

# **A Metabolomics Approach for Targeted Isolation and Production of Bioactive Secondary Metabolites in Microbial Isolates from Extreme Environments**

SUBMITTED IN THE FULFILMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE  
FACULTY OF SCIENCE THE UNIVERSITY OF STRATHCLYDE

BY

**Bela Maguie Pereira Sanches**

BSc (Hons), MSc (Hons)

Strathclyde Institute of Pharmacy and Biomedical Sciences  
University of Strathclyde  
Glasgow  
United Kingdom

April 2019

# Author's Declaration

This thesis is the result of the author's original research, except where it is specifically indicated. It has been composed by me and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

# Acknowledgements

The past years have been the most challenging of my life, not only in an intellectual matter but also in terms of physical and psychological wellbeing. Definitely, the PhD journey was difficult and if you add to this the “unforeseen” incident of being diagnosed with cancer, it made all even harder. Honestly, I thought that I was never going to get to this point of writing the acknowledgements of my thesis, thus, I am really happy and thankful that I was able to finish my PhD.

I would want to express my genuine gratitude to my supervisor, Dr RuAngelie Edrabel for encouraging me to pursue this scientific path, for believing in my skills and potential since we’ve met five years ago during my Erasmus internship and for her patience, motivation and endless knowledge. I want to thank as well to my second supervisor, Dr Paul Herron, although the genomic work was a brief, he was helpful. I would like to thank Mrs Louise Young for the guidance during the biological assays, especially to Miss Grainne Abbott for her continuous support during the laboratory work as well for her incredible support during my illness times. I will never forget the times you visited me in the hospital and how empathetic you were with my situation. Many thanks to Mr. Craig Irving from the department of Pure and Applied Chemistry for running my NMR samples and for Mr. Gavin Bain for helping me in the optical rotation experiments. I would like to thank Prof. Elsa Bronze da Rocha for her passion for science and excellent supervision during the undergraduate research project in Faculdade de Farmácia da Universidade do Porto, growing in me the desire to follow a career in research.

I would like to thank for the emergency grant from the Funds for Women Graduates and bursary from The Thomas and Margaret Roddan Trust, without this crucial financial help during a problematic period of my life, it was not going to be possible to finish my PhD.

I would like to express my gratitude to all the labmates from Dr RuAn’s research group, Natural Products Metabolomics Group, during the past years, especially to Dr

Lynsey McIntyre whom teach me all the essential techniques and methods that I've applied during my PhD and to Kirsty Black who became one of my best friends, supporting me with kindness, joyfulness and laughter during all times inside an outside the lab, I am so grateful for your friendship and sweet madness. I also want to thank Dr Jana Schniete and Tiago Melo Santos for their patience and assistance with the genomic work.

I am extremely grateful for all the medical team that saved my life, the Meadowpark St. Surgery, especially Dr Jemma Callahan for not ignoring my symptoms and referring me for a specialist Dr Graham Mackay, to the amazing surgeons Dr Chong and Dr Quinn you are the best, to the oncologist Dr Alexander McDonald who was so serious at the beginning but now he smiles during appointments, to all the wonderful nurses, especially Aileen Sutherland, your hugs when I go to the appointments are the greatest. You all are the real superheroes!

I have been really blessed with my cherished friends from childhood, high-school, university and from all different circumstances of life. I am sorry I cannot name all here (but you know who you are), but I want to show gratitude to Ana Lourenço, Ana Reis, Andreia Matas, Andreia José, Anaisa Tavares, Bruna André, Eunice Marques, Filipa Fonseca, Helena Vicente, Leonor Neves, Maria Inês Sequeira, Sofia Chante, Tatiana Cunha, among others. You are the best, fill my life with happiness and your friendship is a truly irreplaceable gift.

I want to express my deepest gratitude for my family, most especially, my brother and father. This journey would not have been possible without your never-ending support. Thank you so much for always inspiring me to follow my aspirations and always wanted the best for me. Thank you, father, for being here during my treatment and surgeries, I would not have made it this far without you.

I would like to express my whole-hearted gratitude to my boyfriend, partner, friend, Marco Fernandes, you were the reason I've survived, you've been always there during my darkest times and know me better than anyone else in this world. Additionally, you always believed in me, even when I didn't, and never let me give up. It's impossible

to translate into words how much you mean to me. I also want to thank as well to Marco's mother, Maria de Fátima Fernandes for her unconditional support during this difficult period.

Lastly, I want to dedicate this thesis to the memory of my dearest friend, Helena Gonçalves, a former PhD student from SIPBS, who tragically passed away in January 2017, after a short but courageous battle against acute myeloid leukaemia. I will always remember you.

# List of Tables

Table 1-1 Up-to-date pipeline perspective of the marine pharmaceuticals .....	37
Table 1-2 Overview of databases for the analysis of secondary metabolites and their biosynthetic gene clusters. ....	66
Table 1-3 Overview of computational tools for the analysis of secondary metabolites and their biosynthetic gene clusters.....	69
Table 2-1 Samples analysis concentration and solvents used in the extraction procedure.....	77
Table 2-2 Seeding densities (cell/cm <sup>2</sup> ) for the used cell lines. ....	94
Table 2-3 PTP1B assay buffer preparation. ....	98
Table 2-4 Enzyme buffer preparation. ....	99
Table 3-1 Mobile phase used for first fractionation of SBT531.....	111
Table 3-2 Chromatographic conditions applied to the first fractionation of SBT531 .....	112
Table 3-3 First SBT531 fractionation that afforded 12 fractions. ....	112
Table 3-4 Dereplication table of predicted bioactive metabolites by OPLS-DA loadings.....	133
Table 3-5 Mobile phase used for sub-fractionation of SBT531 fractions .....	138
Table 3-6 Chromatographic conditions applied to the sub-fractionation of SBT531 fractions.....	138
Table 3-7 Sub-fractionation of SBT531 Fraction 4 afforded 12 fractions. ....	139
Table 3-8 Sub-fractionation of SBT531 Fraction 6 and 7 afforded 17 fractions.....	140
Table 3-9 Sub-fractionation of SBT531 Fraction 9 afforded 15 fractions. ....	140
Table 3-10 Isolated compounds from SBT531 fractions.....	146
Table 3-11 Aseanostatin P6.....	147
Table 3-12 <sup>1</sup> H NMR (400 MHz) and <sup>13</sup> C NMR (100 MHz) of aseanostatin P6 in DMSO-d <sub>6</sub> .....	155
Table 3-13 2-hydroxy-14-methylhexadecanoic acid.....	156
Table 3-14 <sup>1</sup> H NMR (400 MHz) and <sup>13</sup> C NMR .....	163
Table 3-15 3-hydroxy-13-methyltetradecanoic acid. ....	165
Table 3-16 <sup>1</sup> H NMR (400 MHz) and <sup>13</sup> C NMR .....	173
Table 3-17 <i>N</i> -phenethylacetamide.....	176
Table 3-18 Cholic acid derivative. ....	178
Table 3-19 - Parameters of experiments performed on SBT531 .....	180
Table 3-20 <i>M. ruestringensis</i> assembly metrics .....	193
Table 3-21 Taxonomic labelling using top families and genera.....	194
Table 4-1 Gradient used for first fractionation of SBT687 .....	216
Table 4-2 Chromatographic conditions applied to the fractionation of SBT687.....	216
Table 4-3 First SBT687 fractionation afforded 8 fractions. ....	216
Table 4-4 Gradient used for the second fractionation of SBT687.....	218
Table 4-5 Second SBT687 fractionation afforded 22 fractions.....	218
Table 4-6 Rf values of the encircled analytes from the TLC plates .....	222

Table 4-7 Dereplication table of predictive possible bioactive metabolites .....	235
Table 4-8 Dereplication table of the manzamines produced by the SBT687 fractions .....	241
Table 4-9 4-epi-manoyl oxide acid. ....	249
Table 4-10 <sup>1</sup> H NMR (600.13 MHz) and <sup>13</sup> C NMR (150.92 MHz) of manoyl oxide acid .....	263
Table 4-11 Manzamines found in SBT687 Fraction 17. ....	266
Table 4-12 AntiSMASH analysis of the biosynthetic gene clusters (BGCs) .....	277
Table 4-13 Biological activity and source of other-diterpenoids. ....	278
Table 4-14 Some of the relevant manzamines isolated since manzamine A in 1986 .....	281
Table 5-1 Gradient elution of crude extract obtained from 3L of SBT692 using Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification. ....	293
Table 5-2 Chromatographic conditions applied on the fractionation of SBT692 crude extract obtained from 3L scale-up .....	293
Table 5-3 Fractionation of SBT692 crude extract from 3L scale-up affording 12 fractions. ....	294
Table 5-4 Gradient elution of SBT692 crude extract obtained from the first 15 L scale-up using the Reveleris™ Flash Chromatography System. ....	296
Table 5-5 Chromatographic conditions applied on fractionation of SBT692 crude extracts obtained from the first 15L scale-up using the Reveleris™ Flash Chromatography System. ....	296
Table 5-6 Fractionation of SBT692 crude extract .....	297
Table 5-7 Step gradient elution of SBT692 crude extract obtained from the 2nd 15L scale-up using Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification. ....	299
Table 5-8 Chromatographic conditions applied to the crude extract .....	299
Table 5-9 Fractionation of SBT692 crude extract .....	300
Table 5-10 Dereplication table of the ighlighted end point metabolites from SBT692 .....	320
Table 5-11 Dereplication table of the predicted bioactive end point metabolites from fraction F9, F11, F12 and F13 from SBT692 second fermentation .....	350

# List of Figures

Figure 1-1 A schematic of the process in drug discovery and development .....	30
Figure 1-2 Discovery of new molecular entities (NMEs) and natural products contribution.....	31
Figure 1-3 Number of new compounds isolated from marine organisms from 2001 to 2013.....	35
Figure 1-4 Flow chart from marine habitat to compound selection .....	45
Figure 1-5 Molecular structure of secondary metabolites isolated from thermophilic bacteria .....	48
Figure 1-6 Scanning electron micrograph of <i>Muricauda ruestringensis</i> B1 .....	51
Figure 1-7 Phylogenetic tree highlighting the position of <i>Muricauda ruestringensis</i> .....	51
Figure 1-8 Metabolomics workflow.....	54
Figure 1-9 DNA sequencing costs along time.....	59
Figure 1-10 Metagenomics workflow. ....	61
Figure 1-11 Methods that can be used to activate silent genes in microorganisms .....	63
Figure 1-12 Overview of the most commonly used and freely accessible tools specialized for the analysis of secondary metabolites .....	65
Figure 2-1 Chromatogram from SBT531 first fractionation using the Biotage Isolera™.....	82
Figure 2-2 Chromatogram from SBT531-Fraction 4 (SBT2311) sub-fractionation .....	82
Figure 2-3 Chromatogram from SBT531-Fraction 9 (SBT2316) sub-fractionation .....	83
Figure 2-4 Chromatogram from SBT 531-Fraction 6+7 (SBT2313+SBT2314) sub-fractionation.....	83
Figure 2-5 Chromatogram from SBT687 first fractionation using the Reveleris™.....	84
Figure 2-6 Chromatogram from SBT687 second fractionation using the Reveleris™ .....	84
Figure 2-7 Chromatogram from SBT692 first fractionation .....	85
Figure 2-8 Chromatogram from SBT 692 second fractionation.....	86
Figure 2-9 The preparation of the serial concentration dilutions of the samples.....	95
Figure 2-10 Preparation of serial concentration dilutions of the standard curve of MG132 for the NFκβ Luciferase assay. ....	98
Figure 2-11 The preparation of the serial concentration dilutions of the standard curve of TFMS for the PTP1B assay. ....	100
Figure 2-12 The preparation of the serial concentration dilutions of the standard curve of acarbose for the alpha-glucosidase assay.....	101
Figure 3-1 <sup>1</sup> H NMR in DMSO-d <sub>6</sub> of an extract of SBT531 obtained by small scale-up.....	109
Figure 3-2 LC-HRMS data of SBT531 obtained by small scale-up fermentation and extracted ion chromatograms.....	110
Figure 3-3 Summary TLC plates of the first fractionation of SBT531 .....	113
Figure 3-4 Pre-saturated <sup>1</sup> H NMR (400 MHz) spectra obtained for SBT531 fractions.....	114
Figure 3-5 Inhibition effect of SBT531 crude extract and fractions F1 to F10 against α-glucosidase	116
Figure 3-6 SBT531 inhibition curve against α-glucosidase at serial concentrations (n=3).....	116
Figure 3-7 Acarbose inhibitor control curve on α-glucosidase assay (n=3).....	117

Figure 3-8 SBT531 Fraction 2 inhibition curve against $\alpha$ -glucosidase .....	117
Figure 3-9 Inhibition effect of SBT531 crude extract and fractions F1 to F10 against PTP1B .....	118
Figure 3-10 SBT531 inhibition curve against PTP1B at serial concentrations (n=3). .....	119
Figure 3-11 PTP1B inhibitor IV curve in the PTP1B assay. ....	119
Figure 3-12 SBT531 Fraction 2 inhibition curve against PTP1B at serial concentrations (n=3). ....	120
Figure 3-13 SBT531 Fraction 3 inhibition curve against PTP1B at serial concentrations (n=3). ....	120
Figure 3-14 Heatmap based on HR-LC/MS data .....	122
Figure 3-15 Base peak plot chromatogram (HR-LC/MS data) of SBT531 .....	123
Figure 3-16 Principal component analysis (PCA) plot of SBT531 .....	125
Figure 3-17 OPLS-DA of SBT531 fractions according to inhibition activity against PTP1B. ....	127
Figure 3-18 OPLS-DA of SBT531 fractions.....	128
Figure 3-19 Permutation tests (100 permutations) for SBT531 fractions for the OPLS-DA model ..	128
Figure 3-20 Intensity base peak plot of LC-HRMS data in negative ionisation mode .....	131
Figure 3-21 Intensity base peak plot of HR-LC/MS data.....	132
Figure 3-22 1H NMR dimensionality reduction overview .....	137
Figure 3-23 Summary TLC plates of the sub-fractionation of SBT531 fraction 4.....	141
Figure 3-24 Summary TLC plates of the sub-fractionation .....	141
Figure 3-25 Summary TLC plates of the Sub-fractionation of SBT531 Fraction 9 .....	142
Figure 3-26 $^1\text{H}$ NMR spectra of SBT531 F4 sub-fractions (SBT2369 to SBT2380) at 400 MHz .....	142
Figure 3-27 $^1\text{H}$ NMR spectra of SBT531 F6+F7 sub-fractions (SBT2410 to SBT2426) at 400 MHz.....	143
Figure 3-28 $^1\text{H}$ NMR spectra of SBT531 F9 sub-fractions (SBT2381 to SBT2395) at 400 MHz .....	144
Figure 3-29 Inhibition effect of SBT531 fraction 4 sub-fractions 11 and 12 against PTP1B .....	145
Figure 3-30 $^1\text{H}$ NMR spectrum of aseanostatin P6 at 400 MHz in DMSO- $d_6$ . ....	150
Figure 3-31 Expansion of $^1\text{H}$ NMR spectrum of aseanostatin P6 at 400 MHz in DMSO- $d_6$ . ....	151
Figure 3-32 $^{13}\text{C}$ spectrum of aseanostatin P6 at 100 MHz superimposed with DEPT-135 .....	152
Figure 3-33 Expansion of $^{13}\text{C}$ spectrum of aseanostatin P6 at 100 MHz superimposed with DEPT-135 .....	152
Figure 3-34 1H-1H COSY spectrum of aseanostatin P6 at 400 MHz in DMSO- $d_6$ .....	153
Figure 3-35 HMBC of aseanostatin P6 at 400 MHz .....	154
Figure 3-36 $^1\text{H}$ NMR spectrum of 2-hydroxy-14-methylhexadecanoic acid.....	159
Figure 3-37 $^{13}\text{C}$ spectrum of 2-hydroxy-14-methylhexadecanoic acid at 100 MHz in DMSO- $d_6$ . ....	160
Figure 3-38 Expansion of $^{13}\text{C}$ spectrum of 2-hydroxy-14-methylhexadecanoic acid .....	160
Figure 3-39 $^1\text{H}$ - $^1\text{H}$ COSY spectrum of 2-hydroxy-14-methylhexadecanoic at 400 MHz .....	161
Figure 3-40 HMBC of 2-hydroxy-14-methylhexadecanoic acid.....	162
Figure 3-41 Superimposed $^1\text{H}$ - $^1\text{H}$ COSY NMR (400 MHz) spectrum in Pyridine- $d_5$ .....	164
Figure 3-42 $^1\text{H}$ NMR spectrum of 3-hydroxy-13-methyltetradecanoic acid .....	169
Figure 3-43 $^{13}\text{C}$ spectrum of 3-hydroxy-13-methyltetradecanoic acid .....	169
Figure 3-44 Expansion of $^{13}\text{C}$ spectrum at 100 MHz of 3-hydroxy-13-methyltetradecanoic acid .....	170
Figure 3-45 $^1\text{H}$ - $^1\text{H}$ COSY spectrum at of 3-hydroxy-13-methyltetradecanoic acid 400 MHz .....	171

Figure 3-46 HMBC of 3-hydroxy-13-methyltetradecanoic acid at 400 MHz .....	172
Figure 3-47 Superimposed <sup>1</sup> H- <sup>1</sup> H COSY NMR (400 MHz) spectrum for MPTA derivatives of 3-hydroxy-13-methyltetradecanoic acid.....	175
Figure 3-48 Principal component analysis (PCA) of metabolic profiles of <i>M. ruestringensis</i> .....	182
Figure 3-49 Principal component analysis (PCA) of the metabolomic profiles of bacteria 02MUR extracts .....	184
Figure 3-50 Principal component analysis (PCA) of the metabolomic profiles of bacteria 03MUR extracts .....	186
Figure 3-51 - Principal component analysis (PCA) of the metabolomic profiles of bacteria 05MUR extracts .....	189
Figure 3-52 Comparison of the chromatograms (negative ionization-LC-HRMS data) of the SBT531 medium scale-up .....	190
Figure 3-53 Comparison of the chromatograms from Biotage MPLC Isolera™ One 2.0.4.....	191
Figure 3-54 <i>M. ruestringensis</i> functional roles and protein families.....	195
Figure 3-55 Biosynthetic gene clusters analysis .....	197
Figure 3-56 Biosynthesis of fatty acids pathway .....	197
Figure 3-57 Metabolism of biotin (vitamin H).....	198
Figure 3-58 Summary diagram of metabolic pathways of biotin and fatty acids biosynthesis .....	199
Figure 3-59 Inhibition effect of SBT587 crude extract against PTP1B .....	200
Figure 3-60 <sup>1</sup> H-NMR spectral data (400 MHz) of SBT587 Day 7.....	201
Figure 3-61 <sup>1</sup> H-NMR spectral data (400 MHz) of SBT531 small scale up.....	202
Figure 3-62 <sup>1</sup> H-NMR spectral data (400 MHz) of SBT587 Day 7 Extract 1, 2 and 3 10% Inoculum.....	202
Figure 3-63 <sup>1</sup> H-NMR spectral data (400 MHz) of SBT531 Fraction 2 .....	203
Figure 3-64 Base peak plot chromatogram (HR-LC/MS data) of SBT587.....	204
Figure 3-65 Base peak plot chromatogram (HR-LC/MS data) of SBT587.....	205
Figure 3-66 Base peak plot chromatogram (HR-LC/MS data) of SBT531 .....	205
Figure 3-67 Metabolites identified in <i>M. ruestringensis</i> strains SBT531 and SBT587.....	206
Figure 4-1 <i>Micromonospora</i> sp. L5 overall genomic description .....	213
Figure 4-2 Proton <sup>1</sup> H NMR (400 MHz) spectra obtained for SBT687 at different scale-up .....	215
Figure 4-3 Proton <sup>1</sup> H NMR (400 MHz) spectra obtained for SBT687 fractions .....	219
Figure 4-4 Summary TLC plates.....	221
Figure 4-5 Heatmap based on HR-LC/MS data .....	223
Figure 4-6 Base peak plot chromatogram (HR-LC/MS) of SBT687 fractions.....	225
Figure 4-7 Structure of loxistatin and 1-vinyl-8-hydroxy-β-carboline.....	225
Figure 4-8 SIMCA analysis of SBT687 fractions mass spectral data .....	228
Figure 4-9 Contribution scores plots for F5, F6, F16, F17, and F18 to indicate the orthogonal VIP .....	229
Figure 4-10 MVDA analysis of mass spectral data of SBT687 fractions .....	234
Figure 4-11 OPLS-DA of SBT687 fractions with only F17 and F18.....	238
Figure 4-12 MVDA analysis of mass spectral data of SBT687 fractions F17 and F18 .....	239

Figure 4-13 Permutation test (100 permutations) for the OPLS-DA model of SBT687 fractions .....	240
Figure 4-14 Heatmap based on <sup>1</sup> H NMR data displaying the distinct metabolic profiles of 21 fractions of SBT687.....	242
Figure 4-15 <sup>1</sup> H- <sup>1</sup> H COSY (400 MHz) spectrum in DMSO-d <sub>6</sub> of the cyclized dipeptide.....	243
Figure 4-16 MVDA analysis of <sup>1</sup> H NMR spectral data of SBT687 fractions .....	245
Figure 4-17 MVDA MS-NMR data fusion analysis of SBT687 fractions .....	248
Figure 4-18 <sup>1</sup> H NMR spectrum of manoyl oxide acid at 600.13 MHz in DMSO-d <sub>6</sub> .....	254
Figure 4-19 Expansion of <sup>1</sup> H NMR spectrum between 0.5 to 1.80 ppm of manoyl oxide acid .....	255
Figure 4-20 JMOD – C spectrum at 150.92 MHz of manoyl oxide acid in DMSO-d <sub>6</sub> .....	255
Figure 4-21 Expansion of DEPT – C spectrum between 15 to 38 ppm at 150.92 MHz of manoyl oxide acid in DMSO-d <sub>6</sub> .....	256
Figure 4-22 <sup>1</sup> H- <sup>1</sup> H COSY (400 MHz) spectrum in DMSO-d <sub>6</sub> .....	257
Figure 4-23 HSQC of manoyl oxide acid at 600 MHz.....	258
Figure 4-24 Expansion of the HSQC of manoyl oxide acid at 600 MHz.....	259
Figure 4-25 HMBC of manoyl oxide acid at 600 MHz.....	260
Figure 4-26 Expansion of HMBC of manoyl oxide acid at 600 MHz.....	261
Figure 4-27 1H-1H NOESY (400 MHz) spectrum in DMSO-d <sub>6</sub> .....	262
Figure 4-28 Compound dolatriol C <sub>20</sub> H <sub>32</sub> O <sub>3</sub> .....	264
Figure 4-29 Predicted <sup>1</sup> H and <sup>13</sup> C NMR spectrum for dolatriol. ....	265
Figure 4-30 Manzamine A and the characteristic β-carboline moiety of manzamines. ....	268
Figure 4-31 Simulated <sup>1</sup> H NMR spectrum for manzamine A. ....	269
Figure 4-32 <sup>1</sup> H NMR spectrum of SBT687 F17 manzamine at 600.13 MHz in DMSO-d <sub>6</sub> .....	270
Figure 4-33 <sup>1</sup> H- <sup>1</sup> H COSY (600.13 MHz) spectrum in DMSO-d <sub>6</sub> .....	271
Figure 4-34 Expansion of <sup>1</sup> H- <sup>1</sup> H COSY (600.13 MHz) spectrum in DMSO-d <sub>6</sub> .....	272
Figure 4-35 Stereochemistry of the manoyl oxide acid.....	275
Figure 4-36 Examples of isolated diterpenes from marine <i>Micromonospora</i> sp.....	276
Figure 4-37 Structure of the glycosylated carotenoid sioxanthin. ....	276
Figure 5-1 Inhibition of TNF-α-NFκB by MG132 .....	289
Figure 5-2 Summary TLC of the first fractionation of SBT692 crude extract obtained from a 3L scale-up with solvent system ACN:Water 1:1 on RP18 plate .....	295
Figure 5-3 Summary TLC of the fractionation of SBT692 crude extract obtained from the first 15L scale-up with solvent system ACN:Water 1:1 on RP18 plates.....	298
Figure 5-4 Summary TLC of the fractionation of crude extract SBT692 obtained from the second 15L scale-up with ACN:Water 1:1 solvent system on RP18 plates.....	301
Figure 5-5 Proton <sup>1</sup> H NMR (400 MHz) spectra obtained for SBT92 crude extracts obtained from different scale-up fermentations in DMSO-d <sub>6</sub> .....	303
Figure 5-6 Proton <sup>1</sup> H NMR (400 MHz) spectra obtained for SBT92 fractions .....	304
Figure 5-7 MG132 inhibition control curve for the NFκβ Luciferase assay done in triplicate (n=3) gave an IC <sub>50</sub> of 1.5μM. ....	305

Figure 5-8 The inhibition effect of SBT692 crude extracts and fractions .....	306
Figure 5-9 Cytotoxicity effect of SBT692 crude extracts and fractions from first fractionation .....	307
Figure 5-10 MCF-7 cancer cell line viability screening .....	308
Figure 5-11 HT29 cancer cell line viability screening .....	309
Figure 5-12 Cell viability screening of SBT692 crude extracts and fractions .....	310
Figure 5-13 LC-HRMS data of SBT692 from the initial small-scale fermentation .....	312
Figure 5-14 Base peak chromatograms (HRMS data) of SBT692 crude extract .....	313
Figure 5-15 Heatmap based of HRMS data displaying distinct metabolic profiles of two batches of fermentations of SBT692 .....	314
Figure 5-16 PCA metabolomic screening of SBT692 .....	316
Figure 5-17 OPLS-DA metabolomic screening of SBT692 .....	319
Figure 5-18 MVDA analysis of NMR data of SBT92 fractions from two fermentation batches .....	324
Figure 5-19 PCA-X scores plot of SBT687 and SBT692 fractions from first and second fermentation batches .....	325
Figure 5-20 MVDA analysis of mass spectral data of SBT92 fractions from first fermentation batch versus SBT687 fraction .....	328
Figure 5-21 MVDA analysis of mass spectral data of SBT92 fractions from second fermentation batch versus SBT687 fraction .....	330
Figure 5-22 MVDA analysis of mass spectral data of SBT92 fractions from first and second fermentation batches versus SBT687 fractions .....	332
Figure 5-23 PCA-X score plot of <sup>1</sup> H NMR data from SBT692 fractions from first and second fermentations and SBT687 fractions .....	333
Figure 5-24 MVDA analysis of <sup>1</sup> HNMR spectral data of SBT92 fractions from first fermentation batch versus SBT687 fraction .....	336
Figure 5-25 MVDA analysis of <sup>1</sup> HNMR spectral data of SBT92 fractions from second fermentation batch versus SBT687 fraction .....	338
Figure 5-26 MVDA analysis of <sup>1</sup> HNMR spectral data of SBT92 fractions from first and second fermentation batches versus SBT687 fractions .....	341
Figure 5-27 PCA metabolic screening of SBT692 (2 <sup>nd</sup> fermentation) .....	343
Figure 5-28 MVDA first analysis of mass spectral data of SBT92 fractions from second fermentation batch according to bioactivity .....	345
Figure 5-29 MVDA second analysis of mass spectral data of SBT92 fractions from second fermentation batch according to bioactivity against TNF-alpha .....	347
Figure 5-30 Contribution scores plots for F11, F12 and F13 to indicate the orthogonal VIP .....	348
Figure 5-31 Contribution scores plots for F9 to indicate the orthogonal VIP .....	349
Figure 5-32 MVDA analysis of NMR data of SBT92 fractions from second fermentation according to bioactivity .....	353
Figure 5-33 Major secondary metabolites identified from the dereplication of the outlying fraction being produced by the <i>Micromonospora</i> sp. N74 (SBT692) .....	355

Figure 5-34 (continued) Major secondary metabolites identified from the dereplication .....	356
Figure 5-35 Major secondary metabolites identified from the dereplication of the slightly bioactive fraction .....	359

## **Appendices & Supplemental Material**

App 1 <sup>1</sup> H NMR (500 MHz) spectrum for N-phenethylacetamide.....	393
App 2 <sup>1</sup> H- <sup>1</sup> H COSY NMR (500 MHz) spectrum for N-phenethylacetamide.....	394
App 3 <sup>13</sup> C NMR (500 MHz) spectrum for N-phenethylacetamide.....	395
App 4 <sup>1</sup> H- <sup>13</sup> C HMBC NMR (500 MHz) spectrum for N-phenethylacetamide.....	396
App 5 <sup>1</sup> H- <sup>13</sup> C HSQC NMR (500 MHz) spectrum for N-phenethylacetamide.....	397
App 6 <sup>1</sup> H NMR (400 MHz) spectrum for cholic acid derivative .....	398
App 7 Stacked <sup>13</sup> C (1) and DEPT (2) NMR (100 MHz) spectra for cholic acid derivative .....	399
App 8 <sup>1</sup> H- <sup>1</sup> H COSY NMR (400 MHz) spectrum for cholic acid derivative .....	400
App 9 <sup>1</sup> H- <sup>13</sup> C HMBC NMR (400 MHz) spectrum for cholic acid derivative .....	401
App 10 <sup>1</sup> H- <sup>13</sup> C HMQC NMR (400 MHz) spectrum for cholic acid derivative .....	402

# Table of Contents

Author's Declaration.....	ii
Acknowledgements.....	iii
List of Tables.....	vi
List of Figures.....	viii
Table of Contents.....	xiv
Abstract.....	xx
Abbreviations.....	xxii
Chapters.....	xxvii
Chapter 1.....	28
1.1 Introduction.....	29
1.2 Drug discovery.....	30
1.2.1 The re-emergence of natural products.....	30
1.3 Marine bacteria as a wealthy resource of secondary metabolites.....	34
1.3.1 Extreme marine microorganisms and their ability to survive in harshest environments.....	46
1.3.2 Secondary metabolites isolated from thermophilic bacteria.....	47
1.3.3 <i>Flavobacteriaceae</i> family.....	49
1.3.4 Micromonosporaceae family.....	52
1.4 Metabolomics technologies and platforms.....	53
1.4.1 Mass spectrometry (MS).....	55
1.4.2 Nuclear magnetic resonance (NMR).....	56
1.4.3 Multivariate approaches and manual-based metabolite structure elucidation.....	56
1.5 Mining genomes for natural products discovery.....	58
1.5.1 Silent (cryptic) biosynthetic gene clusters (BGC).....	62

1.5.2	Genome sequencing platforms and <i>in silico</i> approaches .....	64
1.6	Aims of the project .....	70
1.7	Objectives of the project.....	71
Chapter 2	.....	72
2.1	General Reagents and solvents .....	73
2.2	General equipment .....	73
2.2.1	Microbiology equipment.....	74
2.3	Bacterial culturing .....	74
2.3.1	<i>Muricauda ruestringensis</i> (SBT531 and SBT587) .....	74
2.3.2	<i>Micromonospora sp.</i> N17 and N74 (SBT687 and SBT692).....	75
2.4	Extraction .....	77
2.5	Liquid Chromatography – Mass Spectrometry (LC-MS) .....	78
2.5.1	Material and instruments.....	78
2.5.2	Methods.....	78
2.6	Medium pressure liquid chromatography (MPLC) and preparative high-performance liquid chromatography (prep-HPLC).....	79
2.6.1	Material and instruments.....	79
2.6.2	Methods.....	81
2.7	Thin layer chromatography (TLC) and preparative TLC.....	86
2.7.1	Materials and instruments .....	86
2.7.2	Methods.....	87
2.8	Nuclear Magnetic Resonance .....	88
2.8.1	Material and instruments.....	88
2.8.2	Methods.....	88
2.9	Software.....	89
2.9.1	Data analysis tools for LC-MS data.....	89

2.9.2	NMR.....	91
2.10	Optical Rotation.....	91
2.11	Bioassay screening.....	92
2.11.1	Materials, reagents, instruments and software.....	92
2.11.2	Cell culture: media preparation, cell splitting and seeding.....	93
2.11.3	Sample preparation.....	95
2.11.4	Alamar Blue assay: cell viability and cytotoxicity.....	96
2.11.5	NFκB Luciferase method.....	96
2.11.6	PTP1B assay.....	98
2.11.7	Alpha-glucosidase assay.....	100
2.12	Genomics.....	102
2.12.1	DNA extraction.....	102
2.12.2	DNA quality/quantity controls.....	103
2.12.3	Sequencing and bioinformatic analysis.....	104
Chapter 3	.....	105
3.1	Introduction.....	106
3.1.1	Aim(s).....	108
3.2	Small scale-up fermentation and extraction of SBT531.....	108
3.3	Medium scale-up fermentation, extraction and fractionation of SBT531.....	111
3.3.1	Fractionation of SBT531.....	111
3.4	Metabolomic-guided screening of SBT531.....	121
3.4.1	Metabolomic profiling of fractions of the crude extract of SBT531.....	121
3.5	Metabolomic-guided sub-fractionation of SBT531 fractions F4, F6+F7 and F9.....	138
3.5.1	Biological activity of SBT531 subfractions.....	145
3.6	Identification and structure elucidation of SBT531 compounds.....	145

3.6.1	Aseanostatin P6.....	147
3.6.2	2-Hydroxy-14-methylhexadecanoic acid.....	156
3.6.3	3-Hydroxy-13-methyltetradecanoic acid .....	165
3.6.4	<i>N</i> -Phenethylacetamide.....	176
3.6.5	Cholic acid derivative .....	178
3.7	Bioreactor fermentation of SBT531 .....	180
3.8	Genomics profile of SBT531 .....	192
3.8.1	Sequencing and genome annotations .....	192
3.8.2	Analysis of secondary metabolites biosynthetic gene clusters .....	192
3.9	Small scale-up fermentation and extraction of SBT587 .....	200
3.9.1	NMR and HR-LC/MS comparisons with SBT531 .....	201
3.10	Discussion.....	206
Chapter 4.....		211
4.1	Introduction .....	212
4.2	Small and Medium scale up fermentation, extraction and fractionation of SBT687 .....	214
4.2.1	Scaled-up fermentation of SBT687.....	214
4.2.2	Fractionation of SBT687.....	215
4.3	Metabolomic guided screening of SBT687.....	222
4.3.1	Metabolomic profiling of fractions of the crude extract of SBT687 .	222
4.4	Identification and structure elucidation of SBT687 compounds.....	249
4.4.1	4-epi-manoyl oxide acid.....	249
4.4.2	Manzamines .....	266
4.5	Discussion .....	273
4.5.1	Diterpenes as promising insecticidal agents .....	273

4.5.2	Manzamines the $\beta$ -carboline alkaloid compounds with significant insecticidal/antifeeding activity .....	280
Chapter 5	.....	287
5.1	Introduction .....	288
5.1.1	Anti-cancer BioScreening Campaign.....	288
5.1.2	Aim(s) .....	291
5.2	Small and Medium scale up fermentation, extraction and fractionation of SBT692 .....	292
5.2.1	Scaled-up fermentation of SBT692.....	292
5.2.2	Fractionation of SBT692.....	292
5.3	Metabolomic screening of SBT692.....	311
5.3.1	Metabolomic profiling of the crude extract and fractions of SBT692 for both 15 L scale-up fermentations.....	311
5.3.2	Metabolite profile comparison between SBT692 and SBT687 .....	325
5.3.3	Multivariate analysis of slightly bioactive fractions obtained from the second fermentation batch of SBT692.....	341
5.4	Discussion .....	354
Chapter 6	.....	361
6.1	Conclusions .....	362
6.1.1	<i>Muricauda ruestringensis</i> (SBT531 and SBT587) metabolomic profile and biological importance of isolated metabolites as potential targets for metabolic diseases.....	362
6.1.2	<i>Micromonospora</i> sp. N17 (SBT687) and N74 (SBT692) potential agrochemical and anti-cancer biological activity from isolated/dereplicated compounds and metabolomic profile.....	363
6.2	Future work .....	364
List of References	.....	366
Appendices & Supplemental Material	.....	392

Appendix I: NMR data of <i>N</i> -phenethylacetamide .....	393
Appendix II: NMR data of a cholic acid derivative .....	398
Appendix III: NMR data of 4-epi-manoyl oxide acid.....	403

## Abstract

Marine microorganisms produce unique secondary metabolites, which are responsible for a variety of biologically active molecules with a wide range of pharmaceutical properties. There is a lack of novel effective drugs for metabolic diseases, cancer and parasite infections as well as TNF-alpha inhibitors hence; underexplored marine bacteria could be an important source for new bioactive molecules.

Two strains of *Muricauda ruestringensis* (SBT531 and SBT587) were isolated from geothermal intertidal pools in Iceland and two strains of *Micromonospora* sp. N17 and N74 (SBT 687 and SBT692) were isolated from the Mediterranean sponge *Phorbastenacior* from the Santorini volcanic complex of Crete. Bioactive metabolites production in thermophile strains of *M. ruestringensis* (SBT531 and SBT587) and *Micromonospora* sp. (SBT687 and SBT692) were identified and isolated, by using a metabolomics approach to analyse the liquid chromatography-high resolution mass spectrometry (LC-HRMS) and nuclear magnetic resonance (NMR) data sets. The LC-HRMS data was processed by using the modified version of Mzmine 2.10 software, dereplicated with an in-house EXCEL macro coupled to the AntiMarin and Dictionary of Natural Products (DNP) database to be statistically evaluated by multivariate analysis in SIMCA v15.02. Orthogonal partial least squares discriminant analysis (OPLS-DA) in SIMCA was used to predict and pinpoint the biologically active secondary metabolites. Up-scaling was optimised to increase the production yield of the target metabolites. Additionally, specific bioassay screening determined the activity, if any, from the crude extracts, fractions and isolated compounds.

The total ethyl acetate organic extracts of SBT531 and SBT587 were active inhibitors in target-based functional assays: alpha-glucosidase and protein-tyrosine phosphatase 1B (PTP1B) that are a therapeutic target for the treatment of diabetes and other metabolic syndromes. The fractionation of SBT531 afforded a series of bioactive alpha and beta-hydroxy acid derivatives and allowed the definition of a preliminary structure-activity relationship based on their relative potency. Aseanostatin P6 (13-methyltetradecanoic acid) was the major compound identified, followed by two derivatives, 3-hydroxy-13-methyltetradecanoic acid and 2-hydroxy-14-

methylhexadecanoic acid. On the other hand, the initial organic crude extracts of SBT687 and SBT692 showed inhibition activity against sea lice and inhibition of TNF-alpha respectively, however after further fermentation scale-up the fractions were missing the inhibition activity.

## Abbreviations

<b>ABBV-399</b>	Telisotuzumab vedotin
<b>ABT-414</b>	Depatuxizumab mafodotin
<b>ACN</b>	Acetonitrile
<b>ACP</b>	Acyl carrier protein
<b>ADC</b>	Antibody drug conjugate
<b>AP-6</b>	Aseanostatin p6
<b>APC</b>	Adenomatous polyposis coli protein
<b>ApoE</b>	Apolipoprotein E
<b>Ara-A</b>	Vidarabine
<b>Ara-C</b>	Cytarabine
<b>AXIN2</b>	Axin-2
<b>BAX</b>	Apoptosis regulator bax
<b>BC</b>	Biotin carboxyl
<b>BCCP</b>	Biotin carboxyl carrier protein
<b>BGC</b>	Biosynthetic gene cluster
<b>BHA</b>	Beta hydroxy fatty acid derivatives
<b>bp</b>	Base pairs
<b>bs</b>	Broad singlet
<b>BSA</b>	Bovine serum albumin
<b>C10</b>	Monoterpenes
<b>CaCo2</b>	Human colon adenocarcinoma cell line
<b>CDCl3</b>	Deuterated chloroform
<b>CDX-011</b>	Glembatumumab vedotin
<b>CE</b>	Capillary electrophoresis
<b>CH</b>	Methine group
<b>CH2</b>	Ethyl group
<b>CH3</b>	Methyl group
<b>COSY</b>	Correlated spectroscopy
<b>CRC</b>	Colorectal cancer

<b>CYP</b>	Cytochrome p450
<b>DCC</b>	Netrin receptor dcc
<b>DCDS-4501A</b>	Polatuzumab vedotin
<b>DCM</b>	Dichloromethane
<b>DCW</b>	Dry cell weight
<b>DiFMUP</b>	6,89-difluoro-4-methylumbelliferyl phosphate
<b>diTPS</b>	Diterpene synthase
<b>DMEM</b>	Dulbecco's modified eagle medium
<b>DMSO</b>	Dimethyl sulfoxide
<b>DMXBA</b>	3-(2,4-dimethoxybenzylidene) anabaseine
<b>DNP</b>	Dictionary of natural products
<b>E7389</b>	Eribulin mesylate
<b>EC</b>	Enzyme nomenclature
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>ELSD</b>	Evaporative light scattering detector
<b>ERBB2</b>	Receptor tyrosine-protein kinase ERBB-2
<b>ESI</b>	Electrospray ionization
<b>ET-743</b>	Trabectedin
<b>EtOAc</b>	Ethyl acetate
<b>fabH</b>	3-oxoacyl-[acp] synthase 3
<b>FBS</b>	Foetal bovine serum
<b>FDA</b>	Food and drug administration
<b>G2PLA5</b>	Putative glycosyltransferase epsj
<b>G2PNR5</b>	3-oxoacyl-[acp] synthase 2
<b>G2PNW2</b>	Peptidase s9
<b>GC</b>	Gas chromatography
<b>GGPP</b>	Geranylgeranyldiphosphate
<b>GI50</b>	50% reduction in proliferation of cancer cells
<b>HBSS</b>	Hanks' balanced salt solution
<b>HCA</b>	Hierarchical clustering
<b>HCT116</b>	Human colorectal carcinoma cell line
<b>HEK293</b>	Human embryonic kidney cell line

<b>HMBC</b>	Heteronuclear multiple bond correlation
<b>HMQC</b>	Heteronuclear multiple quantum coherence
<b>HPLC</b>	High performance liquid chromatography
<b>HR-LC/MS</b>	High resolution liquid chromatograph mass spectrometer
<b>HSQC</b>	Heteronuclear single quantum correlation
<b>HSV-1</b>	Herpes simplex virus type 1
<b>HT29</b>	Human colorectal adenocarcinoma cell line
<b>HTS</b>	High-throughput screening
<b>IκB-alpha</b>	NF-kappa-b inhibitor alpha
<b>IL-1</b>	Interleukin-1
<b>KEGG</b>	Kyoto Encyclopedia of Genes and Genomes
<b>Ki</b>	Inhibitory constant
<b>KRAS</b>	GTPase KRAS
<b>LC50</b>	50% lethal concentration
<b>LC-MS</b>	Liquid chromatography-mass spectrometry
<b>m/z</b>	Mass/charge
<b>M199</b>	Medium 199
<b>MB</b>	Megabase
<b>MCF-7</b>	Human breast adenocarcinoma
<b>MDA-MB-231</b>	Human breast adenocarcinoma
<b>MeOH</b>	Methanol
<b>MG132</b>	Carbobenzoxy-leu-leu-leucinal, proteasome inhibitor
<b>min</b>	Time minutes
<b>MLH1</b>	DNA mismatch repair protein mlh1
<b>MMAE</b>	Monomethylauristatin e
<b>MMAF</b>	Monomethylauristatin f
<b>MPLC</b>	Medium pressure liquid chromatography
<b>MS</b>	Mass spectrometry
<b>MSH2</b>	Muts protein homolog 2
<b>MW</b>	Molecular weight
<b>NCTC</b>	Human normal keratinocytes
<b>NEAA</b>	Non-essential amino acid

<b>NFkB</b>	Nuclear factor NF-kappa-b p105 subunit
<b>NFKBIA</b>	NF-kappa-b inhibitor alpha
<b>NGS</b>	Next-generation sequencing
<b>NMEs</b>	New molecular entities
<b>NMR</b>	Nuclear magnetic resonance
<b>NP</b>	Natural products
<b>NPD</b>	Natural products discovery
<b>NPI-0052</b>	Marizomib
<b>NPI-2358</b>	Plinabulin
<b>NPMG</b>	Natural products metabolomics group
<b>NRPS</b>	Nonribosomal peptides
<b>NSCLC</b>	Non-small-cell lung carcinoma
<b>nt.</b>	Nucleotides
<b>OD</b>	Optical density
<b>OPLS-DA</b>	Orthogonal partial least-squares regression and discriminant analysis
<b>P388D1</b>	Mice lymphoma cell line
<b>p50</b>	Nuclear factor NF-kappa-b p105 subunit
<b>p65</b>	Transcription factor p65
<b>PCA</b>	Principal component analysis
<b>PDA</b>	Photodiode array
<b>PKS</b>	Polyketide synthases
<b>PLS</b>	Partial least squares regression
<b>PM01183</b>	Lurbinectedin
<b>PM184</b>	Anti-tumour drug pm184
<b>PPARA</b>	Peroxisome proliferator activated receptor alpha
<b>PTP1B</b>	Protein-tyrosine phosphatase 1b
<b>QC</b>	Quality control
<b>R2</b>	Coefficient of determination
<b>RAST</b>	Rapid annotation using subsystem technology
<b>Rf</b>	Retention factor
<b>rpm</b>	Revolutions per minute

<b>SBDD</b>	Structure-based drug design
<b>SBSPKS</b>	Structure-based sequence analysis of polyketide synthases
<b>SD</b>	Standard deviation
<b>Ser</b>	Serine
<b>SGN-35</b>	Brentuximab vedotin
<b>SGN-CD19A</b>	Denintuzumab mafodotin
<b>SIDR</b>	Strathclyde Institute for Drugs Research
<b>SIPBS</b>	Strathclyde Institute of Pharmacy and Biomedical Sciences
<b>SKBR-3</b>	Human breast adenocarcinoma cell line
<b>SMAD4</b>	Mothers against decapentaplegic homolog 4
<b>SPE</b>	Solid-phase extraction
<b>STRs</b>	Short-tandem repeats
<b>SW-480</b>	Human colon adenocarcinoma cell line
<b>T84</b>	Human lung-derived colorectal carcinoma cell line
<b>TAE</b>	TridacetateEDTA
<b>TFMS</b>	Bis(4-trifluoromethylsulfonamidophenyl)-1,4-diisopropylbenzine
<b>TFs</b>	Transcription factors
<b>TGFBR2</b>	TGF-beta receptor type-2
<b>TGFBR2</b>	TGF-beta receptor type-2
<b>TLC</b>	Thin-layer chromatography
<b>TNBC</b>	Triple negative breast carcinomas
<b>TNF-alpha</b>	Tumour necrosis factor-alpha
<b>TOCSY</b>	Total correlation spectroscopy
<b>TP53</b>	Cellular tumour antigen p53
<b>TRPA1</b>	Transient receptor potential cation channel subfamily a member 1
<b>TRPM8</b>	Transient receptor potential cation channel subfamily m member 8
<b>ub. med.</b>	Ubiquitin mediated
<b>v/v</b>	Volume/volume
<b>VIP</b>	Variable influence/importance in projection
<b>vvm</b>	Volume of air/per unit of medium/per unit of time
<b>WES</b>	Whole-exome sequencing
<b>WGS</b>	Whole-genome sequencing



# Chapter 1



GENERAL INTRODUCTION

## 1.1 Introduction

Screening for new compounds with pharmacological interest for a specific disease or a disease class has a long history of success and demerit cases. For instance, adoption of high-throughput screening (HTS) methods for early-stage drug discovery directly yielded cyclosporin A, a fungal-derived immunosuppressant medication and a mold-derived agent—mevastatin used to normalise cholesterol levels. Likewise, drug discovery success cases via development of new drug candidates relying upon a structure-based drug design (SBDD) shed-to-light dorzolamide, a topical ophthalmic agent applied in the treatment of glaucoma. Dorzolamide acts by reducing the production of aqueous humour via inhibition of carbonic anhydrases with consequent intraocular pressure improvement. This method also drove to uncover imatinib that is used today in cancer chemotherapy, specifically employed in the treatment of many leukaemia subtypes. Other drugs detected by the same process is vemurafenib, a BRAF inhibitor used as chemotherapeutic agent in late-stages of melanoma. To thrive the path for novel compounds research or boost new drug repurposing programmes, it becomes apparent that an ideal workflow for earlier drug discovery should rely on a whole range of detection tools passing through coupled or used-in-parallel analytical platforms (Roy, 2018).

Since there is an unmet need to develop new metabolomics workflows based on investigation of natural occurring metabolites to enhance drug discovery, the present project focused on the identification of secondary metabolites with interesting bioactivity from two different strains of thermophilic marine bacteria, *Muricauda ruestringensis* and *Micromonospora sp.* A known metabolomic workflow was applied to the mentioned bacteria strains which have been not study by this approach previously. Briefly, this involved *in vitro* cultivation of marine bacteria specimens and further characterization of their bioactive secondary metabolites via a workflow of fractionation, mass spectrometry and nuclear magnetic resonance spectroscopy.

## 1.2 Drug discovery

The development of a new drug, from target identification to the presentation of a final product followed by approval for prescription to the general public by a governmental/ local state authority (e.g. FDA—The Food and Drug Administration, EMA—The European Medicines Agency) involves a multi-step procedure which can easily revolve around 12 to 15 years and posing extremely high costs for companies (Mohs and Greig, 2017, DiMasi et al., 2010, Hughes et al., 2011). This process (Figure 1.1) starts with basic research that includes lead identification, synthesis scale-up and *in vitro* pharmacology, followed by preclinical development, which consist of specific activities such as studies of absorption, distribution, excretion, and metabolism, as well as screening for activity on cytochrome P450 (CYP) liver enzymes, assay metabolites for acute pharmacology and *in vitro* toxicity (Pereira and Aires-de-Sousa, 2018). Clinical trials include Phase I, II, and III that comprises ethics review, *in vivo* metabolism assessment, and as a final point the regulatory filings for new drug application (DiMasi et al., 2010, Hughes et al., 2011).

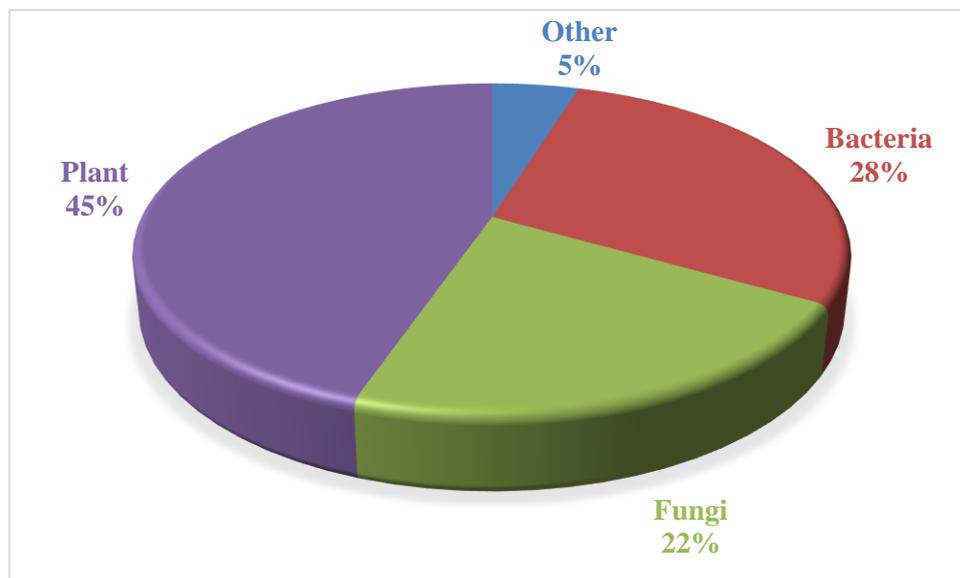


**Figure 1-1 A schematic of the process in drug discovery and development.** Adapted from (Hughes et al., 2011).

### 1.2.1 The re-emergence of natural products

Nature always has been a supply of medical products, with more than 80% of drug substances obtained from natural products (NPs) (Harvey, 2008, McChesney et al., 2007). NPs are compounds, with potential biological activity, derived (in part or as a whole) from natural sources like plants, terrestrial and marine microorganisms (Baker et al., 2007, Cragg and Newman, 2013). To date; plants, fungi and bacteria are the class of organisms that yielded the most comprehensive number of novel molecular

entities (Figure 1.2) either as directly isolated NPs or as initial leads for the synthesis of bioactive molecules (Cragg and Newman, 2013, Onaka et al., 2011, Yang et al., 2014).



**Figure 1-2 Discovery of new molecular entities (NMEs) and natural products contribution.** The figure shows natural product new molecular entities (NMEs) from 1931 to 2013. The cumulative percentage of Food and drug administration (FDA) – approved natural product NMEs, categorized by environmental source (Patridge et al., 2016).

NPs have played an important global role in the treatment and prevention of human diseases, while they have a long history of clinical use, better patient acceptance and tolerance (Chen et al., 2015, Kumar et al., 2014). Moreover, NPs have been and continue to be the origin for a substantial portion of human therapeutics with more than 90% of current therapeutic classes are derived from a natural product prototype (McChesney et al., 2007). Looking back to the period from 1981 to 2006, of the total of 1184 compounds approved for Human disease intervention, only 30% were of synthetic origin while roughly 33% were represented by NPs; the remaining is reasoned to be of either of semi-synthetic or synthetic NPs derivatives origin (Patridge et al., 2016, Romano et al., 2017).

Since the early '90s, many pharmaceutical companies entirely abandoned the approach of using NPs as the main source of lead compounds (i.e. compounds with pharmaceutical/ biological activity) due to raised difficulties in finding supplies of raw precursors, as well as due to technical and analytical issues concerning molecular

screening and characterization (McChesney et al., 2007). These issues were in part solved by the adoption of state-of-the-art approaches, such as high-throughput screening (HTS), *in silico* molecular modeling, combinatorial synthesis between other methodologies, which in overall contributed to the establishment of small molecules synthetic libraries (McChesney et al., 2007). Nevertheless, recently, the Nobel Prize in Physiology or Medicine of 2015 was shared by two leading investigators, William C. Campbell and Satoshi Omura regarding the discovery of the bacterial-sourced natural product, avermectin; in which its derivative, ivermectin can be applied to reduce the incidence of an array of infectious conditions related with parasitic worms such as onchocerciasis (river blindness) and lymphatic filariasis (elephantiasis). Another awardee is Tu Youyou, for her discovery of artemisinin, a plant-derived natural product as a new therapy for malaria. Both discoveries were a great boost in the renewed positivity for sourcing lead compounds from natural products (NPs) (Tambo et al., 2015). Such event shed light upon the entire field, that a New Golden Age in NPs discovery is arising, reviving the earlier 50's and 60's, which enclosed the glorious discovery of broad range antibiotics such as penicillin and streptomycin (Shen, 2015).

The value of NPs can be assessed using diverse criteria such as the rate of introduction of new chemical entities of large structural diversity, including serving as templates for semi-synthetic and total synthetic modification, the frequency of use in the treatment of disease and the number of diseases treated or prevented by these substances (Onaka et al., 2011). Regardless of the recognized opulent record of NPs in drug discovery and their unique structural diversity, there are still several problems associated with NPs (Amirkia and Heinrich, 2015). Characteristic limitations of these compounds are low solubility or chemical instability, which especially obstruct the development of any formulation type drugs. Furthermore, many NPs are complex structures while the concentrations of the active constituents in crude extracts are largely unknown (Baker et al., 2007, Chen et al., 2015). After the isolation of the pure compounds, it is essential to elucidate the chemical structure, which is an enduring and challenging process (McChesney et al., 2007). Major developments in fermentation optimization, purification and structure elucidation enabled faster access to higher

yields of pure compounds (Lam, 2007, Eldridge et al., 2002). Combination of separation and high-throughput technologies involving high-pressure liquid chromatography (HPLC) and solid phase extraction (SPE) with nuclear magnetic resonance (NMR) and mass spectrometry (MS) revealed a significant impact in decreasing dereplication time, isolation and structure elucidation of natural products (NPs) in crude extracts (Lam, 2007, Exarchou et al., 2005).

The field is evolving thanks to combinatorial approaches using latest advances in genomics, metagenomics and cheminformatics which can impact on the rapidly identification of gene clusters encoding bioactive NPs, a fundamental step for understanding natural products biosynthesis (Shen, 2015). The molecular biology-based approaches enhance the elucidation of enzymatic phases involved in the biosynthesis of microbial secondary metabolites applying the identification of biosynthetic gene clusters (Cacho et al., 2014). Currently the NPs research programs are inspired by the understanding of the molecular and enzymatic basis of secondary metabolites biosynthesis and the relationship between the secondary metabolite's molecules and the correspondent biosynthetic gene clusters (Walsh and Fischbach, 2010).

Natural Products Discovery (NPD) is employed by Pharma and Academia to produce compounds that can tackle bacterial and fungal infections, develop novel compounds for cancer therapy, and prevent organ allografts rejection as well hypercholesteremia (Dias et al., 2012). NPs discovery has been driven either by bioactivity-guided fractionation of crude broth extracts, or via chemical screening after isolation of metabolites with desirable spectroscopic properties (Bachmann et al., 2014). Between late 90s to early 2000s, several of the bigger pharmaceutical companies withdrew their NPD programmes and sold their collections of screening extracts, because of the advent of HTS that at the time was promoted as a better method for drug discovery. The conventional extract-based screening was believed to result in the continuous re-discovery of previously isolated compounds and was economically and synthetically challenging (Ojima, 2008, Newman, 2008). Molecular target-based drug discovery, which applies combinatorial libraries to obtain efficient "*hits*", replaced the "*classical natural product chemistry*" for the past twenty years (Ojima, 2008). The advancements

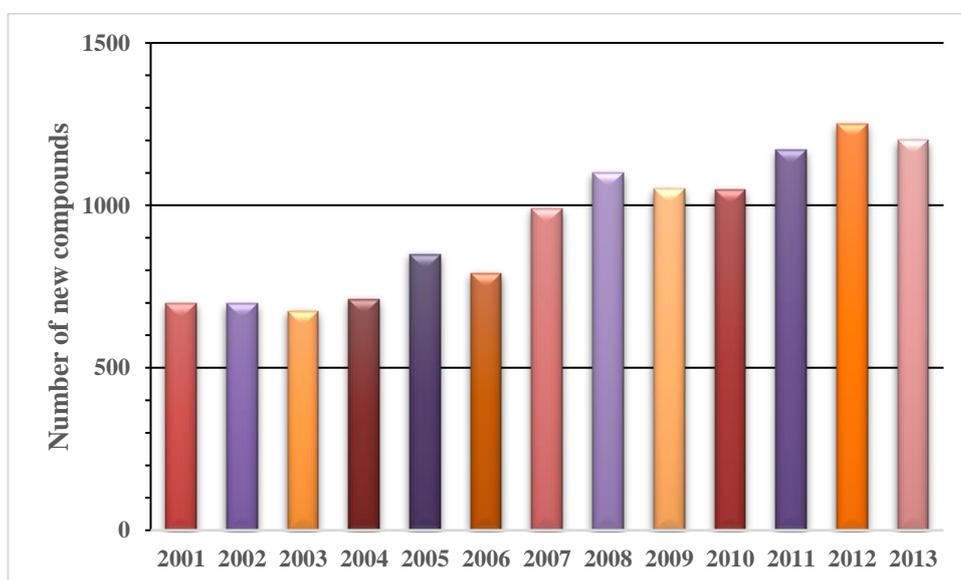
in analytical instrumentation and sophisticated hyphenation of separation techniques with highly sensitive detectors have allowed for better detection of small molecule compounds measurable in biological systems, such as, primary and secondary metabolites and unquestionably will now be used to advance the discovery of natural product chemistry to identify potential novel drugs candidates (Newman, 2008). The combinatorial chemistry allowed the ability to produce drug-like molecules, increasing the size of the known chemical space, DNA sequencing was responsible for the identification of new drug targets, improvements in the elucidation of three-dimensional protein structures by X-ray crystallography helped the identification of lead compounds via structure-guided strategies and computational drug design and screening were employed with laboratories becoming progressively computerized and automated (Pereira and Aires-de-Sousa, 2018, Eder and Herrling, 2016). The effective drug development from new structural leads from nature requires multidisciplinary collaborations (Cragg and Newman, 2013). Integration of natural products chemistry with synthetic chemistry, medicinal chemistry, pharmacology, biology, and associated disciplines ensures a greater chance to advance natural product leads into therapeutic drugs (Galm and Shen, 2007).

### **1.3 Marine bacteria as a wealthy resource of secondary metabolites**

Oceans represent the largest reservoir of ecological, chemical and biological diversity as well as have borne most of the biological activities on our planet, having a great potential for the development of new NPs (Bhatnagar and Kim, 2010, Dalmaso et al., 2015, Romano et al., 2017). Several biologically active compounds with varying degrees of action have been isolated from marine sources throughout the years (Figure 1.3) and until 2018 more than 28,000 marine natural products have been reported derived from a variety of marine sources (Pereira and Aires-de-Sousa, 2018).

Marine environment is a rich source of bioactive compounds, favoring the production of great variety of novel molecules, many of which are unique in terms of diversity,

and structural and functional features not found in terrestrial sources (Romano et al., 2017). The marine environment also represents a largely unexplored source for isolation of new microbes such as bacteria, fungi, actinomycetes, microalgae-cyanobacteria and diatoms, that are producers of potent bioactive secondary metabolites (Bhatnagar and Kim, 2010). Exploitation of their secondary metabolism has guaranteed the discovery of compounds with chemical characteristics and biological properties not existing in screening libraries of synthetic compounds (Marinelli et al., 2015). The systematic investigation of marine environments as sources of novel biologically active agents began around the 1970s and throughout the years an increasing number of new compounds from marine sources were isolated (Cragg and Newman, 2013, Romano et al., 2017).



**Figure 1-3 Number of new compounds isolated from marine organisms from 2001 to 2013.** Adapted from: (Romano et al., 2017).

During the last decades of the 20th century, research on natural products with special emphasis on the ability of marine organism to produce secondary metabolites, revealed the importance of these substances for the treatment of several human diseases. Although numerous bioactive substances have been identified, only recently the first marine drugs were approved (Imhoff et al., 2011). Marine natural products chemists and pharmacologists were responsible for the development of several molecules now available on market or being assessed in clinical and pre-clinical trials. By March 2018, FDA approved seven marine-derived drugs (four anticancer, one antiviral, one

pain control, and one for hypertriglyceridemia) and several ones in clinical trials: six compounds in Phase III, ten in Phase II and six in Phase I as shown on Table **1.1** (Mayer et al., 2010, Gerwick and Moore, 2012).

**Table 1-1 Up-to-date pipeline perspective of the marine pharmaceuticals.** Adapted from: (Mayer et al., 2010, Gerwick and Moore, 2012)

Clinical status	Compound name	Trademark	Marine organism	Chemical class	Company or Institution	Disease area
Approved	Trabectedin (ET-743)	Yondelis® (2015)	Tunicate	Alkaloid	Pharmamar	Cancer: Soft Tissue Sarcoma and Ovarian
	Brentuximab vedotin (SGN-35)	Adcetris® (2011)	Mollusk/ Cyanobacterium	Antibody Conjugate Monomethylauristatin E (MMAE)	Drug Seattle Genetics	Cancer: Anaplastic large T-cell systemic malignant lymphoma, Hodgkin's disease
	Eribulin Mesylate (E7389)	Halaven® (2010)	Sponge	Macrolide	Elisai Inc.	Cancer: Metastatic Breast Cancer  Hypertriglyceridemia

	Omega-3-acid ethyl esters	Lovaza® (2004)	Fish	Omega-3-fatty acids	GlaxoSmithKline	Pain: Severe Chronic Pain
	Ziconotide	Prialt® (2004)	Cone snail	Peptide	Jazz Pharmaceutical	Antiviral: Herpes Simplex Virus
	Vidarabine, (Ara-A)	Vira-A® (1976)	Sponge	Nucleoside	Mochida Pharmaceutical Co.	Cancer: Leukemia
	Cytarabine (Ara-C)	Cytosar-U® (1969)	Sponge	Nucleoside	Pfizer	
Phase III	Plinabulin (NPI-2358)	NA	Fungus	Diketopiperazine	BeyondSpring Pharmaceuticals	Cancer: Non-Small Cell Lung, Brain

	Plitidepsin	Aplidin®	Tunicate	Depsipetide	Pharmamar	Cancer: Multiple Myeloma, Leukemia, Lymphoma
	Tetrodotoxin	Tectin®	Pufferfish	Guanidinium alkaloid	Wex Pharmaceutical Inc.	Pain: Chronic Pain
	Lurbinectedin (PM01183)	NA	Tunicate	Alkaloid	Pharmamar	Cancer: Ovarian, Breast, SCLC
	Depatuzumab mafofotin (ABT-414)	NA	Mollusk/ Cyanobacterium	ADC Monomethylauristatin F (MMAF)	AbbVie	Cancer: Glioblastoma, Paediatric Brain tumours
	Polatuzumab vedotin (DCDS-4501A)	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Genentech/Roche	Cancer: Non-Hodgkin lymphoma, Chronic lymphocytic leukemia, Lymphoma, B-Cell, lymphoma follicular
Phase II	GTS-21 (DMXBA)	NA	Worm	Alkaloid	-	Schizophrenia, Alzheimer disease, Attention deficit,

					Hyperactivity disorder, Endotoxemia, Sepsis, Vagal activity Cancer: relapsed/refractory and frontline DLBCL
Denintuzumab mafodotin (SGN-CD19A)	NA	Mollusk/ Cyanobacterium	ADC (MMAF)	Seattle Genetics	
AGS-16C3F	NA	Mollusk/ Cyanobacterium	ADC (MMAF)	Agensys & Astellas Pharma Pharmamar	Cancer: Renal cell carcinoma
Plocabulin (PM184)	NA	Sponge	Polyketide	GenMAb	Cancer: Solid tumours
Tisotumab Vedotin	HuMax®- TF-ADC	Mollusk/ Cyanobacterium	ADC (MMAE)		Cancer: Ovary, Cervix, Endometrium, Bladder, Prostate, Head and Neck,

Enfortumab Vedotin ASG-22ME	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Seattle Genetics	Esophagus, Lung (NSCLC) Cancer: Neoplasms, Metastatic urothelial cancer
Glembatumumab Vedotin (CDX-011)	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Celldex Therapeutics	Cancer: Metastatic breast cancer, metastatic melanoma, Triple negative breast cancer
GSK2857916	NA	Mollusk/ Cyanobacterium	ADC (MMAF)	GalaxoSmithKline	Cancer: Multiple myeloma
Ladiratumumab vedotin	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Seattle Genetics Neurotrope BioScience	Cancer: Breast

						Alzheimer's disease
	Bryostatin	NA	Bryozan	Macrolide lactone		
Phase I	ABBV-085	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Abbvie	Cancer: Solid tumours
	Telisotuzumab					Cancer: Solid tumours
	vedotin	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Abbvie	Cancer: Solid tumours
	(ABBV-399)					
	ABBV-221	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Abbvie& Genetics	Seattle Cancer: Refractory and relapsed Lymphoid malignancy
ASG-67E	NA	Mollusk/ Cyanobacterium	ADC (MMAE)	Astellas& Genetics	Seattle Cancer: Metastatic urothelial Cancer	
ASG-15ME	NA	Mollusk/	ADC (MMAE)			

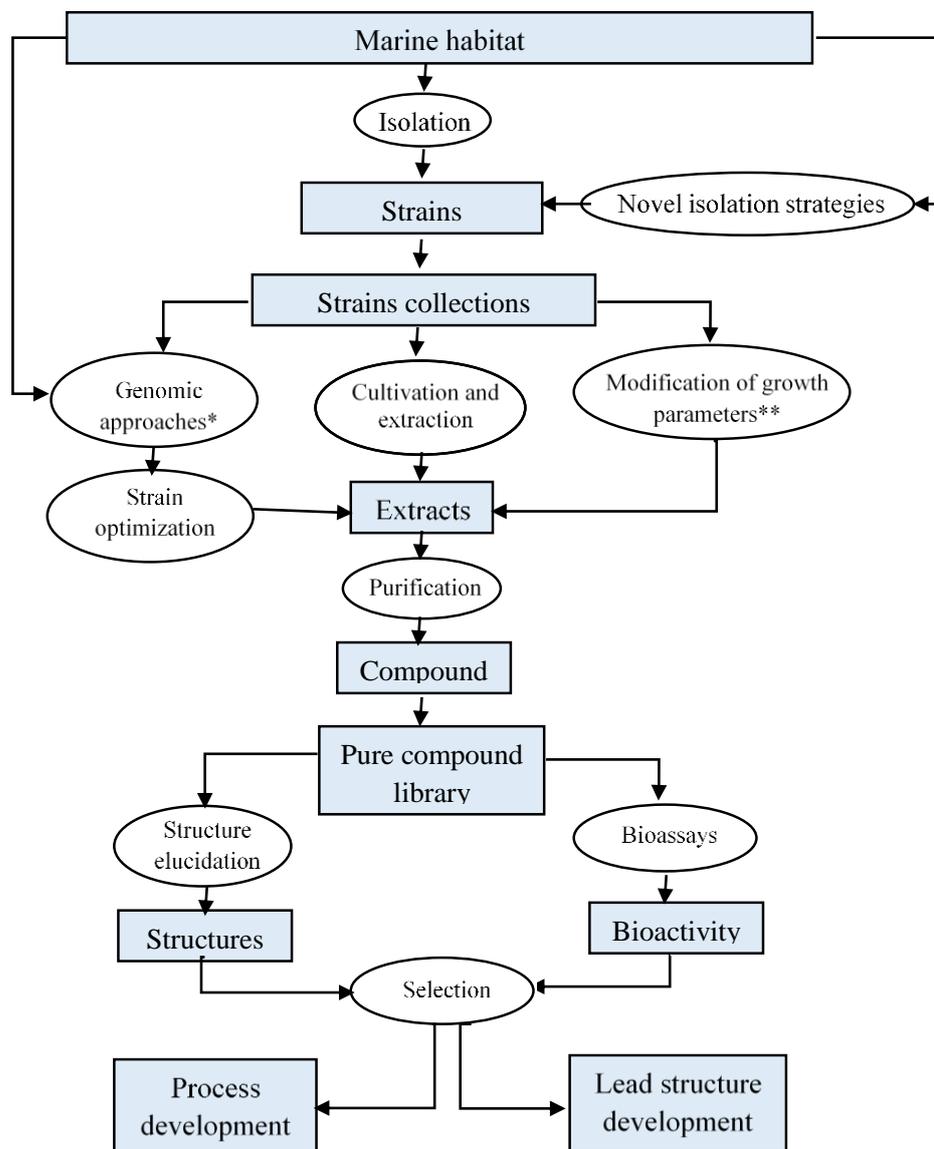
			Cyanobacterium		Astellas& Genetics	Seattle	Cancer: Non-Small cell lung Cancer, Pancreatic, Melanoma, Lymphoma, Multiple Myeloma
Marizomib (Salinosporamid eA; NPI-0052)	NA	Bacterium	Beta-lactone-gamma lactam		Triphase		

**Abbreviations:** **ADC:** Antibody Drug Conjugate; **Ara-A:** 9-β-D-arabinofuranosyladenine; **Ara-C:** Cytosine Arabinoside; **BCMA:** B-cell Maturation Antigen; **CD:** Cluster of Differentiation; **c-Met:** Tyrosine-Protein Kinase Met; **DMXBA:** 3-(2,4-Dimethoxybenzylidene)-Anabeseine; **EGFR:** Epidermal Growth Factor Receptor; **ENPP3:** Ectonucleotide Pyrophosphatase/ Phosphodiesterase Family Member 3; **ET:** Ecteinascidin; **FDA:** Food and Drug Administration; **GCC:** Guanylyl Cyclase C; **GPNUMB:** Glycoprotein nonmetastatic B; **JNK:** c-Jun N-terminal protein kinases; **LIV-1:** Zinc transporter SLC39A6; **LY6E:** Lymphocyte Antigen 6 Complex, Locus E; **MMAE:** Monomethylauristatin E; **MMAF:** Monomethylauristatin F; **NA:** Not Available; **NaPi2b:** Sodium-Dependent Phosphate Transport Protein 2b; **PSMA:** Prostate-Specific Membrane Antigen; **RAC1:** Ras-related C3 botulinum toxin substrate 1; **SLITRK6:** SLIT and NTRK-like protein 6.

Microorganisms, including bacteria, fungi and algae, produce secondary metabolites; which may have some bioactivity; were studied extensively over the past 30 years to isolate and elucidate novel compounds from marine sources (Bhatnagar and Kim, 2010, Kumar et al., 2014). These metabolites may be effective against infectious diseases, conditions of multiple bacterial infections and have also been found to be useful against carcinomas, risk of coronary heart disease, or may act as immune-suppressants to aid in organ transplantation, consequently making the microbial secondary metabolites an enormous source of pharmaceutical importance. Along with the potential sources of natural products, bacteria have proven to be a particularly abundant resource with a small group of taxa accounting for most discovered compounds (Jensen et al., 2005). Marine bacteria are prolific producers of such secondary metabolites as they succeed in harsh oceanic climates. Bacteria produce some secondary metabolites for their defence against other microorganisms and these secondary metabolites serve as a source of bioactive compounds for use in human therapies (Bhatnagar and Kim, 2010).

The value of natural products from marine species has been acknowledged for the past fifty years, but it is only in recent years that there has been a renewed interest in this source of new medicines. Chemical structure and pharmacological characterizations of marine natural products libraries have successfully identified many “hit” compounds (Yang et al., 2014). Structurally unique secondary metabolites have been isolated and identified from marine organisms (Kumar et al., 2014). Bacterial marine natural products are an important source of novel lead structures and the high hit rates in lead drug screening and large-scale structural diversity make marine natural products ideal candidates for drug discovery (Gulder and Moore, 2009, Yang et al., 2014). The biomedical potential of marine bacterial agents, which could be biotechnologically scaled-up, might help overcome the problems of supply and sustainability encountered with microorganisms or marine invertebrates from the sea such as sponges, tunicates, and soft corals. The remarkable capabilities of these microbes to produce chemically unique bioactive molecules are reinforced by the characterization of a wealth of interesting new structures through the application of classical screening and isolation techniques as shown on Figure 1.4., the steps of

further drug development of natural products from marine microorganisms (Gulder and Moore, 2009, Imhoff et al., 2011).



**Figure 1-4 Flow chart from marine habitat to compound selection** for further drug development of natural products from marine microorganisms. \*Genomic approaches include genome mining, metagenomics approaches and molecular manipulation of microbes to activate “silent” gene clusters. \*\*Modification of growth parameters: pH, temperature, and nutrients. In addition, stimulation, co-cultivation and use of epigenetic modulators may modify the secondary metabolite pattern of microorganisms. Adapted from (Imhoff et al., 2011).

### **1.3.1 Extreme marine microorganisms and their ability to survive in harshest environments**

Marine microorganisms have evolved for millions of years in order to survive in the severe surroundings characterized by extreme physical or chemical parameters like pressure, temperature, pH, salinity, oxidative stress, hazardous pathogens, radiation, metal and chemicals, which some of these are toxic (Canganella and Wiegel, 2014, Hu et al., 2016, Nazarenko et al., 2011). Microorganisms living in the deep sea have several unique features specifically adapted for such an extreme environment to be able to live, grow, and thrive in this environment (Horikoshi, 1998). The diverse conditions in marine environments have forced on a bacterial selection leading to new adaptive strategies and therefore the synthesis of new metabolites. Many of these compounds produced by the bacteria are referred as secondary metabolites and its production can be affected by adaptations to the sea conditions (de Carvalho and Fernandes, 2010).

Extremophile is an organism that successfully thrive in an extreme environment being able to adapt and survive to extreme conditions due to their highly flexible metabolism. The term extremophile describes any organism capable of living and growing under extreme conditions, that can refer to physical extremes (temperature, pressure or radiation) but also to geochemical extremes such as salinity and pH, which have attracted the attention of research centres in the pursuit for new potential bioactive substances useful in different fields (Canganella and Wiegel, 2014, Rothschild and Mancinelli, 2001, van den Burg). Although all hyperthermophiles are members of the Archaea and Bacteria, eukaryotes are common among the psychrophiles, acidophiles, alkaliphiles, piezophiles, xerophiles and halophiles (which respectively survive at low temperatures, low pH, high pH, and under extremes of pressure, desiccation and salinity) (Rothschild and Mancinelli, 2001).

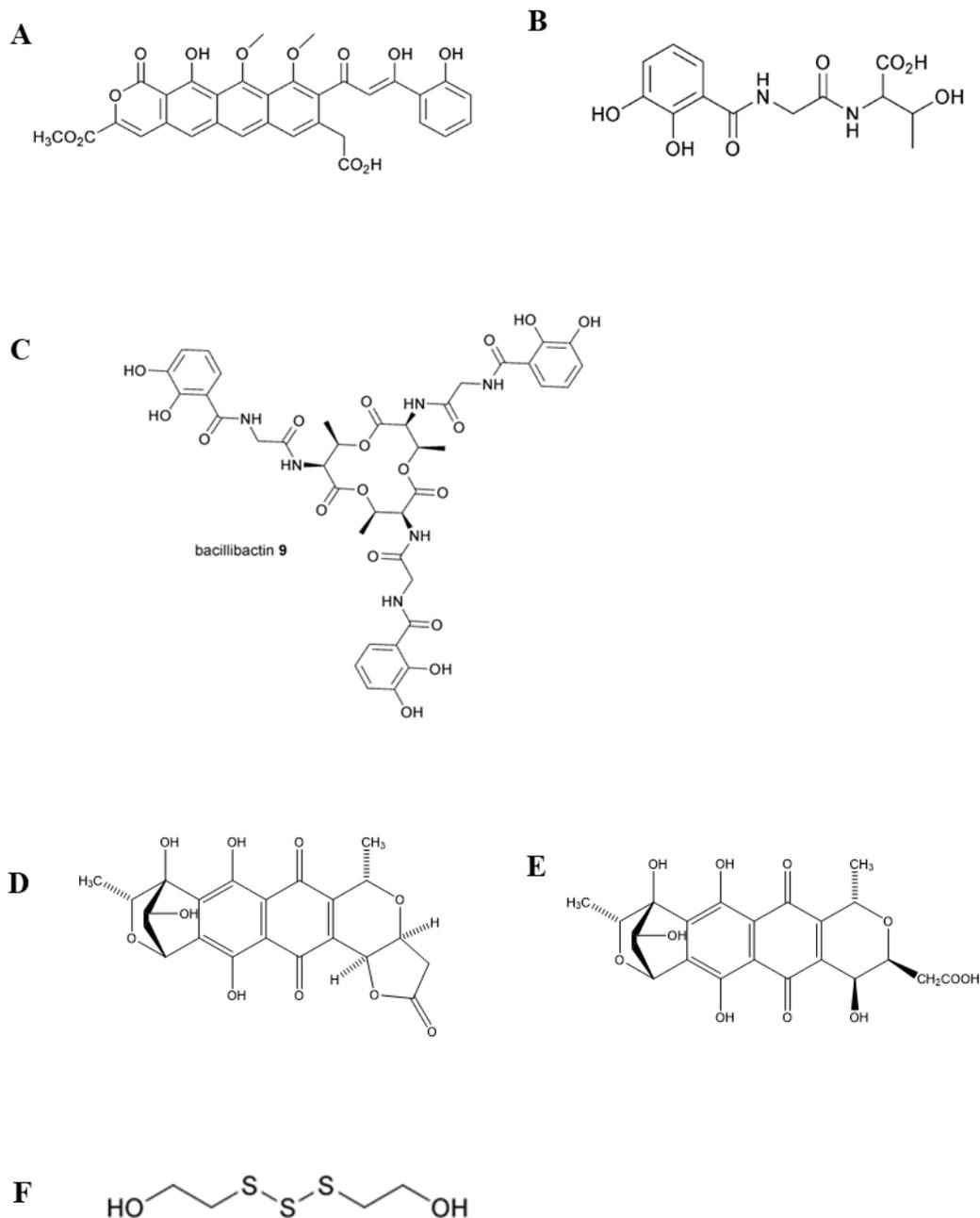
A diversity in extreme environments with an increasing number of extremophiles have been discovered and described. The first extreme environments largely investigated were those characterized by high temperatures above 30°C. These microorganisms can be subdivided based on their optimal temperature, into three main groups: thermophiles that dominate heated environments with an optimal temperature between

30 °C and 64°C and a maximum at 70°C, extreme thermophiles with an optimal temperature between 65°C and 80°C, and finally hyperthermophiles with an optimal temperature above 80°C and a maximum above 90°C (Canganella and Wiegel, 2014, Urbietta et al., 2015, Baker et al., 2001).

### **1.3.2 Secondary metabolites isolated from thermophilic bacteria**

Marine microorganisms have exclusive characteristics due to their need to adapt to extreme environments and that have received them increasing attention to produce biologically active compounds. Extremely important oceanic resources, like marine microorganisms such as bacteria, fungi, actinobacteria and cyanobacteria, produce in their adaption to extreme environmental conditions, structurally novel and pharmacologically active lead compounds (Zhao et al., 2015). These microorganisms act as new and promising potential sources of a vast number of bioactive compounds for functional foods and pharmaceutical development. Important bioactive compounds have been isolated showing activities such as antimicrobial, antitumour, antioxidative, anti-inflammatory, antiviral, antituberculosis, antiparasitic, anthelmintic, antimalarial, antidiabetic, antiprotozoal, anticoagulant and antiplatelet effects (Zhao et al., 2015). Thermophilic bacteria are a prolific source of a variety of secondary metabolites (Figure 1.5), for instance, the isolation of thermorubin (Figure 1.5A), an oxanaphthacene, from a mildly thermophilic actinomycete was first reported in 1964 (Wilson and Brimble, 2009). The compound shown strong bioactivity against Gram-positive bacteria, less against Gram-negative bacteria and slight to no activity against yeast and fungi. Likewise, a siderophore SVK21 (2,3-dihydroxybenzoylglycyl-threonine) (Figure 1.5B) isolated from the thermoresistant bacterium *Bacillus licheniformis* and bacillibactin (Figure 1.5C) produced by *Bacillus subtilis*, both facilitates microbial iron intake (Wilson and Brimble, 2009, Reitz et al., 2017). Another case involved a broad-spectrum antibiotic granaticin (Figure 1.5D) and their derivative—granaticinic acid (Figure 1.5E) isolated from the thermophilic *Streptomyces sp.* XT-11989 which demonstrated strong antibacterial activity against Gram-positive bacteria (Sung et al., 2017). Similarly, the BS-1 (bis(2-hydroxyethyl)

trisulfide) (Figure 1.5F) compound was isolated from *Bacillus stearotherophilus* UK563 in 1991 and was described to have potent cytotoxic effects on mice cell lines mimicking leukaemia, mastocytoma and lymphoma (Wilson and Brimble, 2009, Kohama et al., 1992).



**Figure 1-5 Molecular structure of secondary metabolites isolated from thermophilic bacteria: A) Thermorubin B) SVK21 C) bacillibactin D) granaticin E) granaticinic acid and F) BS-1. Adapted from (Wilson and Brimble, 2009).**

### **1.3.3 *Flavobacteriaceae* family**

The genus *Flavobacterium* was established by Frankland in 1889, belonging to the family *Flavobacteriaceae* in the phylum *Bacteroidetes*, which currently account for more than 50 other genera. These species are Gram-negative, non-spore-forming, strictly aerobic, motile by gliding, pigmented bacteria exhibit menaquinone 6 as their only or major respiratory quinone. They have DNA G þ C contents within the range of 30–41 mol% (Bernardet and Nakagawa, 2006, Zamora et al., 2013).

Members of the *Flavobacteriaceae* family include a group of physiologically diverse and widely distributed environmental microorganisms. This family has been isolated from many different habitats: soil, freshwater and marine environments, food and dairy products and their processing environments, specimens from humans, hospital equipment and devices, diseased dogs and cats, diseased amphibians and reptiles, diseased freshwater and marine fishes, diseased molluscs, crustaceans, and sea urchins, diseased birds, eggs and digestive tract of insects, vacuoles or cytoplasm of amoebae and diseased plants (Bernardet and Nakagawa, 2006, Wańkiewicz and Irzykowska, 2014, Zamora et al., 2013).

Reasonably, applications of members of the family *Flavobacteriaceae* are related to their habitats and to their relationship to the host. Positive aspects include synthesis of several enzymes potentially useful in industry or medicine, turnover of organic matter in soil, water, and sewage plants, decomposition of pesticides and insecticides, destruction of toxic proliferative algae, and symbiosis with various insects. Negative effects include decomposition and defects of food (meat, fish, and dairy products), infections in humans and animals and destruction of valuable algae and vegetables (Bernardet and Nakagawa, 2006).

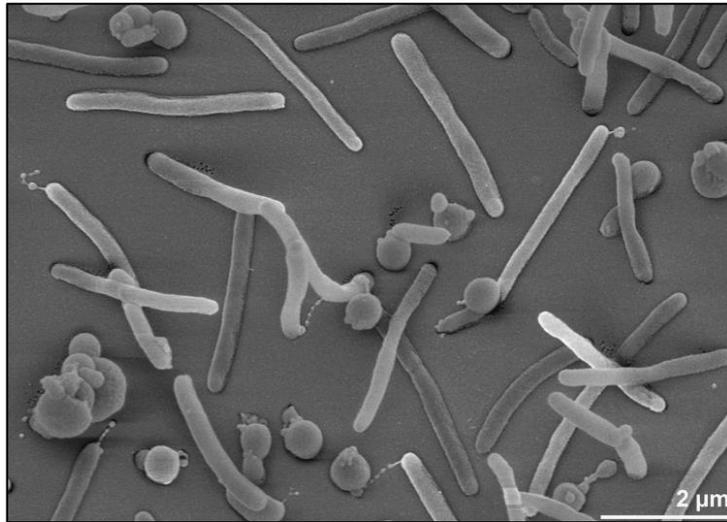
#### **1.3.3.1 *Muricauda ruestringensis***

Although there are no several studies and papers regarding *Muricauda ruestringensis*, it is a bacterium of the *Flavobacteriaceae* family in the phylum *Bacteroidetes*. Preliminary studies of the strain during the EU-FP7 SeaBioTech programme showed inhibition bioactivity against tyrosine-protein phosphatase non-receptor type 1 also

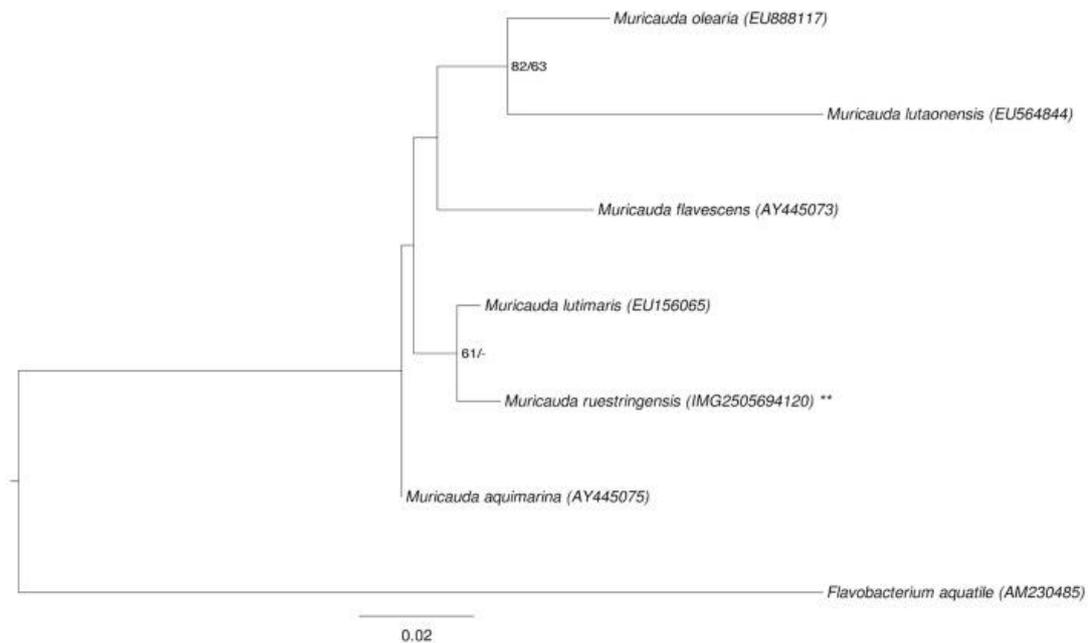
known as protein-tyrosine phosphatase 1B (PTP1B) and at some extent against alpha glucosidase, which makes it extremely interesting in terms of looking into potential new drugs against metabolic diseases such as diabetes type II and obesity that currently represent a major public health issue.

*Muricauda ruestringensis* (Figure 1.6) is a Gram-negative, facultatively anaerobic, and slightly halophilic, with optimum pH for growth 6.5 – 7.5. Seawater is required for growth with mostly polar - located appendages, which seem to be used to connect cells to each other, though not all the cells produced such appendages. It is not able to degrade hexadecane, but able to use different sugars and amino acids for growth, indicating that it probably profits from the lysis or from products like surfactants of other cells in the community (Bruns et al., 2001).

*Muricauda ruestringensis* is a type species of the genus *Muricauda* and the species is of interest because of its isolated position in the genomically unexplored genus *Muricauda*, located in a position of the tree of life (Figure 1.7) containing not many organisms with sequenced genomes. Strain BIT was isolated from a seawater sediment suspension from intertidal sediment at the German North Sea coast, which contained hexadecane as the sole carbon source during the initial cultivation. Later, the organism either turned out to be unable to degrade hexadecane or lost its ability to do so. The genus name was derived from the Latin words muris, for mouse, and cauda, the tail; *Muricauda*, tail of the mouse, referring to the cellular appendages observed on some cells. The species epithet is derived from the Neo-Latin word ruestringensis, pertaining to the former village of Rüstringen, which was destroyed by a tidal wave in 1362 (Huntemann et al., 2012).



**Figure 1-6** Scanning electron micrograph of *Muricauda ruestringensis* B1. Adapted from (Huntemann et al., 2012).



**Figure 1-7** Phylogenetic tree highlighting the position of *Muricauda ruestringensis* relative to the type strains of the other species within the genus *Muricauda*. Adapted from (Huntemann et al., 2012).

All members of the genus *Muricauda* were isolated from marine environments such as tidal flats, intertidal pools, a salt lakes and seawater. Sea tidal flats are broad, low-gradient muddy coastal areas that experience exposure and flooding by seawater

between low and high tides. It is well known that sea-tidal flats are rich in valuable biological resources that play very important roles in the restoration and stabilization of coastal ecosystems and nutrient cycling (Kim et al., 2013).

### **1.3.4 Micromonosporaceae family**

The family Micromonosporaceae currently contains 35 genera and includes Gram-positive microorganisms characterized by three types of sporulating structures, specifically, single spores, spore chains, and sporangia which are borne directly on the substrate hyphae. Spores may be non-motile or motile with tufts of polar flagella. They are aerobic, non-acid fast and mesophilic microorganisms and form non-fragmenting, branched, and septate substrate hyphae; aerial mycelium is absent or scanty. Many strains produce carotenoid mycelial pigments, giving the colonies an orange to red appearance; however blue-green, brown, or purple pigments are also produced (Hirsch and Valdés, 2010, Trujillo et al., 2014).

Micromonosporaceae strains degrade chitin, cellulose, lignin, and pectin, and these microorganisms play an important role in the turnover of organic plant material. Furthermore, many strains produce useful secondary metabolites and enzymes. They have important applications in industry, biotechnology, and agriculture. Members of the family Micromonosporaceae are responsible for the production of gentamicin, vitamin B 12 and alpha-glucosidase inhibitor acarbose, a potent drug used worldwide in the treatment of type-2 diabetes mellitus (McChesney et al., 2007, Kumar et al., 2008, Trujillo et al., 2014).

#### **1.3.4.1 *Micromonospora* sp.**

One of the genera of the family Micromonosporaceae is *Micromonospora*, proposed by Ørskov (1923), which typically exhibit single, non-motile spores on the mycelium substrate and the absence of true aerial mycelium. The spore surface is smooth or ridged. The colonies can be a variety of colours, including white, orange, rose, or brown while the pigments may or may not diffuse into the medium (Zhang et al.,

2012). Members of this genus are Gram-positive, non-acid fast, chemo-organotrophic and usually grow between 20° C and 45° C (Trujillo et al., 2014).

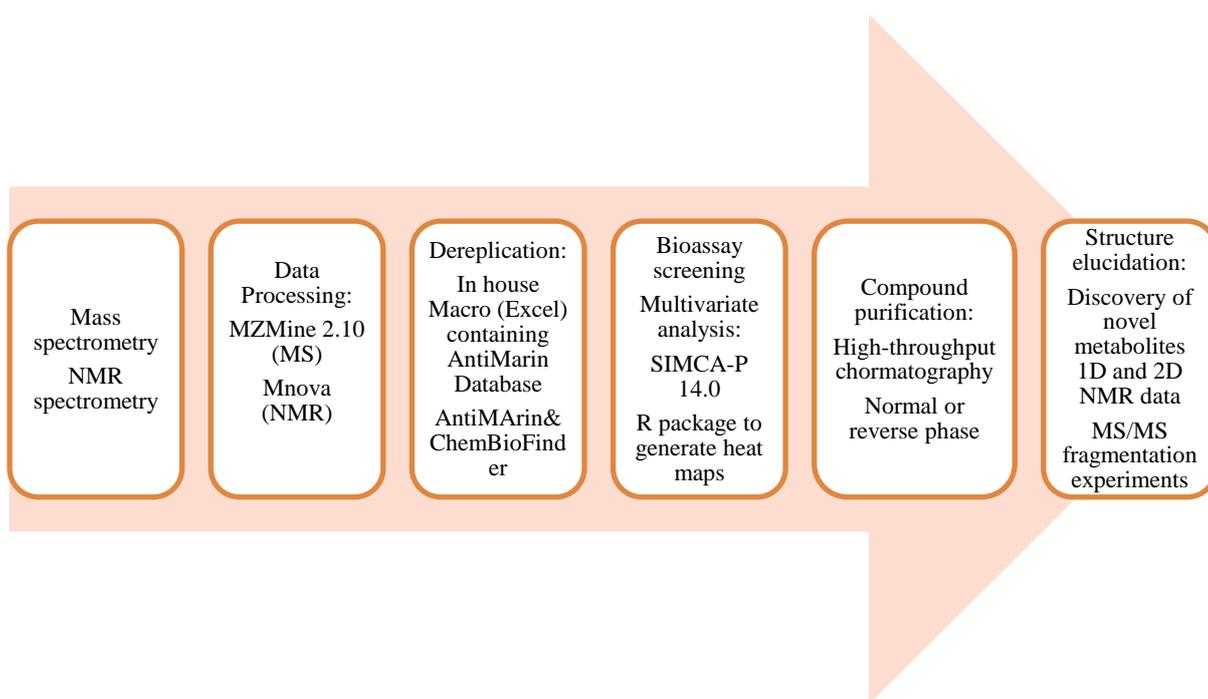
Since 2000, several novel *Micromonospora* species have been described in the last decade and an increasing number of members of the genus have been isolated. Currently, the genus *Micromonospora* contain 81 species with validly published names and is the second richest producer of antibiotics in the order Actinomycetales (Hirsch and Valdés, 2010, Songsumanus et al., 2013, Zhang et al., 2012, Riesco et al., 2018). Usually *Micromonospora* are isolated from different habitats in nature such as peat swamp forest soil, water samples, floodplain meadows, root nodules, marine sediment, plant tissues, sea sand, marine sponges and other environments (Hirsch and Valdés, 2010, Supong et al., 2013, Zhao et al., 2004).

The bacteria, particularly those from the order actinomycetales, commonly known as actinomycetes have proven to be a productive source of new and bioactive compounds. Among the various genera within the actinomycetales, *Micromonospora* has been intensively investigated. *Micromonospora* species are best known for synthesizing antibiotics, especially aminoglycoside, enediyne, and oligosaccharide antibiotics, such as anthraquinones, anthracyclines, alkaloids, and macrolides. The secondary metabolites produced by actinomycetes demonstrate a variety of biological activities including antibacterial, antifungal, antiprotozoal, anthelmintic, antiviral, insecticidal, cytotoxic, antioxidant, and anti-inflammatory (Hirsch and Valdés, 2010, Sousa et al., 2012).

## 1.4 Metabolomics technologies and platforms

The metabolome is a genome final product, which is defined as the total quantitative group of small molecular weight compounds (metabolites) present in a cell or organism that is involved in metabolic reactions (Dunn et al., 2005). Metabolites are small molecules that are chemically transformed during metabolism, providing functional information of cellular state, which serve as direct signatures of a biochemical activity therefore, they are easy to correlate with a phenotype when compared to genes and proteins whose function is subject to epigenetic regulation and

post-translational modifications, respectively (Patti et al., 2012). Metabolomics is part of the omics strategies (genomics, proteomics and transcriptomics) that aims at describing the metabolome qualitatively and quantitatively by applying various analytical platforms and methods (shown in Figure 1.8) (McMurray et al., 1989). Complementary to the other omics approaches, metabolomics has the essential role of linking genotype-phenotype information (Lopes et al., 2017).



**Figure 1-8 Metabolomics workflow** process used in our Research Group, Natural Products Metabolomics Group (NPMG) Adapted from: (Macintyre et al., 2014, Tawfike et al., 2013).

Metabolomics combines analytical chemistry, platform technology, mass spectrometry and NMR spectroscopy with data analysis (Griffiths et al., 2010). Whereas genomics and proteomics are usually studied using microarrays and 2D-gels or mass spectrometry respectively. However, the technique of choice is less evident for metabolomics (Dunn et al., 2005). Nevertheless, MS and NMR usually are the most frequently used methods of detection in analysis of the metabolome (Villas-Boas et al., 2005) and the data generated through this technique is further processed using different software either open-source or commercially available, such as MZmine (Katajamaa and Oresic, 2005), Mnova, MetAlign (Lommen, 2009), MathDAMP

(Baran et al., 2006), MS-DIAL (Tsugawa et al., 2015) and XCMS (Smith et al., 2006). Respective software can be coupled with online or commercially available libraries and databases like the Dictionary of Natural Products (DNP), ChemSpider (Pence and Williams, 2010), MarinLit (Blunt and Munro, 2012) or in-house / custom database to dereplicate secondary metabolites (Harvey et al., 2015). Then the processed data is further subjected to multivariate statistical analysis applying either unsupervised clustering such as partial component analysis (PCA) or supervised clustering such as orthogonal partial least squares discriminant analysis (OPLS-DA) to provide information on the putative bioactive metabolite at the first fractionation step or detect putative biomarkers in a cellular process (Nejadgholi and Bolic, 2015, Robotti and Marengo, 2016, Morlock et al., 2014).

### **1.4.1 Mass spectrometry (MS)**

Mass spectrometry (MS) is the analytical technique of choice in metabolomics for identification and/or quantification of varied classes of metabolites, consisting in the production of gas-phase ions that are then detected and characterised by their mass to charge ratio ( $m/z$ ) (Becker et al., 2012). In a simple way, a mass spectrometer contains a sample inlet, an ion source, a mass analyser and a detector with the functions of introducing the sample into the mass spectrometer, generate gas-phase ions *via* an ionisation technique, separate the ions according to their mass-to-charge ratio ( $m/z$ ) then generating an electric current from the incident ions that is proportional to their abundances (Becker et al., 2012). Gas chromatography (GC), high performance liquid chromatography (HPLC), capillary electrophoresis (CE) allow improved metabolite identification and quantification. The use of MS in metabolomics has important advantages such as requiring small volume of samples and provides highly sensitive detection and metabolite identification *via* mass spectrum interpretation and comparison or molecular formula prediction *via* high resolution mass measurements (Lindon et al., 2000). Moreover, the combination of separation techniques with MS increases the capability of the chemical analysis of highly complex biological samples (Villas-Boas et al., 2005). Nevertheless, MS is also destructive, so the sample is not

recoverable and is a slow method, unlike NMR spectroscopy (Lindon et al., 2000, Wishart, 2016).

### **1.4.2 Nuclear magnetic resonance (NMR)**

Nuclear magnetic resonance (NMR) spectroscopy is a widely used technique for metabolomics studies with many benefits such as being specific and at the same time non-selective, non-destructive, requires no separation or derivatization, fast, offering highly reproducible and quantitative analyses (Dunn et al., 2005, Viant et al., 2003, Wishart, 2016). The NMR spectrum is specific and unique to each compound, and provides a valuable structural information about the components of the analysed sample by combining the information on chemical shift (nature of chemical environment), signal multiplicities (neighbouring signals), homonuclear and heteronuclear coupling constants, integral of the signal (number of protons), spin-spin coupling (number and nature of neighbours and connectivity information) and relaxation or diffusion (size of molecule and large-scale environment of location) (Dunn et al., 2005, Dona et al., 2016). Although the most utilised NMR experiment is the one-dimensional proton (1D- $^1\text{H}$ ) NMR, currently there are available alternative NMR techniques that offer additional chemical and structural information, when in some situations that  $^1\text{H}$  NMR is insufficient to provide information to entirely characterize the metabolite (Villas-Boas et al., 2005, Viant et al., 2003). Additionally, the analysis is non-destructive, and does not require pre-selection of the analysis conditions like choosing the ion source or chromatographic operating conditions (stationary phase, mobile phase, temperature) in LCMS (Dunn et al., 2005).

### **1.4.3 Multivariate approaches and manual-based metabolite structure elucidation**

An extensive variety of statistical methods are currently available, ranging from unsupervised methods, such as, PCA, or hierarchical clustering (HCA), to supervised methodologies like partial least squares (PLS), partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-

DA) (Lindgren et al., 1996, el-Deredy, 1997). Processed MS and NMR data results to a matrix of signal intensities, which can be employed for statistical analytical tools (Spicer et al., 2017). The first step in metabolomics data analysis is using PCA as an initial exploratory and visualization method, that gives an overview of the variability of the dataset as the samples are grouped based on similarity or differences within the group of samples. This enables to detect trends, groupings and outliers (Wold et al., 1987, Dona et al., 2016) through scores and loadings plot. In the scores plot, each point represents an individual sample, while a loadings plot gives the corresponding information about which variables have the greatest contribution to the positioning of the samples on the scores plot and are responsible for the clustering of samples (Holmes et al., 2000). PCA analysis is followed by supervised pattern recognition techniques, which applies the class information of the samples to maximise the separation between different groups of samples and detect the metabolic signatures that contribute to the classifications (Dona et al., 2016). OPLS-DA is the most used supervised methodology, having the same predictive power as PLS but gives better interpretation of the relevant variables than PLS. This methodology provides information about the causes for class separation (Bylesjö et al., 2006)

The manual-based structure elucidation is a complex, challenging and time-consuming process that was applied during this project. For this effect, it combines the information obtained from 1D and 2D NMR experiments with MS data and information from metabolite databases such as DNP and AntiMarin to achieve accurate metabolite identification (Dona et al., 2016). Most of the metabolites observed by NMR spectroscopy and mass spectrometry were identified with information available in diverse databases, however, some of the metabolites are unknown, unidentified or truly novel, which brings difficulties to the elucidation method (Dona et al., 2016). Nevertheless, it is the unknown metabolites that are the target of interest for further isolation work.

## 1.5 Mining genomes for natural products discovery

More than 20 years have passed since the first microbial genome of the *Haemophilus influenzae* (Pfeiffer's bacillus) was fully sequenced (Fleischmann et al., 1995); this event boosted both biological knowledge at the time by characterising the genomic landscape of unknown microbial organisms as well as provided an overall improvement of the DNA sequencing technical aspects and standard operator protocols, which ultimately opened the avenue for the application of shotgun approaches in whole-genome sequencing across many relevant bacterial species and strains (Donkor, 2013, Wang et al., 2016a, Kwong et al., 2015).

The era of technological genomic revolution is a direct result of a remarkable innovation of nucleotide sequencing methodologies that have drastically altered the cost and complexity associated with DNA sequencing (Heather and Chain, 2016). The abilities of DNA sequencers are outperforming current advances in computing technologies as designated by Moore's law (Figure 1.9). This law specifies that the complexity of microchips evaluated by the number of transistors in each square inch per unit cost doubles every two years, while sequencing capabilities are following a roughly exponential scaling behaviour, which is doubling every five months (Stein, 2010, Muir et al., 2016, November, 2018).

