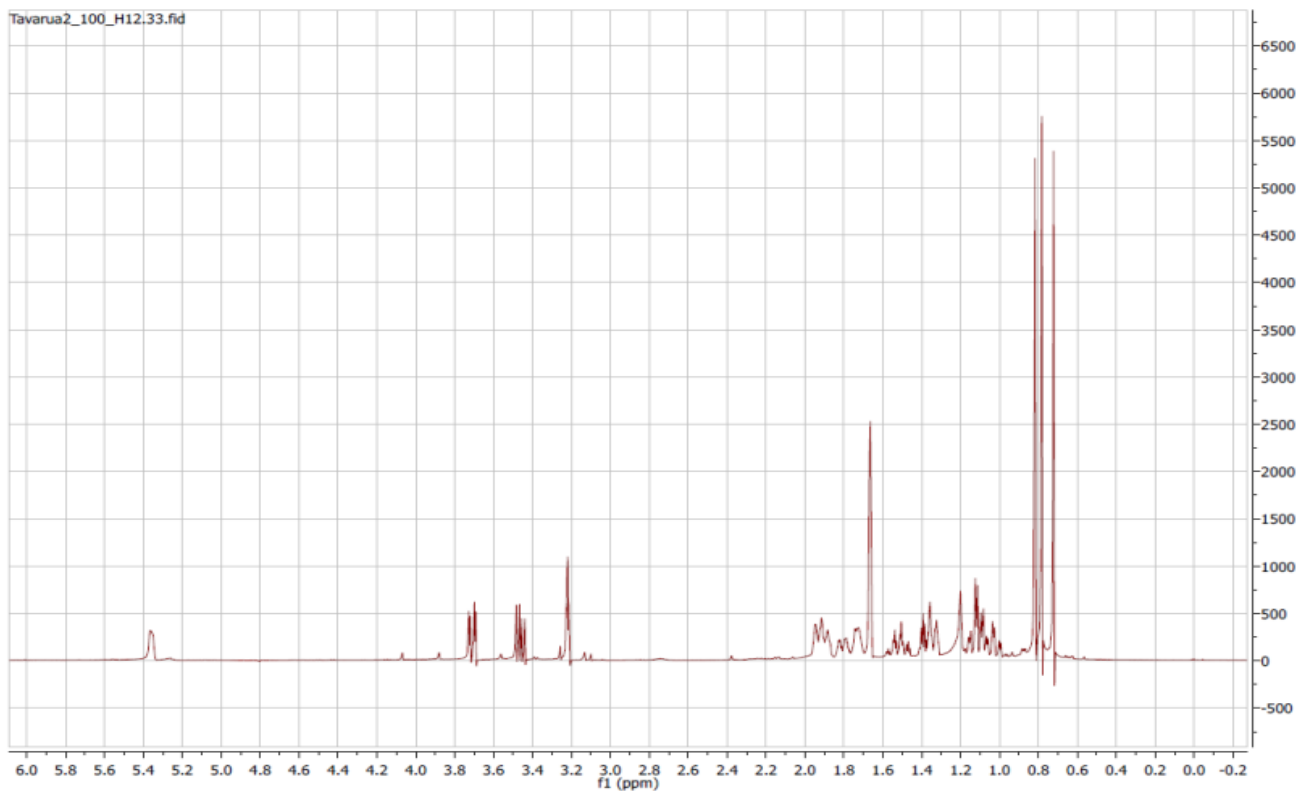


Figure S16. ^1H NMR spectrum of compound **2** in CD_3OD at 600 MHz

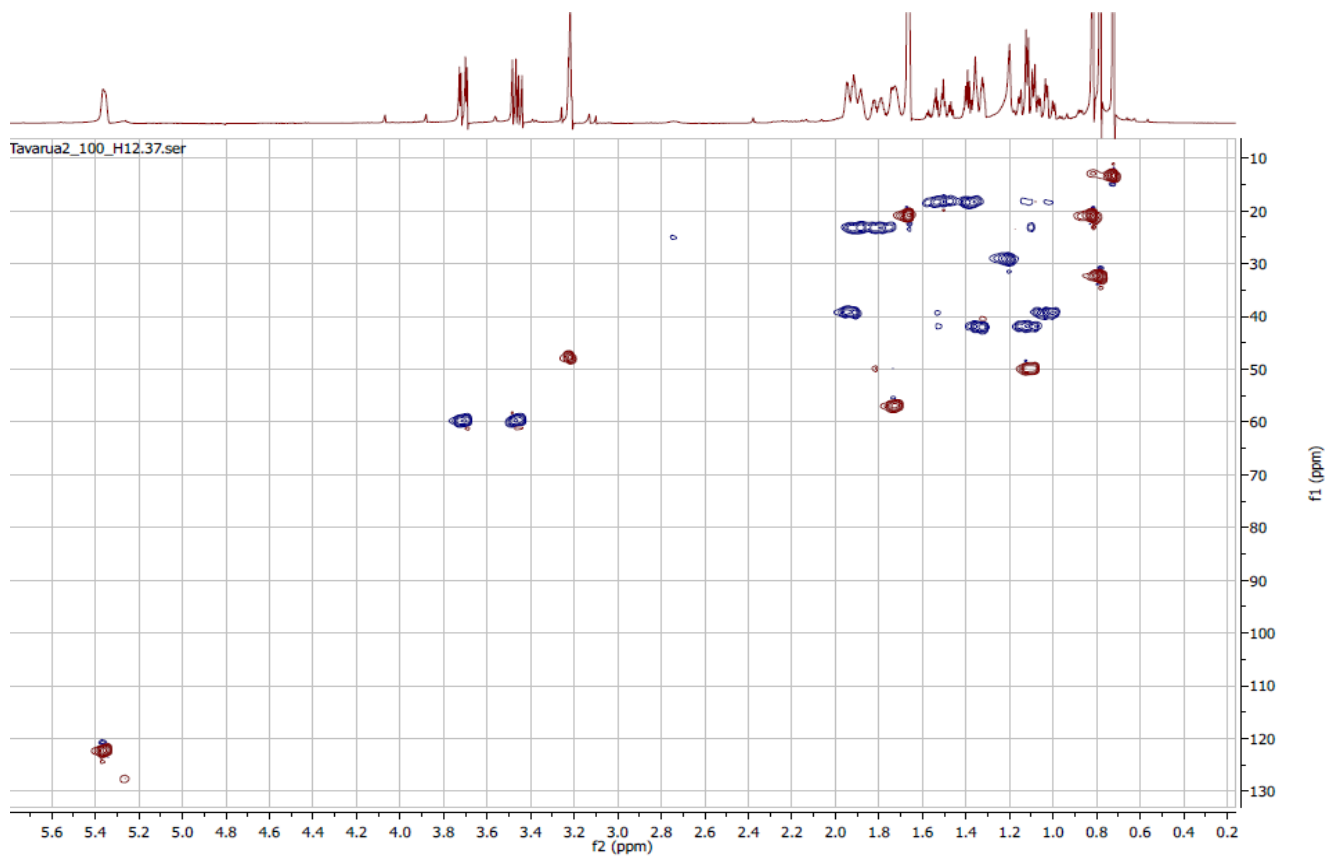


Figure S17. HSQC spectrum of compound **2** in CD_3OD at 600 MHz

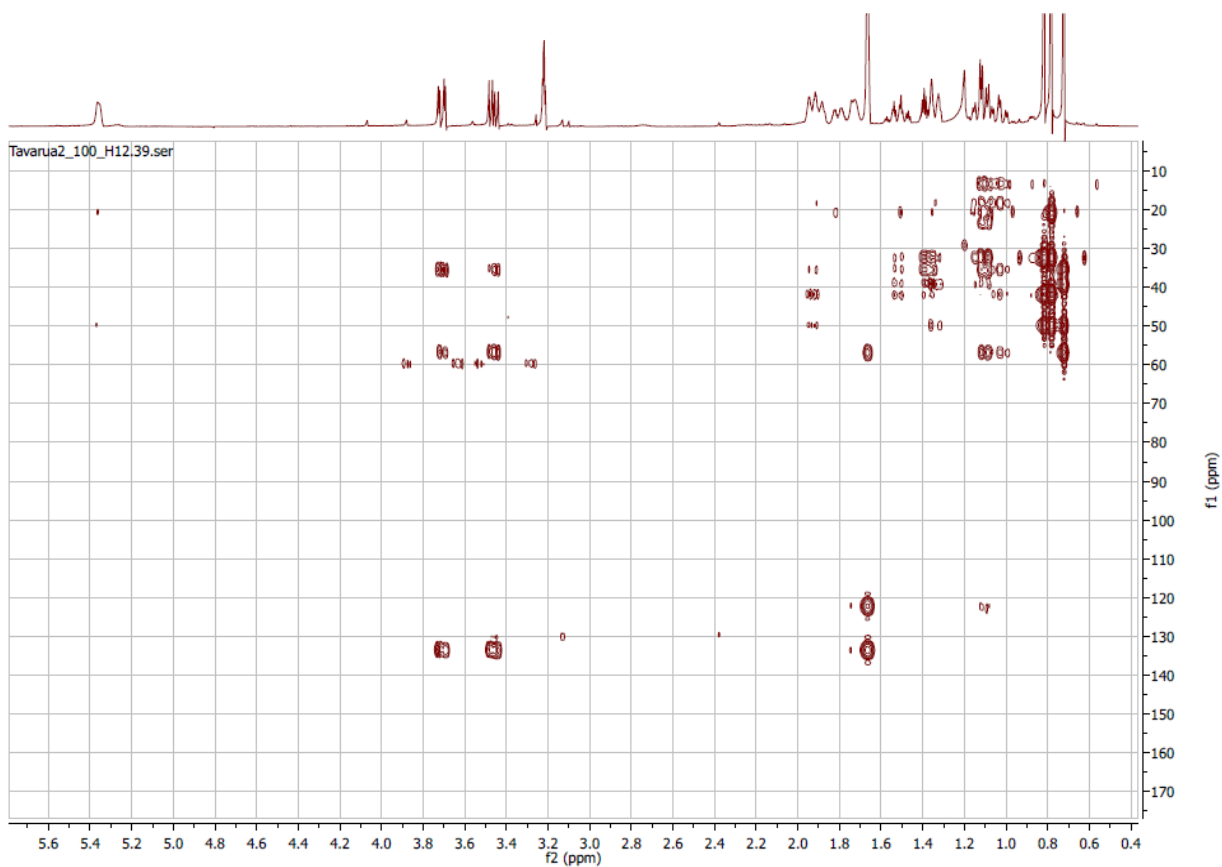


Figure S18. HMBC spectrum of compound **2** in CD₃OD at 600 MHz

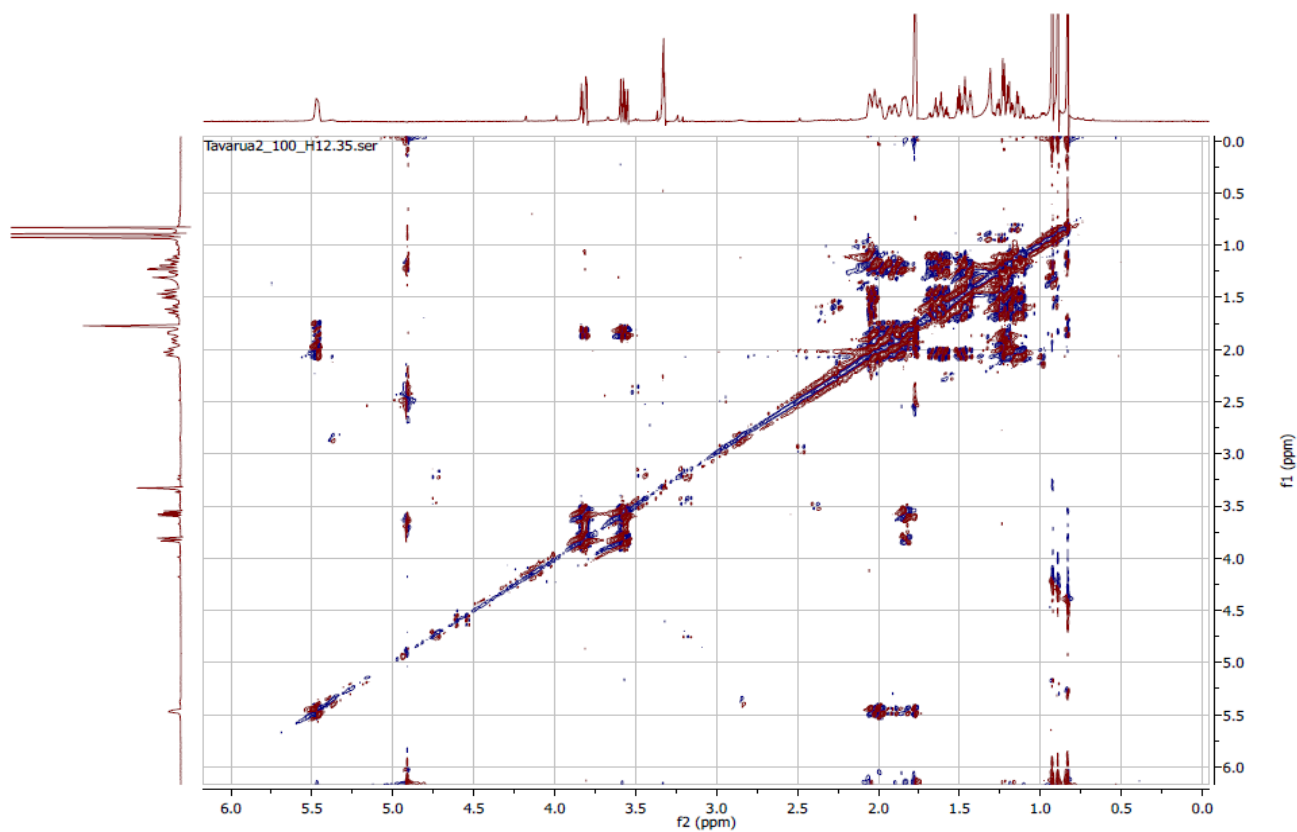