Figure removed due to copyright restrictions

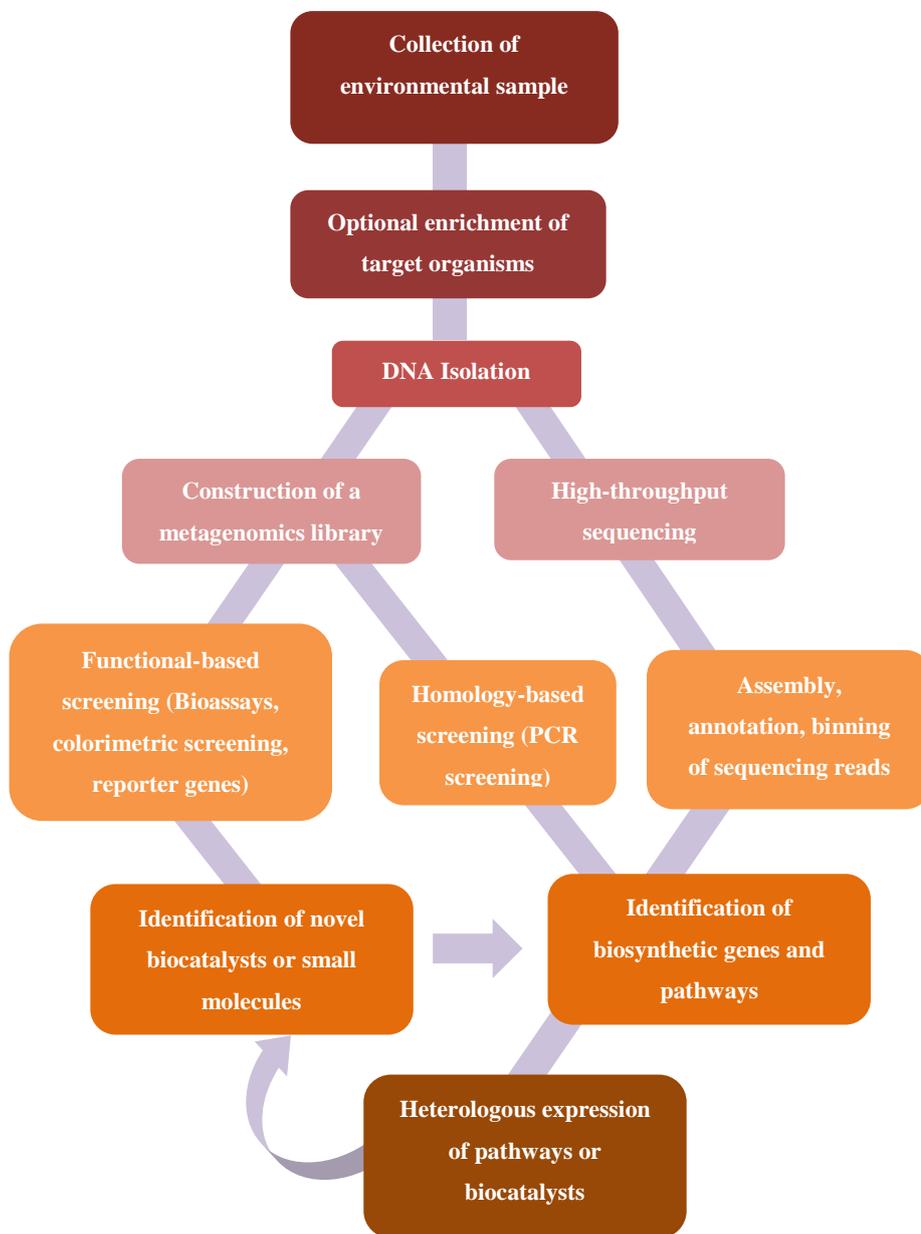
**Figure 1-9 DNA sequencing costs along time.** The cost-accounting data presented here are summarized relative to two metrics and represented in USD dollars: (a) "Cost per Megabase of DNA Sequence" - the cost of determining one megabase (Mb; a million bases) of DNA sequence; (b) "Cost per Genome" - the cost of sequencing a human-sized genome Adapted from: (Wetterstrand, 2018).

The noticeable decrease in the price of sequencing has made bacterial genome sequencing affordable to an increasingly number of laboratories, leading to a democratization of sequencing competences (Shendure and Ji, 2008). The reducing cost of sequencing and the growing number of sequence reads being produced has

yielded a cost modification from sequencing to bioinformatics analysis pipelines, together with quality control (QC), assembly, annotation, and data warehousing (Muir et al., 2016). This fostered the number of sequenced genomes that has dramatically increased over the last decade—there are more than 163344 sequenced microbial genomes currently publicly available in September 2018 (O'Leary et al., 2016), including 27 sequenced *Muricauda* genomes, 92 sequenced *Micromonospora* genomes and houses thousands of metagenome projects (Mukherjee et al., 2017).

The exponential development of available genetic data for microorganisms enabled a promising new period in the discovery of novel NPs (Figure 1.10). Nowadays, it is possible to quickly identify gene clusters encoding bioactive NPs as well to make computer predictions of chemical structure based on the sequence information. The advances in DNA sequencing and bioinformatics technologies are acknowledgeable, especially the use of selective genome scanning tools instead of sequencing the whole microbial genome (Lam, 2007). Presently, it is possible to use genomic information to generate structure predictions that can be used to identify new molecules, by removing the probabilities of re-isolating known compounds (Bode et al., 2002). The structure predictions provide valuable information to guide and help compound purification, structure confirmation, and design ideal fermentation conditions to produce the required NPs, by knowing the biosynthetic gene encoding specific NPs that could only be expressed in certain fermentation conditions. Research groups are applying the integration of genomic approaches and technologies to increase the efficiency of NPs discovery with the decreasing of the discovery process time (Bode et al., 2002, Lam, 2007). Another main frontier in drug discovery is the application of metagenomics (Figure 1-10) and genomic tools which enable the identification of bioactive compounds in uncultivable species that currently constitute most known and unknown microorganisms at sea. Genetic engineering became a great alternative to avoid difficulties of supply by obtaining recombinant strains capable of higher amount of products of interest and may be able to yield novel analogues with characteristics greater to those of the related natural product (Eva-Maria Unsin et al., 2013, Romano et al., 2017). Notwithstanding developments in culturing techniques, the metagenomics will obtain additional interest. The mining for novel marine natural

products and their exploitation for biotechnological applications will continued to be improved by the fast evolving technologies and the higher information of molecular processes and mechanisms responsible for secondary metabolite production (Wilson and Piel, 2013, Romano et al., 2017).



**Figure 1-10 Metagenomics workflow.** Drug discovery pipeline based on a metagenomics approach. Adapted from: (Wilson and Piel, 2013, Romano et al., 2017).

### 1.5.1 Silent (cryptic) biosynthetic gene clusters (BGC)

A biosynthetic gene cluster (BGC) can be defined as a co-localized clustered group of two or more genes in a specific genome which encode a biosynthetic pathway for the production of a particular metabolite, that can comprehend its chemical variants, being highly informative of the encoding class of compounds (Medema et al., 2015). The genome sequencing technological revolution allowed the faster, easier and inexpensive way of acquiring the complete bacterial genome, followed by the development of computational tools to gather and analyse genome sequences. One of the fundamental approaches in microbial (especially bacterium) natural product discovery is genome mining, defined as an alternative approach to more traditional methods for discovery of novel secondary metabolites by information extraction from genome sequences (Trivella and de Felicio, 2018). Recently, there has been a combined effort of the scientific community to report the hold-ups in the discovery of new bioactive natural products (NP). Genome sequencing and mining, genetic or chemical manipulation of microbial growth, and mass spectrometry (MS)-based metabolomics are the foremost developing areas in this field and comprise the tripod for contemporary natural product discovery (Trivella and de Felicio, 2018). Undeniably, furthestmost sequenced microorganisms and plants with relatively large genomes, comprehend dozens or more blueprints for the biosynthesis of secondary metabolites (Blin et al., 2013). Furthermore, computerised bioinformatics platforms nowadays simplify the semi-automated prediction of natural products by encoding secondary metabolic blueprints (Boddy, 2014). Nevertheless, the identification of genome-encoded secondary metabolism is one of the first stage in the genome mining process. Definitely, genome mining widens the total spectrum of the updated essential dogma of molecular biology with bioinformatics prediction of gene and pathway role, the control of gene expression and translation, and the identification and structural elucidation of new metabolites from within the metabolome of the producing organisms (Bachmann et al., 2014). Therefore, genome mining studies frequently become more than uniquely a natural product discovery programme as they involve a complete understanding and management of cellular molecular systems (Figure 1.11). The genome mining importance ranges well beyond its potential to completely avoid

the “chance” element of the discovery of secondary metabolites method. For example, understanding the linking between metabolites, that characterise one of the end points of the essential dogma, and the gene sequences that encode the metabolites can afford awareness into the elementary biology of producing organisms as discrete individuals and as members of the microbiota of their environment. Additionally, numerous if not most of secondary metabolites have roles in interspecies, intergeneric and/or inter-kingdom chemical ecological associations (Bachmann et al., 2014).

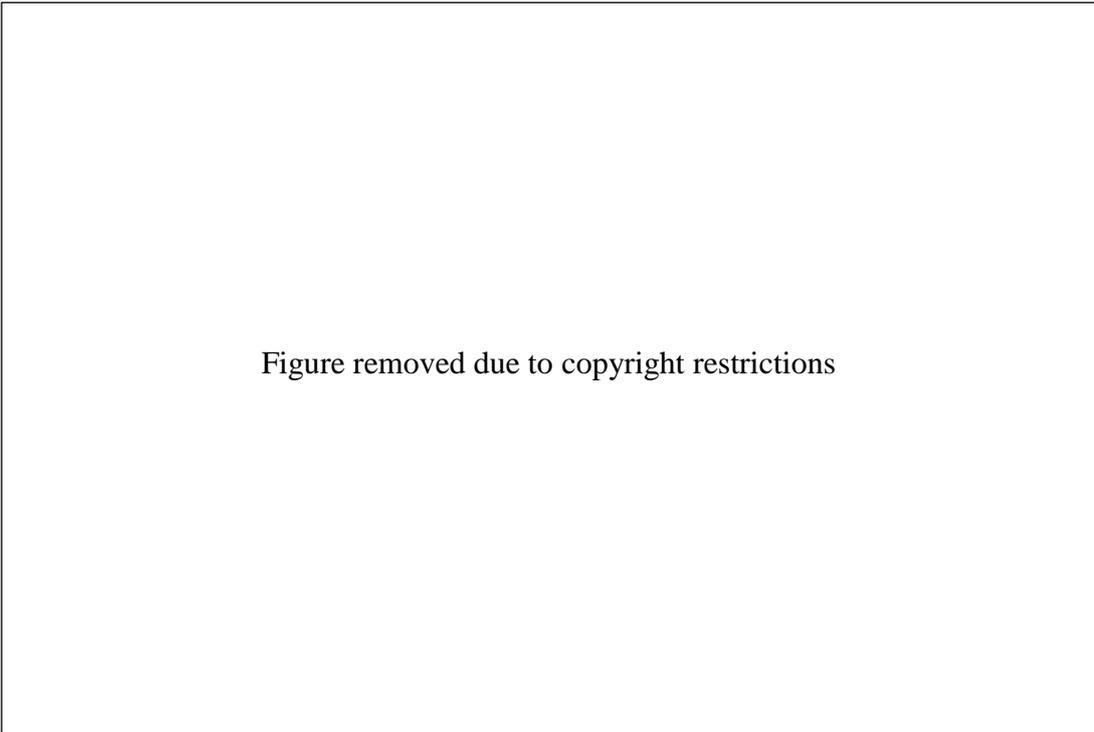


Figure removed due to copyright restrictions

**Figure 1-11 Methods that can be used to activate silent genes in microorganisms.** From: (Walsh and Tang, 2017). **PTF**: promoter transcription factor, **TF**: transcription factor, **PBPG**: promoter biosynthetic pathway, **BPG1**: biosynthetic pathway gene 1, **BPG2**: biosynthetic pathway gene 2, **BPG3**: biosynthetic pathway gene 3, **Pcons**: constitutive promoter.

## 1.5.2 Genome sequencing platforms and *in silico* approaches

The advances in DNA sequencing, particularly in Next Generation Sequencing (NGS), have significantly improved the quantity and quality of genomic information that can be obtained from microbial samples (Zhang et al., 2011). The reduced cost of NGS as well as the increase in high-throughput whole genome sequencing (WGS), as well as other NGS applications such as whole exome sequencing (WES) or RNA-Seq, a possible and reliable approach for studying bacterial transcriptomes has been progressively being developed. However, there are still some challenges such, as contaminations, data storage, management, analysis, and interpretation, which have to be considered for the proper use of this technology in NPs research (Su et al., 2011).

Currently millions of putative BGCs have been identified, which have potential to produce a vast number of novel natural products. A typical workflow for a modern NPs discovery strategy begins with identification of silent BGCs using bioinformatics tools (Figure 1.12), followed by activation of target BGCs using synthetic biology strategies and structural characterization of the corresponding natural products. Based on the implementation strategies, BGCs repositories (Table 1-2), computational tools (Table 1-3) for identification of BGCs mainly fall into two groups. The first group is rule-based tools, such as ClusterScan (Volpe et al., 2018), NP.searcher (Wang et al., 2009), NaPDoS (Ziemert et al., 2012), SEARCHPKS (Yadav et al., 2003), and SMURF (Khaldi et al., 2010) for mining polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS), BAGEL (de Jong et al., 2006) for RiPPs, and antiSMASH (Medema et al., 2011). These tools use predefined rulesets to identify signature genes that are associated with biosynthetic pathways annotated in literature or reference databases resources (Table 1.2). As a result, they can predict known classes of BGCs with high confidence and positive rates but cannot readily identify BGCs that use completely different enzymatic mechanisms. Notable among them is antiSMASH, which in its 3rd version can detect 44 different classes of BGCs, and upon integration with ClusterFinder (Cimermanic et al., 2014), which uses a rule-independent algorithm could further provide alternative options for automatic genome-scale

predictions of BGCs of both known and novel classes (Ren et al., 2017, Weber et al., 2015, Rutledge and Challis, 2015).



**Figure 1-12 Overview of the most commonly used and freely accessible tools specialized for the analysis of secondary metabolites and their pathways.** From: (Weber and Kim, 2016).

**Table 1-2 Overview of databases for the analysis of secondary metabolites and their biosynthetic gene clusters.**

<b>General scope compound databases</b>	<b>Web Server URL</b>	<b>Reference</b>	<b>No. of molecules</b>	<b>Bioactivities</b>	<b>Free use</b>
ChEBI	www.ebi.ac.uk/chebi	(Hastings et al., 2013)			
ChEMBL	www.ebi.ac.uk/chembl	(Bento et al., 2014)			
PubChem	pubchem.ncbi.nlm.nih.gov	(Wang et al., 2009)			
ChemSpider	chemspider.com	(Pence and Williams, 2010)			
<b>NPs databases</b>					
NORINE	bioinfo.lifl.fr/norine	(Caboche et al., 2008)			
StreptomeDB	www.pharmaceutical-bioinformatics.de/streptom-edb	(Lucas et al., 2013)			
Dictionary of Natural Products	dnp.chemnetbase.com	(Buckingham, 1997)	>230k (>153k)	yes	no
KNAPSAcK	kanaya.naist.jp/KNAPSAcK	(Nakamura et al., 2014)			
Super Natural II	bioinformatics.charite.de/supernatural	(Banerjee et al., 2015)	>325k	yes	yes
Reaxys	www.reaxys.com		>220k	yes	no
UNPD	pkuxxj.pku.edu.cn/UNPD	(Gu et al., 2013)	>229k (>167k)	no	yes
TCM database@Taiwan	tcm.cmu.edu.tw	(Chen, 2011)	>60k (~50k)	yes	yes
TCMID	www.megabionet.org/tcmid		>13k (>11k)	yes	yes

Chem-TCM	www.chemtcm.com		>12k	yes	no
HIT	lifecenter.sgst.cn/hit		>700 (>400)	yes	yes
HIM	binfo.shmtu.edu.cn:8080/him		~1300 (~700)	yes (focus on toxicity data)	yes
AfroDb	african-compounds.org/about/afrodb		~1000 (~900)	yes	yes
SANCDDB	sancdb.rubi.ru.ac.za		>600 (~600)	no	yes
NANPDB	www.african-compounds.org/nanpdb		>4400 (~3900)	yes	yes
NPACT	crdd.osdd.net/raghava/npact		~1500 (~1400)	yes (focus on anticancer activity)	yes
NPCARE	silver.sejong.ac.kr/npcare		>6500 from online search; >1500 in bulk download (>1500)	yes (focus on anticancer activity)	yes
TIPdb	cwtung.kmu.edu.tw/tipdb		~9000 (~8000)	yes (focus on anticancer, antiplatelet and antituberculosis activity)	yes
StreptomeDB	www.pharmaceutical-bioinformatics.de/streptomedb		~4000 (~3600)	yes	yes
NuBBE database	nubbe.iq.unesp.br/portal/nubbedb.html		>1800, including >1700 plant NPs and >100 microorganism NPs (~1700)	yes (focus on antimicrobial activity)	yes
Carotenoids Database	www.carotenoiddb.jp		>1100	yes	yes
AntiBase	wwwuser.gwdg.de/~hlaatsc/antibase.htm		>40k	yes	no
DMNP	dmnp.chemnetbase.com		>55k (including NP derivatives)	yes	no
MarinLit	pubs.rsc.org/marinlit	(Blunt and Munro, 2012)	>29k	yes	no

*In silico* methods for connecting genomic to chemical have played a vital role in genomics discovery of new natural products. Here, in this introductory section, an overview of available computational tools has been presented (Table 1.3) then to briefly describe a novel computational framework, specifically retro-biosynthetic enumeration of biosynthetic reactions, which can add to the list of computational tools available for linking natural products to their BCG. The forward “Genes to Metabolites” approach is applied by mostly of the existing bioinformatics tools for analysis of secondary metabolite BCG. In contrast to this approach, the reverse/retro-biosynthetic approach, “Metabolites to Genes” would include tallying the several biochemical transformations or enzymatic reactions which would generate the given chemical moiety starting from a set of precursor molecules and identifying enzymatic domains that can hypothetically catalyse the numbered biochemical transformations. The linkage of the biosynthetic genes to secondary metabolites and vice versa can hypothetically support not exclusively in description of new secondary metabolites, as well in redesigning known biosynthetic pathways of secondary metabolites to yield new compounds (Khater et al., 2016, Medema and Fischbach, 2015, Walsh and Fischbach, 2010). The two approaches that can be applied to accomplish the redesigning biosynthetic pathways are the one mentioned before the Forward (Genes to Metabolites) and Reverse/Retro-biosynthetic (Metabolites to Genes) with both created on understanding the development of sequence/structural features of individual domains of biosynthetic pathways (Bachmann, 2010, Cacho et al., 2014). There are multiple benefits from such approaches, not only the understanding of the molecular and enzymatic basis of secondary metabolites biosynthesis (Cacho et al., 2014). For instance, it is possible to create more efficient total synthesis paths for complex secondary metabolites from natural products based on the novel chemical insights attained from the metabolic pathway elucidation. Correspondingly, the information about the BGC can enable the increasing the yield of advantageous secondary metabolites using metabolic engineering (Pickens et al., 2011).

**Table 1-3 Overview of computational tools for the analysis of secondary metabolites and their biosynthetic gene clusters.** Nonribosomal peptides (NRPS), polyketide synthase genes (PKS), structure-based sequence analysis of polyketide synthases (SBSPKS).

Tool/Database	Web Server URL	Reference
<b>BGS's identification and analysis</b>		
antiSMASH	antismash.secondarymetabolites.org	(Medema et al., 2014, Blin et al., 2013)
ClusterFinder	https://github.com/petercim/ClusterFinder	(Cimermancic et al., 2014)
NP.searcher	dna.sherman.lsi.umich.edu	(Wang et al., 2009)
SMURF	jcvi.org/smurf	(Khaldi et al., 2010)
BAGEL	bagel.molgenrug.nl	(de Jong et al., 2006, de Jong et al., 2010, van Heel et al., 2013)
ClustScan	bioserv.pbf.hr/cms	(Starcevic et al., 2008)
NaPDoS	napdos.ucsd.edu	(Ziemert et al., 2012)
eSNaPD	esnapd2.rockefeller.edu	(Reddy et al., 2014)
<b>NRPS-PKS/SBSPKS</b>		
SBSPKS	www.nii.ac.in/sbspks.html	(Anand et al., 2010)
MultiGeneBlast	multigeneblast.sourceforge.net /	(Medema et al., 2013)
<b>Connecting genomic and MS data</b>		
GNPS	gnps.ucsd.edu	(Wang et al., 2016a)
Pep2Path	pep2path.sourceforge.net	(Medema et al., 2014)
RiPPQuest	cyclo.ucsd.edu	(Mohimani et al., 2014a)
NRPQuest	cyclo.ucsd.edu	(Mohimani et al., 2014b)
CycloQuest	cyclo.ucsd.edu	(Mohimani et al., 2011)
<b>Substrate specificity predictions (NRPS/PKS enzymes)</b>		
NRPSPredictor	nrps.informatik.uni-tuebingen.de	(Rausch et al., 2005, Rottig et al., 2011)
LSI Predictor	bioserv7.bioinfo.pbf.hr/LSIpredictor	(Baranasic et al., 2014)
NRPSsp	www.nrpsp.com	(Prieto et al., 2012)
NRPS/PKS substrate predictor	www.cmbi.ru.nl/NRPS-PKS-substrate-predictor	(Khayatt et al., 2013)
<b>Gene cluster databases</b>		
IMG-ABC	https://img.jgi.doe.gov/ABC	(Markowitz et al., 2014)
MIBiG repository	mibig.info	(Medema et al., 2015)
DoBISCUIT	www.bio.nite.go.jp/pks	(Ichikawa et al., 2013)
ClusterMine360	clustermine360.ca	(Conway and Boddy, 2013)
ClustScan DB	csdb.bioserv.pbf.hr/csdb	(Diminic et al., 2013)
MAPSI	gate.smallsoft.co.kr:8008/pks	(Tae et al., 2009)

## 1.6 Aims of the project

The project was designed to identify the bioactive metabolites production in the thermophile strains of *Muricauda ruestringensis* (SBT531 and SBT587) and *Micromonospora sp.* (SBT687 and SBT692). Using metabolomics tools, fermentation and growth parameters were optimised to increase the production yield of the respective bioactive target metabolites. The bioactive metabolites were also targeted using a metabolomics approach.

High resolution mass spectrometry in combination with liquid chromatography (LC-HRMS), nuclear magnetic resonance ( $^1\text{H}$  NMR, CARBON and CARBON-DEPT/JMOD) and 2D-Correlation Spectroscopy (COSY, HMBC, HMQC, HSQC, and NOESY) were also performed. Bacterial extracts were prepared for specific bioassays testing. The spectral data were processed using several software's, MestReNova 10.0, MzMine 2.10, and statistically analyzed SIMCA 15.0.2, then dereplicated by employing an In-house database management Macro by EXCEL. The metabolite search and obtained "hits" from the database were further supported by using ChemBioFinder version 13 while all structures were drawn using ChemDraw 16.0. DNP and AntiMarin database provided details on the putative identities of all metabolites in the bacterial extracts and the number of remaining unknowns for the extracts.

## 1.7 Objectives of the project

The following objectives were accomplished:

- Up-scaling of the bacterial culture to optimise production of the desired bioactive compounds;
- The use of liquid chromatography coupled to a high-resolution Fourier transform mass spectrometry (LC-HRFTMS) and offline NMR to identify and isolate compounds from *Muricauda ruestringensis* (SBT531 and SBT587) and *Micromonospora sp.* (SBT687 and SBT692);
- Dereplication of bacterial extracts from *Muricauda ruestringensis* (SBT531 and SBT587) and *Micromonospora sp.* (SBT687 and SBT692) by using LC-HRFTMS and NMR data;
- The use of bioassays to determine the activity, if any, of compounds isolated from *Muricauda ruestringensis* (SBT531 and SBT587) and *Micromonospora sp.* (SBT687 and SBT692);
- The use of LC-HRFTMS and NMR to analyse extracts and compounds from *Muricauda ruestringensis* (SBT531 and SBT587) and *Micromonospora sp.* (SBT687 and SBT692);
- Isolation of pure compounds from *Muricauda ruestringensis* (SBT531 and SBT587) and *Micromonospora sp.* (SBT687 and SBT692).

## Chapter 2



### MATERIALS AND GENERAL METHODS

## 2.1 General Reagents and solvents

Analytical grade acetone, methanol (MeOH), ethyl acetate (EtOAc), dichloromethane (DCM), acetonitrile (ACN), absolute ethanol, HPLC grade EtOAc and HPLC grade ACN, HPLC grade n-hexane, HPLC grade MeOH, HPLC grade Isopropanol were purchased from VWR chemicals, France. HPLC grade DCM and HPLC grade absolute ethanol were obtained from Fisher Scientific, UK. HPLC grade acetone and formic acid were purchased from Sigma-Aldrich Co., Germany HPLC grade water was obtained directly from the laboratory through a direct Q-3 water purifier system (Millipore, Watford, UK). The solvents used for NMR were dimethyl sulphoxide (DMSO), deuterated DMSO-d<sub>6</sub>, deuterated chloroform (CDCl<sub>3</sub>), deuterated pyridine-d<sub>5</sub>, (*S*)-(+)-methoxy- $\alpha$ -trifluoro-methylphenylacetyl chloride ((*S*)-(+)-MTPA-Cl) and (*R*)-(-)-methoxy- $\alpha$ -trifluoro-methylphenylacetyl chloride ((*R*)-(-)-MTPA-Cl) used for Mosher ester analysis, were purchased from Sigma-Aldrich Co.,USA. Celite<sup>®</sup> S from Sigma-Aldrich Co.,USA was also used.

## 2.2 General equipment

The analytical miller (model: IKA A11 Basic) was purchased from IKA, Germany. The rotary evaporator model number R-110 and R-3 were from BÜCHI, Switzerland. The Ultrawave sonicator was from Scientific Laboratory Supplies, Ltd. The Force 7 centrifuge was from Fisher Scientific. The handheld UV lamp (UVGL-55 UV Lamp) was from UVP, Cambridge, UK. The heat gun HL 2010 E Type 3482 was from Steinel, USA. The Stuart<sup>®</sup> block heater SBH 130D/3 was from Bibby Scientific Ltd., Staffordshire, UK. The freeze dryer, model Christ Alpha 2-4 was from Martin Christ Gefriertrocknungsanlagen GmbH, Germany. The pH meter was purchased from Jenway, UK. The optical rotation was measured on a 341 Polarimeter from PerkinElmer, Inc., USA

## 2.2.1 Microbiology equipment

The laminar flow hood, BioMAT2, was purchased from Medical Air Technology, UK. The stand incubator Incu-160s (27°C) or 160c (14 °C) was used to store inoculated agar plates was from SciQuip Ltd, Shropshire, UK. The 15L double deck incubator and rotatory shaker Incu-Shake FL24-1R and 5L benchtop incubator and rotatory shaker Incu-Shake MAXI were obtained from SciQuip Ltd, Shropshire, UK. The homogenizer IKA T18 Basic Ultra Turrax and handheld homogenizer Ultra Turrax T8 were from IKA Labortechnik Germany.

## 2.3 Bacterial culturing

During the present study, the microorganisms investigated were all collected during the SeaBioTech project which was funded by the European Commission within its FP7 Programme, with Grant Number 311932.

### 2.3.1 *Muricauda ruestringensis* (SBT531 and SBT587)

The two strains of the bacteria *Muricauda ruestringensis* (referred to as SBT 531 and SBT587) were obtained from the intertidal pools in Iceland. Bacterial cultures were optimized in 100mL Difco™ Marine Broth from Becton, Dickinson and Company Sparks, USA in twelve 200mL shake flasks. The chosen media was identical to those used by MATIS (Iceland), who preliminarily isolated and cultivated the bioactive strain. The bacterial cultures were inoculated with two types of inoculum. The first set of six flasks was inoculated with 10% inoculum obtained from a 1mL glycerol-marine broth stock culture. The second set was inoculated with single colonies from marine agar plates. In triplicates, the cultures were incubated at 45 °C, the growth temperature similar to the natural Icelandic habitat of the microorganism and shaken at 120 rpm. These culture conditions were similar already established in MATIS. The incubation was stopped, and the cultures were harvested at different growth phases after 1, 3, 5, and 7 days. Comparisons were made in terms of changes in metabolite profiles and bioactivity at varying incubation periods to determine the optimum growth

phase to extract the secondary metabolites. Then, liquid cultures of SBT531 were scaled-up to 15 L (30 x 500 mL) in thirty-one-litre conical flasks using the optimised culture conditions based on optimum growth, metabolite profiles and bioactivity, by inoculating single colonies in marine broth, then incubated for 7 days at 45 °C and shaken at 120 rpm. At the end of the incubation period, the cultures were extracted, which led to the isolation of small molecules and pure compounds, presented in Chapter 3 of the thesis.

The results of the analysis on the metabolite production and chemical profile for SBT 587 came out to be identical to SBT 531. Hence, no further optimisation experiments and fractionation were performed on SBT 587 after its small-scale fermentation.

### **2.3.2 *Micromonospora sp.* N17 and N74 (SBT687 and SBT692)**

The bacterium *Micromonospora sp.* N17 (referred to as SBT 687) and *Micromonospora sp.* N74 (SBT692) were isolated from the Mediterranean sponge *Phorbas tenacior* from the Santorini volcanic complex of Crete. Bacterial cultures were grown at 30 °C and shaken at 120 rpm in ISP2 medium containing sea salt at 25 g/L. ISP2 medium contained Bacto -Yeast extract (4.0g/L) from Oxoid, UK, Bacto-Malt extract (10.0g/L) from Sigma-Aldrich Co.,USA, D-(+)-Glucose monohydrate 99% (4.0g/L) from Alfa Aesar, England, sea salt (25g/L) Royal Nature Advanced Pro Formula Salt by Royal Nature, Israel, distilled water and bacto-agar(20.0g/L) from Oxoid, UK, in case of preparing agar plates). The chosen media was identical to those used by the research group at the University of Wuerzburg (Germany), who preliminarily isolated and cultivated the bioactive strain.

Bacterial cultures were optimized in 100mL ISP2 media with 0.025% w/v salt in twelve 200mL shake flasks. The bacterial cultures were inoculated with two types of inoculum. The first set of six flasks was inoculated with 10% inoculum obtained from a 1 mL glycerol ISP2 broth stock culture. The second set was inoculated with single colonies from marine agar plates. Again, in triplicates, the cultures were incubated at

30 °C, the optimal growth temperature similar to the microorganisms' natural habitat in Crete, and shaken at 120 rpm, which was the similar shaker conditions used in Wuerzburg. The incubation was stopped, and the cultures were harvested at different growth phases after 1, 3, 5, and 7 and 10 days. Comparisons were made in terms of changes in their metabolite profiles and bioactivity at varying incubation periods. The same optimized conditions were established for SBT687 and SBT692 except for the incubation period that took 7 days for SBT687 and 10 days for SBT692.

Based on optimal biomass of the cultures, as well as metabolite profiles and bioactivity of the extracts, the optimum growth phase conditions were achieved to harvest the cultures, the liquid cultures of SBT687 and SBT692 were then scaled-up to 15 L (30 x 500 mL) in thirty one-litre conical flasks, by inoculation with 10% (v/v) inoculum in ISP2 medium, then incubated at 30 °C and 120 rpm for 7 and 10 days, respectively.

## 2.4 Extraction

After the culture growth was terminated, about 500 mL of HPLC grade EtOAC was added to the bacterial cultures and left overnight. Then, samples were individually homogenized with an Ultra-turrax T 18 basic homogeniser, filtered using a Buchner funnel with 110mm Fisherbrand filters (Fisher Scientific, Hemel Hempstead, UK) and transferred to a 1L separating funnel. Followed by liquid-liquid partitioning of the aqueous phase with 500 mL of EtOAC three times. The organic solvent was evaporated in a rotary evaporator under vacuo; and the dried residue was reconstituted in a minimum volume of EtOAC to be transferred to a tared storage vial then dried over nitrogen. The extracts were then concentrated and stored at 4 °C till further work up. The dried fractions were then reconstituted with the appropriate solvent for mass spectrometry, NMR spectroscopy and bioassay screening as shown on Table 2.1.

**Table 2-1 Samples analysis concentration and solvents used in the extraction procedure.**

Analysis	Concentration	Solvent
<b>LC-HRMS</b>	1 mg/mL	100% Methanol or 80:20 Methanol/DCM
<b>NMR</b>	5 mg/600 $\mu$ L < 5mg/200 $\mu$ L	DMSO-d6 Or other suitable solvent
<b>Bioassay screening</b>	10 mg/mL	DMSO

## **2.5 Liquid Chromatography – Mass Spectrometry (LC-MS)**

### **2.5.1 Material and instruments**

HPLC grade solvents that included MeOH, DCM, ACN, FA, and water were used. For crude extracts, fractions and pure compounds, both positive and negative mode experiments were carried out using an Exactive mass spectrometer with an electrospray ionization source attached to an Accela 600 HPLC pump with Accela autosampler and UV/Vis detector from Thermo Scientific, Bremen, Germany. Mass spectrometry was carried out over a mass range of 100-2000 Da in the positive and negative ionization modes using a spray voltage of 4.5 kV and capillary temperature at 270°C. Injecting volume used for each sample was 10 µL. A flow rate of 300µL/min was utilised. The column used was an ACE Excel 3 C18 column (3µm x 150mm x 3mm) from Hichrom Limited, Reading, UK.

### **2.5.2 Methods**

All samples, crude extracts, fractions and pure compounds, were prepared at a concentration of 1 mg/mL in 80:20 MeOH:DCM or 100% MeOH, depending on the solubility characteristics of the samples. A solvent and media blank were also included in MS runs. A binary gradient method was utilized. The two solvents consisted of A (water and 0.1% formic acid) and B (ACN and 0.1% formic acid). The gradient was carried out for 45 min culminating with 90 A % and 10% B at 0 min following a 30 min gradient to 0% A and 100% B, which remained isocratic until the 45<sup>th</sup> min after which the column was further re-equilibrated back to 90 A % and 10% B for another 9 min. The UV absorption wavelength was set at 254 nm, the sample tray temperature was maintained at 4°C and the column maintained at 20 °C. The samples were run sequentially, with solvent and media blanks analysed first. HR-LCMS data was acquired using Xcalibur version 2.2 from Thermo Scientific, Bremen, Germany, and processed using the software described in section 2.9.

## **2.6 Medium pressure liquid chromatography (MPLC) and preparative high-performance liquid chromatography (prep-HPLC)**

### **2.6.1 Material and instruments**

The solvents and reagents used were HPLC grade MeOH, DCM, EtOAc, ACN, FA, acetone, water and were purchased as described in section 2.1. Medium pressure and preparative liquid chromatography systems were used to isolate and purify the bioactive secondary metabolites. The elution rate is higher than those used in quantitative high-pressure liquid chromatography to attain a high-throughput separation of the compounds from higher loading concentrations of mg to g levels. To determine which solvent mixture and column to be used in the experiment, separation of the different metabolites were optimised using different combinations of eluting solvent systems on reverse (C18) and normal (silica) phased thin layer chromatography (TLC). Details of the TLC procedure could be read in section 2.7.

One of the systems used was the Biotage Isolera™ One 2.0.4 Spektra flash purification system from Biotage, Uppsala, Sweden. This had a photodiode array (PDA) detector covering an UV wavelength range of 200-400 nm but did not have an evaporative light scattering detector (ELSD). The Biotage had four solvent channels allowing four solvents to be used in a binary gradient during a single run. Different Biotage® SNAP loading cartridges were used according to individual sample characteristics, such as the amount, polarity, complexity and solubility. Details on the respective samples will be presented in the results and discussion chapters). The sample extract was dissolved in a suitable solvent prior to loading it to a samplet. The samplet was then dried to get rid of the carrier solvent and was placed over the column.

Another system used was the Reveleris® Flash Forward from Grace Davison Sciences, UK. This system has an advantage over the Biotage as it has two detectors, an ELSD and a UV detector from which a certain working wavelength can be chosen between the given ranges of 200-500 nm. This guarantees superior sensitivity, selectivity and

detection of all peaks including UV-inactive compounds. This system also allowed a binary solvent gradient, but four different solvents could be used in a single run because there are four independent solvent channels. The dissolved sample was mixed with Celite<sup>®</sup> S and the carrier solvent was evaporated. The sample-loaded celite was packed into a solid sample loader, the size of which was dependent on the amount of sample and the size of the preparative column to be used that were obtained from Reveleris, USA.

The appropriate size of the column and the flow rate, which ranges from 4 to 200 mL were set according to the amount and complexity of the respective samples in terms of the number of compounds to be separated. The pressure was automatically adjusted according to the size of column used. The collector was built into the system and trays recognised by the software. At final of the run the chromatograms could be saved and printed.

The prep-HPLC system used was the Reveleris<sup>®</sup> Prep HPLC Flash forward system from Grace Davison Discovery Sciences, US. The conditions were similar to the Reveleris<sup>®</sup> Flash Forward as this system has also two detectors, an ELSD and a UV detector from which a certain working wavelength can be chosen between the given ranges of 200-500 nm, the pressure was automatically adjusted according to the size of the column used. However, the flow rate used could be lower (1 to 200 mL/ min) than was recommended on the column brochure, the ELSD threshold is higher than 5 mv, liquid samples were injected up to 5 mL volumes directly into the system *via* a 6-port loop injection valve. The column used was VisionHT C18 Highload 5µm from Dr.Maisch, Germany with 150 mm x 10mm dimensions.. Prior to injection, a sample was dissolved into a suitable solvent, which is usually the same as the initial solvent system of the chromatographic gradient. The mixture of the samples and solvents were filtered through 0.22 µm Millex<sup>®</sup> GP from Merck Millipore, Ireland.

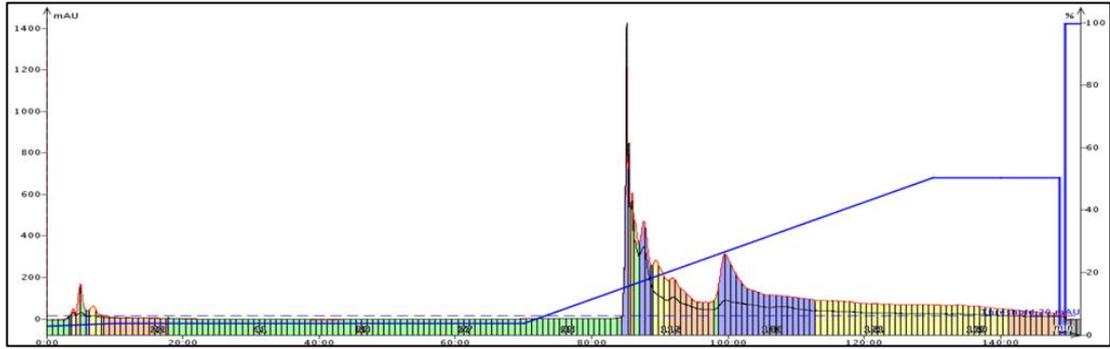
At the end of the respective runs, columns were then flushed with 100mL of 1:1 MeOH:Acetone and collected as the wash. Fractions were then pooled using UV read-outs for the respective peaks from the Chromatography System and their similarities on a RP18 reverse or Si60 normal phase TLC plate eluted with the appropriate solvent

system. Reverse phase plates were eluted with 1:1 ratio of ACN:H<sub>2</sub>O. Aliquots of each fraction were prepared for MS (0.1 to 1 mg), NMR (where extract weight was at least 1 mg) and bioactivity assays (10mg/mL) as described in detail in Chapters 3, 4 and 5 under results and discussion.

## **2.6.2 Methods**

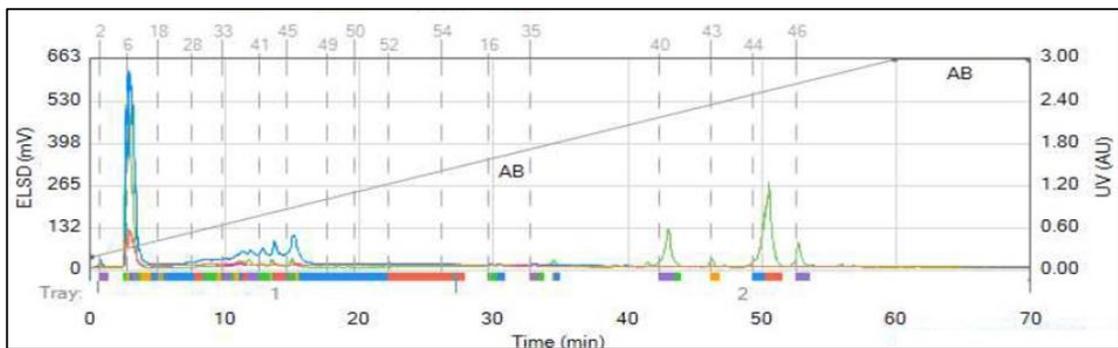
### **2.6.2.1 *Muricauda ruestringensis* (SBT531)**

First fractionation was performed on 1.3086 g of crude extract using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification system. The sample was loaded on a normal phase SNAP HP-Sil 25g cartridge with a 3g sample. Gradient elution method was used with DCM (A) and MeOH (B) as solvent system. The gradient was carried out for 160 min starting with 2% B at 0 min to 3% B in 5 min, which was held for 70 min, then increased to 50% B at 70 min, which remained isocratic for the rest of the run (Figure 2.1). The flow rate was at 12mL/min and maximum collection volume was set at 10mL. The system marked the respective test tubes based on the eluting peaks at real time as determined by UV from 254 to 280 nm. The column was then flushed with 100mL of 1:1 MeOH:Acetone and this was collected as the wash. Fractions were then pooled using UV read-outs from the Isolera™ One 2.0.4 Spektra flash purification system and similarities in TLC on Si60 normal phase plates using DCM and MeOH as solvent system.

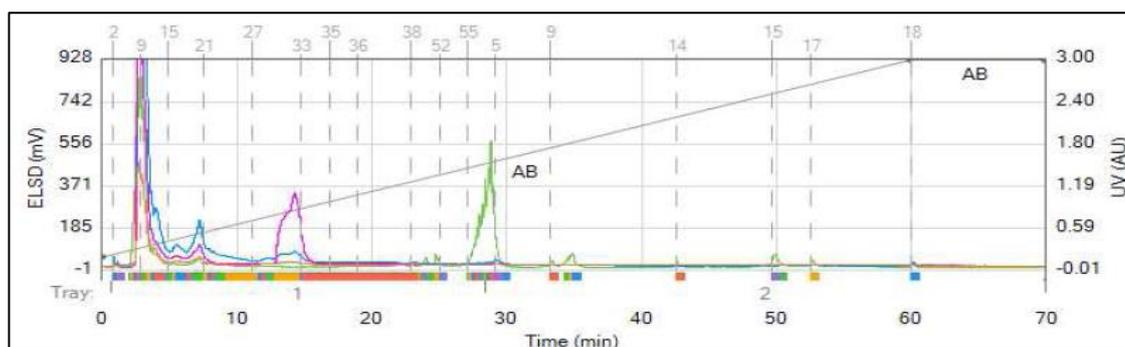


**Figure 2-1 Chromatogram from SBT531 first fractionation using the Biotage Isolera™ One 2.0.4 Spektra flash purification system.**

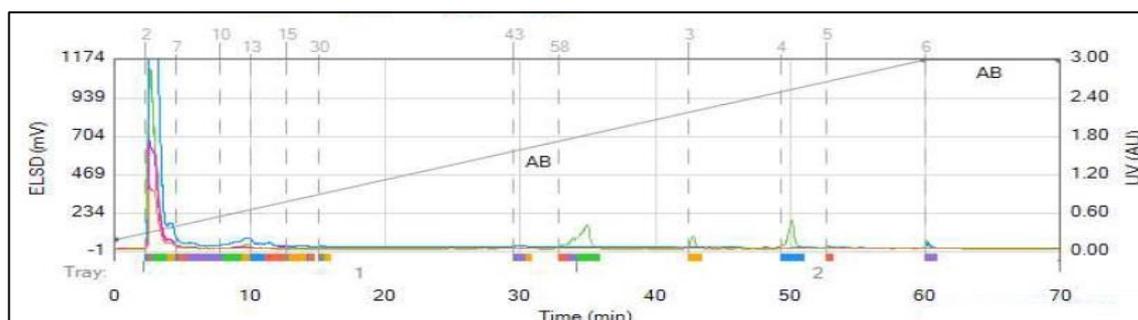
Fraction 4 (SBT2311) Figure 2.2, Fraction 9 (SBT2316) Figure 2.3, Fractions 6 and 7 combined (SBT2313 and SBT 2314) Figure 2.4, were subjected to further purification using the Reveleris Prep HPLC system. The column vision HT C18 5u was used on a binary gradient elution method using water with 0.1% FA (A) and ACN with 0.1% FA (B) as solvent systems. The gradient was carried out for 70 min with a flow rate of 5mL/min as follows: culminating with 5% B at 0min to 100% B at 60min, which was then held for 10 min. Samples were dissolved in less than 5mL of the starting eluents and filtered prior to injection. It is important to use an injection volume less than the 5 mL injection loop volume.



**Figure 2-2 Chromatogram from SBT531-Fraction 4 (SBT2311) sub-fractionation using the Reveleris Prep HPLC system.**



**Figure 2-3 Chromatogram from SBT531-Fraction 9 (SBT2316) sub-fractionation using the Reveleris Prep HPLC system.**

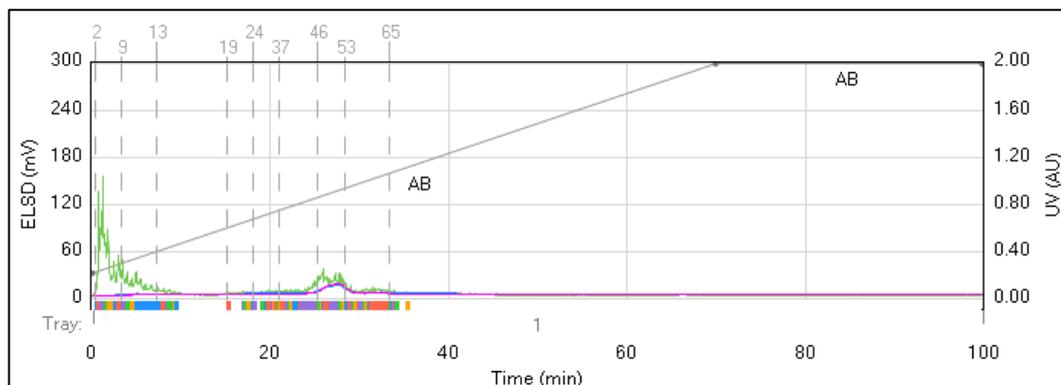


**Figure 2-4 Chromatogram from SBT 531-Fraction 6+7 (SBT2313+SBT2314) sub-fractionation using the Reveleris Prep HPLC system.**

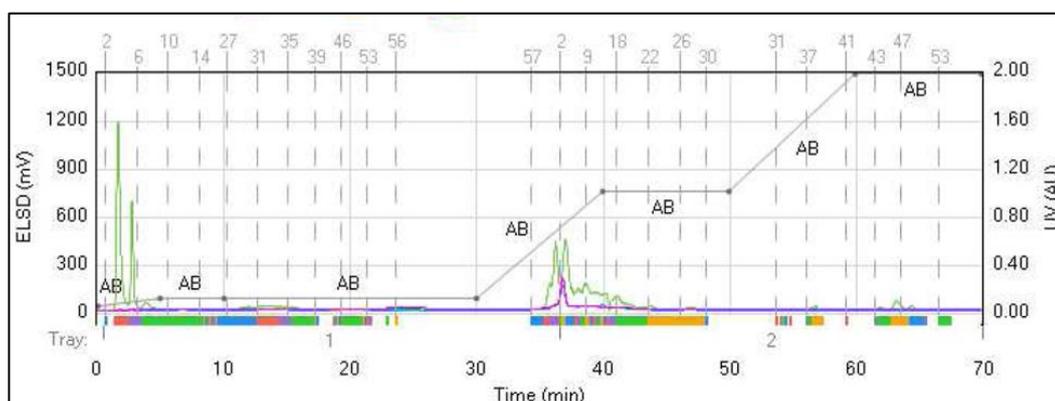
### **2.6.2.2 *Micromonospora sp.* N17 (SBT687)**

First fractionation was performed on 131.9 mg of crude extract using the Reveleris™ Flash Chromatography System using a reverse phase C18 column 40g and gradient elution method using water (A) and ACN (B). The gradient was carried out for 100 min as follows;10% to 100% B from 0 to 70 min, which was held at 100% B for another 30 min for the rest of the run (Figure 2.5). The flow rate was 20mL/min and fractions were collected at a maximum volume of 15mL each. Fractions were then pooled by also using their chromatogram similarities on a reverse phase TLC plate.

The total yield of all fraction were sent to PharmaQ (Norway) to perform bioactivity inhibition assays against parasite *Lepeophtheirus sp.* (sea lice).



**Figure 2-5 Chromatogram from SBT687 first fractionation using the Reveleris™ Flash Chromatography System.**

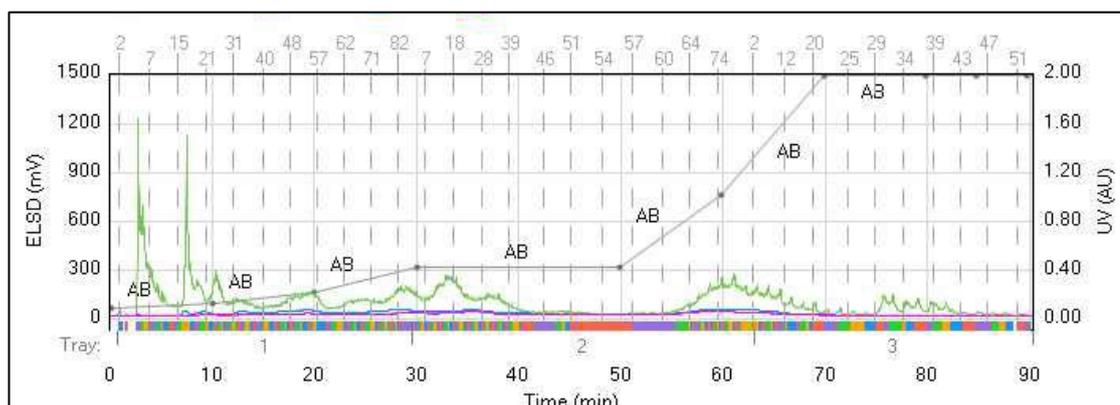


**Figure 2-6 Chromatogram from SBT687 second fractionation using the Reveleris™ Flash Chromatography System.**

A second fractionation was performed on 396.1 mg of crude extract using the Reveleris™ Flash Chromatography System with a 40g C18 reverse phase column by gradient elution method using water (A) and ACN (B). The gradient was carried out for 70 min as followed; 2% to 5% B at 0 to 5 min, which was then held for 25 min at 5% B, then gradient increased to 50% B for 10 min to the 40<sup>th</sup> min then to 100% B to the 60<sup>th</sup> min which was then held for another 10 min to the end of the run (Figure 2.6). The flow rate was 40mL/min and fractions were collected to a maximum volume of 25mL each.

### 2.6.2.3 *Micromonospora* sp. N74(SBT692)

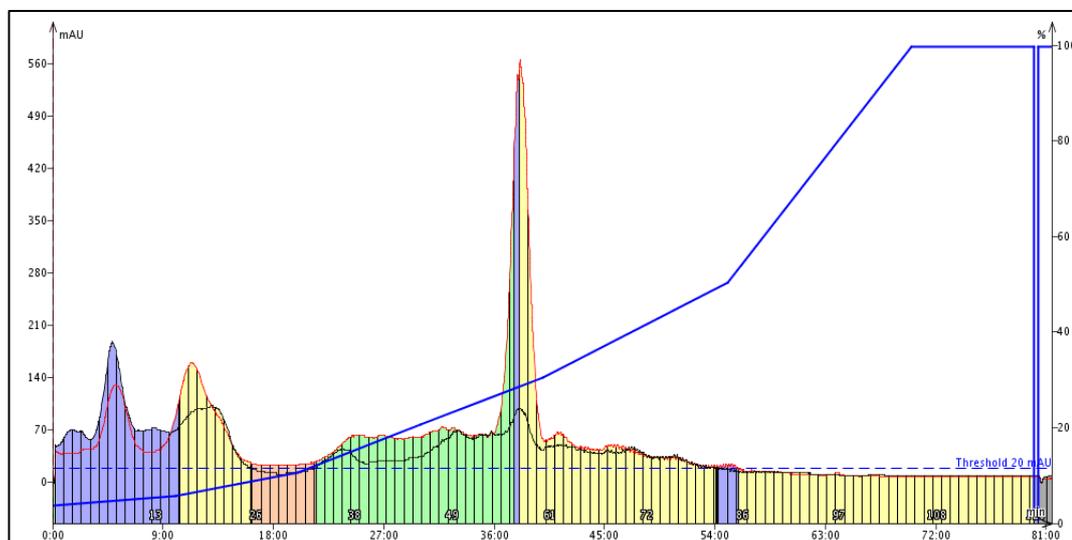
First fractionation of the 15L scale-up was performed on 1.3g of crude extract using the Reveleris™ Flash Chromatography System using a reverse phase C18 column 40g and gradient elution method using water and 0.1% FA (A) and ACN and 0.1% FA (B). The gradient was carried out for 95 min and the program followed; at 0 min B=3% ,going to 5% at 10 min which was held for 10 min, going 10% at 20 min during 10 min, going to 20% at 30 min which was held during 20 min, going to 50% at 50 min which was held during 10 min, going to 100% at 60 min which was held the rest of the run (Figure 2.7). The flow rate was 20mL/min and test-tubes containing a maximum volume of 25mL were collected based on the real time indication of eluting compounds, determined by UV.



**Figure 2-7 Chromatogram from SBT692 first fractionation using the Reveleris™ Flash Chromatography System.**

Second fractionation of the 15L scale-up was performed on 1.1 g of crude extract using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification system on a reverse phase SNAP Kp-C18-Hs 30g cartridge with a 3g C18 sample. A gradient elution method was used with 0.1% FA in water (A) and 0.1% FA in ACN (B). The gradient was carried out for 85 min with a stepwise increase of solvent B every 10 minutes. The run started with 3% to 5% B for the first 10 min, then increased from 5% to 10% B in the next 10 min or the 20<sup>th</sup> min, followed by 10% to 20% B at 30 min, 20% to 30% at 40 min, 30% to 50% at 55 min, 50% to 100% at 70 min, which was held for 10 min during the rest of the run (Figure 2.8). The flow rate was 25mL/min and maximum

collecting volume was 18mL. Fractions were monitored on Si60 normal phase plates using DCM and MeOH as solvent system.



**Figure 2-8 Chromatogram from SBT 692 second fractionation using the Biotage Isolera™ One 2.0.4 Spektra flash purification system.**

## **2.7 Thin layer chromatography (TLC) and preparative TLC**

### **2.7.1 Materials and instruments**

TLC was performed on normal phase silica gel 60 F<sub>254</sub> aluminium-plates and on reverse phase silica gel 60 RP-18 F<sub>254</sub>S aluminium plates. Preparative TLC were done on TLC plates silica gel 60 F<sub>254</sub> 20x20 cm aluminium sheets and were all purchased from Merck KGaA, Germany. The capillary tubes used to load the samples on the TLC plates were obtained from Hirschmann, Germany. The handheld UV lamp used as a detector has two wavelengths at 254 (short UV) and 365 nm (long UV). The spraying reagent as described below was prepared with sulfuric acid from Fisher Scientific, UK, acetic acid from Sigma-Aldrich, US and Anisaldehyde from Acros Organics. The other

reagents used to dissolve the sample and to the solvent systems vary according to the individual sample characteristics and are detailed in section 2.1.

## 2.7.2 Methods

The TLC allows a good and relatively simple separation of the compounds of an extract or a fraction. Frequently, the solvents used in normal-phase are DCM, MeOH, EtOAc and n-hexane while in reverse phase is ACN and water. Sometimes drops of formic or sulfuric acid was added to enhance the separation. The samples were dissolved in suitable solvents (usually MeOH or acetone) and a small volume was spotted 1 cm above the bottom edge of the TLC plate. Chromatograms were developed until the eluting solvent reached around 5 cm of length of the plate. The developed plates were then dried and were observed under the UV lamp. Any band observed was marked with a carbon pencil, usually using the letters S (short), L(long) and B (blue), V(violet), O(orange), G(Green), R(red) for the wavelength used and different colours observed. TLC plates were then sprayed with Anisaldehyde/H<sub>2</sub>SO<sub>4</sub> reagent (mixture of 0.5 mL of anisaldehyde, 85 mL of methanol, 10 mL glacial acetic acid and 5 mL of concentrated H<sub>2</sub>SO<sub>4</sub> added slowly to give a final volume of 100.5 mL) and heated at 170°C-200°C using a heat gun that allows the visualization of organic compounds without UV absorbance such as steroids, terpenes, phenolic acids among others carbonyl group containing molecules. The R<sub>f</sub> values of the observed bands that corresponded to the analyte migration were calculated according to the following equation:

$$R_f \text{ value} = \frac{\text{distance of analyte migration}}{\text{distance of mobile phase migration}}$$

A preparative TLC was used when the amount of sample was too low (<50 mg) to perform further fractionation by prep-HPLC. An analytical TLC was first accomplished to get the suitable solvent system for the best chromatogram separation for the preparative experiment. The TLC chamber was equilibrated with the mobile

phase for at least 10 min by placing a filter paper inside the chamber. Equilibration was completed when the filter paper is saturated with the solvent system. Meanwhile, the samples were spotted along a line 2 cm above the bottom edge of each TLC plate. After elution the bands were observed under UV light and were marked using a carbon pencil as described above. All individual eluted bands were recovered by cutting the marked bands into several pieces and immersed then stirred in 100 mL acetone overnight. The stirring step was repeated twice, each time for 1 hour, and the supernatant was filtered and collected in a tared round bottom flask to dry using the rotary evaporator. The dried residue was reconstituted in a suitable solvent to be transferred to a weighed vial, then dried under nitrogen prior to further analysis.

## **2.8 Nuclear Magnetic Resonance**

### **2.8.1 Material and instruments**

NMR experiments were carried out on a Jeol-LA400 FT-NMR spectrometer system equipped with a 40TH5AT/FG probe (JEOL, Tokyo, Japan) in Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde. Selected samples were measured on the Bruker Biospin GmbH-NMR instrument from Bruker, Germany, equipped with a 5mm BB-1H/19F/D probe, both AVIIIHD500 and AV600 are equipped with a BCU-05 unit for automatic cooling of probes to 0° C, from the Department of Pure and Applied Chemistry, University of Strathclyde.

### **2.8.2 Methods**

Crude extracts and fractions were prepared by dissolving 5 mg of bacterial extract in 600 µL DMSO-d<sub>6</sub> (Sigma-Aldrich, Dorset, UK) or other suitable deuterated solvent. These were transferred to 5 mm 7" 400 MHz NMR tubes (Norell, US. Wilmad®). For fractions and pure compounds generated with limited sample weights (<5 mg), the total fraction was dissolved in 200 µL DMSO-d<sub>6</sub> (Sigma-Aldrich, Dorset, UK) and transferred to a 3mm 7" 400 MHz NMR tube (GPE scientific, Bedfordshire, UK).

For Mosher ester experiments for determining the absolute configuration of stereogenic carbinol carbons, 1 mg aliquots of each sample were dissolved in 700  $\mu\text{L}$  of pyridine- $d_5$ , transferred to two NMR tubes and had their  $^1\text{H}$  and  $^1\text{H}$ - $^1\text{H}$  COSY spectra measured. Afterwards, 5  $\mu\text{L}$  of (*S*)-(+)-MTPA-Cl and ((*R*))-(-)-MTPA-Cl were added to the respective aliquots. The NMR tubes were shaken thoroughly and were allowed to stand at room temperature for 72 hours. The reaction was monitored by  $^1\text{H}$  and  $^1\text{H}$ - $^1\text{H}$  COSY every 24 hours. Changes in chemical shifts of the stereogenic carbinol protons for the respective chiral reagents were then recorded at the end of the reaction. All experiments were processed by MestreNova (Mnova) 10.0.2 developed by Mestrelab Research, Spain. Chemical shifts are given in ppm and coupling constants in Hz.

## 2.9 Software

### 2.9.1 Data analysis tools for LC-MS data

Raw data were initially sliced into two data sets based on the ionization mode (positive and negative modes) using the MassConvert tool from ProteoWizard. The sliced data sets were imported and processed in modified version of Mzmine 2.10 using predefined settings to extract features from the raw data. The following data processing steps were carried out using MZmine; peak detection, (mass detection and chromatographic builder), deconvolution, deisotoping, filtering, alignment and gap filling. Identification of adducts and complexes and formula prediction steps were carried out to predict possible molecular formulae for each feature and to minimize mis-assignment of features by eliminating adducts and complexes.

Data was then exported as a CSV file for further clean-up. An In-house EXCEL Macro was developed by Dr Tong Zhang and was adapted for both dereplication and metabolomics study as described by Macintyre et al., (2014). An accustomed library was created by employing an algorithm to use the molecular formula data set from

Antibase® (February 2013), Marinlit® (September 2013), and DNP (December 2016) from which the monoisotopic exact masses were recalculated. The customized library was used instead of the manually curated Antibase, MarinLit and DNP databases, which do not differentiate between monoisotopic, average, and most abundant mass. The created library was then coupled to MZmine and employed as the custom database for peak identification and dereplication. “Hits” and unidentified peaks were double checked against the MS raw data in Xcalibur 2.2 (Thermo, Scientific, Germany). A CSV file was also used to generate a heat map from programming software R version x64 3.0.3 using a script utilizing the g-plot package (R Foundation for Statistical Computing, Austria). The in-house macro was utilized to match the  $m/z$  ion peaks in each bacterial extract with compounds in the database (using RT and  $m/z$  threshold of 3ppm) to provide details on the putative identities of all metabolites in each bacterial extract and the number of remaining unknowns for each extract.

The data was analysed using SIMCA V 15.0.2. For the MS data, the MZmine ID number was merged with the ionization mode to generate a unique primary ID, while the other variables like retention time,  $m/z$ , and molecular weight were considered as secondary IDs. Principal component analysis (PCA), an unsupervised statistical analysis method was initially used to analyse the data sets. A supervised statistical analysis method with orthogonal partial least squares discriminant analysis (OPLS-DA) was afterwards utilised to compare groups and discriminate metabolites according to a known variable that classified the groupings. Both methods were done applying the Pareto scaling and the models were validated based on multiple correlation ( $R^2$ ) and cross-validation ( $Q^2$ ) coefficients as well as by permutation tests for the supervised method. The In-house EXCEL Macro described above was then employed to identify the top 20 features (ranked by peak intensity) and corresponding putative identities in each sample by creating individual CSV files for each extract. Hits from the database were accessed using ChemBioFinder version 13 (PerkinElmer Informatics, Cambridge, UK).

## 2.9.2 NMR

The data obtained was processed with MestReNova (Mnova10.0) software to confirm chemical structures.  $^1\text{H}$  spectrums were processed. The baseline was corrected by manual phasing and by using the Whittaker Smoother. Gaussian was set to 1 Hz for apodization. Chemical shifts were given in ppm and coupling constants in Hz. The NMR data was also statistically analysed using SIMCA V 15.0.2 For the NMR data the chemical shifts in ppm was used to generate the unique primary ID while there were no other secondary IDs considered.

## 2.10 Optical Rotation

Compounds with chiral centres were required to determine their optical rotation. Stereoisomers could rotate linearly polarised light counter-clockwise (-) or clockwise (+). Optical rotation experiment was performed with Mr Gavin Brain from the Department of Chemistry, University of Strathclyde, based on a standard operating procedure for measuring optical rotation on the Perkin Elmer 341 polarimeter (PerkinElmer Inc., US) at 589 nm. The compound samples were dissolved in MeOH or other suitable solvent (spectral grade) to a concentration of 2mg/mL. The specific optical rotation for each sample is calculated using the **equation**:

$$[\alpha]_{\lambda}^T = \frac{100\alpha}{lC}$$

Where  $\alpha$  is the average of first ten readings of the rotation value,  $l$  the cell volume in mL (the value of  $l$  for the micro test cell is 1) and  $C$  is the concentration in g/mL, measures done at 20° C.

## **2.11 Bioassay screening**

### **2.11.1 Materials, reagents, instruments and software**

The cell lines used in this project were human normal keratinocytes (NCTC), human colon carcinoma derived from lung metastasis (T84), human colon adenocarcinoma primary tumour (HT29), human colon adenocarcinoma (CaCo2), human breast adenocarcinoma (MCF-7), human breast adenocarcinoma (MDA-MB-231). All cell lines were deposited and maintained at the University of Strathclyde. Medium 199 (M199) and geneticin was purchased from Life Technologies, Thermofisher, UK. Eagle's medium (DMEM), Eagle's red phenol free (DMEM) and TrypLE™ Express were purchased from Gibco BRL, UK. Foetal bovine serum (FBS), L-glutamine, penicillin/streptomycin solution and blastidicin were from Invitrogen, UK. Hanks' balanced salt solution (HBSS), non-essential amino acid (NEAA), TNF alpha, MG132 and Triton X-100 were purchased from Sigma-Aldrich, UK. Bright-Glo was purchased from Promega, UK. AlamarBlue® was obtained from Biorad. For the experiments in this project, 25 cm<sup>2</sup>, 75 cm<sup>2</sup> and 150 cm<sup>2</sup> sterilised tissue culture flasks, falcon tubes, 96-well black plates with clear bottom with lid were from Corning Incorporated, USA. Clear 96-well tissue culture plates were from TPP®, Switzerland. Three different types of micro 96-well plates, ones with a clear bottom, others half area clear and half area solid black area plates were purchased from Greiner bio-one, Germany.

The cell lines were all grown and incubated in Napco 5410 incubator from Napco, USA or HERA cell 150 from Thermo Scientific, UK, at 37°C in a humidified atmosphere saturated with 5% CO<sub>2</sub>. The cabinet used was the SterilGard biological safety cabinet from The Baker Company, USA. The IEC Medispin centrifuge was purchased from Thermo Scientific, Germany. The Bright-Line haemocytometer was obtained from Reichert, US. The water bath was from Clifton, UK. The microscope was from Olympus Optical, Japan. The Wallac Victor 2 (PerkinElmer, UK) was used for measuring the fluorescence, ex560/em590nm and ex360/em460nm, absorbance at 405nm and luminescence and was from PerkinElmer, UK. The calculations were performed by Microsoft EXCEL 2019 and graph bar charts, dilution and dose response

curves were designed by non-linear regression analysis using GraphPad Prism V 6.0 developed by GraphPad Software, US.

### **2.11.2 Cell culture: media preparation, cell splitting and seeding**

All cell lines were stored long term in liquid nitrogen. The NCTC medium was cultured in M199 medium that was supplemented with 100 IU/mL/100 µg/mL penicillin/streptomycin and 10% (v/v) FBS. Cell lines T84, HT29, MCF-7, MDA-MB-231 were cultured in DMEM supplemented with 1% (v/v) L-glutamine and the same as for M199. CaCo2 was also cultured in DMEM but supplemented with additional 1% of NEAA. The pH of all media was adjusted between 7.35-7.45 by adding drops of sterile sodium hydroxide 0.1N. After collecting the cell lines from storage cryogenic dewars, they were thawed in a water bath at 37°C and then were transferred to a 25 cm<sup>2</sup> seeding flask with 5 mL of medium and grown at 37°C in a CO<sub>2</sub> incubator. The cells were checked every day to verify the growth and if necessary, the medium was changed to feed the cells to promote the growth. The cells were sub-cultured by trypsinisation twice a week, every 3 – 4 days and maintained at 37°C in a humidified atmosphere saturated with 5% CO<sub>2</sub> until the cells reached 80% confluency. After checking the confluency under the microscope, the medium was removed from the flask containing the cells. The cells were then washed twice with HBSS, removing and discarding the HBSS after each wash. TrypLE Express was then added to the flask to detach the cells from the flask wall. The cell lines were incubated for 2-10 min at 37°C. The incubation period depends on the used cell line. Pre-warmed medium was added to the flask to stop the action of TrypLE Express. Cells were then transferred to a 15 mL falcon tube then centrifuged for 2 mins. After centrifugation, the supernatant was discarded, and the remaining cell pellet was resuspended in pre-warmed medium at 37°C. The volume of medium used to resuspend the cell pellet was proportional to the size of the pellet. Meanwhile, the haemocytometer was cleaned with 70% alcohol and 15 to 20 µL of the cell suspension was added to the chamber to count the number of cells. The volume of the cell suspension transferred to a new seeding flask was calculated according to the following **equation**:

$$v = \frac{\text{seeding density} \left( \frac{\text{cells}}{\text{cm}^2} \right)}{\text{cell count} \times 10^4 \left( \frac{\text{cells}}{\text{mL}} \right)} \times \text{area of the flask (cm}^2\text{)}$$

The volume of the cell suspension used to seed a 96-well plate for the different assays was calculated according to the following **equation**:

$$v1 = \frac{\text{seeding density} \left( \frac{\text{cells}}{\text{cm}^2} \right)}{\text{cell count} \times 10^4 \left( \frac{\text{cells}}{\text{mL}} \right)} \times v2$$

Where *v1* is the volume of cell suspension in mL to be added to the new falcon tube completed to *v2*, whilst *v2* represents the total needed volume of medium in mL used for the 96-well plate with a varying number of samples. The flask seeding densities for the cell lines vary according to the cell growth and were listed on Table 2.2.

**Table 2-2 Seeding densities (cell/cm<sup>2</sup>) for the used cell lines.**

Cell line	NCTC	T84	HT29	CACO2	MCF-7	MDA-MB-231
<b>Flask</b>	0.4 -0.8 x 10 <sup>4</sup>	1 -3 x 10 <sup>4</sup>	1 -3 x 10 <sup>4</sup>	1 x 10 <sup>4</sup>	0.5x10 <sup>4</sup>	0.5x10 <sup>4</sup>
<b>96-well plate</b>	5 x 10 <sup>4</sup> or 2 x 10 <sup>4</sup> (NFkβ or cytotox)	10 x 10 <sup>4</sup>	10 x 10 <sup>4</sup>	10 x 10 <sup>4</sup>	10 x 10 <sup>4</sup>	10 x 10 <sup>4</sup>

Initially all fractions were tested at a concentration of 30 µg/mL for all the performed assays. To obtain the dilution curves, samples were tested from 0.003 to 100 µM by serial dilution. After a certain number of hours of incubation time, fluorescence, absorbance and luminescence were read on a Wallac Victor 2. Further statistical procedures were carried on GraphPad Prism v6.0.

### 2.11.3 Sample preparation

The samples were prepared by dissolving the required weight of sample in an appropriate quantity of DMSO, to give a concentration of 10 mg/mL or 1 mg/100 $\mu$ L. These samples were then stored as stock samples and were kept at -20°C in a sealed polypropylene plate. This plate was the stock solution used for serial dilution for all the assays. Initially, all the samples were tested at one concentration of 30 $\mu$ g/mL (1.2  $\mu$ L sample from the stock solution with 99  $\mu$ L of medium) to evaluate their *in vitro* activity. However, where there was elicited activity, these samples were added in serial dilutions ranging from 0.003  $\mu$ M to 30  $\mu$ M to obtain a dilution curve (Figure 2.9).

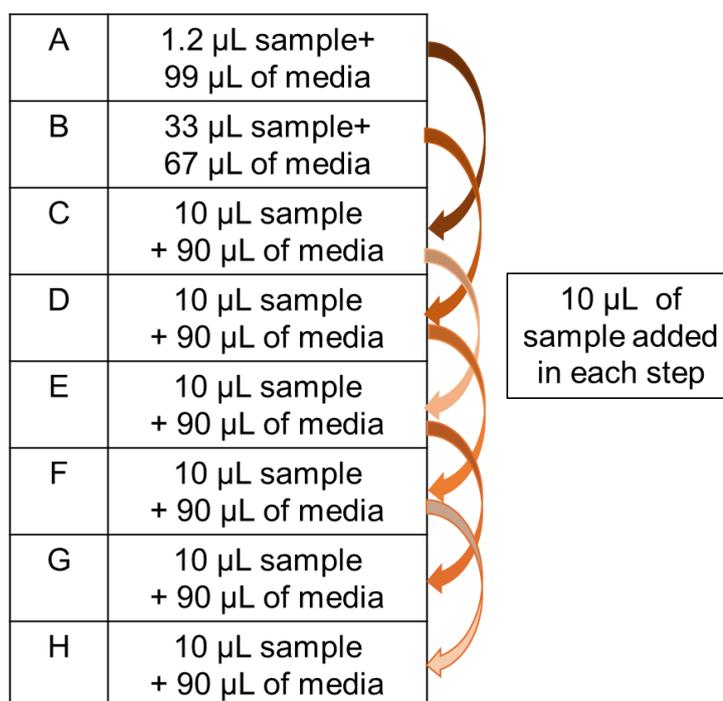


Figure 2-9 The preparation of the serial concentration dilutions of the samples.

#### **2.11.4 Alamar Blue assay: cell viability and cytotoxicity**

Crude extracts, fractions, and pure compounds from SBT531, SBT687 and SBT692 strains were tested on T84, HT29, CaCo2, MCF-7 and MDA-MB-231 cell lines as well as on normal NCTC cell line, in order to test their cytotoxicity using the Alamar Blue assay. Cell viability and consequent cytotoxicity to cancer cells were tested by following the same principles and steps used on normal cells. All samples were tested in independent triplicate assays and the viability percentages of control were calculated. The crude extract or fraction was considered to have cytotoxic effect if it resulted in  $\leq 40\%$  viability of the tested cells with results confirmed by microscopic visualisation. For the initial screening, the crude extracts and fractions were tested at a concentration of  $30\mu\text{g/mL}$ .

The positive control (column 1 of the 96-well plate) consisted of untreated cells with 100% viability, which means wells containing only cells with medium, and the negative control contained cells treated with 0.1% of Triton X-100 (column 12). All samples were added to the cell lines to test at a final concentration of  $30\mu\text{g/mL}$  in a 96-well clear plate and incubated for 48 hours at  $37^{\circ}\text{C}$  in a humidified atmosphere saturated with 5%  $\text{CO}_2$ . In parallel, T84 and HT29 cell lines were incubated for 72 hours with sample. Alamar Blue, in a final concentration of 10% (v/v), was then added to the seeding plate containing the cell line with the test samples to assess the cell viability. The test plates were incubated for 4 to 6 hours after which, fluorescence measurements were taken at 560 nm and 590 nm wavelengths using a Wallac Victor microplate reader. The seeding plates were observed under the microscope throughout the incubation period, and perceptible cell changes were recorded to compare with the statistical results.

#### **2.11.5 NF $\kappa$ B Luciferase method.**

The NF $\kappa$ B Luciferase (inflammatory marker) assay was performed to test the inhibition of the NF $\kappa$ B pathway ability of SBT687 and SBT692 extracts and fractions on NCTC cells. This cell line was previously genetically modified and designed for monitoring

the NF $\kappa$ B signal transduction pathways. The cell line contained a luciferase gene that when activated, it induced the transcription of the luciferase reporter gene and can be measured by luminescence. NCTC cells were treated with 0.1% blasticidin (20  $\mu$ l in 20 mL) and 0.8% geneticin (160  $\mu$ l in 20  $\mu$ l) on every 2-3 passages to select the cells containing the modified genes. On the first day of the experiment, a black clear bottom plate was seeded with  $1 \times 10^4$  cells/well in a volume of 200 $\mu$ L of normal growth medium per well and incubated for 48 hours at 37°C in a humidified atmosphere saturated with 5% CO<sub>2</sub>. FBS (0%) phenol red-free DMEM was used to prepare the samples. After 72h, the medium was removed with a multichannel pipette and 50 $\mu$ l /well of 0% FBS phenol red-free DMEM was added. The cells must be in serum-free medium for at least thirty minutes prior to adding the test samples to the plate. Meanwhile, TNF $\alpha$  medium was prepared by adding 1.5  $\mu$ l of 10  $\mu$ g/mL of TNF  $\alpha$  to 10 mL 0% FBS phenol red-free DMEM medium. 10ml of TNF  $\alpha$  is used per plate to make a final concentration 0.75 ng/mL in each well. The TNF $\alpha$  medium was used to prepare the samples and MG132 control.

The MG132 curve was designated on column 12 (A12 to H12). MG132 was prepared in the TNF $\alpha$  medium. In well position A12, 5  $\mu$ l of 10mM MG132 stock solution was added to 245  $\mu$ l of TNF $\alpha$  medium, to make a final concentration of 100  $\mu$ M MG132. In well position B12, 1.5  $\mu$ l of 10mM stock of MG132 and 249  $\mu$ l TNF $\alpha$  medium was added to make a final concentration of 30  $\mu$ M MG132. Next, 1 in 10 serial dilutions of MG132 were performed. 25 $\mu$ L of MG132 from a previously prepared well was added to 225  $\mu$ L TNF $\alpha$  medium (Figure **2.10**). To column 1 (A to D,) 0% FBS phenol red-free DMEM (100  $\mu$ l) was added. To column 1 (E to H), 50  $\mu$ L of 0% FBS phenol red free DMEM and 50  $\mu$ L of TNF $\alpha$  medium were added, giving a total of 100  $\mu$ L. After 4h of incubation at 37 °C, the medium was removed from the wells and 50  $\mu$ L of 1:1 Bright-Glo: 0% FBS phenol red free DMEM was added to each well. The plate was wrapped in foil and incubated in the dark at room temperature for ten minutes. Wallac Victor 2 was used to read the luminance Iso96lum and results were analysed using GraphPad Prism v6.0.

A12	5 $\mu$ L of MG132+ 245 $\mu$ L of media
B12	1.5 $\mu$ L of MG132+ 249 $\mu$ L of media
C12	25 $\mu$ L of MG132+ 225 $\mu$ L of media
D12	25 $\mu$ L of MG132+ 225 $\mu$ L of media
E12	25 $\mu$ L of MG132+ 225 $\mu$ L of media
F12	25 $\mu$ L of MG132+ 225 $\mu$ L of media
G12	25 $\mu$ L of MG132+ 225 $\mu$ L of media
H12	25 $\mu$ L of MG132+ 225 $\mu$ L of media

**Figure 2-10 Preparation of serial concentration dilutions of the standard curve of MG132 for the NF $\kappa$ B Luciferase assay.**

### 2.11.6 PTP1B assay

The PTP1B assay was performed on SBT531 extracts, fractions and pure compounds and on SBT587 extracts. The assay buffer was freshly prepared for each experiment accordingly as shown in Table 2.3.

**Table 2-3 PTP1B assay buffer preparation.**

Reagent	Weight/Volume	Supplier/code
Catalase*	12.5 mg	Sigma C1345
HEPES (25mM)	297 mg	Sigma H3375
Sodium chloride (50mM)	146 mg	Sigma S9625
Dithiothreitol (2mM)	15.4 mg	Sigma D5545
EDTA (2.5mM)	36 mg	Sigma E1644
BSA (0.01mg/mL)	50 $\mu$ L of 10mg/mL	Sigma A2153

\*The catalase was added first and directly weighed into the buffer stock bottle followed by the other ingredients. 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid-d18 (**HEPES**), Ethylene-diamine-tetra acetic acid (**EDTA**), Bovine serum albumin (**BSA**). The PTP1B assay buffer was made up to 50 mL with distilled water at a pH of 7.2 (acceptable range: 7.18-7.22) at 37°C.

The enzyme buffer was prepared accordingly as shown in Table 2.4.

**Table 2-4 Enzyme buffer preparation.**

Reagent	Weight/Volume	Supplier/code
<b>HEPES</b> (50mM)	595 mg	Sigma H3375
<b>Dithiothreitol</b> (5mM)	38.5mg	Sigma D5545
<b>EDTA</b> (1mM)	14.6 mg	Sigma E1644
<b>NP-40</b> (0.05%)	25 $\mu$ L	Sigma 98379

4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid-d18 (**HEPES**), Ethylene-diamine-tetra acetic acid (**EDTA**), Nonidet™ P-40 (**NP-40**). The enzyme buffer was made up to 50 mL with distilled water with the pH adjusted to 7.2.

The PTP1B enzyme (Sigma P6244-50UG) was prepared with aliquots of 13  $\mu$ M stock into 20  $\mu$ l lots, made to 52 nM concentration, i.e. 20  $\mu$ l of 13  $\mu$ M in 5 mL enzyme buffer and aliquoted into 1 mL lots to aliquot into 100  $\mu$ l per tube. A 100  $\mu$ L of the 52 nM aliquoted PTP1B enzyme was added to 2.4 mL of PTP1B enzyme buffer per plate giving a final assay concentration of 1nM. The 6,89-difluoro-4-methylumbelliferyl phosphate (DiFMUP) solution was prepared by adding 1.71 mL of DMSO to 5mg of DiFMUP to make a concentration of 10 mM. For every plate, 5 $\mu$ L of 10mM DiFMUP was prepared in 1.25 mL of assay buffer. Bis(4-trifluoromethylsulfonamidophenyl)-1,4-diisopropylbenzine (Protein Tyrosine Phosphatase Inhibiter IV) (TFMS) (Calbiochem 540211) was prepared with 10mg of the inhibitor dissolved in 1.64mL of DMSO for a 10mM concentration.

Column 12 of the dilution plate provided the TFMS curve. On well A12, with a concentration of 400 $\mu$ M of TFMS (to make 100  $\mu$ M), 4 $\mu$ L of 10mM of TFMS was added to 96  $\mu$ L of assay buffer. In well B12, with a concentration of 120 $\mu$ M (final 30  $\mu$ M), 1.2 $\mu$ L of 10mM of TFMS was added to 99  $\mu$ L of assay buffer. From wells C12 to H12, 1 in 10 serial dilutions were performed using a multiple-channel pipette by taking 10  $\mu$ L of the TFMS mixed with the assay buffer from the previously prepared well into 90  $\mu$ L of assay buffer. To illustrate, 10  $\mu$ L was taken from A12/B12 to C12/D12, and then 10  $\mu$ L from C12/D12 to E12/F12 and finally taking 10  $\mu$ L from E12/F12 to G12/H12 (Figure 2.11).

A12	4 $\mu$ L TFMS+ 96 $\mu$ L assay buffer
B12	1.2 $\mu$ L TFMS+ 99 $\mu$ L assay buffer
C12	10 $\mu$ L TFMS + 90 $\mu$ L assay buffer
D12	10 $\mu$ L TFMS + 90 $\mu$ L assay buffer
E12	10 $\mu$ L TFMS + 90 $\mu$ L assay buffer
F12	10 $\mu$ L TFMS + 90 $\mu$ L assay buffer
G12	10 $\mu$ L TFMS + 90 $\mu$ L assay buffer
H12	10 $\mu$ L TFMS + 90 $\mu$ L assay buffer

10  $\mu$ L of  
TFMS added  
in each step

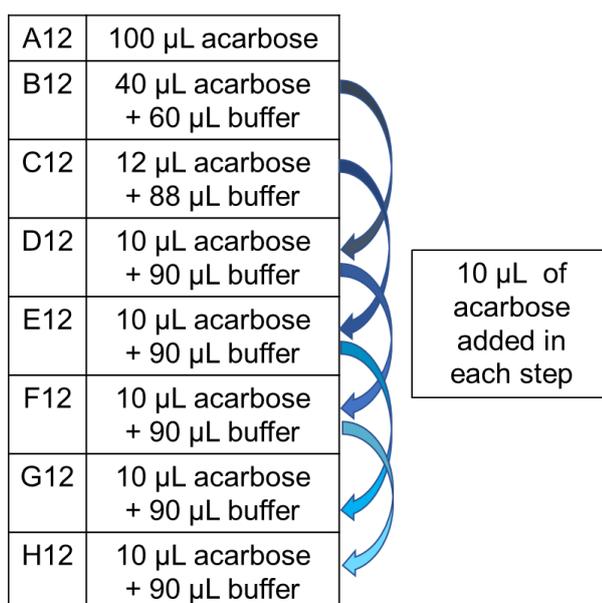
**Figure 2-11** The preparation of the serial concentration dilutions of the standard curve of TFMS for the PTP1B assay.

A half-area 96-well black plate was used for this assay method. To column 1 (positive control), 10  $\mu$ L of buffer was added. To column 12 (negative control), 10  $\mu$ L of TFMS was added to obtain the standard curve. To columns 2-11, 10  $\mu$ L of sample was added to 20  $\mu$ L PTP1B enzyme for each well. The plate was wrapped in cling film and incubated for thirty minutes at 37°C. After the incubation period, 10  $\mu$ L of DiFMUP substrate was added to each well and incubated for 10 minutes at 37°C. Wallac Victor 2 was used to read the fluorescence at ex360/em460nm and results were analysed using GraphPad Prism v6.0.

### 2.11.7 Alpha-glucosidase assay

The alpha-glucosidase screening assay was performed on SBT531 extracts, fractions and pure compounds as well as on SBT587 extracts. Solution A was prepared by dissolving 13.9 g of sodium phosphate monobasic dihydrate (Sigma 04269) in distilled water. Solution B was prepared by dissolving 26.8 g of sodium phosphate dibasic heptahydrate (Sigma S9390) in distilled water. The 0.1mM assay phosphate buffer was

prepared with 25.5mL of Solution A and 24.5mL of Solution B, then made up to 100mL using distilled water at pH 6.8. The yeast  $\alpha$ -glucosidase (EC 3.2.1.20) (Sigma G0660) at 0.2 units/mL was prepared by adding 13 $\mu$ L of the stock solution into 2.5mL of buffer. The substrate consisted of 1.8mg p-nitrophenyl- $\alpha$ -D-glucopyranoside (Sigma N1377) dissolved in 1500 $\mu$ L of buffer giving a concentration of 4mM. Acarbose (Sigma A8989-1G) solution was prepared by dissolving 13mg of acarbose in 200  $\mu$ L of buffer to give a concentration of 100mM then sonicated. Column 12 of the dilution plate was the standard curve for acarbose. Into well A12, 100  $\mu$ L of 100mM acarbose was added. To well B12, 40  $\mu$ L of 100mM acarbose was added into 60  $\mu$ L of buffer. In well C12, 12  $\mu$ L of 100mM acarbose was added to 88  $\mu$ L of buffer and from D12 to H12, 1 in 10 serial dilutions were performed using a multiple-channel pipette by adding 10  $\mu$ L of acarbose from the previous well to 90  $\mu$ L of buffer. To illustrate, by taking 10  $\mu$ L from B12/C12 into D12/E12, and then 10  $\mu$ L from D12/E12 into F12/G12 and finally taking 10  $\mu$ L from F12 into H12 (Figure 2.12).



**Figure 2-12 The preparation of the serial concentration dilutions of the standard curve of acarbose for the alpha-glucosidase assay.**

For the assay, a half-area 96-well clear plate was used. To the well positions from A2-H11, 10  $\mu$ L of sample was added, and in wells A12 to H12, 10  $\mu$ L of acarbose standard

was added. In wells A1 to D1, 10  $\mu$ L of buffer was added, and in wells E1 to H1 30  $\mu$ L of assay buffer was added. Into wells A1 to D1 and A2 to Hq12 was added 20  $\mu$ L of yeast  $\alpha$ -glucosidase enzyme. Subsequently, there was an incubation time of 10 min at 37°C, then 10  $\mu$ L of the substrate was added into all wells and again incubated for 10 min at 37°C. Wallac Victor 2 was used to read the absorbance at 405nm, and results were analysed using GraphPad Prism V 6.0.

## **2.12 Genomics**

### **2.12.1 DNA extraction**

For the extraction of the SBT531 DNA, isolate II genomic DNA kit and the protocol from Bioline, a Meridian Life Science® Company, was used. The sample was initially prepared by adding up to 5 mL of liquid broth bacterial culture to a falcon tube, then centrifuged for 5 min at 8000xg and the supernatant was then removed. On the pre-lysis step, the pellet was resuspended in 180 $\mu$ L lysis buffer GL and 25  $\mu$ L of Proteinase K solution then vigorously vortex and transferred to a sterile 1 mL Eppendorf cap. The sample was then incubated for 1-3 hours at 56°C, until completely lysed by vortexing occasionally. Lyses was completed when the sample solution clears up. After incubation, 200  $\mu$ L of Lysis buffer G3 was added, then vortexed and incubated at 70°C for 10 min to further lysis the sample. The sample was then centrifuged at 16.000 x g for 5 min, then the supernatant was transferred to a new Eppendorf cap. This was followed by the addition of 210  $\mu$ L of ethanol and the sample was again vortexed to adjust the DNA binding conditions. The sample lysate was transferred to a new eppendorf cap that was provided in the kit, and again centrifuged at 11.000 x g for 1 min then the supernatant was discarded. The silica membrane was washed with 500  $\mu$ L of Wash Buffer GW1 then with 600  $\mu$ L of Wash Buffer GW2. The sample was centrifuged two times at 11.000xg for 1 min, then the supernatant was discarded. The silica membrane was dried and centrifuged at 11.000 x g for 1 min. Meanwhile, purite water was heated to 70°C then filtered and 50 to 75  $\mu$ L of the heated purite water was added to the sample and incubated at room temperature for 3 minutes then centrifuged at 11.000 or 16.000 x g for 1 minute to elute the isolated DNA.

## 2.12.2 DNA quality/quantity controls

The DNA was quantified by reading the nucleic acid concentration with the Nanodrop 2000/2000L spectrophotometer, Thermo Scientific. Initially, 2  $\mu\text{L}$  of water was added twice to do the blank. The amount of DNA in  $\text{ng}/\mu\text{L}$  was determined from 2  $\mu\text{L}$  of the sample by measuring the absorbance at 260/280 and 260/230. The ratio of absorbance at 260/280 was used to assess the purity of DNA and RNA. The ratio must be approximately 1.8 for a “pure” DNA. The lower ratio is evidence of the presence of protein, phenol or other contaminants. The ratio of absorbance at 260/230 was used as a secondary measure of nucleic acid purity and the expected values must be in the range of 2.0-2.2. If the ratio is lower, this indicates the presence of contaminants that absorb at 230 nm. The DNA was subjected to another quality control by running an agarose gel with Fisher reagents. The gel was added to a gel tank from Jencons, run on a BioRad chamber and visualized with a visualizer from Ingenius, Syngene. The agarose gel was prepared by dissolving 0.6 g of agarose with 1  $\mu\text{L}$  of ethidium bromide in 50  $\mu\text{L}$  of 1x diluted TrisAcetateEDTA (TAE) 1x buffer, then the gel was poured into the gel electrophoresis tank and solidified. After that, the plate was filled with the TAE 1x diluted buffer and 6  $\mu\text{L}$  of HyperLadder II and was introduced to one of the wells of the plate. Isolated DNA sample was prepared by mixing 3  $\mu\text{L}$  DNA with 1  $\mu\text{L}$  dye and 2  $\mu\text{L}$   $\text{dH}_2\text{O}$ , then the mixture was introduced to the wells with a 6  $\mu\text{L}$  blank. The gel was run for 3-40 min under a 100 volts voltage. After the run was finished, the gel was examined under the 302/312 nm UV-B. The last quality control was performed with Qubit dsDNA HS assay kit from Invitrogen by Life Technologies and results were read on the Qubit 2.0 Fluorometer, Invitrogen by Life Technologies. The sample and standards were added to labelled 0.5 mL tubes. The Qubit working solution was initially prepared by diluting the Qubit dsDNA HS reagent at 1:200 in Qubit dsDNA HS Buffer and the final volume of each tube must be 200  $\mu\text{L}$ . 190  $\mu\text{L}$  of Qubit working solution was added to each of the standard tubes while 10  $\mu\text{L}$  of each Qubit standard was added to the appropriate tube, then briefly vortexed. Qubit working solution (180–199  $\mu\text{L}$ ) was incorporated to the sample tubes so that the final volume is 200  $\mu\text{L}$  prior to adding the sample (can be from 1-20  $\mu\text{L}$ ) then again briefly vortexed. The samples tubes were incubated at room temperature for 2 minutes. The standards and samples

were analysed under a Qubit 2.0 Fluorometer. The resulting concentrations were given in ng/mL, which corresponded to the concentration after the dilution step. The concentration of the sample was calculated by using the following **equation**:

$$\textit{Concentration of the sample} = QF \textit{ value} \times \frac{200}{x}$$

where QF is the value given by the Qubit 2.0 Fluorometer, x is the volume of the sample added to the assay tube in  $\mu\text{L}$ . The results were in ng/mL.

### **2.12.3 Sequencing and bioinformatic analysis**

The isolated DNA after passing the quantity and quality control described in section 2.12.2, was sent for sequencing to the MicrobesNG sequencing project, IMI-School of Biosciences at University of Birmingham, performed using next-generation sequencing with Illumina short read technology establishing a minimum coverage of 30x. The obtained sequence was analysed repeatedly by using several computational tools such as, Rapid Annotation using Subsystem Technology version 2.0 (RAST) (<http://rast.nmpdr.org/rast.cgi>) to do high-quality genome annotations. The Antibiotics and Secondary Metabolite Analysis SHellstand-alone (antiSMASH) (<https://antismash.secondarymetabolites.org>) was utilised for automatic genomic identification and analysis of BGCs. While the Basic Local Alignment Tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to compare the hits to available databases. Results are presented in Chapter 3.

## Chapter 3

*Muricauda ruestringensis* SBT531: SEARCH FOR NEW POTENTIAL  
DRUGS AGAINST METABOLIC DISEASES

### 3.1 Introduction

There is a lack of research studies and published papers regarding secondary metabolites of *Muricauda ruestringensis* and its potential bioactivity. Nevertheless, preliminary studies of the strain during the EU-FP7 SeaBioTech programme showed inhibition bioactivity against tyrosine-protein phosphatase non-receptor type 1 also known as protein-tyrosine phosphatase 1B (PTP1B) and at some extent against alpha glucosidase, which makes it extremely interesting in terms of looking into potential new drugs against metabolic diseases such as diabetes type II and obesity that currently represent a major public health issue.

Diabetes mellitus is a chronic endocrine metabolic disease that is characterized by hyperglycaemia due to insulin deficiency or insulin resistance, being currently the seventh leading cause of death worldwide (Klaman et al., 2000, Shah et al., 2018). Diabetes mellitus is divided in four types, including type 1 or insulin-dependent, type 2, gestational diabetes and drug-induced or chronic pancreatitis-induced diabetes (Jiang et al., 2012). The overwhelming majority of the diabetic population has type 2 diabetes (over 90%), which requires efficient, safe and cost-effective oral hypoglycemic therapy (Shah et al., 2018). Obesity can be responsible for important health concerns, which in addition is the major cause of type 2 diabetes, contributing to half of the detected cases (Jiang et al., 2012, Sarmiento Quintero et al., 2016).

PTP1B is an intracellular phosphorylating enzyme encoded by the PTPN1 gene in humans, considered to play an important role in the negative regulation of insulin signal transduction pathway (Lu et al., 2008, Ali et al., 2017). The increase in the bioactivity of the PTP1B enzyme was implicated in the development of breast cancer so its inhibition has been studied in the search for new anti-breast cancer agents (Elchebly et al., 1999). The selective inhibition of PTP1B has been extensively documented as a potential drug target for the treatment of type 2 diabetes and obesity. PTP1B inhibitors prevents dephosphorylation of activated insulin receptor (Sohn et al., 2013). However, to date there is still no approved selective PTP1B inhibitor, as a few compounds have been tested in clinical trials, but none of them have progressed into a clinical trial and/or patent registration (Heneberg, 2009, Zhang et al., 2016a). A

high number of PTP1B inhibitors, either synthetic or isolated natural products, have been developed and investigated for their ability to stimulate insulin signaling (Tamrakar et al., 2014). The first reported natural product with PTP1B inhibitory activity was sulfuricin 236, a bicyclic sesquiterpene with an aliphatic side-chain and a terminal sulfate group isolated from a deep-water sponge from the genus *Ircinia* in Andros, Bahamas (Cebula et al., 1997). Between 2010 to 2016, 118 marine natural products derived from sponges, algae, soft corals and fungi were described for their PTP1B inhibitory activity (Zhou et al., 2017). Majority of the reported compounds were phenolics, which contains one or more hydroxyl groups that can be classified into different groups such as, flavonoids, bromophenols, phenolic acids, or phenolics containing furan or pyran rings, coumarins, lignans and other miscellaneous phenolics. Besides phenolics and steroids, terpenes were also described, consisting of a diverse number of isoprene units classified as hemiterpenes, monoterpenes, sesquiterpenes, sesterpenes, triterpenes, tetraterpenes, polyterpenes (Jiang et al., 2012). Structure activity relationship studies proposed that fewer polar substituents on the structure skeletons can translate into increased activity, while the addition of one hydroxyl group may lead to less activity or the different position of the hydroxyl substituents may translate into different inhibitory activities as well. Moreover, further studies are required to evaluate how and where the different substituents on the different structures would determine better activity and selectivity profiles in inhibiting PTP1B (Jiang et al., 2012).

Alpha ( $\alpha$ )-glucosidase (EC 3.2.1.20,  $\alpha$ -D-glucoside glucohydrolase) is an intestinal enzyme responsible for the glycogen metabolism of higher organisms, nutrient absorption and processing of bacteria causing post-prandial hyperglycaemia while its inhibition resulted on the decrease of circulating glucose levels and rate of absorption (Kim et al., 2017b). The first identified alpha-glucosidase inhibitor was acarbose that was derived from *Actinoplanas uthaences* and was introduced to the market in the 1990s (Thareja et al., 2012). Miglitol is another example of an alpha-glucosidase inhibitor, which was obtained from various *Bacillus* and *Streptomyces* strains but is now currently produced by oxidation of 1-amino-D-sorbitol to 6-amino-L-sorbose by fermentation of *Gluconobacter oxydans* (Pei et al., 2004, Sulistiyani et al., 2016). Another example is voglibose synthesized from valiomamine, which was isolated from

a fermentation broth of *Streptomyces hydroscopicus subsp. limoneus* (Rios et al., 2015, Santos et al., 2018, Thareja et al., 2012, Pei et al., 2004, Sulistiyani et al., 2016). Along with other anti-diabetic drugs, the alpha-glucosidase inhibitors have been proven to be of a significant advantage due to their localized action with minimal absorption hence limiting the systemic side effects (Ghani, 2015). It is noticeable that in spite of the large number of publications on a wide chemical diversity of alpha-glucosidase inhibitors as new promising therapeutic agents, there is still a lack of further research on their development into new drugs (Ghani, 2015).

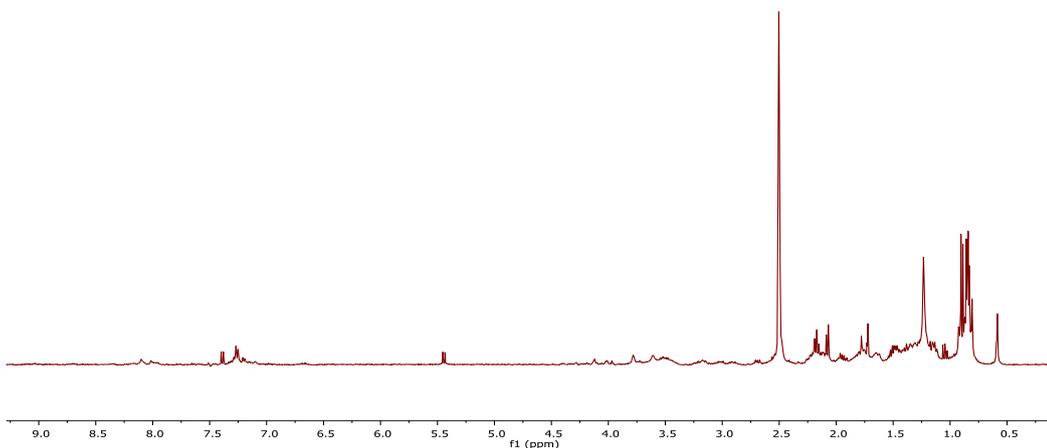
### **3.1.1 Aim(s)**

Here, *M. ruestringensis* SBT531 isolated from intertidal pools from Iceland was used for: *i*) method optimisation for increasing batch production (small- to large scale-up); *ii*) multivariate analysis of metabolic-guided screens; *iii*) bioassay-guided (assess inhibition effect against PTP1B and  $\alpha$ -glucosidase) fractionation for further characterisation (LC-MS/MS and NMR-based) of relevant compounds; *iv*) association of isolated bioactive compounds with biosynthetic gene clusters (BGC); *v*) comparison of metabolic profiles of two strains of *M. ruestringensis* (SBT531 vs. SBT587).

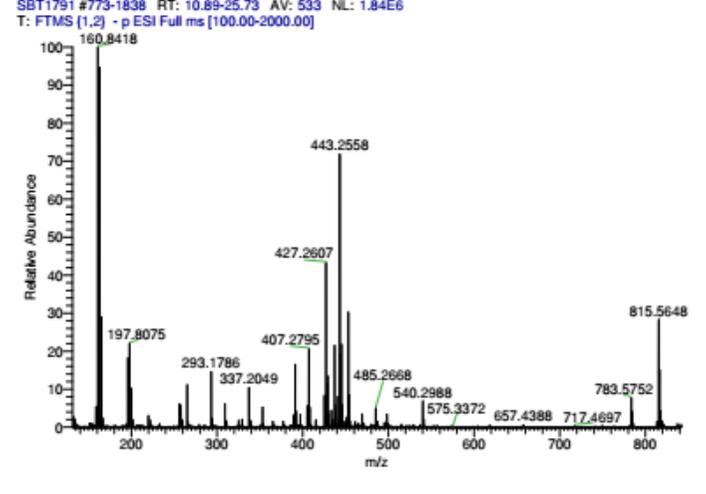
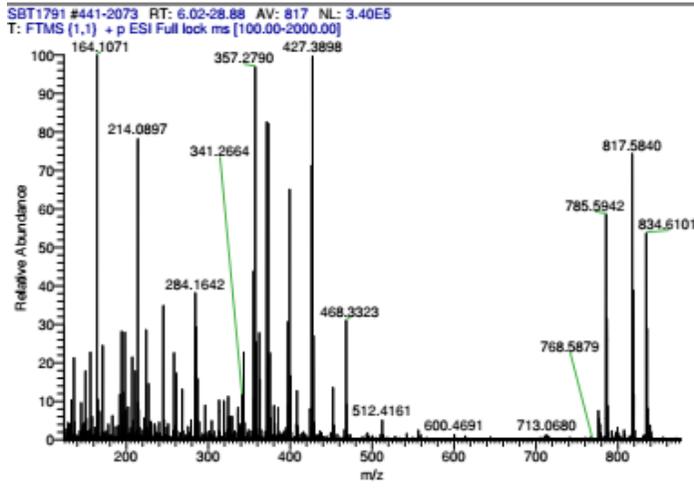
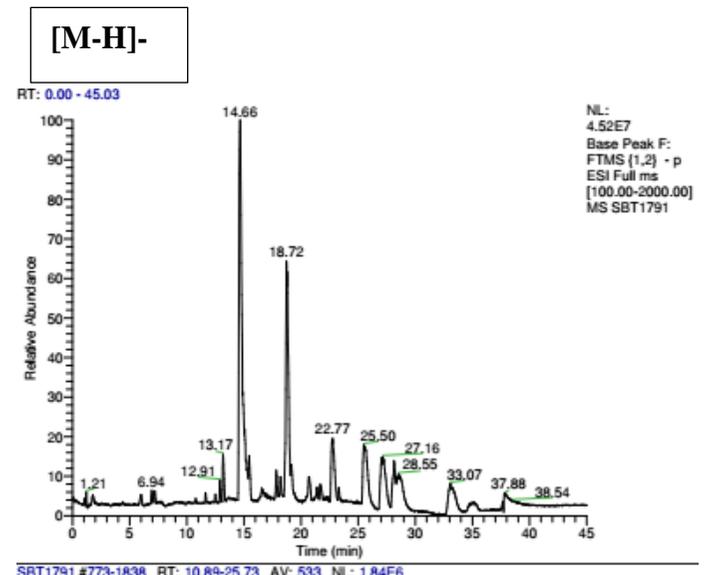
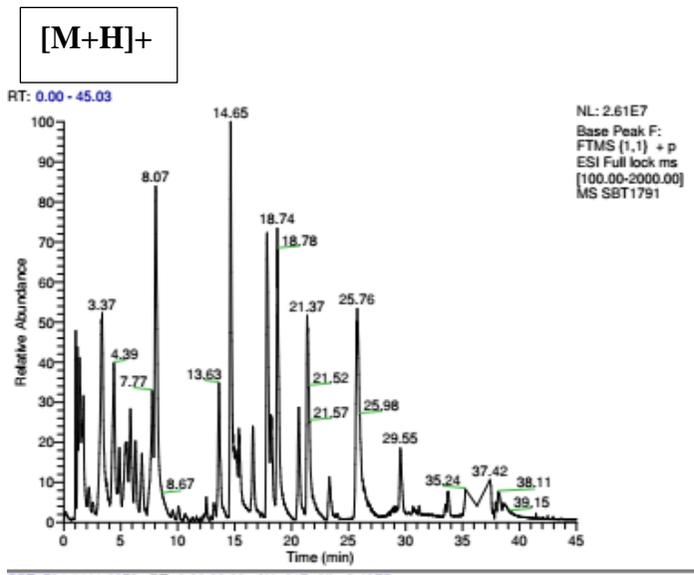
## **3.2 Small scale-up fermentation and extraction of SBT531**

The detailed methods concerning the fermentation and extraction of *M. ruestringensis* were presented on Chapter 2. The initial small-scale fermentation was performed at Matis Ltd, and Icelandic Food and Biotech R&D institute founded in 2007 while the initial bioactivity screening for alpha-glucosidase and PTP1B inhibition were performed at Strathclyde Institute for Drugs Research (SIDR) by Mrs Louise Young and Miss Grainne Abbott. Optimal growth conditions were achieved by inoculation of single colonies on a Difco Marine Broth medium and incubated for 7 days growth at

45 °C while shaken at 120 rpm. A second small scale up fermentation (4 x 250 mL) was performed in SIPBS following the fermentation conditions established in Matis Ltd. After a similar chemical profile was achieved, a medium scale up fermentation was pursued to afford extracts that were subjected to  $^1\text{H}$  NMR (Figure 3.1) and HR-LC/MS (Figure 3.2) analysis. Spectral data of  $^1\text{H}$  NMR exhibited signals ranging from 0.5 to 7.5 ppm, with most signals on the aliphatic region from 0.5 to 2.25 ppm, some signals on the olefinic region 5 to 5.5 ppm and on the aromatic region from 7 to 7.5 ppm denoting the presence of phenyl ring systems. Bioassay screening indicated that SBT531 was an alpha glucosidase inhibitor at 85% at 30  $\mu\text{g}/\text{mL}$  with an  $\text{IC}_{50}$  value of 25  $\mu\text{g}/\text{mL}$  and a PTP1B inhibitor at 100% at 30  $\mu\text{g}/\text{mL}$  with an  $\text{IC}_{50}$  value of 1.1  $\mu\text{g}/\text{mL}$ .