Figure S19. COSY spectrum of compound **2** in CD₃OD at 600 MHz

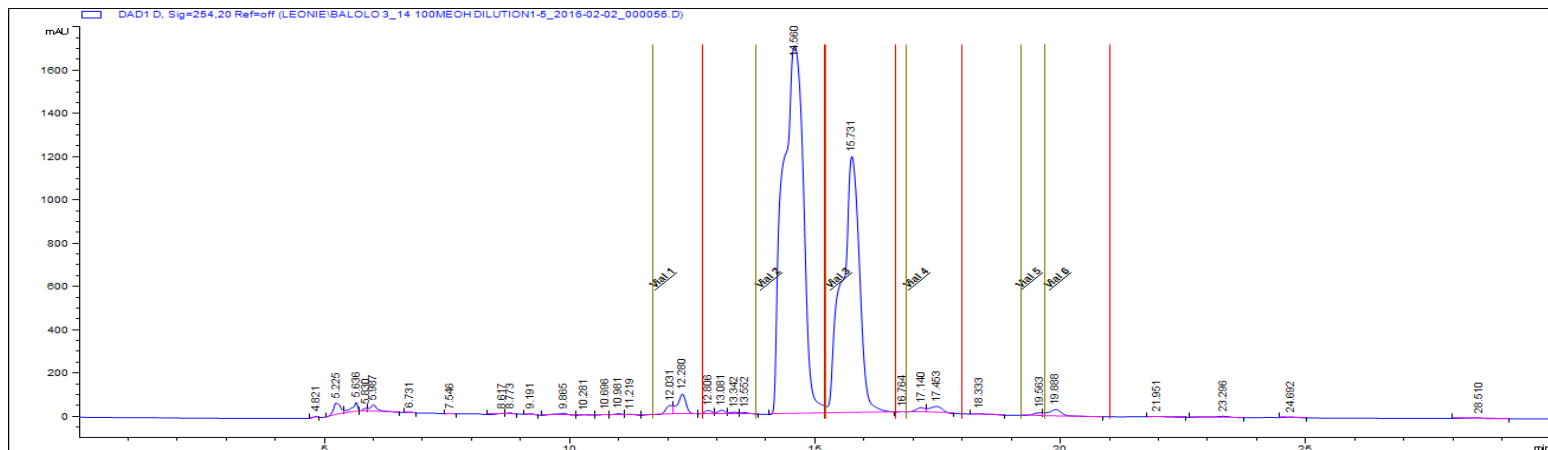


Figure S20. HPLC chromatogram of Balolo 3 SPE 100% MeOH

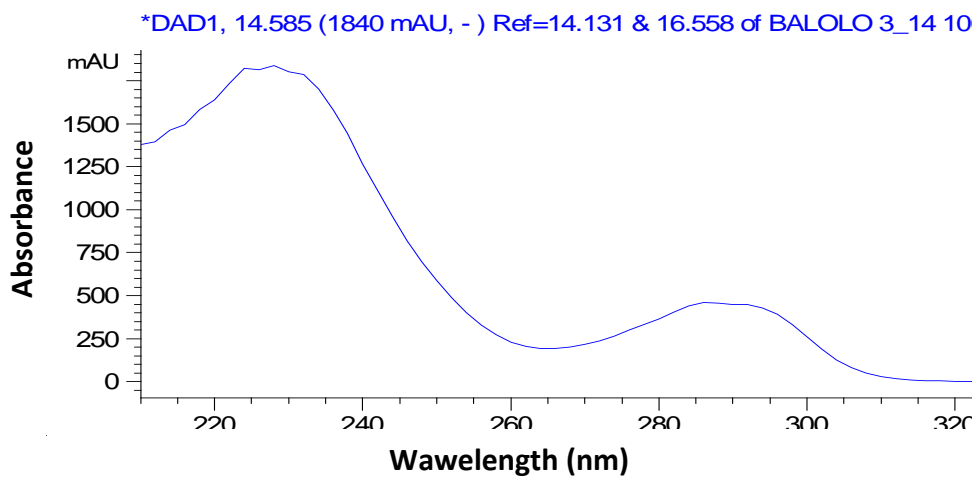


Figure S21. UV/Vis spectra of compound 3 in MeOH

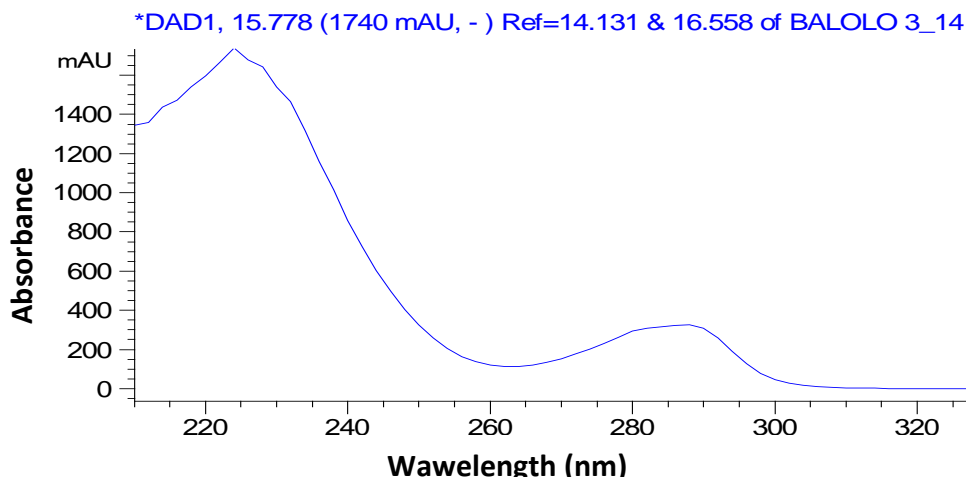
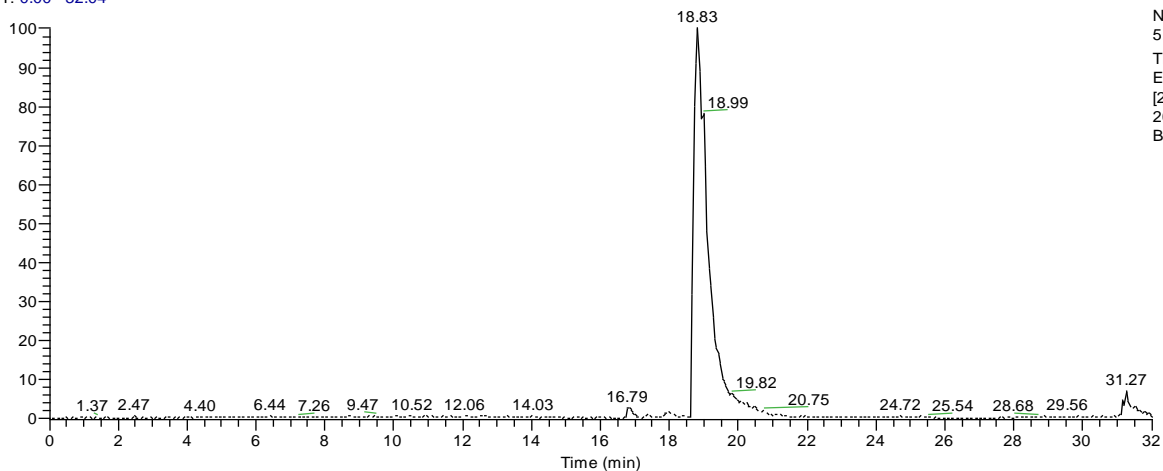
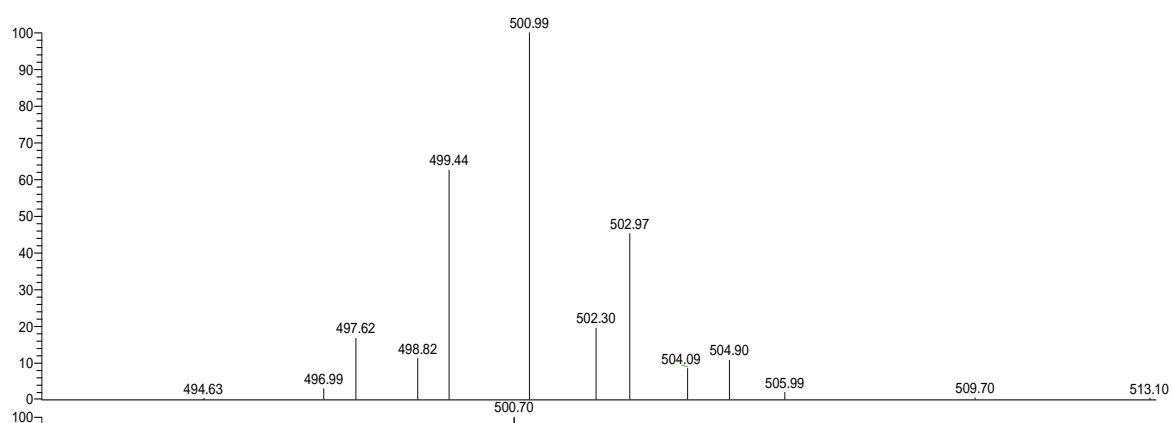


Figure S22. UV/Vis spectra of compound 4 in MeOH

RT: 0.00 - 32.04



NL: 5.83E5
TIC F: ITMS - c
ESI Full ms
[200.00-
2000.00] MS
B314-100-H2



NL: 1.11E5
B314-100-
H2#1029 RT:
18.88 AV: 1 F:
ITMS - c ESI Full
ms
[200.00-2000.00]

NL: 3.76E5

Figure S23. LC-ESIMS spectrum of the compound 3

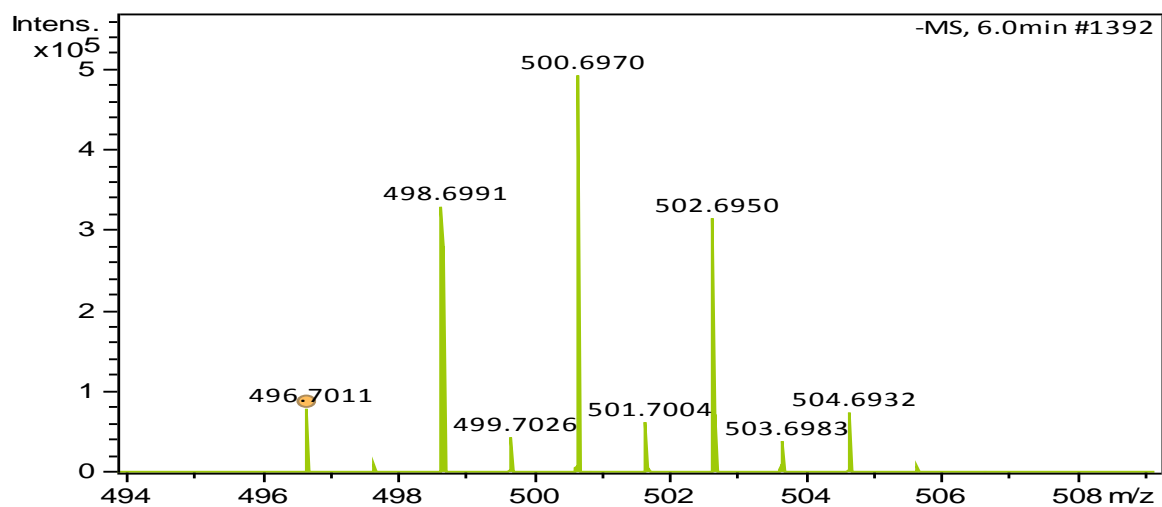


Figure S24. HRTOFMS spectrum of compound 3

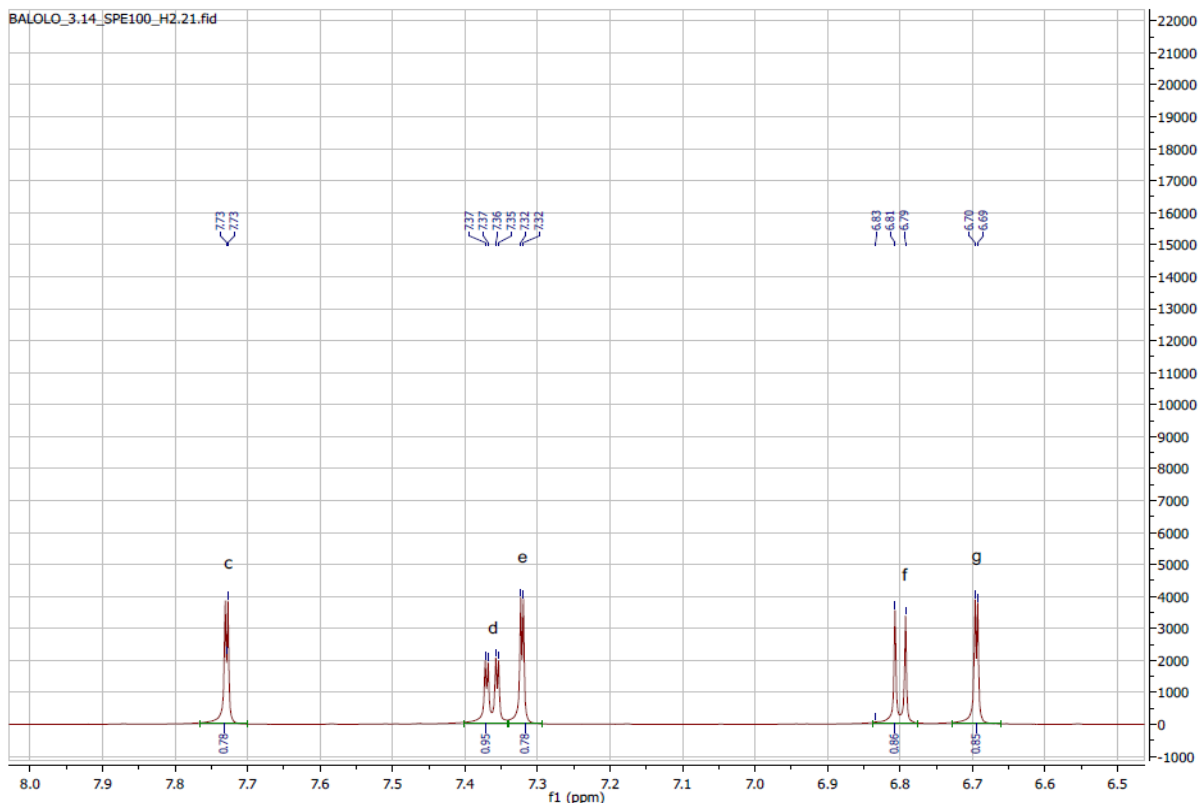


Figure S25. ^1H NMR spectrum of compound **3** in CD_3OD at 600 MHz