**Figure 3-1**  $^1\text{H}$  NMR in  $\text{DMSO-d}_6$  of an extract of SBT531 obtained by small scale-up ( $\text{DMSO-d}_6$  peak at 2.5ppm).



**Figure 3-2 LC-HRMS data of SBT531 obtained by small scale-up fermentation and extracted ion chromatogram on positive and negative ionisation mode.**

## 3.3 Medium scale-up fermentation, extraction and fractionation of SBT531

### 3.3.1 Fractionation of SBT531

Liquid cultures of SBT531 were scaled-up to 15 L (30 x 500 ml) using the optimised culture conditions by inoculating single colonies on Difco Marine Broth medium, incubated for 7 days at 45 °C and shaken at 120 rpm as described in 2.3.1. After 7 days of incubation it was extracted with ethyl acetate as described in 2.4. This solvent was selected after the optimisation of the solvent extraction at the beginning of the project on which the bioactivity was mainly found on the ethyl acetate extracts. The yield of the obtained total organic extract was 1.3086 g. The ethyl acetate extract was fractionated using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification system with a gradient elution method using DCM (A) and MeOH (B) as presented in 2.6.2.1. The mobile system is shown in detail in Table 3.1 and chromatographic conditions are listed in Table 3.2.

**Table 3-1 Mobile phase used for first fractionation of SBT531** on a Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification system.

<b>time (min)</b>	<b>% Solvent B</b>
<b>0</b>	2
<b>5</b>	2
<b>15</b>	3
<b>75</b>	3
<b>135</b>	3-50
<b>160</b>	50
<b>Total time=160</b>	

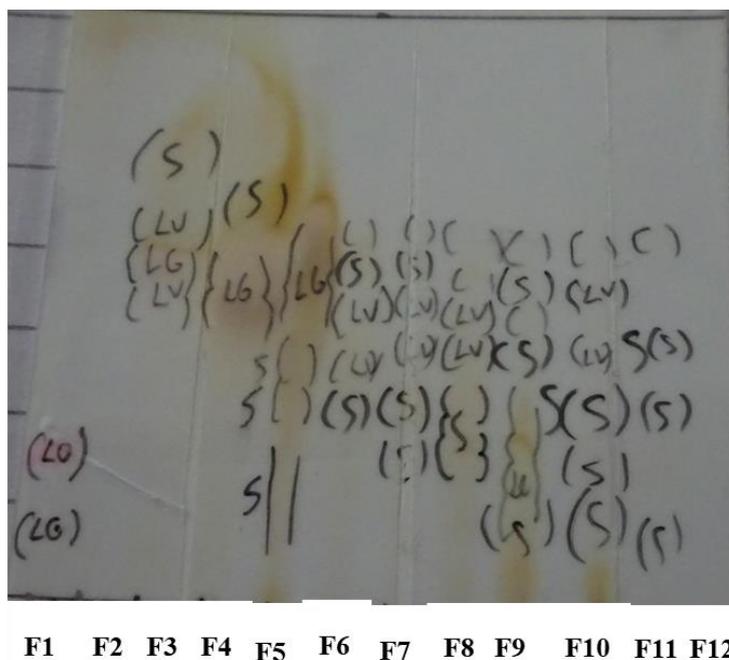
**Table 3-2 Chromatographic conditions applied to the first fractionation of SBT531**  
Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification system.

<b>Column</b>	SNAP HP-Sil 25 g
<b>Samplet</b>	3g
<b>Solvents</b>	DCM (A) and MeOH (B)
<b>Flow rate</b>	12 mL/min
<b>Detection mode</b>	UV1+UV2
<b>Start Threshold</b>	20 mAU
<b>UV1 Wavelength</b>	254 nm
<b>UV2 Wavelength</b>	280 nm

The first fractionation afforded 187 test tubes, which were pooled according to the similarities of their TLC profile that yielded 12 fractions (F1-F12) (shown on Figure 3.2). The pooled fractions were then analysed by NMR (Figure 3.3), HR-LC/MS (shown on metabolomics section 3.4) and bioassay screening as shown on Table 3.3.

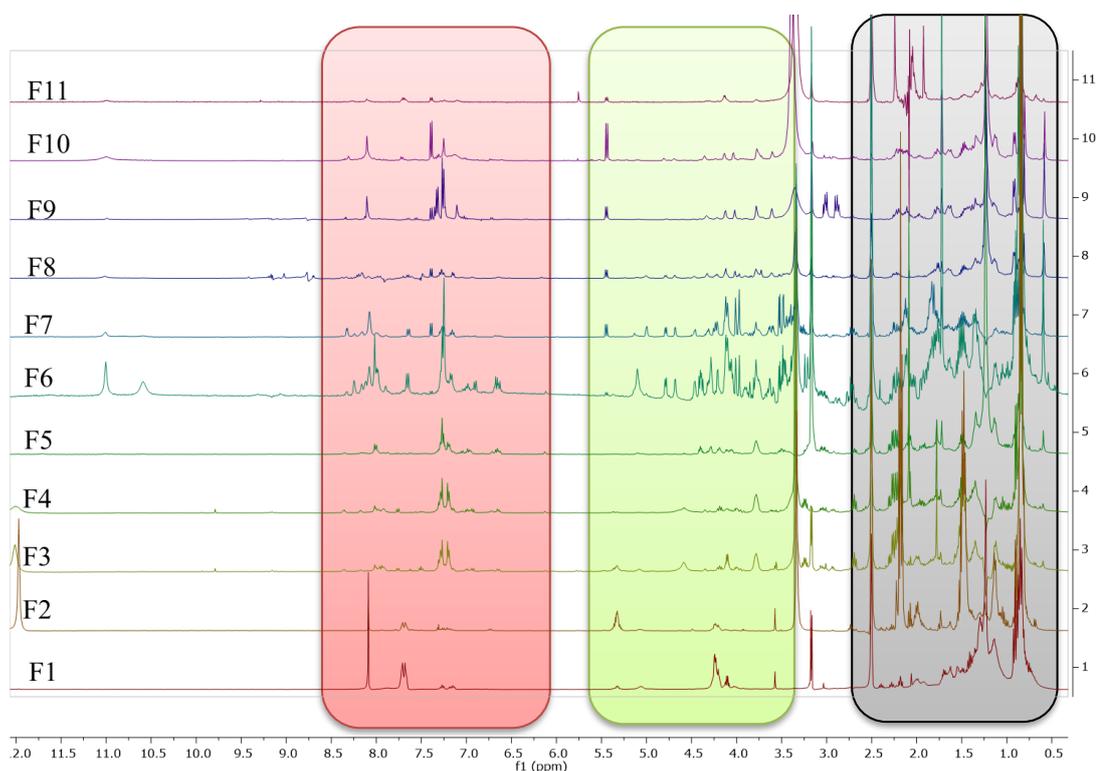
**Table 3-3 First SBT531 fractionation that afforded 12 fractions.**

<b>Fraction</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
F1	1-10	17.3
F2	11-104	198.7
F3	105	37.1
F4	106-107	118.4
F5	108-109	133
F6	110-111	51
F7	112	33.6
F8	113-120	98
F9	121-135	159.2
F10	136-184	180.7
F11	185-187	1.8
F12	Wash run	0.2



**Figure 3-3 Summary TLC plates of the first fractionation of SBT531** with mobile phase DCM:MeOH 9:1. Brown spots visualized after spraying with anisaldehyde spraying agent. S (Short), LG (Long Green), LO (Long Orange), LV (Long Violet).

The fractions were analysed by  $^1\text{H}$  NMR spectroscopy (Figure 3.4) along with HR-LC-mass spectrometry to have a general perspective of the type of compounds being produced and to prioritise which fractions would proceed for further purification work. Analysis of the  $^1\text{H}$  NMR spectral data showed majority of the signals in the aliphatic (0 – 2.5 ppm), olefinic (3.5 – 5.5 ppm) and aromatic (6 – 8.5 ppm) regions. Furthermore, signals corresponding to fatty acids were detected (1 – 1.50 ppm) with characteristic long chain methylene units that were the most relevant structural units elucidated for SBT531. Fractions F2, F3 and F4 had similar spectral features, however, F2 and F4 had a clearer spectrum in comparison to F3. Fractions F6 and F7 were combined after visualization of the NMR data and TLC summary plate, as both fractions were found to be identical.



**Figure 3-4 Pre-saturated  $^1\text{H}$  NMR (400 MHz) spectra obtained for SBT531 fractions.** Numbers on Y axis indicate the fraction number. Solvent used was  $\text{DMSO-}d_6$  (peak at 2.5 ppm). Highlighted in blue is the aliphatic region, green for the olefinic region, and in red is the aromatic region.

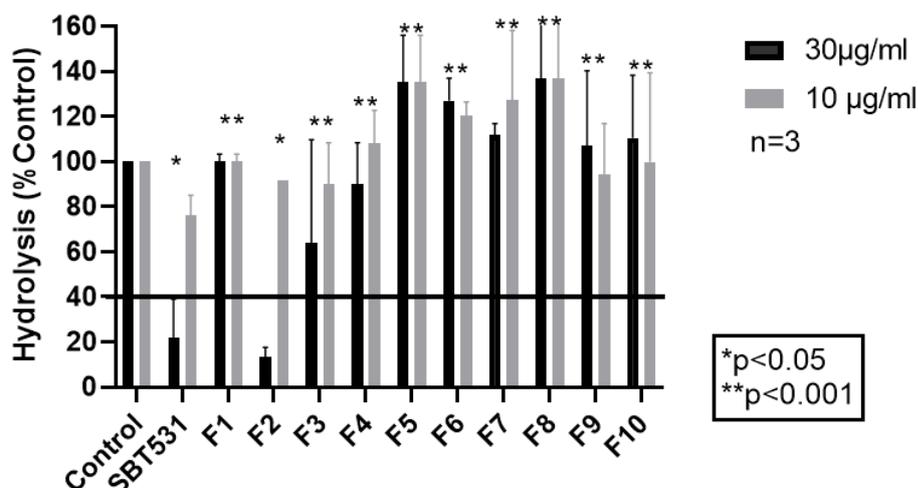
### 3.3.1.1 Biological activity of SBT531 crude extract and fractions

The crude extract and fractions were assayed for their inhibitory activity against  $\alpha$ -glucosidase and PTP1B. SBT531 crude extract exhibited its bioactivity on  $\alpha$ -glucosidase causing less than the 40% enzyme activity threshold at 30  $\mu\text{g/mL}$  on a triplicate assay ( $n=3$ ) but was found to be inactive at 10  $\mu\text{g/mL}$  (Figure 3.5). When tested on a serial dilution of the concentration level, an inhibitory constant ( $K_i$ ) of 11  $\mu\text{g/mL}$  was calculated. This curve was considered ambiguous because the fit does not follow the typical sigmoid inhibition control curve as that observed for acarbose (Figures 3.6 and 3.7). The ambiguity is important depending on the research goal. If the goal is to interpolate unknowns from a standard curve, the interpolations will still be useful but if the goal is to learn about the data by inspecting the values of the

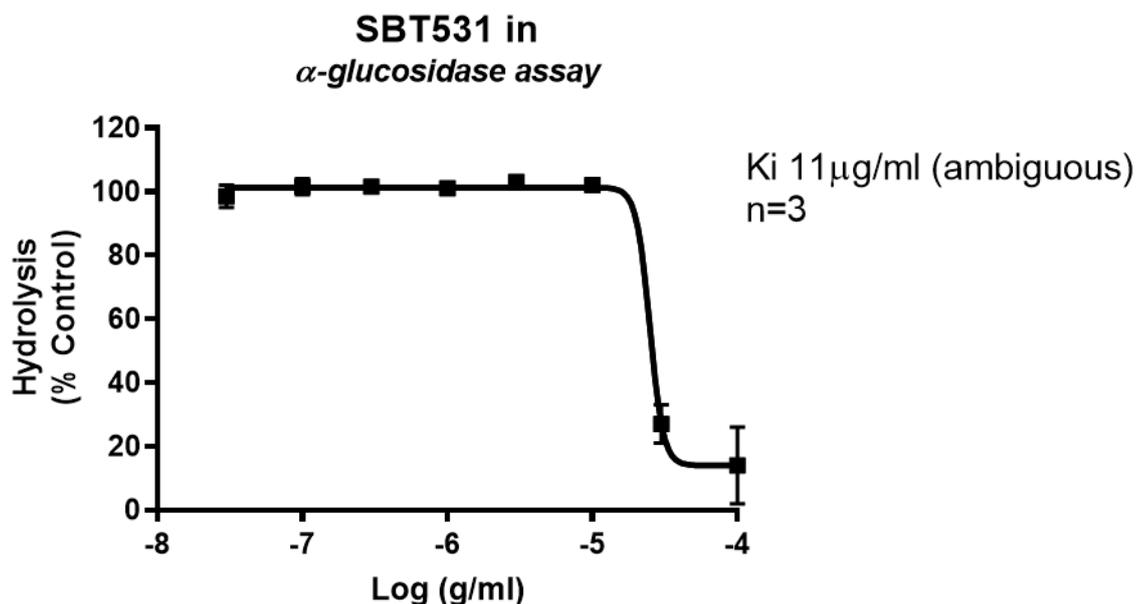
parameters, at least one value will not be reliable. The acarbose inhibition control curve for the alpha - glucosidase assay done in triplicate gave a  $K_i$  value of 10  $\mu\text{g/mL}$  (Figure 3.7). SBT531 fractions were also tested and only Fraction 2 showed inhibitory activity at 30  $\mu\text{g/mL}$  but not at 10  $\mu\text{g/mL}$  as with the crude extract (Figure 3.5). Then when Fraction 2 was assayed at different concentrations by serial dilution, it was not possible to generate an inhibition curve to generate the subsequent  $K_i$  value, so it was not considered a true valid inhibition result (Figure 3.8). The  $K_i$  value characterises the concentration required to form an enzyme-inhibitor complex, where a lower  $K_i$  value will indicate more effective inhibition against alpha glucosidase and/or PTP1B enzymes in the development of preventive and therapeutic agents (Ali et al., 2017).

Chromatographic fractionation is a purification technique applied to isolate and purify the targeted bioactive metabolites from the crude extract. The crude extract itself holds unique chemical properties that is responsible for the bioactivity. The extract comprised of a diverse metabolite profile, which can act synergistically by the interaction of the metabolites producing a combined bioactive effect greater than the sum of their separate effects. Nevertheless, when the crude extract is subjected to fractionation and purification, it can miss the previous reported bioactivity. Loss or change in bioactivity can also be due to variations in concentrations of the bioactive metabolites in the different fractions. The concentration of the bioactive metabolite can decrease after fractionation rendering the fraction inactive when bioactivity is concentration-dependent and not only due to the mere presence or absence of the metabolite. These factors all together can explain the reasons why inhibition activity could disappear after fractionation and was only present on the crude extract.

**SBT531 crude extract and fractions  
in  $\alpha$ -glucosidase assay**



**Figure 3-5** Inhibition effect of SBT531 crude extract and fractions F1 to F10 against  $\alpha$ -glucosidase at 30 and 10  $\mu$ g/mL (n=3). Error bars represent the standard deviations (SD) of three replicates (n=3) and values were significant at \* $p$ <0.05 and \*\* $p$ <0.001 by 2way ANOVA multiple comparisons with the control.



**Figure 3-6** SBT531 inhibition curve against  $\alpha$ -glucosidase at serial concentrations (n=3).

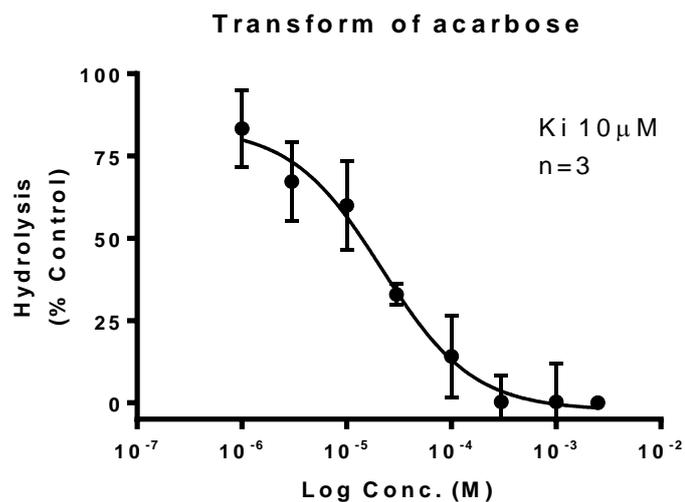


Figure 3-7 Acarbose inhibitor control curve on  $\alpha$ -glucosidase assay (n=3).

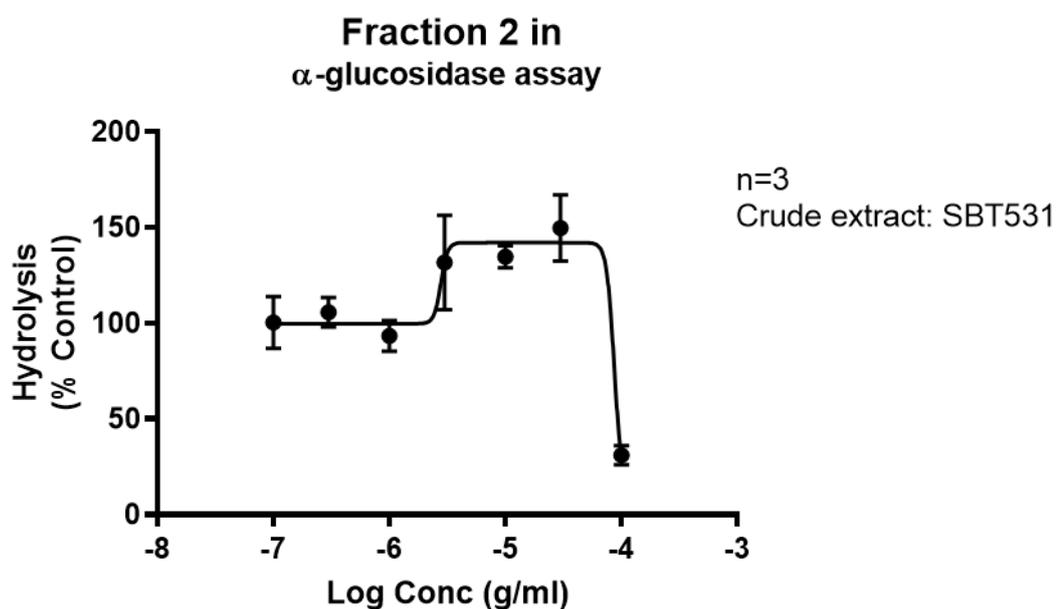
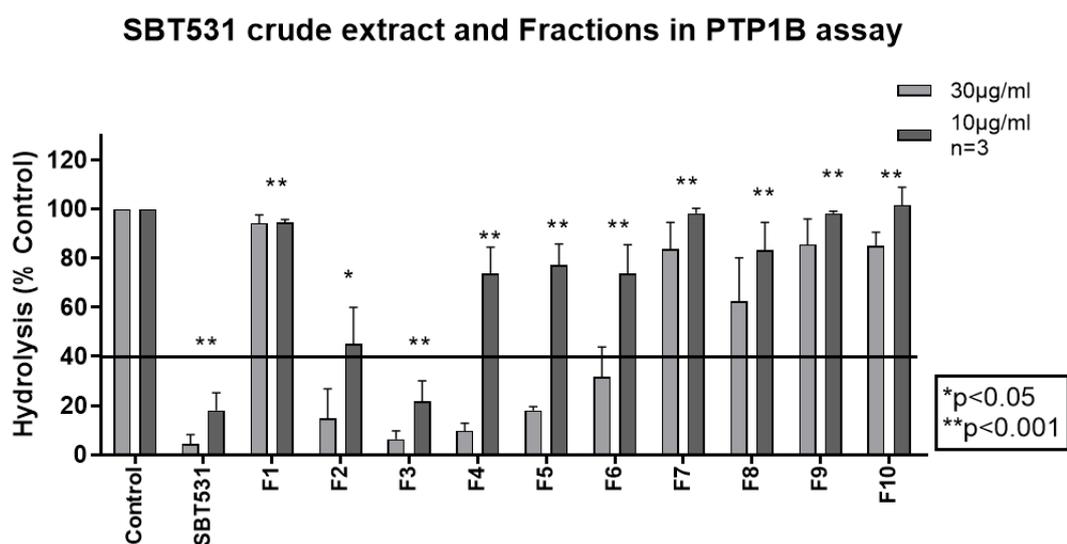


Figure 3-8 SBT531 Fraction 2 inhibition curve against  $\alpha$ -glucosidase at serial concentrations (n=3).

SBT531 crude extract showed inhibition effect against PTP1B at concentrations of 30 and 10  $\mu$ g/mL (Figure 3.9). Serial dilution of its concentrations afforded a  $K_i$  at 1.1

$\mu\text{g/mL}$  (Figure 3.10). SBT531 fractions (F1 to F10) were also assayed for inhibition effect against PTP1B at concentrations of 30 and 10  $\mu\text{g/mL}$  in a triplicate assay ( $n=3$ ) (Figure 3.9). Inhibitions for Fractions 2 and 3 gave  $K_i$  values of 3.0 and 3.1  $\mu\text{g/mL}$ , respectively (Figures 3.12 and 3.13). The PTP1B inhibitor IV curve gave a  $K_i$  value of 0.7  $\mu\text{g/mL}$  performed in triplicates ( $n=3$ ) (Figure 3.11) showing the validity of the model.



**Figure 3-9 Inhibition effect of SBT531 crude extract and fractions F1 to F10 against PTP1B at 30 and 10  $\mu\text{g/mL}$  ( $n=3$ ).** Error bars represent the standard deviations (SD) of three replicates ( $n=3$ ) and values were significant at  $*p<0.05$  and  $**p<0.001$  by 2way ANOVA multiple comparisons with the control.

### SBT531 in the PTP1B Assay

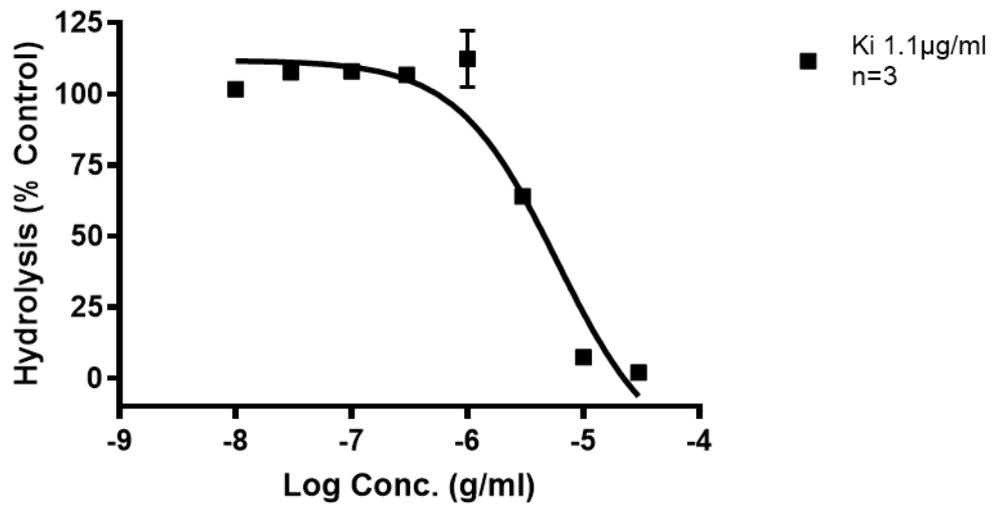


Figure 3-10 SBT531 inhibition curve against PTP1B at serial concentrations (n=3).

### PTP1B Inhibitor IV in the PTP1B Assay

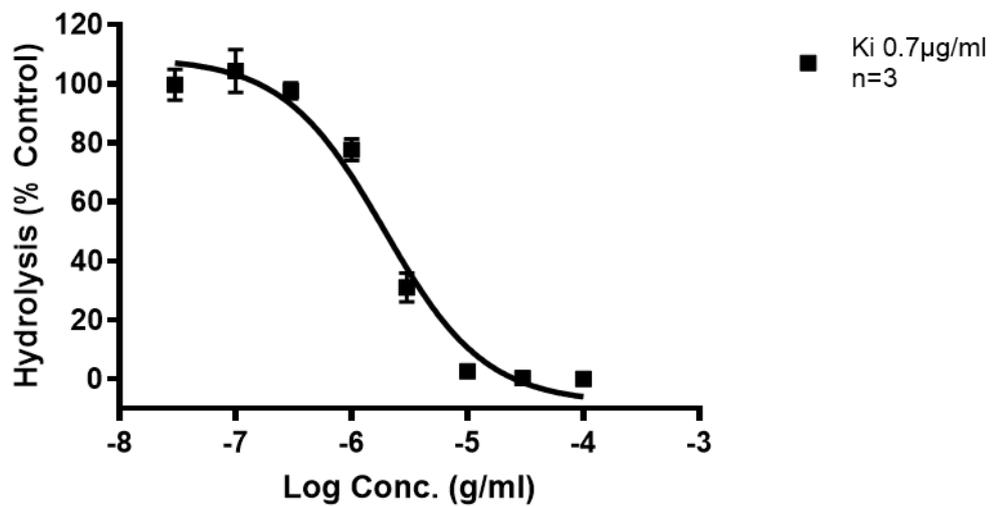


Figure 3-11 PTP1B inhibitor IV curve in the PTP1B assay.

### Fraction 2 in the PTP1B Assay

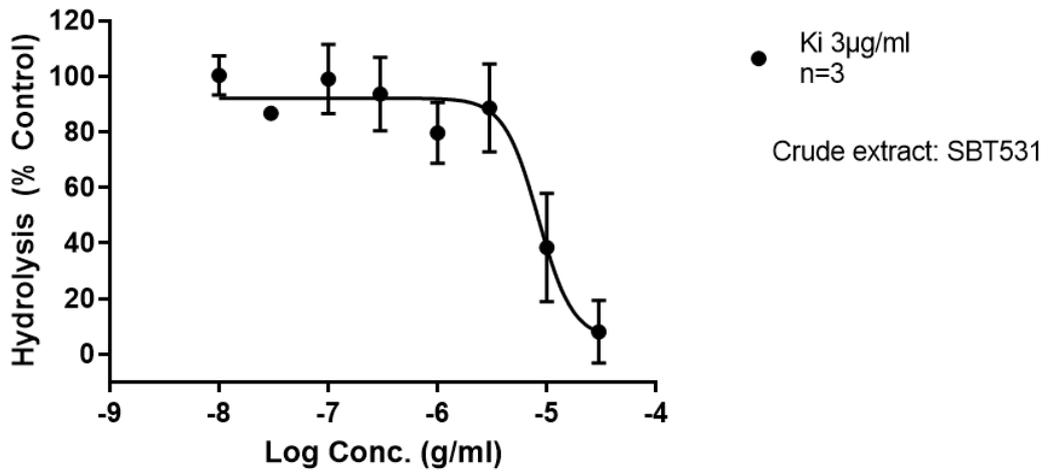


Figure 3-12 SBT531 Fraction 2 inhibition curve against PTP1B at serial concentrations (n=3).

### Fraction 3 in the PTP1B Assay

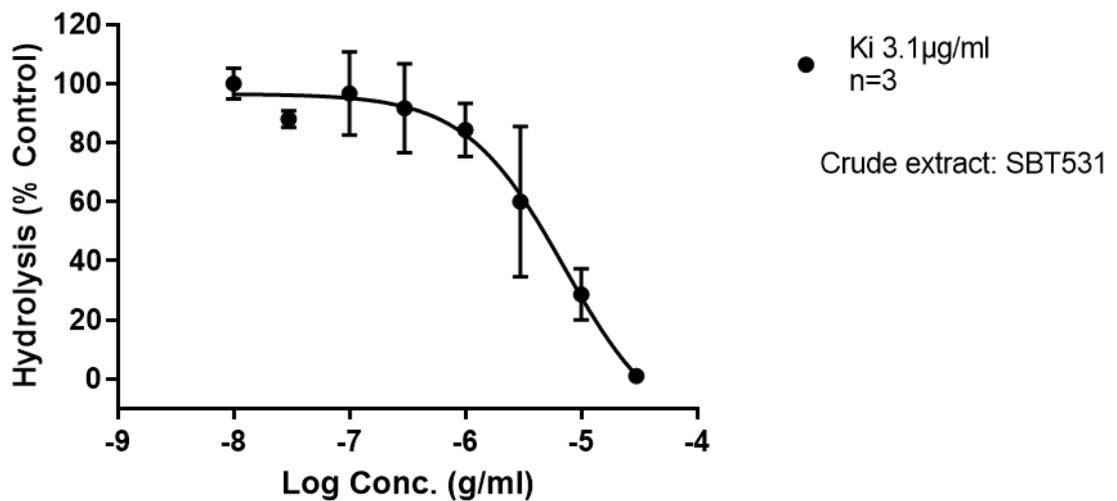


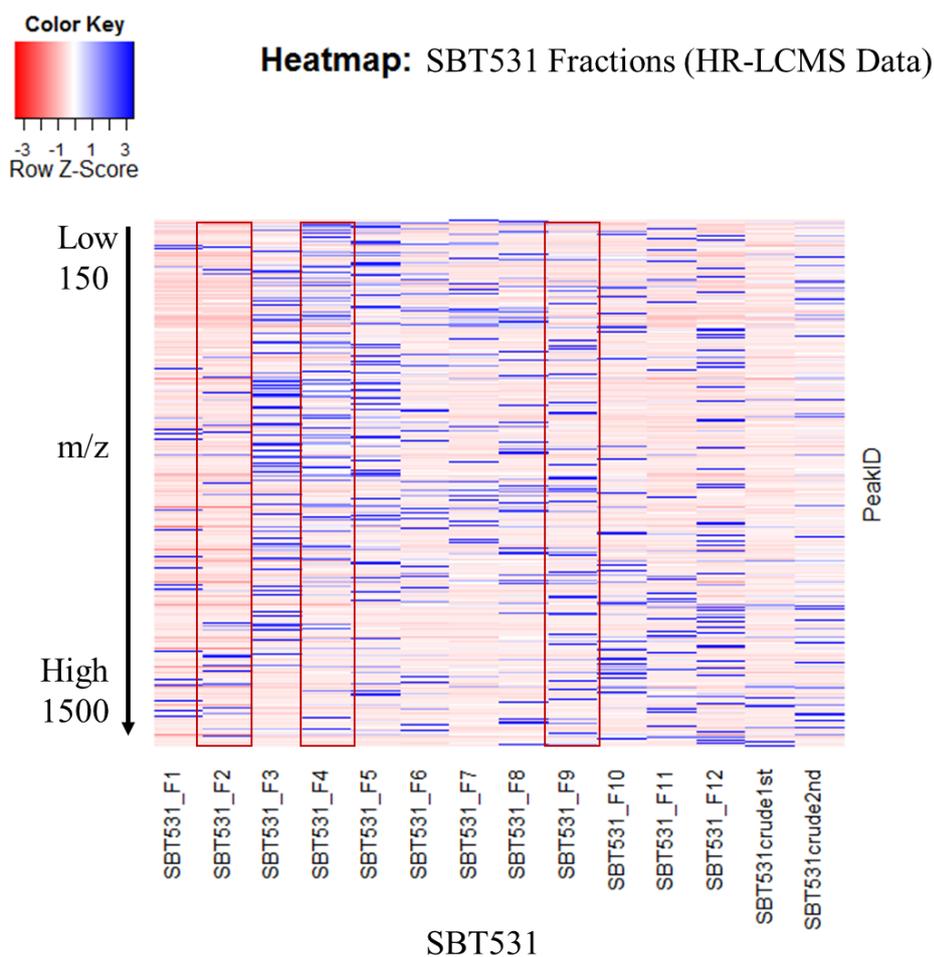
Figure 3-13 SBT531 Fraction 3 inhibition curve against PTP1B at serial concentrations (n=3).

## 3.4 Metabolomic-guided screening of SBT531

### 3.4.1 Metabolomic profiling of fractions of the crude extract of SBT531

#### 3.4.1.1 HR-LC/MS Data

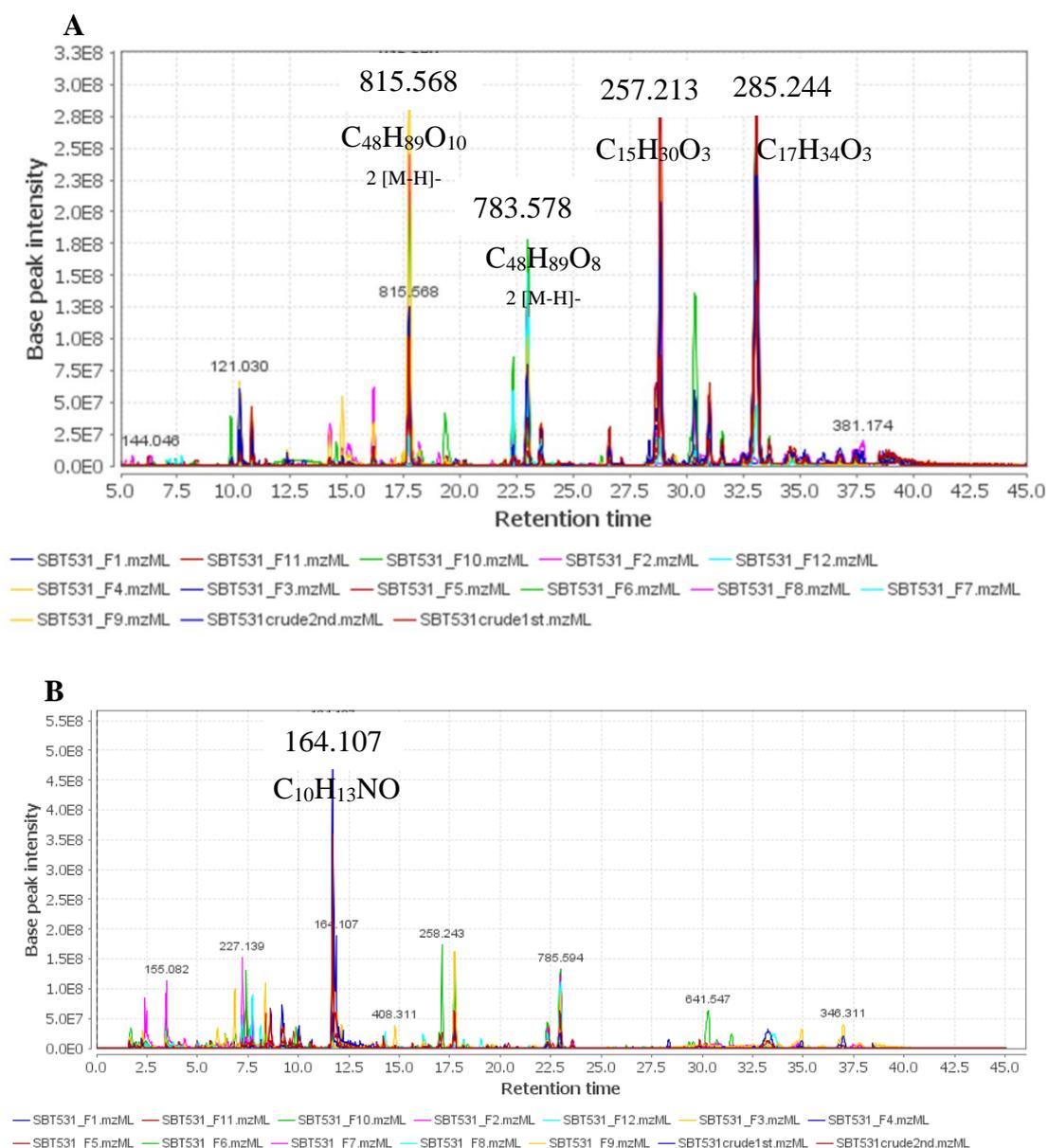
After the HR-LC/MS data was obtained on the crude extract, fractions and sub-fractions, the data was processed as described in Chapter 2, section 2.9.1. The mass spectral data was processed, dereplicated, subjected to multivariate analysis. Statistical plots were generated that included PCA and OPLS-DA scores and loadings scatter plots. S-plots were also generated from the OPLS-DA loadings scatter plot, which were validated by a permutation test. A heatmap based on the HR-LCMS data was put together that allowed the visualization of the chemical diversity of respective fractions. The intensity of the blue bands specified the increase in concentrations of a metabolite in the sample while the absence or decrease in the concentration of a metabolite was indicated by red bands. The heatmap on Figure 3.14 shows that F3, F4, F5 and F12 produced more metabolites as evidenced by the greater number and intensity of the blue bands. While F3, F4, and F5 gave rise to a larger quantity of low-molecular weight metabolites, F12 had an increase in density of higher molecular weight compounds. Other fractions afforded certain specific metabolites as shown by the red bands while the emergence of new blue bands were observed.



**Figure 3-14 Heatmap based on HR-LC/MS data displaying distinct metabolic profiles of 12 fractions and 2 crude extracts.** Boxed in red are fractions from where compounds were later isolated.

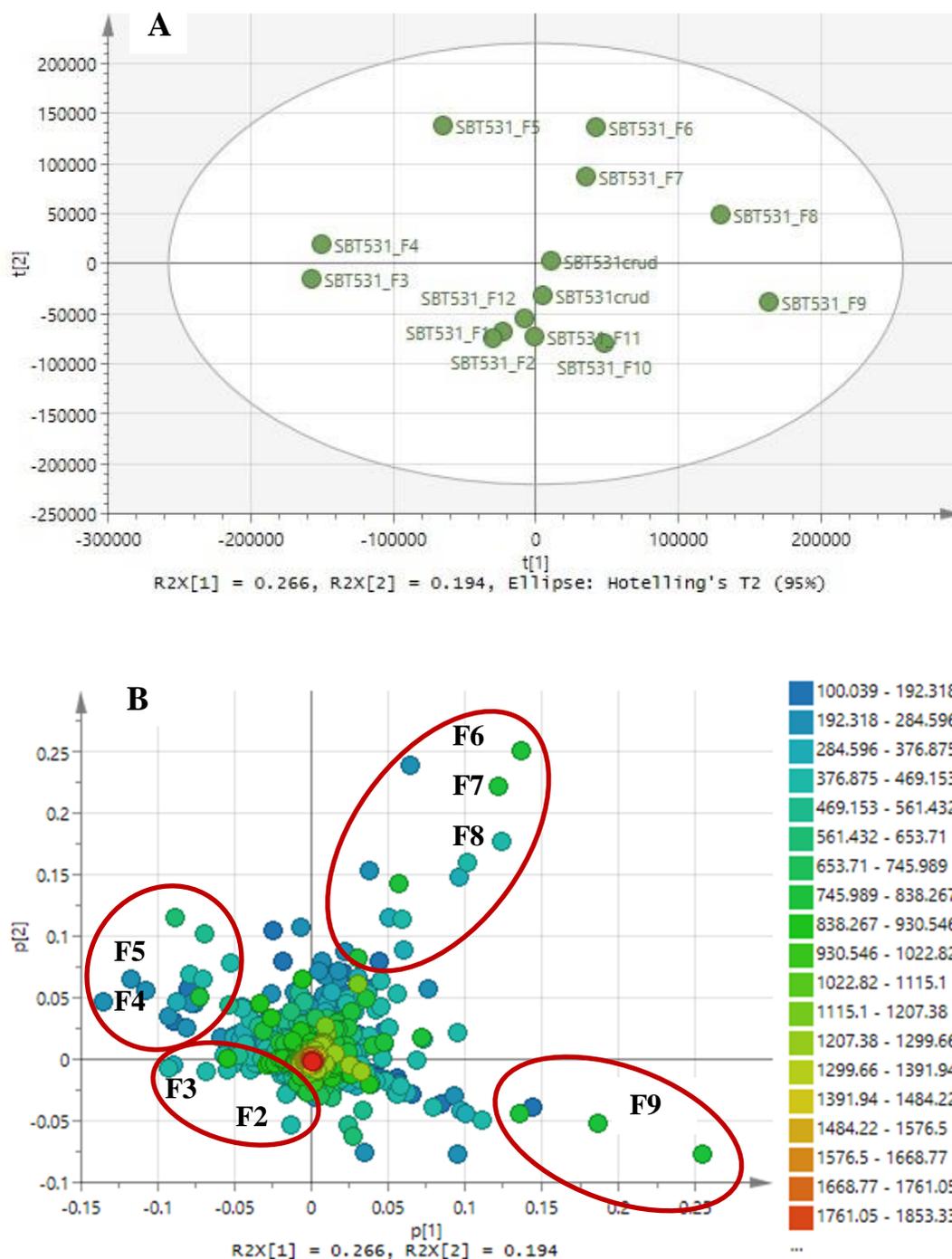
The HR-LC/MS data was processed using MZMine from which base peak chromatograms, base peak intensity and scatter plots were generated that allowed additional comprehensive profiling of the chemical composition of each fraction in comparison to the crude extract. Furthermore, this also later permitted correlating the mass spectral data with those of the NMR results. Visualization of the base peak plot chromatograms on negative and positive ionization mode shown on Figure 3.15A and B illustrated that the main compounds being produced were hydroxylated fatty acids,

cholic acid derivatives and *N*-phenethylacetamide that was confirmed with further 2D NMR experiments.



**Figure 3-15 Base peak plot chromatogram (HR-LC/MS data) of SBT531 crude extract and fractions in negative (A) and positive (B) ionization mode created on MzMine 2.10 modified version.**

PCA was performed on the HR-LC/MS data of 12 isolates and two crude extracts. PCA plots were generated from four components, a  $R^2X$  (coefficient of determination) value of 0.621 was achieved using Pareto scaling, indicating a good model for a metabolomics data set to fit ( $R^2X > 0.5$ ) and  $Q^2$  of 0.207 which specifies the predictive ability of the model. The difference between  $R^2X$  and  $Q^2$  was 0.414, which should be less than 0.3 indicating a slightly over fitted dataset. The presence of outliers and diversity on chemical profiles of fractions could explain the larger difference between  $R^2X$  and  $Q^2$  or low  $Q^2$  value. PCA demonstrated patterns, trends and enabled the visualization of any outliers in the dataset. The fractions were clustering in different quadrants according with their chemical similarities. Fractions F4 and F5 were located on the upper left quadrant and F3 on the bottom left quadrant. Fractions F6, F7, and F8 were located on the upper right quadrant while F9 on the bottom right quadrant (Figure 3.16A). The loadings plot represented the molecular ion masses of the detected secondary metabolites of samples found on the corresponding quadrant position on the scores plot (Figure 3.16B). By comparing the results with the bioassay screening described in section 3.3.1.1, F2 to F6, showed inhibition effect against PTP1B, however, F2 and F3 were the most relevant fractions showing inhibition on both concentrations at 30 and 10  $\mu\text{g/mL}$ , which were located on the same bottom left quadrant, displaying similarities of their chemical profiles with metabolites sharing similar  $m/z$  between 284.596 to 376.875.

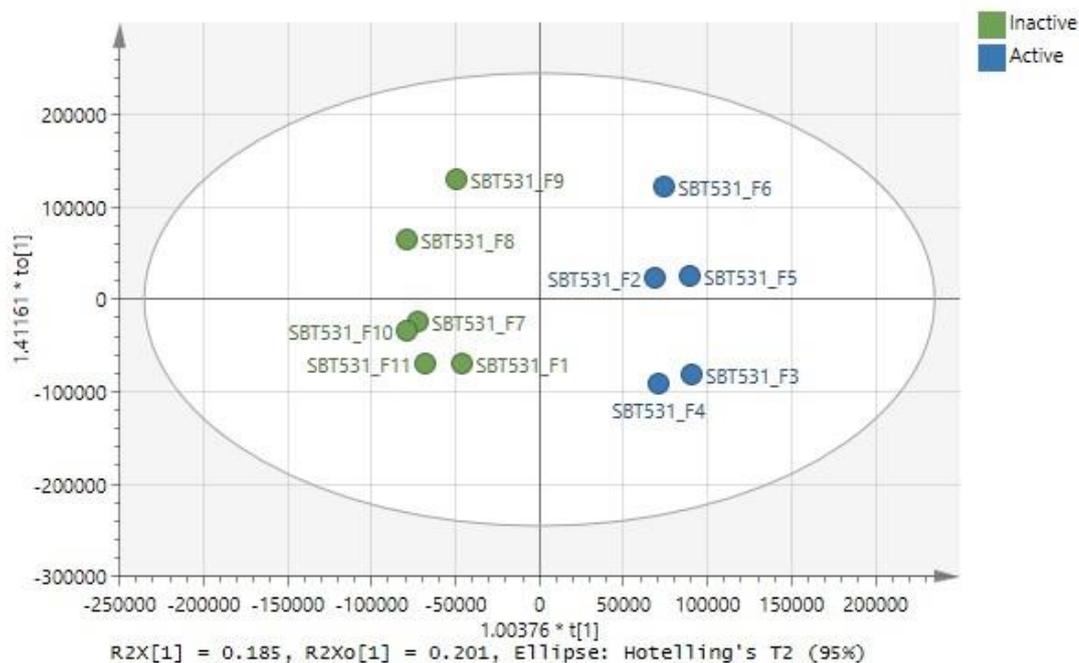


**Figure 3-16 Principal component analysis (PCA) plot of SBT531. (A) PCA-X scores plot of SBT531 fractions. (B) Loadings plot acquired from the PCA-X model.**

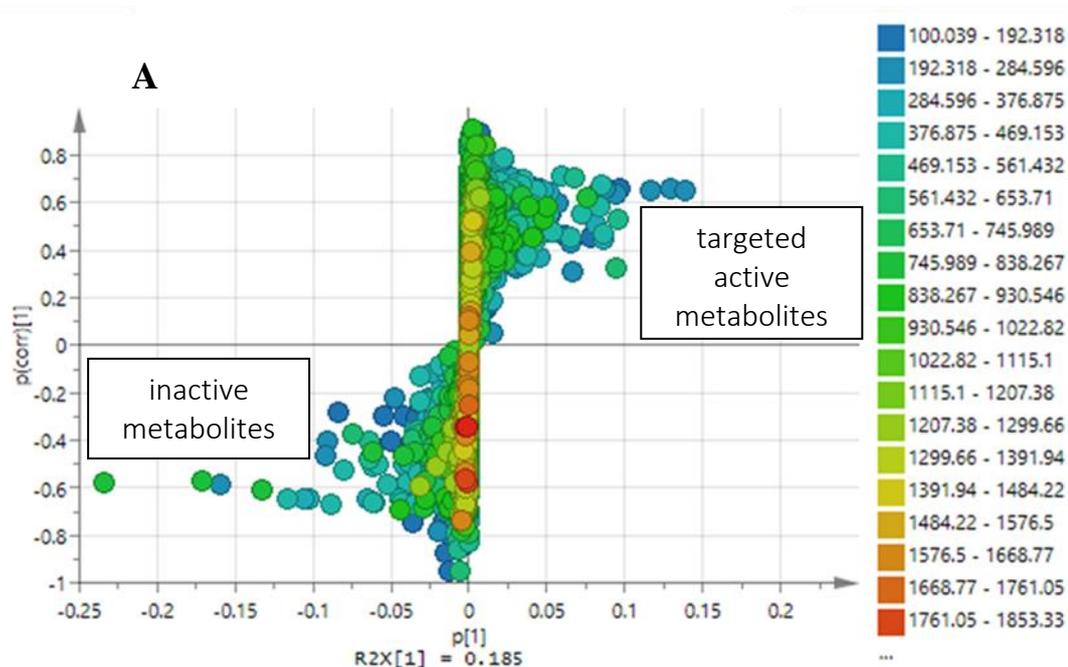
To be able to predict and understand the type of chemistry that would be responsible for the inhibition bioactivity of the SBT531 fractions, a supervised multivariate

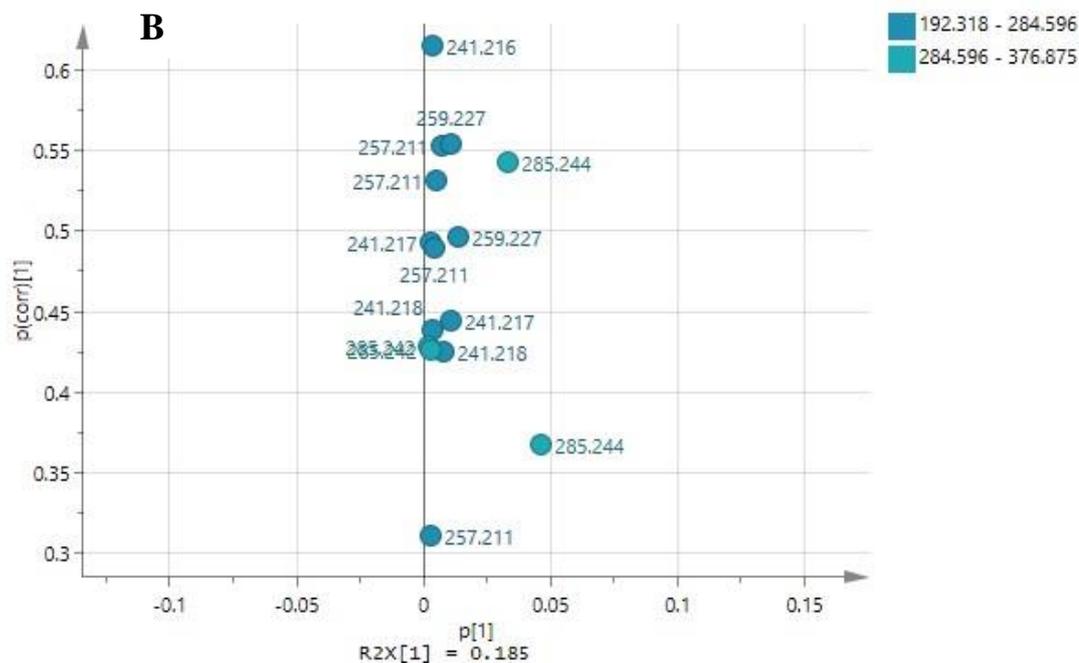
analysis was accomplished by subjecting the HR-LC/MS data sets to OPLS-DA (Orthogonal Partial least squares - Discriminant Analysis or also known as Orthogonal Projections to Latent Structures). Therefore, this approach was used because separate groups of observations are defined by the user, according to the aim of the study and are clustered in the scores plot, whilst the variables in the loading plots are grouped accordingly so they are responsible for the observations separation in the scores plot. The sample extracts were grouped into two classes: active vs. inactive (Y variables) and by using a loadings S-plot, respective metabolites and functional groups (X variables) could be pinpointed to be responsible for the bioactivity. This statistical model would assist in targeting the bioactive natural products for further isolation work. It is possible to realize the active vs inactive fractions in Figure 3.17, while the individual metabolites belonging to each group of are shown in the S-plot (Figure 3.18A). An expansion showing the position of the isolated metabolites on the S-plot showed the target putative compounds for the active fractions in Figure 3.18B, aseanostatin P6 (molecular formula:  $C_{15}H_{30}O_2$ ,  $m/z$ : 241.2177  $[M-H]^-$ ), two fatty acid hydroxylated derivatives, 2-hydroxy-14-methylhexadecanoic acid (molecular formula:  $C_{17}H_{34}O_3$ ,  $m/z$ : 285.2438  $[M-H]^-$ ), and 3-hydroxy-13-methyltetradecanoic acid (molecular formula:  $C_{15}H_{30}O_3$ ,  $m/z$ : 257.2125  $[M-H]^-$ ). This was preceded by a permutation test (Figure 3.19) to authenticate the validity of the model, the  $R^2Y$  was 0.974 and  $Q^2$  was 0.385, while the  $Q^2Y$  intercept was -0.27. These values specified good fitting and good prediction as the  $R^2Y$  and  $Q^2$  values were close to 1 and 0.5, respectively, while the  $Q^2Y$  intercept was -0.27, which is less than 0 indicating the validity of the permutation test. Furthermore, the difference between  $Q^2$  and  $R^2Y$  was 0.589 which is more than 0.3, indicating overfitting (the ideal difference is less than 0.3). The  $R^2X(1)$  was 0.185 which means that the variation between groups is 18.5% and  $R^2X0(1)$  was 0.201 with a variation within groups of 20.1% showing a high diversity in chemical profiles between fractions, as well the existence of more unique fractions that fail to cluster. The targeted end point metabolites from the S-plot were dereplicated and was cross-matched with the Antimarin and DNP databases to provide structural details about the targeted metabolites as listed in Table 3.4. Moreover, it was not possible to isolate all the compounds shown on the dereplication table. Although the peak intensity was high (E+7 to E+9), the ionisation capability of the respective

metabolites limits the isolation potential and may not reflect the accurate concentration of the metabolites that could be lower than expected to be physically isolated.

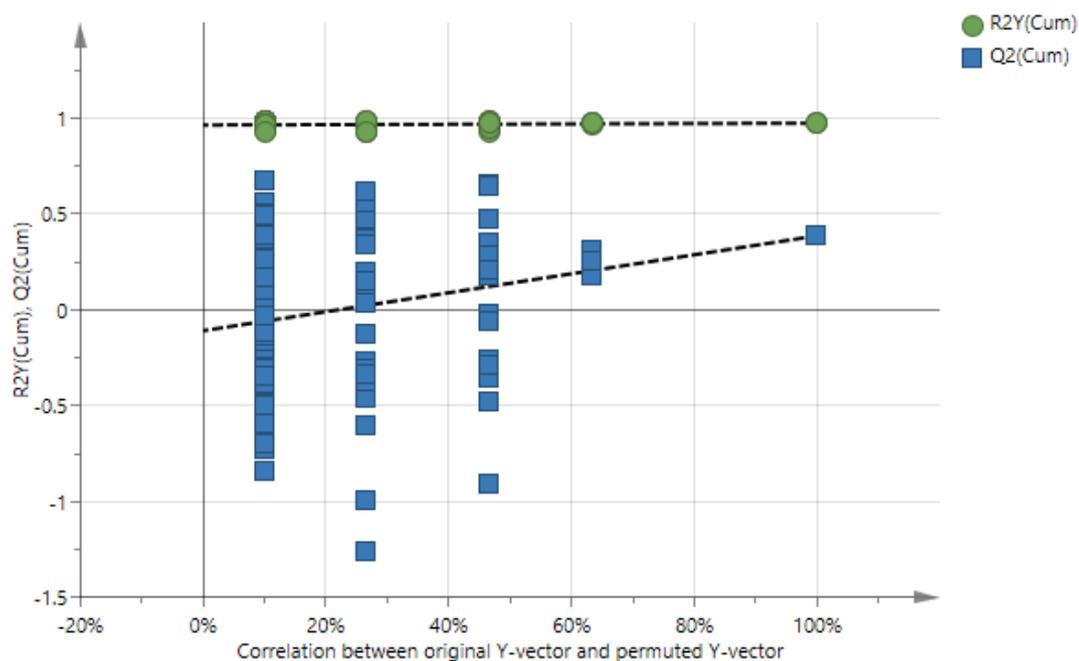


**Figure 3-17 OPLS-DA of SBT531 fractions according to inhibition activity against PTPB1B.**





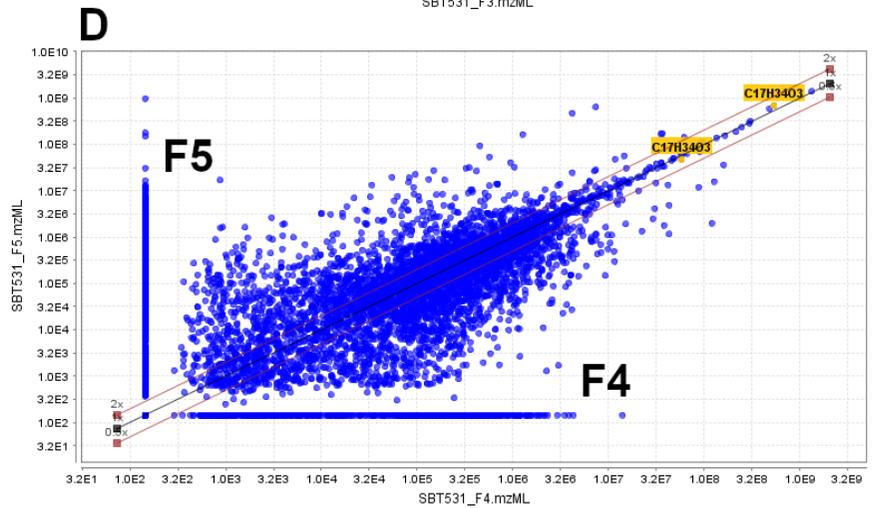
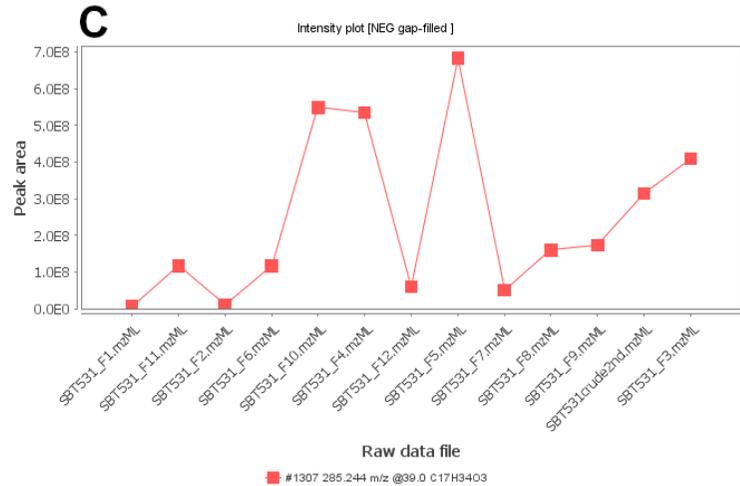
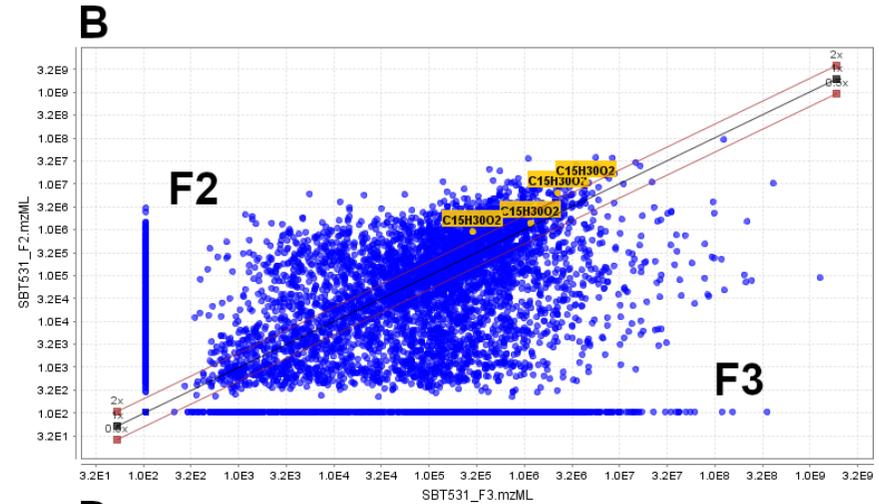
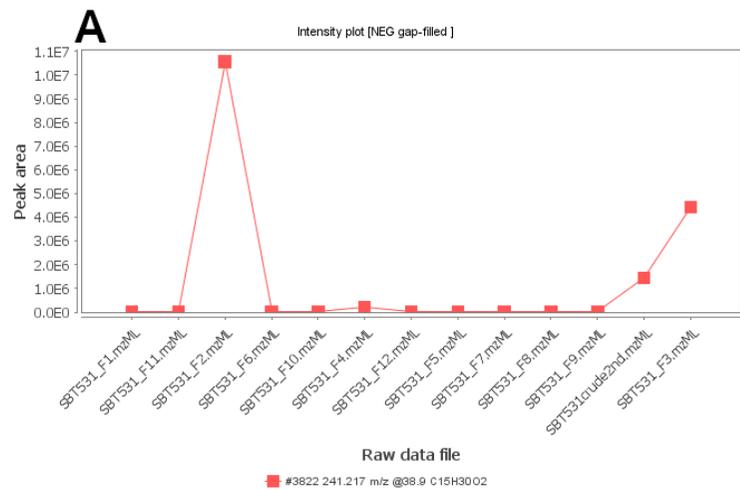
**Figure 3-18 OPLS-DA of SBT531 fractions.** (A) S-plot for SBT531 fractions acquired from an OPLS-DA model (Figure 3.24) for their inhibition activity against PTPB1B. (B) Position of the isolated active metabolites on the S-plot (upper right quadrant) indicating the target putative compounds for the active fractions labelled with their *m/z*.



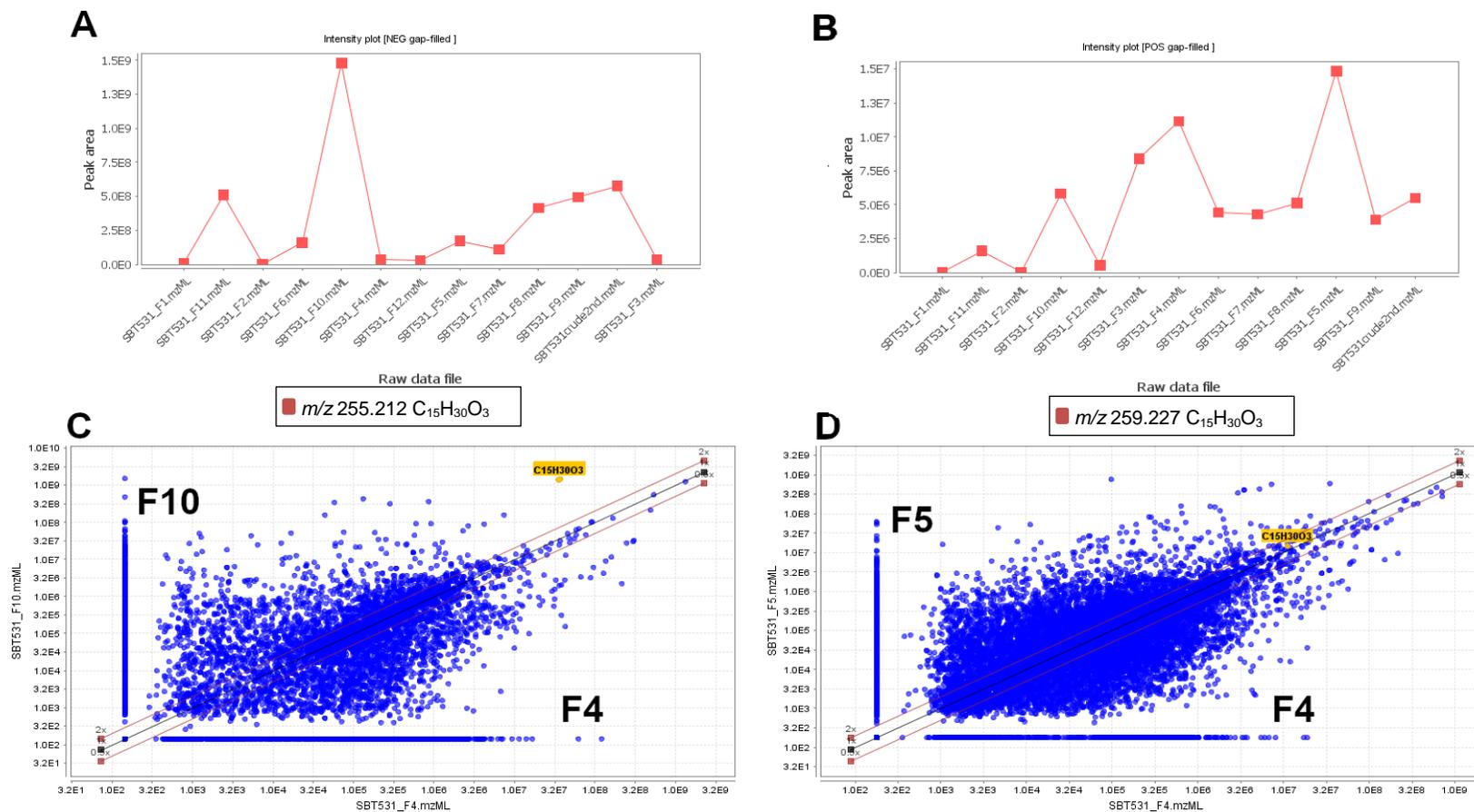
**Figure 3-19 Permutation tests (100 permutations) for SBT531 fractions for the OPLS-DA model** of their inhibition activity against PTP1B, where Q2Y is less than zero.

As shown by the intensities of the base peaks in the negative ionisation mode (Figure **3.20A**) of aseanostatin P6 ( $C_{15}H_{30}O_2$ ,  $m/z$  241.217), this metabolite was mainly found in F2 at highest concentration, then followed by F3. Both fractions F2 and F3 showed inhibition activity. However, the compound was isolated from F2, as it was a purer fraction than F3, which contained a more complex mixture of compounds. Aseanostatin P6 was only detected in the negative mode because of the compound's low ionisation capability in electrospray (ESI)-MS. The scatter plot (Figure **3.20B**) compared the metabolites found in F2 and F3 that exhibited the richness of both fractions in terms of chemical composition. The presence of  $C_{15}H_{30}O_2$  was found in both F2 and F3 as the ions representing the compound were positioned on the diagonal indicating the metabolites occurring in both fractions. The intensity base peak plot in the negative ionisation mode (Figure **3.20C**) for 2-hydroxy-14-methylhexadecanoic acid ( $C_{17}H_{34}O_3$ ,  $m/z$  285.244), showed that all fractions and crude extract except F1, F2 and F7 were producing this compound. A higher concentration of the compound was visible in F5, followed by F4 and F10. Comparison of F4 and F5, in terms of their chemical composition and diversity, was shown by a scatter plot (Figure **3.20D**) indicating that both fractions are rich and diverse in their metabolite composition. The metabolite  $C_{17}H_{34}O_3$  was found to be present in both fractions as the molecular ion was positioned on the diagonal. The compound 2-hydroxy-14-methylhexadecanoic acid was later isolated from F4 due to its higher purity in the fraction. For 3-hydroxy-13-methyltetradecanoic acid ( $C_{15}H_{30}O_3$ , at  $m/z$  257.212) the intensity base peak plot in the negative (Figure **3.21A**) and positive (Figure **3.21B**) ionisation mode were also construed. The difference in the higher ionisation capability of 3-hydroxy-13-methyltetradecanoic acid to aseanostatin P6, relies on the extra hydroxyl group in the latter. In the negative mode, the compound  $C_{15}H_{30}O_3$  was relatively highest in F10 and has a lower concentration in F4. Moreover, fractions F5, F6, F7, F8, F9 and F11 also yielded  $C_{15}H_{30}O_3$ . Comparing F4 and F10 on the scatter plot (Figure **3.21C**), confirmed the higher concentration of  $C_{15}H_{30}O_3$  in F10 as it was shown above the diagonal representing F10. On the positive mode, all fractions indicated to produce the metabolite except F1, F2 and F12 with the highest intensity peaks on F4 and F5, which was also indicated on the scatter plot (Figure **3.21D**) to be present in both fractions. The concentrations of  $C_{15}H_{30}O_3$  in all fractions that yielded the metabolite

were in general of higher intensity on the negative mode. The negative ionisation mode is more appropriate for compounds with carboxyl or hydroxyl groups, whereas the positive ionisation mode is more suitable for compounds with amino, amides or carbonyls substituents, which have stronger proton affinity (Chai et al., 2015, Dunn et al., 2013, Creydt and Fischer, 2017). Furthermore, F4, from which  $C_{15}H_{30}O_3$  was also isolated, provided elevated purity and afforded higher intensity peaks both on the negative and positive mode. From all fractions where  $C_{15}H_{30}O_3$  was detected, not all were considered active. The bioactivity can be concentration-dependent, albeit the compound was present on the HR-LCMS TIC, but when comparing the NMR spectra, the compound was not detectable in all fractions. The retention times of the target metabolite also varied for the different fractions, which could signify the presence of isomers that could account for the differences in bioactivity.



**Figure 3-20 Intensity base peak plot of LC-HRMS data in negative ionisation mode of  $C_{15}H_{30}O_2$  (A) and  $C_{17}H_{34}O_3$  (C) in SBT531 crude extract and fractions and scatter plot comparisons of F3 (x-axis) and F2 (y-axis) negative ionisation mode regarding the presence of  $C_{15}H_{30}O_2$  (B) and F4 (x-axis) with F5 (y-axis) of  $C_{17}H_{34}O_3$  (D).**



**Figure 3-21 Intensity base peak plot of HR-LC/MS data in negative (A) and positive (B) ionisation mode of  $C_{15}H_{30}O_3$  in SBT531 crude extract and fractions and scatter plot comparing of F4 (x-axis) and F10 (y-axis) negative ionisation mode (C) and F4 (x-axis) and F5 (y-axis) positive ionisation mode (D) regarding the presence of  $C_{15}H_{30}O_3$ .**

**Table 3-4 Dereplication table of predicted bioactive metabolites by OPLS-DA loadings S-plot.** Highlighted rows represent the isolated compounds from SBT531.

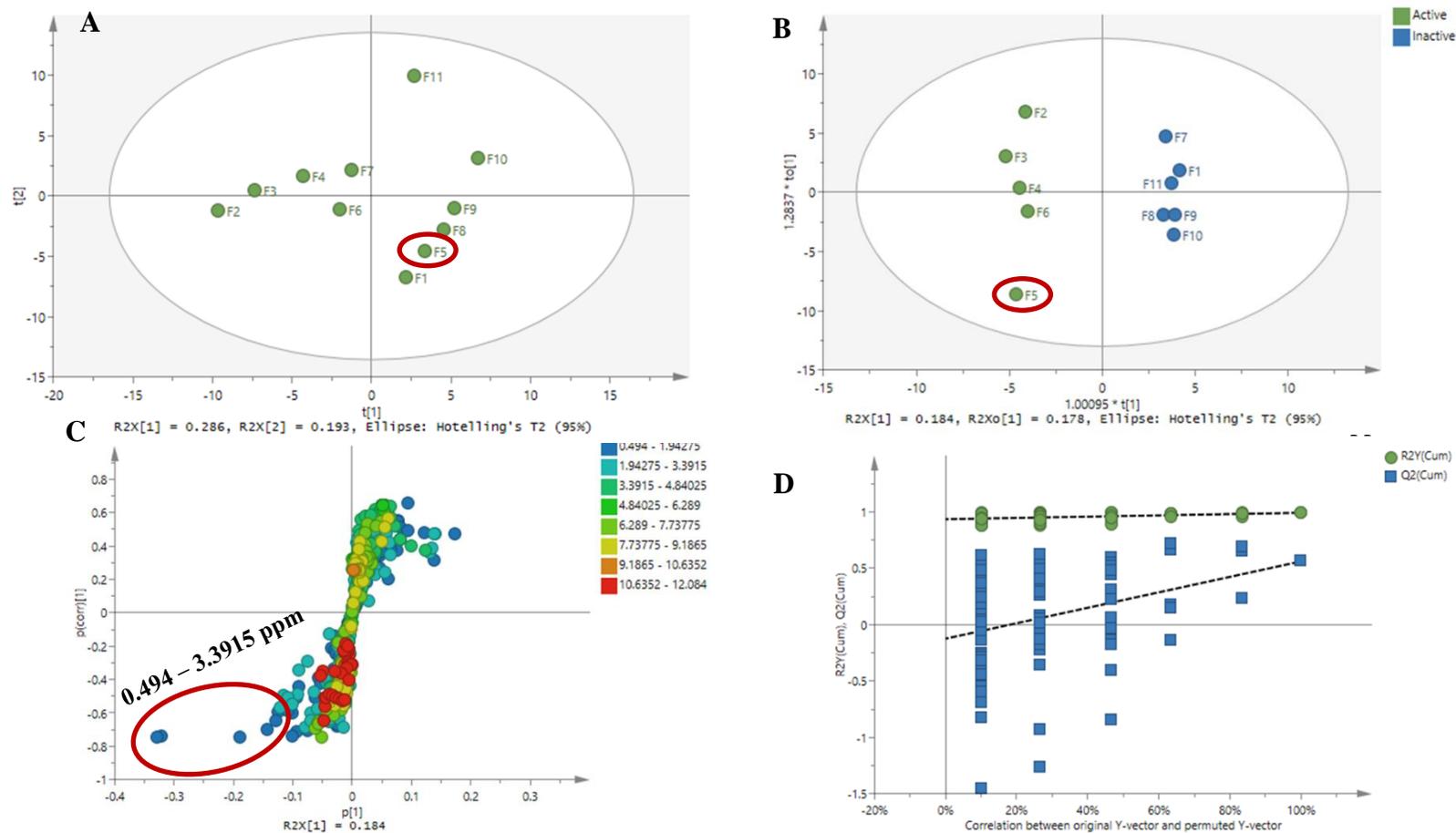
<b>MzMineID</b>	<b>m/z</b>	<b>Rt (min)</b>	<b>Peak Area</b>	<b>MW</b>	<b>Molecular Formula</b>	<b>Name</b>	<b>Source</b>
P_2849	180.1019	7.67	1.37E+08	179.0946	C <sub>10</sub> H <sub>13</sub> NO <sub>2</sub>	4'-tolyl-3-aminopropanoate	Actinobacteria <i>Streptomyces sp</i> Porifera <i>Xestospongia sp</i>
P_16318	227.1389	7.94	1.67E+08	226.1316	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>3</sub>	cyclo(D-cis-Hyp-L-Leu) cyclo[L-(4-hydroxyprolinyl)-L-leucine] cyclo[L-(4-hydroxyprolinyl)-D-leucine]	Actinobacteria <i>Streptomyces sp</i> Cnidaria <i>Palythoa sp</i>
P_9314	165.1103	11.69	3.29E+08	164.1030	C <sub>3</sub> H <sub>12</sub> N <sub>6</sub> O <sub>2</sub> C <sub>5</sub> H <sub>14</sub> N <sub>3</sub> O <sub>3</sub>	No hits	No hits
N_5093	815.5682	14.25	8.68E+07	816.5755	C <sub>48</sub> H <sub>80</sub> O <sub>10</sub>	No hits	No hits
N_3458	815.5679	17.73	2.65E+09	816.5752			
P_865	211.0865	19.33	5.83E+07	210.0792	C <sub>13</sub> H <sub>10</sub> N <sub>2</sub> O	pyocyanin	Proteobacteria <i>Pseudomonas aeruginosa</i>
N_5091	783.5783	22.33	5.44E+08	784.5856	C <sub>48</sub> H <sub>80</sub> O <sub>8</sub>	No hits	No hits
N_3466	783.5782	22.97	1.47E+09	784.5855			
P_9315	393.2999	22.97	7.51E+07	392.2927	C <sub>24</sub> H <sub>40</sub> O <sub>4</sub>	muquibilin sigmosceptrellin-A,B,C	<i>Porifera Prianos sp.</i>

						Mycaperoxide	<i>Porifera Sigmosceptrella laevis</i>
N_3462	515.4321	28.80	5.84E+08	516.4394	C <sub>30</sub> H <sub>60</sub> O <sub>6</sub>	No hits	No hits
N_3459	571.495	33.08	1.39E+09	572.5023	C <sub>34</sub> H <sub>68</sub> O <sub>6</sub>	No hits	No hits
P_1622	340.2635	33.29	7.13E+08	339.2562	C <sub>21</sub> H <sub>31</sub> N <sub>4</sub> C <sub>23</sub> H <sub>33</sub> NO	No hits	No hits
N_2023	241.2174	38.90	1.06E+07	242.2246	C <sub>15</sub> H <sub>30</sub> O <sub>2</sub>	iso-13-methyltetradecanoic acid 13-methylmyristic acid	<i>Bacillus sp. bmg287-af7</i> <i>Micrococcus sp.</i> <i>arctic ice bacterium ANT V/2 370</i>
P_6206	164.107	38.94	6970326	163.0997	C <sub>10</sub> H <sub>13</sub> NO	No hits	No hits
P_5870	164.107	39.67	4.56E+07				
N_1307	285.2436	38.96	2.51E+08	286.2509	C <sub>17</sub> H <sub>34</sub> O <sub>3</sub>	No hits	No hits
N_2025	285.2438	34.52	5.91E+07	286.2511			
P_9566	259.2265	38.97	6927812	258.2192	C <sub>15</sub> H <sub>30</sub> O <sub>3</sub>	2-methoxytetradecanoic acid	<i>Porifera Callyspongia fallax</i>
P_3193	259.2266	28.81	1.48E+07	258.2194			
N_2542	257.2124	30.42	5.56E+08	258.2197		2-methoxy-13-methyltetradecanoic acid	

### 3.4.1.2 NMR Data

The  $^1\text{H}$  NMR dataset was treated similarly as the mass spectral data. PCA was performed on the  $^1\text{H}$  NMR data of 11 fractions, one fraction was not analysed due to its insufficient yield. PCA plots were generated from four components, a  $R^2X$  value of 0.62 was achieved using Pareto scaling, indicating a good fit for a metabolomics dataset and  $Q^2$  of 0.117 which specified the predictive ability of the model. The difference between  $R^2X$  and  $Q^2$  was 0.503, which should be less than 0.3 to confirm complete absence of overfitting. The high diversity on the spectral profiles between fractions could account for the larger difference between  $R^2X$  and  $Q^2$ . It was possible to visualize on the PCA-X scores plots (Figure 3.22A) the similarities between the fractions that showed inhibition effect against PTP1B such as F2 with F3, F4, and F6. However, the active fraction F5 occurred in another quadrant. As shown in Figure 3.22A, F5 is encircled in red. In the PCA scores plot, F5 clustered with the inactive fractions which could signify similarity in their chemical profile but a possible change in functional group or stereochemistry could have affected the bioactivity of F5. Meaning that F5 is chemically unique in comparison to the other active fractions, which was also reflected on the OPLS-DA scores plot (Figure 3.22B) as F5 did not cluster together with the rest of the other active fractions. It is possible to differentiate the active vs inactive fractions in Figure 3.22B through the occurrence of unique functional groups or structural chemical shifts between the classes as shown in Figure 3.30C. The  $R^2X(1)$  was 0.191 which means that the variation between groups is 19.1% and  $R^2X0(1)$  is 0.178 with a variations within groups of 17.8% showing that  $R^2X(1) > R^2X0(1)$ . The variation difference between the groups is higher than within groups as must be expected, indicating that the metabolites found in the active fractions share similar structural features and/or functional groups. The S-plot (Figure 3.22C) showed that the bioactive fractions contained structural features with chemical shifts mostly between 0.494 and 3.3915, on the aliphatic hydroxylated region that corresponds to the fatty acid later isolated and elucidated compounds in this study. This was validated by a permutation test (Figure 3.22D),  $R^2Y$  was 0.992 and  $Q^2$  was 0.561, while  $Q^2Y$  intercept was at -0.124. These values specified a good fit and good predictability as

the  $R^2Y$  and  $Q^2$  values were close to 1 and  $>0.5$ , respectively, while the  $Q^2Y$  intercept was at  $-0.124$ , which is less than 0 indicating passing the permutation test. Furthermore, the difference between  $Q^2$  and  $R^2Y$  was  $0.431$  which was still greater than  $0.3$ , indicating an overfit data (the ideal difference must be less than  $0.3$ ).



**Figure 3-22 1H NMR dimensionality reduction overview. (A)** PCA-X score plot of SBT531 fractions **(B)** OPLS-DA of SBT531 fractions according to inhibition activity against PTPB1B **(C)** S-plot for SBT531 fractions acquired from an OPLS-DA model **(B)** for their inhibition activity against PTPB1B. **(D)** Permutation tests (100 permutations) for SBT531 fractions for the OPLS-DA model of their inhibition activity against PTPB1B, where Q2Y is less than zero. F5 is encircled due to its unusual clustering behaviour in terms of its bioactivity.

### 3.5 Metabolomic-guided sub-fractionation of SBT531 fractions F4, F6+F7 and F9

Based on metabolomic profiling results, fractions of SBT531 were selected for further purification work. This included Fraction 4, combined Fractions 6 and 7, and Fraction 9 at 118.4, 84.6 and 159.2 mg, respectively. Fractionation was performed on the Reveleris Prep HPLC system over a column Vision HT C18 5 $\mu$  with a binary gradient elution method using water with 0.1% FA (A) and ACN with 0.1% FA (B) as described under section 2.6.2.1. The mobile system and chromatographic conditions are detailed in Tables 3.5 and 3.6, respectively.

**Table 3-5 Mobile phase used for sub-fractionation of SBT531 fractions F4, F6+F7 and F9 using the Reveleris Prep HPLC system.**

Length time (min)	% Solvent B
0	5
60	5-100
10	100
<b>Total time=70</b>	

**Table 3-6 Chromatographic conditions applied to the sub-fractionation of SBT531 fractions F4, F6+F7 and F9 using the Reveleris Prep HPLC system.**

<b>Column</b>	Vision HT C18 5 $\mu$
<b>Sample</b>	5 mL liquid
<b>Solvents</b>	Water with 0.1% FA (A) and ACN with 0.1% FA (B)
<b>Flow rate</b>	5 mL/min

<b>Detection mode</b>	UV1+UV2
<b>ELSD Threshold</b>	20 mV
<b>UV Threshold</b>	0.05 AU
<b>UV1 Wavelength</b>	210 nm
<b>UV2 Wavelength</b>	254 nm
<b>UV3 Wavelength</b>	280 nm

The Sub-fractionation of F4 resulted in 104 test tubes, which were pooled according to the similarities of their TLC profile, yielding a total of 12 fractions (F1-F12) as represented on Table 3.7 and Figure 3.23. Combined fractions F6 and F7 gave rise to 64 test tubes, yielding a total of 17 fractions (F1 to F17) listed on Table 3.8 and Figure 3.24. F9 afforded 76 test tubes, yielding a total of 15 fractions (F1 to F15) shown on Table 3.9 and Figure 3.25. The sub-fractions were analysed by HR-LC/MS, NMR (Figures 3.26, 3.27 and 3.28) and prepared for bioassay screening. Sub-fractions with complex TLC profiles and very low yield of less than a milligram were not subjected to NMR analysis.

**Table 3-7 Sub-fractionation of SBT531 Fraction 4 afforded 12 fractions.**

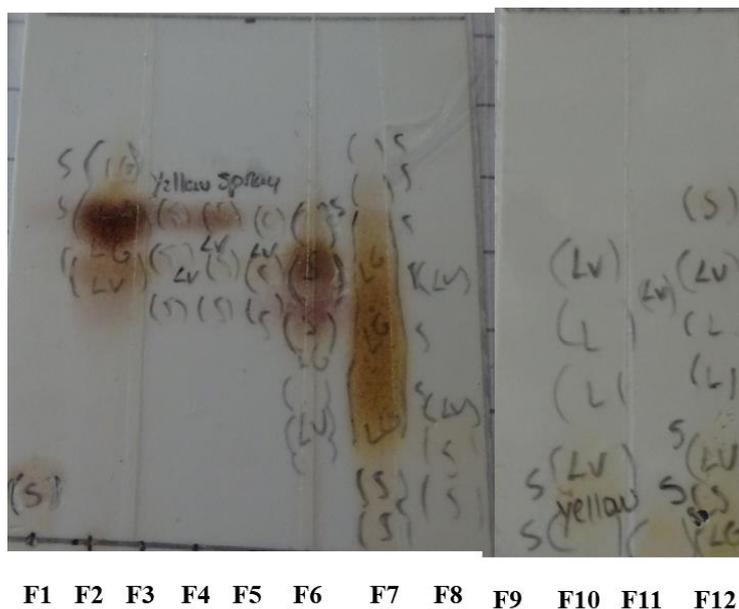
<b>Sub-fraction/</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
subF1	1-6	0.6
subF2	7-10	7.3
subF3	11-29	0.9
subF4	30-34	0.8
subF5	35-40	1.9
subF6	41-47	2.6
subF7	48-71	1.9
subF8	72-85	0.6
subF9	86-91	0.4
subF10	92-97	0.5
subF11	98-104	9.4
subF12	Wash run	28.9

**Table 3-8 Sub-fractionation of SBT531 Fraction 6 and 7 afforded 17 fractions.**

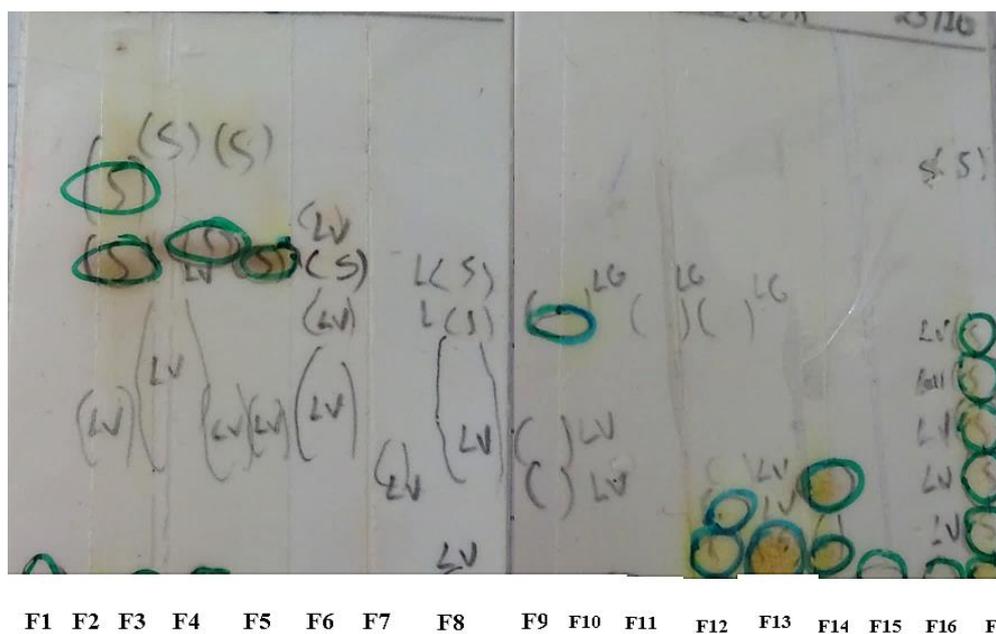
<b>Sub-fraction/</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
subF1	1-3	0.6
subF2	4	18.7
subF3	5	2.3
subF4	6-9	1.9
subF5	10-12	2.6
subF6	13-15	0.9
subF7	16-18	0.4
subF8	19-35	0.8
subF9	36-41	0.3
subF10	42-46	1.7
subF11	47-53	0.4
subF12	54-58	0.8
subF13	59-60	3.9
subF14	61	0.7
subF15	62	1.4
subF16	63-64	0.4
subF17	Wash run	4.8

**Table 3-9 Sub-fractionation of SBT531 Fraction 9 afforded 15 fractions.**

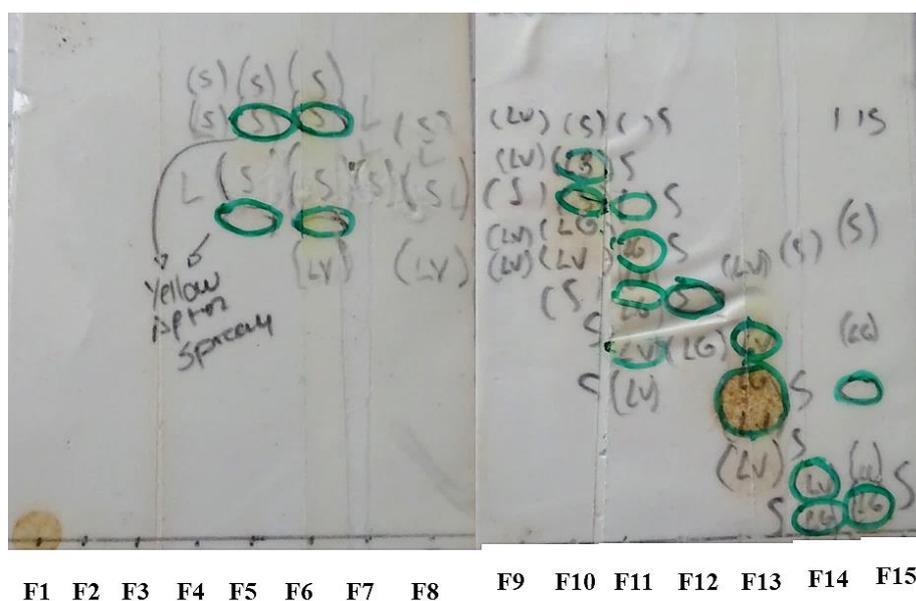
<b>Sub-fraction/</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
subF1	1-3	0.6
subF2	4	0.2
subF3	5-8	0.2
subF4	9	0.1
subF5	10-11	14.4
subF6	12-14	4.6
subF7	15-19	2
subF8	20-22	1.8
subF9	23-24	1.5
subF10	25-33	2.7
subF11	34-36	1.5
subF12	37-49	2
subF13	50-70	15.5
subF14	71-76	2.2
subF15	Wash run	6.6



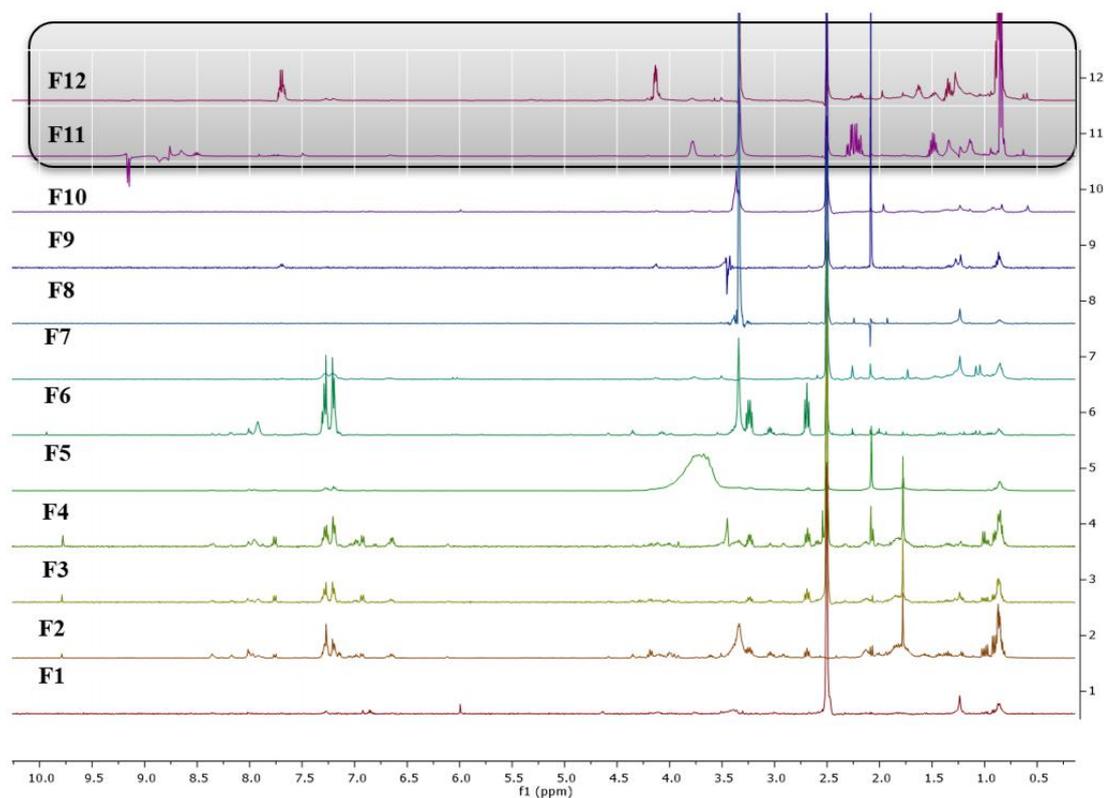
**Figure 3-23 Summary TLC plates of the sub-fractionation of SBT531 fraction 4** on RP18 with mobile phase solvents ACN:Water 1:1. Brown spots were visualized after spraying with anisaldehyde. S (Shortwave), LG (Longwave Green), LV (Longwave Violet).



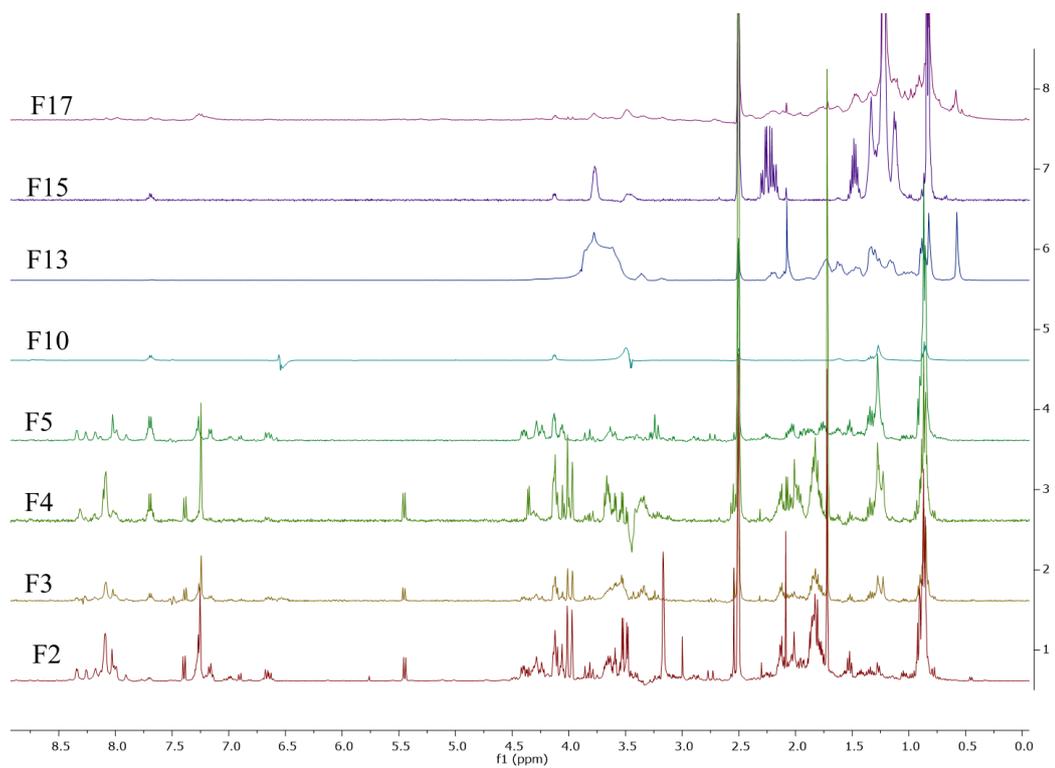
**Figure 3-24 Summary TLC plates of the sub-fractionation of combined SBT531 fractions 6 and 7** on RP18 with mobile phase solvents ACN:Water 1:1. Brown spots encircled with green were visualized after spraying with anisaldehyde. S (Shortwave), LG (Longwave Green), LV (Longwave Violet).



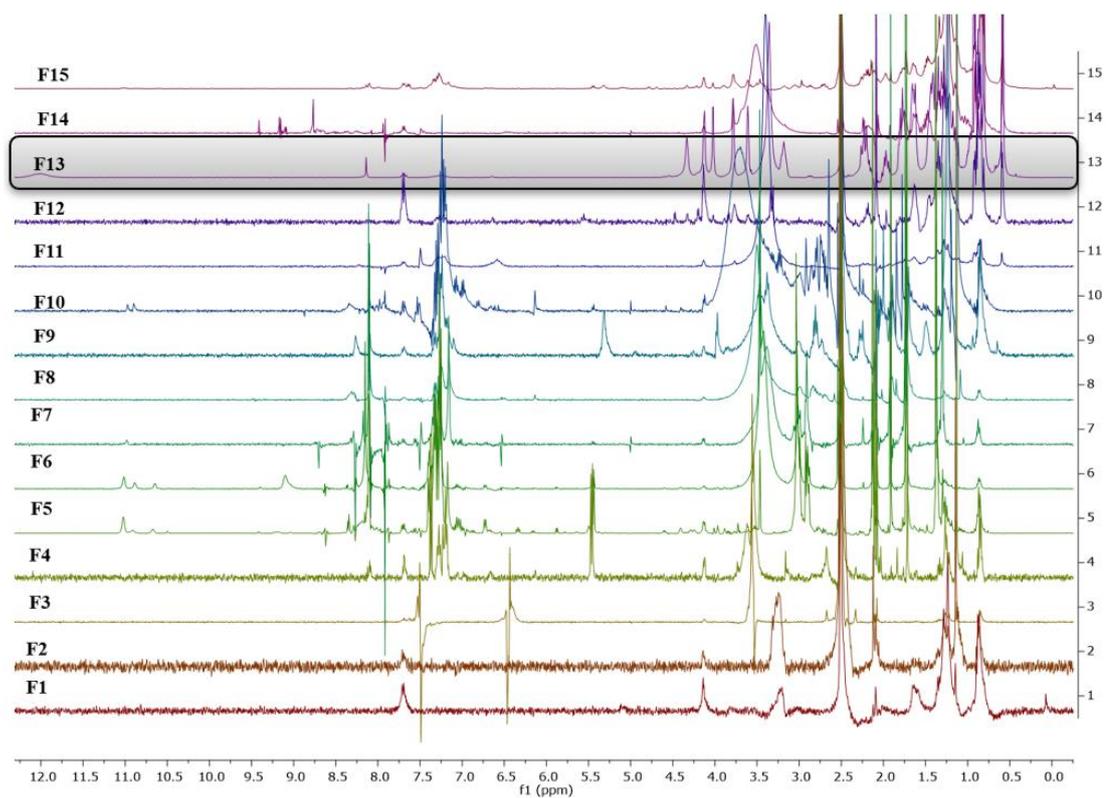
**Figure 3-25 Summary TLC plates of the Sub-fractionation of SBT531 Fraction 9 on RP18 with mobile phase solvents ACN:Water 1:1. Brown spots encircled in green were visualized after spraying with anisaldehyde. S (Shortwave), LG (Longwave Green), LV (Longwave Violet), LY (Longwave Yellow).**



**Figure 3-26  $^1\text{H}$  NMR spectra of SBT531 F4 sub-fractions (SBT2369 to SBT2380) at 400 MHz in  $\text{DMSO-}d_6$  (peak at 2.5 ppm). The coloured box shows the sub-fractions that were subjected to further NMR experiments to structure elucidation.**



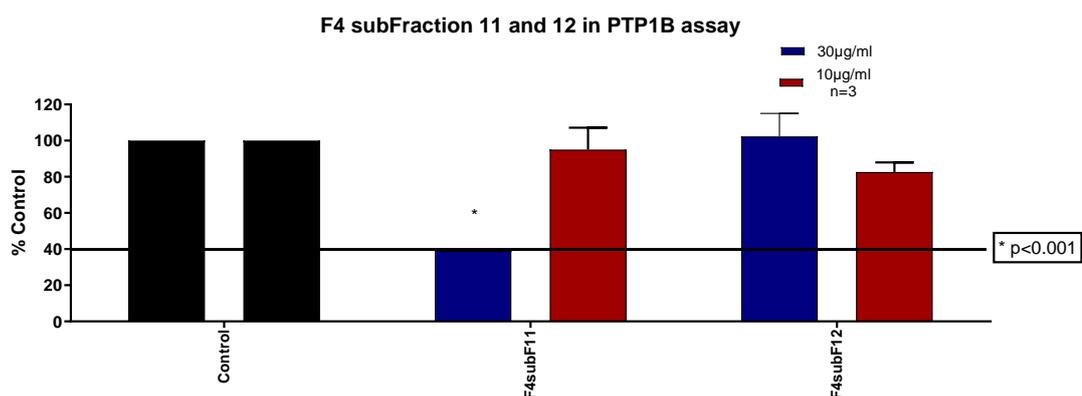
**Figure 3-27** <sup>1</sup>H NMR spectra of SBT531 F6+F7 sub-fractions (SBT2410 to SBT2426) at 400 MHz in DMSO-*d*<sub>6</sub> (peak at 2.5 ppm). It was not possible to perform <sup>1</sup>H NMR on all 17 fractions due to insufficient yield.



**Figure 3-28** <sup>1</sup>H NMR spectra of SBT531 F9 sub-fractions (SBT2381 to SBT2395) at 400 MHz in DMSO-*d*<sub>6</sub> (peak at 2.5 ppm). The coloured box shows the sub-fraction that was subjected to further NMR experiments to structure elucidation.

### 3.5.1 Biological activity of SBT531 subfractions

Hydroxylated fatty acid aseanostatin derivatives were isolated from SBT531 - F4. Aseanostatin P6 was isolated from F2 and afforded an inhibition effect against PTP1B. However, only F4 subfraction 11 showed an inhibition effect at 30  $\mu\text{g/mL}$  (Figure 3.29), which will be discussed later this chapter.



**Figure 3-29** Inhibition effect of SBT531 fraction 4 sub-fractions 11 and 12 against PTP1B at 30 and 10  $\mu\text{g/mL}$ . Error bars represent the standard deviations (SD) of three replicates ( $n=3$ ), and values were significant at  $*p<0.001$  by Two-way ANOVA multiple comparisons with the control.