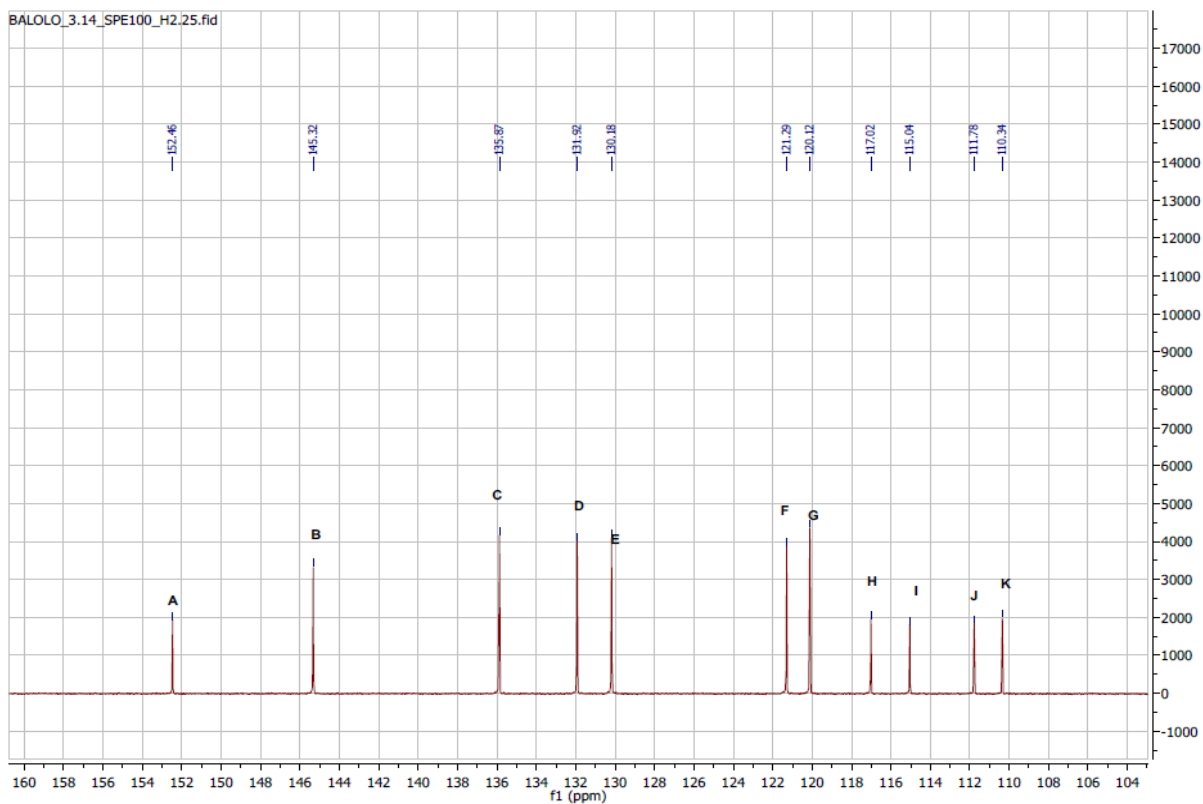


Figure S26. ^{13}C NMR spectrum of compound **3** in CD_3OD at 600 MHz

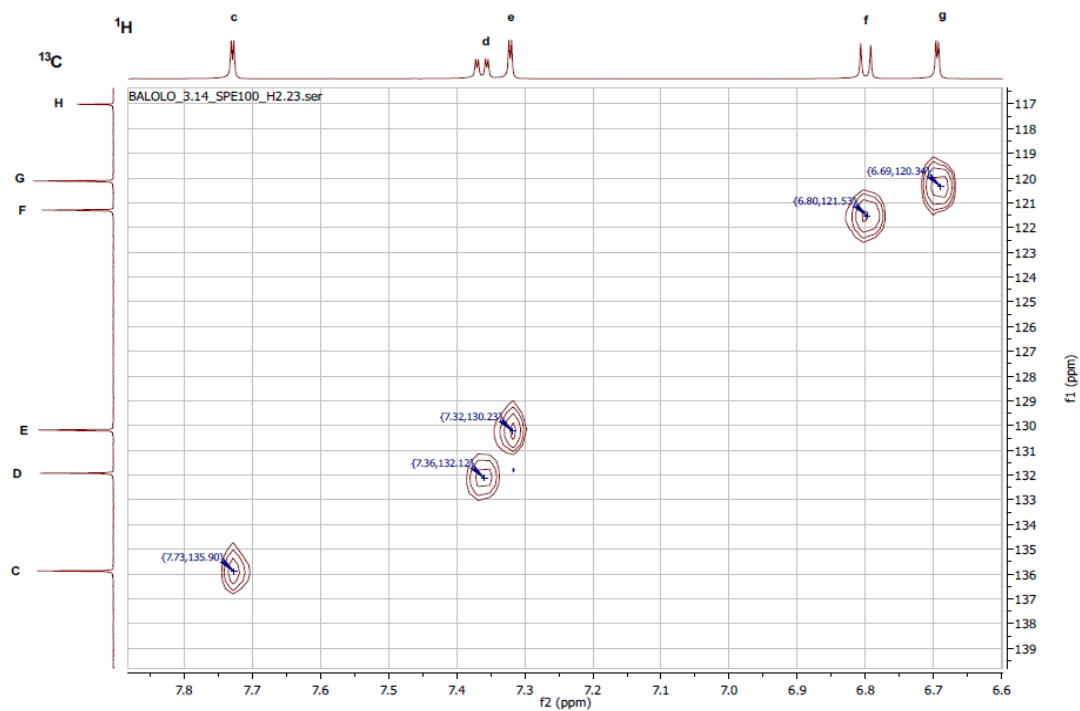


Figure S27. HSQC spectrum of compound 3 in CD_3OD at 600 MHz

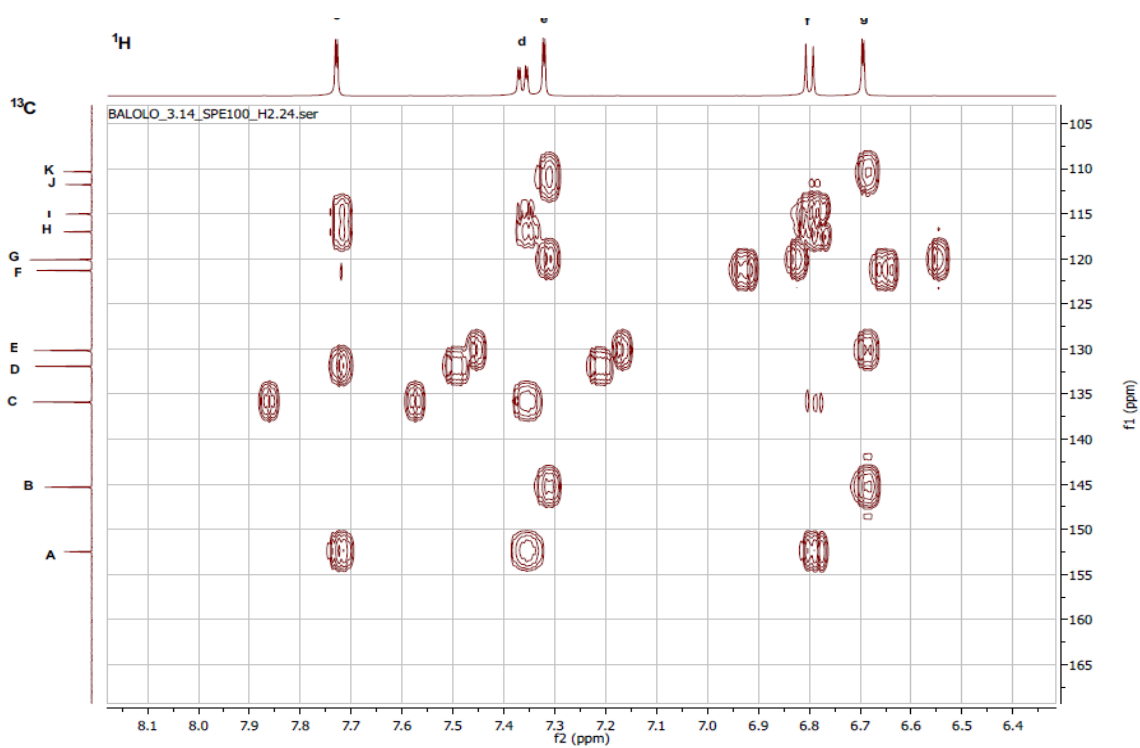


Figure S28. HMBC spectrum of compound 3 in CD_3OD at 600 MHz

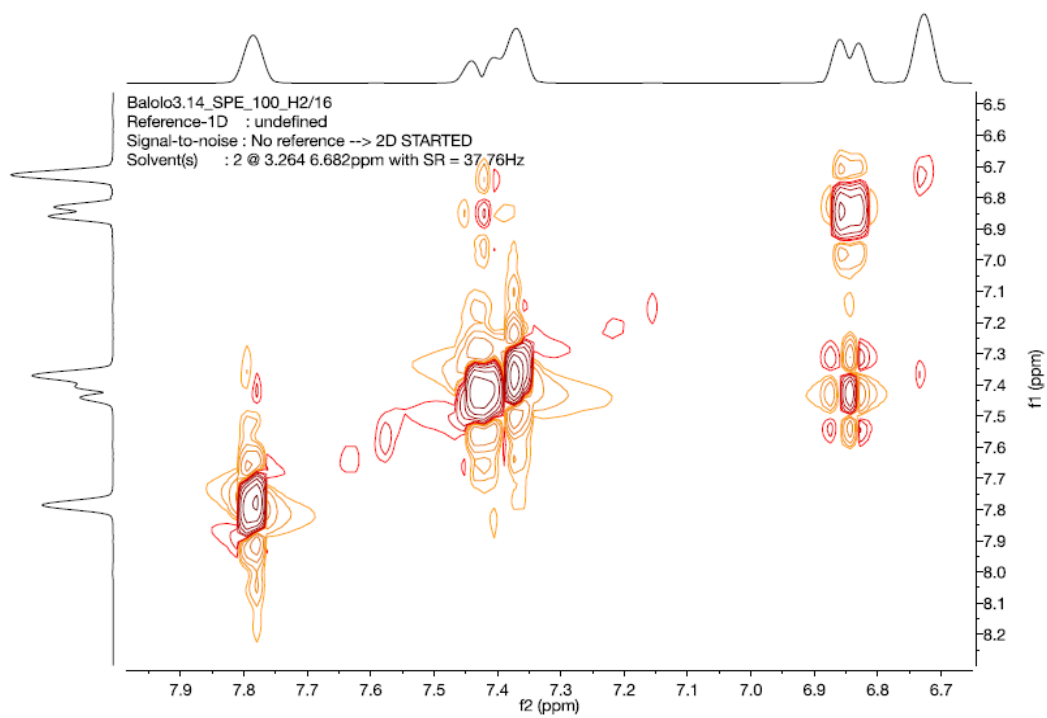


Figure S29. COSY spectrum of compound **3** in CD₃OD at 600 MHz

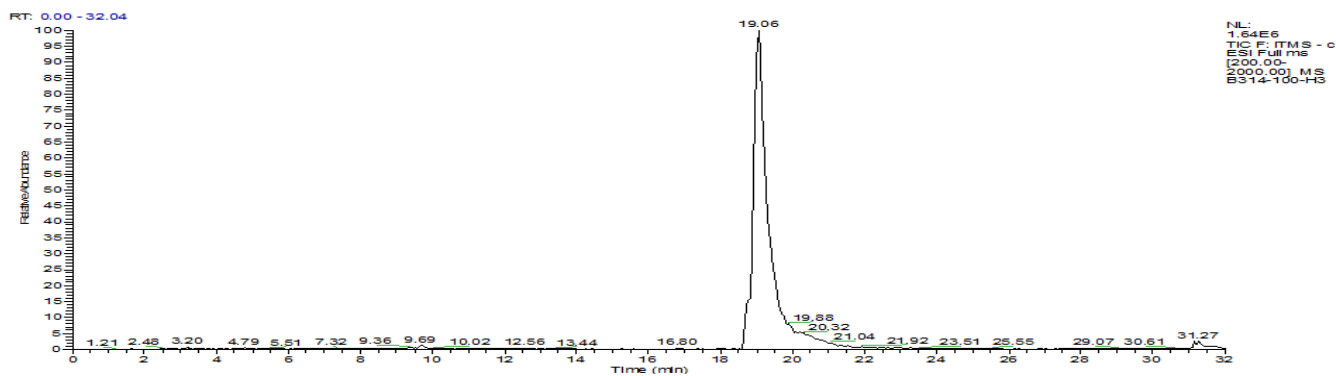
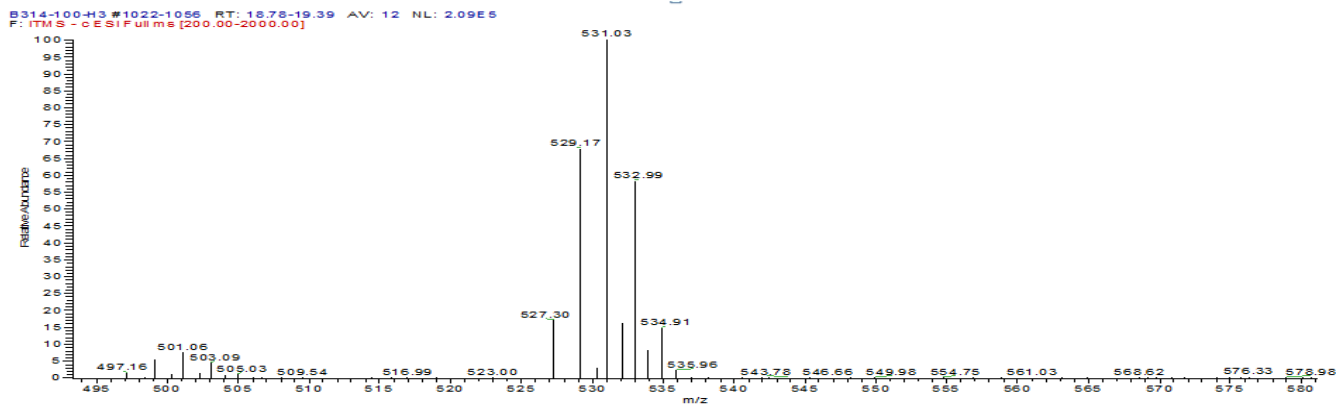