## 3.6 Identification and structure elucidation of SBT531 compounds

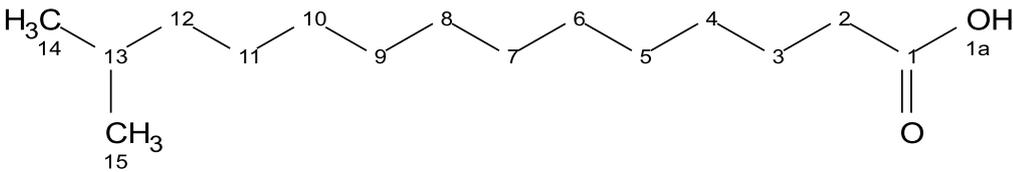
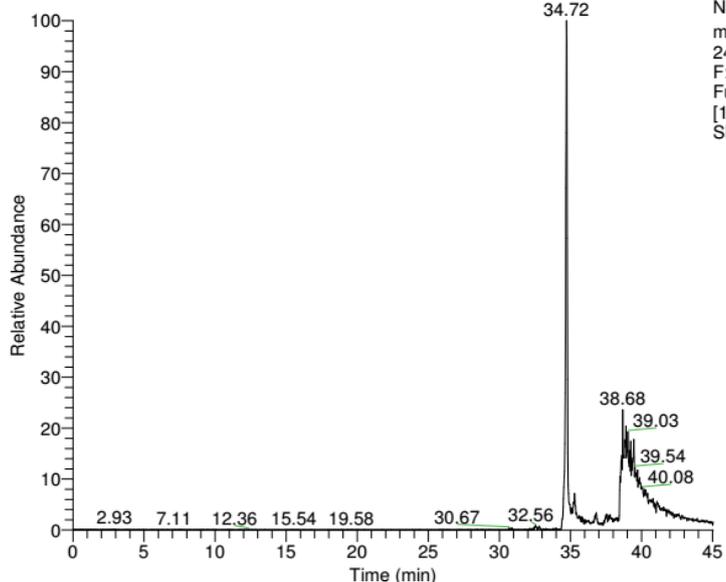
The fractionation and Sub-fractionation of SBT531 afforded the purification of bioactive metabolites. As a result, five compounds were obtained from the scale-up fermentation work (Table 3.10). HR-LC/MS-based metabolomics guided a targeted isolation work of the secondary metabolites. The isolated compounds were elucidated by employing 1D and 2D NMR.

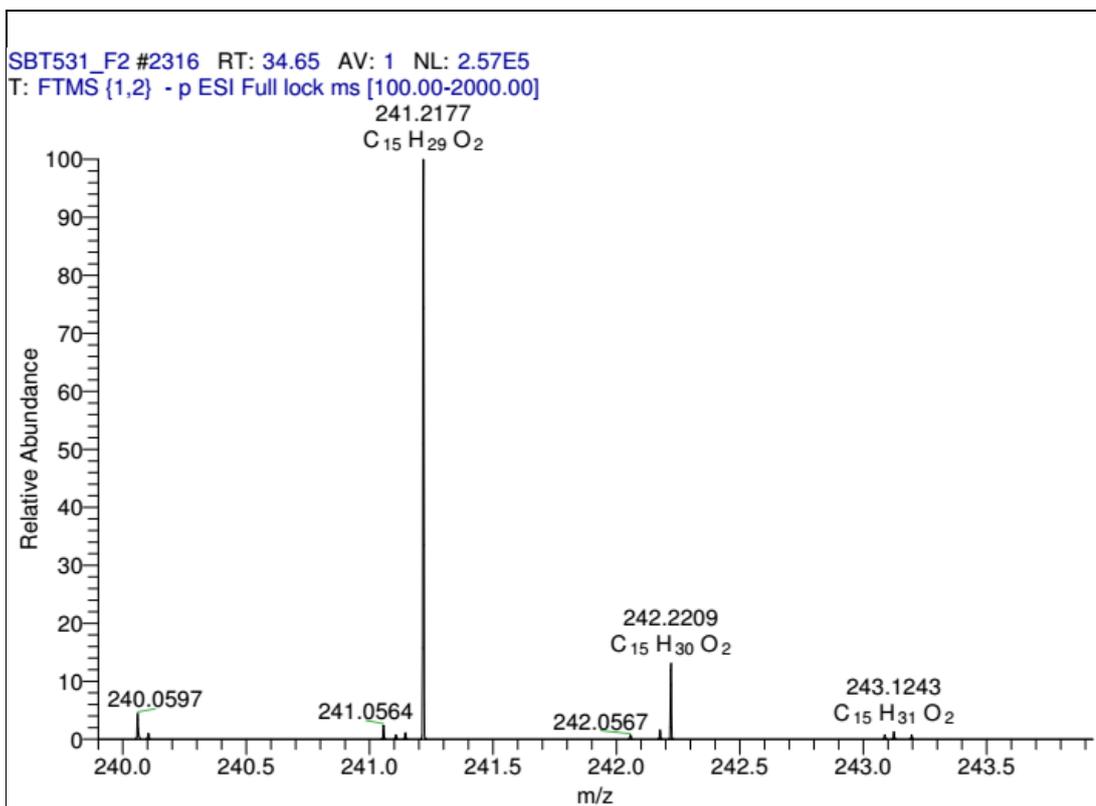
**Table 3-10 Isolated compounds from SBT531 fractions.**

<b>Name</b>	<b>MF</b>	<b>New / Known</b>	<b><i>m/z</i></b>	<b>MWt</b>	<b>Fraction Source</b>	<b>Weight (mg)</b>
aseanostatin P6	C <sub>15</sub> H <sub>30</sub> O <sub>2</sub>	Known	241.2174	242.403	SBT531_F2 SBT2309	198.7
3-hydroxy-13- methyltetradecanoic acid	C <sub>15</sub> H <sub>30</sub> O <sub>3</sub>	Known	257.2125	258.40	SBT531_F4_subF11 SBT2379	9.4
2-hydroxy-14- methylhexadecanoic acid	C <sub>17</sub> H <sub>34</sub> O <sub>3</sub>	Known	285.2438	286.46	SBT531_F4_subF12 SBT2380	28.9
cholic acid derivative	C <sub>24</sub> H <sub>40</sub> O <sub>4</sub>	Known	391.2856	392.58	SBT531_F9_subF13	15.5
<i>N</i> -phenethylacetamide	C <sub>10</sub> H <sub>13</sub> NO	Known	164.1070	163.22	SBT531_F4_subF2	7.3

### 3.6.1 Aseanostatin P6

Table 3-11 Aseanostatin P6.

<p><b>Compound name: aseanostatin P6 (Known):</b></p> <p>Fraction: 2 Retention time (min): 34.72 Synonym(s):</p> <ul style="list-style-type: none"><li>• 13 – methyltetradecanoic acid</li><li>• 13 – methylmyristic acid</li></ul> <p>Source: <i>Muricauda ruestringensis</i> Amount of sample: 198.7 mg Percent yield: 15% Physical description: Dark brown oily sample Molecular formula: C<sub>15</sub>H<sub>30</sub>O<sub>2</sub> Molecular weight: 242.403 g/mol Exact mass: 242.2246</p>

<p><b>LC – HRFTMS spectra [M-H]-</b> RT: 0.00 - 45.04</p>  <p>NL: 6.05E5 m/z= 241.2100-241.2200 F: FTMS {1,2} -p ESI Full ms [100.00-2000.00] MS SBT531_F2</p>



Aseanostatin P6 was isolated from fraction 2 at a yield of 15.18% (198.7 mg). Based on the high-resolution mass spectral data, the ESI peak eluted at 34.72 min and an ion peak was observed in the negative mode at  $m/z$  241.2177  $[M-H]^-$ . This revealed an exact mass of 242.2246 g/mol, which established the molecular formula of C<sub>15</sub>H<sub>30</sub>O<sub>2</sub> (Table 3.11). The degrees of unsaturation in a molecule can be calculated from the molecular formula applying the following formula:

$$\text{Degree of unsaturation} = \frac{2C + 2 + N - H - X}{2}$$

Where **C**=number of carbons, **N**=number of nitrogen's, **H**=number of hydrogens, **X**=number of halogens.

The degree of unsaturation for C<sub>15</sub>H<sub>30</sub>O<sub>2</sub> was 1 and that double bond was represented on the carboxylic acid shown on the most downfield signal in the <sup>13</sup>C NMR while the carboxylic OH singlet can be discerned at 11.97. The <sup>1</sup>H NMR spectrum (Figure 3.30 and 3.31) of aseanostatin P6 showed proton signals from 0.84 to 11.97 ppm. However, most signals were seen from 0 to 2.2 ppm on the aliphatic region. It contained two

methyl ( $CH_3$ ) doublets ( $J = 6.70$  Hz) at  $\delta_H$  0.84-0.85 and were assigned to  $CH_3$ -14 and  $CH_3$ -15. A broad singlet (bs) was observed at  $\delta_H$  1.24, which integrated for nine methylene ( $CH_2$ ) units due to their similar chemical environment and corresponded to  $CH_2$ -4 to  $CH_2$ -12. Overlapping multiplets (m) for methylene  $CH_2$ -3 and methine  $CH$ -13 were detected between 1.45 to 1.51 ppm. Additionally, at  $\delta_H$  2.18, a methylene triplet (t,  $J = 7.33$  Hz) was observed for  $CH_2$ -2.  $^{13}C$  and DEPT (Figure 3.32 and 3.33) was also performed to confirm the carbon assignments. The most downfield signal at  $\delta_C$  175.06 was assigned to position C-1, indicating the deshielding effect of the electron withdrawing carboxylic acid -COOH. On the other hand, the most upfield signal at  $\delta_C$  23.09 was assigned to the two methyl signals for C-14 and C-15.

Moreover, other 2D experiments were performed, such as  $^1H$ - $^1H$  COSY (Figure 3.34) and  $^1H$ - $^{13}C$  HMBC (Figure 3.35). Through  $^1H$ - $^1H$  COSY, it was possible to identify the correlations between  $\delta_H$  0.84, the methyl groups on position 14 and 15 with  $\delta_H$  1.49 for the methine group on position 13. The methylene group at  $\delta_H$  1.49 for position 3 correlated with the methylene unit on position 2 at  $\delta_H$  2.18. Cross peaks were also observed between the overlapping methylene and methine units on positions 3 and 13 at  $\delta_H$  1.49 with the methylene groups in positions 4 and 12, respectively at  $\delta_H$  1.24.

Furthermore,  $^1H$ - $^{13}C$  HMBC allowed to complete the carbon assignments and structure. The correlations of  $\delta_H$  2.18 to  $\delta_C$  25.07 ( $^3J$ ) and  $\delta_H$  2.18 to  $\delta_C$  29.32 ( $^2J$ ), marked the positions for C-4 and C-3, respectively. Also, by detecting the CH direct signals, it can be deduced that the peak at  $\delta_C$  34.23 is C-2. The position of C-13 and C-12 was established by the  $^3J$  and  $^2J$  HMBC cross peaks with  $CH_3$ -14 and  $CH_3$ -15 at  $\delta_H$  0.84 with  $\delta_C$  39.05 and  $\delta_C$  27.96, respectively. Additionally, the positions for C-5 to C-11 were assigned in ascending order of ppm, with three overlapping signals, as  $\delta_C$  27.36, 29.12, 29.49, 29.60, 29.60, 29.60, 29.60 and 29.88. The presented  $^1H$  and  $^{13}C$  NMR spectral data of the isolated compound was compared with the literature (Sarpe and Kulkarni, 2011) which confirmed the elucidation of the structure for aseanostatin P6 (Table 3.12).

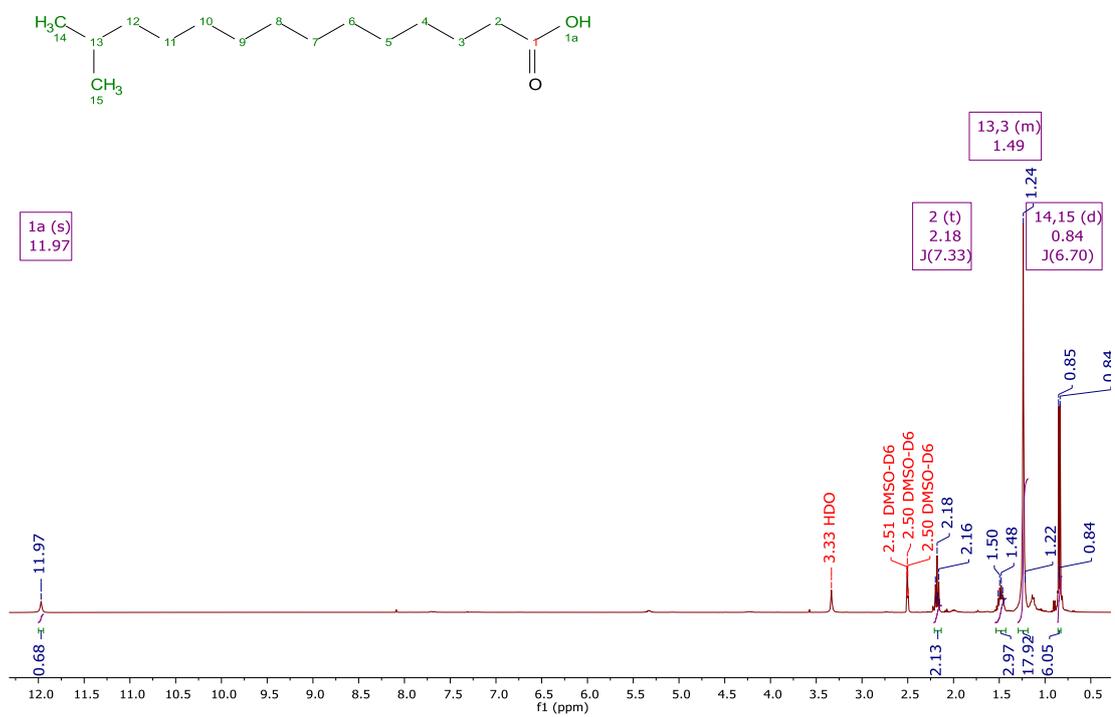


Figure 3-30 <sup>1</sup>H NMR spectrum of aseanostatin P6 at 400 MHz in DMSO-d<sub>6</sub>.

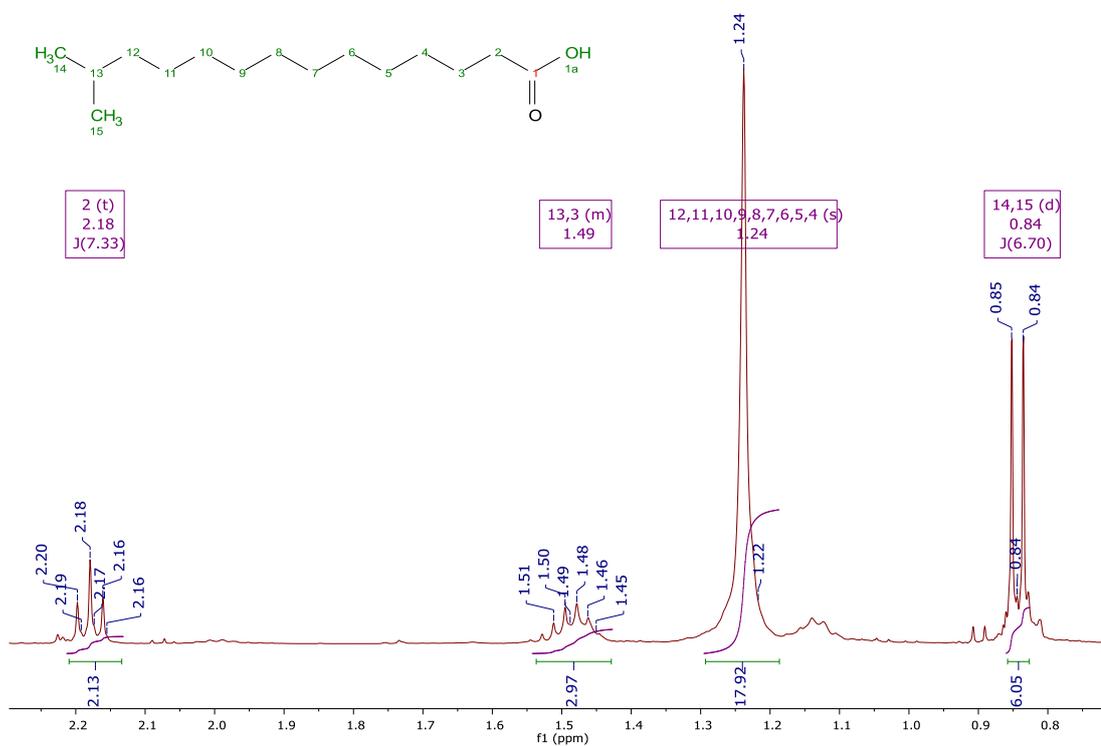


Figure 3-31 Expansion of  $^1\text{H}$  NMR spectrum of aseanostatin P6 at 400 MHz in  $\text{DMSO-d}_6$ .

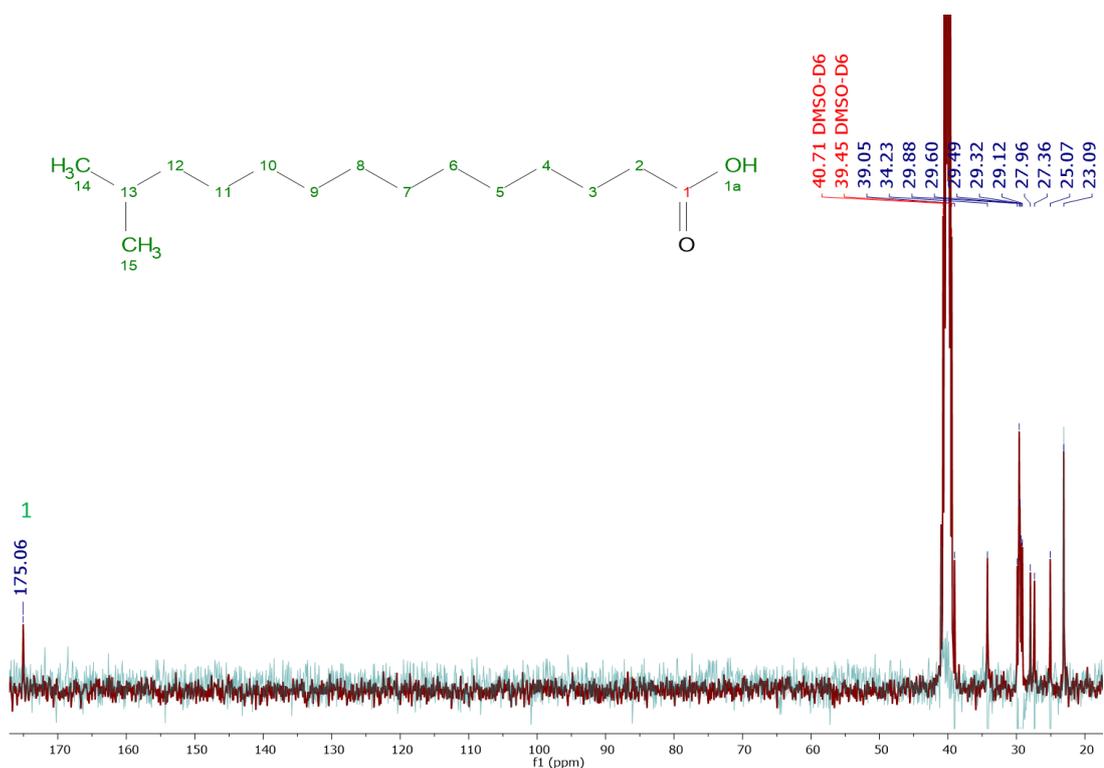


Figure 3-32  $^{13}\text{C}$  spectrum of aseanostatin P6 at 100 MHz superimposed with DEPT-135 (blue line) in  $\text{DMSO-d}_6$ .

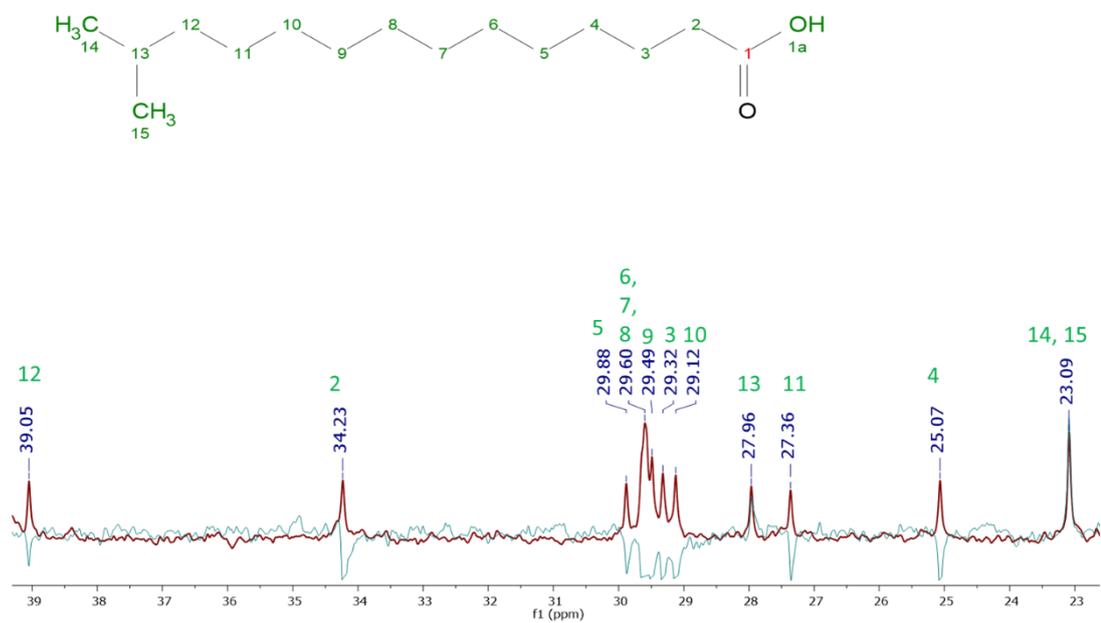


Figure 3-33 Expansion of  $^{13}\text{C}$  spectrum of aseanostatin P6 at 100 MHz superimposed with DEPT-135 (blue line) in  $\text{DMSO-d}_6$ .

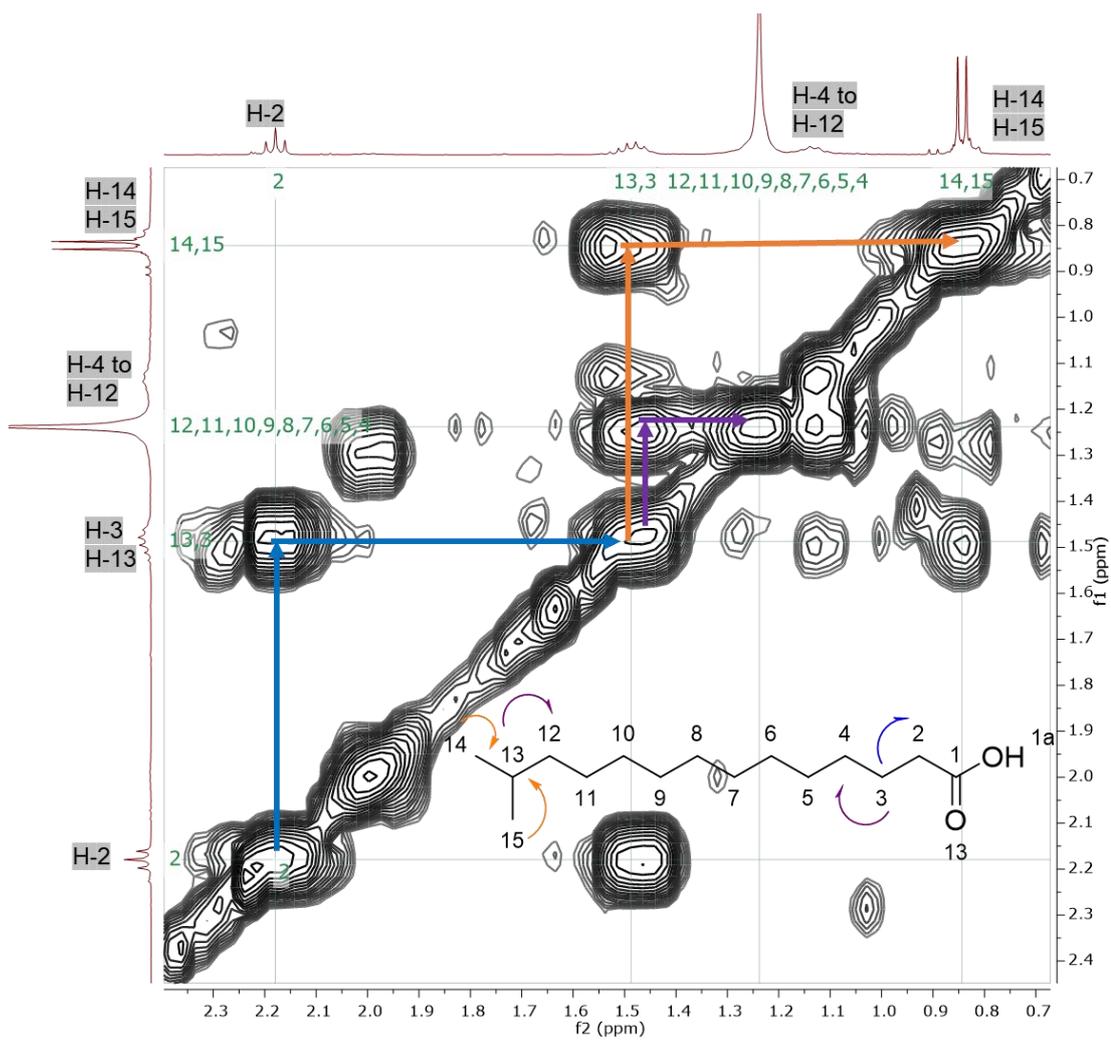


Figure 3-34 1H-1H COSY spectrum of aseanostatin P6 at 400 MHz in DMSO-d6 showed partial correlations in the substructure.

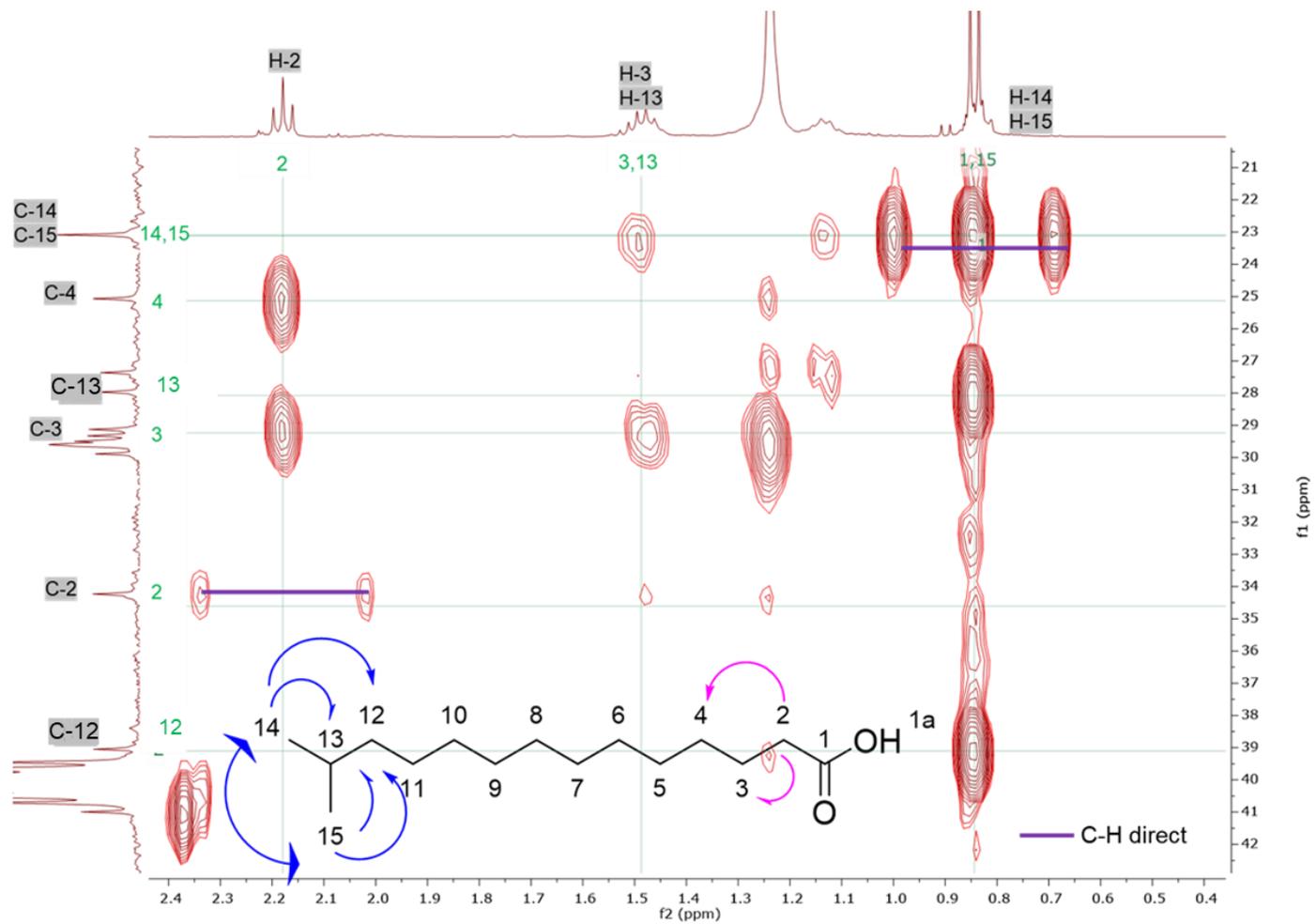


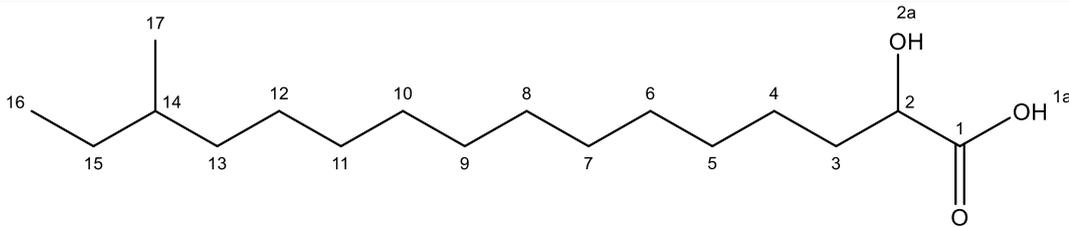
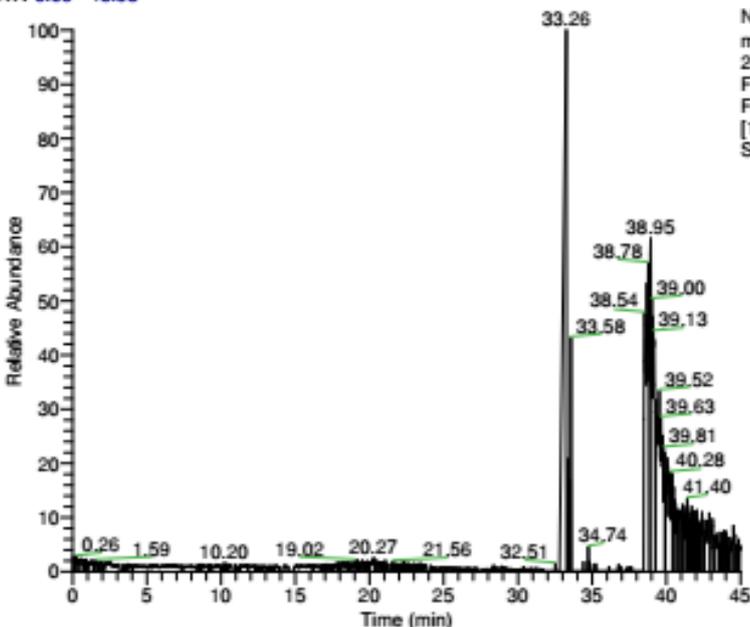
Figure 3-35 HMBC of aseanostatin P6 at 400 MHz. The X and Y – axes correspond to the proton and carbon spectra, respectively.

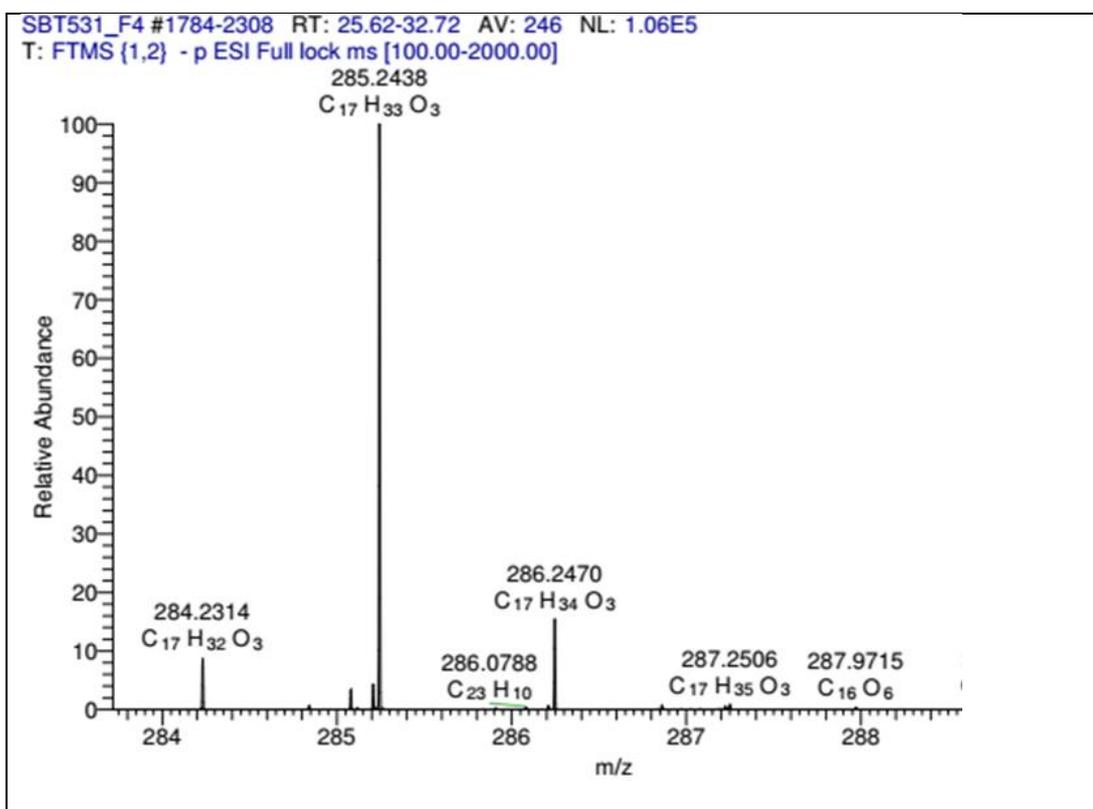
**Table 3-12 <sup>1</sup>H NMR (400 MHz) and <sup>13</sup>C NMR (100 MHz) of aseanostatin P6 in DMSO-d<sub>6</sub> in comparison with literature data** (Sarpe and Kulkarni, 2011). Full assignment was deduced from HMBC data (DMSO-d<sub>6</sub>, 400 MHz).

Atom No.	aseanostatin P6 in DMSO-d <sub>6</sub>		Literature (Sarpe and Kulkarni, 2011) in CDCl <sub>3</sub>	
	δ <sub>H</sub> (multiplicity, <i>J</i> in Hz) (400 MHz)	δ <sub>C</sub> (multiplicity) (100 MHz)	δ <sub>H</sub> (multiplicity, <i>J</i> in Hz) (400 MHz)	δ <sub>C</sub> (100 MHz)
<b>1</b>		175.06 (C)		180.5
<b>2</b>	2.18 (t, <i>J</i> =7.3)	34.23(CH <sub>2</sub> )	2.34 (t, <i>J</i> =7.5)	34.30
<b>3</b>	1.49(m)	29.32(CH <sub>2</sub> )	1.51 (h, <i>J</i> =6.6)	29.40
<b>4</b>		25.07(CH <sub>2</sub> )		24.09
<b>5</b>		27.36(CH <sub>2</sub> )		30.10
<b>6</b>		29.12(CH <sub>2</sub> )		29.90
<b>7</b>		29.49(CH <sub>2</sub> )	1.25-1.32 (m)	29.84
<b>8</b>	1.24 (bs)	29.60(CH <sub>2</sub> )		29.79
<b>9</b>		29.60(CH <sub>2</sub> )		29.60
<b>10</b>		29.60(CH <sub>2</sub> )		29.30
<b>11</b>		29.88(CH <sub>2</sub> )		27.60
<b>12</b>		39.05(CH <sub>2</sub> )	1.12-1.17 (m)	28.20
<b>13</b>	1.49 (m)	27.96(CH)	1.63 (q, <i>J</i> =7.5)	39.20
<b>14</b>	0.84 (d, <i>J</i> =6.7)	23.09(CH <sub>3</sub> )	0.86 (d, <i>J</i> =6.6)	22.80
<b>15</b>	0.84 (d, <i>J</i> =6.7)	23.09(CH <sub>3</sub> )	0.86 (d, <i>J</i> =6.6)	22.80
<b>1a</b>	11.97 (s)			
<b>OH</b>				

### 3.6.2 2-Hydroxy-14-methylhexadecanoic acid

Table 3-13 2-hydroxy-14-methylhexadecanoic acid.

<b>Compound name: 2-hydroxy-14-methylhexadecanoic acid (Known)</b>
Fraction: 4 subFraction 12
Retention time (min): 33.26
Synonym(s): <ul style="list-style-type: none"><li>• <math>\alpha</math>-hydroxyanteisoheptadecanoic acid</li><li>• anteiso-C<sub>17:0</sub> 2-OH</li></ul>
Source: <i>Muricauda ruestringensis</i> (SBT531)
Amount of sample: 28.9 mg
Percent yield: 24%
Physical description: brown sticky powder
Molecular formula: C <sub>17</sub> H <sub>34</sub> O <sub>3</sub>
Molecular weight: 286.454 g/mol
Exact mass: 286.2508
$[\alpha]_D^{20}$ -4 (c 0.1 MeOH)

<b>LC – HRFTMS spectra [M-H]-</b> RT: 0.00 - 45.03




The 2-hydroxy-14-methylhexadecanoic acid was isolated from fraction 4 subfraction 12 with a yield of 24.4% (28.9 mg). Based on the high-resolution mass spectral data, the ESI peak was eluted at 33.26 min and afforded an ion peak at  $m/z$  285.2438  $[M-H]^-$  in the negative mode. This revealed an exact mass of 286.2508 g/mol, which established the molecular formula of C<sub>17</sub>H<sub>34</sub>O<sub>3</sub> (Table 3.13).

The degree of unsaturation for C<sub>17</sub>H<sub>34</sub>O<sub>3</sub> was 1 and that double bond was present on the carboxylic acid moiety as shown by the most downfield signals on the <sup>13</sup>C NMR spectrum. The <sup>1</sup>H NMR spectrum (Figure 3.36) of 2-hydroxy-14-methylhexadecanoic acid showed proton signals from 0.85 to 4.14 ppm on the aliphatic region with a lot of overlapping signals. The presence of impurities turned difficult to accurately assign the integrals. A methine doublet of doublet ( $J=5.73, 3.47$  Hz) was exhibited at  $\delta_H$  4.14 and was assigned to CH-2, the geminal proton to the hydroxyl group showing a deshielding effect. A pentet (p,  $J=5.90$  Hz) signal at  $\delta_H$  1.63 was assigned to CH<sub>2</sub>-3. Overlapping methylene and methine signals were observed between 1.33-1.39 ppm as a multiplet (m), which were assigned to CH-14 and CH<sub>2</sub>-15. Ten overlapping methylene units resonated between 1.23 and 1.30 ppm as multiplets (m), assigned to

$CH_2$ -4 to  $CH_2$ -13. It contained two overlapping methyl multiplets (m) at  $\delta_H$  0.85 and 0.87 corresponding to  $CH_3$ -16 and  $CH_3$ -17, respectively that consisted of an overlapping triplet and a doublet.  $^{13}C$  NMR (Figure 3.37 and 3.38) was also performed to confirm the carbon assignments. The most downfield signal at  $\delta_C$  167.55 was assigned to position C-1, indicating the deshielding effect by an electron withdrawing carboxylic acid substituent. The chemical shift, when compared to the carboxylic acid unit found in aseanostatin P6, went upfield due to the presence of the hydroxyl substituent at position 2a that was responsible for an electron donating shielding effect on C-1. Moreover, signals were observed for a hydroxyl bearing carbon at  $\delta_C$  67.96 that was assigned to C-2 while C-3, the vicinal methylene carbon to the hydroxyl carbon at C-2, was assigned at  $\delta_C$  38.64. On the other hand, the most upfield signals at  $\delta_C$  11.37 and 14.47 were assigned to the two methyl signals C-16 and C-17, respectively. The chemical shifts of the terminal methyl units in this structure went upfield when compared to the ones in aseanostatin P6. The C-16 is neighboring a methylene unit at C-15, while C-17 is neighboring a methine unit at C-14 which explains the difference of the chemical shifts. The chemical shift of C-16 shifted upfield (shielding effect) due to a higher electron donating effect of the methylene ( $CH_2$ ) versus methine ( $CH$ ) (Tonelli and Schilling, 1981, Heffner et al., 1986).

As above, further 2D experiments were performed, such as  $^1H$ - $^1H$  COSY (Figure 3.39) and  $^1H$ - $^{13}C$  HMBC (Figure 3.40). Correlations between protons  $CH$ -2 and  $CH_2$ -3 as well as between  $CH_2$ -3 and  $CH_2$ -4 were observed through  $^1H$ - $^1H$  COSY. COSY cross peaks were detected between  $CH_3$ -16 and  $CH_2$ -15,  $CH_2$ -15 to  $CH$ -14 and finally from  $CH$ -14 to  $CH_3$ -17. The HMBC (Figure 3.40) enabled to determine the complete structure as well to complete the carbon assignments of the present structure. The methine proton at C-2 illustrated strong correlations with C-1 ( $\delta_C$  167.55), C-3 ( $\delta_C$  38.64), C-4 ( $\delta_C$  30.36) and C-5 ( $\delta_C$  23.81). The methylene unit at C-3 showed HMBC correlations with C-2, C-4 and C-5. C-14, C-15 and C-13 resonating at 37.50, 29.63 and 23.09 ppm, respectively. On the other hand, the methyl protons on C-17 and C-16 correlated with C-13, C-14 and C-14, C-15, respectively. Furthermore, the positions of C-6 to C-12 were assigned in ascending order of resonances with some overlapping signals as  $\delta_C$  22.98, 27.96, 28.93, 29.56, 29.56, 29.59 and 29.59, respectively. It was

also possible to see the correlations between the methylene protons at 1.23-1.30 ppm with the assigned carbons. It is worth mentioning the presence of impurities with aromatic signals on the  $^1\text{H}$  NMR spectrum (Figure 3.36) between 7.5-8.0 ppm, as well on the  $^{13}\text{C}$  NMR spectrum (Figure 3.37) between 135-140 ppm, that not belong to the compound being analyzed.

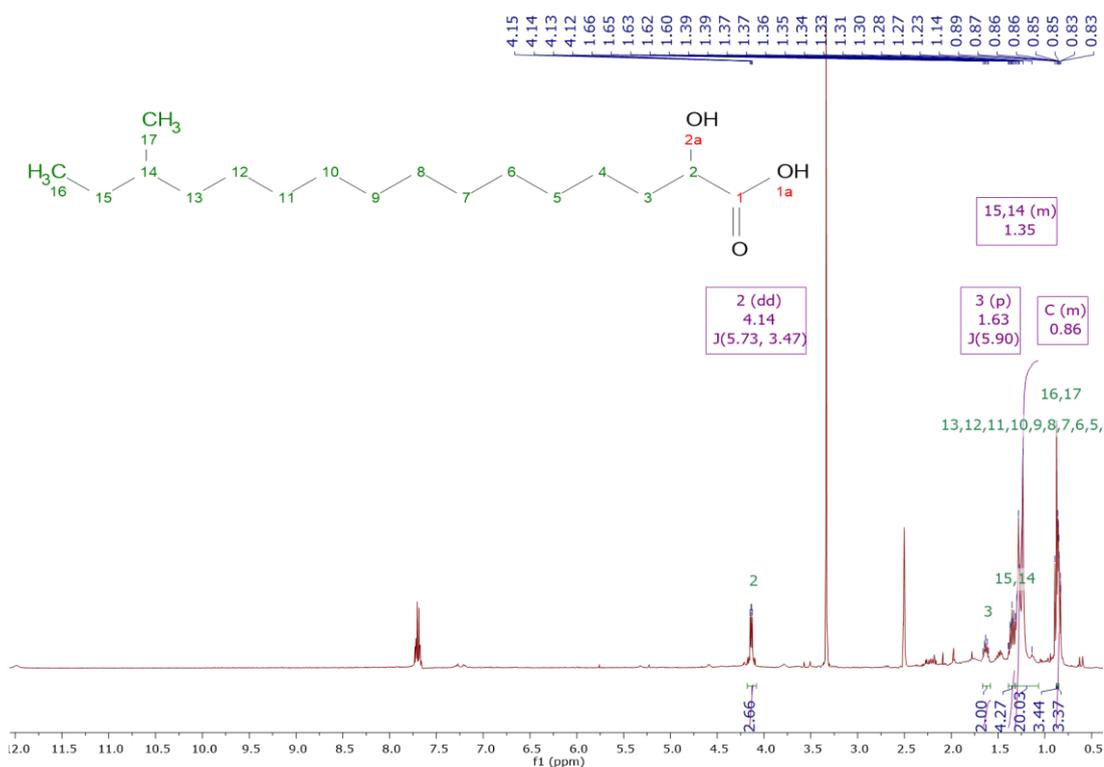


Figure 3-36  $^1\text{H}$  NMR spectrum of 2-hydroxy-14-methylhexadecanoic acid at 400 MHz in DMSO- $d_6$ .

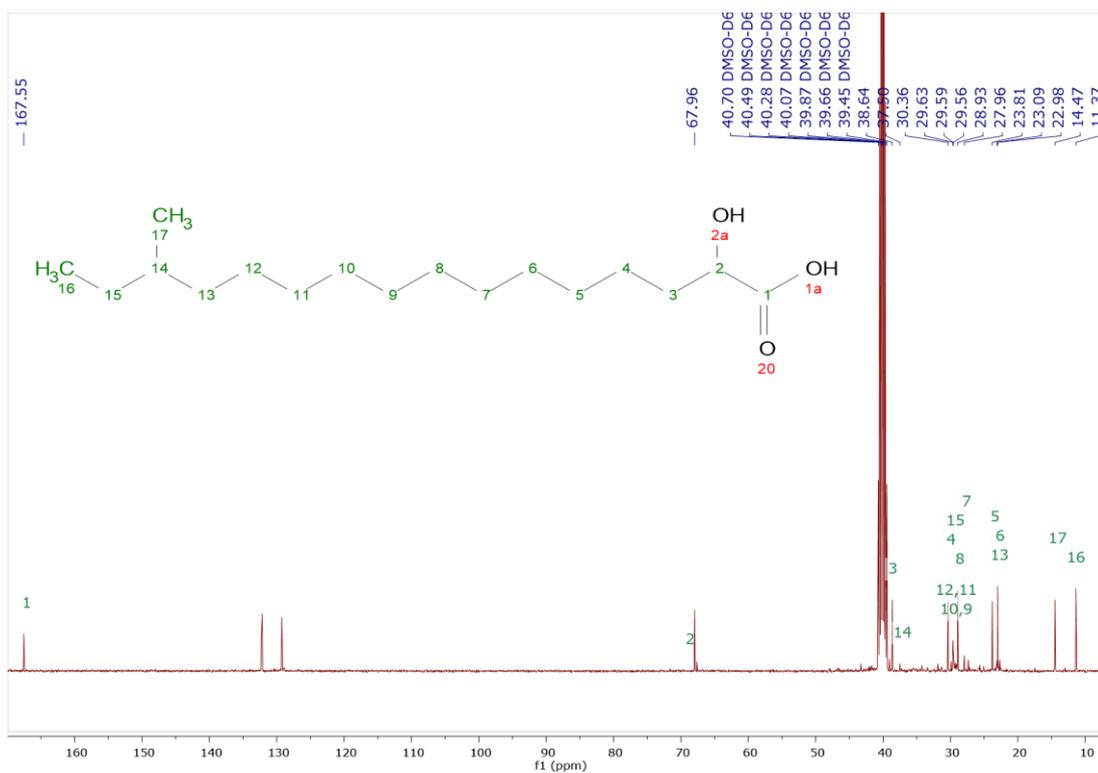


Figure 3-37  $^{13}\text{C}$  spectrum of 2-hydroxy-14-methylhexadecanoic acid at 100 MHz in  $\text{DMSO-}d_6$ .

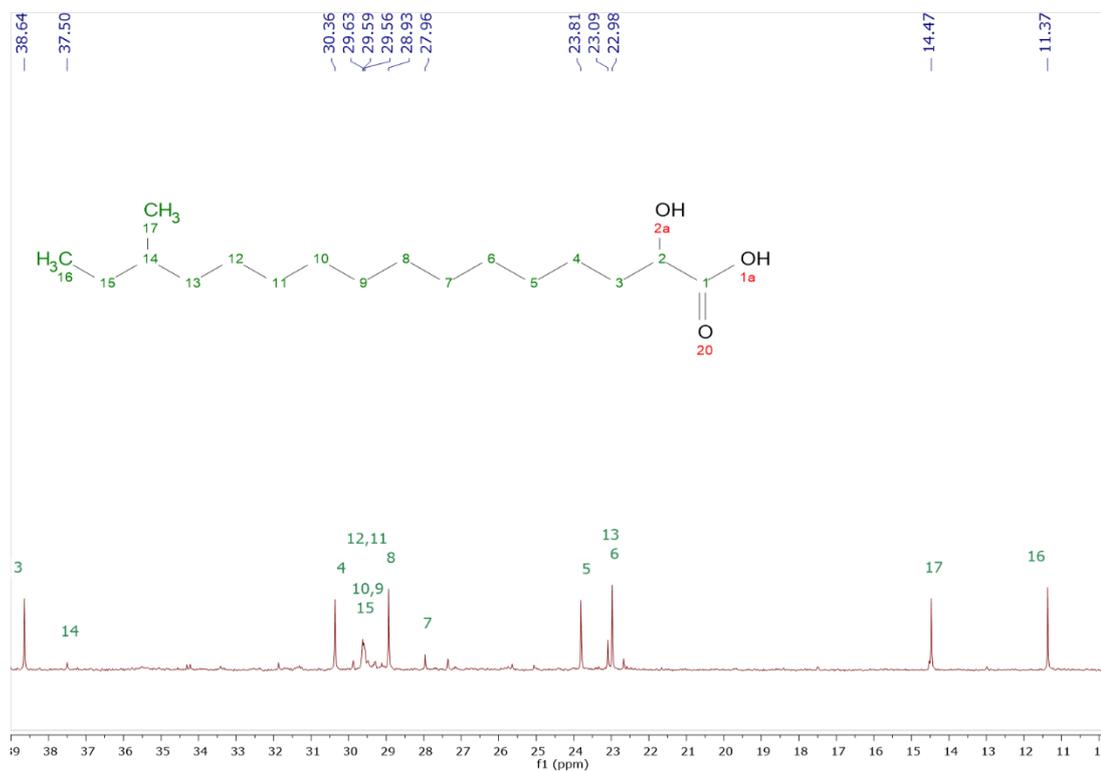
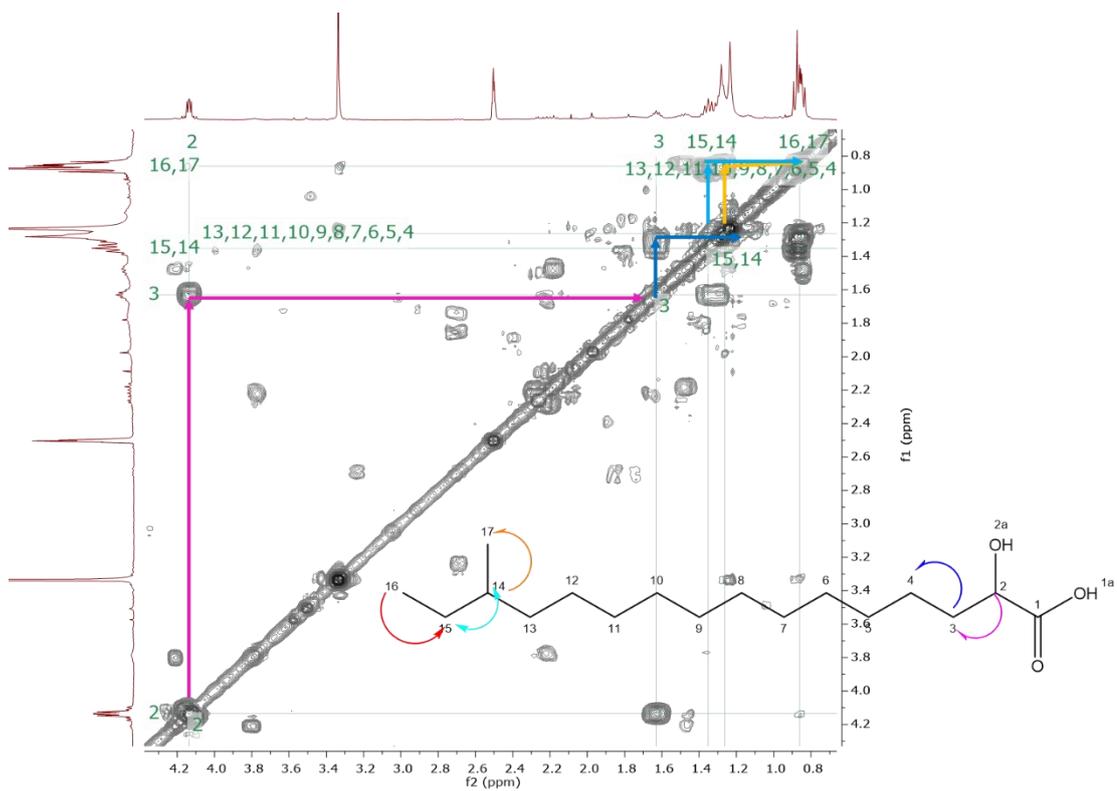
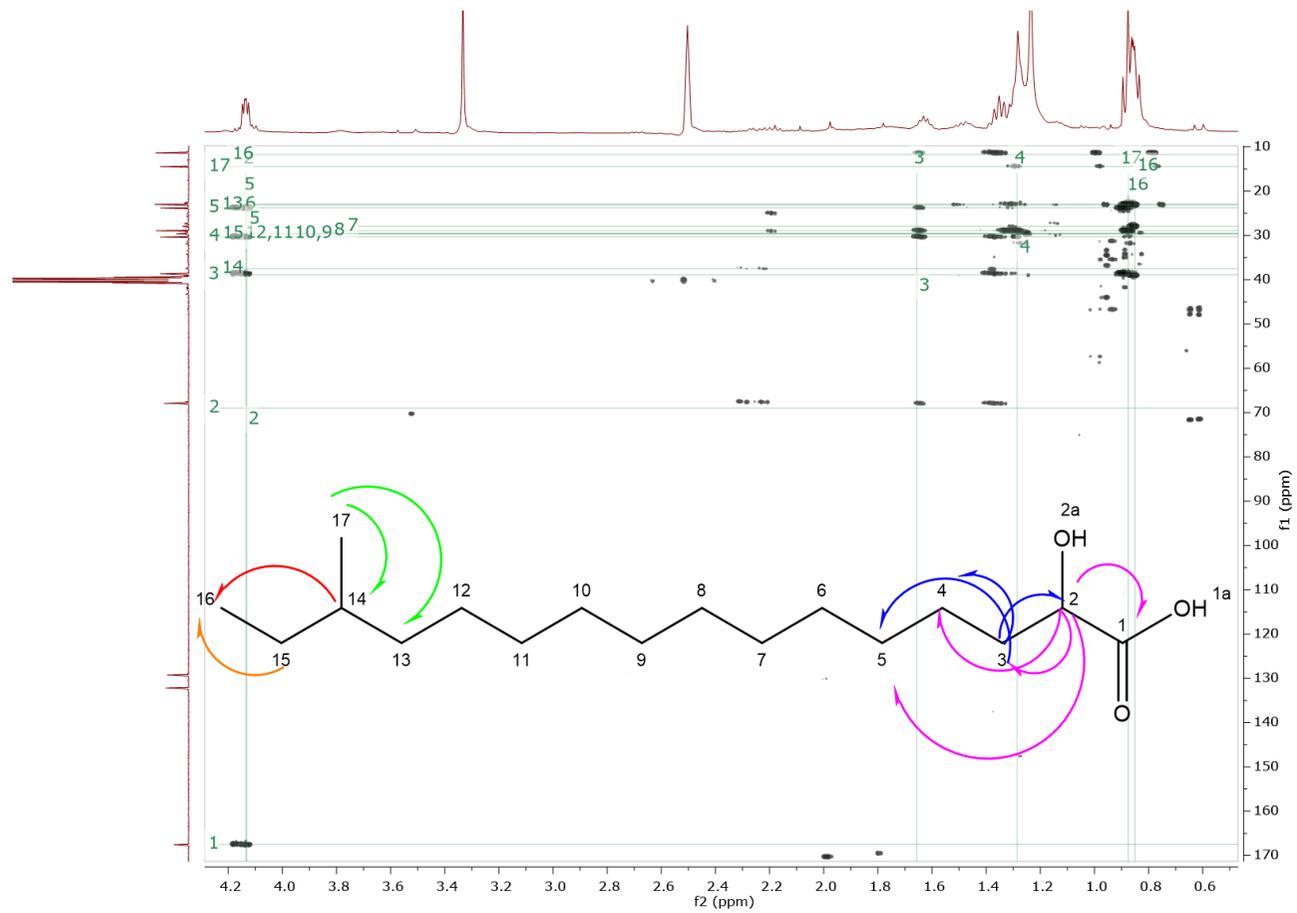


Figure 3-38 Expansion of  $^{13}\text{C}$  spectrum of 2-hydroxy-14-methylhexadecanoic acid at 100 MHz in  $\text{DMSO-}d_6$ .



**Figure 3-39**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of 2-hydroxy-14-methylhexadecanoic at 400 MHz in  $\text{DMSO-d}_6$  showing partial correlations in the substructure.



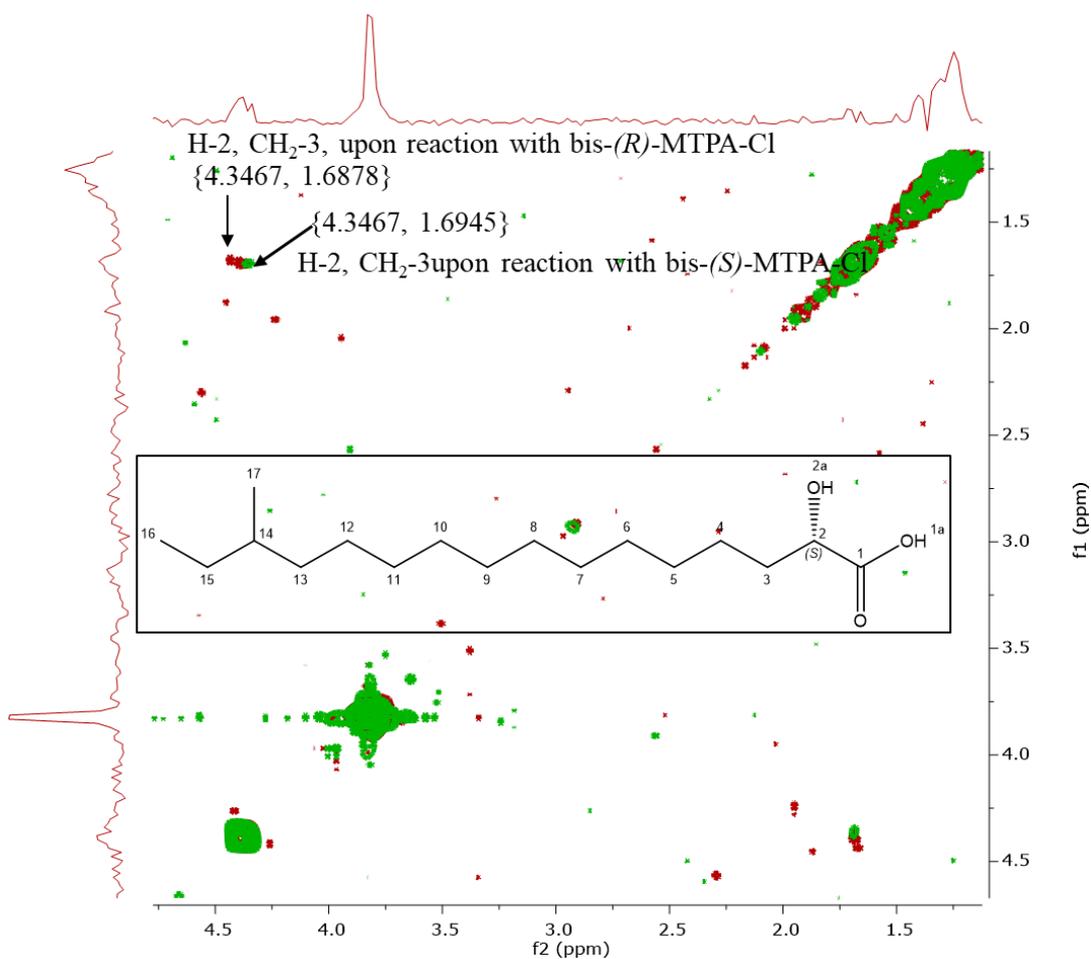
**Figure 3-40** HMBC of 2-hydroxy-14-methylhexadecanoic acid at 600 MHz. The X and Y – axes correspond to the proton  $^1\text{H}$  (600 MHz) and carbon spectra  $^{13}\text{C}$  (150 MHz), respectively.

**Table 3-14**  $^1\text{H}$  NMR (400 MHz) and  $^{13}\text{C}$  NMR (100 MHz) of 2-hydroxy-14-methylhexadecanoic acid in DMSO- $d_6$  in comparison with literature data (Carballeira et al., 2002). Full assignment was deduced from HMBC (DMSO- $d_6$ , 400 MHz).

Atom No.	2-hydroxy-14-methylhexadecanoic acid in DMSO- $d_6$		Literature (Carballeira et al, 2002) in $\text{CDCl}_3$	
	$\delta_{\text{H}}$ (multiplicity, $J$ in Hz) (400 MHz)	$\delta_{\text{C}}$ (multiplicity) (100 MHz)	$\delta_{\text{H}}$ (multiplicity, $J$ in Hz) (500.1 MHz)	$\delta_{\text{C}}$ (multiplicity) (125.8 MHz)
1a-OH		167.55(C)		173.4
2	4.14 (dd, $J=5.73, 3.47$ )	67.96(CH)	3.76	80.7
3	1.63 (p, $J=5.90$ )	38.64(CH <sub>2</sub> )	1.70	38.8
4	1.26(m)	30.36(CH <sub>2</sub> )	1.33-1.25	25.1
5		23.81(CH <sub>2</sub> )		27.01
6		22.98(CH <sub>2</sub> )		29.4
7		27.96(CH <sub>2</sub> )		29.5
8		28.93(CH <sub>2</sub> )		29.6
9		29.56(CH <sub>2</sub> )		29.69
10		29.56(CH <sub>2</sub> )		29.71
11		29.59(CH <sub>2</sub> )		30.0
12	29.59(CH <sub>2</sub> )	32.8		
13	23.09(CH <sub>2</sub> )	39.0		
14	1.35 (m)	37.50(CH)	1.65	34.4
15	1.35 (m)	29.63(CH <sub>2</sub> )		
16	0.85 (m)	11.37(CH <sub>3</sub> )	0.83	11.4
17	0.87 (m)	14.47(CH <sub>3</sub> )	0.85	19.2

The optical rotation was  $[\alpha]_{\text{D}}^{20} -4$  (c 0.1 MeOH), giving a R configuration, (2*R*)-2-hydroxy-14-methylhexadecanoic acid. However, due to the small magnitude of the optical rotation, this could indicate the presence of a mixture of *R* and *S* enantiomers. The absolute stereochemistry was confirmed by the Mosher method. The stereochemistry of the chiral centre C-2 was achieved by observing the changes in the chemical shifts of the neighbouring protons between the bis-(*R*)-MTPA-Cl and the bis-

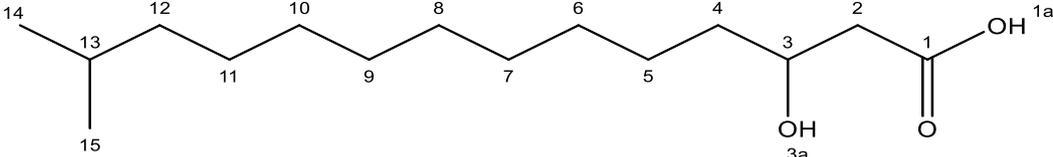
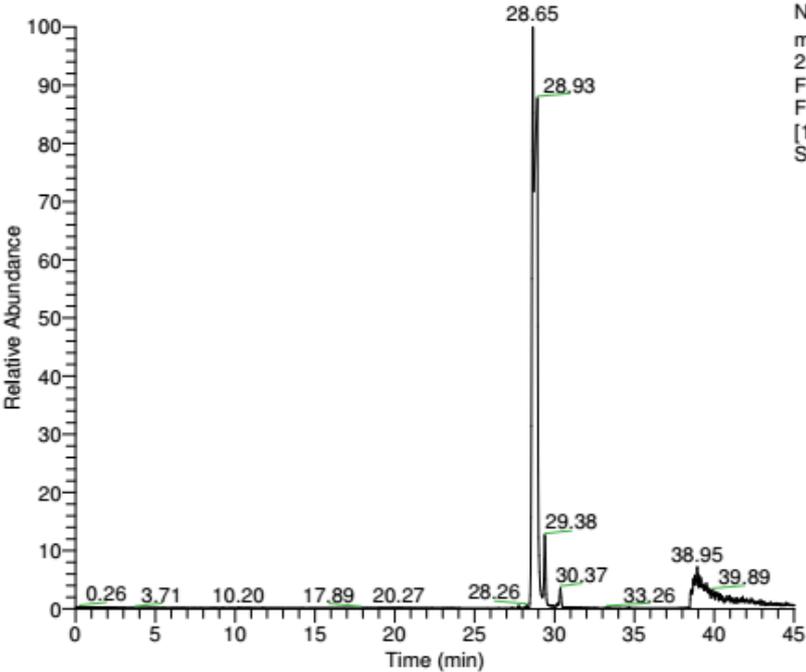
(*S*)-MTPA-Cl derivatives of the 2-hydroxy-14-methylhexadecanoic acid. The chemical shift of the methylene ( $CH_2$ -3) protons was shifted downfield to  $\delta_H$  1.6878 upon reacting with bis-(*R*)-MTPA-Cl reagent (red spectrum) and to  $\delta_H$  1.6945 upon reacting with the bis-(*S*)-MTPA-Cl reagent (green spectrum). The  $\Delta\delta^{SR}(\delta S-\delta R)$  value for  $CH_2$ -3 was calculated to be +0.0067 (+2.68 Hz) as shown in Figure 3.41. The positive value marked the position of the methylene C-3 to be at the right of the chiral centre C-2. By assigning the H-2 backwards while the carboxylate carbon C-1 was positioned left on the chiral centre establishing an *S* configuration.



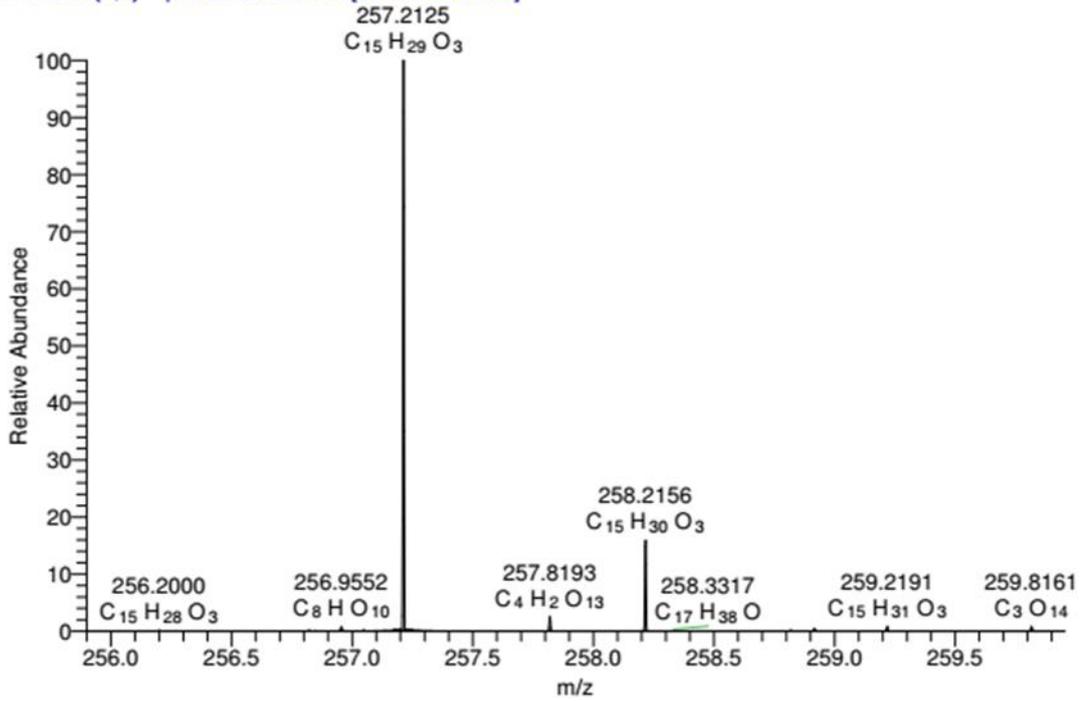
**Figure 3-41 Superimposed  $^1H$ - $^1H$  COSY NMR (400 MHz) spectrum in Pyridine- $d_5$ .** MTPA derivatives of 2-hydroxy-15-methylhexadecanoic acid measured 72 hours after reaction with both bis - (*R*) - MTPA-Cl (red spectrum) and bis- (*S*)-MTPA-Cl reagents (green spectrum). The *S* configuration was established.

### 3.6.3 3-Hydroxy-13-methyltetradecanoic acid

Table 3-15 3-hydroxy-13-methyltetradecanoic acid.

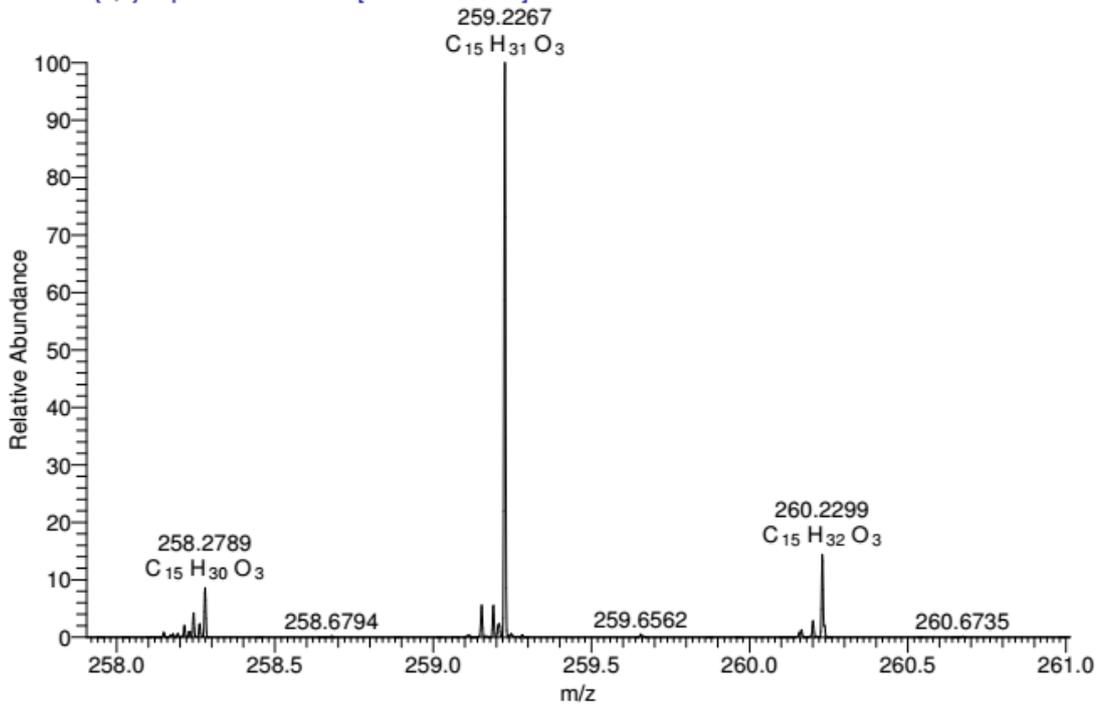
<b>Compound name: 3-hydroxy-13-methyltetradecanoic acid (known)</b>
Fraction: 4 subFraction 11 Retention time (min): 28.65 Synonym(s): iso-C <sub>15:0</sub> 3-OH Source: <i>Muricauda ruestringensis</i> (SBT531) Amount of sample: 9.4 mg Percent yield: 8% Physical description: white sticky powder Molecular formula: C <sub>15</sub> H <sub>30</sub> O <sub>3</sub> Molecular weight: 258.40 g/mol Exact mass: 258.2195 [ $\alpha$ ] <sub>D</sub> <sup>20</sup> +68(c 0.05 MeOH)

<b>LC – HRFTMS spectra [M-H]<sup>-</sup></b> RT: 0.00 - 45.03
 <p>NL: 5.70E7 m/z= 257.2000-257.4000 F: FTMS {1,2} - p ESI Full lock ms [100.00-2000.00] MS SBT531_F4</p>

SBT531\_F4 #1163-3068 RT: 16.84-42.77 AV: 910 NL: 7.29E5  
T: FTMS {1,2} - p ESI Full lock ms [100.00-2000.00]



[M+H]<sup>+</sup>

SBT531\_F4 #1754-2302 RT: 25.20-32.63 AV: 274 NL: 2.77E4  
T: FTMS {1,1} + p ESI Full lock ms [100.00-2000.00]



The 3-hydroxy-13-methyltetradecanoic acid was isolated from fraction 4 subfraction 11 with a yield of 7.94% (9.4 mg). Based on the high-resolution mass spectral data, the ESI peak was eluted at 28.65 min with ion peaks at  $m/z$  257.2125  $[M-H]^-$  in the negative mode and 259.2267  $[M+H]^+$  in the positive mode. This revealed the exact mass at 258.2195 g/mol, which established the molecular formula of  $C_{15}H_{30}O_3$  (Table 3.15).

The degree of unsaturation for  $C_{15}H_{30}O_3$  was again 1 and that the double bond was present on the carboxylic acid as shown by the most downfield signal on the  $^{13}C$  NMR spectrum. The  $^1H$  NMR spectrum (Figure 3.42) of 3-hydroxy-13-methyltetradecanoic acid showed proton signals from 0.82 to 3.81 ppm on the aliphatic region, which was quite comparable to that of 2-hydroxy-14-methylhexadecanoic acid. Similar resonances were observed like the two methyl doublets ( $J = 6.60$  Hz) at  $\delta_H$  0.84-0.85. A methylene multiplet was also observed between 1.11-1.16 ppm while six methylene units resonated at  $\delta_H$  1.24 as a broad singlet or overlapping multiplets. Two overlapping methylene units was also visualized as a broad singlet (bs) between 1.32-1.34 ppm. Additionally, a methine heptet ( $J = 6.57$  Hz) resonated between 1.45 and 1.54 ppm and was assigned to position CH-13. The two doublet of doublets at  $\delta_H$  2.21 and 2.28, with coupling constants of  $J = 14.73, 7.78$  and  $14.77, 5.18$  Hz, respectively, were assigned to the two methylene hydrogens adjacent to the chiral centre. One hydroxyl-bearing methine proton was shown at 3.78 ppm as a multiplet.

The  $^{13}C$  NMR spectra (Figure 3.43 and Figure 3.44) afforded a signal at  $\delta_C$  173.62 for the carboxylic acid carbon at position C-1. The hydroxyl-bearing carbon C-3 was observed at  $\delta_C$  67.63 while its vicinal methylene carbon at C-2 between the carboxylic acid at C-1 and hydroxyl carbon at C-3 was observed at  $\delta_C$  43.31. In addition, C-14 and C-15 were observed as methyl carbons yielding the most upfield resonance at  $\delta_C$  23.10.

2D experiments were performed, such as  $^1H$ - $^1H$  COSY (Figure 3.45) and  $^1H$ - $^{13}C$  HMBC (Figure 3.47) to confirm the structure of 3-hydroxy-13-methyltetradecanoic acid.  $^1H$ - $^1H$  COSY showed correlations between CH-3 at  $\delta_H$  3.78 with  $CH_2$ -4 at  $\delta_H$

1.34 and the geminal methylene protons H-2A and H-2B at 2.21 and 2.28 ppm. Additionally, both  $CH_3$ -14 and  $CH_3$ -15 at  $\delta_H$  0.84 correlated with  $CH$ -13 at  $\delta_H$  1.50. The HMBC data (Figure 3.46) completed the structure elucidation of the metabolite. Carbons C-1 and C-3 were assigned at  $\delta_C$  173.62 and 67.63, respectively, which illustrated strong correlations with the geminal methylene protons of C-2 at  $\delta_H$  2.21 and 2.28. C-2 was assigned as  $\delta_C$  43.31, the third most downfield signal, which corresponded to the carbon between carboxylic and hydroxylic-bearing carbon. The methyl carbons C-14 and C-15 correlated with the methine protons of C-13 at 1.50 ppm, with the methylene protons of C-11 at 1.14 ppm and with each other's methyl protons at  $\delta_H$  0.84-0.85 and carbon at  $\delta_C$  23.10. Also, by the visualization of the CH direct signals, was deduced the position C-11 as  $\delta_C$  27.36. Additionally, the positions of C-6 to C-10 were assigned in ascending order of resonances as  $\delta_C$  29.49, 29.59, 29.63, 29.66 and 29.89, respectively and was possible to see each other correlations between the methylene protons at 1.24 ppm with the assigned carbons. Moreover, there were visible weak correlations to the methylene carbon C-12 and methine carbon C-13 at  $\delta_C$  27.97 and 39.05, respectively. Lastly, the C-4 and C-5 were observed at 25.64 and 37.50, respectively.

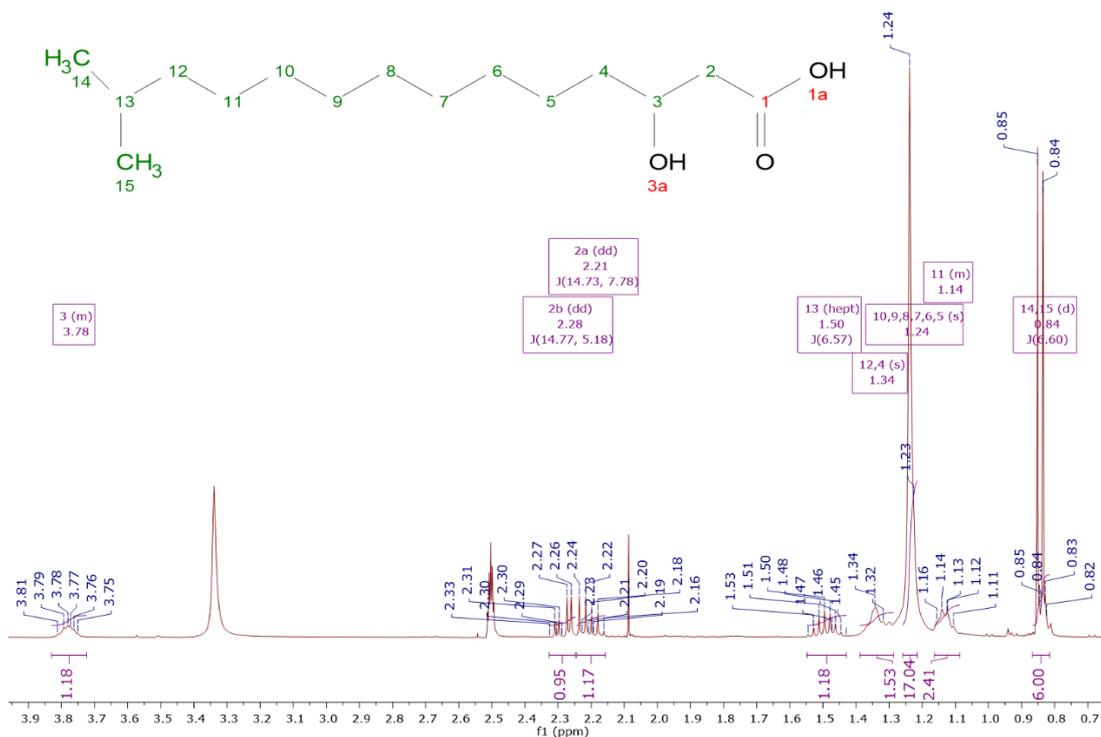


Figure 3-42  $^1\text{H}$  NMR spectrum of 3-hydroxy-13-methyltetradecanoic acid at 400 MHz in  $\text{DMSO-d}_6$ .

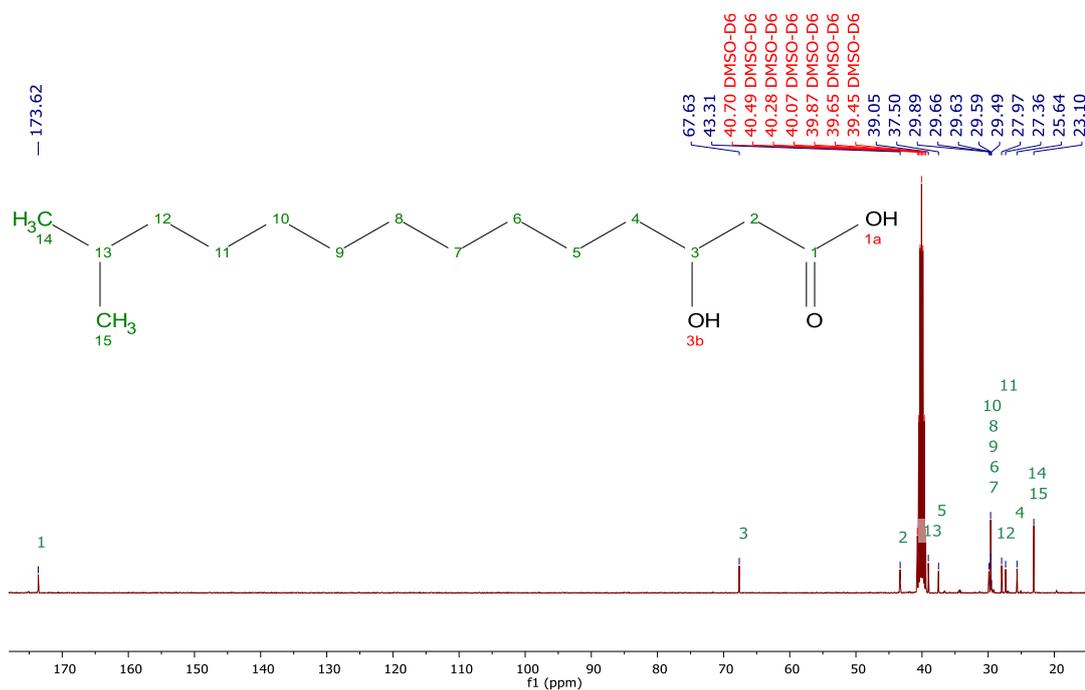


Figure 3-43  $^{13}\text{C}$  spectrum of 3-hydroxy-13-methyltetradecanoic acid at 100 MHz in  $\text{DMSO-d}_6$ .

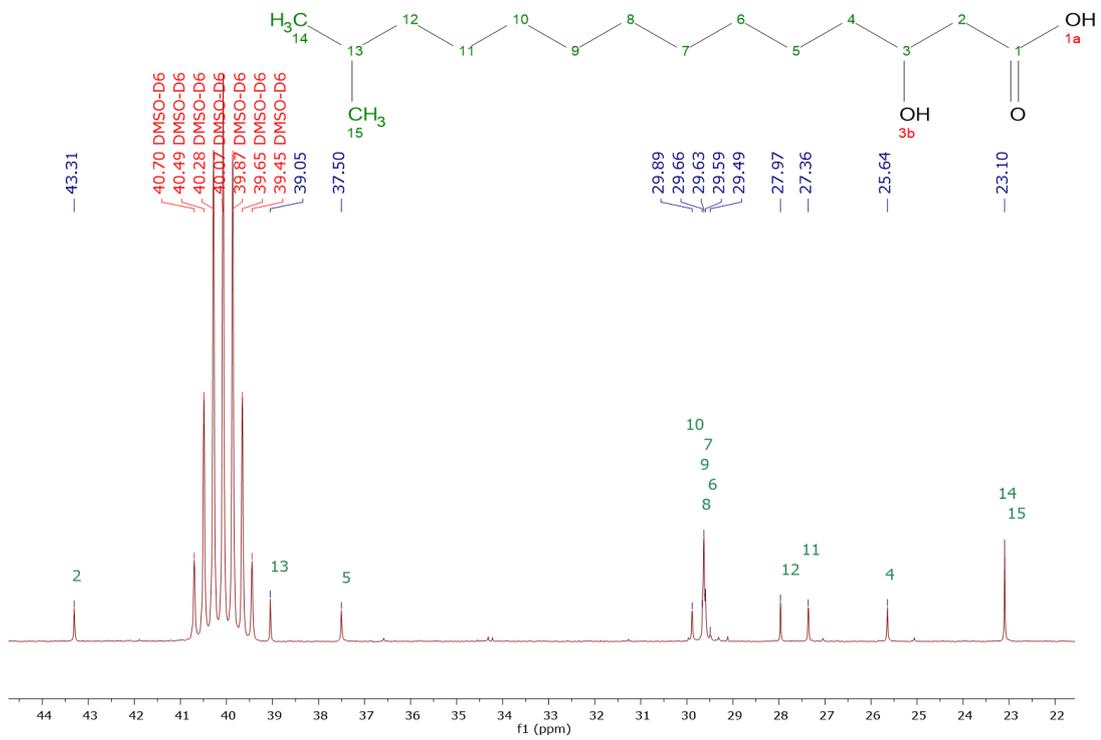


Figure 3-44 Expansion of  $^{13}\text{C}$  spectrum at 100 MHz of 3-hydroxy-13-methyltetradecanoic acid in  $\text{DMSO-d}_6$ .

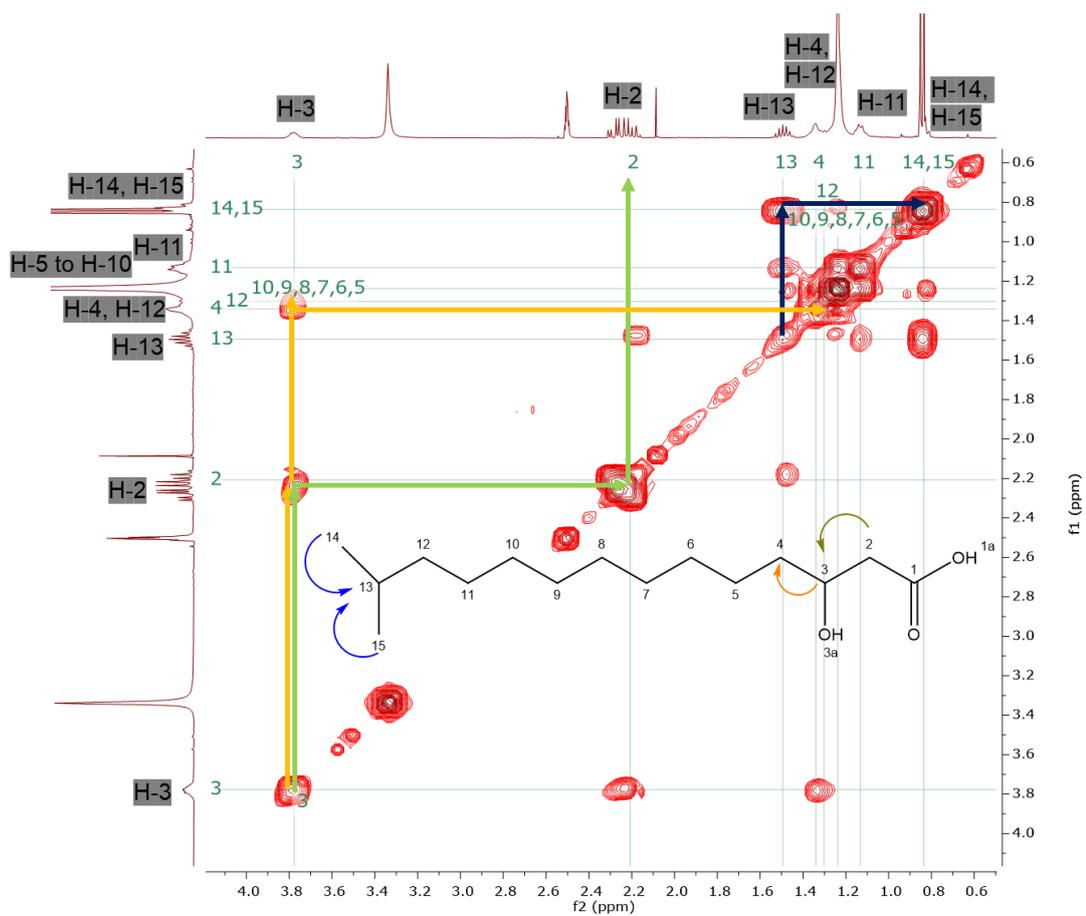
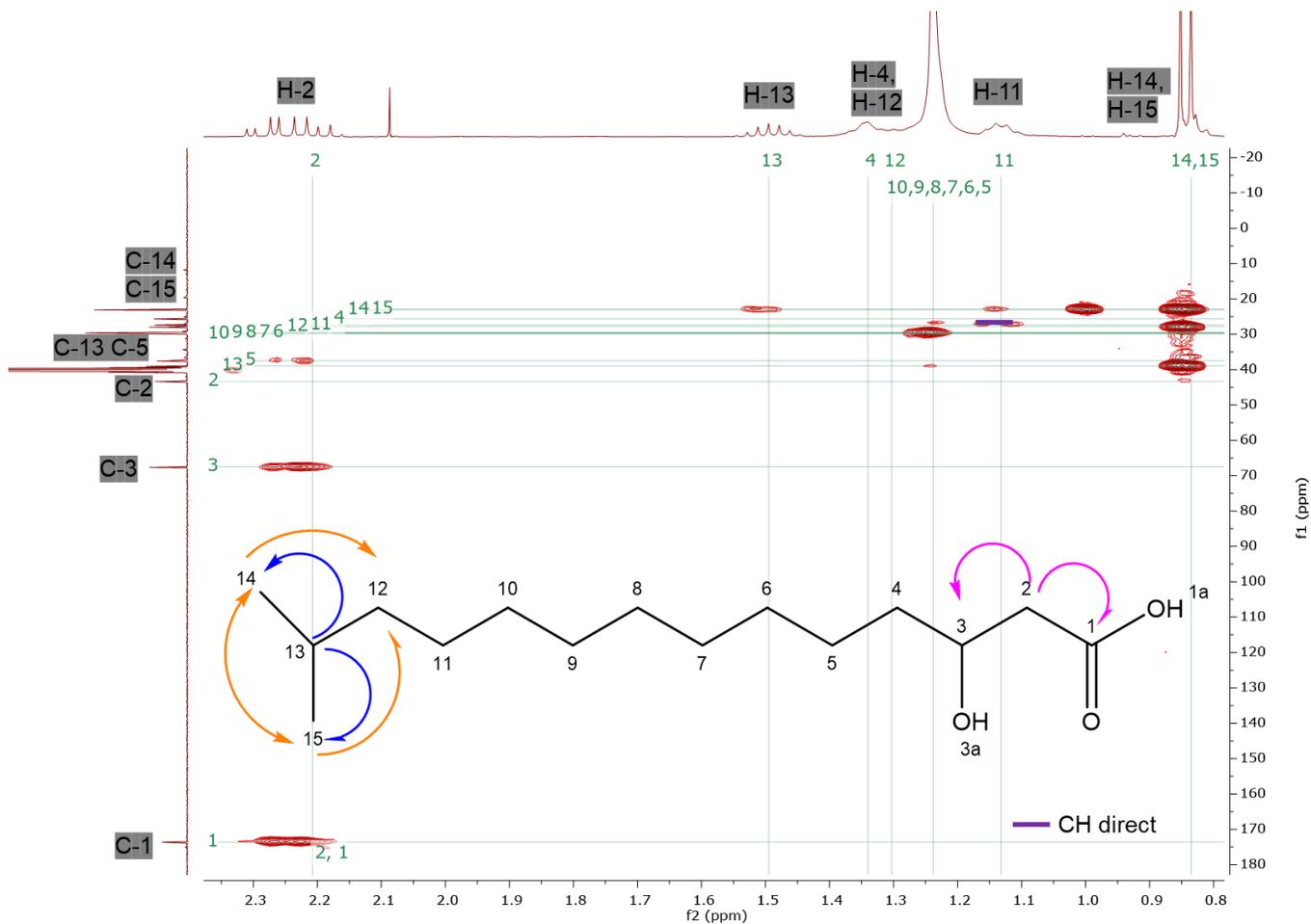


Figure 3-45  $^1\text{H}$ - $^1\text{H}$  COSY spectrum of 3-hydroxy-13-methyltetradecanoic acid 400 MHz in  $\text{DMSO-d}_6$  showed partial correlations in the substructure.



**Figure 3-46** HMBC of 3-hydroxy-13-methyltetradecanoic acid at 400 MHz. The X and Y-axes correspond to the proton  $^1\text{H}$  and carbon spectra  $^{13}\text{C}$ , respectively.

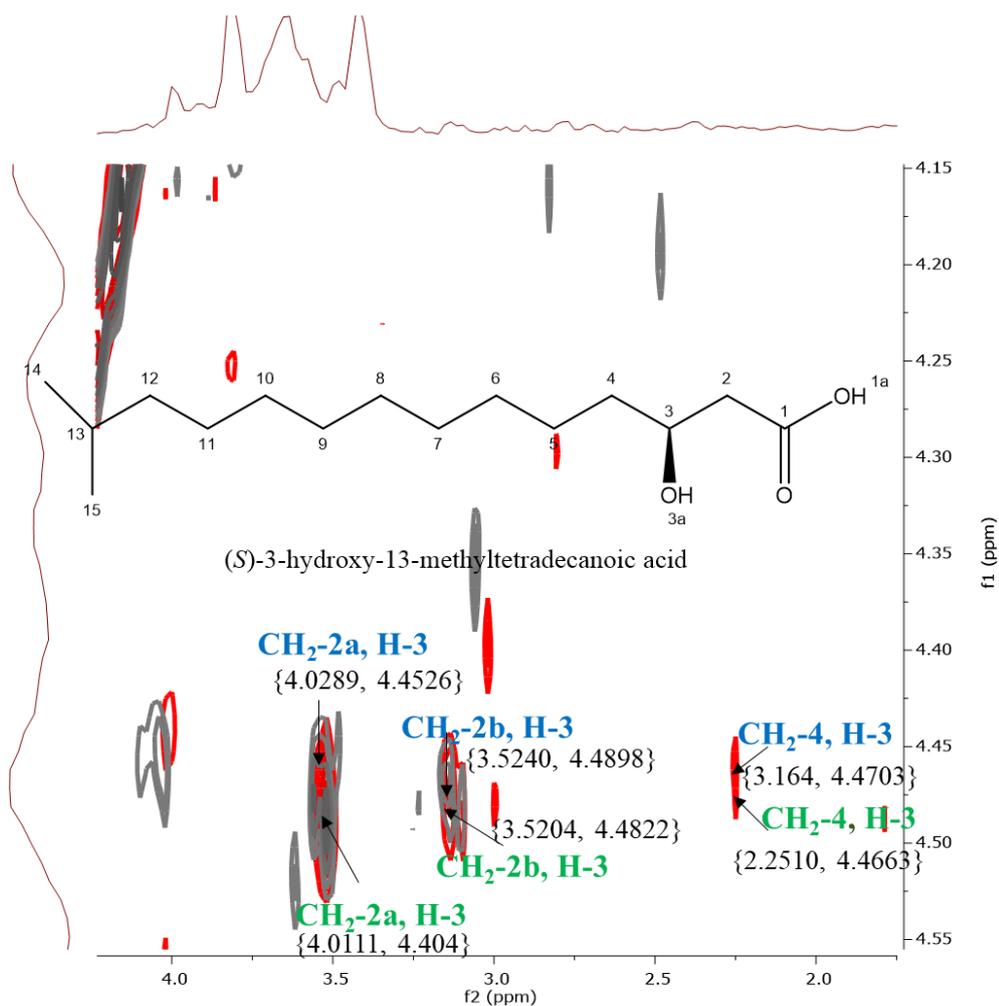
**Table 3-16**  $^1\text{H}$  NMR (400 MHz) and  $^{13}\text{C}$  NMR (100 MHz) of 3-hydroxy-13-methyltetradecanoic acid in DMSO- $d_6$  comparison with literature data (Hashizume et al., 2004, Yanai and Hiramoto, 1999). Full assignment was deduced from HMBC (DMSO- $d_6$ , 400 MHz).

Atom No.	3-hydroxy-13-methyltetradecanoic acid in DMSO- $d_6$	Literature (Hashizume et al, 2004) in $\text{CDCl}_3$	Literature (Yanai and Harimoto, 1999) in $\text{CDCl}_3$
	$\delta_{\text{H}}$ (multiplicity, $J$ in Hz) (400 MHz)	$\delta_{\text{C}}$ (multiplicity) (100 MHz)	$\delta_{\text{H}}$ (multiplicity, $J$ in Hz) (400 MHz)
<b>1</b>		173.63 (C)	
<b>2A</b>	2.21(dd, $J=14.73$ , 7.78)	43.31 ( $\text{CH}_2$ )	2.31(dd, $J=16.6$ , 9.0)
<b>2B</b>	2.28(dd, $J=14.77$ , 5.18)		2.40(dd, $J=16.6$ , 3.2)
<b>3</b>	3.78(m)	67.63 (CH)	3.96(m)
<b>4</b>	1.34 (s)	25.64( $\text{CH}_2$ )	} 1.09-1.19(m)
<b>5</b>		37.50( $\text{CH}_2$ )	
<b>6</b>	} 1.24(s)	29.49( $\text{CH}_2$ )	} 1.04-1.52(m)
<b>7</b>		29.59( $\text{CH}_2$ )	
<b>8</b>		29.63( $\text{CH}_2$ )	
<b>9</b>		29.66( $\text{CH}_2$ )	
<b>10</b>		29.89( $\text{CH}_2$ )	} 1.19-1.39(m)
<b>11</b>	1.14(m)	27.36( $\text{CH}_2$ )	
<b>12</b>	1.34 (s)	27.97 ( $\text{CH}_2$ )	1.09-1.19(m)
<b>13</b>	1.50(hept, $J=6.57$ )	39.05(CH)	1.39-1.60(m)
<b>14</b>	0.84 (d, $J=6.20$ )	23.10( $\text{CH}_3$ )	0.78(d, $J=6.6$ )
<b>15</b>	0.84 (d, $J=6.60$ )	23.10( $\text{CH}_3$ )	0.78(d, $J=6.6$ )
<b>1a</b>			
<b>OH</b>			
<b>3a</b>			
<b>OH</b>			

The literature presented different values for optical rotation, depending if it was a *R* or *S* configuration, ranging from  $[\alpha]_{\text{D}}^{25}-170$  (c 0.1 MeOH) to  $[\alpha]_{\text{D}}^{25}+13.6$  (c 1.05  $\text{CHCl}_3$ ) (Hashizume et al., 2004, Zhang et al., 2008b, Hiramoto et al., 1996). The optical rotation was  $[\alpha]_{\text{D}}^{20}+68$  (c 0.05 MeOH), giving a *S* configuration, (3*S*)-3-hydroxy-13-methyltetradecanoic acid, which was confirmed by Mosher experiments. The stereochemistry of the chiral centre C-3 was achieved by observing the changes in the

chemical shifts of the neighbouring protons and by comparing the bis-(R)-MTPA-Cl and the bis-(S)-MTPA-Cl derivatives of the 3-hydroxy-13-methyltetradecanoic acid. The chemical shift of the methylene (CH<sub>2</sub>-2a) and (CH<sub>2</sub>-2b) was shifted downfield  $\delta_{\text{H}}$  4.0111 and  $\delta_{\text{H}}$  3.5204 upon reacting with bis-(R)-MTPA-Cl reagent and to  $\delta_{\text{H}}$  4.0289 and  $\delta_{\text{H}}$  3.5240 upon reacting with the bis-(S)-MTPA-Cl reagent. The  $\Delta\delta^{\text{SR}}(\delta\text{S}-\delta\text{R})$  value for CH<sub>2</sub>-2a and CH<sub>2</sub>-2b was calculated to be +0.0178 and +0.0036, respectively. The chemical shift of the methylene (CH<sub>2</sub>-4) protons was shifted downfield to  $\delta_{\text{H}}$  2.2510 upon reacting with bis-(R)-MTPA-Cl reagent and to  $\delta_{\text{H}}$  3.1464 upon reacting with the bis-(S)-MTPA-Cl reagent (Figure 3.47). The  $\Delta\delta^{\text{SR}}(\delta\text{S}-\delta\text{R})$  value for CH<sub>2</sub>-4 was calculated to be +0.8954. The positive value marked the position of the methylene CH<sub>2</sub>-2a, CH<sub>2</sub>-2b and CH<sub>2</sub>-4 at the right of the chiral centre C-3, establishing a *S* configuration.

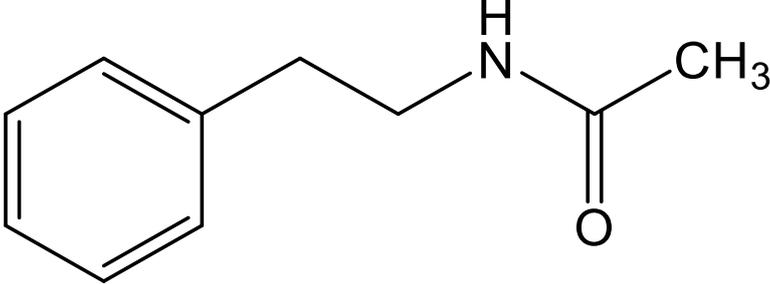
This compound was first isolated from *Flavobacterium meningosepticum*, a Gram-negative bacterium from the *Flavobacteriaceae* family, in 1976 described by Yano et al (Yano et al., 1976), interestingly, the same family as *Muricauda ruestringensis* (SBT531).



**Figure 3-47 Superimposed <sup>1</sup>H-<sup>1</sup>H COSY NMR (400 MHz) spectrum for MPTA derivatives of 3-hydroxy-13-methyltetradecanoic acid.** Measured 48 hours after reaction with both bis- (*R*) - MTPA-Cl and bis- (*S*)-MTPA-Cl reagents, measured in Pyridine-*d*<sub>5</sub>. Blue labels are for couplings upon reaction with bis- (*S*)-MTPA-Cl, while green labels are for the couplings upon reaction with bis- (*R*) - MTPA-Cl.

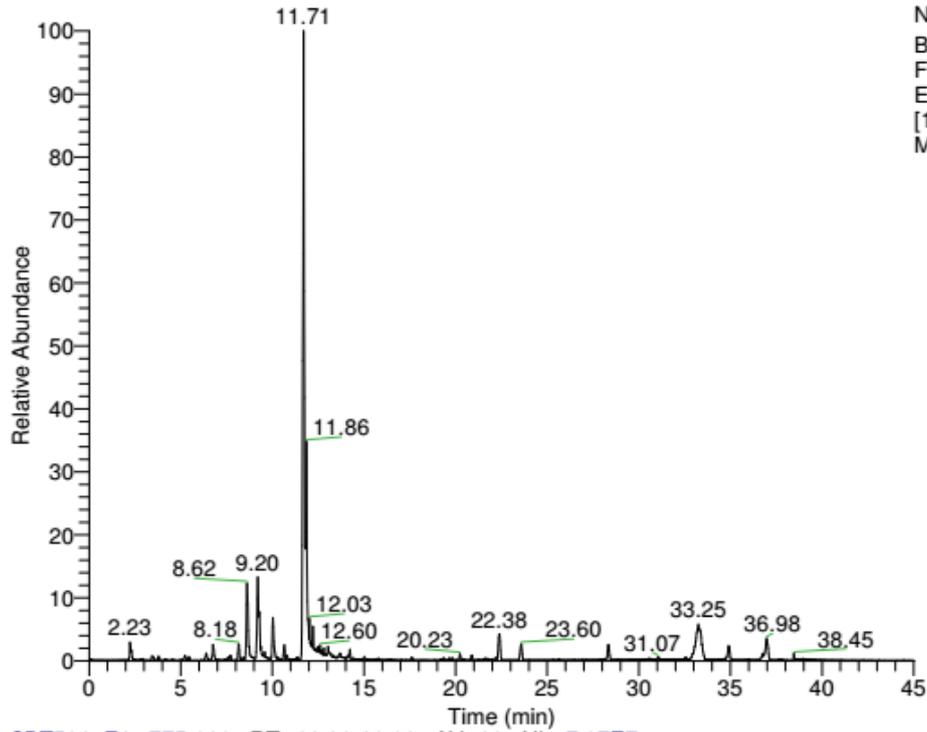
### 3.6.4 N-Phenethylacetamide

Table 3-17 N-phenethylacetamide.

<p><b>Compound name: N-phenethylacetamide (known)</b></p> <p>Fraction: 4 subFraction 2 Retention time (min): 11.71 Synonym(s):</p> <ul style="list-style-type: none"><li>• N-(2-phenylethyl)acetamide</li><li>• N-acetyl-2-phenylethylamine</li></ul> <p>Source: <i>Muricauda ruestringensis</i> (SBT531) Amount of sample: 7.3 mg Percent yield: 6% Physical description: brownish oily sample Molecular formula: C<sub>10</sub>H<sub>13</sub>NO Molecular weight: 163.22 g/mol Exact mass: 163.1070</p>
 <p>The image shows the chemical structure of N-phenethylacetamide. It consists of a benzene ring attached to a two-carbon ethyl chain. The terminal carbon of this chain is bonded to a nitrogen atom, which is also bonded to a hydrogen atom and an acetyl group (a carbonyl group bonded to a methyl group).</p>

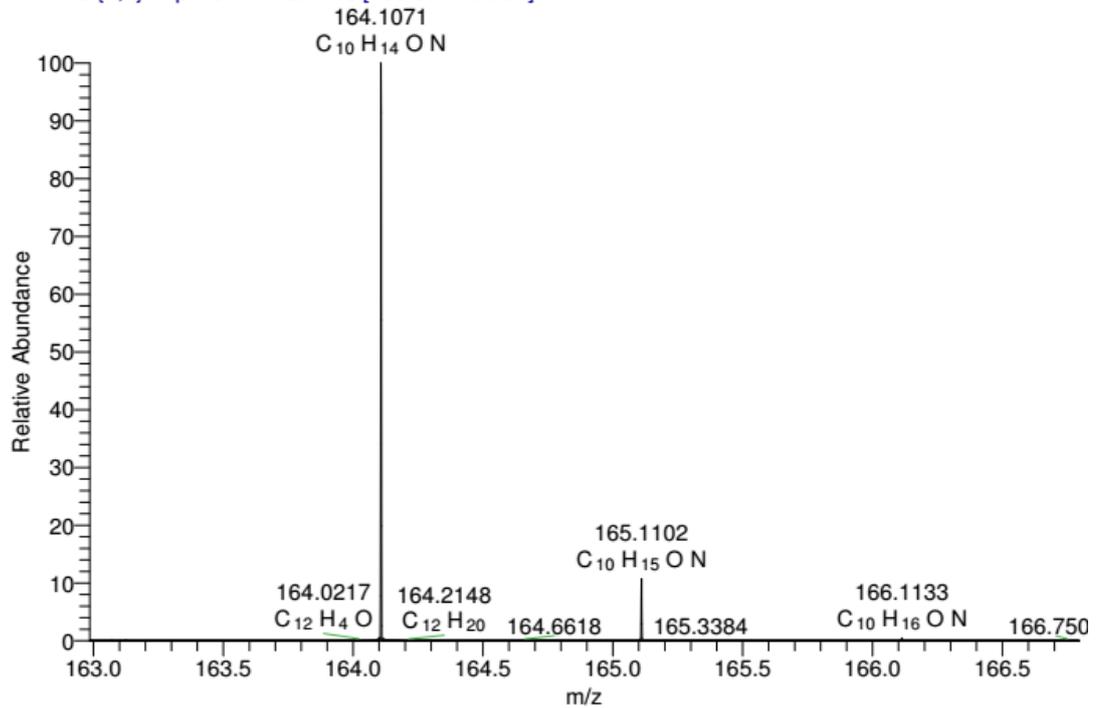
# LC – HRFTMS spectra [M+H]<sup>+</sup>

RT: 0.00 - 45.03



NL: 5.41E8  
Base Peak F:  
FTMS {1,1} + p  
ESI Full ms  
[100.00-2000.00]  
MS SBT531\_F4

SBT531\_F4 #775-839 RT: 11.36-12.22 AV: 33 NL: 7.87E7  
T: FTMS {1,1} + p ESI Full lock ms [100.00-2000.00]

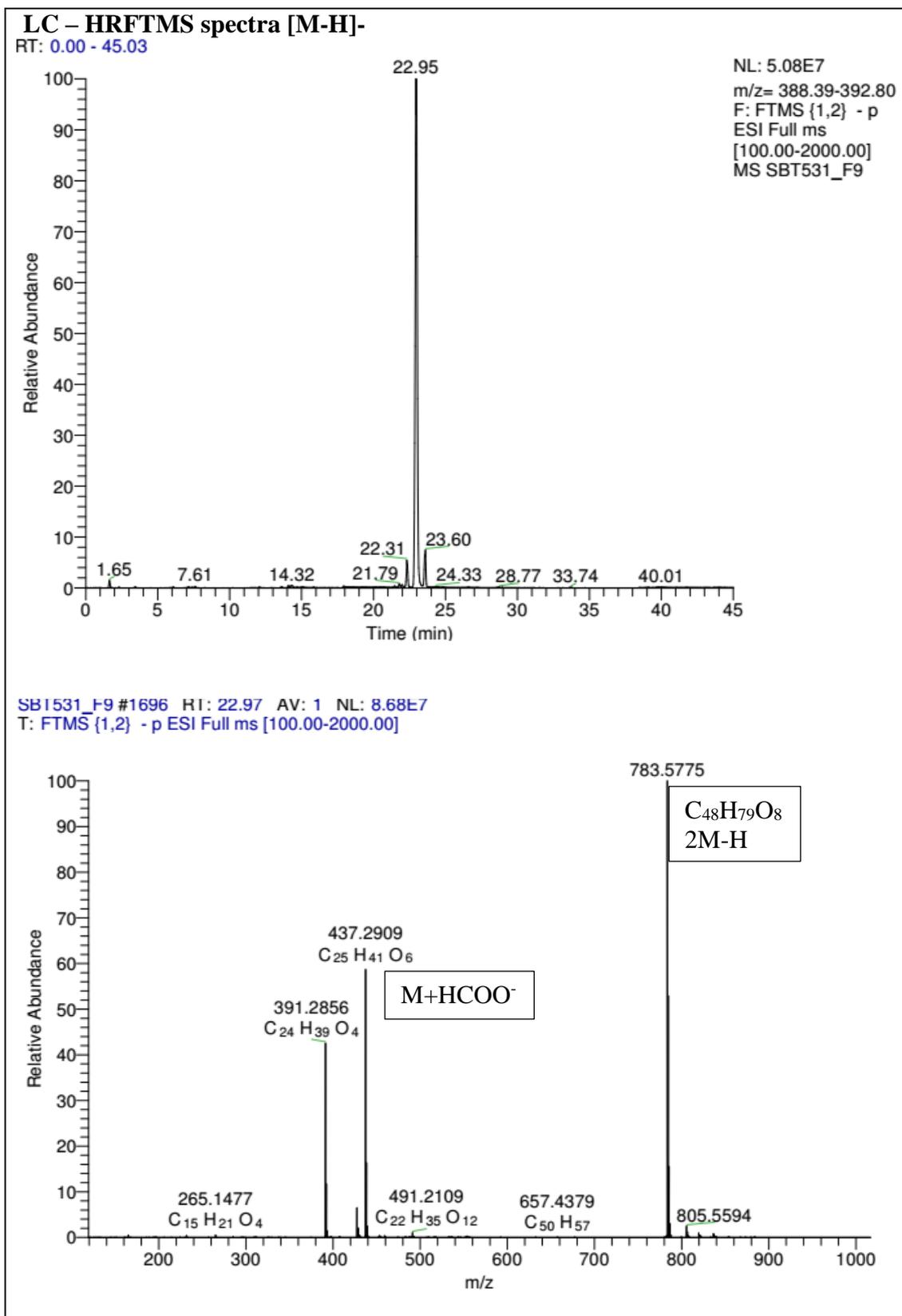


*N*-phenethylacetamide was isolated from fraction 4 subfraction 2 with a yield of 6.17% (7.3 mg). Based on the high-resolution mass spectral data, the ESI peak eluted at 11.71 min yielded an ion peak at  $m/z$  164.1071  $[M+H]^+$  in the positive mode. This suggested the molecular weight of 163.22 g/mol, which established the molecular formula of  $C_{10}H_{13}NO$  (Table 3.17). In comparison to the isolated saturated fatty acids, *N*-phenethylacetamide was found to be inactive in the assays performed in this study. The NMR spectra data used to elucidate the structure is present on the Appendices I.

### 3.6.5 Cholic acid derivative

**Table 3-18 Cholic acid derivative.**

<p><b>Compound name (known):</b> cholic acid derivative</p> <p>Fraction: 9 subFraction 13</p> <p>Retention time (min): 22.95</p> <p>Synonym(s):</p> <ul style="list-style-type: none"> <li>(4R)-4-[(3R,5R,8R,9S,10S,12S,13R,14S,17R)-3,12-dihydroxy-10,13-dimethyl-2,3,4,5,6,7,8,9,11,12,14,15,16,17-tetradecahydro-1H-cyclopenta[a]phenanthren-17-yl]pentanoic acid</li> </ul> <p>Source: <i>Muricauda ruestringensis</i> (SBT531)</p> <p>Amount of sample: 15.5 mg</p> <p>Percent yield: 10 %</p> <p>Physical description: brownish oily sample</p> <p>Molecular formula: <math>C_{24}H_{40}O_4</math></p> <p>Molecular weight: 392.58 g/mol</p> <p>Exact mass: 392.2927</p>



The cholic acid derivative was isolated from fraction 9 subfraction 13 with a yield of 9.74% (15.5 mg). Based on the high-resolution mass spectral data, the ESI peak eluting

at 22.95 min gave an ion peak at  $m/z$  391.2856  $[M-H]^-$  in the negative mode. This suggested the molecular weight of 392.58 g/mol, which established the molecular formula of  $C_{24}H_{40}O_4$  (Table 3.18). Cholic acid was found inactive in all the assays accomplished in this study. The NMR spectra data used to elucidate the structure is present on the Appendices II.

### 3.7 Bioreactor fermentation of SBT531

A large scale-up of SBT531 on the bioreactor was performed by Dr Christina Viegelmann in SIPBS at the Fermentation Centre during the EU-FP7 programme. Throughout the bioreactor experiments, variable measurements associated with microbial growth conditions were changed and controlled, which included: the amount of dissolved oxygen in the liquid phase (%), temperature ( $^{\circ}C$ ), nutrient concentrations (g/L), pH, agitation (rpm), aeration (vvm), medium salinity (%), etc. It was determined that SBT531 could be cultivated at  $37^{\circ}C$  in the bioreactor, instead of at  $45^{\circ}C$  as in the shake flasks experiments. Table 3.19 shows the process variables of the experiments performed on SBT531.

**Table 3-19 - Parameters of experiments performed on SBT531**

Experiment Number	Scale	Medium	Temperature	pH	DO	Agitation	Aeration
01MUR	1-L	Difco MB	37 $^{\circ}C$	Uncontrolled	Uncontrolled	120rpm	1vvm
				7.5			
02MUR	1-L	Difco MB (12 mL bolus of 5g Bacto Peptone + 1 g Yeast Extract added at 24 hours)	37 $^{\circ}C$	7.5	Uncontrolled	120rpm	1vvm
03MUR	1-L	Difco MB (Feed of 5g Bacto Peptone + 1g Yeast Extract added per day with a flow rate of 5 mL/hr)	37 $^{\circ}C$	7.5	Uncontrolled	120rpm	1vvm

---

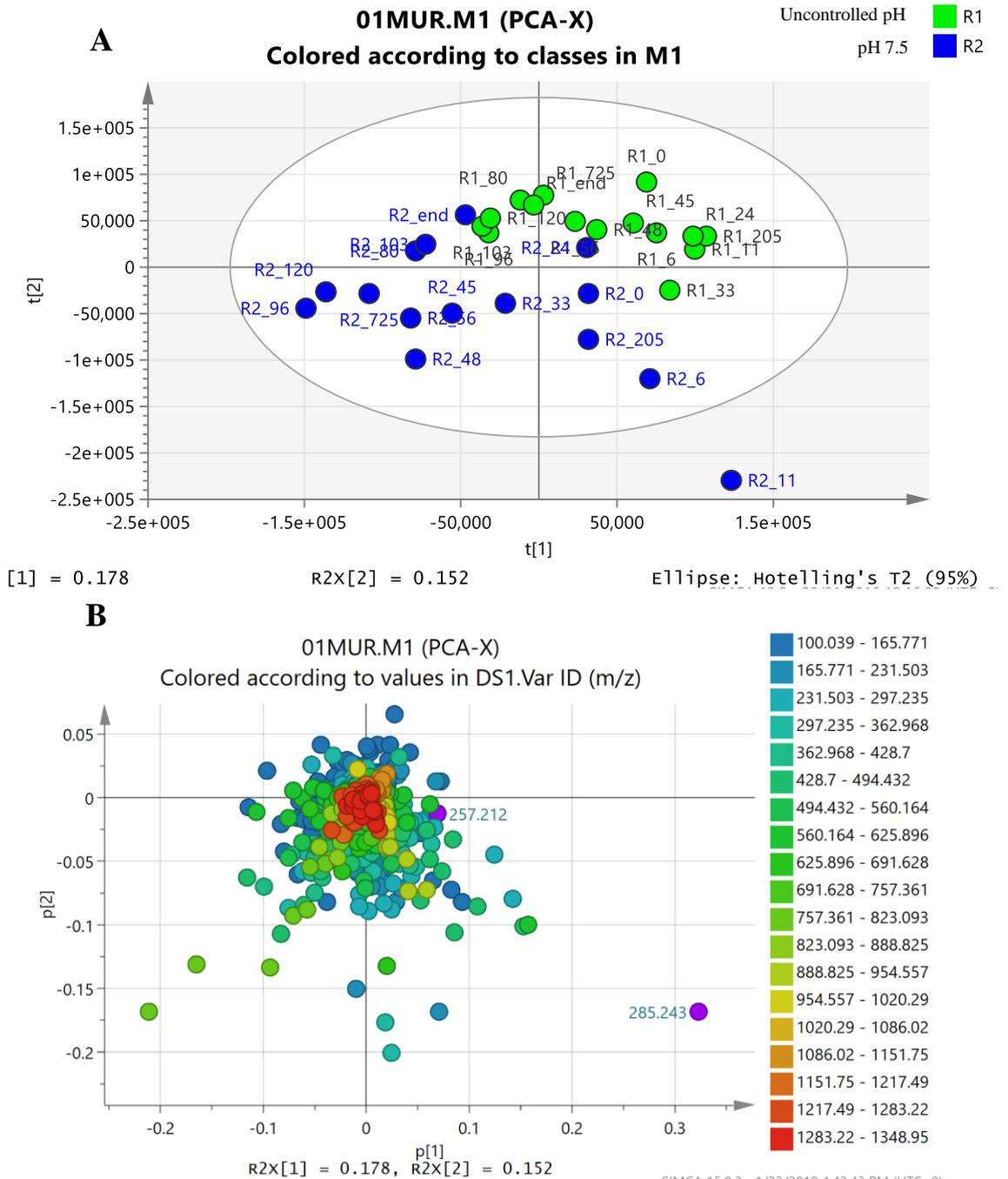
starting at 24  
hrs)

04MUR (OmniLog cell density experime nt)	96- well plate	Difco MB	37°C	Uncontr olled	Uncontrolled	Uncontr olled	Uncont rolled
05MUR	10-L	Difco MB (120 mL bolus of 50g Bacto Peptone + 10 g Yeast Extract added at 24 hours)	37°C	7.5	Uncontrolled	120rpm	1vvm

---

The first bioreactor experiment on SBT531 (01MUR) was performed to compare the growth of the culture when the pH was controlled or uncontrolled. In Reactor 1, the pH was not controlled whereas in Reactor 2, the pH was controlled at 7.5, which was the pH of the inoculum. The growth of the cultures was measured by OD and dry cell weight (DCW). It was possible that the measurements were affected by the production of the brown precipitate, which was also detected in the shake flask cultures. Due to the constant agitation of the cultures, the brown precipitate was not observed in the bioreactors but became more obvious when the DCWs were measured.

Metabolomic analysis using mass spectrometry on the samples from Reactors 1 and 2 revealed that there was indeed a difference in the metabolic profiles of the bacteria when the pH was controlled compared to when it was uncontrolled, as seen in the PCA plot in Figure 3.48A. Three target compounds were identified: aseanostatin P6 ( $C_{15}H_{30}O_2$ ,  $m/z$  241.2173 [M-H]<sup>-</sup>), and the two derivatives of aseanostatin ( $C_{15}H_{30}O_3$ ,  $m/z$  257.2122 [M-H]<sup>-</sup> and  $C_{17}H_{34}O_3$ ,  $m/z$  285.2435 [M-H]<sup>-</sup>) that were also isolated from the shake flask cultures (Figure 3.48B).



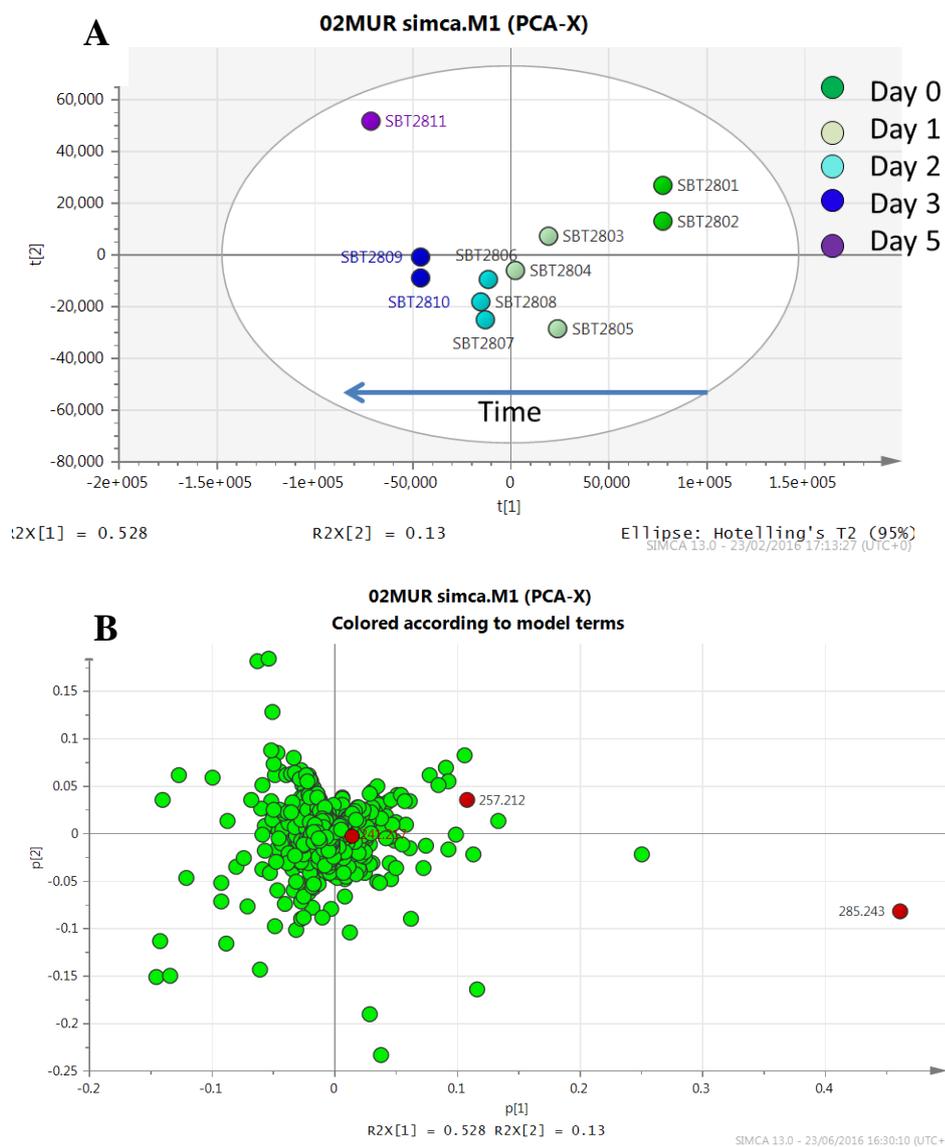
**Figure 3-48 Principal component analysis (PCA) of metabolic profiles of *M. ruestringensis*.** (A) PCA scores plot showing the clustering of samples from Reactor 1 (green), where pH was uncontrolled, and Reactor 2 (blue) where pH was controlled at 7.5. (B) PCA loadings plot coloured according to *m/z*. The target metabolites (285.243 [M-H]<sup>-</sup>, and 257.212 [M-H]<sup>-</sup>) are highlighted in purple.

The two principal components that caused the most variation in the samples, appear to be time (t[1]) and the bioreactors (t[2]). All the samples from Reactor 1 are in the upper half of the plot, whereas those from Reactor 2 are predominantly in the bottom half of the plot. The samples in the right side of the plot are primarily the samples from earlier fermentation time (0-20.5 hours) whereas the later samples, particularly from Reactor 2, are in the left side of the plot. Although not as evident for Reactor 1, it is still clear that the samples taken at earlier time points are on the right side whereas those taken towards the end of the fermentation are on the left side of the scores plot. The loadings plot (Figure 3.48B) shows the targeted metabolites in purple circles. The aseanostatin derivative with  $m/z$  285.2435 [M-H]<sup>-</sup> was in the lower right quadrant and clearly was predominant in Reactor 2. However, three peaks were detected in the chromatogram having a mass of  $m/z$  257.2122 [M-H]<sup>-</sup>. This suggests that different isomers of this compound are being produced in the bioreactors, (3R)-3-hydroxy-13-methyltetradecanoic acid or/and (3S)-3-hydroxy-13-methyltetradecanoic acid. The derivative with an  $m/z$  of 285.2435 [M-H]<sup>-</sup> had a higher intensity compared to the derivative with an  $m/z$  of 257.2122 [M-H]<sup>-</sup>. Aseanostatin P6, despite being a major metabolite according to the NMR data, ionised poorly. Based on the peak area, it was evident that the compounds were produced in greater quantities in Reactor 1, when pH was not controlled.

The second experiment, 02MUR, used the same parameters as Reactor 2 of 01MUR, in which the pH was controlled at 7.5, as the growth (OD<sub>600</sub>) was higher in this bioreactor. A 12-mL bolus containing 5g of bacteriological peptone and 1g of yeast extract was added at 24 hours, which was the end of the exponential phase. A 24-hour sample was obtained prior to the addition of the bolus and another sample was taken at 25 hours after the addition of the bolus.

Metabolomic analysis of the extracts was performed using LC-MS and multivariate analysis. The PCA scores plot (Figure 3.49A) showed a clear trend as time progresses, indicating that the metabolic profile of the bacteria changes each day. The PCA loading plot (Figure 3.49B) shows which metabolites are predominant in the different days.

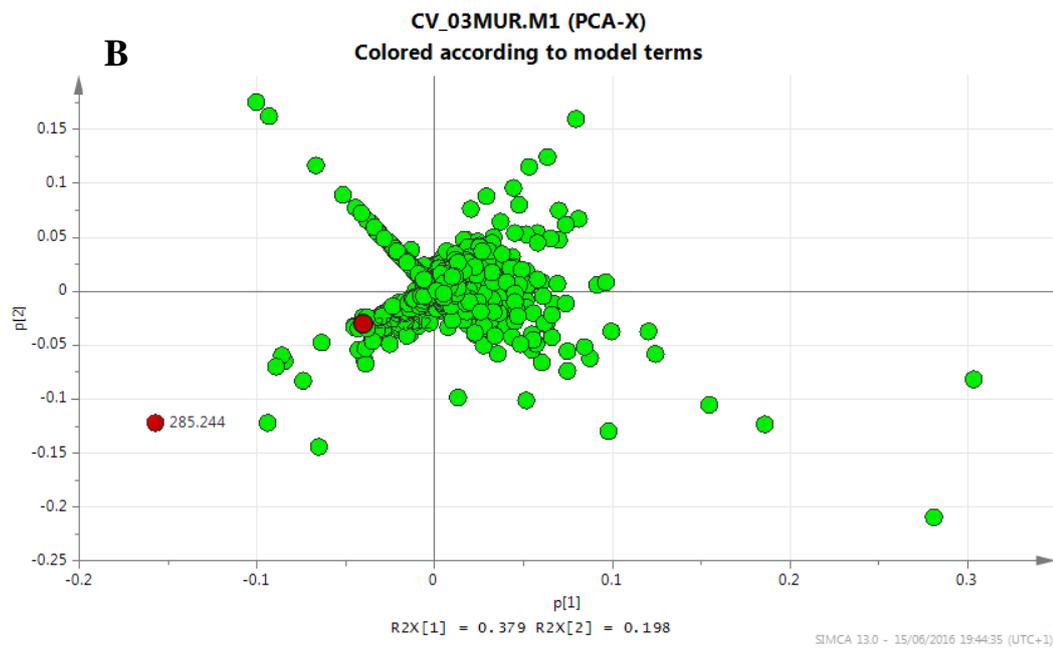
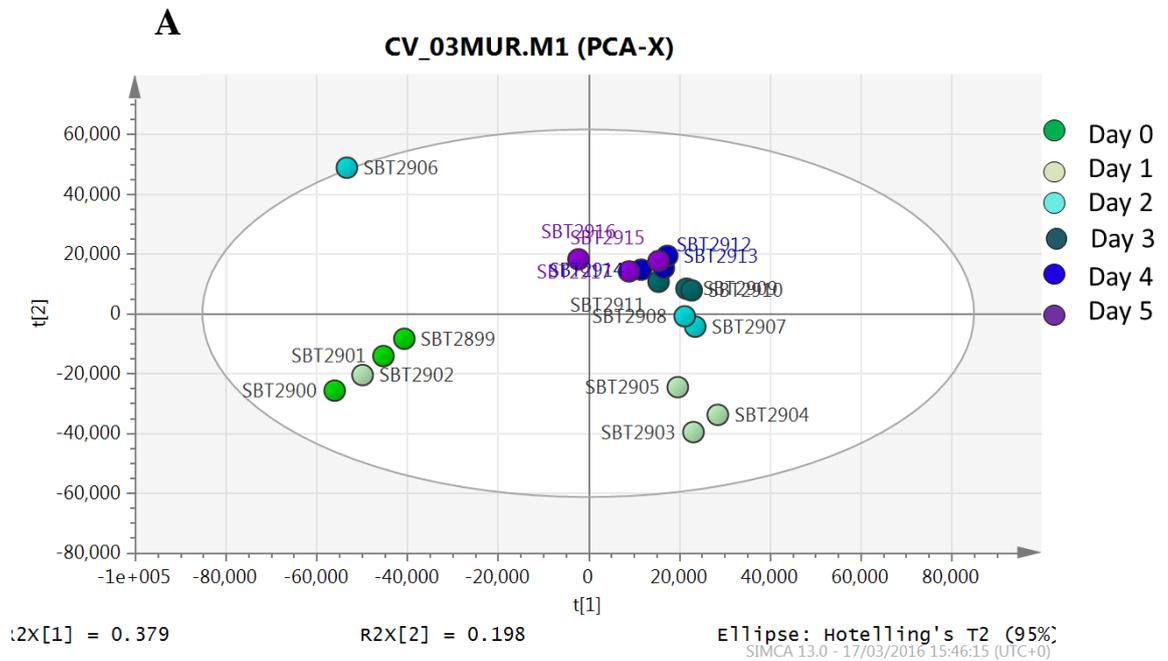
The target metabolites, highlighted in red, are clearly produced early in the fermentation.



**Figure 3-49 Principal component analysis (PCA) of the metabolomic profiles of bacteria 02MUR extracts. (A) PCA scores plot of 02MUR extracts coloured by the day the sample was taken, and (B) PCA loading plot highlighting the target compounds in red.**

The next experiment, 03MUR, was like 02MUR except the additional media components were introduced more slowly into the reactor at a rate of 5g of bacteriological peptone and 1g of yeast extract per 24 hours, which is the length of the exponential phase. The feeding of the fresh medium did not significantly increase the growth of SBT531 compared to when the bolus was added in 02MUR. It is possible that this is due to changes in the osmolarity of the broth, as no marine salts were fed into the reactor. Previous studies on *M. ruestringensis* (Bruns *et al.*, 2001) revealed that it grows at salt concentrations of 0.5-9.0% NaCl with an optimum growth at 3%. The continuous feeding of the peptone and yeast extract at 5 mL/hr could have resulted in the dilution of the salt present in the broth. Metabolomic analysis of the extracts was performed with the PCA scores plot (Figure 3.50A) showing a clear trend as time progresses. The PCA loading plot (Figure 3.50B) showed again that the targeted metabolites were being produced in earlier time points of fermentation.

The PCA plot in Figure 3.50A shows that extracts obtained from Day 0 and Day 1 form distinct clusters, but the Day 2 until Day 5 extracts all clustered together, except for C<sub>17</sub>H<sub>34</sub>O<sub>3</sub> (48 hours) which was an outlier. The target metabolites, aside from aseanostatin P6 which was unable to be detected in the MS due to its poor ionisation, were found in the loading plot (Figure 3.50B) in the lower left quadrant, indicating that they were predominant in the earlier samples. The production of the target metabolites always occurred during the exponential growth phase and appeared largely unaffected by the addition of fresh media to the cultures. The Difco<sup>®</sup> Marine Broth (MB) used for SBT531 did not have any carbon source aside from the sugars present in peptone and yeast extract. It was hypothesised that the addition of a carbon source would extend the exponential growth phase of SBT531, resulting in increased product yields.



**Figure 3-50 Principal component analysis (PCA) of the metabolomic profiles of bacteria 03MUR extracts.** (A) PCA scores plot of 03MUR extracts coloured by the day each sample was taken, and (B) PCA loading plot highlighting the target compounds, m/z 285.243 and 257.2122 [M-H]<sup>-</sup>.

Optimisation of the carbon source and osmolyte concentrations was attempted using phenotypic microarray testing (SOP#37) using the OmniLog (BIOLOG) (04MUR).

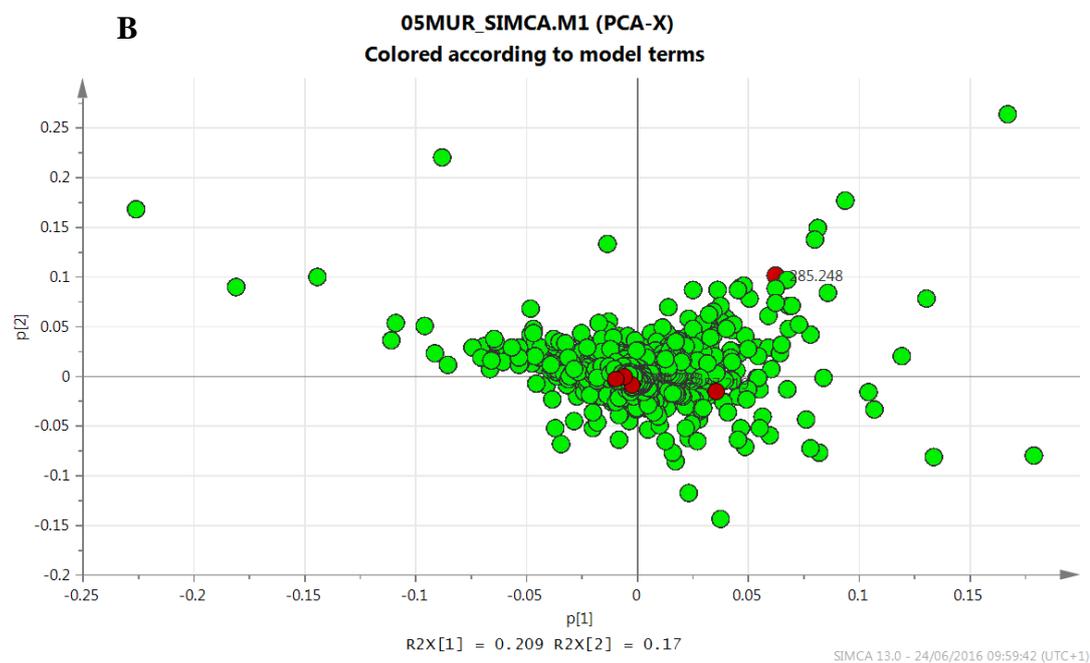
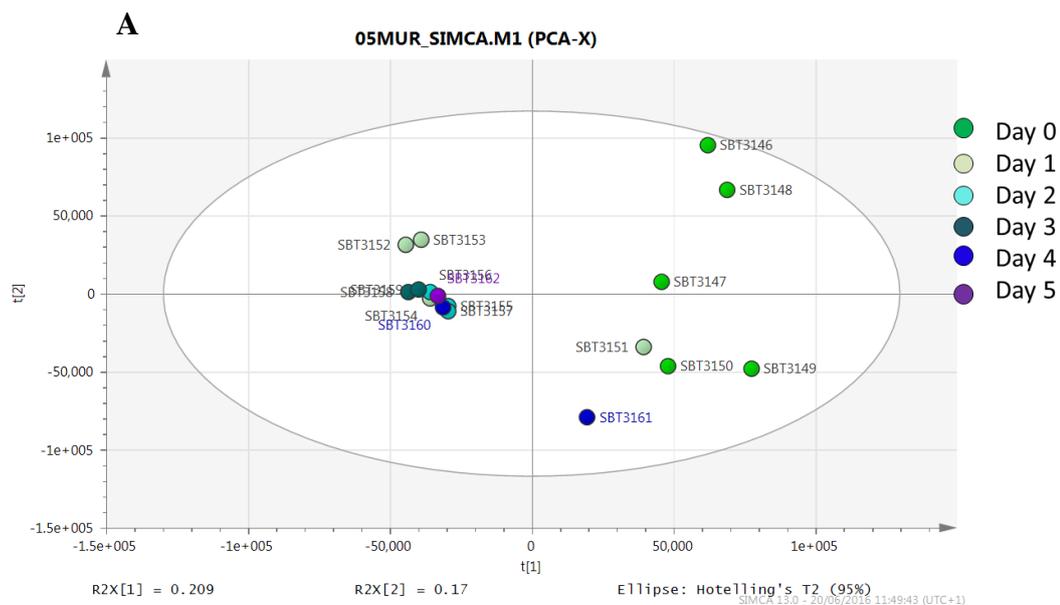
No growth was achieved in initial experiments using the PM9 microplate™ (osmolytes) and dye mixes A and H, following the protocol for Gram-negative organisms recommended by BIOLOG. Both dyes, A and H, were tested by growing SBT531 in Difco® MB with the addition of the dyes to determine if the dyes inhibited the growth of the bacteria. However, it was proven that the dyes did not affect the growth of SBT531. The inoculating fluid provided by BIOLOG was most likely not suitable for cultivation of SBT531. In the course of method development for cultivation of SBT531 using the OmniLog, the bacteria were grown in 20 mL solutions of 5g/L of bacteriological peptone and 1g/L of yeast extract containing varying quantities of salt (1% NaCl, 3% NaCl, 1% marine salts and 3.14% marine salts). No growth was observed in the broth containing NaCl but without the marine salts (MS), while only minimal growth was observed at 24 hours in 1% MS. The solution containing 3.14% MS was identical in composition to Difco® MB in that it was composed of the same quantities of bacteriological peptone and yeast extract in addition to the 3.14% marine salts, however, the growth achieved was less than that achieved in the Difco® MB. The reason for this is not yet known and requires further studies. Microbial growth requires several chemical elements and physical factors such as temperature, pH and osmotic pressure (Zengler and Zaramela, 2018, Schaechter, 2015). Regarding the salt concentration, it is clear from this project that SBT531 requires approximately 3.14% marine salts in order to grow, and that it requires salts other than NaCl. It is also known that it is a facultatively anaerobic microorganism that is capable of growth under both aerobic and anaerobic conditions and slightly halophilic (Bruns et al., 2001, Huntemann et al., 2012). Further experiments may be performed to determine the salts and the concentrations of salts necessary for the optimal growth of SBT531 as well as other obligatory requirements of growth of microorganism.

BIOLOG also recommended that different cell densities be tested in order to determine the best density for use in the 96-well plates. Commencing the experiment with a cell density that was too high resulted in all the redox dye being reduced too rapidly and the growth curve not being evident. A too-low cell density would result in a slow or no growth, as the dyes may sometimes inhibit the growth of the cells if the quantity of cells is too low (Bochner et al., 2011). SBT531 was cultivated in Difco® MB in a 96-

well plate at different cell densities. Purple colour was immediately observed in the wells containing Dye A and an OD<sub>590</sub> of 1. This cell density was evidently too high. Dye H gave better peak shapes in general, and a starting OD<sub>590</sub> of 0.032 gave a good growth curve as a small lag phase could be seen before the exponential phase. It was observed that wells in Row A and Row H had different peak shapes compared to the other replicates which look more similar to each other. This is likely to be due to evaporation of the solution at the edges of the plate, particularly as the plate was incubated for 120 hours at 37°C. This should be taken into consideration when planning future experiments.

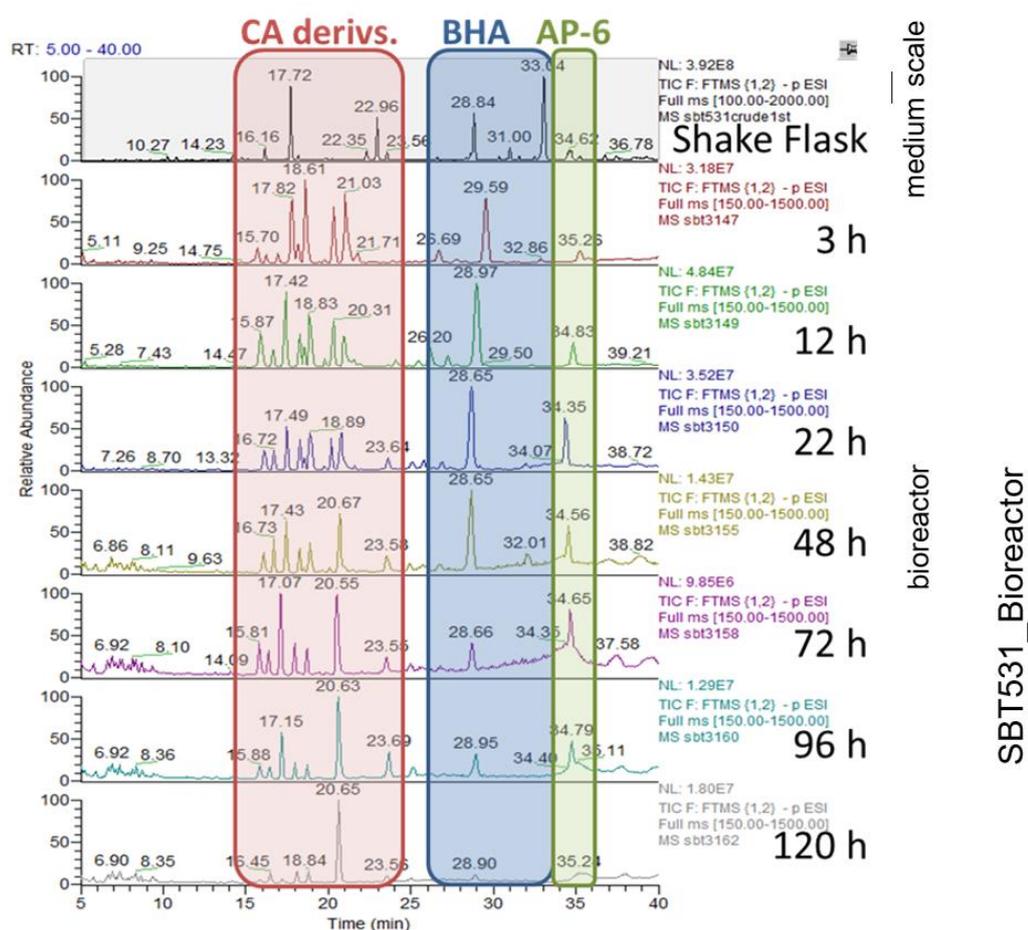
Although carbon source optimisation using BIOLOG technology was unable to be completed, SBT531 was upscaled to 10-L using the Biostat-C bioreactor (Sartorius). This experiment followed the same parameters as 02MUR, in that a bolus containing 50g of bacteriological peptone and 10g of YE was added at the end of the exponential phase. The growth of SBT531 in the 10-L culture was similar to the growth in the 1-L cultures, and the end of the exponential phase was also reached at 24 hours.

Analysis of the MS and NMR data showed that the target metabolites continued to be produced at the 10-L scale. Multivariate analysis of the MS data once again revealed the Day 0 extracts (SBT3146-3150) had distinct chemistries compared to the other extracts, which clustered together (Figure **3.51A**). This showed that the metabolic profile of the bacteria during the exponential phase is significantly different from the other growth phases, and that the addition of the peptone and YE at the end of the exponential phase does not result in a continuation of this metabolic profile, despite prolonging the exponential phase and increasing the biomass. The  $\beta$ -hydroxy fatty acids were produced in a similar pattern to the previous fermentations in the 1-L bioreactors (Figure **3.51B**).

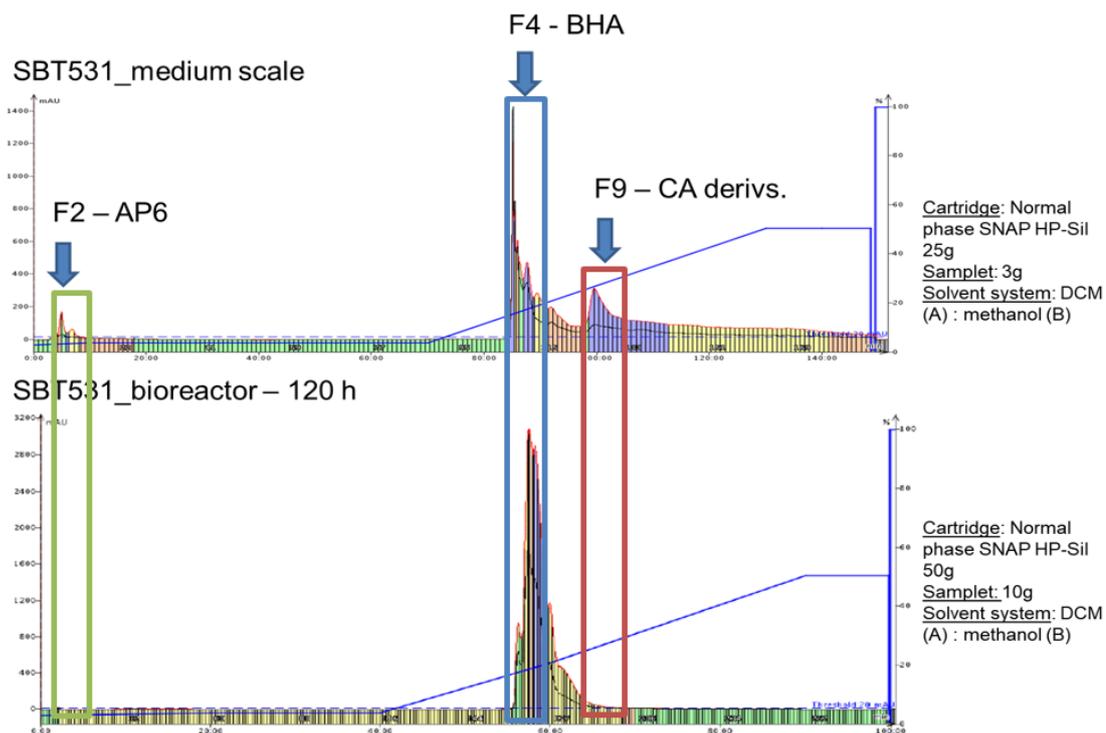


**Figure 3-51 - Principal component analysis (PCA) of the metabolomic profiles of bacteria 05MUR extracts. (A) PCA scores plot of 05MUR extracts coloured by the day the sample was taken, and (B) PCA loadings plot highlighting the target compounds.**

In the SBT531 shake flasks, the major compound being produced is aseanostatin P6. However, in the bioreactors, other derivatives of  $\beta$ -hydroxy fatty acids were being produced as the fermentation time progressed. In the medium scale, aseanostatin P6 was a major component as shown on the NMR but did not ionise well as observed on the HR-LC/MS data. The optimum yield of aseanostatin P6 occurred in the exponential phase during early times of fermentation between 24 to 48 hours and for  $\beta$ -hydroxy fatty acids between 12 to 48 hours, whereas the main production of cholic acid derivatives was obtained later time-point of fermentation after 72 hours. The SBT531 bioreactor at 120 hours, still showed the presence of aseanostatin P6 in small quantities, however together with the production of other components which included cholic acid derivatives and other  $\beta$ -hydroxy fatty acids (Figure 3.52 and 3.53).



**Figure 3-52 Comparison of the chromatograms (negative ionization-LC-HRMS data) of the SBT531 medium scale-up (shake flasks) and the different growth times (3,12,22,48,72,96 and 120h) of SBT531 large scale-up (bioreactor). Highlighted in green box the aseanostatin P6 (AP-6) peak, in blue box the beta hydroxy fatty acid derivatives (BHA) and in red the colic acid derivatives (CA derivs.).**



**Figure 3-53 Comparison of the chromatograms from Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification system of the SBT531 medium sale-up (shake flasks) and bioreactor at 120h. Highlighted in green box the aseanostatin P6 (AP-6) peak, in blue box the beta hydroxy fatty acid derivatives (BHA) and in red the colic acid derivatives (CA derivs.).**

## 3.8 Genomics profile of SBT531

### 3.8.1 Sequencing and genome annotations

Whole genome sequencing of *Muricauda ruestringensis* was performed using next-generation sequencing with Illumina short read technology (MicrobesNG, Birmingham) and establishing a minimum coverage of 30x. The reads were trimmed using Trimmomatic (Bolger et al., 2014) and the quality was assessed using *in-house* scripts combined with the following software: Samtools (Li et al., 2009), BedTools (Quinlan and Hall, 2010) and bwa-mem (Li and Durbin, 2009). Genome annotations were conducted in Prokka 1.11 (Seemann, 2014) and further quality assessment of the assembly in QCAST 3.1 (Gurevich et al., 2013) with evaluation of taxonomic (taxonomic distribution) sequence classification using Kraken (Wood and Salzberg, 2014).

Curation of genomic data (annotation) was ensured via the curation of subsystems in SEED/ RAST webserver (Overbeek et al., 2014) with further downstream mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000). Annotation and analysis of secondary metabolite biosynthesis gene clusters (BGC) in *M. ruestringensis* genome sequences was assessed using antiSMASH (Medema et al., 2011) webserver.

### 3.8.2 Analysis of secondary metabolites biosynthetic gene clusters

The genome assembling of *M. ruestringensis* strain SBT531 contains nine scaffolds / contigs with more than 1000bp with 3121 coding sequences (Table 3.20). Genome assembling metrics and quality control (QC) parameters and confirmation of taxonomic families, genera and species (Table 3.21).

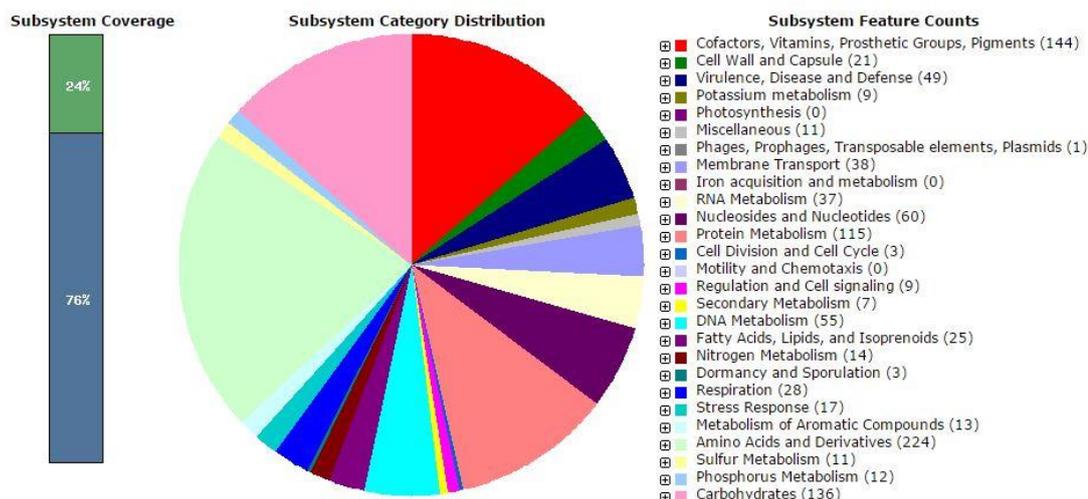
**Table 3-20 *M. ruestringensis* assembly metrics** All sequences in submitted assemblies are initially treated as 'contigs / scaffolds'. Each contig was split on runs of 25 or more 'N' characters to form 'contigs'. Any contigs that could be split in this way were regarded as 'scaffolded contigs' Scaffolds that didn't have runs of 25 or more 'N' characters were also counted as 'unscaffolded contigs'. N50 scaffold/contig length is calculated by summing lengths of scaffolds/contigs from the longest to the shortest and determining at what point you reach 50% of the total assembly size. The length of the scaffold/contig at that point is the N50 length. The L50 measure is the number of scaffolds/contigs that are greater than, or equal to, the N50 length. The N50 is defined as the minimum contig length needed to cover 50% of the genome. L50 is the number of contigs whose summed length is N50. % GC-content (guanine-cytosine). All the assembly metrics were calculated using QUAST (Gurevich et al., 2013) and \*annotations using RAST(Overbeek et al., 2014). Base pairs (bp).

<b>Statistics</b>	<b>12943wD7_Mur (ref.ID)</b>
<b># contigs (&gt;= 0 bp)</b>	37
<b># contigs (&gt;= 1000 bp)</b>	9
<b>Largest contig</b>	1,477,579
<b>Total length</b>	3,264,533
<b>Total length (&gt;= 0 bp)</b>	3,272,505
<b>Total length (&gt;= 1000 bp)</b>	3,264,533
<b>N50</b>	1,424,516
<b>N75</b>	1,424,516
<b>L50</b>	2
<b>L75</b>	2
<b>GC (%)</b>	44.85
<b># coding sequences*</b>	3121
<b># RNAs (including rRNA &amp; tRNA)*</b>	42
<b># Mismatches</b>	0

**Table 3-21 Taxonomic labelling using top families and genera.** Sequence-based classification system calculated using the software Kraken (Wood and Salzberg, 2014).

<b>Unclassified (%)</b>		<b>94.60</b>
Most frequent Family (%)	<i>Flavobacteriaceae</i>	3.88
2nd most frequent Family (%)	<i>Cyclobacteriaceae</i>	0.13
Most frequent genus (%)	<i>Muricauda</i>	1.26
2nd Most frequent genus (%)	<i>Zobellia</i>	0.62
<b>Most frequent species (%)</b>	<i>Muricauda ruestringensis</i>	<b>1.26</b>
Escherichia coli (%)	<i>Escherichia coli</i>	0.00

Distribution of the protein families and associated functional roles in *M. ruestringensis* strain SBT531 genome (Figure 3.54). Three major groups including Amino Acids and Derivatives (224 feature counts), Cofactors, Vitamins, Prosthetic Groups, Pigments (144 feature counts, which covers biotin biosynthesis) and Carbohydrates (136 feature counts).

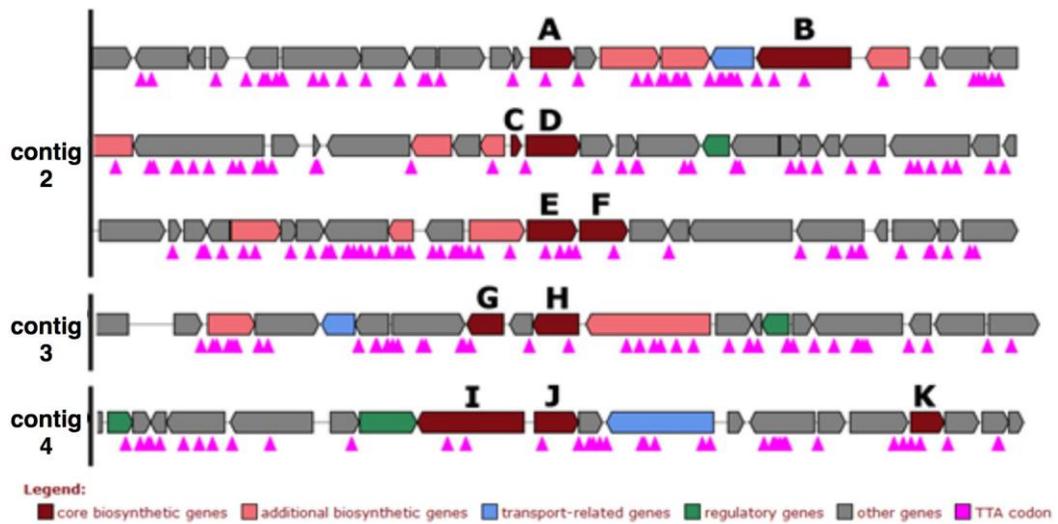


**Figure 3-54** *M. ruestringensis* functional roles and protein families. Subsystems collections retrieved from SEED viewer (Overbeek et al., 2005). Major groups include Amino Acids and Derivatives (224), Cofactors, Vitamins, Prosthetic Groups, Pigments (144) and Carbohydrates (136).

The analysis of the biosynthetic gene clusters (BGCs) of *M. ruestringensis* (Figure 3.55) yielded the follow protein families as follow: 3-oxoacyl-[acp] synthase 3 (fabH, UniProtKB: G2PNX4/G2PMD7) (Figure 3.55A, J), prolyl endopeptidase/oligopeptidase (G2PNW2) (Figure 3.55B), acyl carrier protein AcpP (Figure 3.55C), 3-oxoacyl-[acp] synthase 2 (G2PNR5) (Figure 3.55D and E), biotin synthase (Figure 3.55F), putative glycosyltransferase EpsJ (G2PLA5) (Figure 3.55G), 3-oxoacyl-[acp] synthase 3 protein 1 (fabH, UniProtKB: G2PNX4/G2PMD7) (Figure 3.55H), dipeptidyl-peptidase 5 (Figure 3.55I) and peptidase S9 (G2PNW2) (Figure 3.55K).

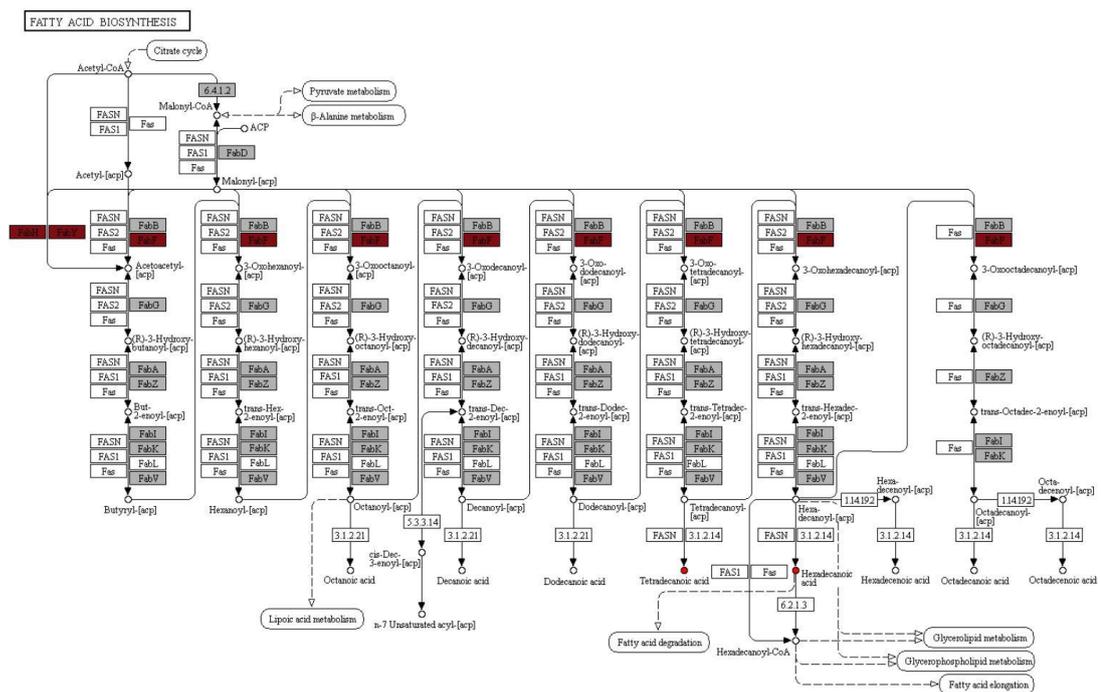
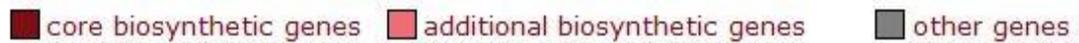
There are three known mechanisms for the initiation of fatty acid biosynthesis in bacteria. Firstly, FabH (3-ketoacyl-ACP synthase III) catalyses the condensation of acetyl-CoA with malonyl-ACP to yield acetoacetyl-ACP (Figure 3.56). In the second pathway, the acetate moiety is first transferred from acetyl-CoA to acetyl-ACP by the transacylase activity of FabH. The acetyl-ACP is then condensed with malonyl-ACP by FabB (synthase I) or alternatively by FabF (synthase II). The third pathway involves the decarboxylation of malonyl-ACP by FabH, FabB or FabF to form acetyl-ACP followed by subsequent condensation with malonyl-ACP. The evidence for the existence of these pathways and their relative contributions to the initiation of fatty acid biosynthesis is a current area of research.

Biotin (vitamin H or vitamin B7) (Figure 3.57) is the essential cofactor of biotin-dependent carboxylases, such as pyruvate carboxylase and acetyl-CoA carboxylase. Mammals cannot synthesize biotin, while in bacteria, fungi, and plants it is synthesized from pimelate thioester through different pathways. In bacteria and many organisms, pimelate thioester is derived from malonyl-ACP. The pathway starts with the methylation to malonyl-ACP methyl ester, followed by the fatty acid chain elongation cycle to form pimeloyl-ACP methyl ester, which is then demethylated to form pimeloyl-ACP. Pimeloyl-ACP is converted to biotin through the final four steps in the biotin bicyclic ring assembly, which are conserved among biotin-producing organisms. In *B. subtilis*, biotin is derived from pimeloyl-ACP formed by oxidative cleavage of long-chain acyl-ACPs. Some bacteria synthesize biotin from pimeloyl-CoA derived from pimelate. Biotin is covalently attached to biotin-dependent carboxylase by biotin protein ligase, also known as holocarboxylase synthase in mammals, to form an active holocarboxylase. After degradation of the biotinylated carboxylase into biocytin, it is further degraded by biotinidase to release free biotin, which is recycled in holocarboxylase synthesis. Biotin is catabolized by beta-oxidation of the valeric acid side chain or oxidation of sulfur in the heterocyclic ring.

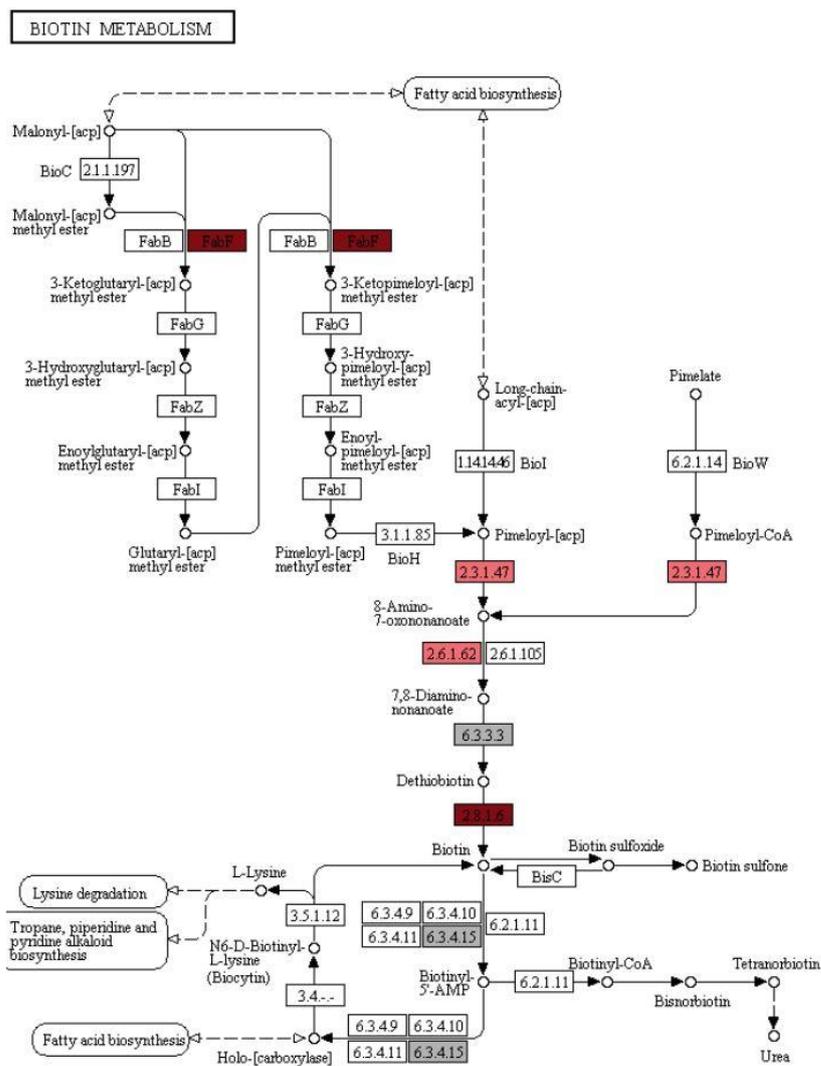


**Figure 3-55 Biosynthetic gene clusters analysis.** Analysis performed with antiSMASH. Annotated protein families as follow: 3-oxoacyl-[acp] synthase 3 (A, J), Prolyl endopeptidase (B), Acyl carrier protein AcpP (C), 3-oxoacyl-[acp] synthase 2 (D and E), Biotin synthase (F), putative glycosyltransferase EpsJ (G), 3-oxoacyl-[acp] synthase 3 protein 1 (H), Dipeptidyl-peptidase 5 (I) and Peptidase S9 (K). Acyl-carrier-protein (acp).

**Legend:**



**Figure 3-56 Biosynthesis of fatty acids pathway.** Mapping of the gene clusters found in the antiSMASH analysis using a reference pathway (EC, enzyme nomenclature) map from KEGG database (<https://www.genome.jp/kegg/pathway/map/ec00061.html>). Tetradeconoic acid and hexadecanoic acid are displayed as full colored red circles. fabH (3-oxoacyl-[acp] synthase 3), fabF (3-oxoacyl-[acp] synthase 2).



**Legend:**

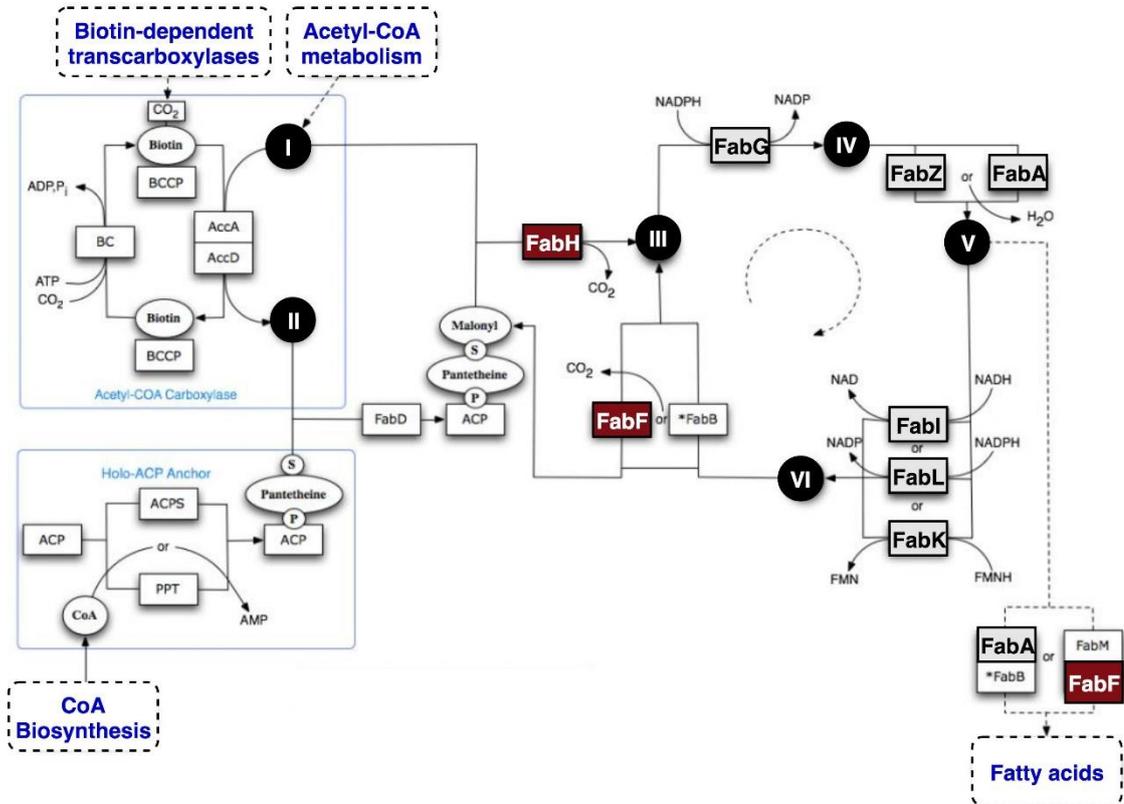
■ core biosynthetic genes    ■ additional biosynthetic genes    ■ other genes

**Figure 3-57 Metabolism of biotin (vitamin H).** Mapping of the gene clusters found in the antiSMASH analysis using a reference pathway (EC, enzyme nomenclature) map from KEGG database (<https://www.genome.jp/kegg/pathway/map/ec00780.html>). fabH (3-oxoacyl-[acp] synthase 3), fabF (3-oxoacyl-[acp] synthase 2).

**Legend:**

■ core biosynthetic genes

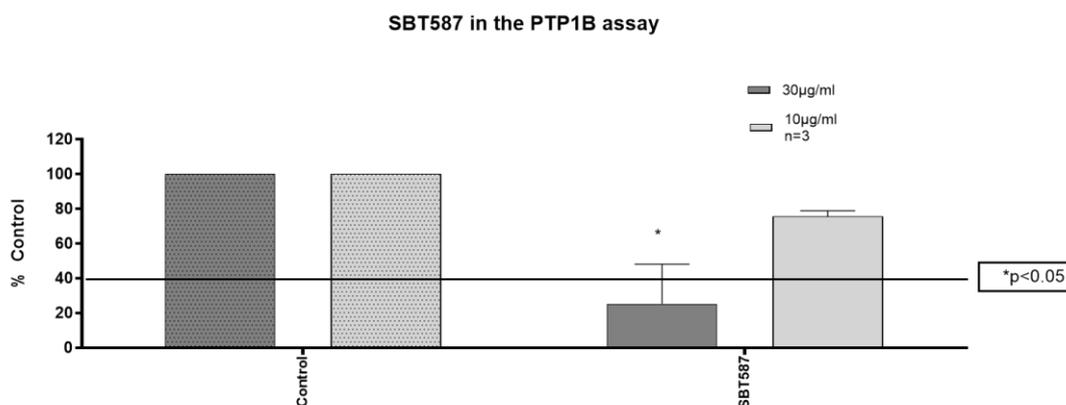
■ other genes



**Figure 3-58 Summary diagram of metabolic pathways of biotin and fatty acids biosynthesis.** The following intermediates are represented: acetyl-CoA (I), malonyl-CoA (II), beta-ketoacyl-ACP (III), beta-hydroxyacyl-ACP (IV), trans-2-enoyl-ACP (V) and acyl-ACP (VI). Acyl carrier protein (ACP), biotin carboxyl (BC) carrier protein (BCCP). FabH (3-oxoacyl-[ACP] synthase 3), FabF (3-oxoacyl-[ACP] synthase 2).

### 3.9 Small scale-up fermentation and extraction of SBT587

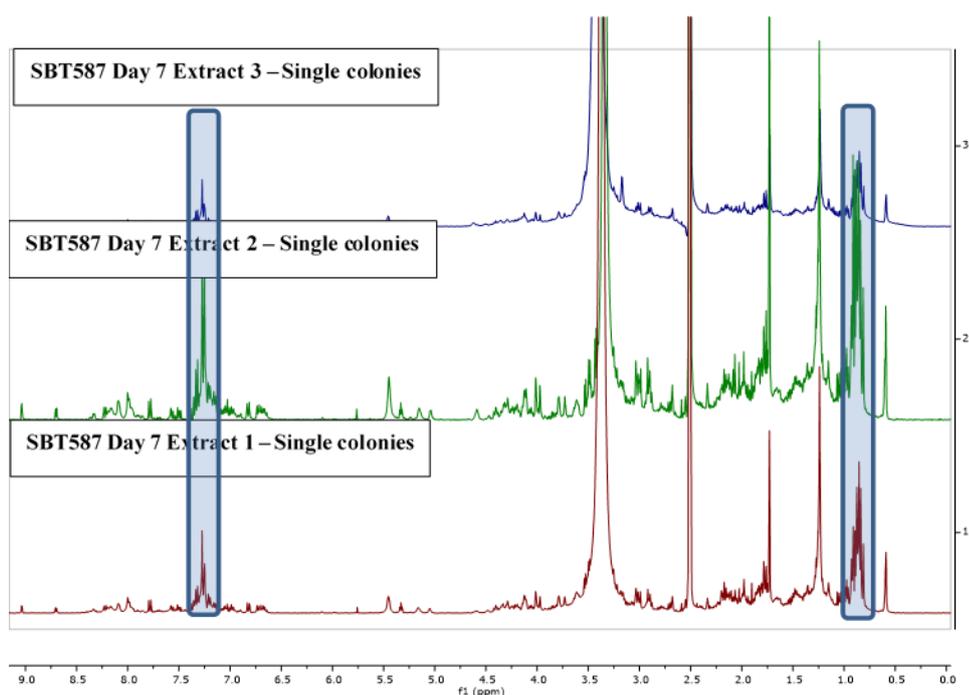
A second bacterium *M. ruestringensis* isolate (SBT587) was obtained from the intertidal pools in Iceland. Bacterial cultures were optimized, incubated and harvested at different growth phases for a controlled number of days 1, 3, 5, or 7 in Difco Marine Broth in shake flasks (12 x 200 ml) inoculated with 10% inoculum and single colonies at 45 °C, 120 rpm. Comparisons were made in terms of metabolite production by varying culture period, comparing metabolite profiles at 1, 3, 5 or 7 days. SBT 587 was found to produce the same type of compounds as SBT 531, thus there were no chemical novelty. In this case, no further scale-up fermentation and fractionation was done on SBT 587. The initial bioactivity screening for alpha-glucosidase and PTP1B inhibition was also performed on the microbial isolate SBT587. SBT587 did not inhibit an alpha glucosidase activity but exhibited activity against PTP1B at 98% inhibition at a concentration of 30 µg/mL (Figure 3.59).



**Figure 3-59 Inhibition effect of SBT587 crude extract against PTP1B at 30 and 10 µg/mL (n=3).** Error bars represent the standard deviations (SD) of three replicates (n=3), and values were significant at \* $p < 0.05$  by Two-way ANOVA multiple comparisons with the control.

### 3.9.1 NMR and HR-LC/MS comparisons with SBT531

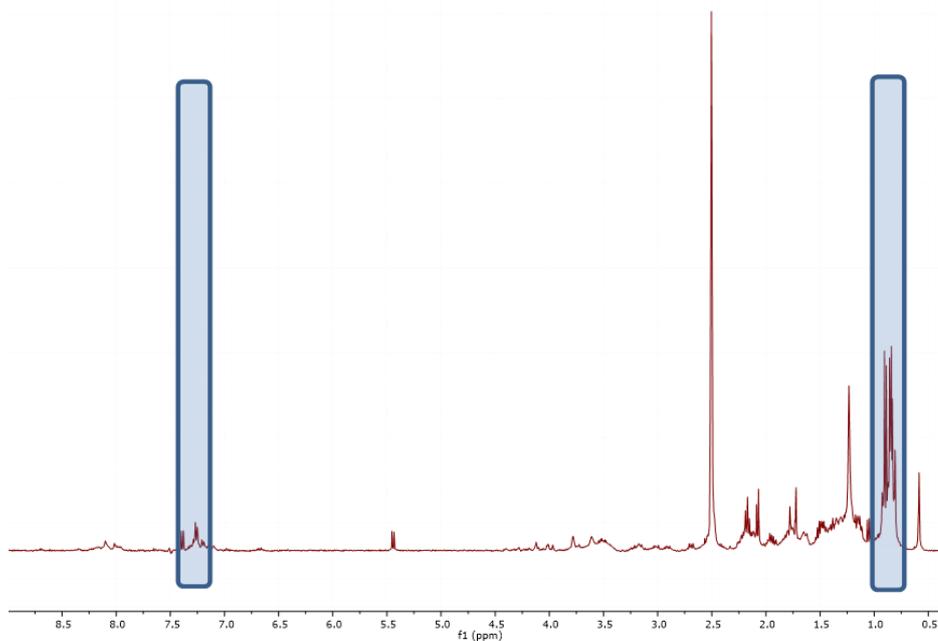
Different extracts of SBT587 were prepared for NMR and HR-LC/MS analysis to determine the metabolite production and if it was comparable to SBT531. Comparing the proton spectral data of SBT587 single colonies at day 7 extracts 1, 2, and 3 (Figure 3.60) with SBT531 small scale at day 7 (Figure 3.61), it is perceptible that the spectra of both strains shared identical peaks, on the aliphatic region between 0.5-1 ppm the characteristic doublet of the fatty acid derivatives of aseanostatin P6 and on the aromatic between 7-7.5 ppm the aromatic rings of the cholic acid derivatives. Furthermore, when analysing the proton NMR spectral data of SBT587, 10% inoculum at day 7 extracts 1, 2 and 3 (Figure 3.62) with SBT531 Fraction2 (Figure 3.63) it is possible to detect the doublet between 0.5-1 ppm and the triplet between 2-2.5 ppm, which are typical or characteristic resonances for aseanostatin P6.



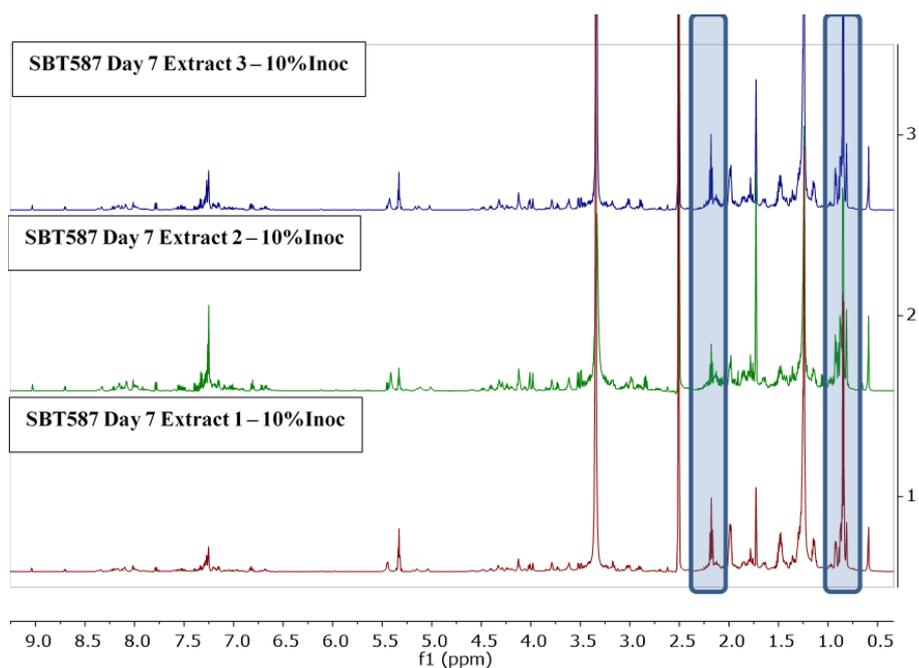
**Figure 3-60**  $^1\text{H-NMR}$  spectral data (400 MHz) of SBT587 Day 7 Extract 1, 2 and 3 Single colonies. The blues boxes showed the fatty acid derivatives signals and cholic acid

derivatives

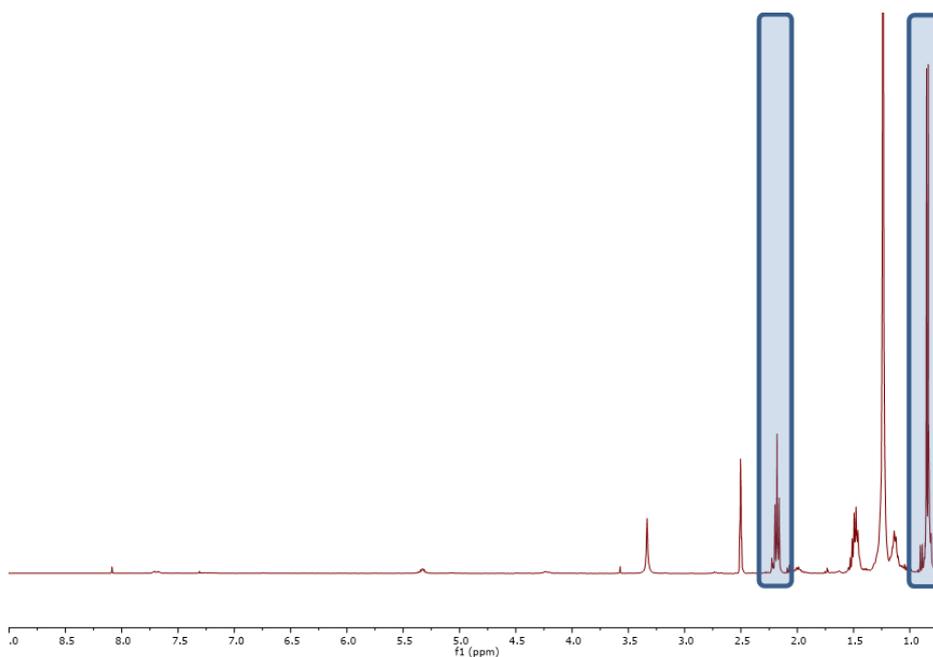
signals.



**Figure 3-61**  $^1\text{H-NMR}$  spectral data (400 MHz) of SBT531 small scale up. The blues boxes showed the fatty acid derivatives signals and cholic acid derivatives signals.

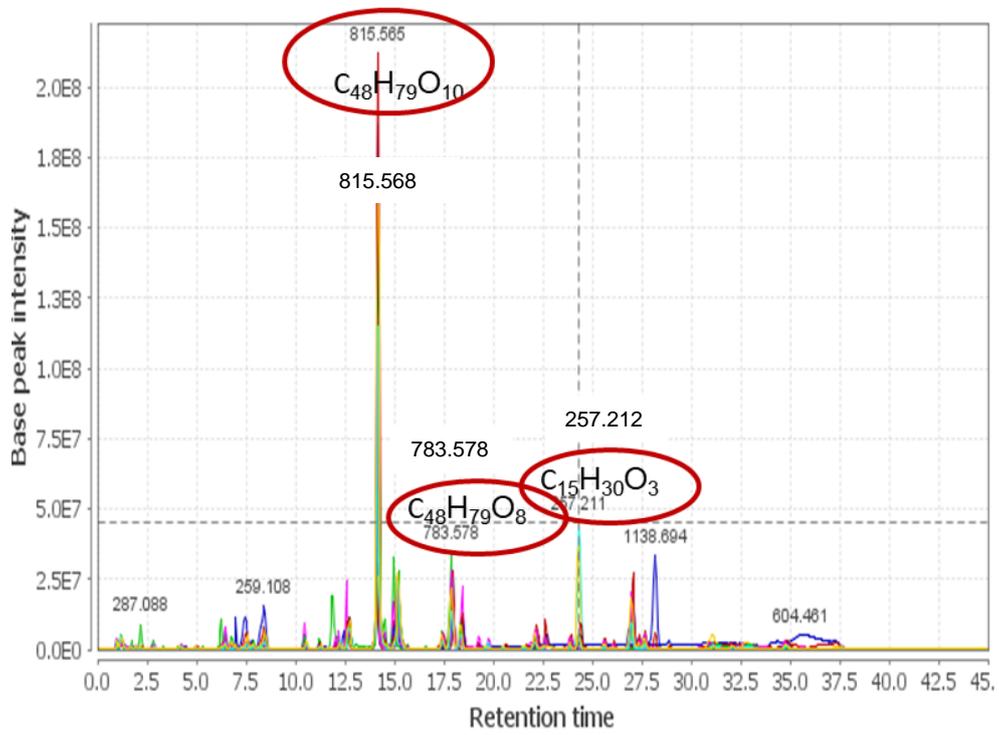


**Figure 3-62**  $^1\text{H-NMR}$  spectral data (400 MHz) of SBT587 Day 7 Extract 1, 2 and 3 10% Inoculum. The blues boxes showed the fatty acid derivatives signals, characteristic of Aseanostatin P6.



**Figure 3-63**  $^1\text{H-NMR}$  spectral data (400 MHz) of SBT531 Fraction 2. The blues boxes showed the fatty acid derivatives signals characteristic of Aseanostatin P6.

Analysing the MS data results regarding the production of secondary metabolite, the base peak plot of the chromatograms from the negative mode of ionization for SBT587 inoculated with 10% inoculum (Figure 3.64) and single colonies (Figure 3.65) were found to be comparable to those of SBT531 Fractions 4, 6, 7 and 9 (Figure 3.66). It can be confirmed that both strains are producing similar set of compounds, although the retention time is slightly different.



**Figure 3-64 Base peak plot chromatogram (HR-LC/MS data) of SBT587 10% Inoculum metabolites (Negative ionization) created on MzMine 2.10 modified version.**

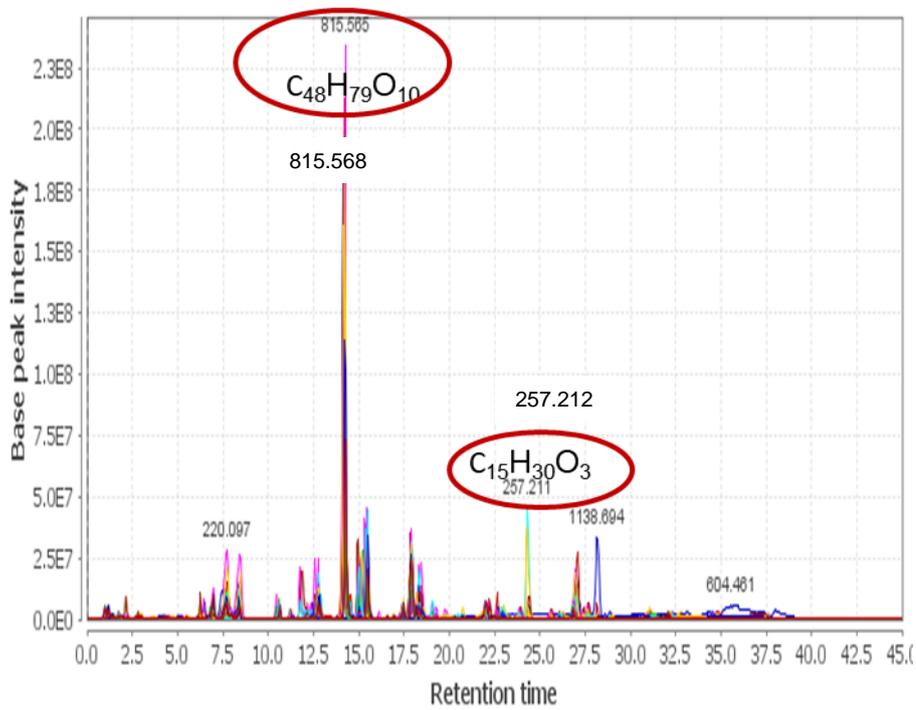


Figure 3-65 Base peak plot chromatogram (HR-LC/MS data) of SBT587 Single Colonies metabolites (Negative ionization) created on MzMine 2.10 modified version.

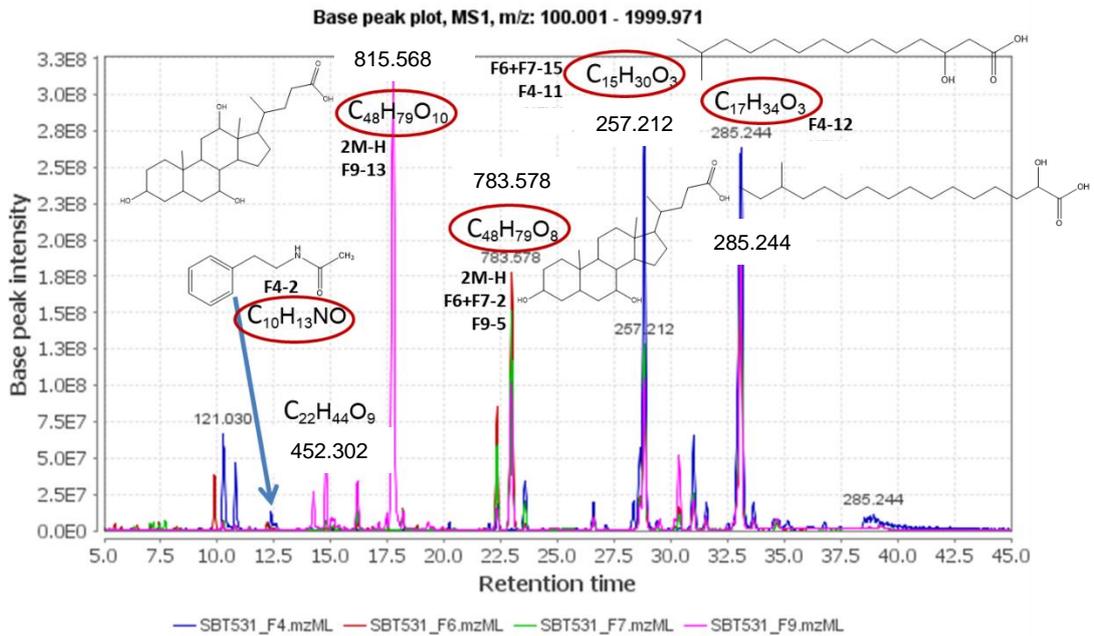


Figure 3-66 Base peak plot chromatogram (HR-LC/MS data) of SBT531 Fractions 4, 6, 7 and 9 metabolites (Negative ionization) created on MzMine 2.10 modified version.

## Discussion

In the present work, different scale-up fermentations were performed on *M. ruestringensis* strains (SBT531 and SBT587). A metabolomics workflow for drug discovery was conducted prior to extraction and fractionation. NMR and HR-LC/MS were used as analytical platforms to achieve the metabolomics analysis. The extracts and fractions were also subjected to a battery of bioassays, from which bioactive fractions were selected for further compound purification for structure elucidation. Potential biosynthetic gene clusters (BGCs) based on previous assessment of *M. ruestringensis* genomic profile was also studied. Similar secondary metabolites were produced during the scale-up steps. Moreover, the elucidated and isolated compounds were mainly branched hydroxylated fatty acids, *N*-phenethylacetamide and a cholic acid derivative as shown in Figure 3.67.

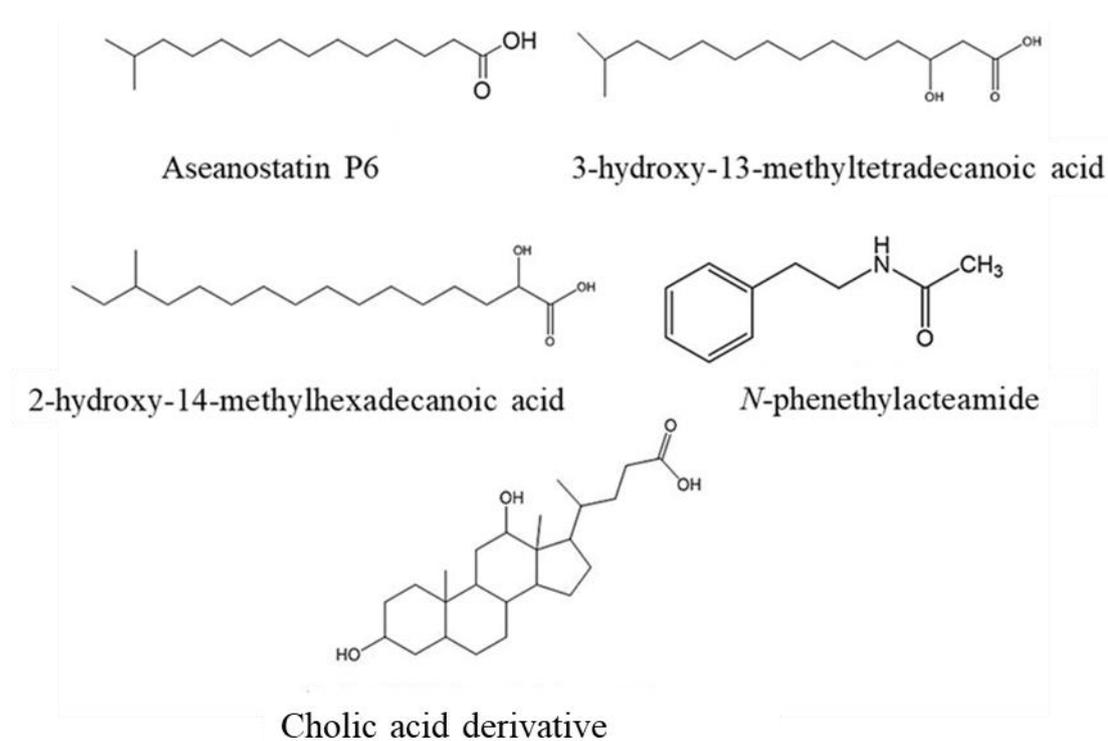


Figure 3-67 Metabolites identified in *M. ruestringensis* strains SBT531 and SBT587.

Both *M. ruestringensis* strains, SBT531 and SBT587, were isolated from intertidal pools in Iceland, therefore, was expected that both strains secondary metabolites were analogous and synonymous, which was confirmed by their similar NMR and HR-LC/MS data. SBT531 and SBT587 both produced the hydroxylated fatty acid derivative, 3-hydroxy-13-methyltetradecanoic acid, with an  $m/z$  of 257.212 [M-H]<sup>-</sup> and aseanostatin P6. Aseanostatin P6 was the major metabolite as indicated by its NMR spectrum, but it ionised poorly and was not visible on the total ion chromatograms of SBT587. Cholic acid derivatives with an  $m/z$  of 815.568 [2M-H]<sup>-</sup> and  $m/z$  of 783.578 [2M-H]<sup>-</sup> were also being produced but with a reduced intensity when compared to the fatty acids. The profile of hydroxy fatty acids is useful to differentiate and grouping certain gram-negative bacteria like *M. ruestringensis*. Both 2-hydroxy and 3-hydroxy fatty acids are found as constituents of cells and their distribution differs depending on the types of microorganisms (Yamanaka et al., 1988). The occurrence of branched-chain hydroxy fatty acids in gram-negative bacteria is rather specific and useful in their comparative biology and phenotypic systematics.

Regarding the Bioreactor fermentation, although the scaling up of SBT531 from 1-L to a 10-L bioreactor resulted in identical growth and similar production of the aseanostatin derivatives, it also resulted in the production of aseanostatin P6 isomers. It is currently unknown why aseanostatin P6 and its derivatives are produced mainly in the exponential phase. aseanostatin P6 is a cellular fatty acid that is known to be a component of the cell wall of *M. ruestringensis*. It comprises 15.52% of the fatty acid profile of *M. ruestringensis* strain BIT (Bruns et al., 2001) and is responsible for the membrane fluidity (Kaneda, 1991). In fact, being an extremophile requires ability of protective measures to extreme environments in which the cell outer-membrane plays an important role as it is the main target of adverse conditions (Kaneda, 1991). Further studies on the optimisation of carbon, nitrogen, and salt concentration of the media and as well gene expression analysis from metabolic approach might yield a better understanding of the role of aseanostatin P6 and its derivatives in SBT531.

Detailed chemical profiles about certain bacteria are well described on the literature, nevertheless current knowledge regarding *M. ruestringensis* is limited. Literature

search concerning production of secondary metabolites in the genus *Muricauda* is quite limited. Nevertheless, carotenoid pigments such as zeaxanthin, a xanthophyll (also known as phyloxanthins) are produced (Prabhu et al., 2014) in many member of the *Muricauda* genera owning antioxidant and anticancer properties. Another example of a bioactive isolated compound from the Genus *Muricauda* was MomL, an *N*-acyl homoserine lactone (Tang et al., 2015) that could be used as a therapeutic agent against *Pseudomonas aeruginosa* since it seems to decrease the virulence in a *Caenorhabditis* infection model.

Aseanostatin P6 or 13-methyltetradecanoic acid, C<sub>15</sub>H<sub>30</sub>O<sub>2</sub>, was isolated from SBT531 fraction 2. Fraction 2 exhibited inhibition activity against alpha-glucosidase (30 µg/mL) and PTP1B (30 and 10 µg/mL, Ki 3 µg/mL) with less than 40% inhibition (≤40% threshold) which can be related with the mentioned isolated compound. It was firstly isolated from the external fatty tissue of sheep in 1952 (Hansen et al., 1953). Since then, the production and isolation of this fatty acid was reported in many bacteria, for instance *Propionibacterium propionicum* (previously known as *Arachnia propionica*) (Cummins and Moss, 1990, O'Donnell et al., 1985), *Porphyromonas circumdentaria* sp. nov. (Love et al., 1992), *Porphyromonas canoris* sp. nov. (Love et al., 1994) *Friedmanniella luteola* sp. nov., *Friedmanniella lucida* sp. nov., *Friedmanniella okinawensis* sp. nov., *Friedmanniella sagamiharensis* sp. nov. (Iwai et al., 2010), *Granulicoccus phenolivorans* gen. nov., sp. nov. (Maszenan et al., 2007) *Chryseobacterium soldanellicola* sp. nov., *Chryseobacterium taeanense* sp. nov. (Park et al., 2006), *Sphingobacterium* sp. (Naka et al., 2003), *Bacillus sphaericus* (Siegel et al., 1997), *Flavobacterium ranacida* (Faung et al., 1996), *Stigmatella aurantiaca* and *Myxococcus xanthus* (Dickschat et al., 2005). Furthermore, this compound was investigated as an important anticancer agent known as inducing apoptosis in several types of human cancer cell lines. Reported literature showed inhibition of tumour cell growth by inducing apoptosis in T-cell lymphomas *in vitro* and *in vivo* by down-regulating phospho-AKT and activating caspase-3 (Cai et al., 2013). Moreover, 13-methyltetradecanoic acid induces mitochondrial dysfunction in bladder cancer cell lines through modulation of the AKT/MAPK signaling pathways (Lin et al., 2012). This compound is also involved in the disruption of mitochondrial integrity and

inducing apoptosis via caspase-independent death pathways reported in human breast cancer cell lines (SKBR-3) (Wongtangintharn et al., 2005). Other research articles report antineoplastic activity in many Human-derived cellular lines: DU-145 (prostate carcinoma), K-562 (leukemia), HCT-116 (colon carcinoma), NCI-H1688 (small lung cell carcinoma), SNU-423 (liver carcinoma), MCF7 (mammary adenocarcinoma), BxPC-3 (pancreatic adenocarcinoma), and NCI-SNU-1 (gastric carcinoma) (Yang et al., 2000). Additionally, reports shown that aseanostatin P6 have anti-inflammatory, antioxidant, antithrombotic effects and can stabilize membranes with a neuroprotective effect on focal cerebral ischemia / reperfusion injury models (Yu et al., 2016).

3-Hydroxy-13-methyltetradecanoic acid,  $C_{15}H_{30}O_3$ , was first isolated from *Flavobacterium meningosepticum*, a Gram-negative bacterium from the *Flavobacteriaceae* family, in 1976 described by Yano *et al* (Yano et al., 1976), interestingly, the same family as *M. ruestringensis*. This compound was isolated from several bacteria within different genus, for instance, *Echinicola* (Zhang et al., 2016a), *Flavobacterium* (Sheu et al., 2016), *Sphingobacterium* (Li et al., 2016), *Aquimarina* (Wang et al., 2016b), *Algoriphagus* (Kohli et al., 2016), *Maribacter* (Park et al., 2016), *Lysobacter* (Miess et al., 2016) and *Hymenobacter* (Liu et al., 2016). The associated bioactivity was concerning insecticidal and antimicrobial activity against Gram-positive bacteria (Morgan et al., 2009, Hashizume et al., 2001, Hashizume et al., 2004), stimulator of the apolipoprotein E (ApoE) secretion from human hepatoma Hep G2 cells (Yanai et al., 2000) importance demonstrated with studies in hyperlipidaemic rabbits, in which ApoE was responsible for the prevention of progression of the atherosclerosis and also decreasing cholesterol levels (Yanai and Hiramoto, 1999).

2-Hydroxy-14-methylhexadecanoic acid,  $C_{17}H_{34}O_3$ , is the less studied of the three isolated fatty acids. It was principally reported being produced by a bacteria belonging to the order Actinomycetales (Mingma et al., 2016, Yano et al., 1970, Yano et al., 1969). The two hydroxylated fatty acid derivatives from aseanostatin P6 were isolated, 3-hydroxy-13-methyltetradecanoic acid,  $C_{15}H_{30}O_3$  from SBT531sub-fraction 11, and

2-hydroxy-14-methylhexadecanoic acid,  $C_{17}H_{34}O_3$  from SBT531 sub-fraction 12. Fraction 4 mL and F4 subFraction 11 both showed inhibition activity against PTP1B at a concentration of 30  $\mu\text{g/mL}$  (123.8  $\mu\text{M}$ ) but not at 10  $\mu\text{g/mL}$  (41.2  $\mu\text{M}$ ), which is not such a potent bioactivity as it should be as lowest as possible.

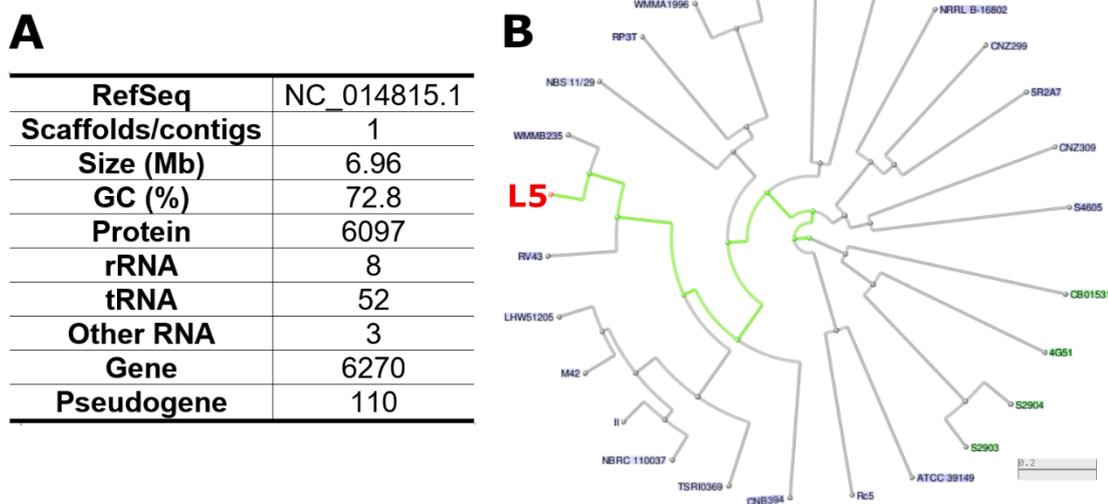
Comparing the three isolated fatty acids in terms of their inhibition ability against PTP1B, the inhibition activity putatively decreased or disappeared with the elongation and/or hydroxylation and methylation of the chain. Further research about structure activity relationship on other derivatives is required to confirm this hypothetical evidence.

## Chapter 4

*Micromonospora* sp. N17 (SBT687): SEARCH FOR  
NEW POTENTIAL AGROCHEMICAL LEADS

## 4.1 Introduction

The genus *Micromonospora* is a Gram-positive, filamentous and sporulating actinobacteria, which was first described by Ørksø in 1923 (Carro et al., 2018, Riesco et al., 2018). Currently the genus has 81 species with valid published names (<http://www.bacterio.net/micromonospora.html>), most of them only discovered during the last decade. This genus holds many uncharacterised isolates (Figure 1) available at NCBI ([www.ncbi.nlm.nih.gov/genome/13707](http://www.ncbi.nlm.nih.gov/genome/13707)). Here, in this chapter *Micromonospora* sp. L5 (Figure 4.1A, B) will be used as a template for further study of its biosynthetic gene clusters (BGCs) since the target bacteria isolated in this project, *Micromonospora* sp. N17 (SBT 687) was not sequenced. Strains from the genus *Micromonospora* have been widely isolated from soil, freshwater, and marine habitats that included plant and animal tissues as well as other sources (Hirsch and Valdés, 2010, Riesco et al., 2018, Carro et al., 2018). *Micromonospora* species have been described for being a potential source of new antibacterials (Charan et al., 2004, Berdy, 2005) but the genus has also been important for the discovery of other bioactive molecules with antitumoral activity (Igarashi et al., 2007), antifungal (Ismet et al., 2004), antiviral, insecticidal (Hirsch and Valdés, 2010) and also for producing vitamin B12 (Leviton and Hargrove, 1952), and potentially can be used for probiotic in aquaculture (Das et al., 2008).