Figure S30. LC-ESIMS spectrum of compound 4

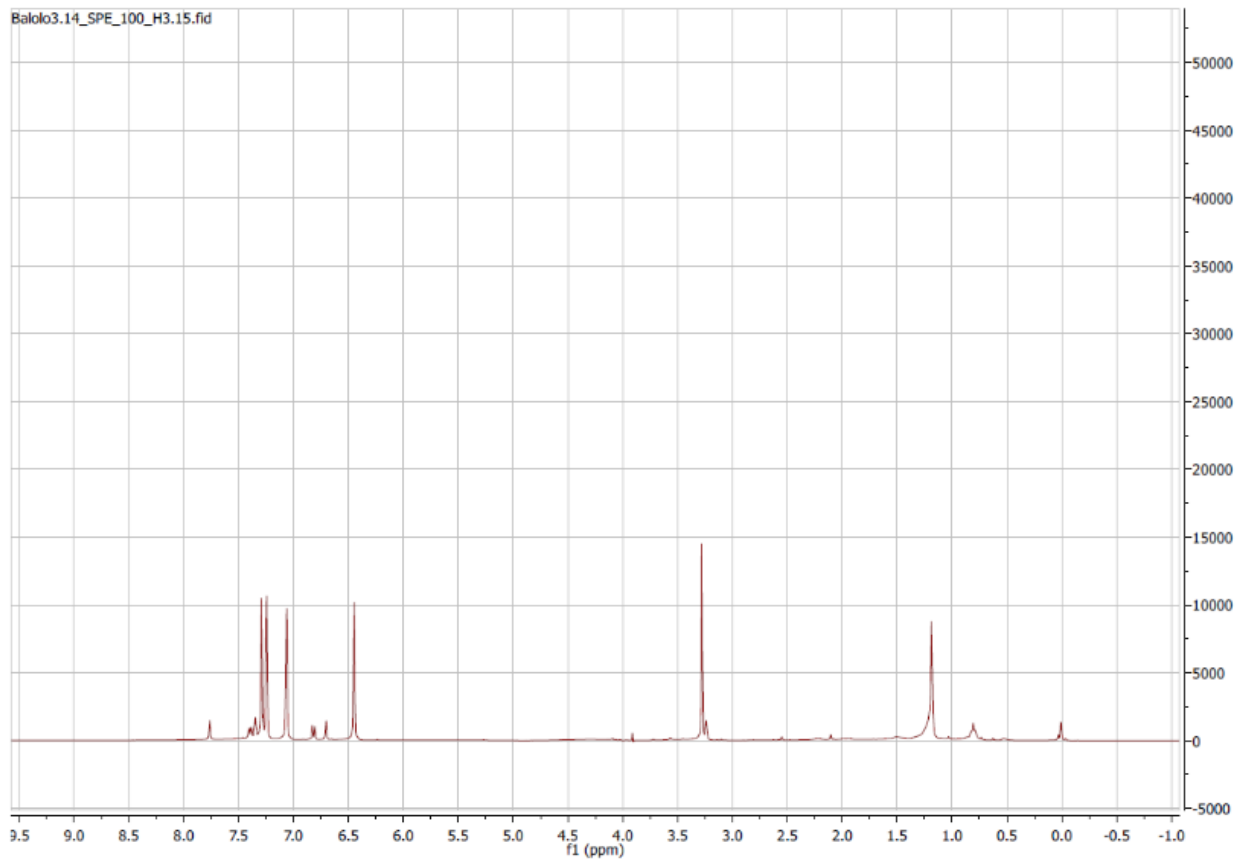


Figure S31. ¹H NMR spectrum of compound 4 in CD₃OD at 600 MHz

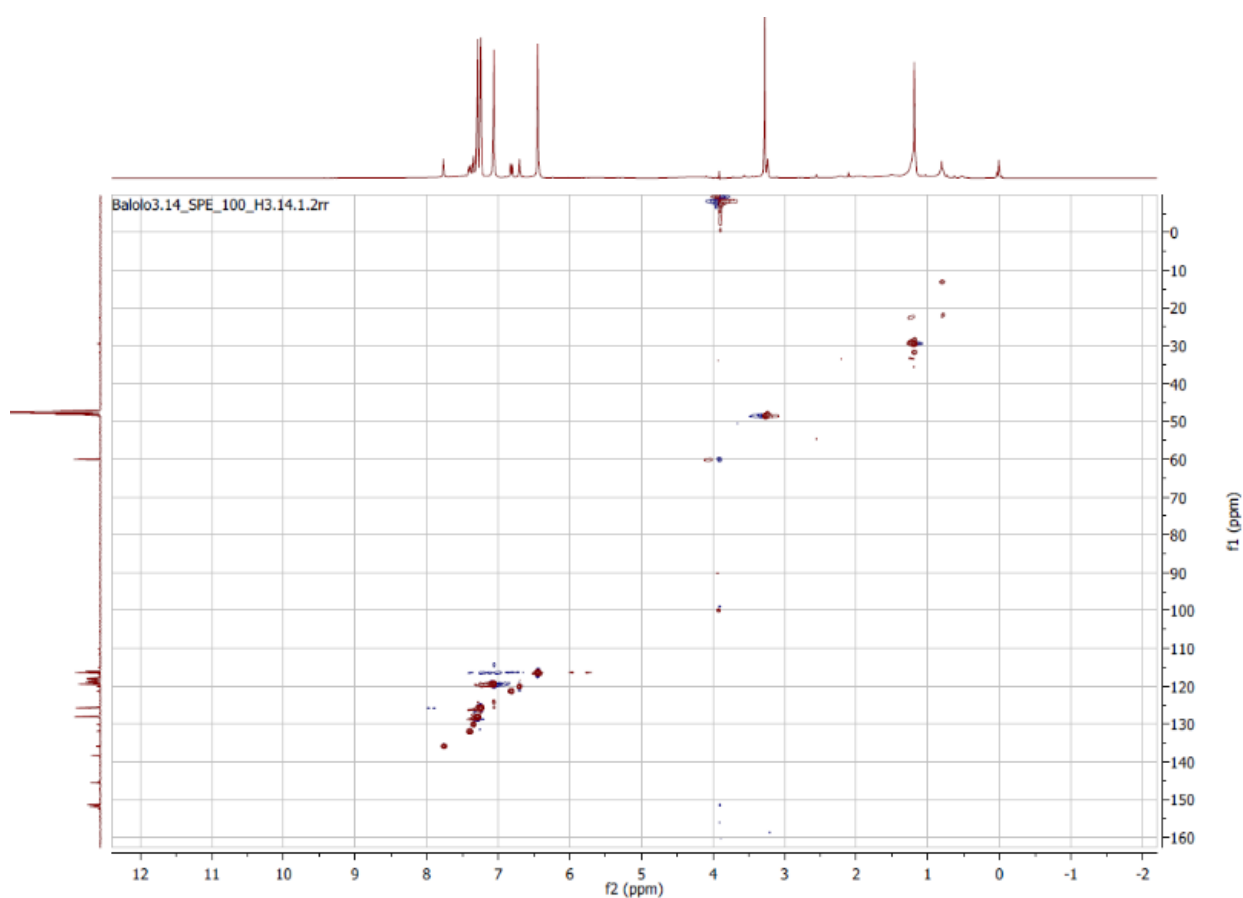


Figure S32. HSQC spectrum of compound **4** in CD₃OD at 600 MHz

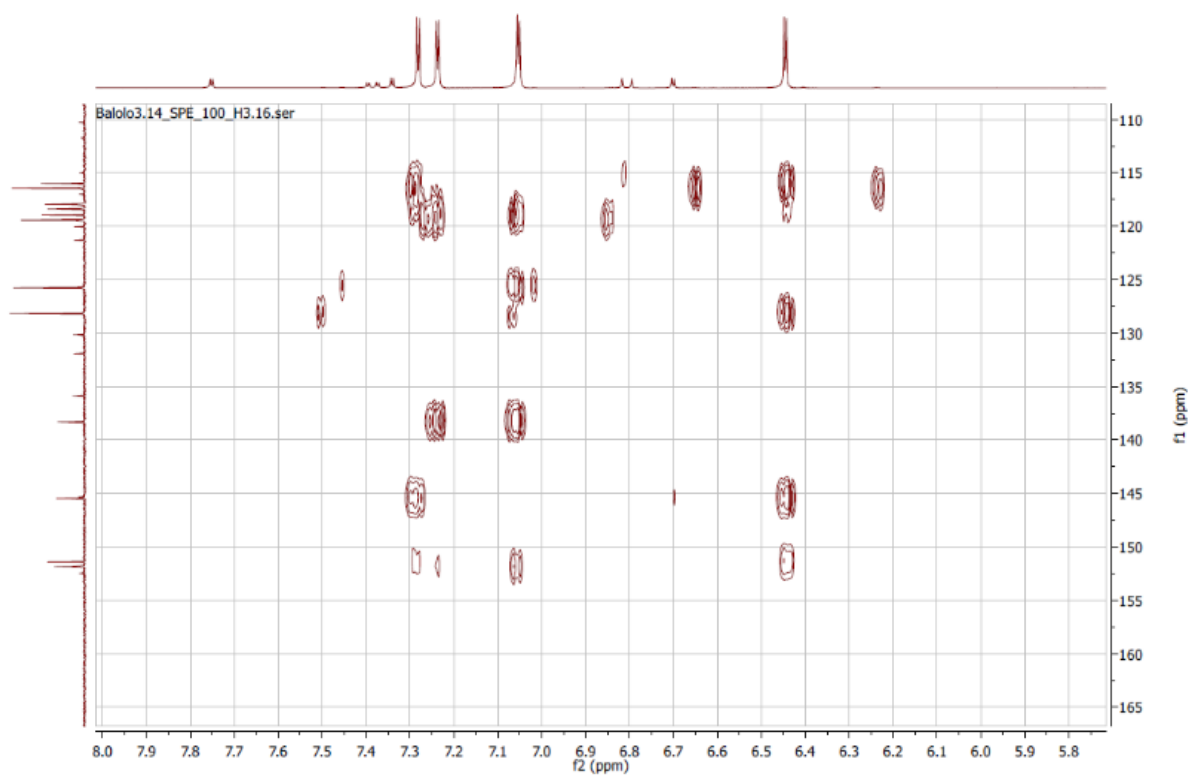


Figure S33. HMBC spectrum of compound **4** in CD₃OD at 600 MHz

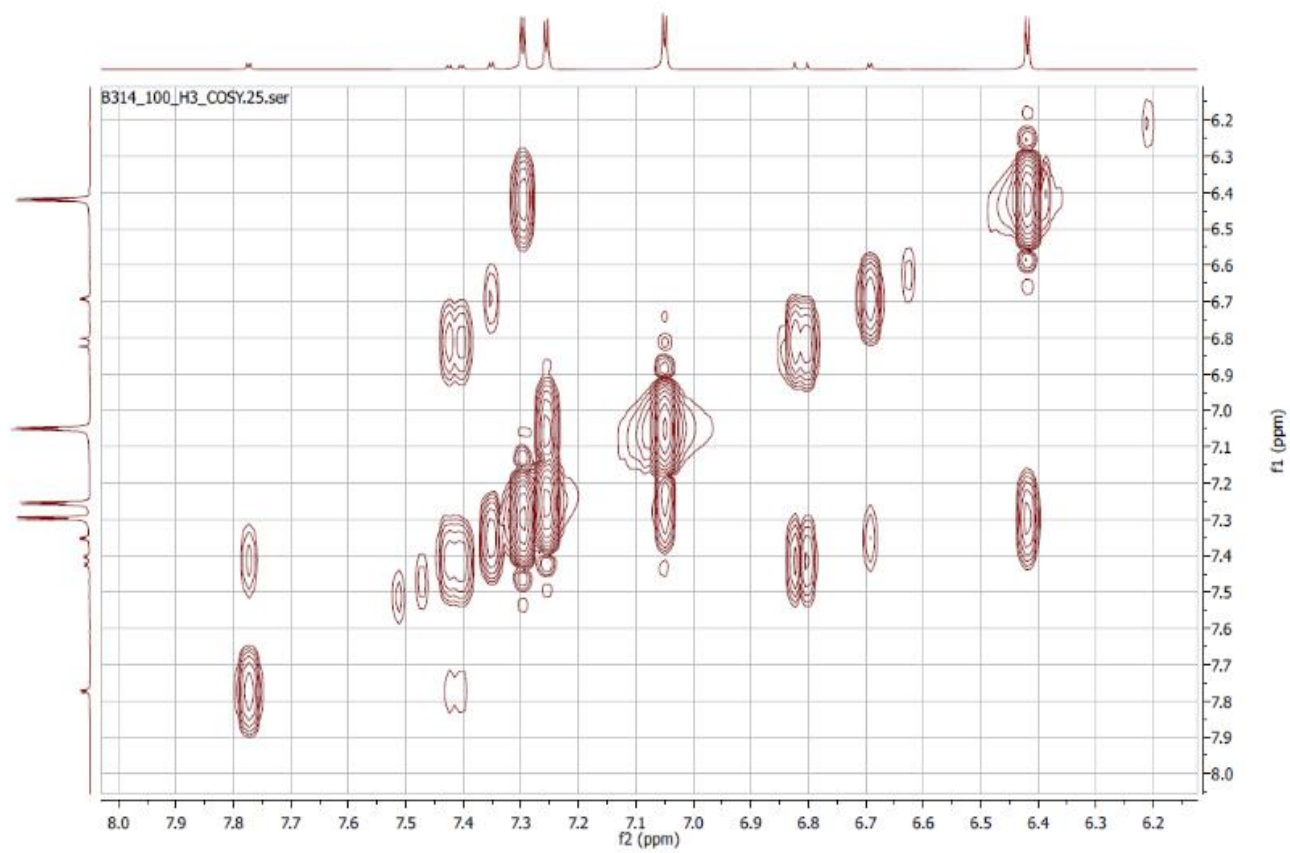


Figure S34. COSY spectrum of compound **4** in CD₃OD at 600 MHz

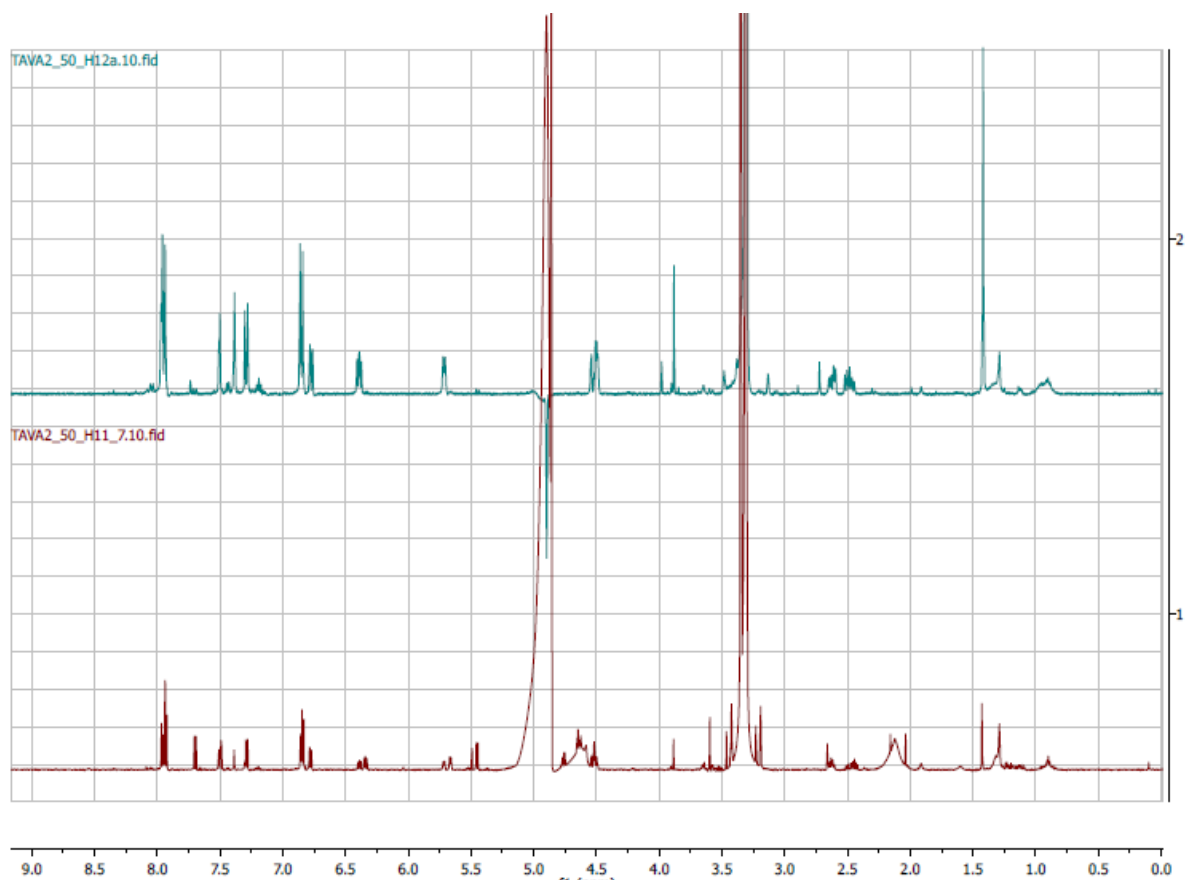
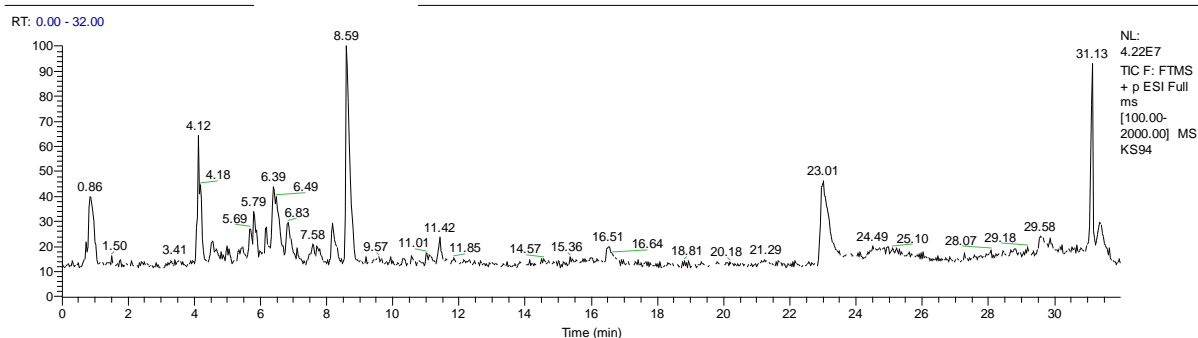