**Figure 4-1 *Micromonospora* sp. L5 overall genomic description (A) and circular tree dendrogram (B) based on genomic BLAST of 26 isolates from an uncharacterised population of *Micromonospora*.**

The bacterium *Micromonospora* sp. N17 (SBT 687) was isolated from the Mediterranean sponge *Phorbas tenacior* from the Santorini volcanic complex of Crete. The high diversity of microorganisms isolated from the marine sponge *Phorbas tenacior* frequently found in the Mediterranean Sea are responsible for a great variety of biologically active secondary metabolites (Dupont et al., 2014). Reported natural products from both the sponge and its associated microorganism included alkaloids (Rudi et al., 1994), macrolides (Searle and Molinski, 1995), terpenoids (McNally and Capon, 2001, Jang et al., 2008, Zhang et al., 2008a), steroids (Morinaka et al., 2007, Morinaka et al., 2009), peptides or modified peptides with potential antimicrobial, antioxidant and antiplasmodial activities (Dupont et al., 2014, Fehmida et al., 2017). Preliminary bioassay screening of SBT687 demonstrated 100% growth inhibition against the parasite *Lepeophtheirus* sp. (sea lice) at 75 µg/mL. The bioscreening result of SBT687 against sea lice exhibited the insecticidal or pesticidal potential of its secondary metabolites. The characterization of new lead compounds against the mentioned infesting parasite holds great interest for the salmon farming industry. Sea lice infestation is becoming a major concern for the salmon industry due to the harmful effect in health, production and market value of the fish (Jimenez et al., 2013). Sea lice is a common name for many species of marine ectoparasitic copepods that exists naturally in the ocean. However, the copepod population can exponentially grow and

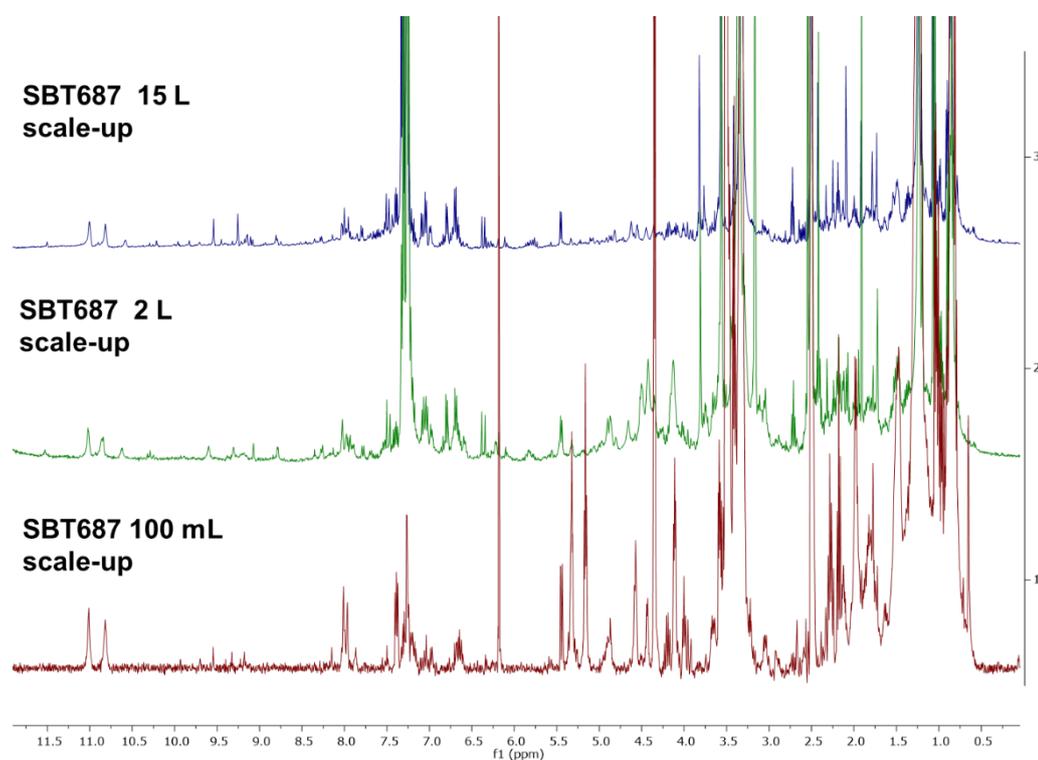
spread in fish farming sites (Wright et al., 2016). The infesting parasite entails the greatest challenge for fish farming. Chemical therapy, biological methods, and operational measures are amongst the approaches that have been used to control sea lice population and infestation (Dumas, 2016). Short-duration immersion treatments used chemicals like pyrethroid-cypermethrin or a combination of pyrethroid-deltamethrin and organophosphates azamethiphos in high-concentrations (Olsvik et al., 2014, Jimenez et al., 2013). On the other hand, macrocyclic lactone emamectin benzoate and the benzoyl urea diflubenzuron and teflubenzuron were used for in-feed treatments (Murray, 2016). Frequently, fishes were also exposed to hydrogen peroxide, a non-specific oxidising agent (Vera and Migaud, 2016, Helgesen and Horsberg, 2013). Chemical treatments can be stressful for the fish and/or even consist a potential threat to the surrounding wild life while resistant parasite strains can still be passed further to stronger generations (Grant, 2002, Helgesen and Horsberg, 2013). Recently, increasing resistance to chemical therapy has been reported (Helgesen and Horsberg, 2013, Carmona-Antoñanzas et al., 2016). Therefore, there is a current need for investigation of novel intervention strategies and novel chemicals to control of sea lice infestations, which is heavily dependent on effective insecticide chemotherapeutants (Wagner et al., 2008, Helgesen and Horsberg, 2013, Jimenez et al., 2013).

## **4.2 Small and Medium scale up fermentation, extraction and fractionation of SBT687**

### **4.2.1 Scaled-up fermentation of SBT687**

Methods relating to fermentation and extraction were described in detail under Chapter 2. The initial small-scale fermentation of the bacterium *Micromonospora sp.* N17 (SBT 687) in ISP2 agar plates and 100 mL liquid broth was done in Julius-von-Sachs-Institut für Biowissenschaften, University of Wuerzburg, Germany. Fermentation conditions for SBT687 were optimised using 10% (v/v) inoculum on ISP2 liquid broth then incubated for 7 days at 30° C with shaking at 120 rpm. Using the optimised culture conditions, SBT687 was then scaled-up to 2L (5 x 400 ml). Fractions from the 2L

culture were sent to PHARMAQ for bioactivity screening for anti-sea lice activity (Lesa assay). PHARMAQ AS (part of Zoetis) that was founded in 2014 and based in Overhalla, Norway, is a veterinary company that manufactures health products for the aquaculture industry. A second medium scale-up to 15L (30 x 500 mL) was also accomplished. Fractions from the culture extract were further subjected to NMR, LC-HRMS, bioassay screening and metabolomics profiling analysis. The  $^1\text{H}$ -NMR stacked spectra of the different SBT687 fermentation scale-up are presented in Figure 4.2, illustrating the stability of the bacterium to be consistent in its metabolite production.



**Figure 4-2 Proton  $^1\text{H}$  NMR (400 MHz) spectra obtained for SBT687 at different scale-up from 100 mL to 2L then 15L. No noticeable changes were observed during the scale-up. Solvent used was  $\text{DMSO-}d_6$  (peak at 2.5 ppm).**

#### 4.2.2 Fractionation of SBT687

Bacterial cultures (liquid cultures) of SBT687 were scaled-up to 2 L (5 x 400 ml) using optimised conditions. After seven days of incubation, bacterial cultures were extracted using ethyl acetate. Total organic extract from the obtained yield was 131.9 mg.

Further fractionation of the ethyl acetate extract was performed using the Reveleris™ Flash Chromatography System enduring 100-min isocratic gradient elution with water (A) and ACN (B) on a reverse-phase 40g C18 column cartridge. The gradient elution HPLC is shown in Table 4.1 and detailed chromatographic conditions in Table 4.2. All yield of the fractions were sent to PharmaQ (Norway) for bioactivity screening. Full description of this setup can be seen in Chapter 2 (2.3.2 to 2.6.2.2).

**Table 4-1 Gradient used for first fractionation of SBT687 on the Reveleris™ Flash Chromatography System.**

Time (min)	% Solvent B
0	10
70	100
30	100
<b>Total run time=100</b>	

**Table 4-2 Chromatographic conditions applied to the fractionation of SBT687 using the Reveleris™ Flash Chromatography System.**

<b>Column</b>	Reveleris C18 40g
<b>Sample</b>	3 spoons of Celite® S with sample dissolved in MeOH
<b>Solvents</b>	Water (A) and ACN (B)
<b>Flow rate</b>	20 mL/min
<b>Detection mode</b>	UV1+UV2
<b>ELSD Threshold</b>	5 mV
<b>UV Threshold</b>	0.05 AU
<b>UV1 Wavelength</b>	290 nm
<b>UV2 Wavelength</b>	320 nm

The fractionation of SBT687 resulted in 72 fractions, which were pooled according to similarities of their thin layer chromatographic (TLC) profile to yield a total of eight fractions (F1-F8) as displayed in Table 4.3.

**Table 4-3 First SBT687 fractionation afforded 8 fractions.**

<b>Fraction/</b>	<b>Test Tube/Fraction</b>	<b>Yield (mg)</b>
<b>numbers</b>		
F1	1-10	7.9
F2	11-20	12.3
F3	21-30	1.9
F4	31-40	3.1
F5	41-50	1.3
F6	51-60	1.9
F7	61-72	7.5
F8	Wash run	19.5

A second fractionation was performed in the crude extract (396.1 mg) from the 15L culture. Similar chromatographic conditions as the 2L culture extract were applied. A 70-min gradient elution was applied with equilibration at 0 minutes with 2% B, increasing to 5%, then followed by a 70-min gradient to 100% B as presented in Table 4.4. A modified gradient was implemented since during the first chromatographic fractionation, 80% of the peaks were eluting in the first 10 minutes of the run. The initial percentage of B was decreased from 10 to 2% to decelerate the elution of the highly polar compounds in order to increase the retention and enhance the separation of analytes. Aliquots of each fractions were prepared for MS, NMR (when possible due to limited extract weight) and bioactivity assays. The second fractionation round of SBT687 resulted in 112 fractions, which were pooled according to their TLC profile to afford a total of 22 fractions (F1-F22) as shown in Table 4.5. The TLC plates from both fractionations were compared based with their similarities and retention factors (R<sub>f</sub>) values. Detailed description of this setup can be seen in Chapter 2 (2.11 to 2.6.2.2).

**Table 4-4 Gradient used for the second fractionation of SBT687 on the Reveleris™ Flash Chromatography System.**

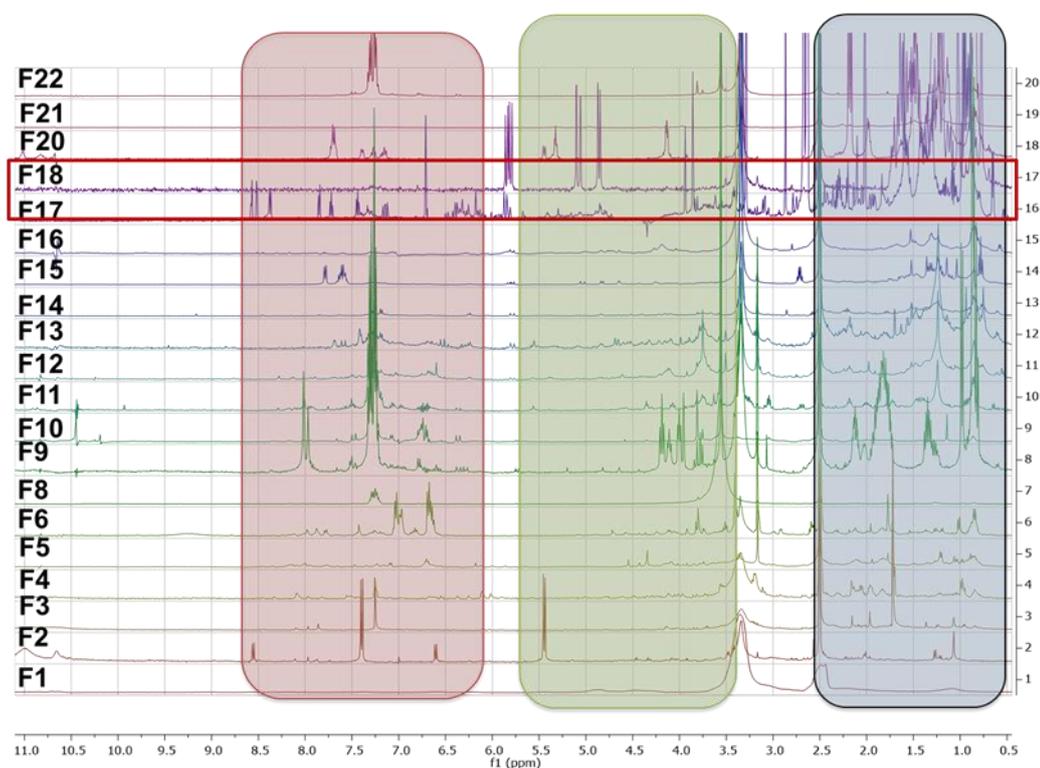
Time (min)	% Solvent B
0	2
5	5
30	5
40	50
50	50
60	100
70	100
<b>Total run time=70 min</b>	

**Table 4-5 Second SBT687 fractionation afforded 22 fractions.**

Fraction	Test Tube/ Fraction numbers	Yield (mg)
F1	1-3	44.3
F2	4	10.8
F3	5-6	6.9
F4	7	2.1
F5	8-25	14.3
F6	26-41	27.8
F7	42-51	5.6
F8	52-56	4
F9	57-59	20.5
F10	60-61	25.3
F11	62-67	29.5
F12	68-71	6.7
F13	72-75	10.9
F14	76-77	7
F15	78-80	6.5
F16	81-85	14.3
F17	86	1.3
F18	87-95	0.6
F19	96-99	4.7
F20	100-112	11.7
F21	Run wash	7.8
F22	Column Flushing (100% ACN)	66.4

#### 4.2.2.1 $^1\text{H}$ -NMR spectra of SBT687 fractions

The fractions were analysed by  $^1\text{H}$  NMR spectroscopy (Figure 4.3) along with HR-LC-mass spectrometry to have a general perception of the type of compounds being produced in terms of their chemistry and to prioritise which fractions, if any, would proceed for further purification work. Analysis of the  $^1\text{H}$  NMR spectral data showed distribution of the signals in the aliphatic (0 – 2.5 ppm), olefinic (3.5 – 5.5 ppm) and aromatic (6 – 8.5 ppm) regions. Moreover, it displays the presence of hydroxylated aliphatic and olefinic compounds such as those found in terpenoids, steroids and fatty acids. Additionally, the presence of aromatic compounds like phenols and alkaloids were indicated by presence of aromatic region signals.

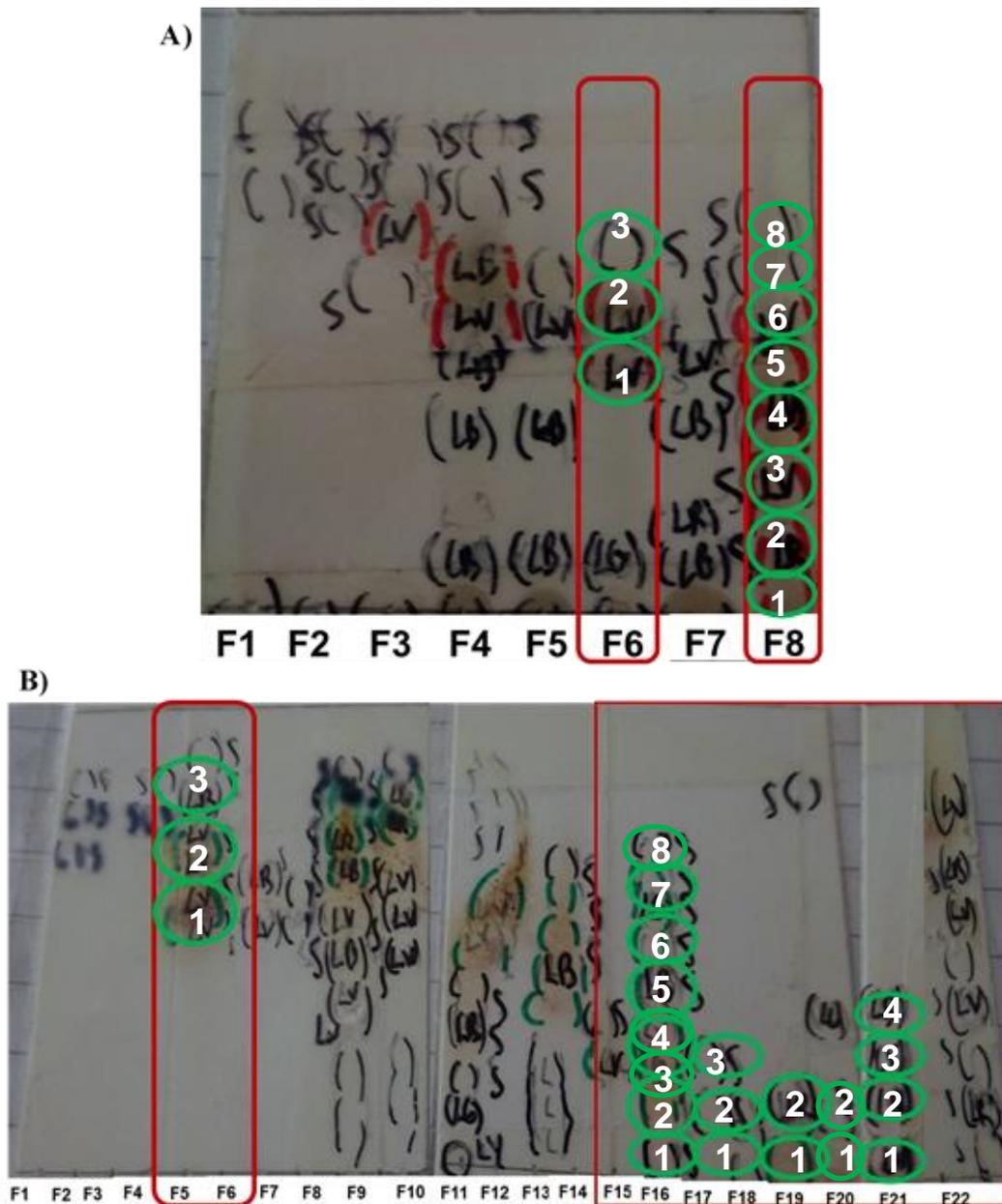


**Figure 4-3 Proton  $^1\text{H}$  NMR (400 MHz) spectra obtained for SBT687 fractions.** Numbers on Y axis indicate the fraction number. Solvent used was DMSO- $d_6$  (peak at 2.5 ppm). Highlighted in blue is the aliphatic region, green for the olefinic region, and in red is the aromatic region. The red box indicates the fractions from which compounds were isolated.

#### 4.2.2.2 Biological activity of SBT687 crude extract and fractions

The preliminary bioassay screening of SBT687 crude extract exhibited 100% inhibition against the parasite *Lepeophtheirus* sp. (sea lice) at 75 µg/mL. Fractions F6 and F8 from the first fractionation afforded 65% and 72% of inhibition against the parasite *Lepeophtheirus* sp. (sea lice), respectively, at 75 µg/mL. Many fractions from the second fractionation weren't sent to Norway (for bioassay screening) due to the end of the project and limitations concerning samples weight. Comparing the TLC profiles of the scale-up batches in terms of their retention factors (Rf) values, absorbance characteristics under the UV and colouration with the anisaldehyde spray reagent, F5/F6 and F15-F22 from the second fractionation of the 15L scale-up were found to be comparable to those of F6 and F8 of the first fractionation from the 2L scale-up (Figure 4.4). However, although F5/F6 of the second fractionation has a similar absorbance pattern under the UV as those of F6, their Rf values did not match. Hypothetically, F15-F22 from the second fractionation must be the active fractions with their comparable Rf values with those from the first fractionation. Rf values of relevant spots were calculated according to the formula defined below and values shown in table 4.6.

where: 
$$Rf \text{ value} = \frac{\text{distance of analyte migration}}{\text{distance of mobile phase migration}}$$



**Figure 4-4 Summary TLC plates** for the fractionation of **A)** the 2L culture (F1 to F8) and **B)** 15L culture (F1 to F22) extracts of SBT687 with mobile phase solvents ACN:Water 1:1 on RP18 plates. Brown spots were visualized after spraying with anisaldehyde reagent. As observed under short and long UV wavelength, the spots were marked S (Short), LB (Long Blue), LG (Long Green), LY (Long Yellow), LV (Long Violet). The red boxes highlighted the active fractions against parasite *Lepeophtheirus sp.* (sea lice) at 75 µg/mL.

**Table 4-6 R<sub>f</sub> values of the encircled analytes from the TLC plates** for the fractionation of **A)** the 2L culture (F1 to F8) and **B)** 15L culture (F1 to F22) extracts of SBT687. As observed under short and long UV wavelength, the spots were marked S (Short), LB (Long Blue), and LV (Long Violet). Brown spots were visualized after spraying with anisaldehyde reagent.

(A) Fraction 2L	R <sub>f</sub> value	(B) Fraction 15L	R <sub>f</sub> value	Spot colour (UV absorbance)	Spot colour (spray reagent)
<b>F6</b>	R <sub>f1</sub> – 0.38	<b>F5</b>	R <sub>f1</sub> – 0.55	short	
	R <sub>f2</sub> – 0.49		R <sub>f2</sub> – 0.67	LV	Brown
	R <sub>f3</sub> – 0.60		R <sub>f3</sub> – 0.82	LV	Brown
<b>F8</b>	R <sub>f1</sub> – 0.05	<b>F16</b>	R <sub>f1</sub> – 0.05		Brown
	R <sub>f2</sub> – 0.13		R <sub>f2</sub> – 0.13	short	Brown
	R <sub>f3</sub> – 0.22		R <sub>f3</sub> – 0.24	short/LB	Brown
	R <sub>f4</sub> – 0.35		R <sub>f4</sub> – 0.38	LV	Brown
	R <sub>f5</sub> – 0.42		R <sub>f5</sub> – 0.45	LB	Brown
	R <sub>f6</sub> – 0.49		R <sub>f6</sub> – 0.49	short	Brown
	R <sub>f7</sub> – 0.58		R <sub>f7</sub> – 0.58	LV	
	R <sub>f8</sub> – 0.61		R <sub>f8</sub> – 0.61	short	
<b>F8</b>	R <sub>f1</sub> – 0.05	<b>F17</b>	R <sub>f1</sub> – 0.05		Brown
	R <sub>f2</sub> – 0.13		R <sub>f2</sub> – 0.13	short	Brown
	R <sub>f3</sub> – 0.22		R <sub>f3</sub> – 0.25	short/LB	Brown
<b>F8</b>	R <sub>f1</sub> – 0.05	<b>F19</b>	R <sub>f1</sub> – 0.05		Brown
	R <sub>f2</sub> – 0.13		R <sub>f2</sub> – 0.13	short	Brown
<b>F8</b>	R <sub>f1</sub> – 0.05	<b>F20</b>	R <sub>f1</sub> – 0.05		Brown
	R <sub>f2</sub> – 0.13		R <sub>f2</sub> – 0.13	short	Brown
	R <sub>f1</sub> – 0.05	<b>F21</b>	R <sub>f1</sub> – 0.05		Brown
	R <sub>f2</sub> – 0.13		R <sub>f2</sub> – 0.14	short	Brown
	R <sub>f3</sub> – 0.22		R <sub>f3</sub> – 0.26	short/LB	Brown
	R <sub>f4</sub> – 0.35		R <sub>f4</sub> – 0.32	LV	Brown

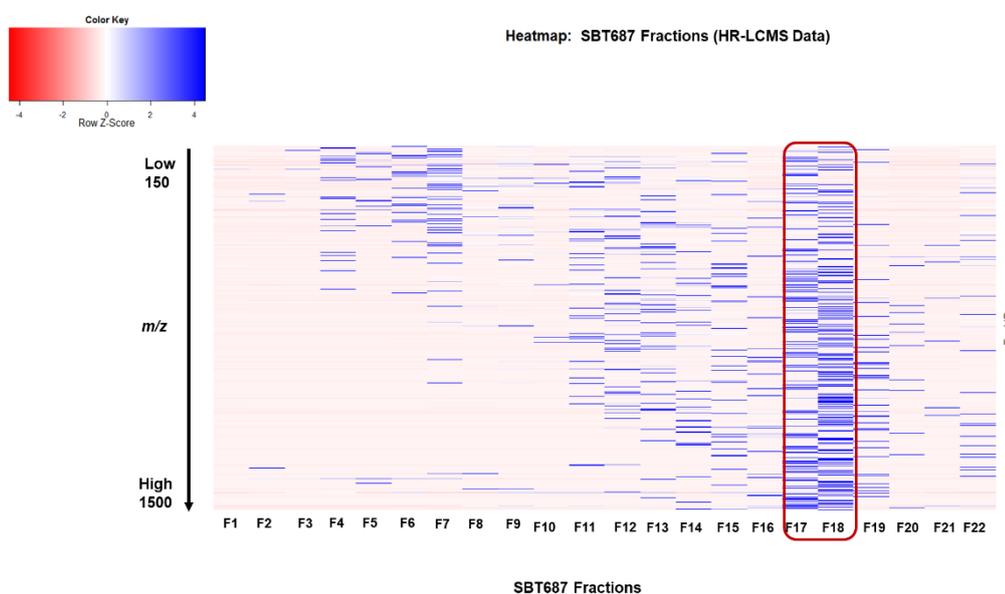
## 4.3 Metabolomic guided screening of SBT687

### 4.3.1 Metabolomic profiling of fractions of the crude extract of SBT687

#### 4.3.1.1 HR-LC/MS data

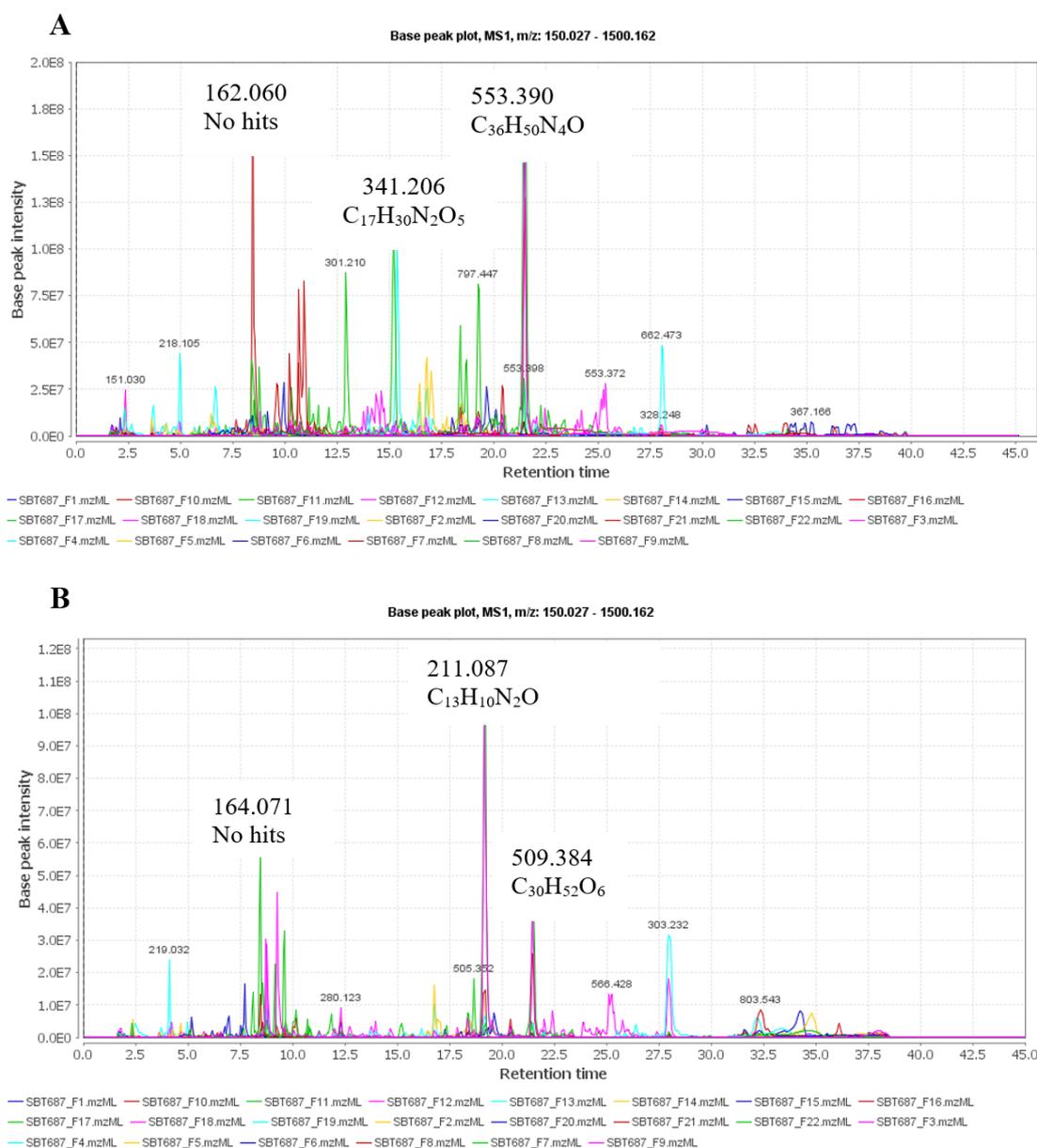
After data acquisition using HR-LC/MS of the crude extract and fractions, datasets were processed as described in Chapter 2 (section 2.9.1). Then implementation of statistical procedures, mainly dimensionality reduction approaches for feature elimination and extraction using, for instance an unsupervised modelling method—

principal component analysis (PCA) and a supervised modelling method—orthogonal projections to latent structures discriminant analysis (OPLS-DA). This yielded the so-called plot-scores (of samples) and loading-plots (of variables) than can be displayed as a PCA biplot and as well as S-plots, which integrates modelled covariance and correlation from the OPLS-DA model in a scatter-plot. Statistical validation of the OPLS-DA model was achieved by resampling the observed data 100 times (permutation test). Additionally, a heatmap (Figure 4.5) was also generated using an initial matrix of  $m/z$  and peak intensities across fractionations (F1-F22). The heatmap allowed the visualization and comparison of the chemical diversity of respective fractions, as well to establish similarities and differences between them. The colour key from red to blue indicated increasing intensity of specific chemical elements in the fractions. The heatmap on shows that F17 and F18 produced a higher diversity of metabolites as evidenced by the greater number of intense blue bands. While the initial fractions F4 to F7 gave rise to a larger quantity of low-molecular weight metabolites, and later fractions F19 to F22 had an increase in density of higher molecular weight compounds. F11 to F16 produced a variety of types of metabolites ranging from lower to higher  $m/z$ . Other fractions afforded certain specific metabolites as shown by the emergence of new blue bands were observed.

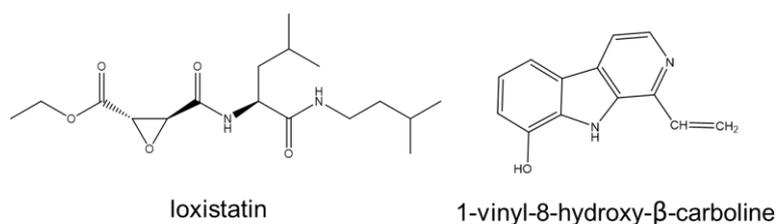


**Figure 4-5 Heatmap based on HR-LC/MS data** displaying distinct metabolic profiles of 22 fractions of SBT687. Boxed in red are the fractions with higher chemical diversity.

The HR-LC/MS data was processed using MZMine modified version 2.10 from which base peak chromatograms and scatter plots could be generated that allowed additional comprehensive profiling of the chemical composition of each fraction in comparison to the crude extract. Furthermore, this also later permitted correlating the mass spectral data with those of the NMR results. Visualization of the base peak plot chromatograms on negative and positive ionization mode as shown on Figure 4.6A and B illustrated the main compounds being produced. The base peak chromatogram on negative ionization mode illustrated high intensity peaks from F17. Apart from the unknown peaks found at  $m/z$  162.060  $[M-H]^-$ , an ion peak at  $m/z$  553.390  $[M-H]^-$  with a molecular formula of  $C_{36}H_{50}N_4O$  was putatively identified as manzamine H (section 4.4.2). Manzamines are a class of polycyclic alkaloids with a fused and bridged membered heterocyclic ring system coupled to a beta-carboline moiety (Radwan et al., 2012). Additionally, another metabolite at  $m/z$  341.207  $[M-H]^-$  that afforded a molecular formula of  $C_{17}H_{30}N_2O_5$ , detected in fractions F2, F4, F8, F9 and F12 to F17 was reputed as loxistatin (Figure 4.7). Loxistatin was earlier described as a cysteine proteinase inhibitor (Satoyoshi, 1992). On the other hand, the positive ionization base peak plot chromatogram showed high intensity peak at  $m/z$  211.087  $[M+H]^+$  with a molecular formula of  $C_{13}H_{10}N_2O$  for the cytotoxic alkaloid 1-vinyl-8-hydroxy- $\beta$ -carboline (Prinsep et al., 1991, Beutler et al., 1993), which was detected in nearly all fractions but with highest intensity in F18. Another major peak, found in F18, was detected at  $m/z$  509.384 with a molecular formula of  $C_{30}H_{52}O_6$  was dereplicated as polyhydroxylated androstane (Kobayashi and Mitsuhashi, 1982).



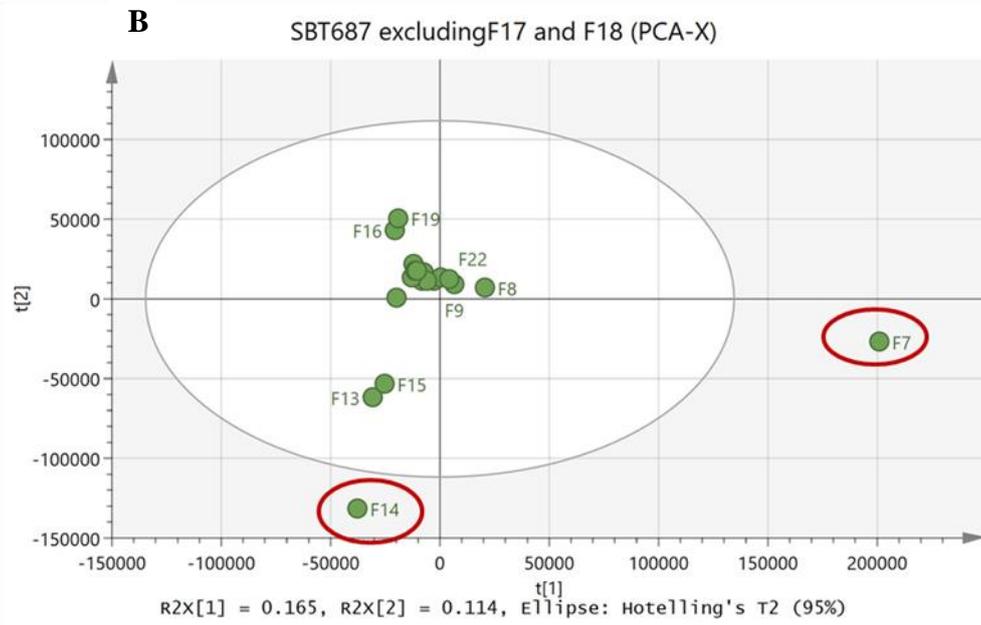
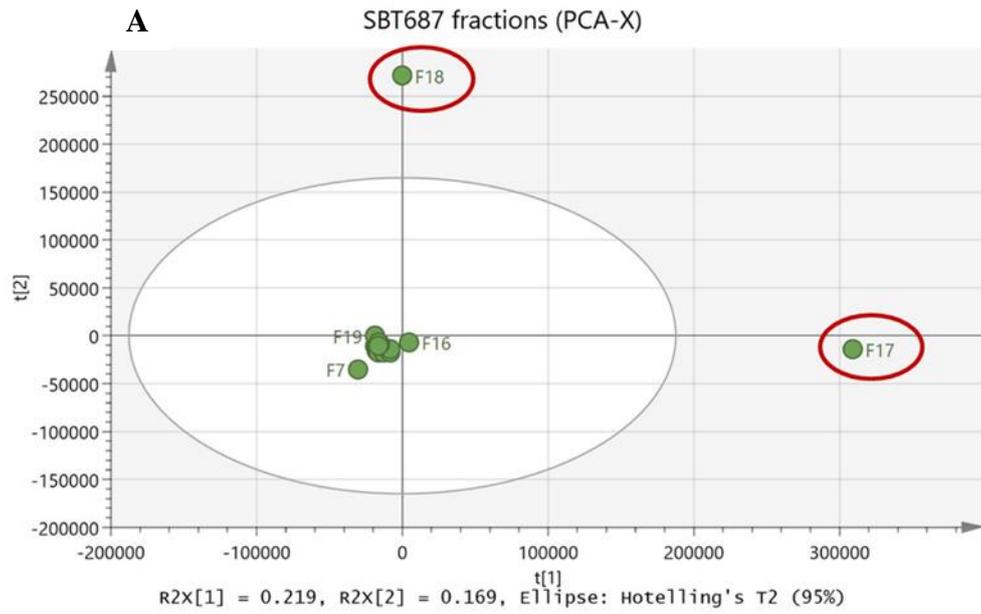
**Figure 4-6 Base peak plot chromatogram (HR-LC/MS) of SBT687 fractions in negative (A) and positive (B) ionization mode created on MzMine 2.10 version.**

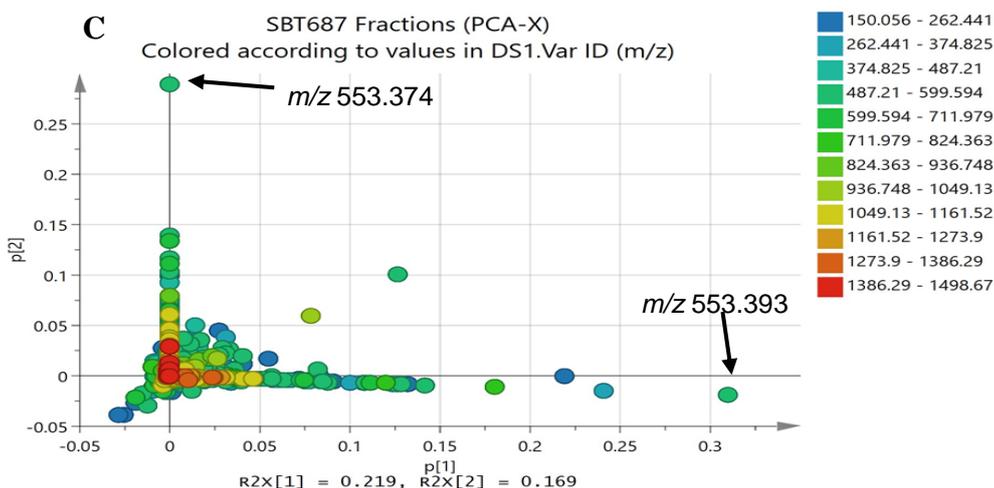


**Figure 4-7 Structure of loxistatin and 1-vinyl-8-hydroxy- $\beta$ -carboline as dereplicated from the negative and positive ionisation mode, respectively.**

PCA was performed on the HR-LC/MS data of 22 isolates. PCA plots were generated from four components, a  $R^2X$  value of 1 was achieved using Pareto scaling, indicating a good model for a metabolomics data set to fit ( $R^2X > 0.5$ , Umetrics) and  $Q^2$  of 1 which specifies good predictive ability of the model. The difference between  $R^2X$  and  $Q^2$  was 0, as both were equal to 1. The initial value of  $R^2X$  was 0.77 and  $Q^2$  -0.53 nevertheless, after adding 21 components the score of 1 for both parameters was achieved. The presence of strong outliers and differences in chemical profiles could be deduced from the  $R^2X[1]$  and  $R^2X[2]$  values, which gives the percent variation between and within clusters, respectively. PCA plots reveal patterns and trends to enable visualization similarities and differences between samples as well as detect outliers in the dataset. Outliers are observed to be situated outside the ellipse. Outlying samples could be indicated by their uniqueness in chemical profile as demonstrated by F17 and F18 encircled in red in Figure 4.8A. In contrast, chemical similarities were exhibited by the rest of the fractions clustering together on the upper left quadrant. In order to confirm the similarity of the clustering fractions, a new PCA scores plot (Figure 4.8B) was generated by excluding F17 and F18. Fractions F14 and F7 became the new outliers, with F9 and F16 located on the upper left quadrant and F13 and F15 were located being on the bottom left quadrant. While the remaining fractions clustered together.

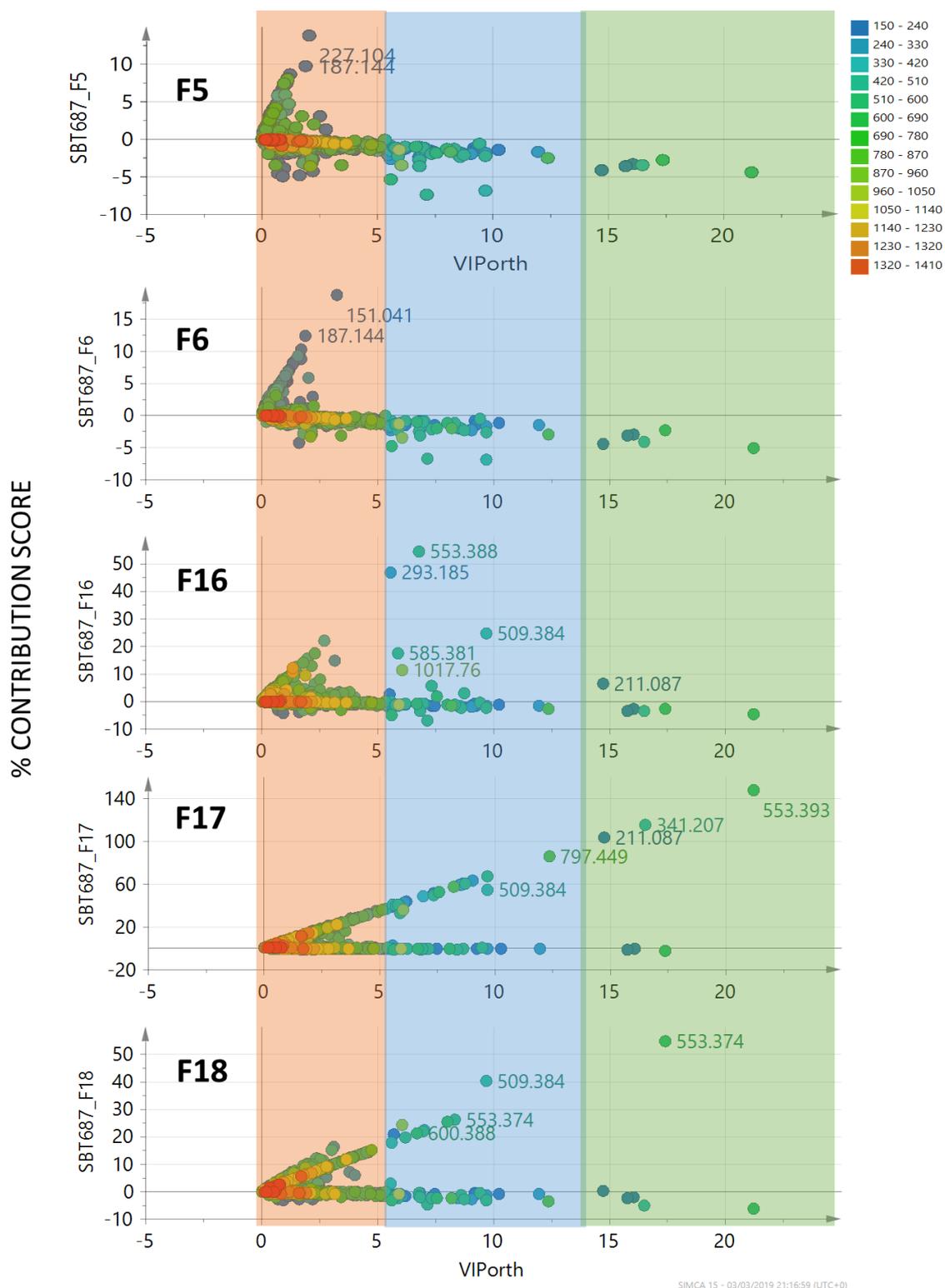
The loadings plot represented the molecular ion masses in  $m/z$  of the secondary metabolites of respective samples found on the corresponding quadrant position on the scores plot. Discriminating metabolites were found with ion peaks at  $m/z$  533.393 and 533.374 for outliers F17 and F18, respectively (Figure 4.8C). The ion peak at  $m/z$  533.393 represented the major ion peak particularly found in F17 was dereplicated as manzamine H, as discussed above and further in section 4.4.2. The ion peak at  $m/z$  533.374 found no hits from the database. A dereplication study identified the metabolites to be terpenes and alkaloids (Table 4.7).





**Figure 4-8 SIMCA analysis of SBT687 fractions mass spectral data.** (A) PCA-X scores plot of SBT687 fractions, showing the outliers F17 and F18 encircled in red. (B) PCA-X scores plot of SBT687 fractions without F17 and F18, showing the new outliers F19 and F7 encircled in red. (C) Loadings plot acquired from the PCA-X (A) model colour coded according to  $m/z$  with labelled “end point” metabolites.

To predict and understand the type of chemistry that would be responsible for the insecticidal bioactivity of SBT687 extract and fractions, a supervised multivariate analysis was accomplished by subjecting the HR-LC/MS data sets to OPLS-DA. The sample fractions were grouped into two classes: active vs. inactive (Y variables) and by using a loading scatter and S-plots, respective metabolites and functional groups (X variables) could be pinpointed to be responsible for the bioactivity. This statistical model would assist in targeting the bioactive natural products for further isolation work. F5 and F6 were included on the active group to confirm the position of their metabolites on the loadings S-plot. VIP (Variable Influence/Importance in Projection) and % contribution scores were employed to statistically validate the predicted bioactive metabolites exhibited as “end point” features on the S-plot specifically those afforded by F5, F6, F16, F17, and F18 (Figure 4.9). The % contribution and VIP (which must be >1) scores must be directly proportional to the probability that a certain metabolite of interest is indeed responsible for the bioactivity. As shown, the predicted bioactive features for F5 and F6 had low VIP and % contribution scores, while those from F16, F17, and F18 gave % contribution scores of greater than 20% while the VIP scores were greater than 5.0.



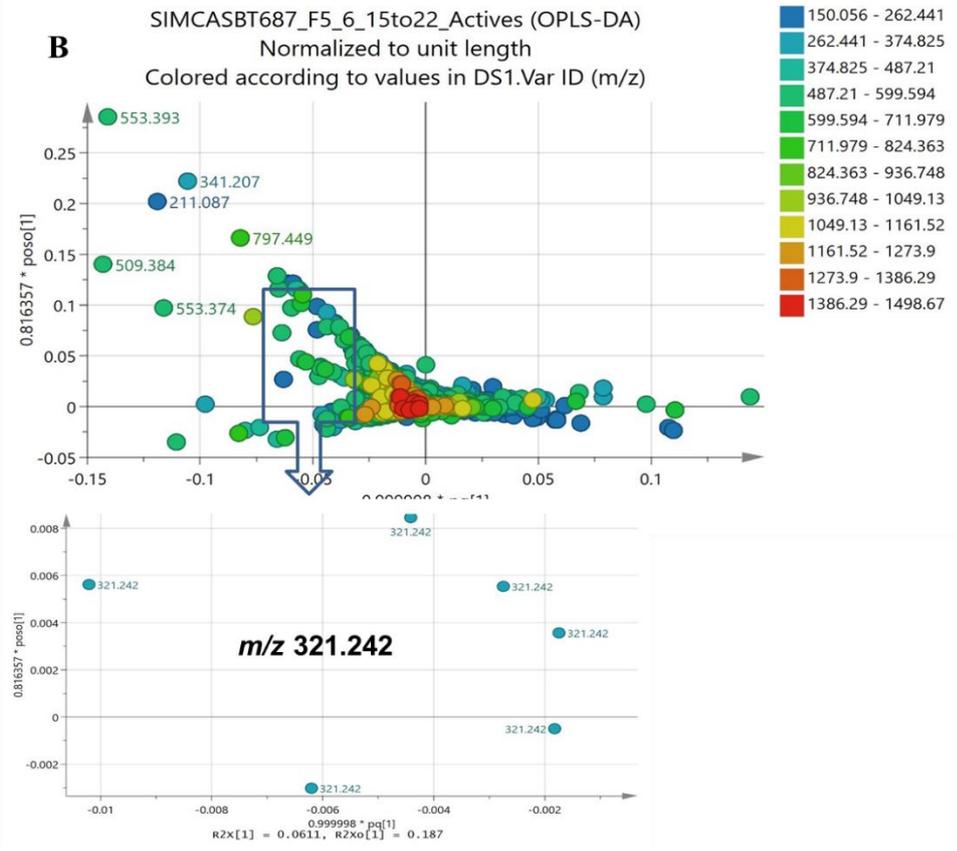
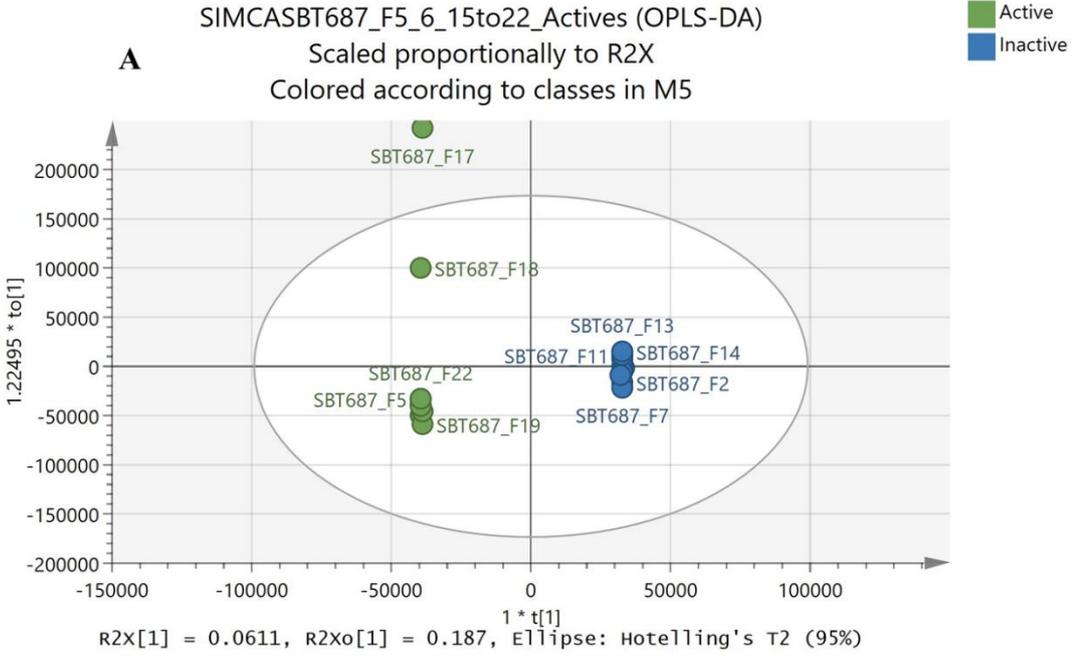
**Figure 4-9 Contribution scores plots for F5, F6, F16, F17, and F18 to indicate the orthogonal VIP (Variable Influence/Importance in Projection) scores on an OPLS model for the “end point” features for the respective fractions on the loadings S-plot as labelled here with their *m/z* ratios. Coloured regions designate increasing orthogonal VIP scores where red < 5, blue = 5 to 15, and green > 15.**

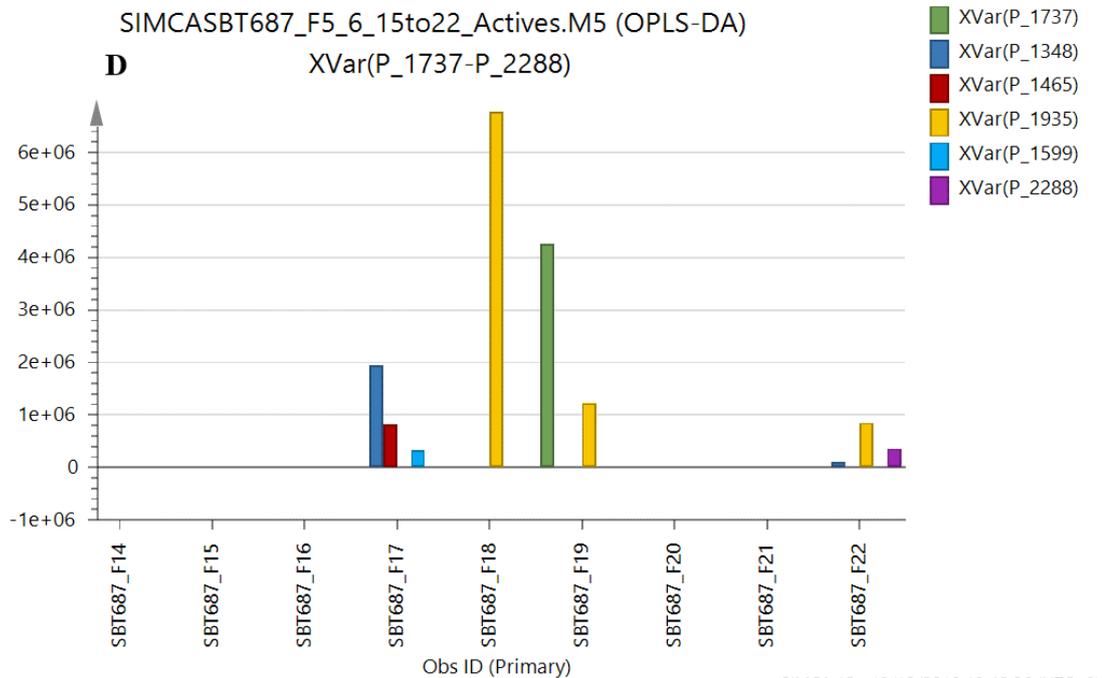
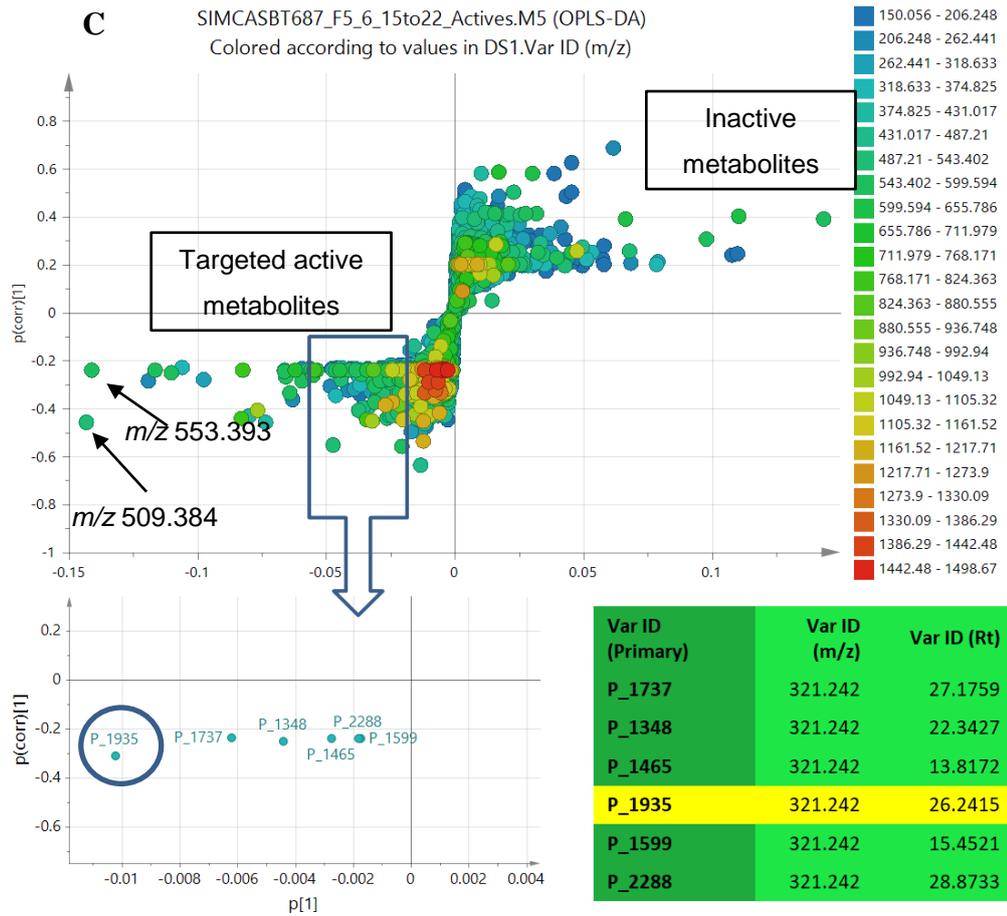
It was possible to differentiate the metabolites from the active vs inactive fractions from the OPLS-DA scores plot shown in Figure 4.10A. F17 remained an outlier as was indicated by the PCA model while F18 on the upper left quadrant was isolated from the other active fractions clustering together on the bottom left quadrant. The individual metabolites belonging to each group are shown on the loadings scatter plot (Figure 4.10B). The OPLS-DA loadings plot exhibited the discriminating active metabolites for F17 and F18 that separated them from the rest of the active fractions, which made the group occupy two different quadrants. The discriminant metabolites were the ion peaks at  $m/z$  553.390  $[M-H]^-$  putatively identified as manzamine H (section 4.4.2), 341.207  $[M-H]^-$  was found to be loxistatin, and 211.087  $[M-H]^-$  putatively identified as the alkaloid 1-vinyl-8-hydroxy- $\beta$ -carboline, and at  $m/z$  509.384  $[M+H]^+$  was dereplicated as polyhydroxylated sterol 1 $\beta$ -3  $\beta$ -5 $\alpha$ , 6  $\beta$  tetrahydroxy sterol. The ion peaks at  $m/z$  797.449  $[M-H]^-$  and 509.384  $[M+H]^+$  did not afford any hits or match from the database. An expansion of the loadings scatter plot was performed to show the ion peaks at  $m/z$  321.232  $[M+H]^+$  (enclosed by the blue box), which was particularly detected in F18.

The discriminating metabolites shown on the active side of the loadings-plot were also indicated on the “end point” of the active side of the S-plot (Figure 4.10C). An expansion of the S-plot (enclosed by the blue box) for the extracted ion peaks at  $m/z$  321.232  $[M+H]^+$ , a major ion particularly detected in F18, confirmed its position to be on the active side of the plot. The metabolites representing the ions peaks found at  $m/z$  321.232 were tabulated with their VarID,  $m/z$  and Rt. A variable column trend plot of the selected ion peaks indicated their relative abundance in the active fractions (Figure 4.10D). The ion peak with a VarID of P\_1935 and a retention time (Rt) of 26.24 min was found at its highest abundance in F18. P\_1935 was putatively identified as dolatriol. However, structure elucidation of the isolated compound as described under section 4.4.1 was proven otherwise as the acid derivative of manoyl oxide. Manoyl oxide derivatives were first reported from *Pinus* needles (Andersson et al., 1990, Elliger et al., 1976, Zinkel et al., 1985) and more recently from liverwort (Guo et al.,

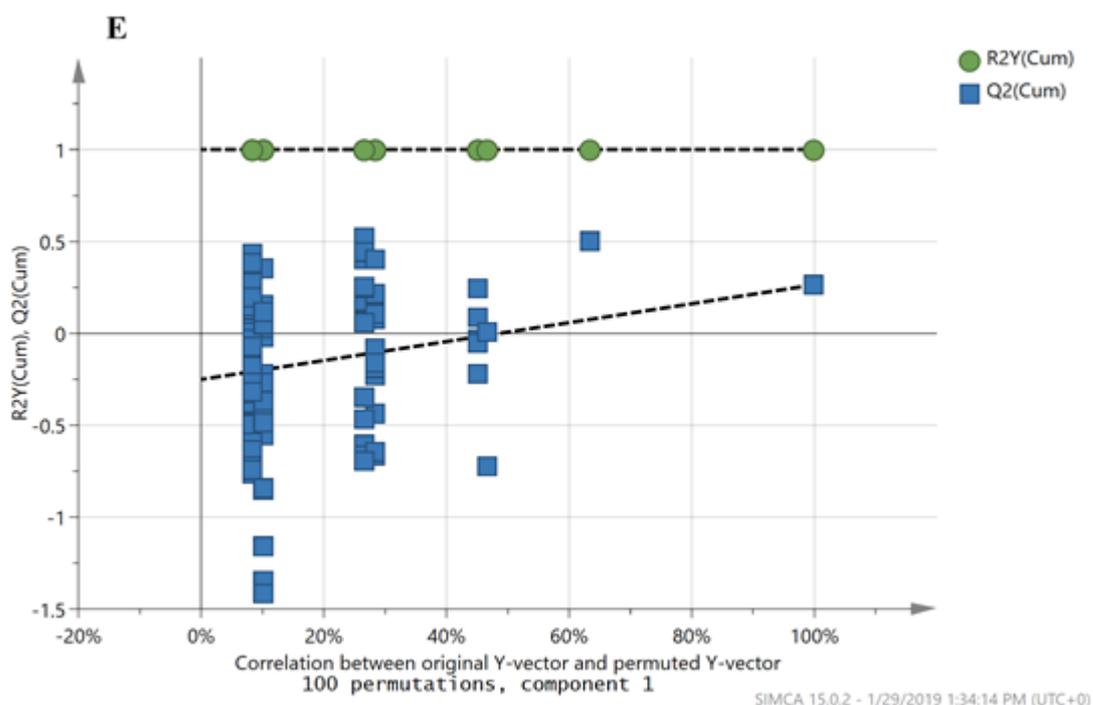
2010). The match was not found through the in-house dereplication macro because all terrestrial sources were filtered out during the search.

The OPLS-DA model was validated by a permutation test (Figure 4.10E). Although the  $R^2Y$  was 1, the  $Q^2$  value was not considered significant at 0.264. These values specified good fitting but poor prediction with  $Q^2 < 0.5$ . Furthermore, the difference between  $Q^2$  and  $R^2Y$  was 0.736 that is greater than 0.3 could indicate overfitting when the ideal difference must be less than 0.3. The resulting  $Q^2Y$  intercept from the permutation test gave a value of  $-0.25$ , which is less than 0 indicating the model's validity. The  $R^2X[1]$  at 0.0615 specified the variation between groups at 6.15% while the  $R^2X_0[1]$  at 0.187 gave an 18.7% variations within the group. The variation within the group is greater than between groups signified the existence of more unique chemical profiles found in fractions F17 and F18 that fail to cluster among the other active fractions. It is normally expected that the variation between groups should be greater than within the groups. The higher percentage of variation within the group could have caused the low predictability score. The targeted bioactive metabolites were dereplicated and cross-matched with the Antimarin and DNP databases to provide structural details as listed in Table 4.7. It is worth mentioning that 70% of the targeted metabolites afforded no hits or match from the database, while 5 of which were significant at  $P < 0.05$ . In comparison to the results of the TLC profiles, F17 and F18 afforded a higher number of metabolites than that of F16 on their MS profiles. On reversed phase TLC, F17 and F18 illustrated the presence of more non-polar components, and hence, the low RF values, which do not reflect the number of eluting components when compared to F16. F16 yielded more polar to semi-polar components demonstrating a better distribution and elution of the metabolites on the chromatogram and hence, F16 revealed a higher number of metabolites on reverse phase TLC.





SIMCA 15 - 12/12/2018 13:45:26 (UTC+0)



**Figure 4-10 MVDA analysis of mass spectral data of SBT687 fractions.** (A) OPLS-DA scores plot of active versus inactive fractions. (B) OPLS-DA loadings scatter plot of active versus inactive metabolites colour coded according to their  $m/z$ , with expansion of the quadrant enclosing the isolated target active metabolite at  $m/z$  321.242 (C) S-plot with expansion of the quadrant enclosing the isolated target active metabolite at  $m/z$  321.242. Table shows VarID,  $m/z$  and  $R_t$  for the extracted ion peaks. Encircled in blue is the major ion peak found in F18. (D) Variable column trend plot for ion peaks at  $m/z$  321.242 to show their relative abundance in the active fractions. (E) Permutation test (100 permutations) for OPLS-DA model of SBT687 fractions on their insecticidal activity against *Lepeophtheirus sp*

**Table 4-7 Dereplication table of predictive possible bioactive metabolites by OPLS-DA loading S-plot.** Highlighted rows represent the isolated compounds from SBT687 fractions.

<b>MzMine ID</b>	<b>m/z</b>	<b>Rt (min)</b>	<b>Peak Area (highest)</b>	<b>MW</b>	<b>Molecular Formula</b>	<b>Name</b>	<b>Source</b>	<b>P values</b>
P_2252	187.1442	6.34	5.95E+07 (F6)	186.1369	no prediction	No hits	No hits	0.12
N_7715	151.0412	9.94	2.37E+08 (F6)	152.0484	no prediction	No hits	No hits	0.29
N_731	341.2070	15.22	1.32E+09 (F17)	342.2147	C <sub>17</sub> H <sub>30</sub> N <sub>2</sub> O <sub>5</sub>	loxistatin	<i>Aspergillus japonicus</i> tpr-64; semisynth.	0.30
N_3941	569.3841	18.40	1.33E+08 (F16)	570.3914	C <sub>35</sub> H <sub>54</sub> O <sub>6</sub>	dihydrocaulerpenyne palmitoleic acid	Chlorophyta <i>Caulerpa</i> <i>prolifera</i>	0.28
N_3944	419.2900	19.23	5.65E+07 (F16)	420.2978	no prediction	No hits	No hits	0.28
N_5471	797.4491	19.23	7.42E+08 (F17)	798.4564	no prediction	No hits	No hits	0.28
N_3943	554.3918	20.37	8.35E+07 (F16)	555.3991	no prediction	No hits	No hits	0.27
P_896	1017.7600	21.23	1.74E+08 (F17)	1016.753	no prediction	No hits	No hits	0.05
N_3947	543.3594	21.47	5.94E+07 (F16)	544.3667	no prediction	No hits	No hits	0.24
N_3940	405.3108	21.47	1.95E+08 (F16)	406.3181	no prediction	No hits	No hits	0.26
N_99	553.3885	21.50	1.21E+09 (F16)	554.3958	C <sub>35</sub> H <sub>54</sub> O <sub>5</sub>	lagunapyrone C	marine actinomycete CNB 984	0.26
N_1599	553.3927	21.52	2.18E+09 (F17)	554.3999	C <sub>36</sub> H <sub>50</sub> N <sub>4</sub> O	manzamine H	Porifera <i>Ircinia</i> sp	0.28
N_7748	553.3745	21.52	1.47E+09 (F18)	554.3817	no prediction	No hits	No hits	0.28

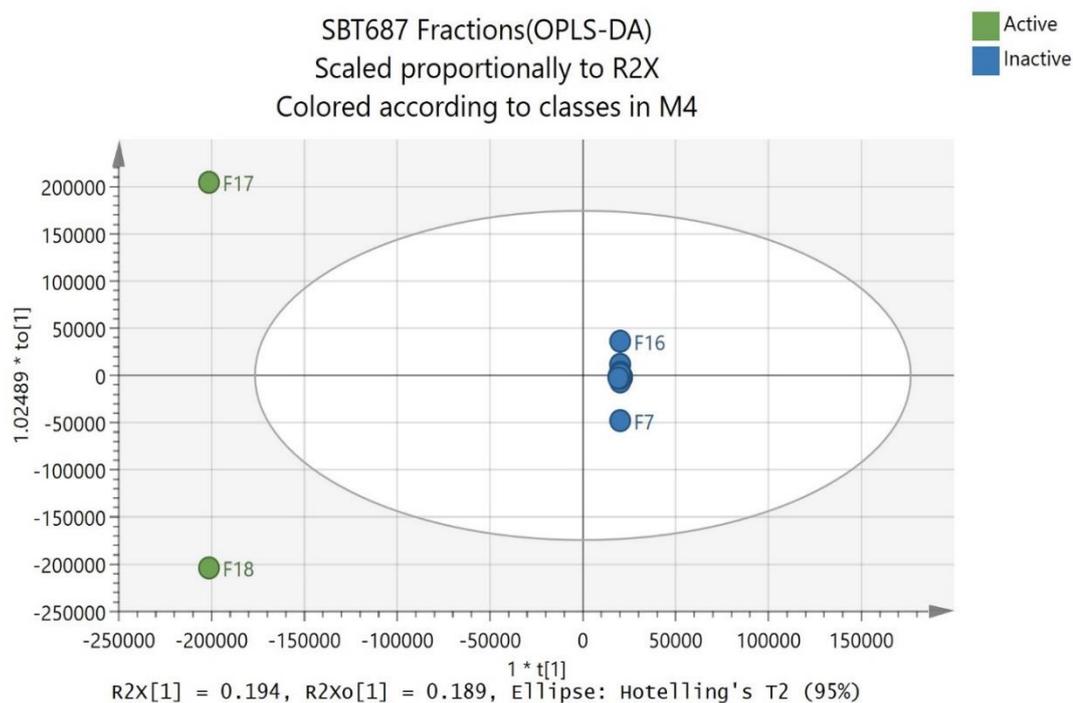
P_669	509.3838	21.74	4.73E+08 (F17)	508.3865	C <sub>30</sub> H <sub>52</sub> O <sub>6</sub>	1β,3β,5α,6β- tetrahydroxy sterol	Cnidaria <i>Sarcophyton glaucum</i>	0.03
N_3370	293.185	22.16	7.51E+08 (F16)	294.1923	no prediction	No hits	No hits	0.20
P_1935	321.242	26.24	6.77+06 (F18)	320.2350	C <sub>20</sub> H <sub>32</sub> O <sub>3</sub>	dolatriol	Mollusca <i>Dolabella auricularia</i>	0.16
N_1684	429.2403	31.22	1.30E+08 (F21)	428.233	no prediction	No hits	No hits	0.03
P_1681	413.2662	31.38	1.69E+08 (F21)	412.2589	no prediction	No hits	No hits	0.04
P_577	803.5434	32.15	1.84E+08 (F21)	802.5361	no prediction	No hits	No hits	0.04

Another OPLS-DA model was generated by grouping the most unique fractions, F17 and F18 as the only active fractions (Figure 4.11). It was possible to realize the active metabolite belonging to F17 as labelled on the S-plot with its Var ID (N\_1599) and  $m/z$  553.393 (Figure 4.12A). An expansion of the S-plot by filtering the  $m/z$  from the putatively dereplicated manzamines derivatives exhibited the distribution of these compounds between the active and inactive fractions (Figure 4.12B). From all the manzamine derivatives, only the ones with  $m/z$  551.377 and 567.372 were in the inactive fractions on the upper right quadrant. A variable column trend plot for N\_1599 disclosed its highest relative abundance to be in F17 (Figure 4.12C). A second variable column trend plot was also generated for the ion peaks putatively identified as manzamine derivatives to determine their relative distribution in the different fractions. All manzamine derivatives were found to be concentrated in F17 while N\_1599 was found to be the most abundant as well (Figure 4.12D).

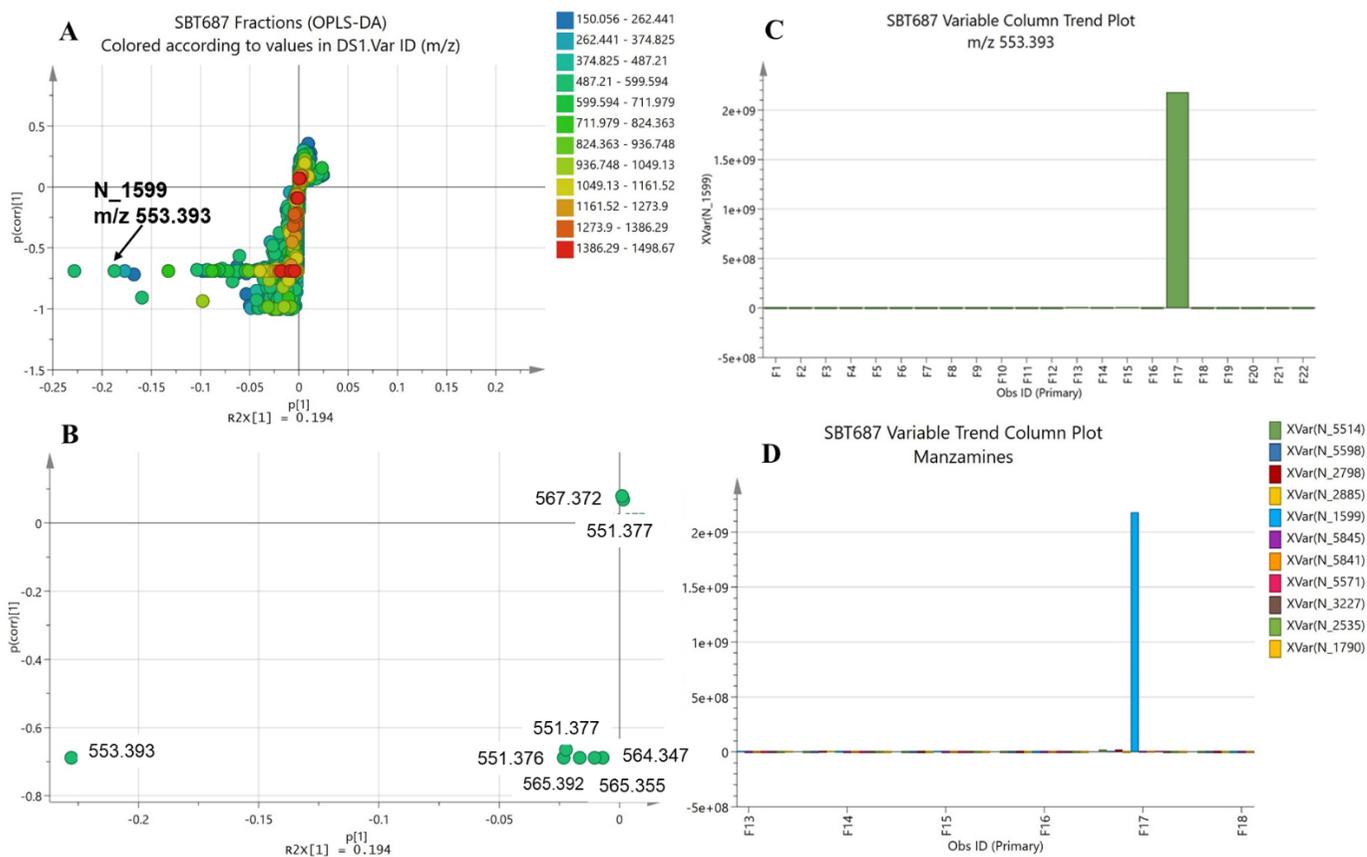
A permutation test (Figure 4.13) was accomplished to authenticate the validity of the current model, the  $R^2Y$  was 1 and  $Q^2$  was 0.911, while the  $Q^2Y$  intercept was at  $-0.33$ . These values specified good fitting and prediction, while the  $Q^2Y$  intercept, which is less than 0 indicated the validity of the OPLS-DA model. Furthermore, the difference between  $Q^2$  and  $R^2Y$  was 0.089 showing again a good fitness and prediction.  $R^2X[1]$  was at 0.194 which means that the variation between groups was 19.4% and  $R^2X_0[1]$  was at 0.189 with a variations within groups of 18.9%. The variation between and within groups was almost equal in this case.

The dereplication Table 4.8 lists all the manzamines produced by SBT687 with their intensities in the respective fractions where they were detected. Interestingly, F17 afforded the highest relative abundance of a mixture of manzamine congeners. However, F16 yielded the highest relative concentration of tetrahydro-8-hydroxymanzamine at  $m/z$  567.3699 and with molecular formula of  $C_{36}H_{48}N_4O_2$ . Manzamines are sponge-derived  $\beta$ -carboline alkaloids that were firstly alleged to be exclusive sponge metabolites (Taylor et al., 2007). However, manzamines have been isolated from different genera of sponges (AlTarabeen et al., 2015, Edrada et al., 1996,

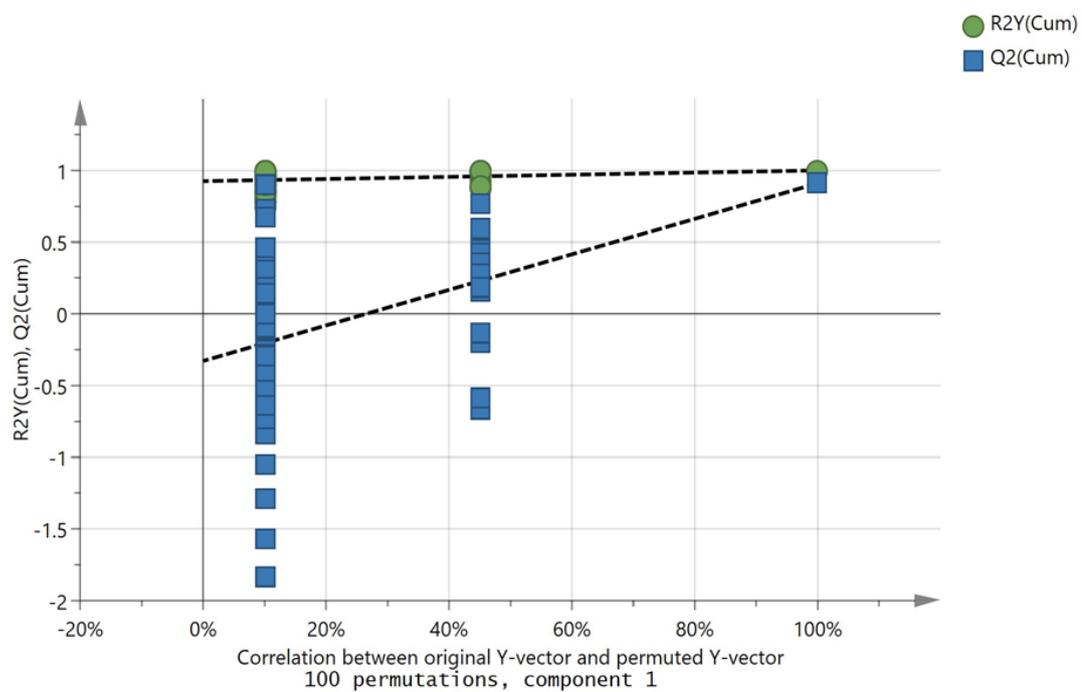
El Sayed et al., 2001, Furusato et al., 2014, Ichiba et al., 1994, Ichiba et al., 1988, Ali et al., 2017, Kobayashi et al., 1994, Kondo et al., 1992, Kubota et al., 2017, Ohtani et al., 1995, Peng et al., 2003, Rao et al., 2006, Rao et al., 2003, Sakai et al., 1986, Sakai et al., 1987, Tsuda et al., 1996, Watanabe et al., 1998, Yamada et al., 2009, Zhou et al., 2000). Nevertheless, recent research has proven that manzamines are microbial metabolites. Manzamines are produced by sponge-derived microorganisms like those of the genus *Micromonospora* (Waters et al., 2014).



**Figure 4-11 OPLS-DA of SBT687 fractions with only F17 and F18 considered for the active group.**



**Figure 4-12 MVDA analysis of mass spectral data of SBT687 fractions F17 and F18 . (A)** S-plot indicating the targeted putative compound labelled with its *m/z* and VarID number for fractions F17 and F18. **(B)** Expansion of the targeted manzamine derivatives labelled with their *m/z* **(C)** Variable column trend plot of the selected metabolite N\_1599 at *m/z* 553.393, showing which fractions were yielding the metabolite. **(D)** Variable column trend plot of all dereplicated manzamine derivatives in SBT687 showing manzamine-yielding fractions.



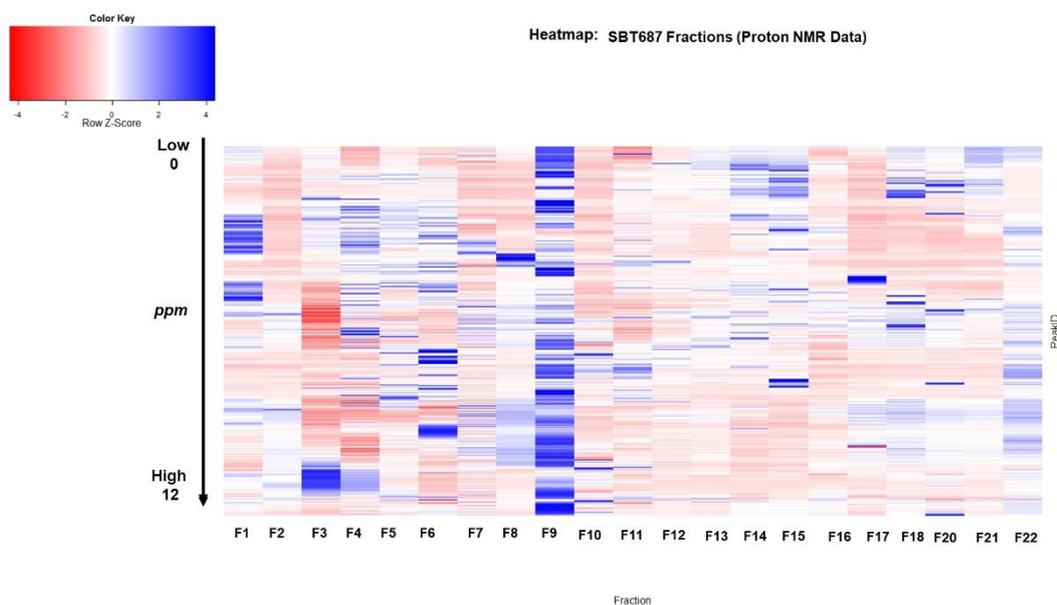
**Figure 4-13** Permutation test (100 permutations) for the OPLS-DA model of SBT687 fractions on their insecticidal activity against *Lepeophtheirus* sp.

**Table 4-8 Dereplication table of the manzamines produced by the SBT687 fractions.** Highlighted row represents the compound found in the S-plot endpoint of the OPLS-DA analysis.

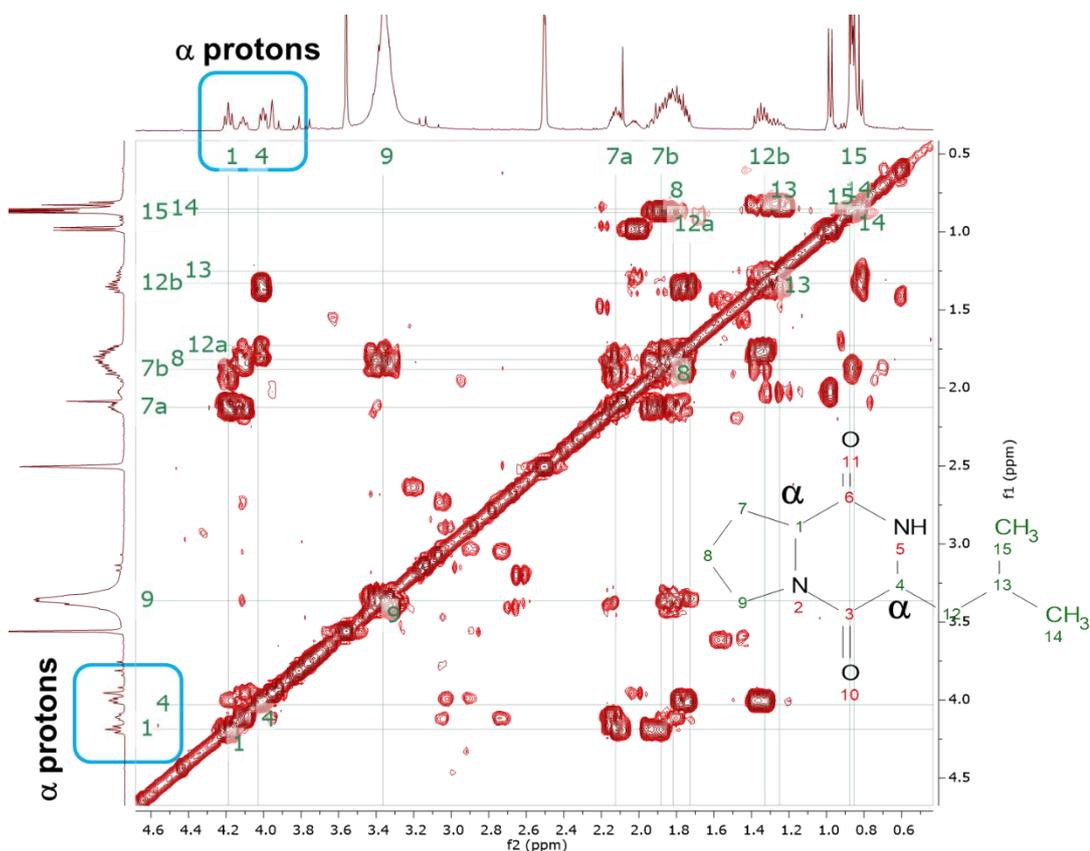
<b>MzMineID</b>	<b>m/z</b>	<b>Rt (min)</b>	<b>Peak Area (highest)</b>	<b>MWt</b>	<b>Molecular Formula</b>	<b>Name</b>	<b>Source</b>
N_5514	551.3763	22.926	2.26E+07(F17)	552.3836	C <sub>36</sub> H <sub>48</sub> N <sub>4</sub> O	manzamine D	Porifera <i>Amphimedon</i> sp
N_5598	551.3765	24.329	8475776(F17)	552.3838			
N_2798	551.3768	21.812	2.24E+07(F17)	552.3841			
N_2885	551.377	18.302	7982102(F14)	552.3843			
N_1599	553.3927	21.522	2.18E+09(F17)	554.3999	C <sub>36</sub> H <sub>50</sub> N <sub>4</sub> O	manzamine H	Porifera <i>Ircinia</i> sp
N_5845	564.3473	22.923	2101300(F17)	565.3545	C <sub>36</sub> H <sub>45</sub> N <sub>4</sub> O <sub>2</sub> <sup>+</sup>	8-hydroxymanzamine A	Porifera <i>Pachypellina</i> sp
N_5841	565.3553	20.106	4589200(F17)	566.3626	C <sub>36</sub> H <sub>46</sub> N <sub>4</sub> O <sub>2</sub>	manzamine J <i>N</i> -oxide	Porifera <i>Xestospongia ashmorica</i>
N_5571	565.392	22.366	1.14E+07(F17)	566.3993	C <sub>37</sub> H <sub>50</sub> N <sub>4</sub> O	N-methyl-epi-manzamine D	Porifera
N_3227	567.3699	22.316	316590.9(F14)	568.3772	C <sub>36</sub> H <sub>48</sub> N <sub>4</sub> O <sub>2</sub>	tetrahydro-8-hydroxymanzamine A	Porifera <i>Cribrochalina</i> sp
N_3495	567.3699	18.681	4.46E+07(F16)	568.3772			
N_2535	567.3719	16.312	1361614(F13)	568.3792			
N_1790	567.3719	15.435	2381570(F13)	568.3792			

### 4.3.1.2 NMR Data

The  $^1\text{H}$  NMR dataset of the fractions was treated correspondingly as the mass spectral data. A heatmap was generated and organized according to chemical shifts to similarly show the differences in chemical profiles of the 21 fractions. Fraction 19 was excluded due to its very low and insufficient yield, which was less than a mg. The heatmap on Figure 4.14 shows that F9 has the greatest number and intensity of the blue bands distributed at a broad range from 0-12 ppm. This fraction was also found to yield diketopiperazines or cyclized dipeptides that includes cyclo-(prolyl-leucine) and cyclo-(prolyl-isoleucine) with crowded peaks in the aliphatic region and particularly at 4 to 5 ppm for the  $\alpha$ -protons as also shown on its COSY spectrum (Figure 4.15). Fraction 18 showed higher intensity peaks on the aliphatic (0 to 3 ppm) and olefinic/sugar (4 to 6 ppm) region while lower intensity peaks were observed in the aromatic (6 to 9 ppm) region. The other fractions displayed a diverse chemical profile with changing intensities of resonances in various regions.



**Figure 4-14** Heatmap based on  $^1\text{H}$  NMR data displaying the distinct metabolic profiles of 21 fractions of SBT687.



**Figure 4-15**  $^1\text{H}$ - $^1\text{H}$  COSY (400 MHz) spectrum in  $\text{DMSO-d}_6$  of the cyclized dipeptide shown in blue boxes are the  $\alpha$  protons that is characteristic for amino acids.

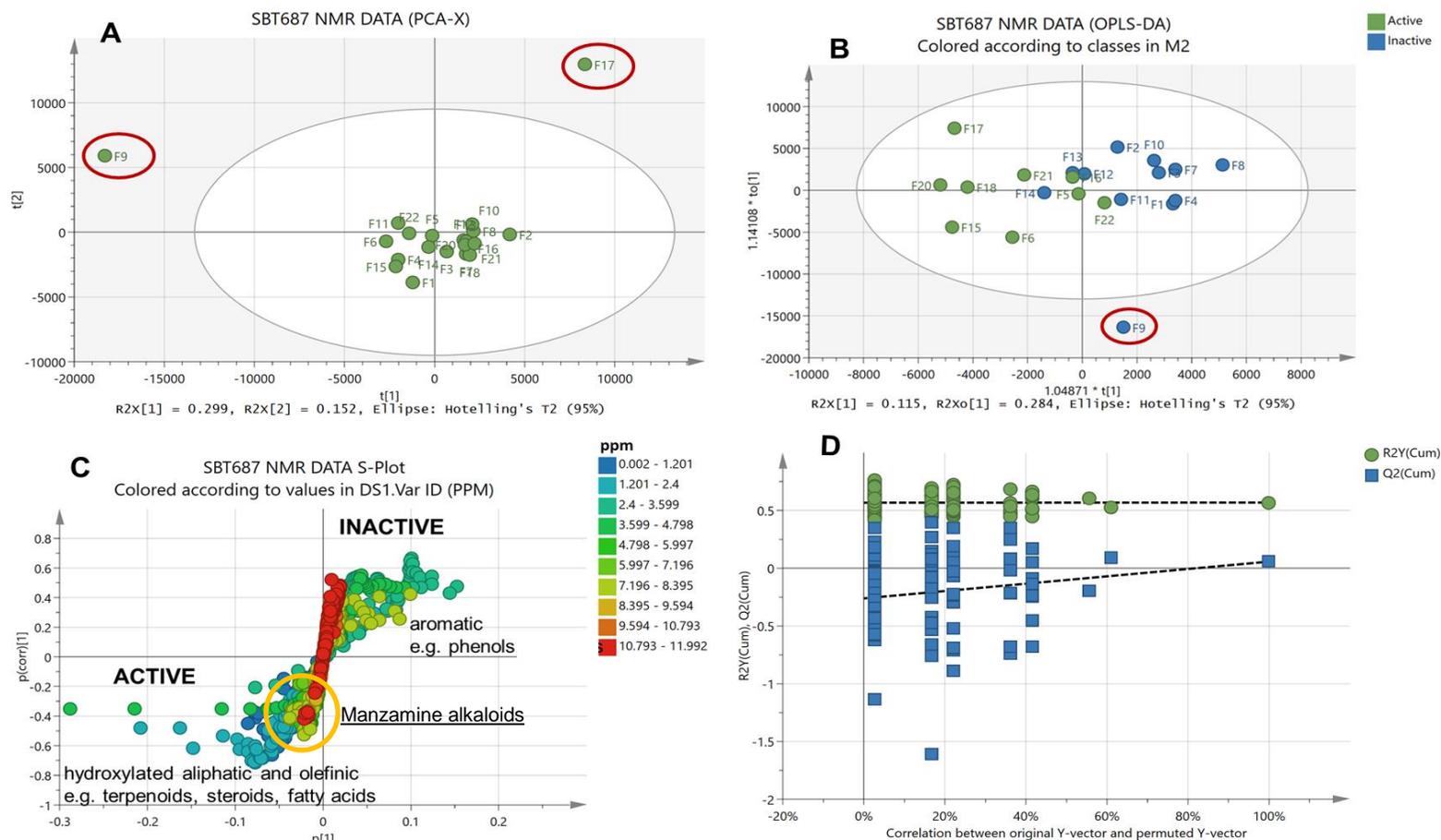
PCA was performed on the  $^1\text{H}$  NMR data of 21 fractions. PCA plots were generated from four components, a  $R^2X$  value of 0.99 was achieved using Pareto scaling, indicating a good fit and  $Q^2$  of 0.54 which specified a good predictive ability of the model for a metabolomics dataset. The difference between  $R^2X$  and  $Q^2$  was 0.459, which should be less than 0.3 to confirm complete absence of overfitting. The high diversity on spectral profiles between fractions could account for the larger difference between  $R^2X$  and  $Q^2$ .

On the PCA-X scores plot (Figure 4.16A), F9 and F17 appeared as outliers while the rest of the others fractions clustered together. The occurrence of unique functional groups or structural chemical shifts between the active and inactive fractions was accomplished by OPLS-DA as shown through a scores plot in Figure 4.16B. F9 was once again an outlier. The  $R^2X[1]$  was 0.115 which means that the variation between groups is 11.5% and  $R^2X_0[1]$  was 0.284 with a variations within groups at 28.4% indicating that  $R^2X_0[1] > R^2X[1]$ . The low percent variation between groups was

evident by the overlapping of active and inactive fractions between and across their respective quadrants. The percent variation within groups was almost double of that between the groups and could be accounted for the dispersed distribution of the fractions, which signified a diversity in structural features.

The S-plot (Figure **4.16C**) showed that the bioactive fractions contained high density of resonances between 1.201 to 5.997 ppm that corresponds to structural features for hydroxylated aliphatic and olefinics as those found in terpenoids, steroids and fatty acids. The inactive fractions were producing mostly aromatic compounds such as phenols with most of the chemical shifts were between 4.798 and 8.395 ppm. Alkaloidal resonances that resembles those of the manzamines can also be observed by closely positioned signals between 7.0 to 12.0 ppm. The data points were found on the active side of the S-plot but still relatively distant from the zero line of the y-axis. The further are the data points from the zero line of the y-axis, the higher is the % confidence interval, which is also affected by the intensity of the resonances. The presence of the manzamines was clearly detected in the mass spectral data of active fractions considering the low yield of F17 at only 1.3 mg. This is a typical classical scenario, where mass spectrometry can simply detect ionisable compounds even at ng levels, while NMR spectroscopy do require mg levels of the respective metabolites in a mixture to be detected.

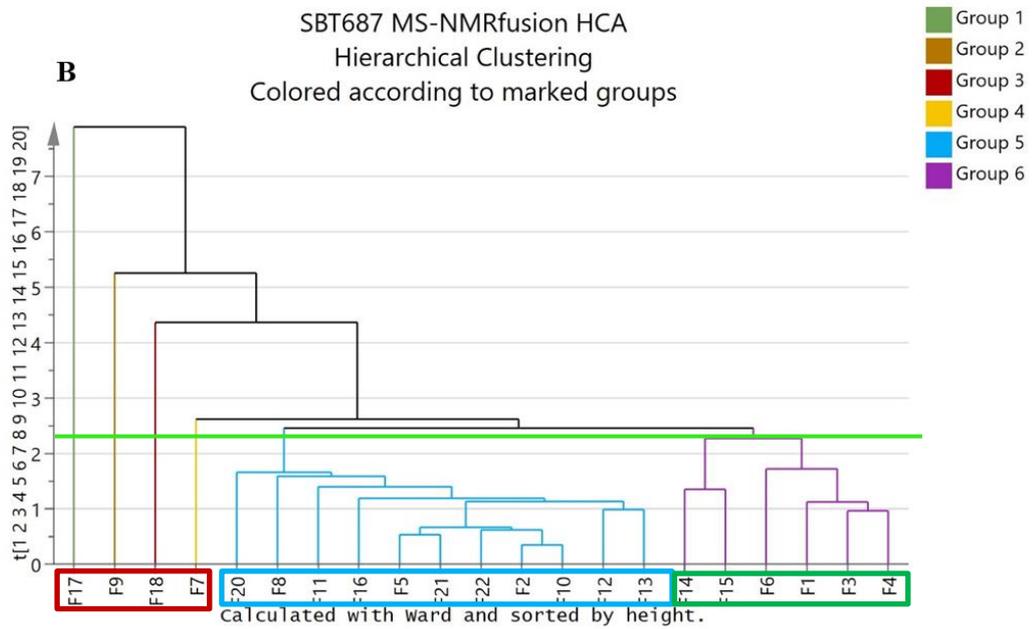
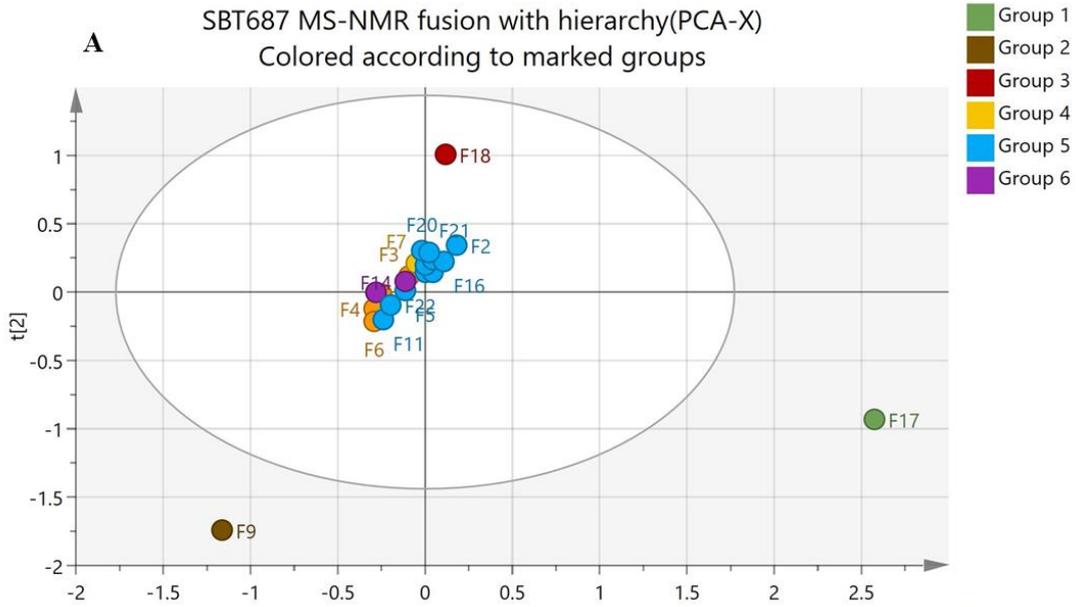
The OPLS-DA model was validated by a permutation test (Figure **4.16D**), where  $R^2Y$  was at 0.568 and  $Q^2$  was at 0.057, while the  $Q^2Y$  intercept was at  $-0.262$ . These values specified a good fit where  $R^2Y > 0.5$  but predictability was not as good. The difference between  $Q^2$  and  $R^2Y$  was 0.511 which was still greater than 0.3, indicating an overfit data. However, the  $Q^2Y$  intercept was less than 0, which indicated a valid OPLS-DA model.

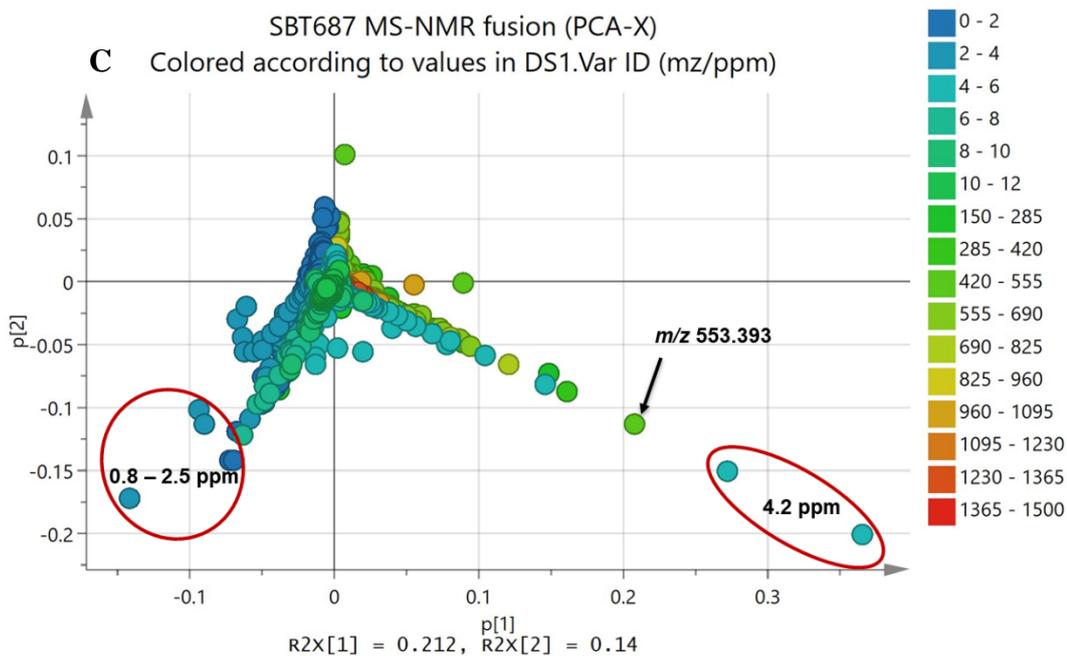


**Figure 4-16 MVDA analysis of <sup>1</sup>H NMR spectral data of SBT687 fractions . (A)** PCA-X score plot of SBT687 fractions with outliers encircled in red **(B)** OPLS-DA of SBT687 fractions according to their insecticidal activity against *Lepeophtheirus* sp. **(C)** S-plot for SBT687 fractions achieved from OPLS-DA showing the main compounds being produced. **(D)** Permutation tests (100 permutations) for the OPLS-DA model of SBT687 fractions of their insecticidal activity against *Lepeophtheirus* sp.

#### 4.3.1.3 HR-LC/MS Data – NMR Data Fusion

The fusion of both spectral dataset from the  $^1\text{H-NMR}$  and HR-LCMS was also subjected to PCA. PCA-X scores and loadings plot as well as hierarchy-clustering analysis (HCA) dendrogram were generated. The PCA-X (Figure 4.17A) was coloured coded according to the marked groups on the HCA (Figure 4.17B) to verify the separation of the outliers from both the NMR and mass spectral datasets. Fusion of the spectral data presented F9 and F17 as the outliers while F18 was positioned on the upper right quadrant isolated from the rest of the other 18 clustering fractions.  $R^2X$  and  $Q^2$  values were of 1, which specified good fitness and predictive ability of the model. The HCA allowed to visualize the best clustering order of the fractions according to the chosen dendrogram cut off. Fractions that clustered together shared the same chemical profile or yielded similar major metabolites. The HCA afforded six groups with F17, F9 and F18 and F7 being the most unique fractions and the rest sharing more similar features. The fractions F20, F8, F11, F6, F21, F22, F2, F10, F12 and F13 form one group while F14, F15, F6, F1, F3 and F4 form another group. The PCA loadings plot were colour coded according to their  $m/z$  and chemical shifts in ppm (Figure 4.17C). Encircled endpoints representing the outliers were labelled with their corresponding chemical shifts. It was demonstrated that the outlier F9 on the lower left quadrant, had the highest density of resonance between 0.8 to 2.5 ppm in comparison to the other fractions. According to these chemical shifts, F9 could be described to yield a higher abundance of aliphatics and olefinic type of compounds. Alternatively, on the opposite quadrant, which represents the second outlier F17, the quadrant is dominated by data points colour coded from lighter blue to darker green representing chemical shifts from 4 to 9 ppm, indicating the presence heteroatom-bearing aliphatics and nitrogen-containing aromatics. These resonances are typical to those of the manzamine alkaloids and this is further confirmed by  $m/z$  553.393 dereplicated as manzamine H found in the same quadrant of the loadings plot. The concatenated data sorted and brought clarity to the positions of the overlapping NMR data points observed in section 4.3.1.2.



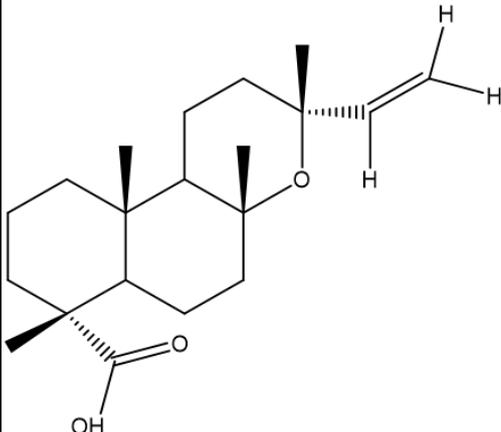


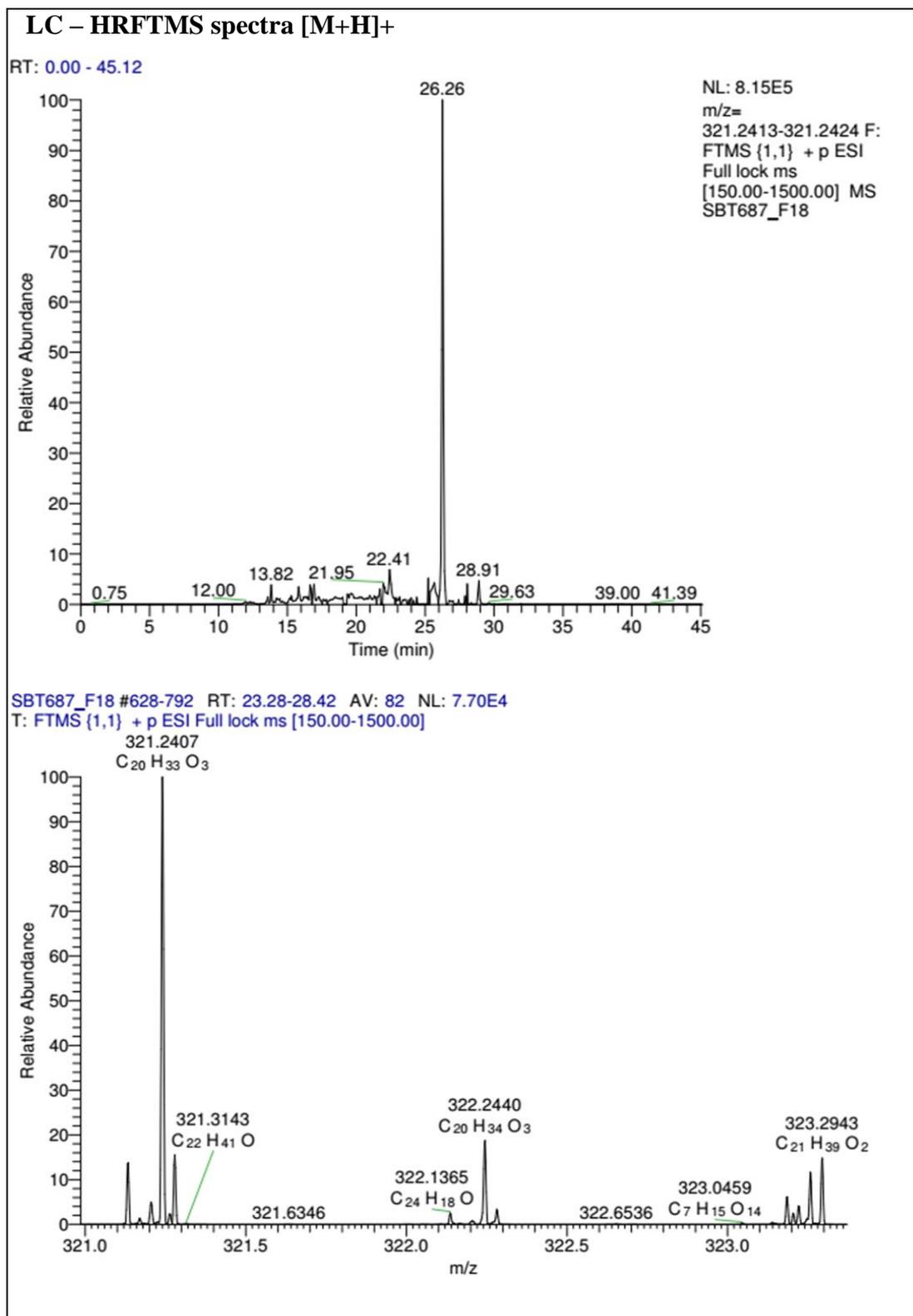
**Figure 4-17 MVDA MS-NMR data fusion analysis of SBT687 fractions.** (A) PCA of fused NMR and MS spectral data for 21 fractions colour coded according to HCA groupings. (B) HCA of SBT687 fractions based on HR-LCMS and  $^1\text{H}$  NMR coloured according to marked groups showing the chemical similarities between fractions and with green line cut-off. The fractions enclosed in a red box represent the group of outlying and isolated fractions while the other two main groups were enclosed by a blue and green box, respectively. (C) Loadings scatter plot coloured according to their  $m/z$  and ppm shifts. Encircled endpoints were labelled with their ppm while one discriminating metabolite was labelled with its  $m/z$ .

## 4.4 Identification and structure elucidation of SBT687 compounds

### 4.4.1 4-epi-manoyl oxide acid

Table 4-9 4-epi-manoyl oxide acid.

Compound name (new compound): 4-epi-manoyl oxide acid
Fraction: Fraction 18 (F18) Retention time: 26.26 Synonym(s): <ul style="list-style-type: none"><li>• 8,13-epoxy-14-labden-19-oic acid</li><li>• 8,13-Epoxyabd-14-en-19-oic [(manoyl oxide)-19-oic] acid</li><li>• (3<i>R</i>,4<i>aR</i>,7<i>R</i>,10<i>aS</i>)-3,4<i>a</i>,7,10<i>a</i>-tetramethyl-3-vinyldodecahydro-1<i>H</i>-benzo[<i>f</i>]chromene-7-carboxylic acid</li></ul> Source: <i>Micromonospora</i> sp. N17 (SBT687) Amount of sample: 0.6 mg Percent yield: 0.2% Physical description: dark yellow oily sample Molecular formula: C <sub>20</sub> H <sub>32</sub> O <sub>3</sub> Molecular weight: 320.466 g/mol Exact mass: 320.2351 [α] <sub>D</sub> <sup>20</sup> Not measured due to insufficient quantities
 <p>Relative stereochemistry shown is based on nOe and after comparison of the NMR spectral data to published data.</p>



Based on the high-resolution mass spectral data, the ESI peak at 26.26 min in the positive mode was found at  $m/z$  321.2407 [M+H]<sup>+</sup>. This revealed the exact mass at 320.2351 g/mol, which established the molecular formula of  $C_{20}H_{32}O_3$  (Table 4.9).

The molecular formula and mass spectra data suggested it was probably a diterpene for a  $C_{20}H_{32}$ . The degree of unsaturation for  $C_{20}H_{32}O_3$  was 5 which corresponded to either to a benzyl unit with an olefinic or an acid substituent; or three aliphatic rings with an olefinic and/or an acid substituent. The  $^{13}C$  and Jmod NMR spectra exhibited signals for a carboxylic acid at ca. 180 ppm and exomethylene moiety at ca. 110 ppm. Alternatively, both  $^1H$  and  $^{13}C$  NMR spectra were devoid of any aromatic resonances with more signals observed in the aliphatic region.

The  $^1H$  NMR spectrum (Figures 4.18 and 4.19) of F18 exhibited four methyl signals within 0.78 to 1.23 ppm. All methyl proton signals were observed as singlets ( $\delta$ 0.78, 1.03, 1.19 and 1.23), which were assigned for H18, H17, H20 and H19. The methyl group signals were also seen in the Jmod spectrum at  $\delta$  15.80, 16.54, 25.70 and 28.78 as C18, C17, C19 and C20, respectively (Figure 4.20 and 4.21). One carboxyl carbon (C16) was observed at  $\delta$ 179.94 and two olefinic carbons for C14 and C15 were observed at 148.59 and 110.32 ppm, respectively. As found in the COSY spectrum (Figure 4.22), two olefinic signals at  $\delta$ 4.85, dd,  $J=10.75, 1.86$  Hz (H15a) and 5.09,  $J=17.16, 2.12$  Hz (H15b) correlated with each other with  $J$  coupling constant of ca. 2Hz, indicating the presence of an exomethylene moiety. Another olefinic methine proton (H14a) was shown at  $\delta$ 5.83,  $J=17.35, 10.71$  and exhibited *trans* ( $J=17.35$ ) and *cis* ( $J=10.71$ ) correlations with both protons H15b and H15a, respectively

Through  $^1H$ - $^{13}C$  HSQC (Figures 4.23 and 4.24), it was possible to complete the assignments of each proton to its corresponding carbon ( $^1J$   $^1H$ - $^{13}C$  bond). The  $CH_3$  and  $CH$  were marked by a red contour (negative phase) while  $CH_2$  was marked by a blue contour (positive phase). The methylene protons for positions 1, 2, 3, 6, 7 and 12 were split to  $H_a$  and  $H_b$  exhibiting resonances at  $\delta$ 0.86 (t,  $J=11.73$ ) and 1.56 (dd,  $J=13.40, 6.35$ ) for H<sub>2</sub>-1; 1.46 (d,  $J=11.18$ ) and 1.56 (dd,  $J=13.40, 6.35$ ) for H<sub>2</sub>-2; 1.46 (d,  $J=11.18$ ) and 1.65 (td,  $J=12.07, 2.90$ ) for H<sub>2</sub>-3; 1.13 (m) and 1.40 (m) for H<sub>2</sub>-6; 1.65 (td,  $J=12.07, 2.90$ ) and 1.35 (m) for H<sub>2</sub>-7; and 1.56 (dd,  $J=13.40, 6.35$ ) and 1.73 (dt,  $J=12.95, 5.20$ ) for H<sub>2</sub>-12. The latter  $CH_2$ s gave  $^1J$  correlations with carbon resonances at  $\delta$ 38.35, 17.62, 36.92, 22.42, 43.13, and 35.60 then assigned for C1, C2, C3, C6, C7, and C12, respectively. Quaternary carbons C4, C8, C10 and C13 with

resonances at  $\delta_{\text{C}}$  46.70, 74.78, 36.26 and 73.17, respectively were not observed in an HSQC spectrum. C8 and C13 resonances shifted downfield due to the deshielding effect of the neighboring electronegative oxygen on the ring. The correlation between  $\delta_{\text{H}}$  1.51 and  $\delta_{\text{C}}$  15.18 was assigned to the methylene CH<sub>2</sub>-1. The methine group for position CH-5 showed correlations between  $\delta_{\text{H}}$  1.65 (td,  $J=12.07, 2.90$ ) and  $\delta_{\text{C}}$  50.96 while CH-9 gave cross peaks between  $\delta_{\text{H}}$  1.28 and  $\delta_{\text{C}}$  56.13.

The <sup>1</sup>H-<sup>13</sup>C HMBC experiment was conducted to establish the connectivity of the substructures within the compound by determining the correlations between neighboring protons and carbons atoms (Figure 4.25 and 4.26). The methyl proton for C-17 at  $\delta_{\text{H}}$  1.03 exhibited strong correlations with adjacent carbons C-16, C-4, C-3 and C-5 confirming the HSQC assignments for the respective carbons. The position of C-18 was further validated by its <sup>1</sup>H cross peaks at  $\delta_{\text{H}}$  0.78 with carbon signals at  $\delta_{\text{C}}$  38.35, 50.96, 36.26, and 56.13 assigned for C-1, C-5, C-10, and C-9, respectively. Furthermore, the position and chemical shift of the quaternary carbon C-14 was confirmed by the correlations of the protons of the exomethylene moiety (H15a, H15b) and olefinic methine unit (H14a) at  $\delta_{\text{H}}$  4.85, 5.09 and 5.83 with  $\delta_{\text{C}}$  73.17. Meanwhile, the methyl moiety at  $\delta_{\text{C}}$  28.78 for C-20 was further confirmed by its correlations with the methylene and methine protons at  $\delta_{\text{H}}$  1.56, 1.73 and 5.83 assigned for C-12 and C-14, respectively. The correlations of  $\delta_{\text{H}}$  1.28 to  $\delta_{\text{C}}$  36.26 (<sup>2</sup> $J$ ),  $\delta_{\text{H}}$  1.28 to  $\delta_{\text{C}}$  74.78 (<sup>2</sup> $J$ ) and  $\delta_{\text{H}}$  1.28 to  $\delta_{\text{C}}$  28.78 (<sup>3</sup> $J$ ), marked the positions for C-10, C-8, and C-19, respectively.

A final experiment was performed to elucidate the relative stereochemistry. Based on the <sup>1</sup>H-<sup>1</sup>H NOESY experiment (Figure 4.27) (full spectrum on **Appendix III**) the methyl unit CH<sub>3</sub>-17 on C-4 gave a nOe on the methyl unit CH<sub>3</sub>-18 on C-10, while CH<sub>3</sub>-18 on C-10 also showed a strong through-space correlation with the CH<sub>3</sub>-19 on C-8. The methyl unit CH<sub>3</sub>-19 on C-8 showed a nOe as well with CH<sub>3</sub>-20 on C-13. The stereo positions of the methyl group on CH<sub>3</sub>-17 on C-4, CH<sub>3</sub>-19 on C-8, CH<sub>3</sub>-18 on C-10, and CH<sub>3</sub>-20 on C-13 all before followed the axial  $\beta$ -positions. The exomethylene proton H15b and methane H14 from the vinyl moiety showed a nOe cross peak with

$CH_3$ -20 on C-13, giving the orientation of C-13 vinyl moiety on the equatorial  $\alpha$ -position, while the H15b is *trans* to H14, H15b is also *cis* to  $CH_3$ -20.

According to the literature, the stereo positions of the methyl groups  $CH_3$ -19 on C-8,  $CH_3$ -18 on C-10, and  $CH_3$ -20 on C-13 followed the axial  $\beta$ -positions while the C-13 vinyl moiety is on the equatorial  $\alpha$ -positions, which were similar to those found in the isolated compound. The occurrence of a carboxylic acid unit (C-16) on C-4, similar to other diterpene derivatives, shielded proton shift of the axial methyl group upfield on C-10 (Bardyshev et al., 1982, Zinkel et al., 1985). This fact can be corroborated with the comparison of the chemical shift of the methyl group singlet on C-18 at  $\delta_H$  0.78 for the isolated compound with that of manoyl oxide at  $\delta_H$  0.85 (Vlad et al., 1966). The presented  $^1H$  and  $^{13}C$  NMR spectral data of the isolated compound was comparable but not identical with that found in the literature (Andersson et al., 1990, Kohda et al., 1983, Bardyshev et al., 1982, Buckwalter et al., 1975, Martín et al., 1980, Zinkel and Clarke, 1985). However, there were some noticeable differences between the chemical shifts presented on the literature and the isolated compound, due to the change in the stereochemistry at position C-4. The carboxyl moiety went on the equatorial  $\alpha$ -position as  $CH_3$ -17 on C-4 took the axial  $\beta$ -position, which instantaneously shielded  $CH_3$ -17 that went upfield to  $\delta_H$  1.03 (−0.25) and  $\delta_C$  16.54 (−8.5) while the carbon at C-4 suffered a deshielding effect by +3.3 ppm when compared to the published manoyl oxide acid as shown in Table 4.10 (Andersson et al., 1990). This change also affected the chemical shifts of methylene protons at position 3, which were deshielded by +0.54 and +0.49 ppm, the methine proton at position 5 was also deshielded by +0.47 ppm but C-5 was shielded by −4.06 ppm and the methylene protons on C-6 by −0.73 and −0.50 ppm, respectively. The structure of F18 was confirmed to be 4-*epi*-manoyl oxide acid, a labdane diterpene acid derivative of manoyl oxide (Table 4.10). This is the first time that this type of compound was isolated from a microorganism. Nevertheless, the isolated compound is considered a new congener as no similar stereoisomer was found in the literature.

However, the yield of the isolated sample (<1mg) was too low to measure the optical rotation of the 4-*epi*-manoyl oxide acid. Published optical rotation ( $[\alpha]_D^{20}$ ) for the

compound ranges between  $-41.5^\circ$  (Zinkel et al., 1985) to  $-39.2^\circ$  (Bardyshev et al., 1982), using solvent as ethanol.

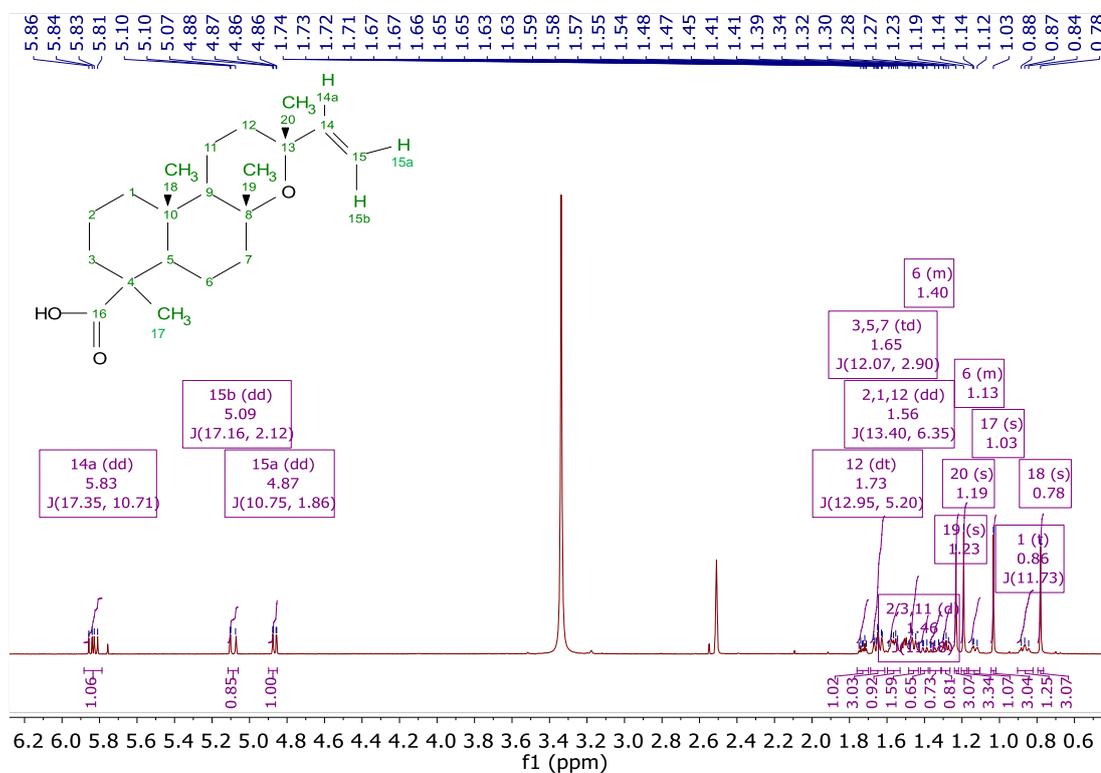


Figure 4-18  $^1\text{H}$  NMR spectrum of manoyl oxide acid at 600.13 MHz in  $\text{DMSO-d}_6$ .

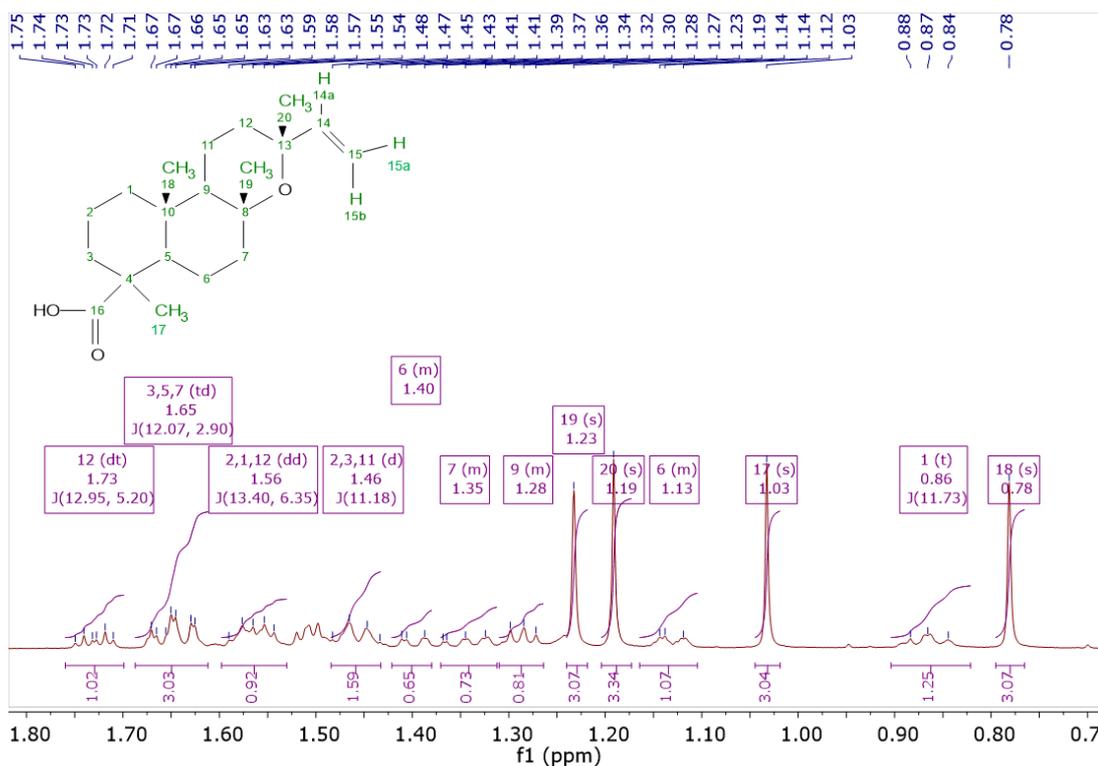


Figure 4-19 Expansion of <sup>1</sup>H NMR spectrum between 0.5 to 1.80 ppm of manoyl oxide acid at 600.13 MHz in DMSO-d<sub>6</sub>.

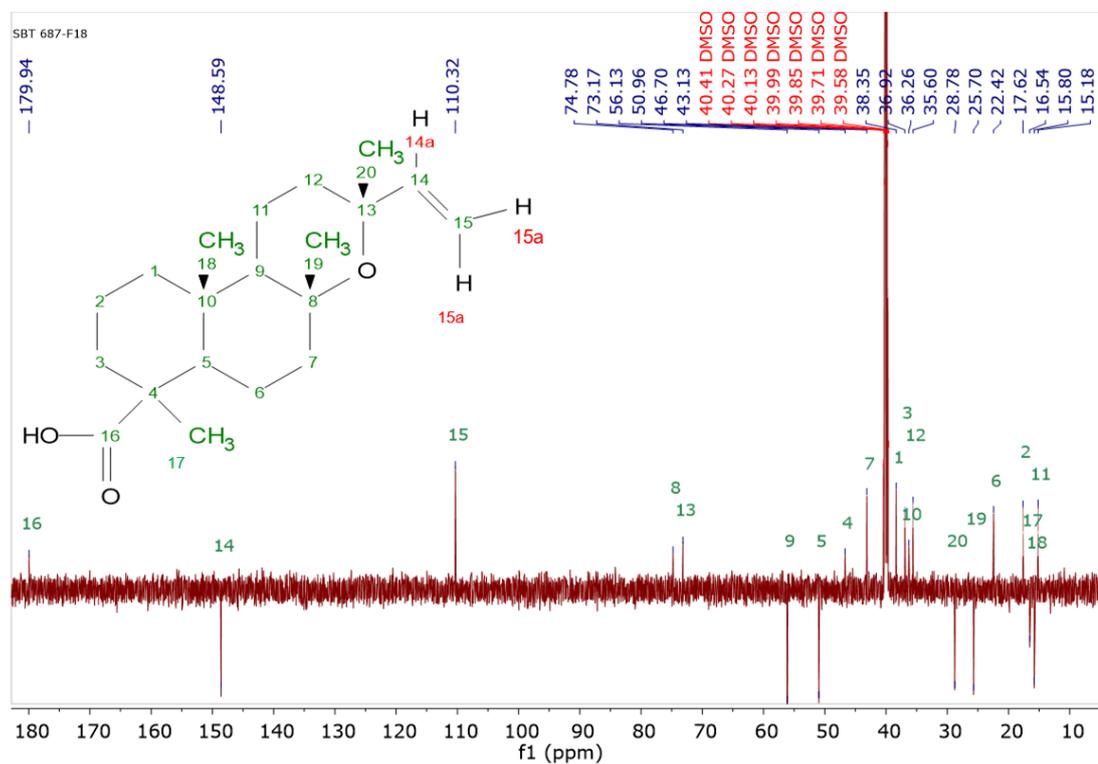
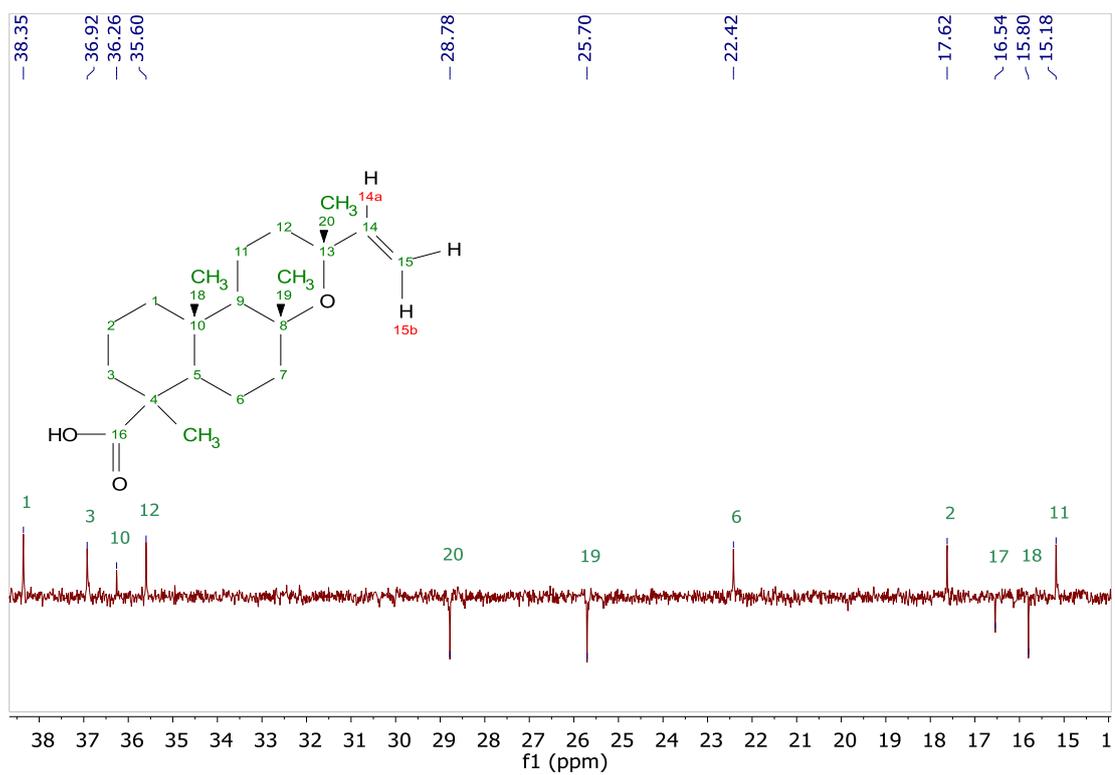
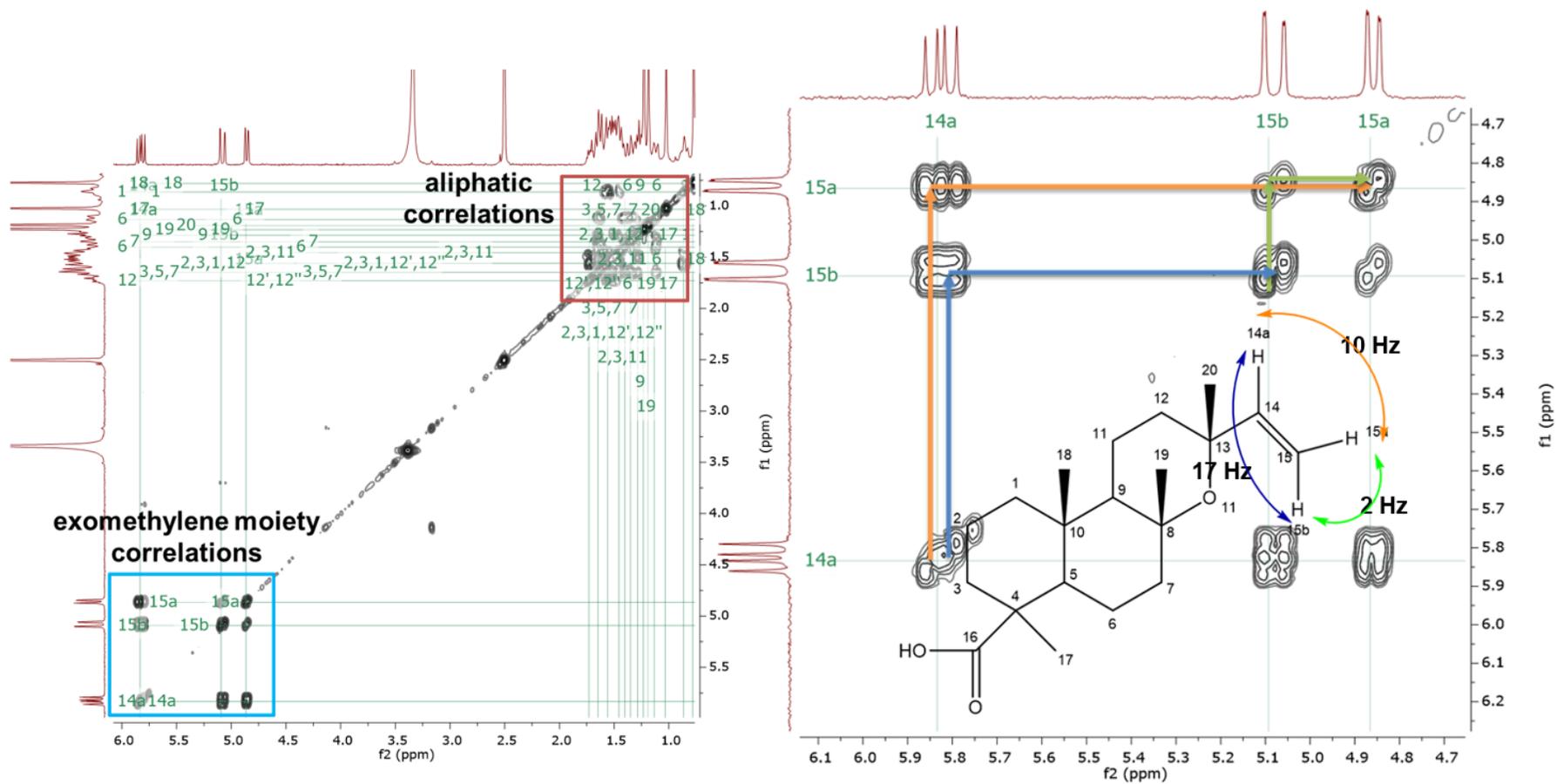


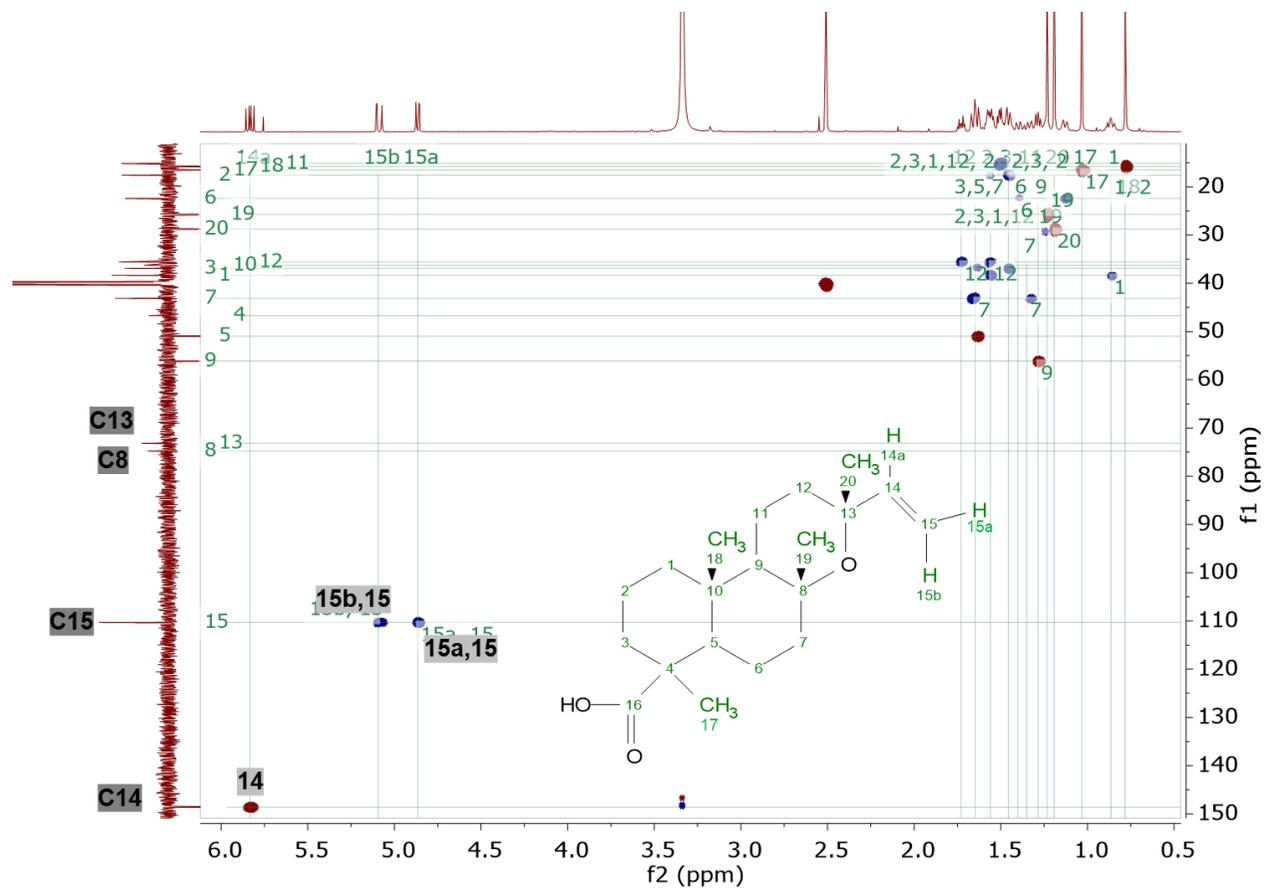
Figure 4-20 JMOD - C spectrum at 150.92 MHz of manoyl oxide acid in DMSO-d<sub>6</sub>.



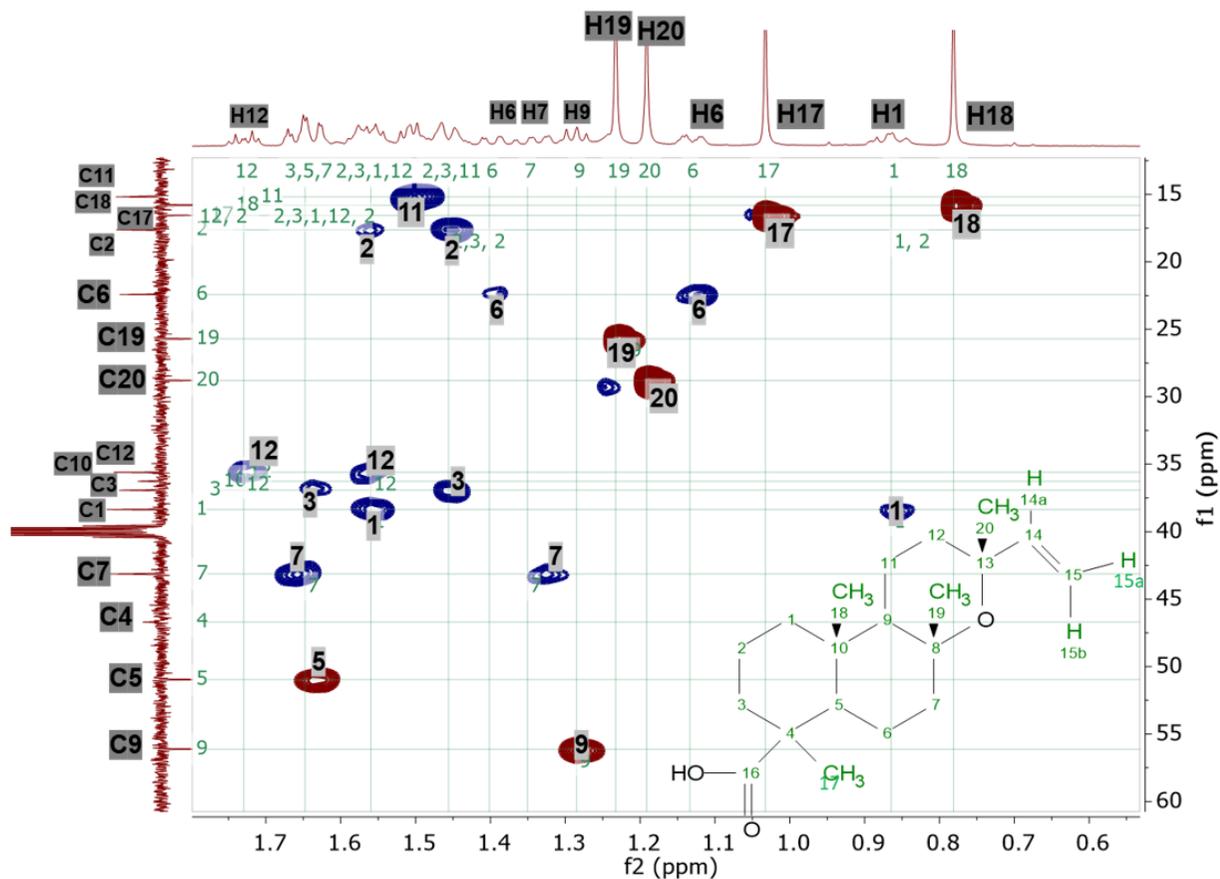
**Figure 4-21** Expansion of DEPT – C spectrum between 15 to 38 ppm at 150.92 MHz of manoyl oxide acid in DMSO-d<sub>6</sub>.



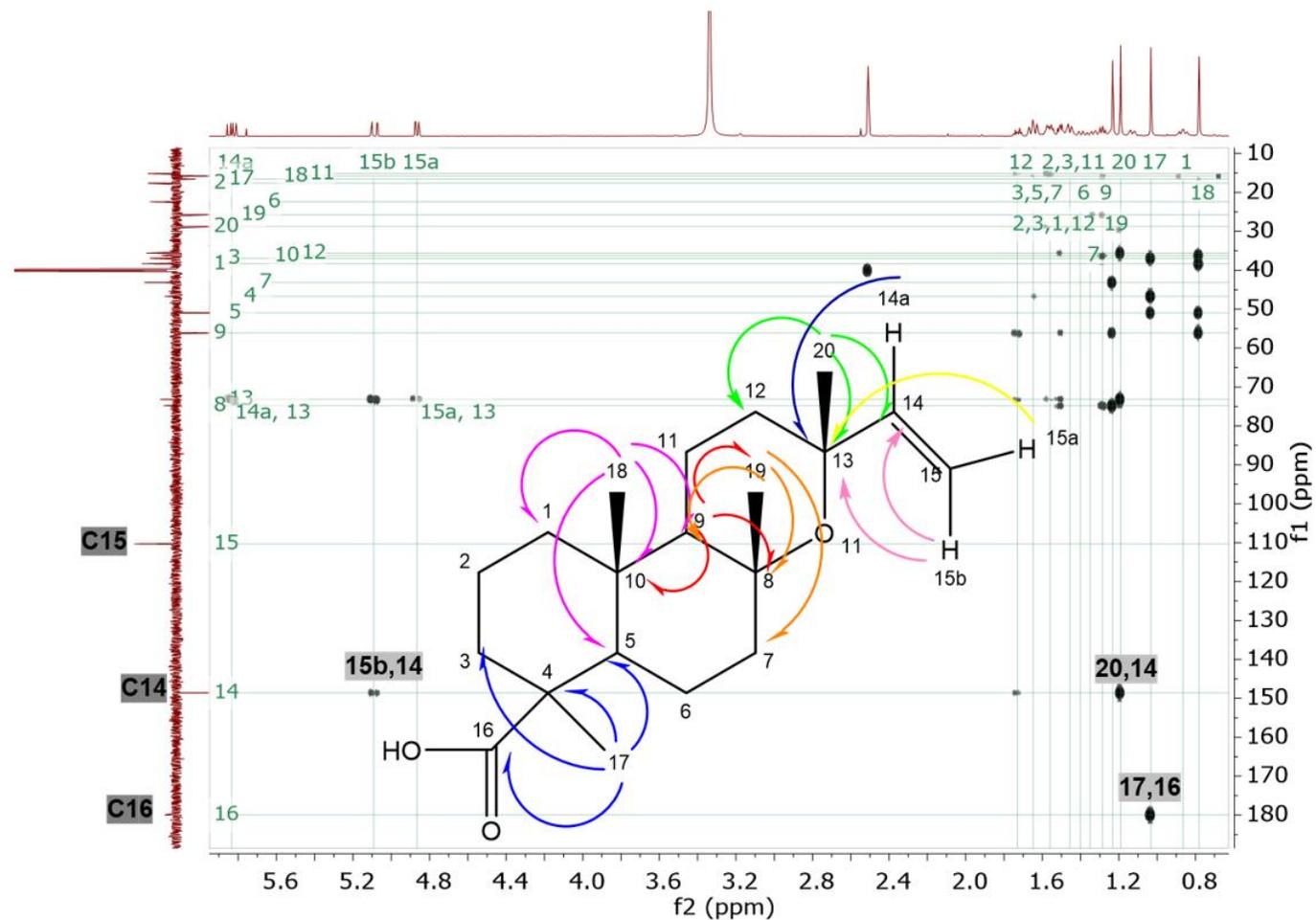
**Figure 4-22**  $^1\text{H}$ - $^1\text{H}$  COSY (400 MHz) spectrum in  $\text{DMSO-d}_6$ . The full spectrum showed the exomethylene olefinic moiety (blue box) and aliphatic (red box) correlations and the expansion showed partial correlations in the olefinic moiety substructure of manoyl oxide acid.



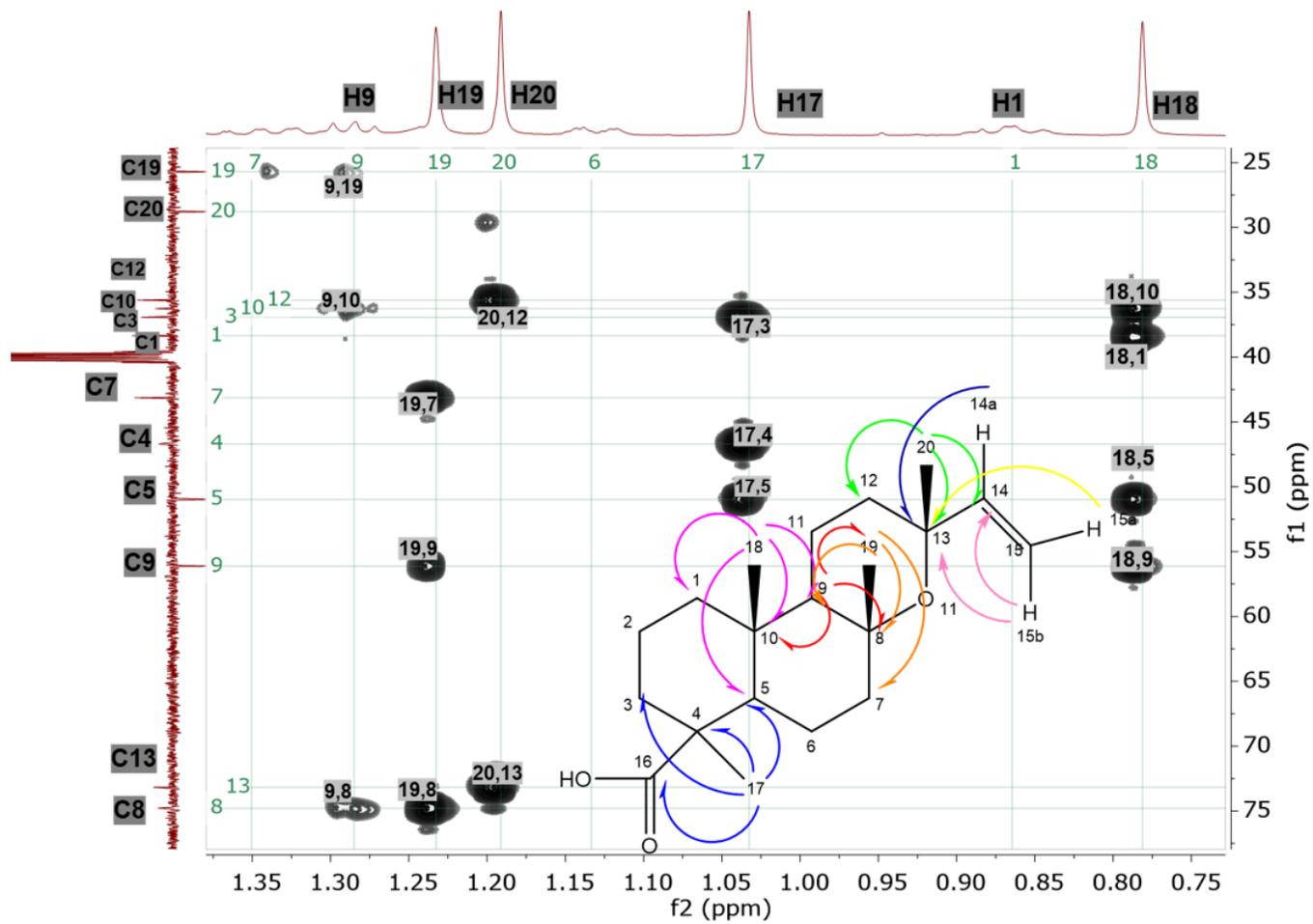
**Figure 4-23 HSQC of manoyl oxide acid at 600 MHz.** The X and Y – axes correspond to the proton (600.13MHz) and carbon (150.92 MHz) spectra, respectively. Red contour represents the CH<sub>3</sub> and CH while blue contour is for CH<sub>2</sub> and C.



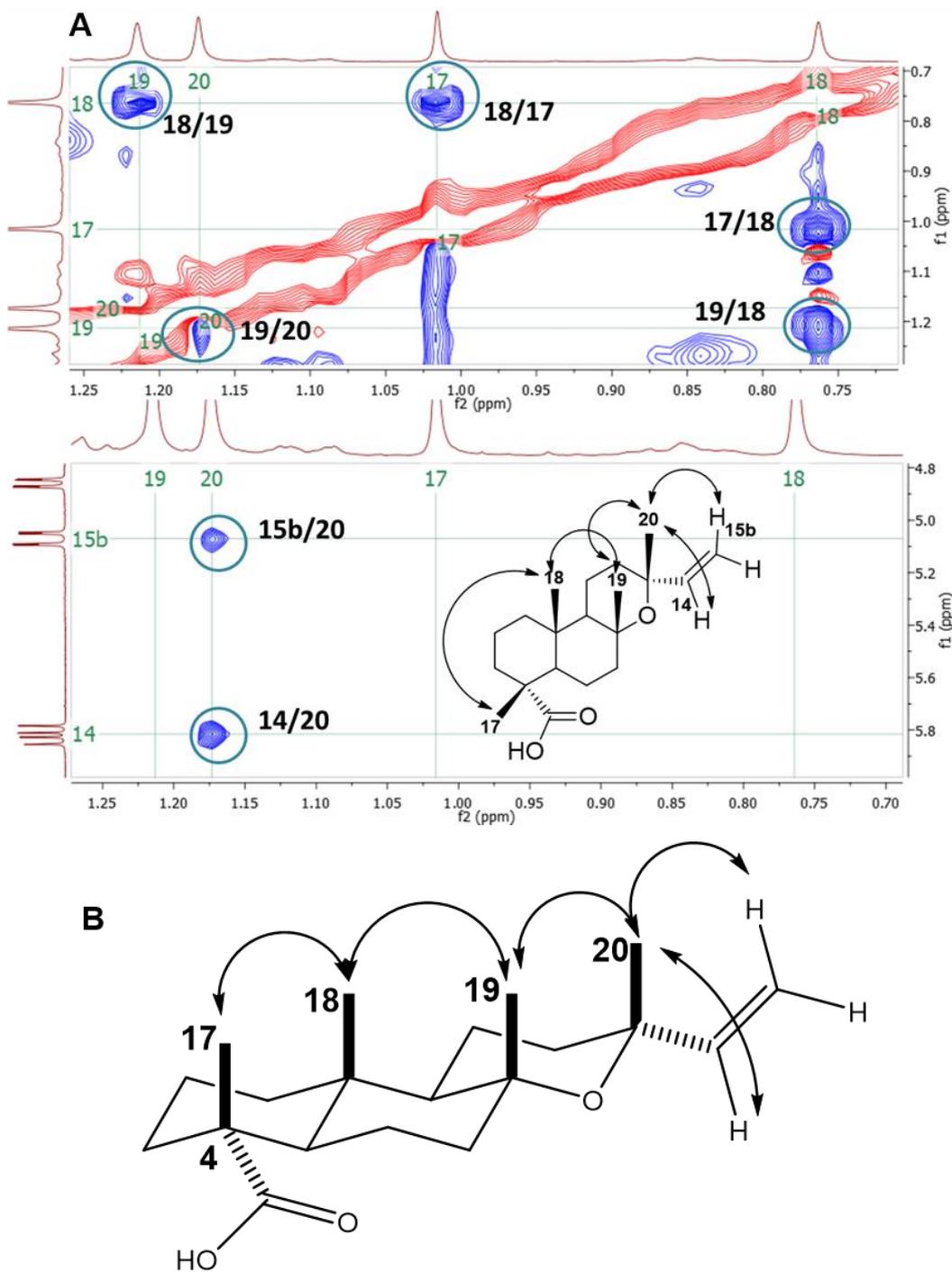
**Figure 4-24** Expansion of the HSQC of manoyl oxide acid at 600 MHz. The X and Y – axes correspond to the proton (600.13MHz) and carbon (150.92 MHz) spectra, respectively. Red contour represents the CH<sub>3</sub> and CH while blue contour is for CH<sub>2</sub> and C.



**Figure 4-25** HMBC of manoyl oxide acid at 600 MHz. The X and Y – axes correspond to the proton  $^1\text{H}$  (600.13 MHz) and carbon spectra  $^{13}\text{C}$  (150.92 MHz), respectively.



**Figure 4-26** Expansion of HMBC of manoyl oxide acid at 600 MHz. The X and Y – axes correspond to the proton  $^1\text{H}$  (600.13 MHz) and carbon spectra  $^{13}\text{C}$  (150.92 MHz), respectively.



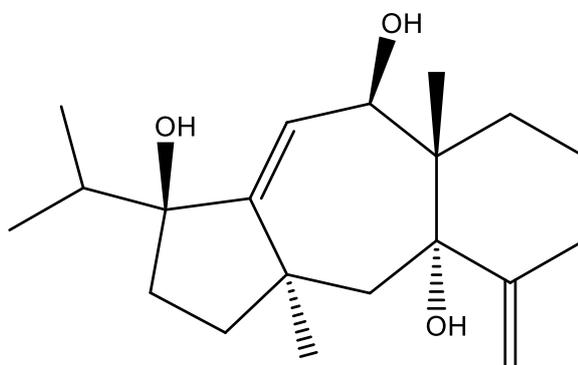
**Figure 4-27**  $^1\text{H}$ - $^1\text{H}$  NOESY (400 MHz) spectrum in  $\text{DMSO-d}_6$ . (A) indicated through space correlation of protons in manoyl oxide acid to establish the relative stereochemistry and below the structure of 4-epi-manoyl oxide acid showing the stereochemistry (B)

**Table 4-10**  $^1\text{H}$  NMR (600.13 MHz) and  $^{13}\text{C}$  NMR (150.92 MHz) of manoyl oxide acid in DMSO- $d_6$  and comparison with literature (Andersson et al., 1990). Full assignment was deduced from HSQC and HMBC (DMSO- $d_6$ , 600 MHz).

Atom No.	in DMSO- $d_6$		Literature (Andersson et al., 1990) in $\text{CDCl}_3$	
	$\delta_{\text{H}}$ (multiplicity, $J$ in Hz) (600.13 MHz)	$\delta_{\text{C}}$ (multiplicity) (150.92 MHz)	$\delta_{\text{C}}$ (multiplicity) (400 MHz)	$\delta_{\text{C}}$ (100 MHz)
1	0.86 (t, $J=11.7$ ) 1.5 (dd, $J=13.4, 6.3$ )	38.35(CH <sub>2</sub> )	0.90 ( $J=13.3, 4.4$ ) 1.67 ( $J=13.3, 3.0$ )	39.13(CH <sub>2</sub> )
2	1.46 (d, $J=11.2$ ) 1.56 (dd, $J=13.4, 6.3$ )	17.62(CH <sub>2</sub> )	1.47 ( $J=13.3, 3.4$ ) 1.87 ( $J=13.5, 4.1$ )	18.99(CH <sub>2</sub> )
3	1.46 (d, $J=11.2$ ) 1.65 (td, $J=12.1, 2.9$ )	36.92(CH <sub>2</sub> )	1.02 ( $J=13.5$ ) 2.14 ( $J=13.5, 4.1$ )	37.84(CH <sub>2</sub> )
4	—	46.70(C)	—	43.67(C)
5	1.65 (td, $J=12.1, 2.9$ )	50.96(CH)	1.18 ( $J=13.0, 3.0$ )	55.02(CH)
6	1.13 (m) 1.40 (m)	22.42(CH <sub>2</sub> )	1.86( $J=14.0, 5.0$ ) 1.90( $J=13.0, 0.5$ )	21.36(CH <sub>2</sub> )
7	1.65 (td, $J=12.1, 2.9$ ) 1.35 (m)	43.13(CH <sub>2</sub> )	1.85( $J=13.0, 3.0$ ) 1.40( $J=13.0, 0.5$ )	43.10(CH <sub>2</sub> )
8	—	74.78(C)	—	74.78(C)
9	1.28 (m)	56.13(CH)	1.32( $J=11.7, 4.1$ )	56.71(CH)
10	—	36.26(C)	—	37.62(C)
11	1.51 (d, $J=7.8, 5.4$ )	15.18(CH <sub>2</sub> )	1.48( $J=6.0, 5.5$ ) 1.60( $J=9.0, 5.5$ )	15.42(CH <sub>2</sub> )
12	1.56 (dd, $J=13.4, 6.3$ ) 1.73 (dt, $J=12.9, 5.2$ )	35.60(CH <sub>2</sub> )	1.64( $J=13.3, 6.0$ ) 1.78( $J=13.3, 5.5$ )	35.59(CH <sub>2</sub> )
13	—	73.17(C)	—	73.26(C)
14	5.83 (dd, $J=17.3, 10.7$ )	148.59(CH)	5.87( $J=17.3, 10.7$ )	147.83(CH)
15	4.85 (dd, $J=10.7, 1.9$ ) 5.09 (dd, $J=17.2, 2.1$ )	110.32(CH <sub>2</sub> )	4.93( $J=10.7, 1.6$ ) 5.15( $J=17.3, 1.6$ )	110.35(CH <sub>2</sub> )
16	—	179.94(C)	—	182.92(C)
17	1.03 (s)	16.54(CH <sub>3</sub> )	1.28	25.06(CH <sub>3</sub> )

<b>18</b>	0.78 (s)	15.80(CH <sub>3</sub> )	0.72	13.11(CH <sub>3</sub> )
<b>19</b>	1.23 (s)	25.70(CH <sub>3</sub> )	1.32	28.62(CH <sub>3</sub> )
<b>20</b>	1.19 (s)	28.78(CH <sub>3</sub> )	1.24	28.77(CH <sub>3</sub> )

According to the dereplication table **4.7** shown on **4.3.1.1**, F18 was putatively identified as dolatriol (Figure **4.28**) with a molecular formula of C<sub>20</sub>H<sub>32</sub>O<sub>3</sub>. Nevertheless, the NMR spectra obtained from F18 was not compatible with the structure of dolatriol. Dolatriol was earlier isolated from Mollusca *Dolabella auricularia* (Pettit et al., 1980).



**Figure 4-28** Compound dolatriol C<sub>20</sub>H<sub>32</sub>O<sub>3</sub>. Relative stereochemistry according to the literature (Saha et al., 1990).

The <sup>1</sup>H and <sup>13</sup>C NMR spectrum prediction (Figure **4.29**) of dolatriol provided invaluable information which demonstrated that this compound did not correspond to the one isolated. In the simulated spectra for dolatriol in DMSO-*d*<sub>6</sub>, three hydroxyl groups in the <sup>1</sup>H NMR exhibited two overlapped singlets at 5.23 ppm and a doublet at 4.39 ppm while the exomethylene was predicted as a pair of doublets at 4.9 and 5.0 ppm and an olefinic methine was revealed at 5.5 ppm. For F18, the exomethylene signals were observed as doublet of doublets and more deshielded downfield.

In the simulated <sup>13</sup>C NMR spectra of dolatriol, the three oxygen-bearing carbons were predicted at 76.26 ppm and 82.25 ppm for two overlapping carbons. For F18, two

quaternary oxygen-bearing carbons were observed at ca. 75 ppm while missing a third signal at ca. 80 ppm as described for dolatriol. The obtained spectra for F18 did not show the signals as predicted for dolatriol but instead exhibited a very downfield signal at 179.94 ppm characteristic of the carboxylic acid group as found in the elucidated structure of 8,13-epoxy-14-labd-en-19-oic acid.

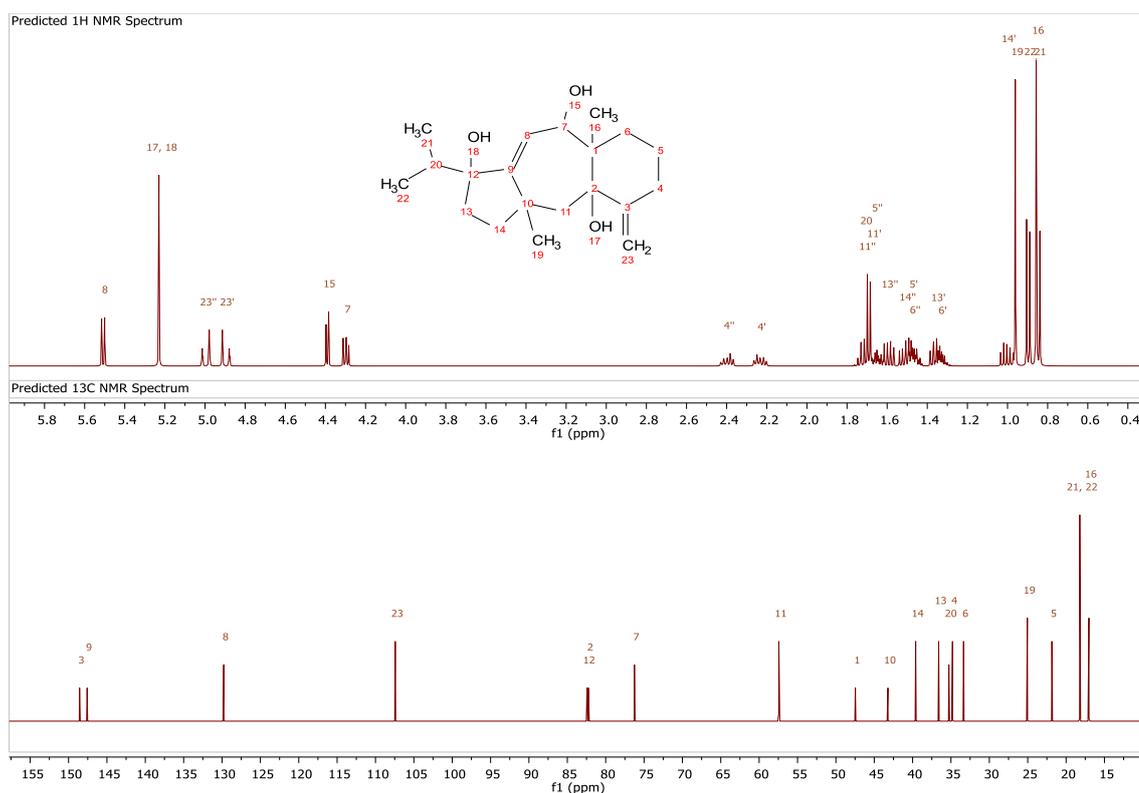
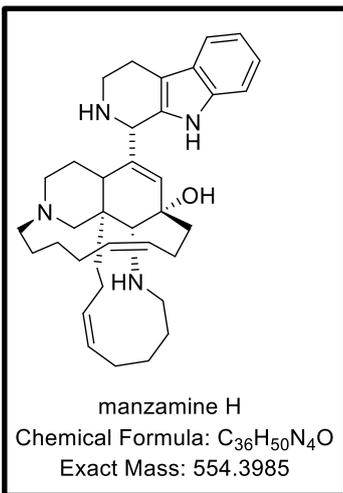
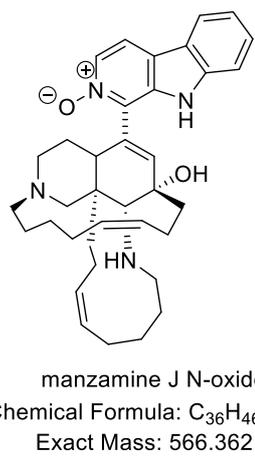
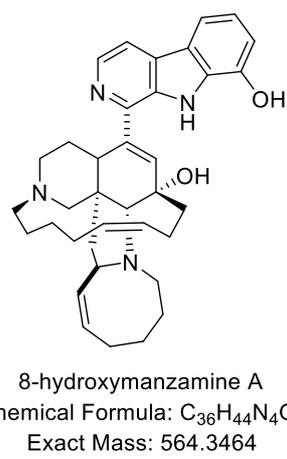
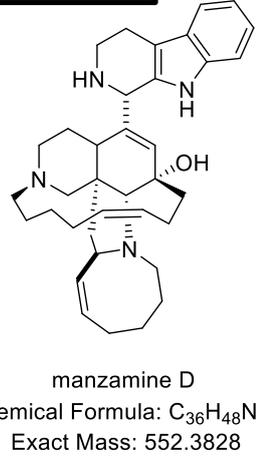
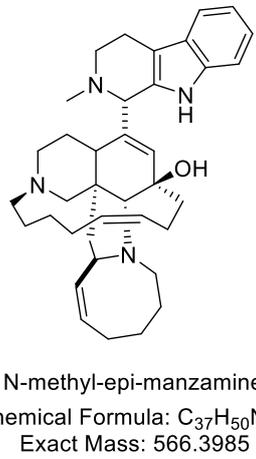


Figure 4-29 Predicted <sup>1</sup>H and <sup>13</sup>C NMR spectrum for dolatriol.

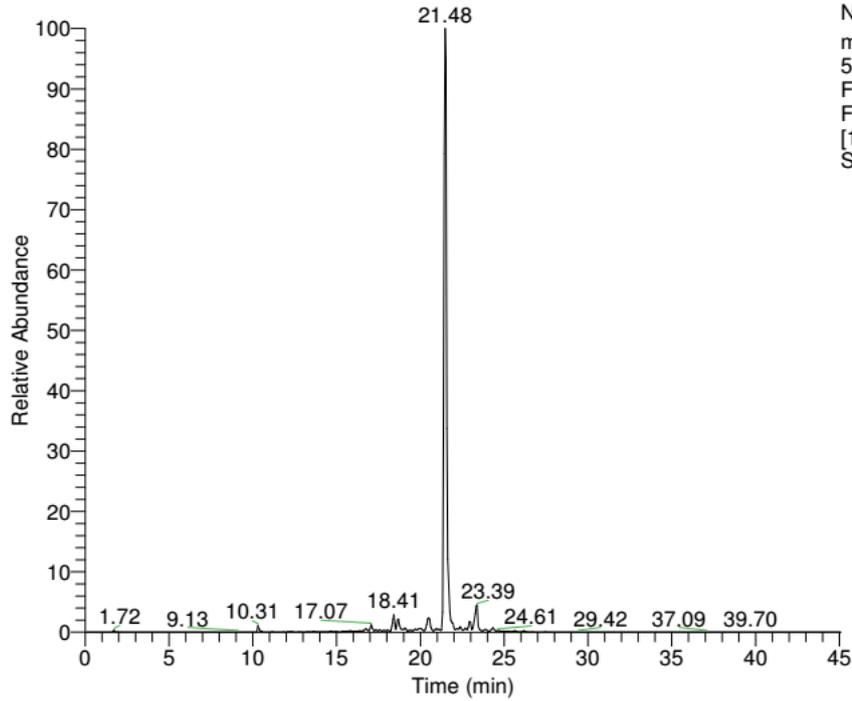
## 4.4.2 Manzamines

Table 4-11 Manzamines found in SBT687 Fraction 17.

Compound name (Known compound): manzamines		
Fraction: Fraction 17 (F17) Source: <i>Micromonospora</i> sp. N17 (SBT687) Amount of sample: 1.3 mg Percent yield: 0.3% Physical description: yellow amorphous powder Major derivative detected: manzamine H		
 <p>manzamine H            Chemical Formula: C<sub>36</sub>H<sub>50</sub>N<sub>4</sub>O            Exact Mass: 554.3985</p>	 <p>manzamine J N-oxide            Chemical Formula: C<sub>36</sub>H<sub>46</sub>N<sub>4</sub>O<sub>2</sub>            Exact Mass: 566.3621</p>	 <p>8-hydroxymanzamine A            Chemical Formula: C<sub>36</sub>H<sub>44</sub>N<sub>4</sub>O<sub>2</sub>            Exact Mass: 564.3464</p>
 <p>manzamine D            Chemical Formula: C<sub>36</sub>H<sub>48</sub>N<sub>4</sub>O            Exact Mass: 552.3828</p>	 <p>N-methyl-epi-manzamine D            Chemical Formula: C<sub>37</sub>H<sub>50</sub>N<sub>4</sub>O            Exact Mass: 566.3985</p>	

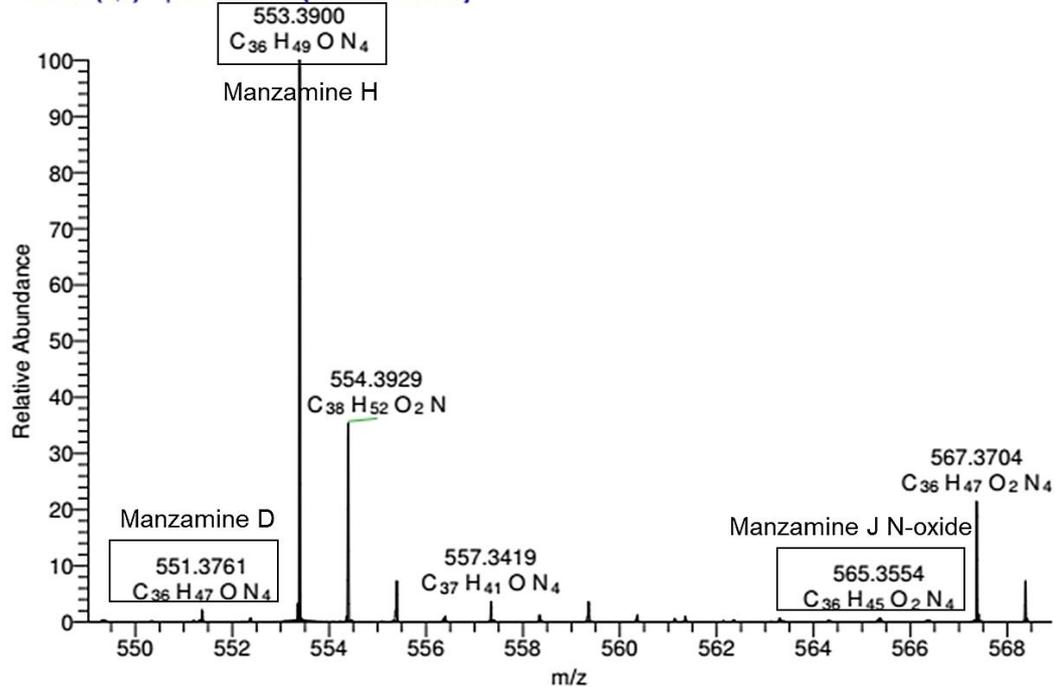
LC – HRFTMS spectra [M-H]-

RT: 0.00 - 45.07



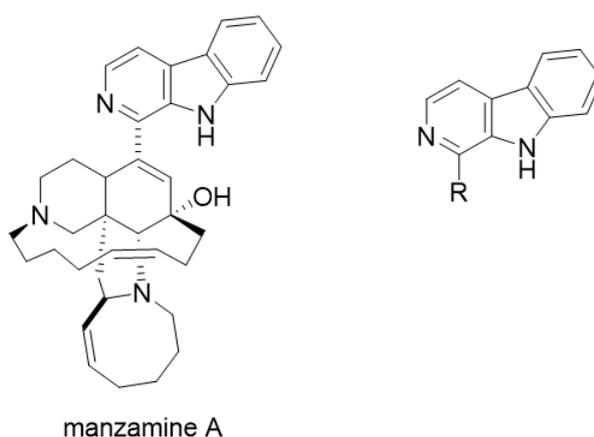
NL: 2.85E8  
m/z=  
551.3758-565.4920 F:  
FTMS {1,2} - p ESI  
Full ms  
[150.00-1500.00] MS  
SBT687\_F17

SBT687\_F17 #203-703 RT: 8.02-25.05 AV: 250 NL: 2.76E6  
T: FTMS {1,2} - p ESI Full ms [150.00-1500.00]



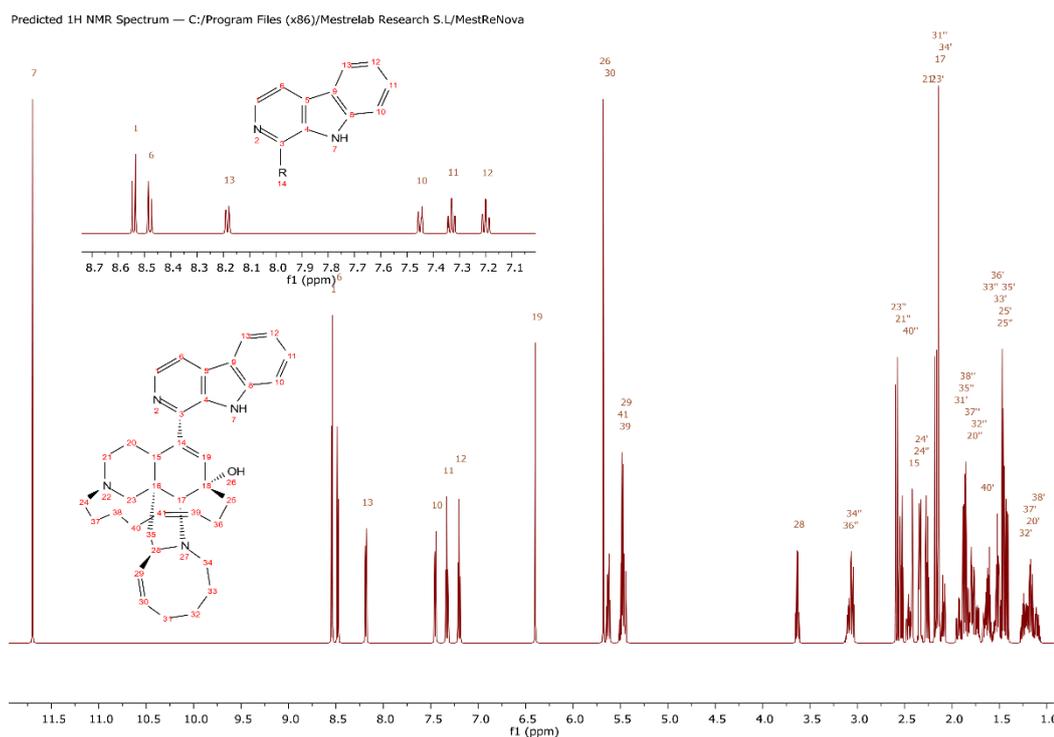
Several manzamines were found in the dereplication of SBT687 fractions particularly in F17 with molecular weights ranging from 552.384 to 568.379 g/mol and retention times between 20.11 to 24.33 minutes. Practically, F17 afforded all the manzamines detected in SBT687, which included manzamine D ( $C_{36}H_{48}N_4O$ ,  $m/z$  551.3761  $[M-H]^-$ , Rt:21.81 min), H ( $C_{36}H_{50}N_4O$ ,  $m/z$  553.3900  $[M-H]^-$ , Rt:21.52 min), J N-oxide ( $C_{36}H_{46}N_4O_2$ ,  $m/z$  565.3554  $[M-H]^-$ , Rt:20.11 min), 8-hydroxymanzamine A ( $C_{36}H_{45}N_4O_2^+$ ,  $m/z$  564.3473  $[M-H]^-$ , Rt:22.93 min), and N-methyl-epi-manzamine D ( $C_{37}H_{50}N_4O$ ,  $m/z$  565.3920  $[M-H]^-$ , Rt:22.37 min) as shown in table 4.11. The highest peak intensity was observed for manzamine H and this metabolite was found on the metabolomic profile analysis presented in section 4.3.1. Tetrahydro-8-hydroxymanzamine A ( $C_{36}H_{48}N_4O_2$ ,  $m/z$  567.3699, Rt: 18.68 min) achieved the highest intensity peak for F16 ( $4.46E+07$ ). Nonetheless, the higher peak intensity did not reflect the true concentration of the metabolite as it was poorly discernible on the  $^1H$  NMR data of F16.

Manzamines are structurally complex alkaloids with an intricate nitrogen-containing polycyclic system, characterized by a fused tetra or pentacyclic ring system attached to a  $\beta$ -carboline moiety (Furusato et al., 2014, Ashok et al., 2014, Rao et al., 2003). Manzamine A (Figure 4.30), the prototype of this group, was the first isolated from an Okinawa sponge of the genus *Haliclona* in 1986 by Higa and co-workers (Radwan et al., 2012).

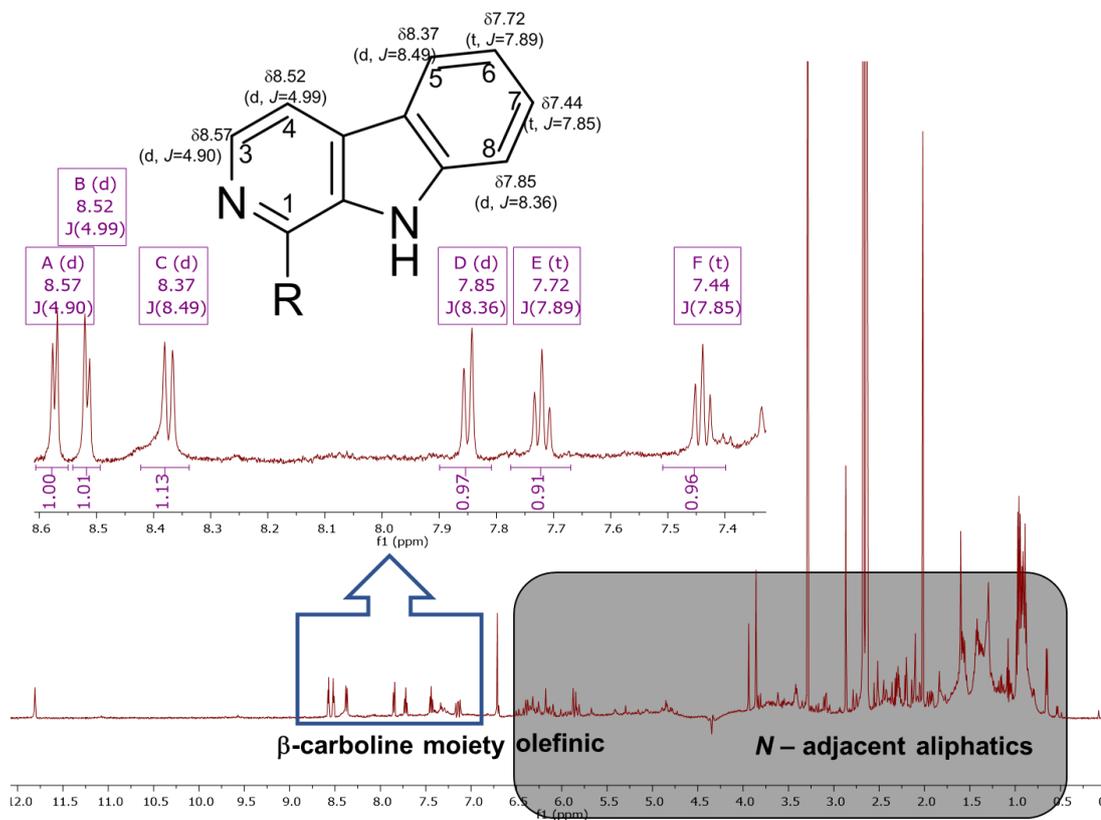


**Figure 4-30** Manzamine A and the characteristic  $\beta$ -carboline moiety of manzamines.

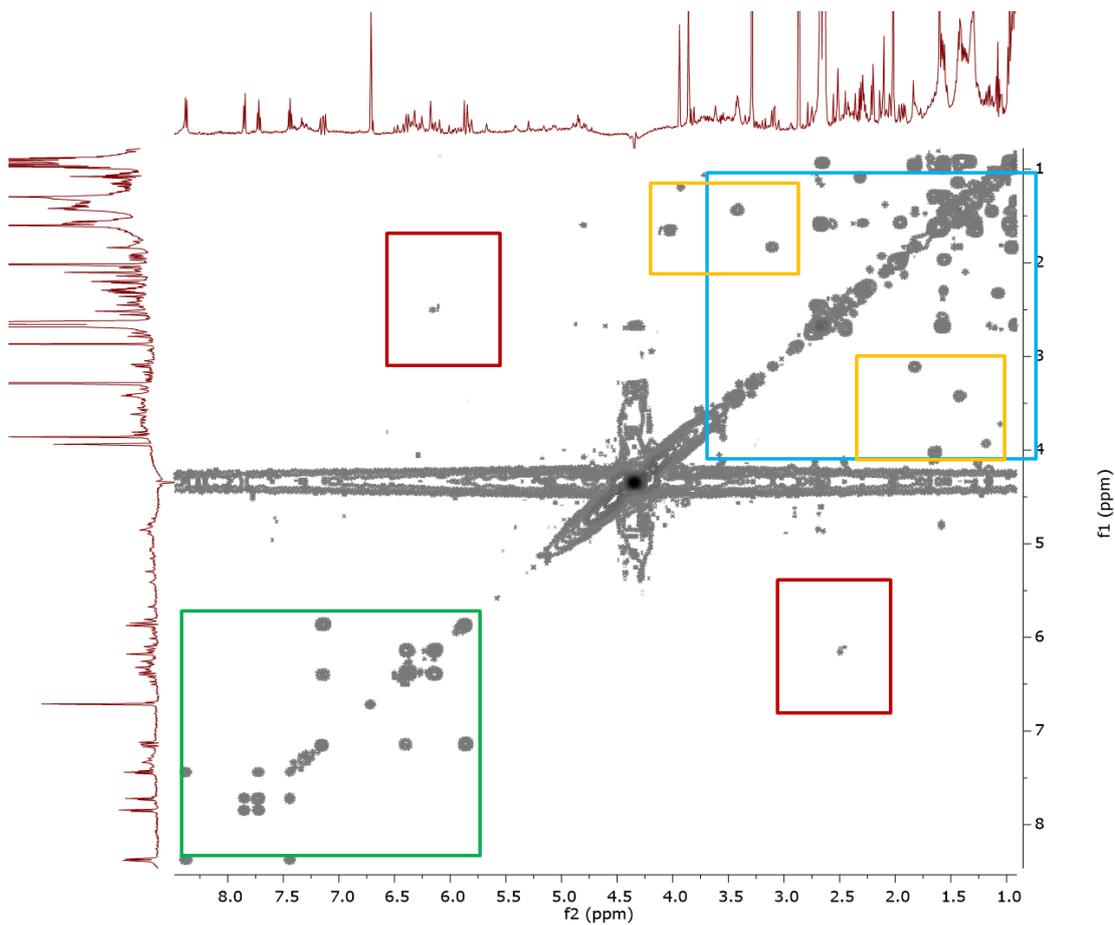
A simulated  $^1\text{H}$  NMR spectrum of manzamine A is depicted in Figure 4.31. The  $^1\text{H}$  NMR spectrum of fraction F17 is presented in Figures 4.32. It was possible to visualize a mixture of different manzamines with crowded overlapped peaks in the aliphatic region and the characteristic aromatic signals of the  $\beta$ -carboline moiety. Nonetheless, the  $^1\text{H}$ - $^1\text{H}$  COSY (Figure 4.33) confirmed the presence of the characteristic signals and correlations of a manzamine in the aromatic and aliphatic region. Furthermore, the correlations within the dominant ABCD and AB system of the  $\beta$ -carboline unit that belongs to manzamine were seen on the expansion of the  $^1\text{H}$ - $^1\text{H}$  COSY (Figure 4.34). Olefinics were observed between 5.5 and 6.5 ppm correlating to the aliphatic region at 2.0 ppm (boxed in red) while the *N*-adjacent protons were found between 3.0 and 4.0 ppm correlating further upfield to the 1.0 to 1.5 ppm region (boxed in yellow).



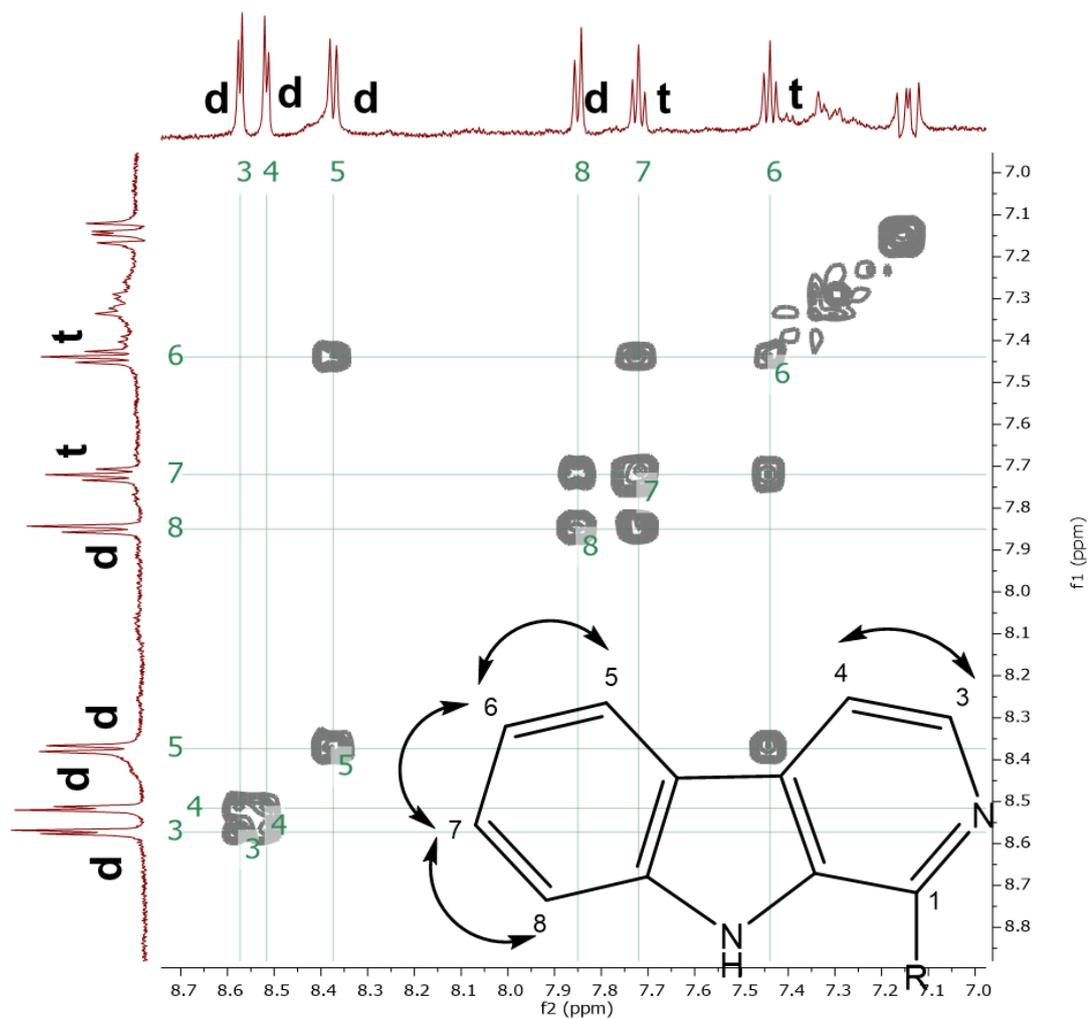
**Figure 4-31 Simulated  $^1\text{H}$  NMR spectrum for manzamine A.**



**Figure 4-32**  $^1\text{H}$  NMR spectrum of SBT687 F17 manzamine at 600.13 MHz in  $\text{DMSO-d}_6$ . Highlighted in blue is the aliphatic region showing the crowded peaks and on the upper left there is an expansion of the signals between 7.4 to 8.6 ppm displaying the signals of the beta carboline moiety.



**Figure 4-33**  $^1\text{H}$ - $^1\text{H}$  COSY (600.13 MHz) spectrum in  $\text{DMSO-d}_6$  showed the spin systems characteristic of manzamines. Green box corresponds to the aromatic region correlations and blue box to the aliphatic region correlations. The red boxes indicate the olefinic systems while the yellow boxes demonstrate the correlation of *N*-adjacent aliphatic protons.



**Figure 4-34 Expansion of  $^1\text{H}$ - $^1\text{H}$  COSY (600.13 MHz) spectrum in  $\text{DMSO-d}_6$  represented the one dominant ABCD system of the beta carboline unit, that belongs to manzamine and correlations within this system with the characteristic doublets (d) and triplets(t) labelled.**

## 4.5 Discussion

In the present work, different scale-up fermentations were performed in *Micromonospora sp.* N17 (SBT687). A metabolomics workflow for drug discovery was conducted prior to extraction and fractionation. NMR and HR-LC/MS were used as analytical platforms to achieve the metabolomics analysis. The crude extract exhibited insecticidal activity with 100% inhibition against parasite *Lepeophtheirus sp.* (sea lice) at 75 µg/mL in preliminary bioassays. The bioassay screening of the fractions did not display insecticidal activity for all fractions. From the fractionation of the 2L culture, fractions F6 and F8 afforded 65% and 72% of inhibition against parasite *Lepeophtheirus sp.* (sea lice), respectively, at 75 µg/mL. From the fractionation of the 15L scale-up, it was not possible to test all fractions due to inadequate yield and time was also insufficient to validate the bioactivity results due to the closing period of the FP7 project with PHARMAQ AS. The inhibitory activity was then deduced by comparison of the thin layer chromatograms of both fractionations from the respective scale-ups. Moreover, using multivariate analysis, the most statistically probable bioactive metabolites were targeted for isolation and elucidation work. The compounds were structurally identified as alkaloids (manzamines) and terpenes (4-epi-manoyl oxide acid) from fractions 17 and 18, respectively.

### 4.5.1 Diterpenes as promising insecticidal agents

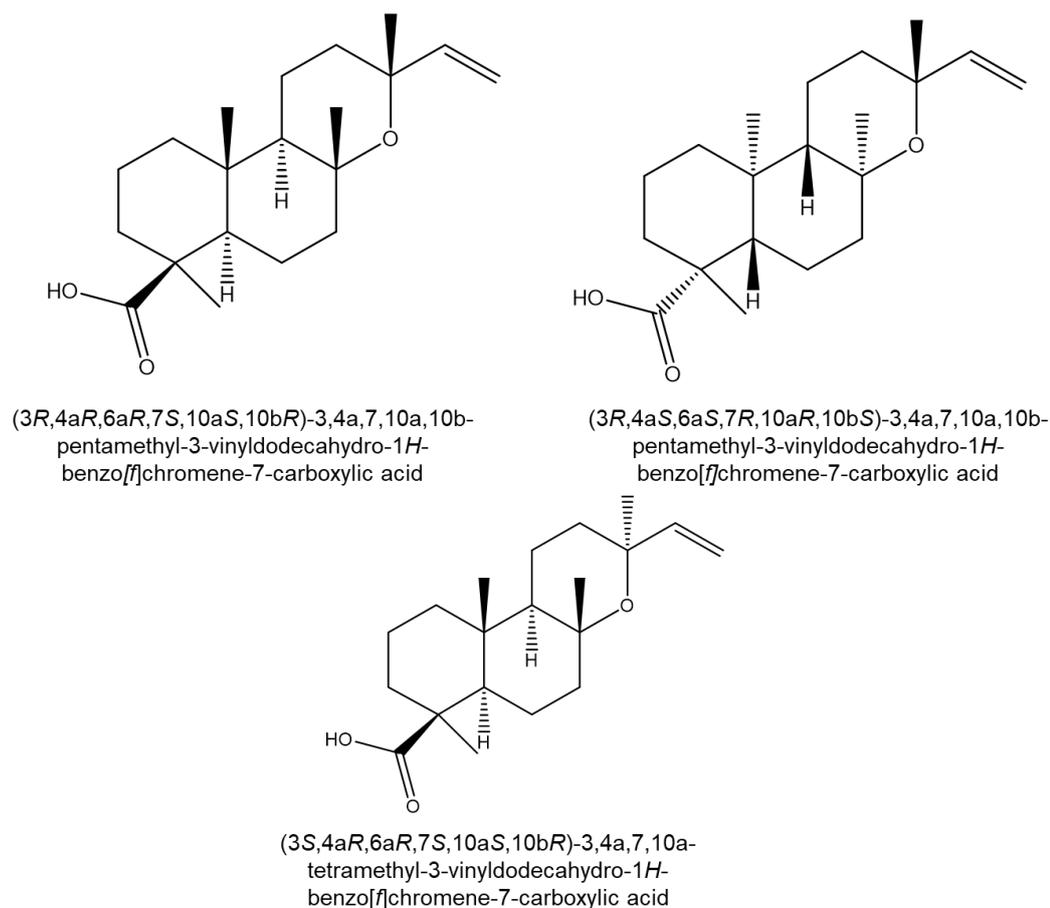
Terpenes are the most abundant structurally diverse class of natural products, followed by alkaloids and phenolics (Singh and Sharma, 2015). The basic unit of terpene consists of an isoprene, 2-methylbuta-1,3-diene (C<sub>5</sub>H<sub>8</sub>). Thus, based on C<sub>5</sub> units, terpenoids are classified in hemiterpenes (C<sub>5</sub>), monoterpenes (C<sub>10</sub>), sesquiterpenes (C<sub>15</sub>), diterpenes (C<sub>20</sub>), sesterpenes (C<sub>25</sub>), triterpenes (C<sub>30</sub>), tetraterpenes (C<sub>40</sub>) and polyterpenes (>C<sub>40</sub>) (Zwenger and Basu, 2008). The terpene isolated from *Micromonospora sp.* N17 Fraction 17 was structurally elucidated as manoyl oxide acid (Andersson et al., 1990, Kohda et al., 1983, Bardyshev et al., 1982, Buckwalter et al., 1975, Zinkel et al., 1985), a labdane diterpenoid acid derivative of manoyl oxide.

Manoyl oxide acid had been synthesized by the oxidation of the labdane diterpene jhanol with Jones' reagent to afford the carboxylic acid congener (Kohda et al., 1983). Previous studies reported that this compound exhibited substantial inhibitory effect on cytokine production from lipopolysaccharide-stimulated human peripheral mononuclear cells in comparison with the reference compound, prednisolone (Duan et al., 1999).

The large superfamily of labdane related diterpenes includes more than 7000 identified natural products characterized by a basic decalin core (Peters, 2010). Their biosynthetic pathway requires the initiating and essential step of a protonation-initiated reaction to cyclization of geranylgeranyldiphosphate (GGPP) to labdane-type diphosphate by a class II diterpene synthase (diTPS), which is then taken up by a class I diTPS to create a hydrocarbon skeletal structure analogous to the labdanes. Through the action of enzymes, such as cytochrome P450s (CYPs), these basic structures were transformed to a varied range of extremely complex molecules (Ignea et al., 2016).

Labdane type diterpenes were isolated/identified from different natural products, particularly higher plants. These compounds were isolated, among others, from pine needles of several species of *Pinus* that includes *P. sylvestris* (Bardyshev et al., 1982), *P. resinosa* (Zinkel and Clarke, 1985), and *P. nigra* (Zinkel et al., 1985); from leaves of *Stevia rebaudiana* (Kohda et al., 1983); roots of *Viguiera hypargyrea*, a Mexican medicinal plant (Marquina et al., 2009); the mangrove plant *Rhizophora mangle* L., (Martins et al., 2017); and from aerial parts of several plants like *Baccharis tola* (Martín et al., 1980), *Eupatorium jhanii* (González et al., 1977), *Leyssera gnaphaloides* L. (Tsichritzis and Jakupovic, 1991), and *Salvia reuterana* (Moridi Farimani and Miran, 2014). However, labdane type diterpenes, such as manoyl oxides, were also isolated from liverwort, a division of non-vascular land plants, which means this class are not only distributed in higher plants (Guo et al., 2010). Even though the labdane type diterpene, manoyl oxide acid, is commonly found in plants, the 4-epimer congener was isolated for the first time from a microbial source, which in this case is from fraction 18 of SBT687. The isolated stereoisomer has never been described

before and is reported for the first time in this study. Figure 4.35 presents the stereochemistry of the different congeners reported in the literature.

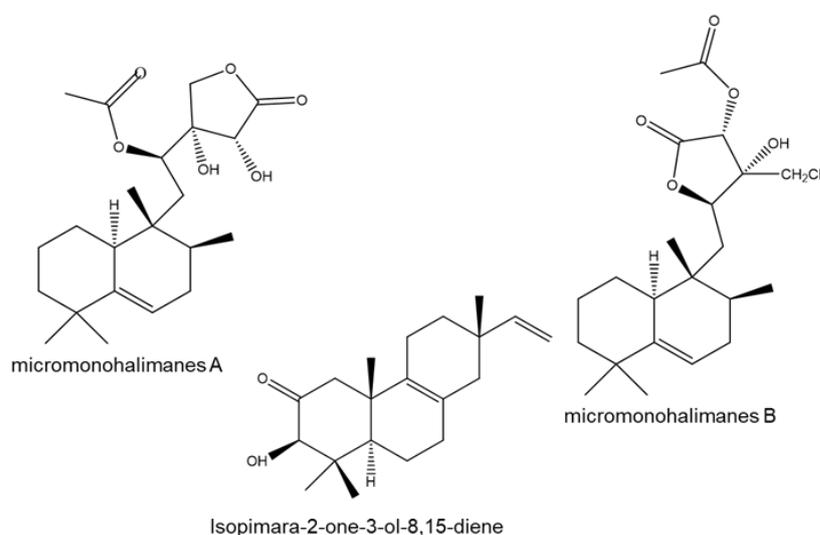


**Figure 4-35 Stereochemistry of the manoyl oxide acid congeners that has been reported in the literature.**

Therefore, the fact that this plant metabolite was found in a microorganism and was recently found in lower plants as well, could suggest that the metabolite is not being produced by the plant but by a microorganism. It is known that the microorganisms have a close interaction with the cell's plant, and taking up secreted metabolites, having a synergistic interaction (Raturi et al., 2013).

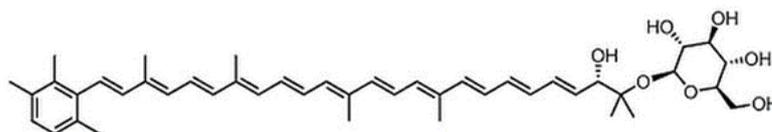
Diterpenoids are predominantly isolated from plant or fungi, but rarely from bacteria (Smanski et al., 2012). However, bacteria, particularly actinomycetes, hold biosynthetic potential as source for terpenoids like it was revealed by advances in microbial genomics (Xie et al., 2014). Nevertheless, it is probable that these genetic pathways remain silent under laboratory growth conditions or remains untargeted by using bioassay-guided fractionation methods designed to isolate bacterial natural

products, which will not detect secondary metabolites at  $\mu\text{g}$  levels (Mullowney et al., 2015). Several diterpenoids have been reported to be isolated from marine *Micromonospora sp.* that includes halimane-type of diterpenoids such as micromonohalimanes A and B, which have moderate antibacterial effects against methicillin-resistant *Staphylococcus aureus* (Zhang et al., 2016b). Another terpene isolated from a marine *Micromonospora* is a  $\Delta^{8,9}$ -pimarane diterpene, isopimara-2-one-3-ol-8,15-diene but was found without significant activity in cytotoxic assays against human ovarian OVCAR4 and Kuramochi cancer cells (Figure 4.36) (Mullowney et al., 2015).



**Figure 4-36** Examples of isolated diterpenes from marine *Micromonospora sp.*

The analysis of the biosynthetic gene clusters (BGCs) of *Micromonospora sp. L5* (Table 4.12) seems enriched for the biosynthetic class of saccharide/terpenes, including sioxanthins (glycosylated carotenoids) (Figure 4.37) and phosphonoglycans; associated respectively to phytoene/squalene synthase and terpene synthase/cyclase metal-binding domain protein families.



**Figure 4-37** Structure of the glycosylated carotenoid sioxanthin.

**Table 4-12 AntiSMASH analysis of the biosynthetic gene clusters (BGCs) of *Micromonospora* sp. L5.** The genomic location/position of each BGC shown as nucleotide (nt.) counts. The identified regions containing the clusters are displayed and numbered in the same order in which they appear on the input nucleotide sequence. Homologous gene clusters are matched based on blast hits against the curated MIBiG repository and ranked by similarity (%). Highly conserved enzyme HMM profiles represented by core-enzymes.

Region	Type	From	To	Most similar known cluster		Similarity	MIBiG BGC-ID
The following regions are from record NC_014815.1 ( <i>Micromonospora</i> sp. L5):							
Region 1	Terpene	221533	242010	Sioxanthin	hybrid	100%	BGC0001087
Region 2	Terpene	313898	334259	Phosphonoglycans	saccharide	3%	BGC0000806
Region 3	Nrps	822679	876586	Zorbamycin	hybrid	4%	BGC0001058
Region 4	T3pks	2761812	2802873	Alkyl-O-Dihydrogeranyl-Methoxyhydroquinones biosynthetic gene	hybrid	71%	BGC0001077
Region 5	Terpene	4058457	4077583				
Region 6	Bacteriocin - Terpene	4219887	4243981	Lymphostin	hybrid	33%	BGC0001006
Region 7	Siderophore	4603155	4615003				
Region 8	Oligosaccharide - T2pks - Other	4644202	4739578	Cinerubin B	polyketide	62%	BGC0000212
Region 9	Transatpks - Nrps - Otherks	4791253	4886022	Leinamycin	hybrid	15%	BGC0001101
Region 10	Betalactone	5007285	5029399	Tautomycetin	polyketide	14%	BGC0000158
Region 11	Oligosaccharide - Nrps - Terpene	5033252	5092158	Lobosamide	t1pks	13%	BGC0001303
Region 12	T2pks	5269338	5341912	Xantholipin	polyketide	14%	BGC0000279
Region 13	Terpene	5524214	5545416	Sioxanthin	hybrid	100%	BGC0001087
Region 14	Nrps - T1pks	6014732	6080342	Bleomycin	hybrid	12%	BGC0000963
Region 15	Otherks - Nrps - T1pks - Siderophore	6169329	6273581	Azicemicin	polyketide	13%	BGC0000202
Region 16	T1pks	6308998	6352726	Tiancimycin	t1pks	19%	BGC0001378
Region 17	T1pks - Nrps	6358138	6596525	Rifamycin	polyketide	35%	BGC0000137

Diterpenes in general own an extensive range of biological activities though, the specific class of labdane diterpenoids appeared as novel lead compounds for current drug discovery with therapeutic properties such as, antimicrobial, antifungal, anticancer, anti-inflammatory, antiparasitic, antiviral, immunomodulatory functions and as well as larval growth inhibitor (Elliger et al., 1976, Tran et al., 2017). Furthermore, several compounds derived from other diterpenoids have been described, with the representatives of this class being tanshinones, carnosic acid and carnosol, forskolin, andrographolide, sclareol, (+)-polyalthic acid and galanal A (Table 4.13)

**Table 4-13 Biological activity and source of other-diterpenoids.**

<b>Name of labdane-type diterpene</b>	<b>Source</b>	<b>Biological activities</b>	<b>Literature</b>
tanshinones	roots of <i>Salvia miltiorrhiza</i>	antioxidant, anti-inflammatory, cardiovascular and cerebrovascular protective effects	(Xu, 2011, Robertson et al., 2014, Ignea et al., 2016)
carnosic acid and carnosol	<i>Salvia spp.</i> <i>Rosmarinus officinalis</i>	antioxidant anti-adipogenic anti-cancer	(Gaya et al., 2013, Danilenko et al., 2003)
forskolin	root cork cells of <i>Coleus forskohlii</i>	positive inotropic agent, vaso and bronchodilatador, weight loss, treatment of open angle glaucoma, anti-hypertensive, protection congestive heart failure	(de Souza et al., 1983, Huerta et al., 2010, Ho and Shi, 1982, Lindner et al., 1978, Caprioli and Sears, 1983, Bristow et al., 1984)
andrographolide	extracts of <i>Andrographis paniculata</i>	anti-inflammatory, anti-hypertensive	(Sheeja et al., 2006, Awang et al., 2012)
sclareol	<i>Salvia officinalis</i>	antiphotoaging	(Park et al., 2016)
(+)-polyalthic acid	<i>Vitex rotundifolia</i>	antimutagenic	(Miyazawa et al., 1995)
galanal A	flower buds of <i>Myoga</i>	anti-inflammatory	(Yamamoto et al., 2014)

One of the described biological effects of the labdane diterpenoids was larval growth inhibition, as mentioned above. Resilient varieties of sunflower (*Helianthus annuus* L.) showed resistance to attacks by larvae of the sunflower moth (*Homeosoma electellum* H.). It was revealed that the extracts of this variety's florets contained greater quantities of larvicidal diterpenoid acids. One of the possible explanations for the mechanism of action for inhibitory larval growth effect could involve the hormonal system of the insect, as the skeletal structure of diterpenes acids, resembles the common steroids, it seemed possible that it could be interfering with the insect steroid metabolism. This interference in any step of hormones metabolism by inhibition of it, would decrease the vitality of the larva (Elliger et al., 1976). There could exist a parallelism between this effect and the potential insecticide or/and pesticide effect of SBT687, particularly against sea lice. Further research could adapt the mechanisms of action of larval growth inhibitors in or order to be a potential target in inhibiting the sea lice infestation

Some diterpenes are known for having insecticidal activity. The following labdane diterpenes 9-acetoxy-2 $\alpha$ ,7 $\alpha$ -dihydroxylabda-8(17),(13Z)-diene, 15-19-diacetoxy-2 $\alpha$ ,7 $\alpha$ -dihydroxylabda-8(17),(13Z)-diene, 7 $\alpha$ ,15,19-triacetoxy-2 $\alpha$ ,7 $\alpha$ -hydroxylabda-8(17),(13Z)-diene, 19-acetoxy-2 $\alpha$ ,7 $\alpha$ -dihydroxylabda-8(17),(13Z)-dien-15-al, 19-acetoxy-7 $\alpha$ ,15-dihydroxylabda-8(17),(13Z)-dien-2-one, 19-acetoxy-2 $\alpha$ , 7 $\alpha$  -dihydroxylabda-14,15-dinorlabd-8(17)-en-13-one and 2 $\alpha$ , 7 $\alpha$ ,15,19-tetrahydroxy-ent-labda-9(17),(13Z)-diene were isolated from the aerial parts of *Hyptis spicigera* and exhibited insecticidal properties against larvae of *Ostrinia nubilalis* Hubner (Fragoso-Serrano et al., 1999). Various chemical classes of terpenoids were also tested against the herbivorous insect *Leptinotarsa decemlineata*, among them the labdane diterpene 2 $\alpha$ ,3 $\alpha$ -dihydroxycativic acid. This compound showed active antifeedant effect and toxic activity to *L. decemlineata*, which confidently, did not present cytotoxicity on mammalian cell lines. A mechanism of action related with membrane-dependent-modifying effects have been proposed for discriminatory cytotoxic effect rather than a general cytotoxic action (Gonzalez Coloma).

Terpenoids being the largest group of natural compounds, provide new inspiring opportunities for drug discovery. In addition to the vast therapeutic effects, terpenoids play an important role in plant interactions thus, the need for further investigation of the potential biological as well eco-chemical functions (Ludwiczuk et al., 2017). The improvement of the regulation of terpene metabolism will enable the manipulation of terpene biosynthetic pathways for enhancement of biotransformation of medicinally important terpenes and production of both known and novel compounds (Singh and Sharma, 2015).

#### **4.5.2 Manzamines the $\beta$ -carboline alkaloid compounds with significant insecticidal/antifeeding activity**

Fraction 17 afforded the identification of a mixture of manzamines (D, H, J *N*-oxide, 8-hydroxymanzamine A and *N*-methyl-epi-manzamine D) with the visualization of characteristic signals of the  $\beta$ -carboline moiety and attached heterocyclic ring. Manzamines are a structurally interesting group characterized by a complex, heterocyclic ring system attached to a  $\beta$ -carboline moiety (Ang et al., 2000). Since the discovery of manzamine A, more than 100 manzamine-related alkaloids have been identified (Furusato et al., 2014). Initially, this class of compounds were isolated from the marine sponge *Haliclona* sp. but over the years, derivatives have been found in numerous distinct genera of marine sponges frequently from the Indian and Pacific oceans (Table 4.14) (Ashok et al., 2014). Marine sponges contain microbes up to 40% of their total biomass by providing the favorable environmental conditions (Valliappan et al., 2014). Amongst the sponge-associated bacteria, the genus *Micromonospora* are one of the dominant producers of bioactive natural products like manzamines (Valliappan et al., 2014). The fact that the manzamines were isolated from a diverse variety of sponges, distantly related, and geographically disparate suggests the possibility that the compounds are actually being produced by the sponge-associated microorganisms, such as a *Micromonospora* species, and not by the individual sponge itself (Rao et al., 2006, Taylor et al., 2007).

The isolation of a genus or strain of bacteria, that is able to produce manzamine derivatives, can be cultured in a large-scale fermenter, which would improve concerns with environmental issues and expedite the route of these compounds through numerous phases of the drug-testing process (Taylor et al., 2007). Growing manzamine-producing bacteria in order to have sufficient yield of manzamines for pharmaceutical and agrochemical research is far more sustainable than harvesting enormous quantities or tonnes of sponges that will instigate environmental concerns and impact (Hill, 2005). Exclusive laboratory synthesis of the manzamines is not sustainable due to the complexity of the chemical structure of the molecule, high costs of production, and intrinsic production of undesirable by-products (Waters et al., 2014). Nevertheless, in 2005, sponge-derived *Micromonospora* was grown at large scale up to 10 liters of fermentation culture with preservation of manzamine yield from a small scale up fermentation (Hill, 2005).