Figure S35. ^1H NMR spectrum of compound **1** HPLC fraction H12 (top) and ^1H NMR spectrum of compound **5** from HPLC fraction H11 (bottom) isolated from Tavarua 2 SPE 50% MeOH-water, in CD_3OD at 600 MHz



KS94 #489 RT: 8.22 AV: 1 NL: 7.00E5
F: FTMS + p ESI Full ms [100.00-2000.00]

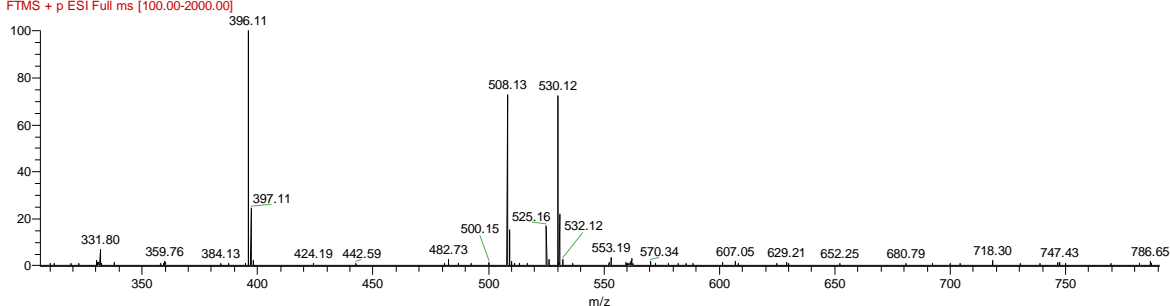
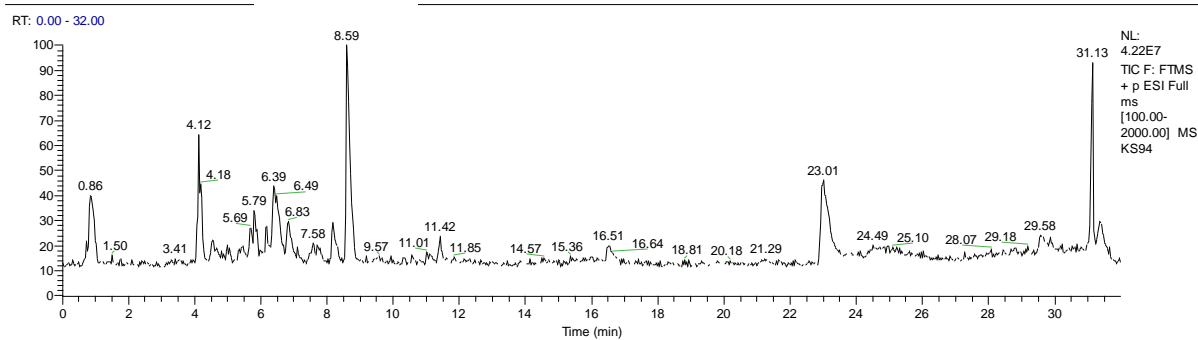


Figure S36. HRMS spectrum of Tavarua 2 SPE 50% MeOH (with fraction H11)



KS94 #409 RT: 6.86 AV: 1 NL: 2.27E6
F: FTMS + p ESI Full ms [100.00-2000.00]

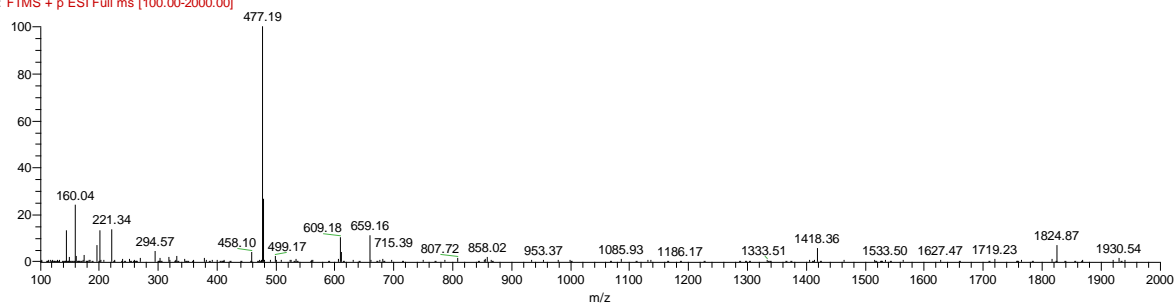


Figure S37. HRMS spectrum of Tavarua 2 SPE 50% MeOH (with fraction H10)

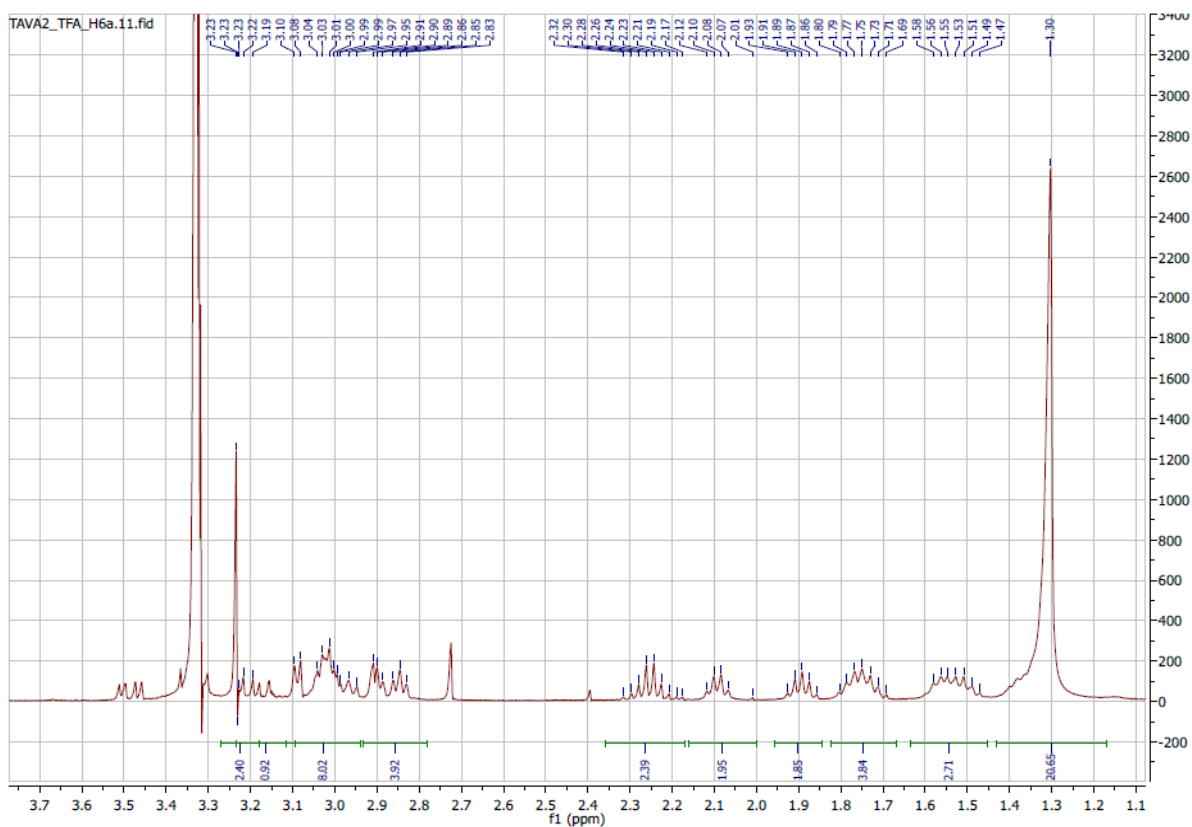


Figure S38. ^1H NMR spectrum of the compound 6 isolated from Tavarua 2 SPE fraction 100% MeOH+TFA in CD_3OD at 600 MHz

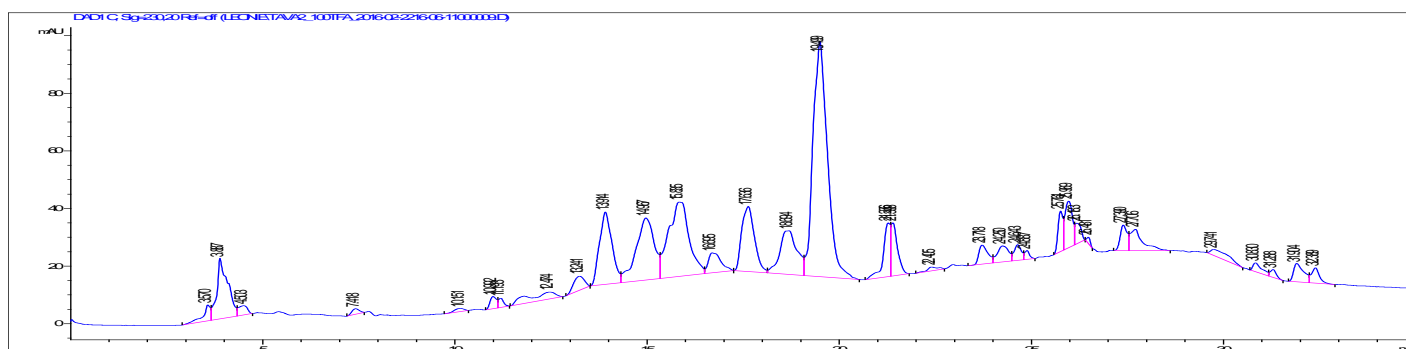


Figure S39. HPLC chromatogram of Tavarua 2 SPE 100% MeOH+TFA

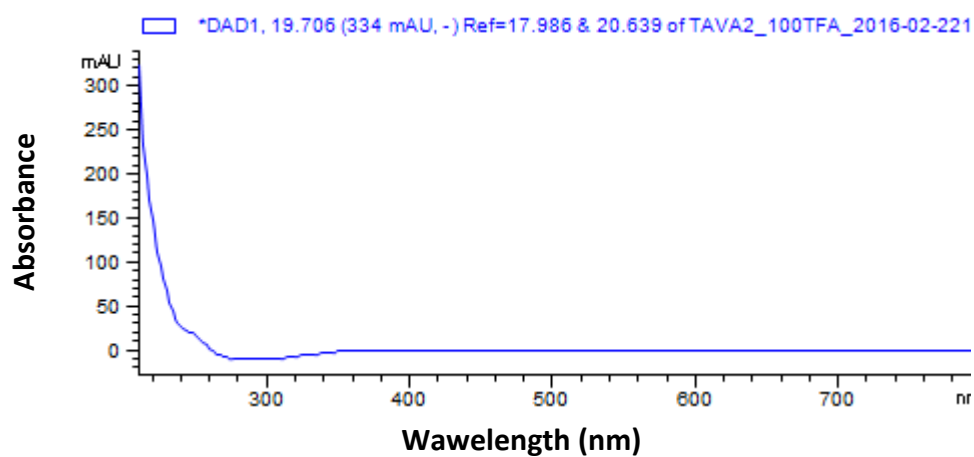


Figure S40. UV/Vis spectra of compound 6 in MeOH

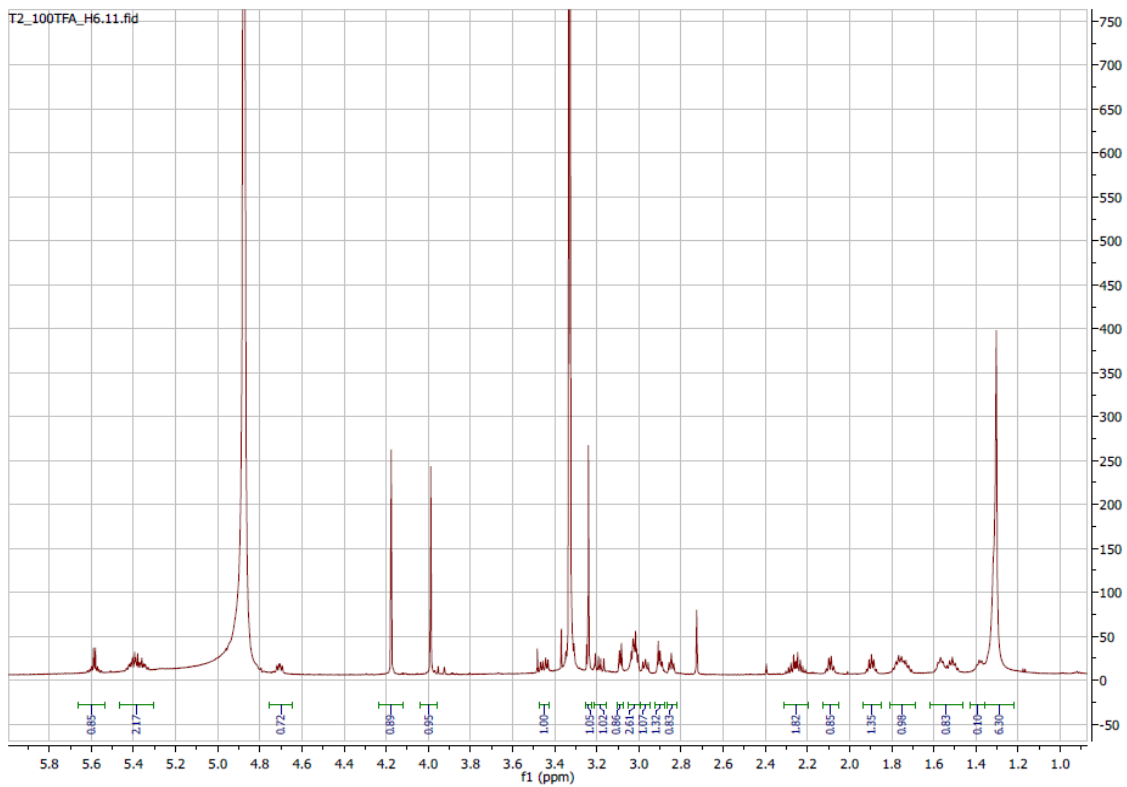


Figure S41. ^1H NMR spectrum of the compound **6** isolated from Tavarua 2 SPE fraction 100% MeOH+TFA in CD_3OD at 600 MHz

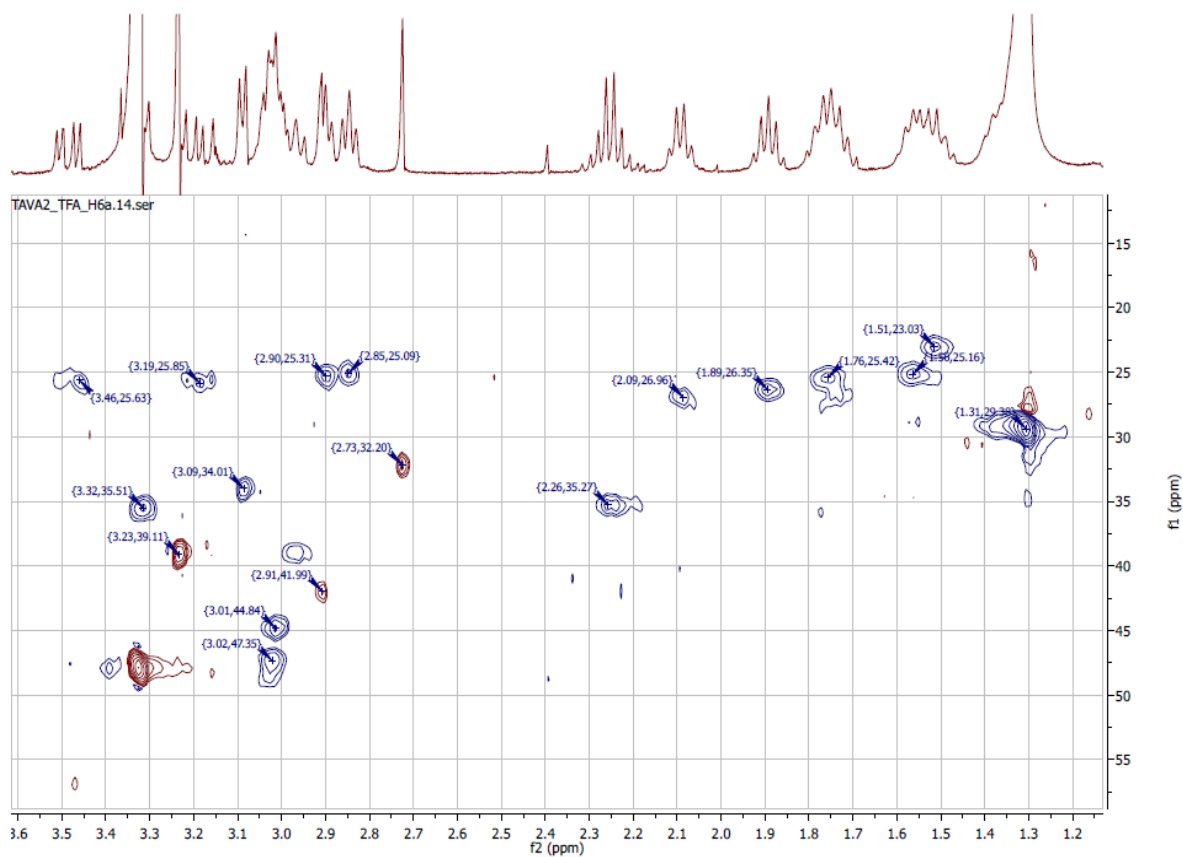


Figure S42. HSQC spectrum of compound **6** in CD_3OD at 600 MHz

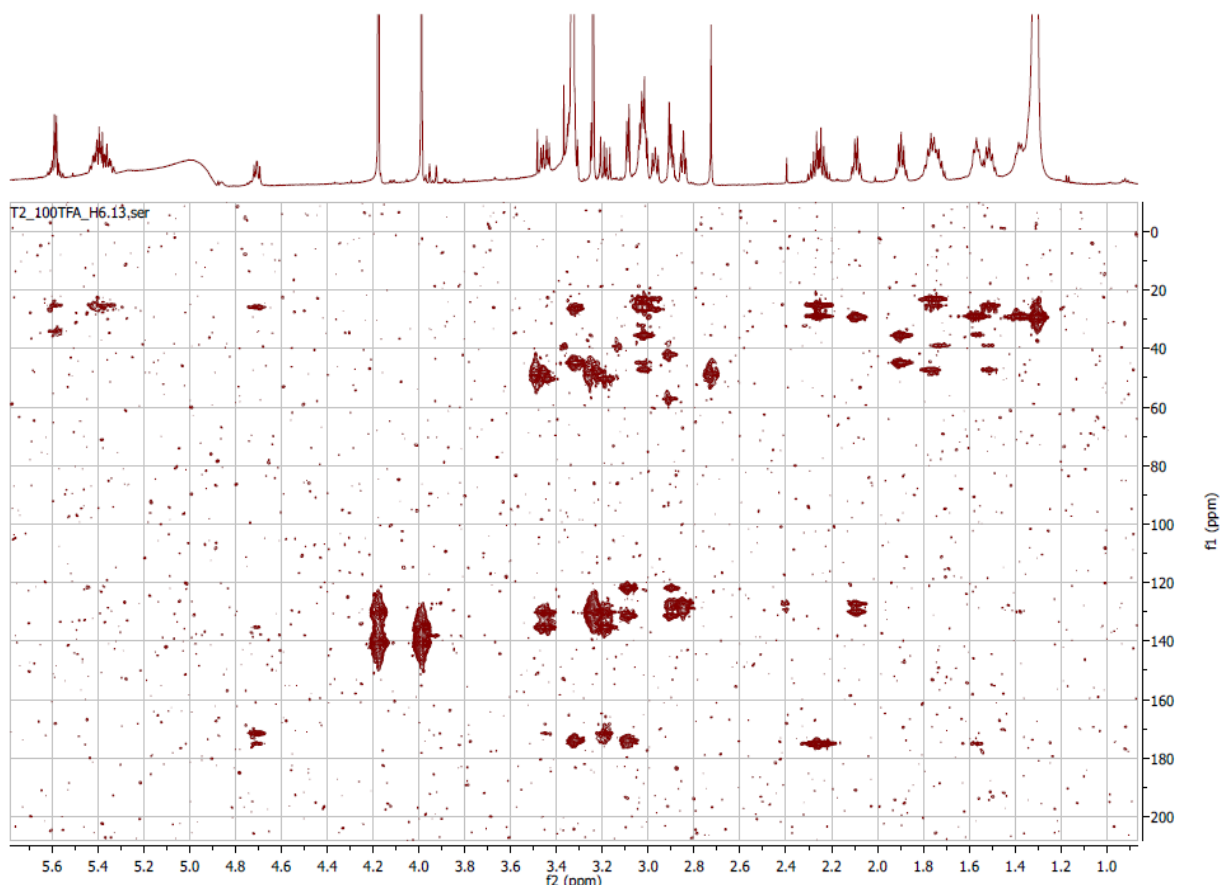


Figure S43. HMBC spectrum of compound **6** in CD₃OD at 600 MHz

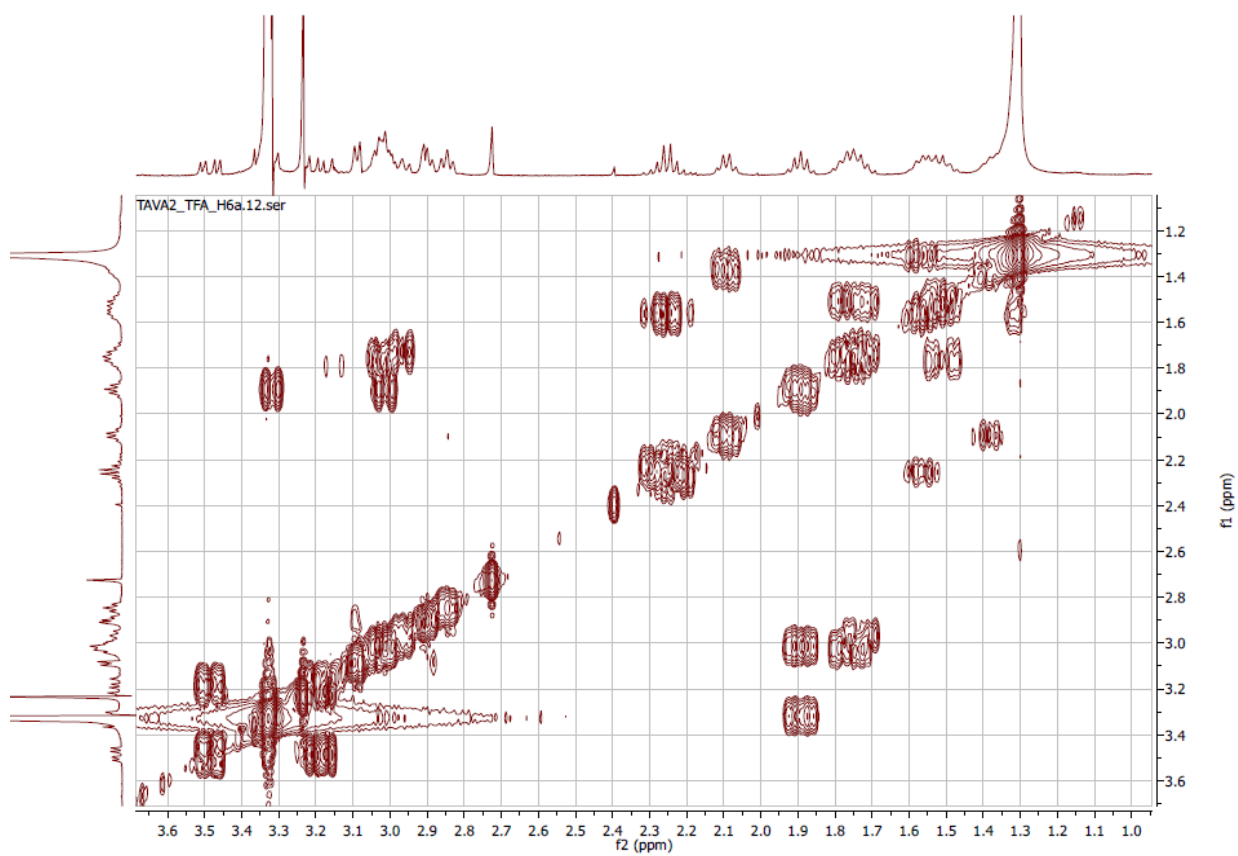


Figure S44. COSY spectrum of compound **6** in CD₃OD at 600 MHz

Léonie PELLISSIER

VALIDATION OF A NATURAL PRODUCTS DEREPLICATION METHOD

Rising incidences of various cancers and ever-growing infections caused by microorganisms are the main issues of modern medicine, coevolving with the respective drug targeting resistances. This “growing health crisis”, as defined by the World Health Organization in 2014, prompts researchers to find new resources of drug leads. Despite the overwhelming number of terrestrial natural products derivatives approved by the Food and Drug Administration as pharmaceuticals, industry has reduced its reliance on it, mostly due to the costs behind high rate of rediscovery of known compounds. However, academics and other scientists continue to explore the natural world turning their efforts towards the marine environment. We contributed to this effort in participating to an EU-funded project: PharmaSea, based on the bio discovery, identification and development of new substances from marine organisms. A key step in marine drug bio discovery is déréplication: the rapid identification of known compounds before isolation. The profusion of data sets and the lack of confirmatory data make this task arduous. Different strategies of dereplication procedures involving Liquid chromatography coupled with High Resolution Mass Spectrometry are thus emerging. In this regard, we investigated the PharmaSea dereplication method on 21 tunicates extracts. This method combines the ACDLabs IntelliXtract software with the MarinLit database from the Royal Society of Chemistry to dereplicate LCMS data. Nuclear Magnetic Resonance was also used to streamline the prioritization of samples to study to validate the method. The most interesting sample according to these strategies was selected for further purification leading to the isolation of seven compounds of which five are new as evidence of the success of the method.

Keywords:

Marine natural products, biodiscovery, dereplication, pharmacochemistry

VALIDATION D'UNE METHODE DE DEREPLICATION DE PRODUITS NATURELS MARINS

L'émergence de résistance aux antibiotiques et aux anticancéreux constitue aujourd'hui une "crise de santé mondiale", comme définie par l'Organisation Mondiale de la Santé en 2014. Cela incite les chercheurs à tenter de développer agents thérapeutiques innovants et plus spécifiques. Malgré la pléiade de dérivés de substances naturelles approuvés sur le marché pharmaceutique par la Food and Drug Administration, l'industrie a restreint son recours à cette source végétale, notamment en raison des coûts engendrés par la fréquente redécouverte de composés connus. Les scientifiques tournent alors désormais leurs efforts vers le milieu marin, encore très peu exploité. Nous avons contribué à cet effort en participant au projet de recherche Européen PharmaSea axé sur la découverte, la caractérisation, et le développement clinique de nouvelles substances pharmaceutiques tirées d'organismes marins. Une étape clé pour réussir dans la découverte de « médicaments de la mer » est la déréplication: l'identification rapide de composés déjà connus avant l'isolement. L'abondance de données et le manque d'informations confirmatives rendent cette tâche difficile. Ainsi, différentes stratégies de déréplication comme la Chromatographie Liquide couplée à la Spectrométrie de Masse Haute Résolution (LC-HRMS) émergent actuellement. Nous avons investigué la méthode de déréplication de données LCMS de PharmaSea sur 21 extraits de tuniciers. Cette méthode est basée sur une combinaison du logiciel ACDLabs IntelliXtract et de la base de données MarinLit de la Royal Society of Chemistry. Grâce à l'adjonction d'une analyse par Résonance Magnétique Nucléaire, nous avons pu classer plus pertinemment les extraits à étudier pour valider la méthode. Les plus prometteurs de nouveauté ont été choisis et purifiés, conduisant à l'isolement de sept composés dont cinq sont nouveaux, preuve de l'efficacité du mode de déréplication utilisé.

Mots clés:

Substances naturelles marines, bio prospection, déréplication, pharmacochimie

PELLISSIER Léonie

[Données à caractère personnel]