**Table 4-14 Some of the relevant manzamines isolated since manzamine A in 1986**  
(Human immunodeficiency virus (HIV-1), Herpes simplex virus 1 e 2 (HSV-I and HSV-II).

<b>Name of manzamine</b>	<b>Source</b>	<b>Biological activity</b>	<b>Literature</b>
manzamine A	<i>Okinawa</i> sponge <i>Haliclona</i> sp.	antitumor, antimalarial, antiparasitic, antileishmanial activity against <i>Mycobacterium tuberculosis</i> , anti-HSV-I, insecticidal, antifungal	(Sakai et al., 1986, Ashok et al., 2014),(Simithy et al., 2018, Peng et al., 2003)
manzamine B	<i>Haliclona</i> sp.	anti-inflammatory, antimalarial	(Sakai et al., 1987, Rao et al., 2006)
manzamine C	<i>Haliclona</i> sp.	neuro anti- inflammatory	(Sakai et al., 1987, Mayer et al., 2005)
manzamine D	<i>Haliclona</i> sp. <i>Ircinia</i> sp. <i>Amphimedon</i> sp.	antitumor, strong activity against <i>Mycobacterium tuberculosis</i>	(Watanabe et al., 1998)
manzamine E	<i>Xestospongia</i> sp.	moderate antimalarial, activity against <i>Mycobacterium tuberculosis</i>	(Ichiba et al., 1988, Rao et al., 2003, Simithy et al., 2018)

manzamine F	<i>Xestospongia sp.</i>	moderate antimalarial, activity against <i>Mycobacterium tuberculosis</i>	(Ichiba et al., 1988, Rao et al., 2006, Simithy et al., 2018)
manzamine H	<i>Ircinia sp.</i>	cytotoxic for leukaemia I210 cells and KB epidermoid carcinoma cells), antimycobacterial	(Kondo et al., 1992)
manzamine J	<i>Ircinia sp.</i>	antimalarial, antiparasitic	(Kondo et al., 1992)
ircinal A	<i>Ircinia sp.</i>	antimalarial, antiparasitic, activity against <i>Mycobacterium tuberculosis</i>	(Kondo et al., 1992, Rao et al., 2006)
ircinal B	<i>Ircinia sp.</i>	antimalarial, antiparasitic	(Kondo et al., 1992)
8-hydroxymanzamine A	<i>Pachypellzna sp.</i> <i>Amphimedon sp.</i> <i>Acanthostrongylophora aff. ingens</i>	strong antimalarial, antitumor, anti-HSV-II, antileishmanial, strong activity against <i>Mycobacterium tuberculosis</i>	(Ichiba et al., 1994),(Simithy et al., 2018)
manzamine Y	<i>Amphimedon sp.</i>	antimalarial	(Kobayashi et al., 1994)
3,4-dihydromanzamine A	<i>Amphimedon sp.</i>	antimalarial	(Kobayashi et al., 1994)
xestomanzamine A	<i>Xestospongia sp.</i>	weak cytotoxicity for KB cell lines	(Kobayashi et al., 1995)
xestomanzamine B	<i>Xestospongia sp.</i>	weak cytotoxicity for KB cell lines	(Kobayashi et al., 1995)
manzamine X	<i>Xestospongia sp.</i>	weak cytotoxicity for KB cell lines	(Kobayashi et al., 1995)
manzamine Y	<i>Xestospongia sp.</i>	weak cytotoxicity for KB cell lines	(Kobayashi et al., 1995)
kauluamine	<i>Prianos sp.</i>	moderate immunosuppressive	(Ohtani et al., 1995)
manzamine L		moderate cytotoxicity for leukaemia I210 cells and KB epidermoid carcinoma cells	(Tsuda et al., 1996)
6-deoxymanzamine-X	<i>Xestospongia sp.</i>	moderate cytotoxicity for L5178y cell lines, anti-insecticidal	(Edrada et al., 1996)
manzamine J N-oxide	<i>Xestospongia sp.</i>	moderate cytotoxicity for L5178y cell lines, anti-insecticidal	(Edrada et al., 1996)

3,4-dihydromanzamine-A-N-oxide	<i>Xestospongia sp.</i>	moderate cytotoxicity for L5178y cell lines, anti-insecticidal	(Edrada et al., 1996)
manzamine A N-oxide	<i>Xestospongia sp.</i>	moderate cytotoxicity for L5178y cell lines	(Edrada et al., 1996)
manzamine M	<i>Amphimedon sp.</i>	moderate cytotoxicity for leukaemia I210 cells	(Watanabe et al., 1998)
epimanzamine D	<i>Palaun</i> sponge	moderate cytotoxicity for HeLa cell lines	(Zhou et al., 2000)
N-methyl epimanzamine D	<i>Palaun</i> sponge	moderate cytotoxicity for HeLa cell lines	(Zhou et al., 2000)
ent-8-hydroxymanzamine-A	<i>Prianos sp.</i>	antiparasitic	(El Sayed et al., 2001)
ent-8-hydroxymanzamine-F	<i>Prianos sp.</i>	antiparasitic	(El Sayed et al., 2001)
neo-kaulamine	<i>Prianos sp.</i>	antiparasitic	(El Sayed et al., 2001)
manadomanzamines A and B	<i>Acanthostrongylophora sp.</i>	strong activity against <i>Mycobacterium tuberculosis</i> , strong activity against HIV-1, moderate activity against AIDS opportunistic infections	(Peng et al., 2003)
zamamidine C	<i>Amphimedon sp.</i>	moderate antimalarial	(Yamada et al., 2009)
acanthomanzamines A-E	<i>Acanthostrongylophora ingens</i>	cytotoxicity for HeLa cells, inhibition of the proteasome	(Furusato et al., 2014)
ircinal E	<i>Acanthostrongylophora ingens</i>	cytotoxicity for murine lymphoma L5178Y cell line	(AlTarabeen et al., 2015)
kepulauamine A	<i>Acanthostrongylophora sp.</i>	antibacterial	(Kim et al., 2017a)
zamamidine D	<i>Amphimedon sp.</i>	antimicrobial	(Kubota et al., 2017)

Manzamines account significant importance as potential pharmaceutical, associated with a great number of substantial biological activities including insecticidal, cytotoxicity, anti-inflammatory, antimicrobial (antibacterial and antifungal), anti-infective, antiparasitic as well with high potential for future clinical applications for malaria, *Mycobacterium tuberculosis*, neurological diseases, HIV, HSV and several types of cancer (Rao et al., 2003, Rao et al., 2006). The mentioned biological activities can suffer modification due to changes in each of the structural moieties, with available literature detailing extensive SAR studies on antimalarial, antileishmanial, anti-HIV, anti-HSV, antitumor activities (Ibrahim et al., 2008, Guzmán et al., 2011, Ashok et al., 2014, Peng et al., 2003, Chatwichien et al., 2015). Nonetheless, there is a lack of SAR studies concerning the insecticidal activity. The insecticidal activity of the *Micromonospora sp.* N17 (SBT687) practically decreased from 100% inhibition to 72% after fractionation, this could be explained that the activity is stronger in the crude extract where all compounds, predominantly the ones identified/mentioned, are organised with a synergistic effect. Further studies about the interaction with terpenes and alkaloids extracted from *Micromonospora sp.* could be interesting to clarify the accurate presence of insecticidal activity. Other additional noteworthy fact is that the described manzamines were identified from a *Micromonospora sp.* isolated from *Phorbas tenacior*, a Mediterranean sponge collected from the Santorini volcanic complex of Crete. Generally, the manzamine-yielding sponges were collected from Indian and Pacific oceans as mentioned earlier, this wide distribution across distantly related sponges, proposes, once again, the putative role of microbes in the biosynthesis of manzamines. Biomimetic fermentation approach embraces substantial potential to produce manzamines, considered to be the key for source challenges and the availability of a sustainable supply for chemically complex precursors which are critical to future research studies.

Antifeedant compounds belong to diverse classes, such as alkaloids, flavonoids and terpenes. Manzamine alkaloids, have been reported to have antifeedant insecticidal activity. There are numerous identified antifeedant mechanisms of action such as chordotonal organ modulator which cause hind leg extension, as well as block femoral chordotonal organ discharge (Taylor-Wells et al., 2018) or could in general act on

specific sensory cells (antifeedant receptors) in the pest causing different effects that prevent or decrease or block insect feeding (Purrington, 2003). The general mechanism of action of manzamine antifeedants could be used as a potential target in inhibiting sea lice infestation, hence, hypothetically SBT687 application as a potential insecticide or pesticide against sea lice.

Beta-carboline compounds are responsible for a high range of biological properties. The fact their structure modification is straightforward, makes these compounds of interest for synthesis and structure-activity relationships with the purpose of novel drugs development with insecticidal activity (Zeng et al., 2010). Harmine compounds belongs to the beta-carboline alkaloids with known reported insecticidal activity (Yao et al., 2004). A series of 1,3-substituted beta-carboline derivatives (tetrahydro- $\beta$ -carboline) from harmine were studied in terms of its insecticidal effect and was acknowledged that 1-phenyl-1,2,3,4-tetrahydro- $\beta$ -carboline-3-carboxylic acid and methyl-1-phenyl- $\beta$ -carboline-3-carboxylate were the most promising compounds, exhibiting high inhibition rates of insect cultured Sf9 cell line *in vitro* and *in vivo* insecticidal activities against instar larvae of mosquitos, *Culex pipiens quinquefasciatus* and mustard aphid, *Lipaphis erysimi* (Zeng et al., 2010). Further study described the synthesis of several 1,2,3,4-tetrahydro- $\beta$ -carboline-3-carboxylic acid derivatives from harmine and evaluated their larvicidal activities, showing low insecticidal activities against lepidopteran pests and mosquito larvae but were higher against oriental armyworm (*Mythimna separate*) and pyrausta nubilis (*Ostrinia nubilalis*) (Song et al., 2014). A recent study published in 2018 synthesised a series of novel  $\beta$ -carboline derivatives from tetrahydro- $\beta$ -carboline ester with hydantoin, thiodiyantoin, and urea as part of their structures or formulation in order to improve and modulate their activity. The insecticidal activity was tested against *Plutella xylostella* and *Culex pipiens pallens* and part of these compounds demonstrated good activity, particularly the ones with hydantoin and urea (Huang et al., 2018). Interestingly, the benzoyl urea diflubenzuron and teflubenzuron are used for in-feed treatments as effective control method of the sea lice (Olsvik et al., 2013, J Branson et al., 2000), which make the beta-carboline compounds containing urea, of interest for future evaluations against sea lice. Structure modifications of the beta-carboline

moiety containing urea with manzamine congeners can be designed to study the potential insecticidal activity against sea lice.

## Chapter 5

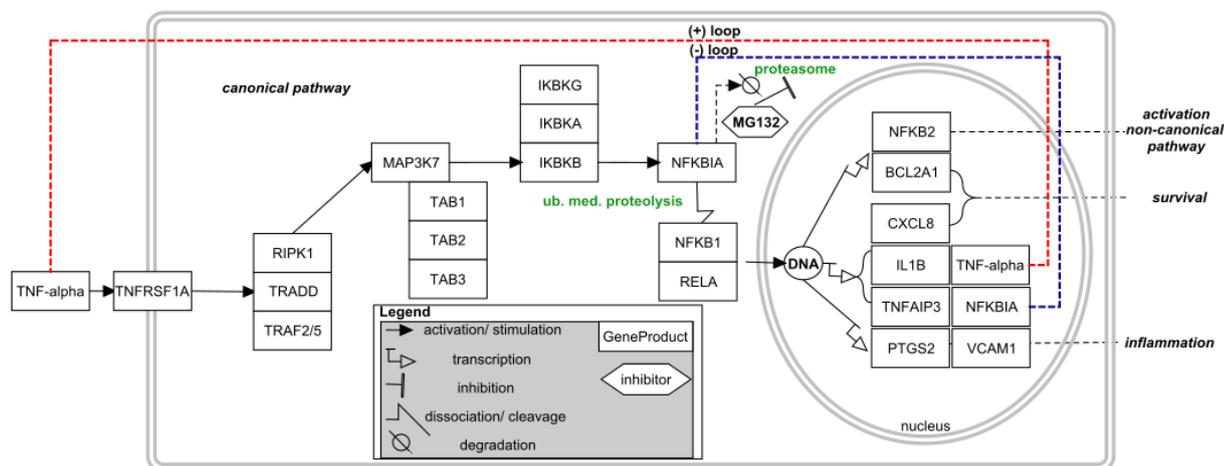
*Micromonospora* sp. N74 (SBT692): SEARCH FOR NEW  
POTENTIAL ANTI-CANCER DRUGS

## 5.1 Introduction

### 5.1.1 Anti-cancer BioScreening Campaign

Worldwide incidence of cancer in 2018 was projected to have risen to 18.1 million new cases and 9.6 million deaths (Bray et al., 2018). Globally, cancer lifetime-prevalence sets that 1/ 5 men and 1/ 6 women will develop the disease, and 1/ 8 men and 1/ 11 women will die from this ailment (Fitzmaurice et al., 2017). A major number of cancer-related deaths worldwide are associated with malignant tumours of the breast, colon and rectum (CRC) (Fitzmaurice et al., 2017). This has led to the development of various therapeutic strategies to prolong survival including improved surgical techniques, anti-angiogenesis therapies and adjuvant/neoadjuvant chemoradiotherapy. Thereby, is of great importance to devise methodologies to not only pharmacologically treat, but also to diagnose these cancer types at an early-stage.

The nuclear factor-kappa B (NF-kappa B, NFkB) belongs to a family of transcription factors (TFs) that function as dimers and regulate genes involved in immunity, inflammation and cell survival (Egan and Toruner, 2006). Several pathways are known to induce NF-kappa B-activation (Oeckinghaus et al., 2011, Perkins, 2007). The canonical pathway (Figure 5.1) is induced by tumour necrosis factor-alpha (TNF-alpha), interleukin-1 (IL-1) or by-products of previous infections of bacterial and/or viral origin. This pathway relies on IKBKB mediated Ikb-alpha (NFKBIA) phosphorylation of amino acids 32 and 36 serine (Ser) residues, leading to its ubiquitination and proteasomal degradation (Mathes et al., 2008), which allows the NFKB1 (p50)/ RELA (p65) NFkB-dimer to enter the nucleus and activate the transcriptional machinery of target genes (Oeckinghaus et al., 2011, Perkins, 2007). Regulation and control of NFkB activity throughout pharmacological intervention, would provide a potential approach for the management of many NFkB-associated human ailments (Kumar et al., 2004).



**Figure 5-1 Inhibition of TNF-alpha-NFkB by MG132.** Activation of NF-kappa B (NFkB) signalling pathway by the cytokine-triggered TNF-alpha in Human keratinocytes (NCTC 2544). NCTC/NFkB-Luciferase cells were treated with MG132 and exposed to TNF-alpha and then undergone luciferase assay to assess NFkB activity. Treatment with MG132 inhibits TNF-alpha-triggered/induced degradation of NFKBIA (I $\kappa$ B-alpha). Ubiquitin mediated (ub. med.); N-benzyloxycarbonyl-L-leucyl-L-leucyl-L-leucinal (MG132); NF-kappa-B transcription factor (RELA-NFKB1 complex).

Concerning progression of non-hereditary CRC, two main mechanism hits (Board, 2019, De Rosa et al., 2015) were identified to yield genetic instability: *i*) chromosomal instability from triggering proto-oncogenes such as GTPase KRas (KRAS) and silencing of tumour suppressing genes such as cellular tumour antigen p53 (TP53), netrin receptor DCC (DCC)/ mothers against DPP homolog 4 (SMAD4), and adenomatous polyposis coli protein (APC); *ii*) microsatellite instability from silencing DNA mismatch repair genes such as MutL protein homolog 1 (MLH1) and/or MutS protein homolog 2 (MSH2) by promoter hypermethylation thereby repressing gene expression, and targeted mutations of genes coding short tandem repeats (STRs), for example TGF-beta receptor type- 2 (GFBR2) and apoptosis regulator BAX (BAX) (Khare and Verma, 2012). Hereditary CRC forms, for instance familial adenomatous polyposis have prevalence of germline defects in APC gene on chromosome 5 (Board, 2019).

Currently, molecular stratification of breast carcinomas are by large guided by the presence or absence of specific hormone receptors such as for estrogen and progesterone, and expression of the receptor tyrosine-protein kinase erbB-2 (ERBB2, HER2) (Wang et al., 2016a, Harrison, 1989). Patients missing or under-expressing all three receptors are designated as suffering from triple negative breast carcinomas (TNBC) therefore chemotherapies relying on hormonal treatment and/or with

herceptin don't have an effect on those groups (Hon et al., 2016, Rakha and Green, 2017). Thus, novel therapies are needed to directly target these receptors.

Selection of *Micromonospora* sp. N17 (SBT692) for further culture optimisation, fractionation work and more intensive bioscreening campaign was due to the positive preliminary screening results obtained from SeaBioTech consortium partners. Pilot-studies on SBT692 crude extracts exhibited activity on TNF-alpha-induced NF-kB (SIPBS, UK) with 88.98% inhibition at 30 µg/mL. It showed weak activity against transient receptor potential cation channels that included subfamily A, member 1 (TRPA1) with 56.09% at 50 µg/mL and subfamily M member 8 (TRPM8) with 36.03% at 50 µg/mL while SBT692 further showed activity on peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ ) with 84.88% at 30 µg/mL (Axxam, Italy). Additionally, SBT692 revealed to potentially inhibit human colorectal carcinoma cell (HCT116) growth by 53% at 30 µg/mL (Horizon, UK).

TRPA1 is a nonselective cation channel well-known for its role as a sensor of xenobiotics and can become active by induction of chemical mediators of inflammation, giving rise to pain and neurogenic inflammation (Rech et al., 2010). Therefore, is an appealing target for therapeutic intervention in pain relief (Noyer et al., 2018) with many antagonists yet to be discovered. Thermoreceptor TRPM8 (Noyer et al., 2018), was earlier seen as a modulator of chronic pain-related pathways, and more recently as a potential molecular target in androgen-regulated prostate cancer (Zhang and Barritt, 2004). Peroxisome proliferator activated receptor alpha (PPAR $\alpha$ ) is a nuclear receptor that modulates hepatocarcinogenesis (observed in rodent models chronically fed with bezafibrates) (Hays et al., 2005). Additionally, PPAR $\alpha$  regulates lipid metabolism, for instance being active under conditions of energy deprivation (Shao et al., 2014).

### 5.1.2 Aim(s)

Here, *Micromonospora* sp. N17 (SBT692) isolated from *P. tenacior* within the Santorini volcanic caldera was employed for initial screening and bioassay-guided fractionation for further characterisation (LC-MS/MS and NMR-based) of relevant compounds in: (i) inhibition of cytokine (TNF-alpha)-triggered NF-kB activation in human keratinocytes (NCTC 2544), (ii) viability of tumour-derived cell models for breast (MCF-7 and MDA-MB-231), colon (HT29 and CaCo2) and advanced colon (T84) cancer.

## **5.2 Small and Medium scale up fermentation, extraction and fractionation of SBT692**

### **5.2.1 Scaled-up fermentation of SBT692**

The detailed methods concerning the fermentation and extraction were explained previously on Chapter 2. The initial small-scale fermentation of the bacterium *Micromonospora* sp. N74 (SBT 687) was performed at Julius-von-Sachs-Institut für Biowissenschaften, University of Wuerzburg, Germany while the bioactivity screening for TNF-alpha inhibition (NFκβ Luciferase Method), cell cytotoxicity and viability were done in SIPBS. The conditions for optimal growth and best chemical profile for SBT692 were achieved with 10% (v/v) inoculum on ISP2 medium and incubation for 10 days at 30° C and 120 rpm. In SIPBS, SBT692 was then scaled up first to 3L (6 x 500 mL) then twice to 15-L in production flasks (30 x 500 mL).

### **5.2.2 Fractionation of SBT692**

After 10 days of incubation, the SBT692 3L scale-up was extracted with ethyl acetate as described in section 2.4. The yield of the obtained total organic extract was 168.3 mg. The ethyl acetate extract was fractionated using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash on a reverse phase column SNAP Kp-C18-HS-12g with a samplet 1.2 g KP-C18-HS-PK20. The fractions were eluted by binary gradient step elution method using water with 0.1% FA (A) and ACN with 0.1% FA (B) as solvents systems as described in section 2.6.2.2. The gradient system is presented in the Table 5.1 and chromatographic conditions are detailed in Table 5.2.

**Table 5-1 Gradient elution of crude extract obtained from 3L of SBT692 using Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification.**

<b>time (min)</b>	<b>% Solvent B</b>
<b>0</b>	5
<b>5</b>	5
<b>15</b>	10
<b>20</b>	20
<b>40</b>	50
<b>50</b>	50
<b>60</b>	100
<b>70</b>	100
<b>Total time=70</b>	

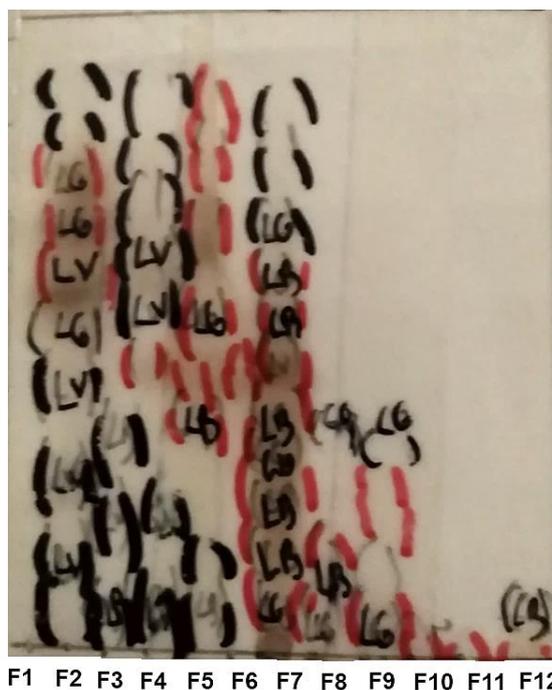
**Table 5-2 Chromatographic conditions applied on the fractionation of SBT692 crude extract obtained from 3L scale-up using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification.**

<b>Column</b>	SNAP Kp-C18-HS-12g
<b>Samplet</b>	1.2 g KP-C18-HS-PK20
<b>Solvents</b>	water with 0.1% FA (A) and ACN with 0.1% FA (B)
<b>Flow rate</b>	12 mL/min
<b>Detection mode</b>	UV1+UV2
<b>Start Threshold</b>	20 mAU
<b>UV1 Wavelength</b>	254 nm
<b>UV2 Wavelength</b>	280 nm

The fractionation of SBT692 resulted to 56 fractions. Fractions with similar TLC profiles were pooled together, yielding a total of 12 fractions (F1-F12) shown on table 5.3 and Figure 5.2.

**Table 5-3 Fractionation of SBT692 crude extract from 3L scale-up affording 12 fractions.**

<b>Fraction</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
F1	1-5	20.3
F2	6-10	6.5
F3	11-15	49.3
F4	16-20	63.3
F5	21-25	86.4
F6	26-30	48
F7	31-35	49.3
F8	36-40	51.5
F9	41-45	43.4
F10	46-50	3.6
F11	51-56	5.9
F12	Column washing	49.5



**Figure 5-2 Summary TLC of the first fractionation of SBT692 crude extract obtained from a 3L scale-up with solvent system ACN:Water 1:1 on RP18 plate.** Brown spots were visualized after spraying with anisaldehyde reagent. As observed under short and long UV wavelength, the spots were marked S (Short), LB (Long Blue) LG (Long Green), LV (Long Violet).

It was observed on the TLC plate from the SBT692 3L scale-up apart from the UV visible compounds, after spraying with anisaldehyde/H<sub>2</sub>SO<sub>4</sub> reagent the presence of organic compounds without UV absorbance that could be steroids, terpenes, phenolic acids among others. Therefore, in order to better isolate these putative interesting compounds, the fractionation of the crude extract (1.3g) from the first 15 L scale-up was performed using the Reveleris™ Flash Chromatography System on a reverse phase C18 column 40g cartridge instead of the Biotage MPLC Isolera™ One 2.0.4 Spektra flash. Since this system holds two detectors, an ELSD and a UV detector from which a certain working wavelength can be chosen (200-500 nm), this guarantees superior sensitivity, selectivity and detection of all peaks including UV-inactive compounds. Elution was done by gradient method using water and 0.1% FA (A) and ACN and 0.1% FA (B) as solvent systems. The run length of time was increased, as well as the polarity (decreasing the initial % of ACN) to improve the separation of the

highly polar compounds. The gradient system is presented in the Table 5.4 and chromatographic conditions are detailed in Table 5.5. Aliquots of each fraction were prepared for MS, NMR and bioactivity assays. Due to limited weights for some fractions, it was not possible to do some of the chemical analysis and bioassay screening procedure.

**Table 5-4 Gradient elution of SBT692 crude extract obtained from the first 15 L scale-up using the Reveleris™ Flash Chromatography System.**

<b>time (min)</b>	<b>% Solvent B</b>
<b>0</b>	3
<b>10</b>	5
<b>20</b>	10
<b>30</b>	20
<b>50</b>	20
<b>60</b>	50
<b>70</b>	100
<b>95</b>	100
<b>Total time=95</b>	

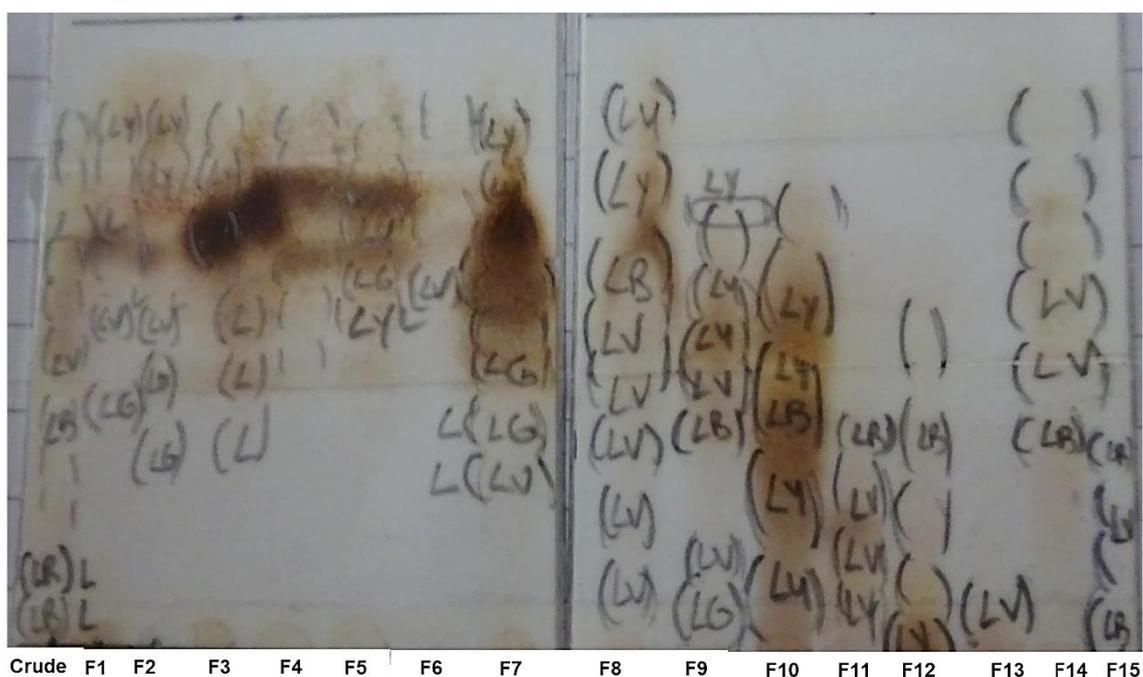
**Table 5-5 Chromatographic conditions applied on fractionation of SBT692 crude extracts obtained from the first 15L scale-up using the Reveleris™ Flash Chromatography System.**

<b>Column</b>	Reverelis C18 40g
<b>Sample</b>	3 spoonful of Celite® S with dissolved sample in MeOH
<b>Solvents</b>	water with 0.1% FA (A) and ACN with 0.1% FA (B)
<b>Flow rate</b>	20 mL/min
<b>Detection mode</b>	UV1+UV2
<b>ELSD Threshold</b>	5 mV
<b>UV Threshold</b>	0.05 AU
<b>UV1 Wavelength</b>	290 nm
<b>UV2 Wavelength</b>	320 nm

The fractionation of SBT692 crude extract obtained from the first 15L scale-up resulted to 216 fractions. Fractions with similar TLC profiles were pooled together, yielding a total of 15 fractions (F1-F22) presented on table 5.6 and Figure 5.3.

**Table 5-6 Fractionation of SBT692 crude extract from the first 15L scale-up afforded 15 fractions.**

<b>Fraction</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
F1	1-15	62
F2	16-19	34
F3	20-31	53.3
F4	32-36	16
F5	37-40	12.8
F6	41-45	14.5
F7	46-105	173
F8	106-144	512.4
F9	145-153	20.3
F10	154-170	42.7
F11	171-184	25.3
F12	185-188	5.4
F13	189-216	28.2
F14	Wash run	12.4
F15	Column washing	7



**Figure 5-3 Summary TLC of the fractionation of SBT692 crude extract obtained from the first 15L scale-up with solvent system ACN:Water 1:1 on RP18 plates.** Brown spots were visualized after with anisaldehyde spraying reagent. As observed under short and long UV wavelength, the spots were marked S (Short), LB (Long Blue), LG (Long Green), LY (Long Yellow), LV (Long Violet).

The fractionation of the second 15 L scale-up was performed on 1.1 g of crude extract using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification with a reverse phase column SNAP Kp-C18-HS-30g with a 3g samplet KP-C18-HS-PK20. Because the Reveleris™ Flash Chromatography System was not available at the time of this fractionation, the Biotage MPLC Isolera™ One 2.0.4 Spektra flash was used, using the same binary gradient elution. Binary gradient elution method used water with 0.1% FA (A) and ACN with 0.1% FA (B) as solvent system presented in the Table 5.7. Chromatographic conditions are shown in Table 5.8.

**Table 5-7 Step gradient elution of SBT692 crude extract obtained from the 2nd 15L scale-up using Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification.**

<b>time (min)</b>	<b>% Solvent B</b>
<b>0</b>	3
<b>5</b>	3
<b>15</b>	10
<b>25</b>	10
<b>35</b>	20
<b>45</b>	30
<b>60</b>	50
<b>75</b>	100
<b>85</b>	100
<b>Total time=85</b>	

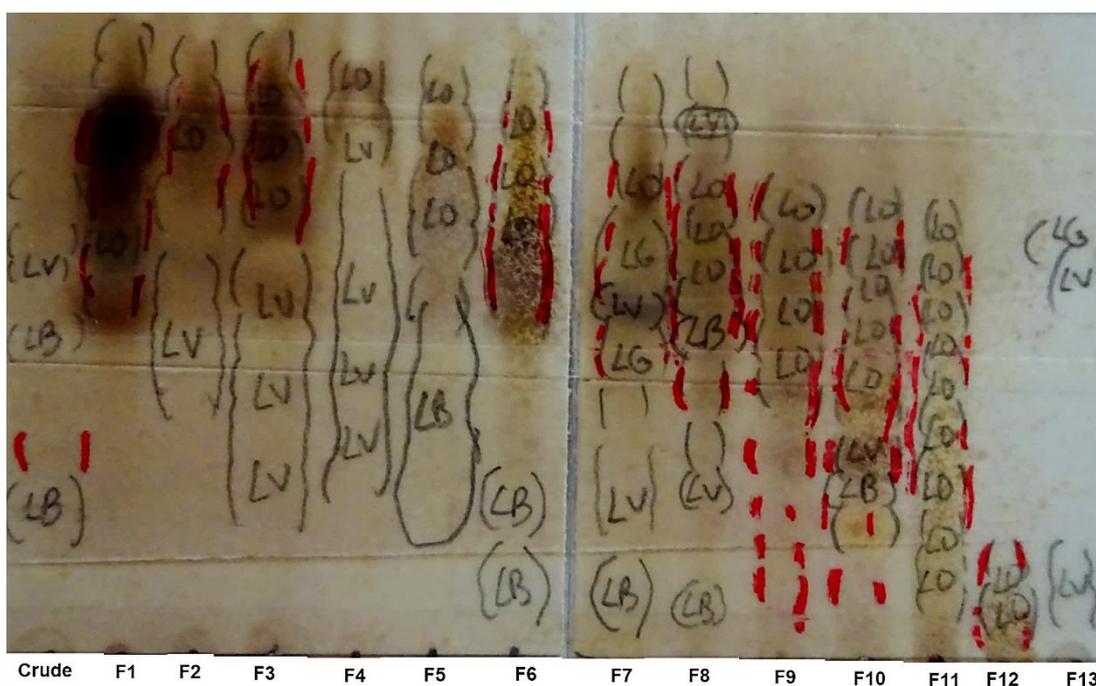
**Table 5-8 Chromatographic conditions applied to the crude extract of the second 15L scale-up of SBT692 using the Biotage MPLC Isolera™ One 2.0.4 Spektra flash purification.**

<b>Column</b>	SNAP Kp-C18-HS-30g
<b>Samplet</b>	3 g KP-C18-HS-PK20
<b>Solvents</b>	water with 0.1% FA (A) and ACN with 0.1% FA (B)
<b>Flow rate</b>	25 mL/min
<b>Detection mode</b>	UV1+UV2
<b>Start</b>	20 mAU
<b>Threshold</b>	
<b>UV1</b>	254 nm
<b>Wavelength</b>	
<b>UV2</b>	280 nm
<b>Wavelength</b>	

The fractionation of the SBT692 crude extract that was obtained from second 15L scale-up resulted to 119 fractions. Again, fractions with similar TLC profile were pooled together, yielding a total of 13 fractions (F1-F13) shown on table 5.9 and Figure 5.4. Fractions from both batches of 15L scale-up were analysed by NMR, LC-HRMS and assayed for their bioactivity.

**Table 5-9 Fractionation of SBT692 crude extract from the second 15L scale-up afforded 13 fractions**

<b>Fraction</b>	<b>Test Tube numbers</b>	<b>Yield (mg)</b>
F1	1-12	63.6
F2	13-21	45.9
F3	22-36	42.7
F4	37-47	323.4
F5	48-50	61.4
F6	51-54	83
F7	55-59	45.6
F8	60-64	14.2
F9	65-71	33.7
F10	72-76	73
F11	77-94	16
F12	95-119	43.7
F13	Column washing	7.4

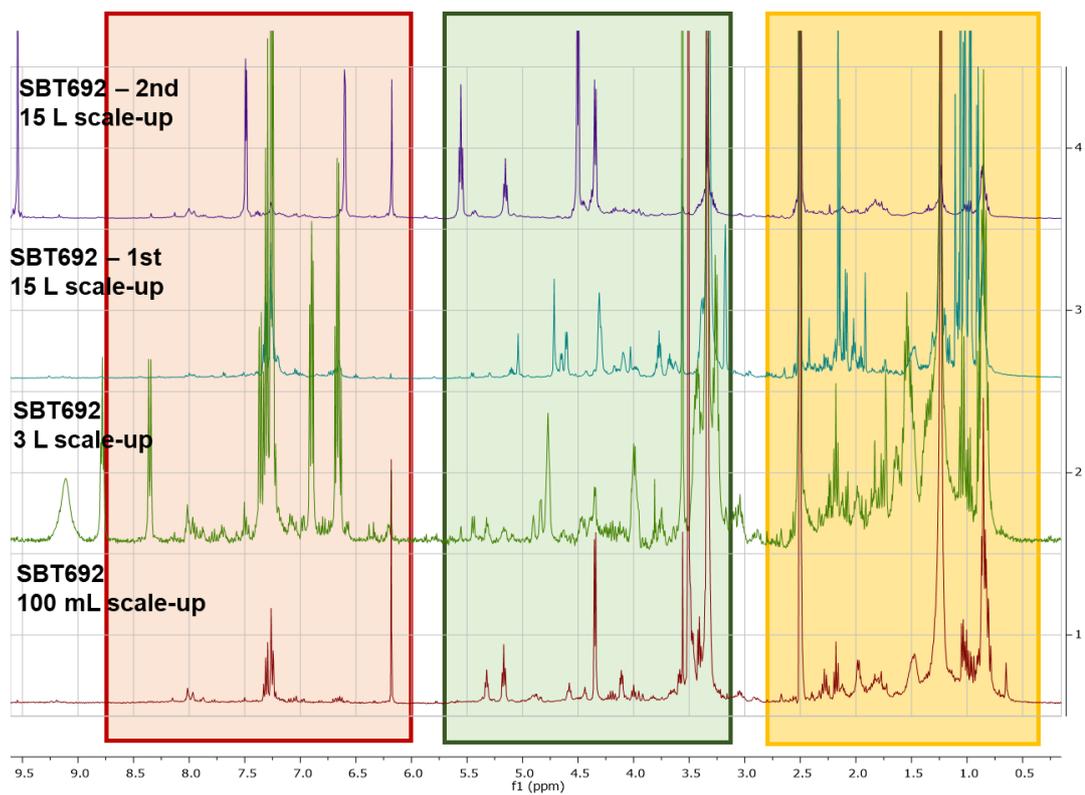


**Figure 5-4 Summary TLC of the fractionation of crude extract SBT692 obtained from the second 15L scale-up with ACN:Water 1:1 solvent system on RP18 plates.** Brown spots were visualized after spraying with anisaldehyde reagent. As observed under short and long UV wavelength, the spots were marked S (Short), LB (Long Blue) LG (Long Green), LV (Long Violet), LO (Orange).

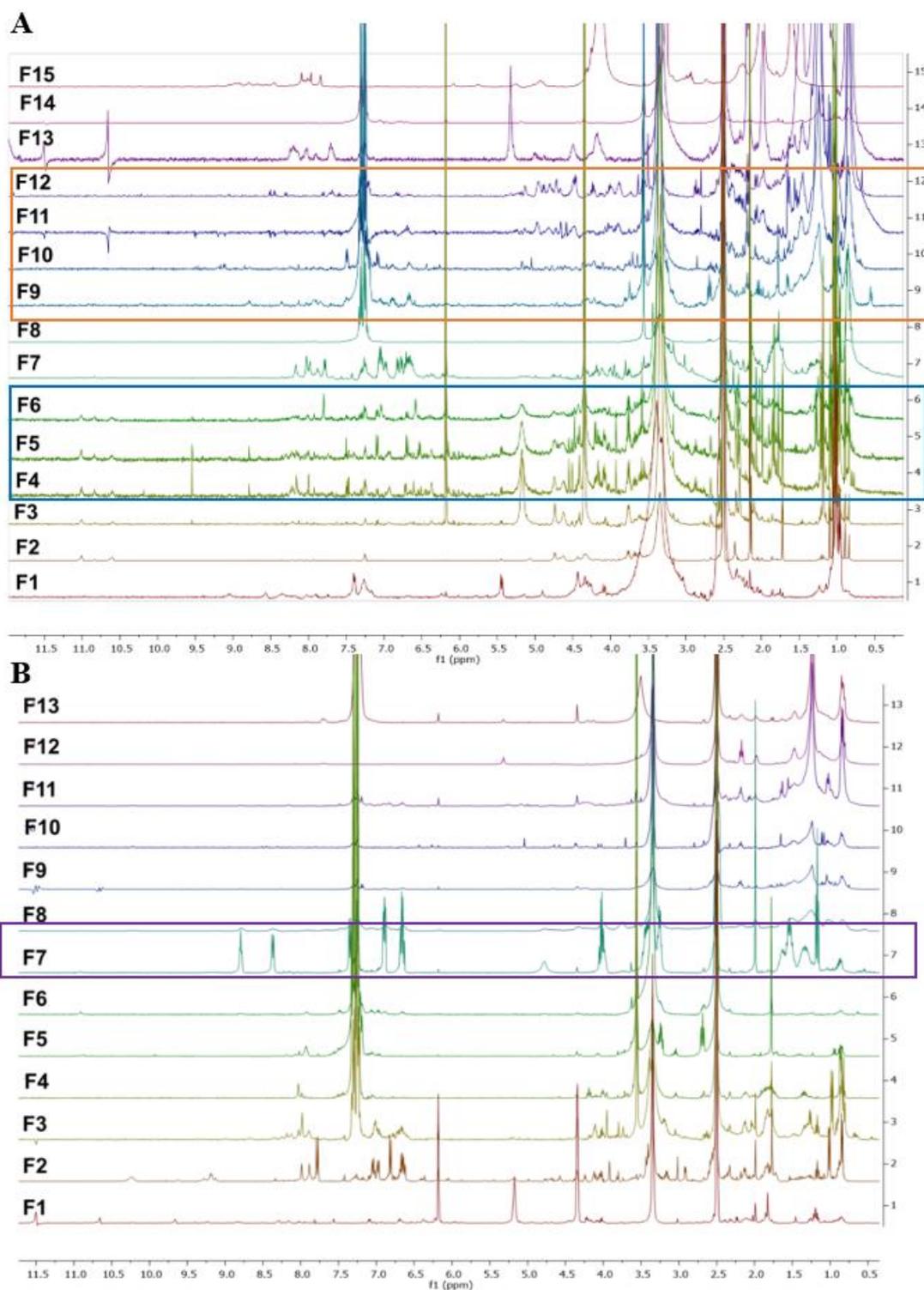
### 5.2.2.1 $^1\text{H}$ NMR spectra of SBT692 fractions

Crude extracts from the initial 100 mL fermentation, 3L scale-up, the first 15L scale-up and second 15L scale-up and fractions thereof were analysed by  $^1\text{H}$  NMR spectroscopy (Figure 5.5 and 5.6) along with LC-high resolution mass spectrometry to provide a comprehensive chemical profile of the metabolites being produced. Analysis of the stacked  $^1\text{H}$  NMR spectra of the crude extracts (Figure 5.5) showed that the chemical profile changed from the 100 mL to the 3L scale up then to the two batches of 15L scale-up. The small scale-up to 3 L exhibited a richer spectrum with crowded and intense peaks in the aliphatic, olefinic and aromatic region. This chemical richness decreased to the following batches of fermentation scale-ups with less peaks and intensities. From the 100 mL to the 3L scale up, there was an increased in resonances in the aromatic region between 6 to 8.5 ppm.

Analysis of the stacked  $^1\text{H}$  NMR spectra of the fractions showed that the compounds being produced were different between batches. Overall, the stacked NMR data exhibited signals in the aliphatic (0 – 2.5 ppm), olefinic (3.5 – 5.5 ppm) and aromatic (6 – 8.5 ppm) regions with some of the regions exhibiting more crowded and overlapping peaks. Fractions from the first fermentation, fractions 4 to 6 (blue box) and 9 to 12 (orange box) showed similar resonances while the others shared lower diversity and intensity of signals (Figure 5.6A). The spectra of F4 to F6 were populated both by aliphatic as well as aromatic resonances. Aliphatic peaks were resonating as complex, overlapped signals with the possible presence of electronegative atoms such as O, N, C=C, and C=O attached to the alkyl group (CH) promoting a deshielding effect. Additionally, the presence of signals that resemble the  $\alpha$  protons of amino acids were detected from 4 to 5.50 ppm. The spectra of F9 to F12 displayed, besides the aliphatic and olefinic signals, aromatic resonances which could be substituted by withdrawing groups that shifted the proton of benzene ring downfield to 7.5 ppm. Moreover, the fractions from the second fermentation exhibited lower diversity and less intense signals compared to those from the first fermentation, with F7 yielding the most unique spectrum. This fraction exhibited signals from fatty acids at 1 to 1.50 ppm characteristic of long chain methylene units and in the aromatic region with downfield signals up to 9 ppm (Figure 5.6B).



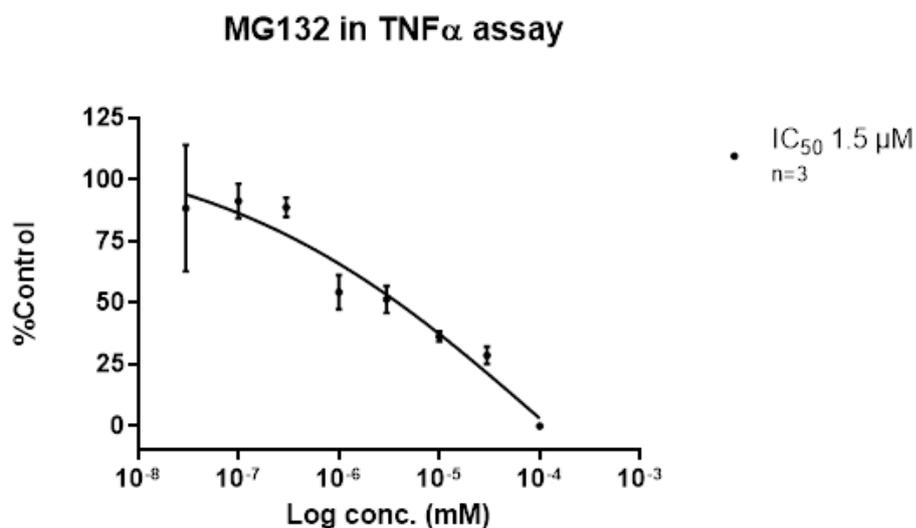
**Figure 5-5 Proton  $^1\text{H}$  NMR (400 MHz) spectra obtained for SBT92 crude extracts obtained from different scale-up fermentations in  $\text{DMSO-}d_6$ .**



**Figure 5-6 Proton  $^1\text{H}$  NMR (400 MHz) spectra obtained for SBT92 fractions.** Numbers on Y axis indicate the fraction number. Solvent used was  $\text{DMSO-}d_6$  (peak at 2.5 ppm). **A)** Fifteen SBT692 fractions from first fermentation and **B)** thirteen SBT692 fractions from second fermentation. Boxed spectra indicate similar and distinct fractions.

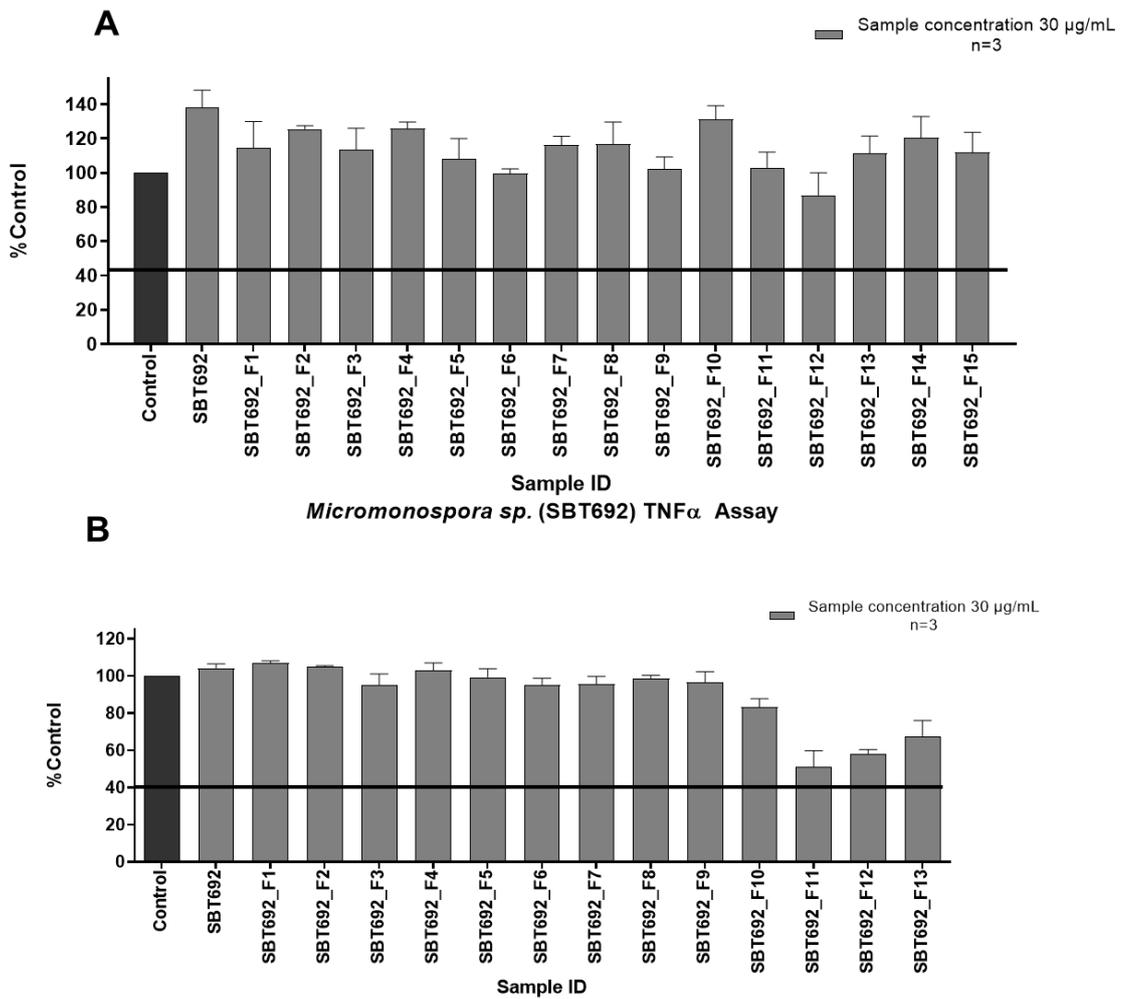
### 5.2.2.2 Biological activity of SBT692 crude extract and fractions

The NF $\kappa$ B Luciferase assay was performed in NCTC cell line, to determine the inhibition effect of the crude extract and fractions of both scale-up batches. MG132 inhibition control gave an IC<sub>50</sub> of 1.5 $\mu$ M (Figure 5.7). The SBT692 crude extract and fractions from both scale-up batches did not exhibit any inhibition effect against TNF-alpha (Figure 5.8). A slight inhibition effect of 40 to 50% was observed against TNF-alpha for F11 to F13 from the 2<sup>nd</sup> 15L scale-up.



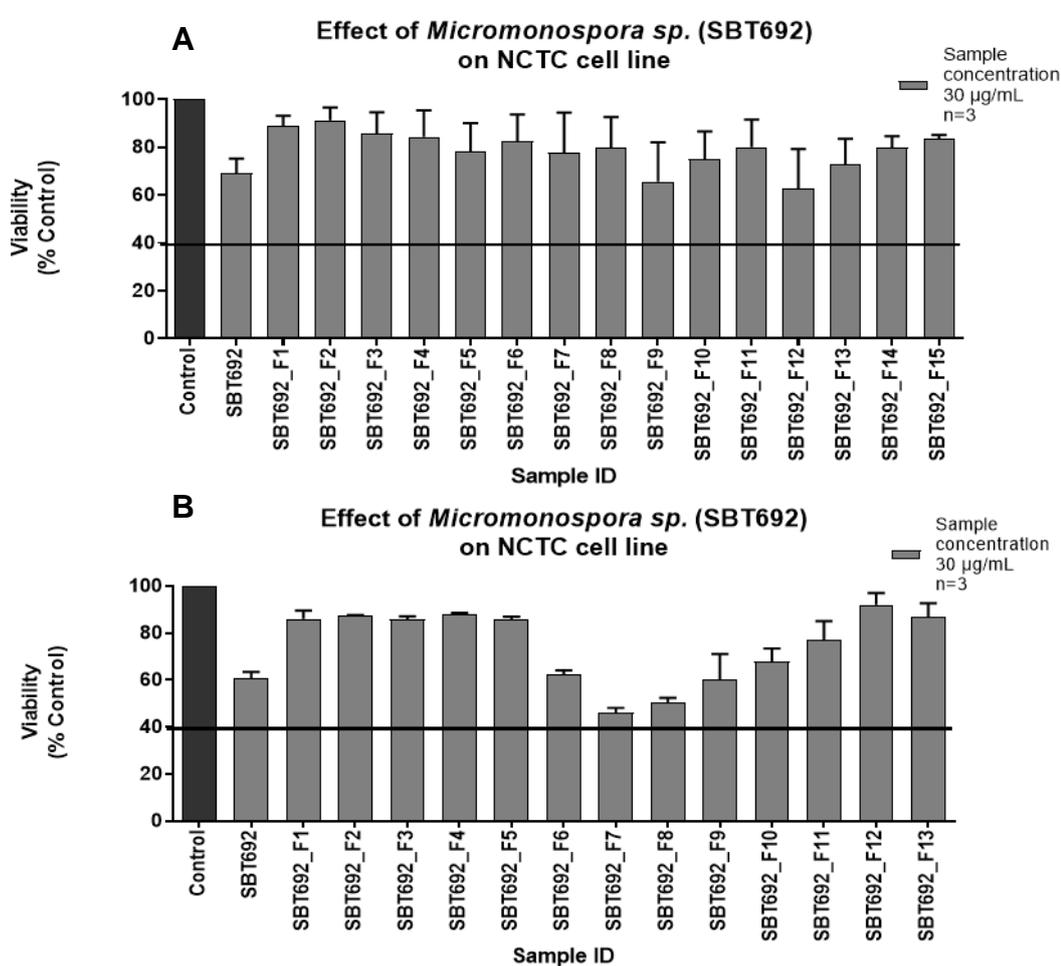
**Figure 5-7** MG132 inhibition control curve for the NF $\kappa$ B Luciferase assay done in triplicate (n=3) gave an IC<sub>50</sub> of 1.5 $\mu$ M.

**Micromonospora sp. (SBT692) TNF $\alpha$  Assay**

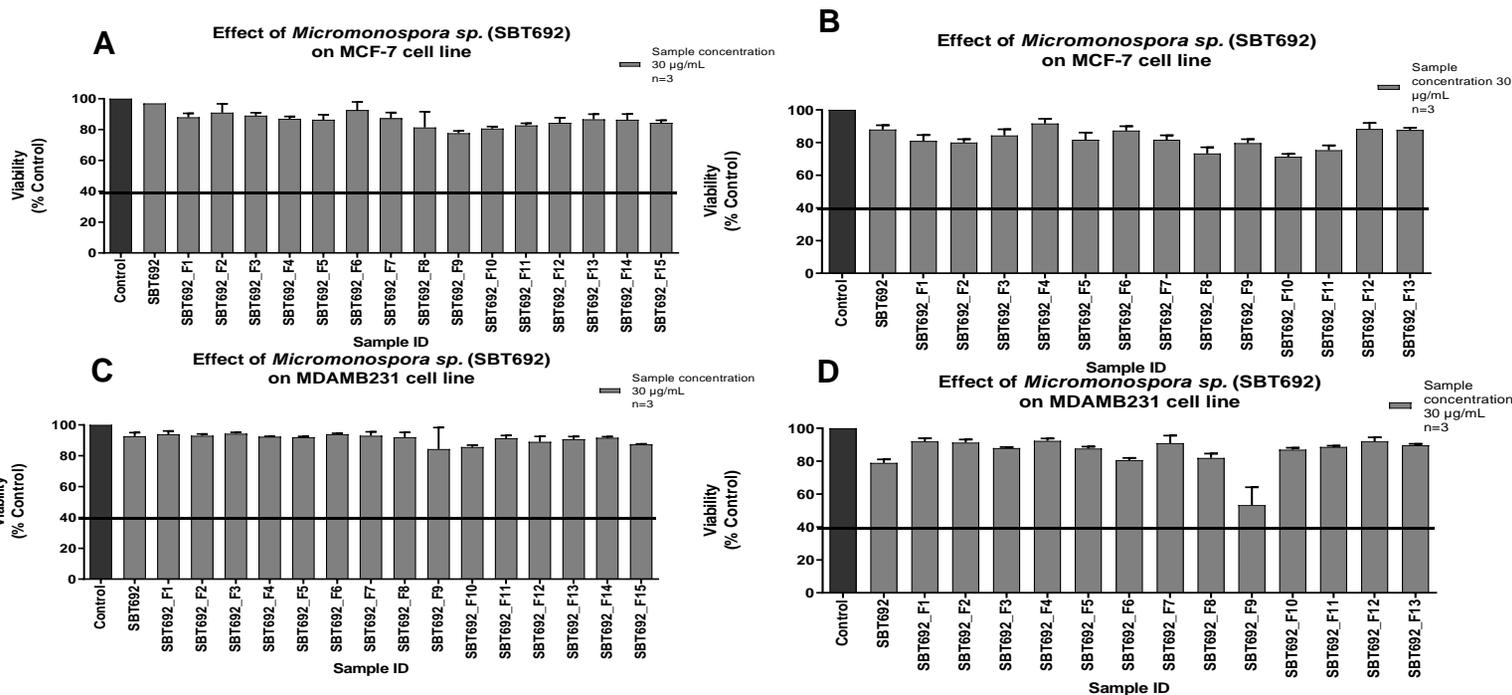


**Figure 5-8** The inhibition effect of SBT692 crude extracts and fractions from first (A) and second (B) 15L scale-ups against TNF alpha in the NFK $\beta$  Luciferase assay at 30 µg/mL tested in NCTC cell line. Error bars represent the standard deviations (SD) of three replicates (n=3) and ANOVA  $p < 0.05$  compared with the control. The black line at 40% indicates the inhibition threshold.

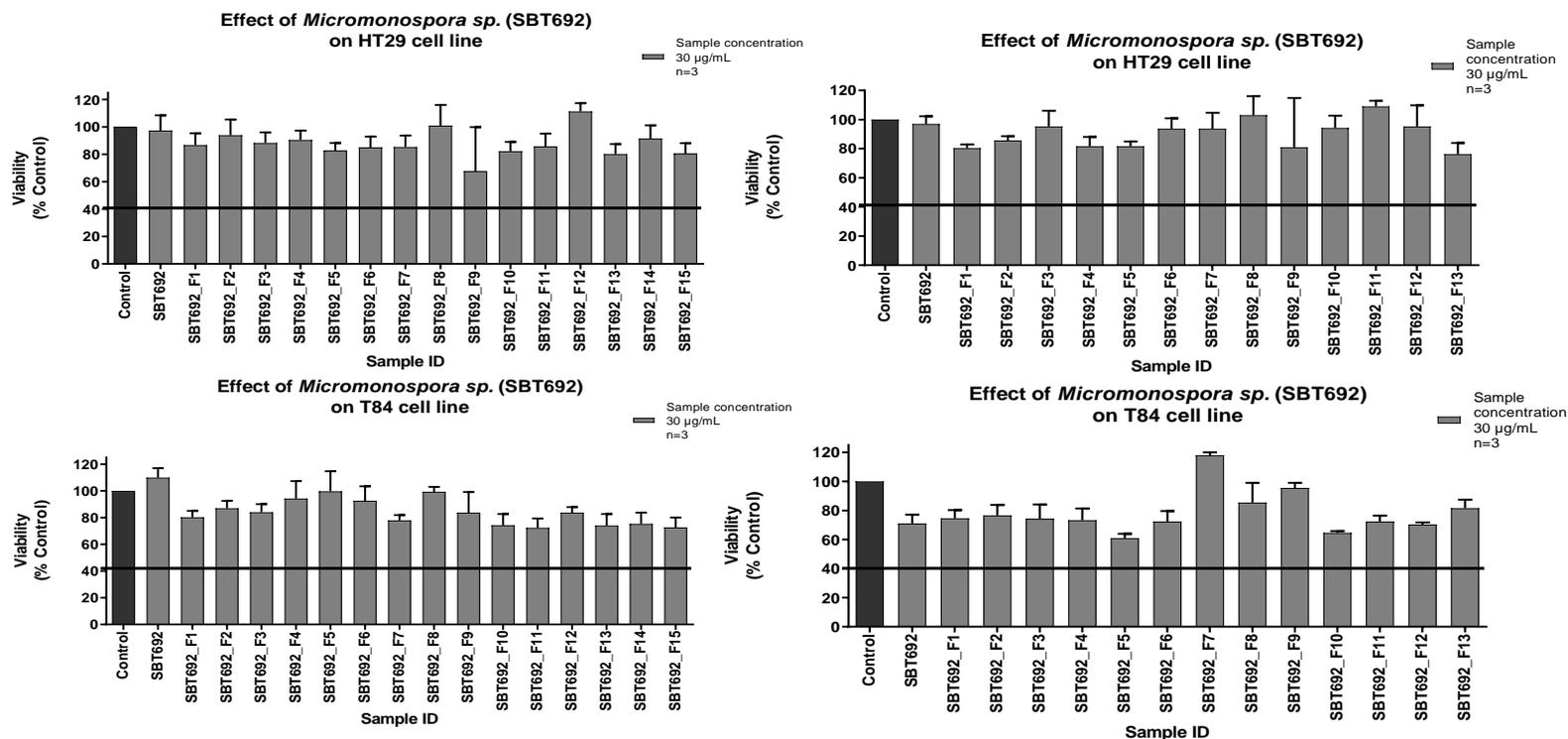
The crude extract and fractions from the first and second 15L scale-up were also assayed for cytotoxicity effect on NCTC cell line (Figure 5.9) and cell viability effect on MCF-7, MDA-MB-231, T84, HT29 and CaCo2 cancer cell lines (Figure 5.10, 5.11 and 5.12) while no cytotoxic effect on NCTC cell line or cell viability effect on the tested cancer cell lines were observed. However, weak cytotoxic activity at 40 to 50% activity against NCTC cell line was observed for F7 and F8 from the 2<sup>nd</sup> 15L scale-up as well.



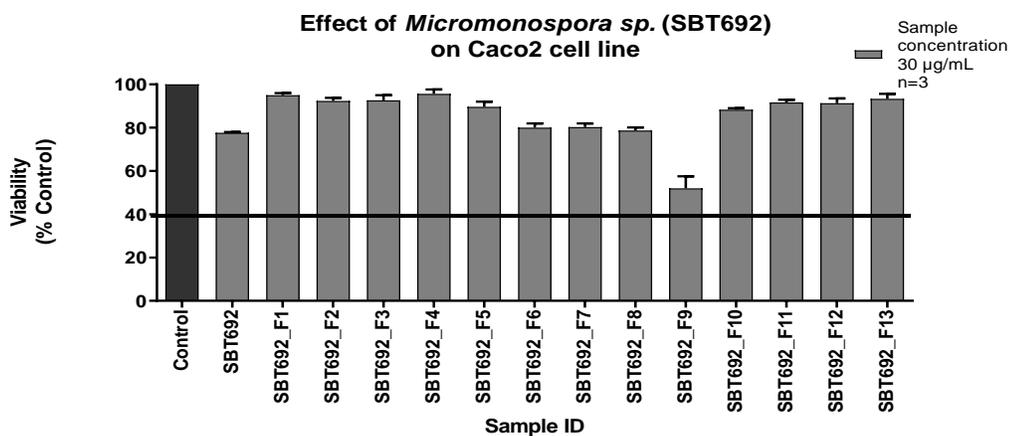
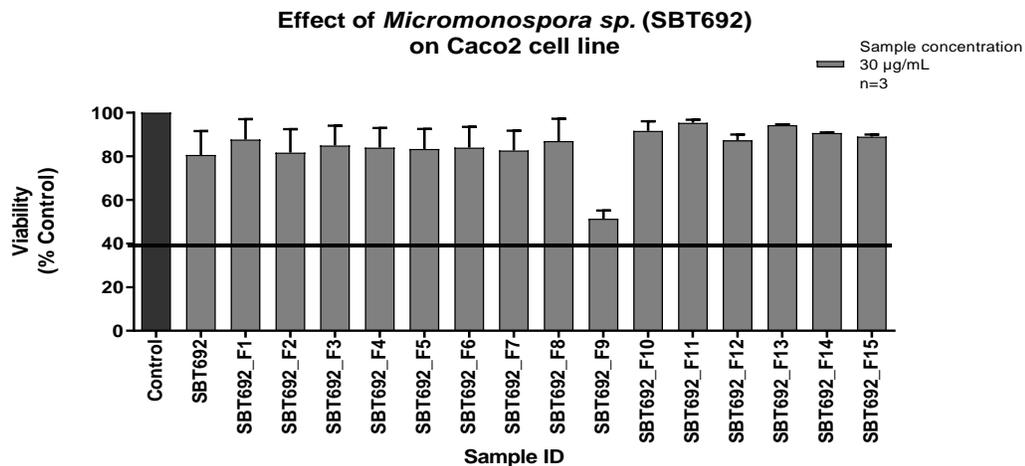
**Figure 5-9 Cytotoxicity effect of SBT692 crude extracts and fractions from first fractionation (A) and second fractionation (B) on NCTC cell line at 30  $\mu\text{g}/\text{mL}$ . Error bars represent the standard deviations (SD) of three replicates (n=3) and ANOVA  $p < 0.05$  compared with the control. The black line at 40% indicates the cytotoxicity threshold**



**Figure 5-10 MCF-7 cancer cell line viability screening** using SBT692 crude extracts and fractions on (A) in first fractionation and (B) second fractionation and on MDA-MB-231 cell line (C) in first fractionation and (D) second fractionation at 30 µg/mL. Error bars represent the standard deviations (SD) of three replicates (n=3) and ANOVA  $p < 0.05$  compared with the control. The black line at 40% indicates the cell viability threshold.



**Figure 5-11 HT29 cancer cell line viability screening** of SBT692 crude extracts and fractions on (A) in first fractionation and (B) second fractionation and on T84 cell line (C) in first fractionation and (D) second fractionation at 30  $\mu\text{g/mL}$ . Error bars represent the standard deviations (SD) of three replicates ( $n=3$ ) and ANOVA  $p<0.05$  compared with the control. The black line at 40% indicates the cell viability threshold.



**Figure 5-12 Cell viability screening of SBT692 crude extracts and fractions** from first fractionation (A) and second fractionation (B) on CaCo2 cell line at 30 µg/mL. Error bars represent the standard deviations (SD) of three replicates (n=3) and ANOVA  $p < 0.05$  compared with the control. The black line at 40% indicates the cell viability threshold.

## 5.3 Metabolomic screening of SBT692

### 5.3.1 Metabolomic profiling of the crude extract and fractions of SBT692 for both 15 L scale-up fermentations

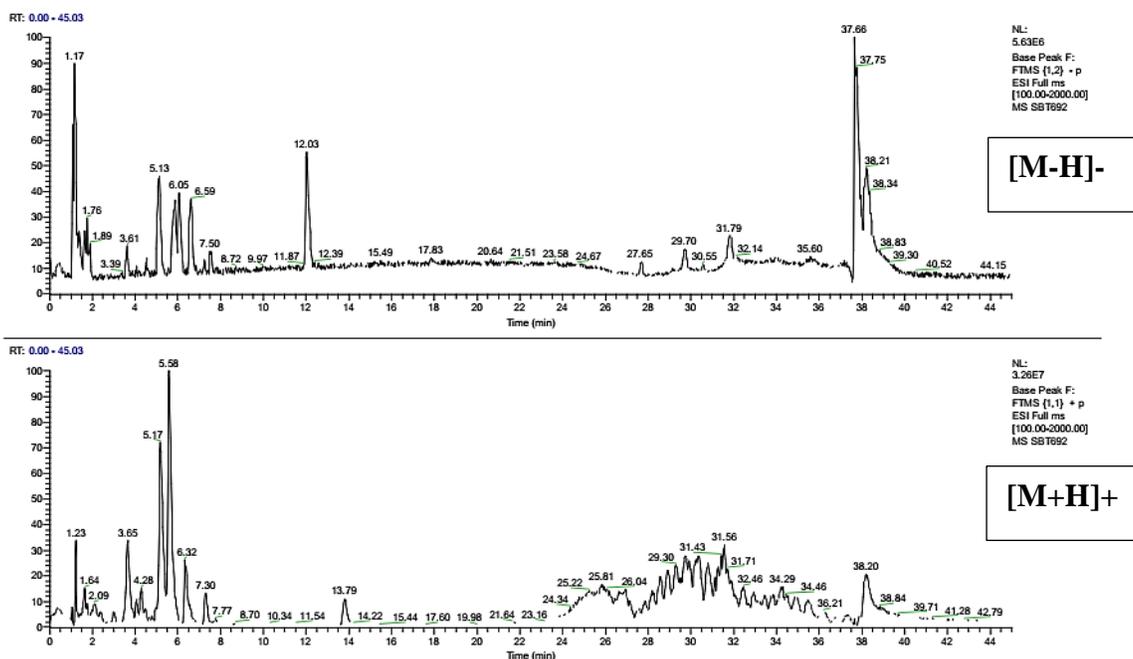
Isolation and purification work of any of the metabolites were not pursued for SBT692 due to loss of bioactivity during scale-up. However multivariate analysis of the production of metabolites were done to attempt to explain the changes occurring in the microbial strains during the scale-up. Due to a similar sponge source for SBT692 and SBT687 that were both taxonomically identified belonging to the genus *Micromonospora*, in this section, differences in metabolic production were also defined.

#### 5.3.1.1 HRMS data

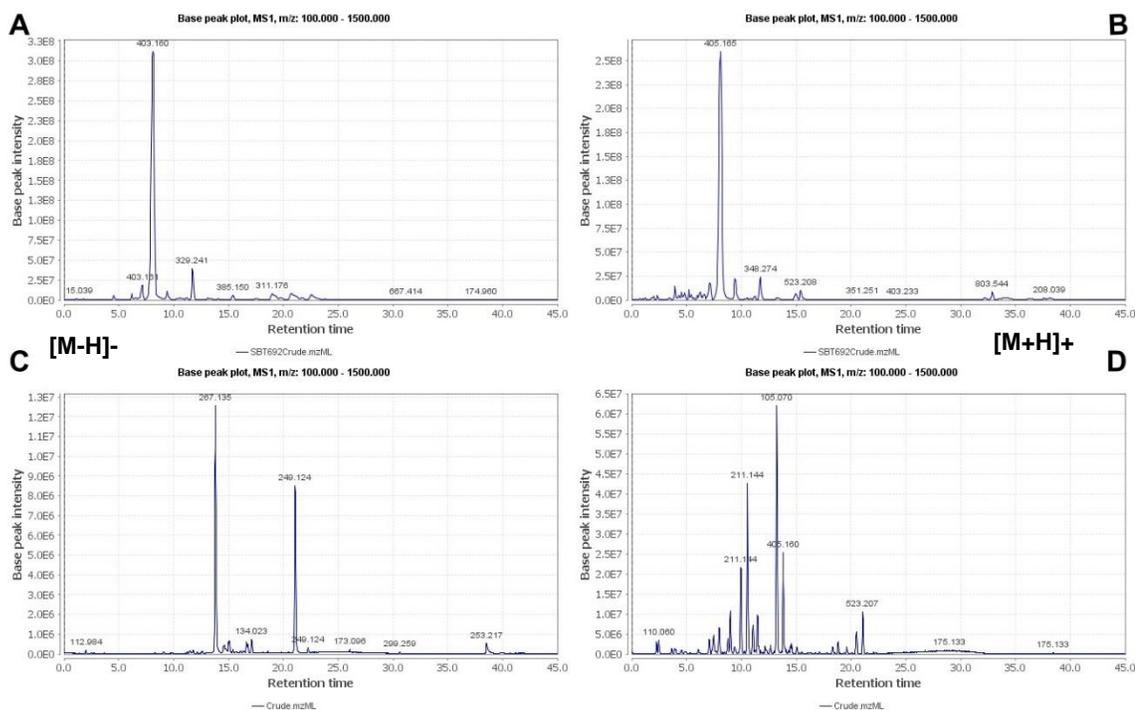
HRMS data was acquired using Xcalibur version 2.2 and was processed using MZMine. Base peak chromatograms were generated which allowed additional comprehensive profiling of the chemical composition and comparison of each SBT692 crude extract from the initial small-scale fermentation to the two batches of 15L scale-up. Base peak chromatograms for both ionization modes (Figure 5.13) of the “bioactive” SBT692 crude extract from the initial small-scale fermentation showed very polar major peaks from 0 to 12 minutes.

Regarding the base peak chromatograms of SBT692 crude extracts from the first 15L scale-up (Figure 5.14), main peaks between 5 to 10 minutes were observed (Figure 5.14A and B). For the second 15L scale-up, the chromatograms were not only different from the first batch but as well as between both ionisation modes (Figure 5.14C and D). The negative ionisation mode displayed increased peak numbers mainly between 10 to 20 minutes while in the positive ionisation mode, peaks were observed between 0 to 25 minutes. The increased in the number of peaks in the negative mode signified an increase

in production of acidic compounds like phenolics and fatty acids in the second fermentation batch. The chromatograms clearly illustrated the differences between the fermentation batches. From the analysis of the HPLC chromatograms, it was quite evident that metabolite production in SBT692 was not stable and reproducible while the bioactivity was lost during scale-up.

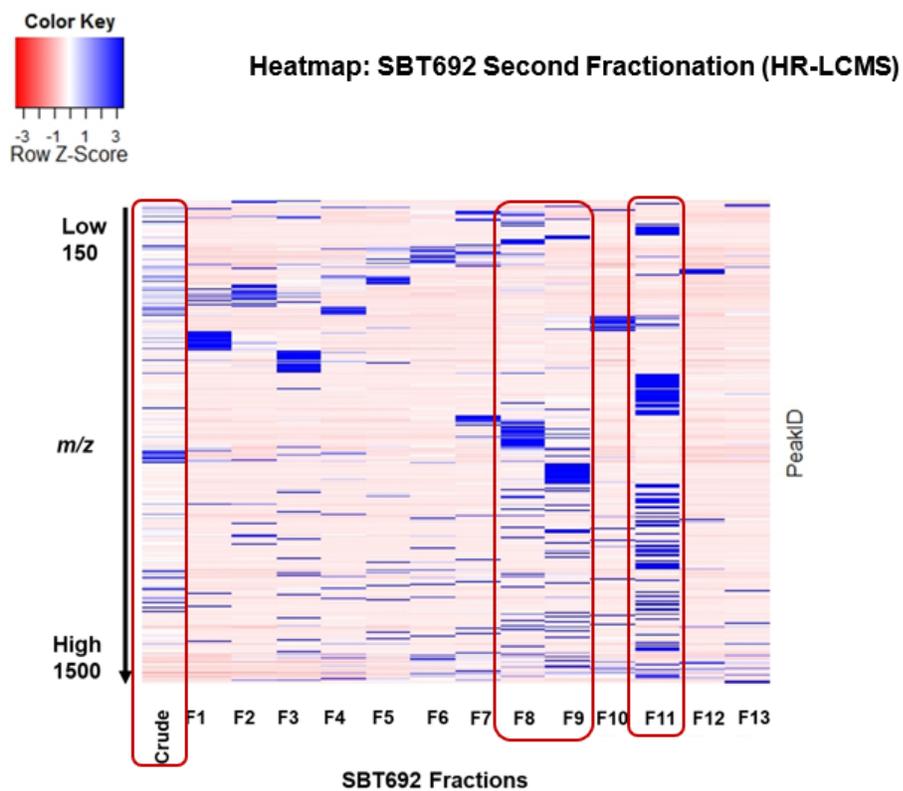
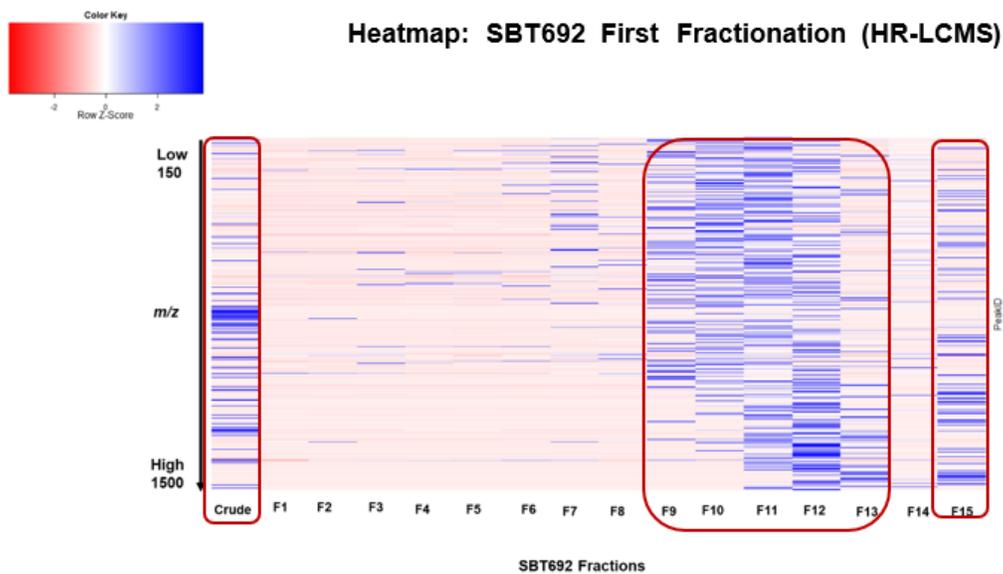


**Figure 5-13** LC-HRMS data of SBT692 from the initial small-scale fermentation and extracted ion chromatograms on negative and positive ionisation mode acquired from Xcalibur version 2.2.



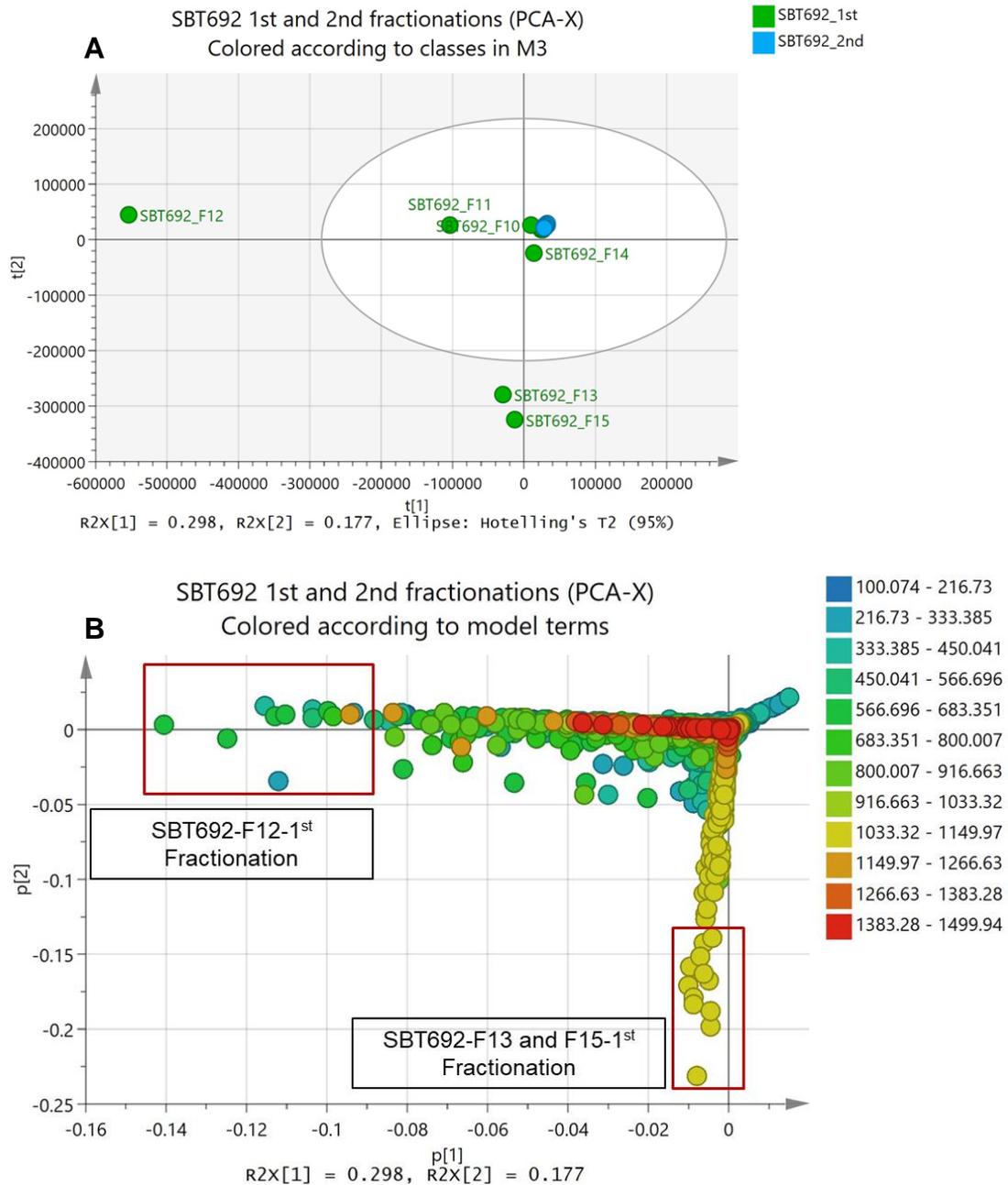
**Figure 5-14 Base peak chromatograms (HRMS data) of SBT692 crude extract of first 15L scale-up in negative (A) and positive (B) ionization mode and of SBT692 crude extract of second fractionation in negative (C) and positive (D) ionization mode created on MzMine 2.10 modified version.**

Heatmaps for the fractions from the first and second 15L scale-up were generated. The heatmap allowed the visualization and comparison of the chemical diversity of respective fractions, as well as to establish similarities and differences between the scale-up batches and between the fractions. The heatmap on Figure 5.15 showed the crude extracts and fractions from both 15L fermentation scale-ups producing different sets of metabolites. Fractions F9 to F13 and F15 from the first scale-up produced higher diversity of metabolites as evidenced by the greater number and intensity of blue bands, while on the second fermentation this was observed in fractions F8, F9 and F11.



**Figure 5-15** Heatmap based of HRMS data displaying distinct metabolic profiles of two batches of fermentations of SBT692. Boxed in red are the crude extracts and fractions showing higher chemical diversity.

PCA was performed on the HRMS data of the two batches of fermentations of SBT692. Using Pareto scaling, PCA plots were generated achieving a  $R^2X$  value of 1 and  $Q^2$  of 0.999, which indicated a well fitted model and good prediction of variables, respectively. The difference between  $R^2X$  and  $Q^2$  was less than 0.3 validating the model. As observed from the PCA scores scatter plot, the fractions from the 2<sup>nd</sup> 15L scale-up were all clustering together that indicated a strong similarity of their chemical profiles (Figure 5.16A). Fractions F12, F13 and F15 from the 1<sup>st</sup> 15L scale-up were outlying while its remaining fractions were clustered together nearer to the group from the 2<sup>nd</sup> scale-up, except for F11 and F14. The findings from the PCA score plot support the information provided by the HPLC chromatograms concerning the differences between the two fermentation batches. Moreover, the loadings scatter plot denoted the molecular ion peaks of the detected secondary metabolites of samples found on the corresponding quadrant position on the scores plot being possible to visualize the discriminating metabolites for F12, F13 and F15, which were highlighted in red boxes in Figure 5.16B. As shown by the distribution of features between the outliers and the group in the middle of the axis, the type of compounds afforded by the respective fractions can be predicted. From the 1<sup>st</sup> 15L scale-up; F12 yielded mid-MW compounds within the 300 to 600 amu range while F13 and F15 exceptionally afforded a higher density of higher MW metabolites with mass ion peaks at  $m/z$  1000 to 1150 amu. The remaining fractions obtained from both 15L scale-up afforded low MW compounds between 100 to 300 amu as well as metabolites with higher MWs greater than 1200 amu.



**Figure 5-16 PCA metabolomic screening of SBT692.** (A) PCA-X scores plot of SBT692 crude extract and fractions from first and second fermentations. The green label is for the first fermentation while the blue stands for the second fermentation of SBT692. (B) Loadings plot acquired from the PCA-X model with metabolites from the outlier's fractions highlighted in red boxes.

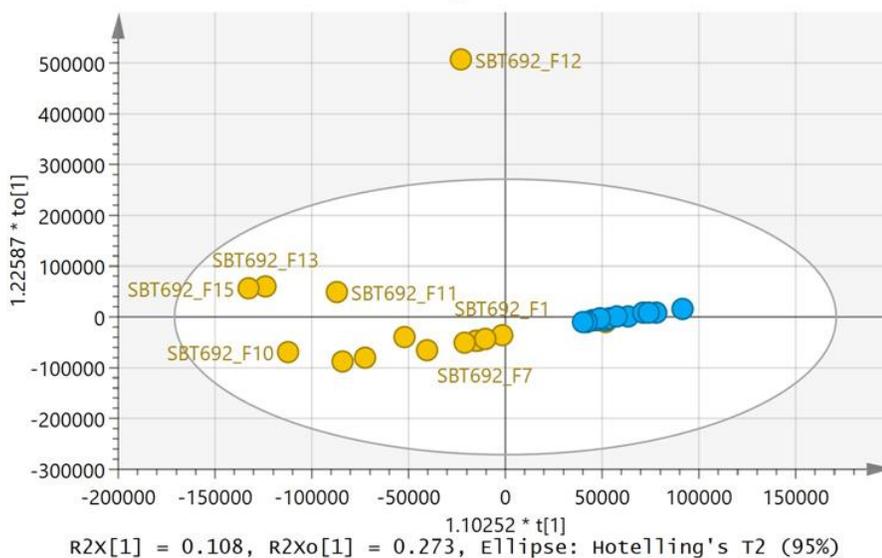
To detect the metabolites contributing for the differences between the two fermentation batches for SBT692, an OPLS-DA model was used (Figure 5.17A) achieving a  $R^2Y$  and  $Q^2$  of 0.648 and 0.534, respectively. Moreover, the difference between  $R^2Y$  and  $Q^2$  was less than 0.3 confirming that the model was not over fitted. The  $R^2X(1)$  was 0.108 which means that the variation between groups is 10.8% while  $R^2X_0(1)$  was 0.273 indicating a variation within groups of 27.3%. The greater variation within the group indicated a higher diversity in chemical profiles between fractions, which was particularly observed for the first 15L scale-up. F12 from the first 15L scale-up was again the outlier while the fractions from the second fermentation batch clustered very tightly close together. Furthermore, to identify the metabolites with the greatest input on the segregation between the fermentation batches, a loadings S-plot was generated (Figure 5.17B). The  $m/z$  features were dereplicated and cross-matched with the in-house database to putatively identify the compounds differentiating the two fermentation batches for SBT692. The loadings S-plot revealed that compounds with MWs between 200 to 450 and 1000 to 1150 amu were unique to the first batch while the second fermentation batch produced a higher density of low MW compounds at  $m/z$  100 to 450 amu (Table 5.10). The higher MW compounds ranging from 1000 to 1150 amu were highly oxygenated as well. The permutation test gave a  $Q^2Y$  intercept at -0.11 which was  $<0$ , further validating the OPLS-DA model (Figure 5.17C).

SBT692 fractions from 1st and 2nd fermentations (OPLS-DA)

**A**

Scaled proportionally to R2X  
Colored according to classes in M6

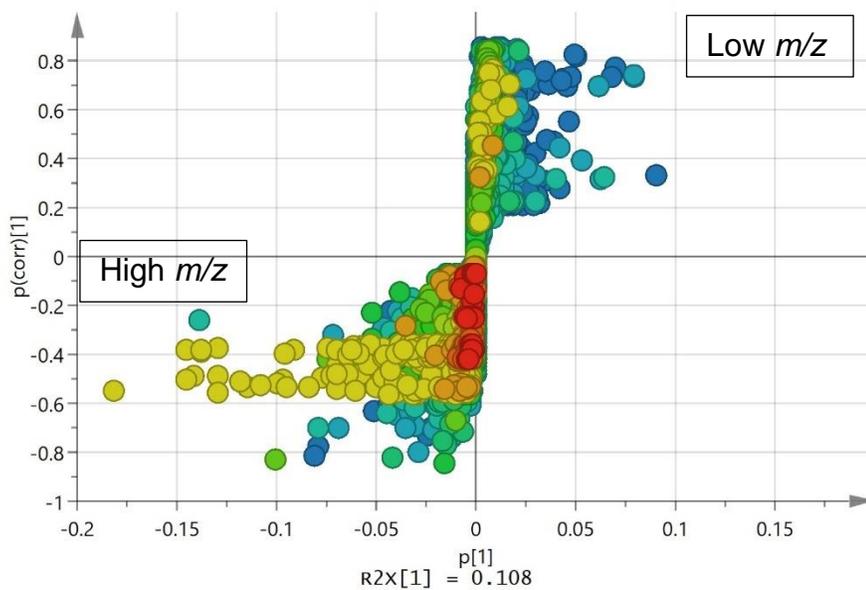
1stFractionation  
2ndFractionation

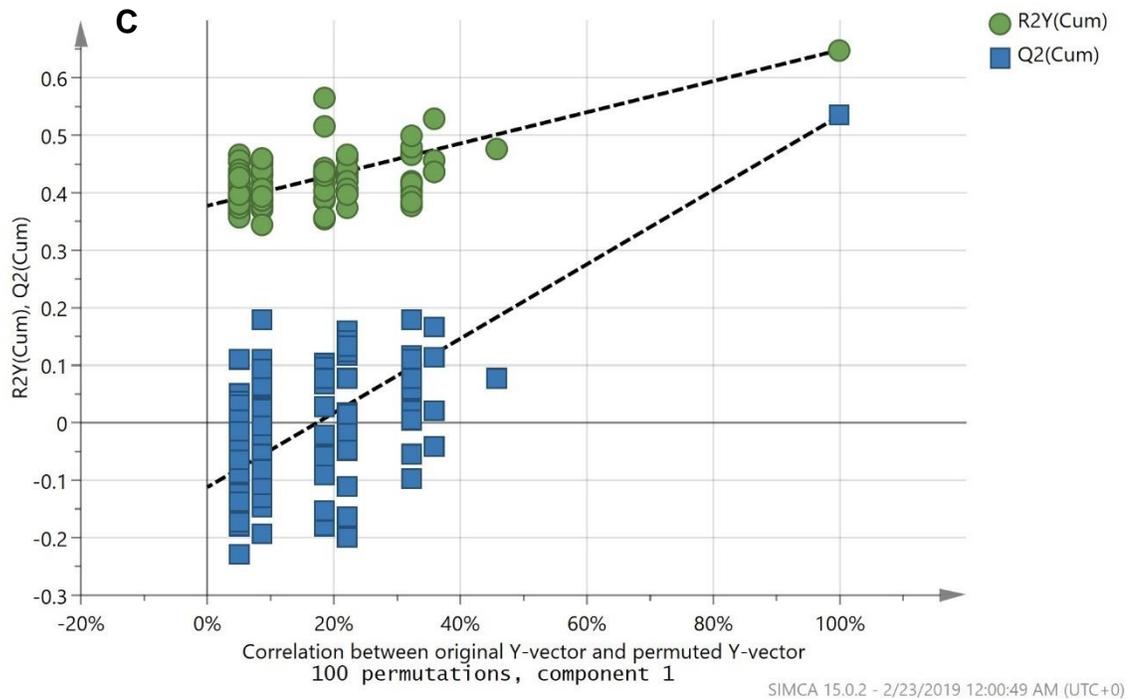


**B**

SBT692 Fractionations (OPLS-DA)  
Colored according to values in DS1.Var ID (m/z)

100.074 - 216.73  
216.73 - 333.385  
333.385 - 450.041  
450.041 - 566.696  
566.696 - 683.351  
683.351 - 800.007  
800.007 - 916.663  
916.663 - 1033.32  
1033.32 - 1149.97  
1149.97 - 1266.63  
1266.63 - 1383.28  
1383.28 - 1499.94





**Figure 5-17 OPLS-DA metabolomic screening of SBT692.** (A) OPLS-DA scores and (B) corresponding loadings S-plot plot of the mass spectral data for SBT692 fractions from the first and second fermentation batches. (C) Permutation tests (100 permutations) to validate the OPLS-DA model of SBT692 differentiating the mass spectral datasets between two fermentation batches.

**Table 5-10 Dereplication table of the ighlighted end point metabolites from SBT692 first (1<sup>st</sup>) and second (2<sup>nd</sup>) fermentation outliers fractions visualized in the loadings scatter plot from the PCA and S-plot endpoints of the OPLS-DA analysis.**

MzMineI D	m/z	Rt (min)	Peak Area (highest)	MWt	Molecular Formula	Name	Source
<b>A) First Fermentation Batch</b>							
P_170	315.1342	5.75	2.55E+05 (1 <sup>st</sup> _F15)	314.1267	C <sub>17</sub> H <sub>18</sub> N <sub>2</sub> O <sub>4</sub>	cribrostatin 5	Porifera <i>Cribrochalina vasculum</i>
P_188	257.2183	7.08	8.24E+08 (1 <sup>st</sup> _F12)	256.1211	C <sub>15</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub>	albonoursin	<i>Streptomyces albulus</i>
P_118	405.1656	8.36	1.77E+06 (1 <sup>st</sup> _F12)	404.1583	C <sub>20</sub> H <sub>24</sub> N <sub>2</sub> O <sub>7</sub>	myxochelin-A	<i>Angiococcus disciformis an d30 (myxobacterales)</i>
P_1770	263.1852	10.20	5.49E+08 (1 <sup>st</sup> _F12)	262.1780	C <sub>13</sub> H <sub>26</sub> O <sub>5</sub>	No hits	No hits
P_1602	219.1591	11.20	4.36E+08 (1 <sup>st</sup> _F12)	218.1518	C <sub>11</sub> H <sub>22</sub> O <sub>4</sub>	No hits	No hits
P_4713	427.2682	17.85	2.35E+08 (1 <sup>st</sup> _F12)	426.2609	C <sub>23</sub> H <sub>38</sub> O <sub>7</sub>	19,23- dihydroxyprotylonolide	<i>Aspergillus sp</i>
P_3020	405.2846	17.87	4.43E+08 (1 <sup>st</sup> _F12)	404.2773	C <sub>21</sub> H <sub>40</sub> O <sub>7</sub>	No hits	No hits
P_4705	387.2738	17.99	6.68E+08 (1 <sup>st</sup> _F12)	386.2666	C <sub>21</sub> H <sub>38</sub> O <sub>6</sub>	6-deoxyerythronolide B	<i>Streptomyces erythreus</i>
P_6748	422.3068	18.43	4.49E+08 (1 <sup>st</sup> _F12)	421.2996	C <sub>28</sub> H <sub>39</sub> NO <sub>2</sub>	No hits	No hits
P_4767	708.4900	20.57	6.42E+08 (1 <sup>st</sup> _F12)	707.4828	C <sub>38</sub> H <sub>62</sub> N <sub>9</sub> O <sub>4</sub> <sup>+</sup>	dinordehydro-batzelladine B	Porifera <i>Clathria calla</i> , Porifera <i>Monanchora arbuscula</i>

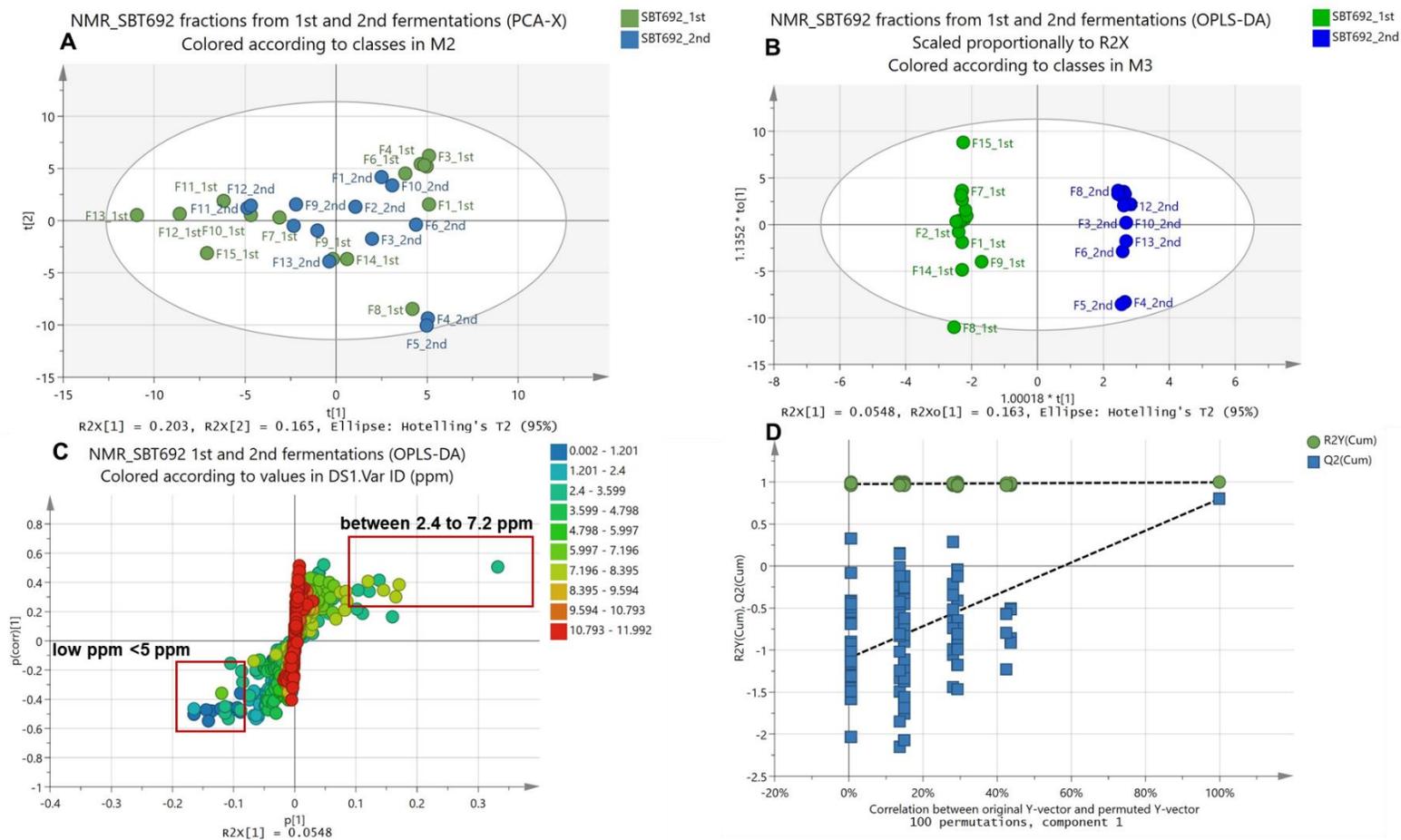
P_4864	664.4642	20.71	4.54E+06 (1 <sup>st</sup> _F12)	663.4569	C <sub>47</sub> H <sub>57</sub> N <sub>3</sub> / C <sub>36</sub> H <sub>57</sub> N <sub>9</sub> O <sub>3</sub> / C <sub>49</sub> H <sub>59</sub> O	No hits	No hits
P_5245	620.4379	20.85	4.59E+06 (1 <sup>st</sup> _F12)	619.4306	C <sub>30</sub> H <sub>49</sub> N <sub>15</sub> / C <sub>34</sub> H <sub>51</sub> N <sub>9</sub> O <sub>2</sub> / C <sub>32</sub> H <sub>61</sub> NO <sub>10</sub>	No hits	No hits
P_5984	576.1112	20.98	1.71E+07 (1 <sup>st</sup> _F12)	575.4039	C <sub>32</sub> H <sub>49</sub> N <sub>9</sub> O/ C <sub>31</sub> H <sub>53</sub> N <sub>5</sub> O <sub>5</sub> / C <sub>30</sub> H <sub>57</sub> NO <sub>9</sub>	No hits	No hits
P_6745	532.3845	21.09	1.62E+06 (1 <sup>st</sup> _F12)	531.3773	C <sub>30</sub> H <sub>45</sub> N <sub>9</sub> / C <sub>25</sub> H <sub>45</sub> N <sub>11</sub> O <sub>2</sub> / C <sub>28</sub> H <sub>53</sub> NO <sub>8</sub>	No hits	No hits
P_4722	617.4266	21.56	4.33E+08 (1 <sup>st</sup> _F12)	616.4193	unpredicted	No hits	No hits
P_5236	810.5582	22.25	4.58E+08 (1 <sup>st</sup> _F12)	809.5509	C <sub>44</sub> H <sub>67</sub> N <sub>13</sub> O <sub>2</sub> C <sub>43</sub> H <sub>71</sub> N <sub>9</sub> O <sub>6</sub> C <sub>38</sub> H <sub>71</sub> N <sub>11</sub> O <sub>8</sub>	No hits	No hits
P_6746	573.4008	22.97	9.44E+08 (1 <sup>st</sup> _F12)	572.4015	C <sub>30</sub> H <sub>56</sub> N <sub>2</sub> O <sub>8</sub>	N-Butylglucosamine ester derivative B	<i>Serratia sp.</i>
P_4749	1198.8410	23.03	4.66E+08 (1 <sup>st</sup> _F12)	1197.834	unpredicted	No hits	No hits
P_4714	613.3931	23.45	4.24E+08 (1 <sup>st</sup> _F12)	612.3858	C <sub>33</sub> H <sub>56</sub> O <sub>10</sub>	echinasteroside C	<i>Echinodermata</i> <i>Echinaster</i> <i>brasiliensis</i>
N_7255	1048.691	32.40	1.25E+08 (1 <sup>st</sup> _F15)	1049.698	C <sub>54</sub> H <sub>95</sub> N <sub>7</sub> O <sub>13</sub>	pumilacidin A	<i>Bacillus subtilis</i> , <i>marine Bacillus</i> <i>pumilus</i> KMM1364
P_9198	1078.739	35.65	1.75E+09 (1 <sup>st</sup> _F15)	1077.732	C <sub>56</sub> H <sub>99</sub> N <sub>7</sub> O <sub>13</sub>	pumilacidin C	No hits
N_8686	1076.723	36.90	2.70E+09 (1 <sup>st</sup> _F15)	1077.730	C <sub>55</sub> H <sub>103</sub> N <sub>3</sub> O <sub>17</sub>	primycin, primycin A1	<i>Streptomyces</i> <i>primycini</i> , <i>Micromonospora</i> <i>galeriensis</i>

N_12376	1090.736	37.24	2.43E+09 (1 <sup>st</sup> _F15)	1091.744	C <sub>56</sub> H <sub>105</sub> N <sub>3</sub> O <sub>17</sub>	primycin A2	<i>Thermopolyspora galeriensis</i>
<b>B) Second Fermentation Batch</b>							
P_919	211.144	9.98	1.34E+08 (2 <sup>nd</sup> _F4)	210.1367	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>2</sub>	(3R,7aR)- cyclo(leucylprolyl); maculosin-6; cyclo-D- Prolyl-D-leucyl	<i>Streptomyces rochei 87051-3, marine bacteria</i>
P_3343	269.1493	13.80	1.18E+08 (2 <sup>nd</sup> _F8)	268.142	C <sub>13</sub> H <sub>20</sub> N <sub>2</sub> O <sub>4</sub>	No hits	No hits
N_168	267.1351	13.82	7.35E+08 (2 <sup>nd</sup> _F8)	268.1424	C <sub>14</sub> H <sub>16</sub> N <sub>6</sub>	No hits	No hits
P_13012	175.1331	13.93	2.5E+06 (2 <sup>nd</sup> _F8)	174.1258	C <sub>19</sub> H <sub>18</sub> O <sub>3</sub>	No hits	No hits
P_2270	105.0702	14.80	8.09E+07 (2 <sup>nd</sup> _F7)	104.0630	C <sub>8</sub> H <sub>8</sub>	No hits	No hits
P_12185	117.0912	15.37	1.25E+06 (2 <sup>nd</sup> _F8)	116.0839	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	No hits	No hits
P_9116	115.0756	22.17	5.99E+07 (2 <sup>nd</sup> _F11)	114.0684	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	No hits	No hits

### 5.3.1.2 NMR Data

The  $^1\text{H}$  NMR spectral data was subjected to multivariate analysis similar to the mass spectral dataset. Using Pareto scaling, PCA achieved a  $R^2X$  value of 0.52, and a  $Q^2$  of 0.197, which specified a well fitted model ( $R^2X \geq 0.5$ ) but not so good prediction of variables with  $Q^2 < 0.5$ . The difference between  $R^2X$  and  $Q^2$  was 0.323, which should be less than 0.3 to confirm complete absence of overfitting. PCA-X scores plots (Figure 5.18A) of fractions from both fermentation batches SBT692 presented a mixed clustering pattern with some overlapping fractions between batches indicating similarity in molecular structures of compounds produced in both batches. On the other hand, the differences between the batches are more difficult to detect in NMR as only components greater than milligram levels can be detected.

Moreover, to distinguish the determinants of clustering variation on the PCA score plot, an OPLS-DA was performed (Figure 5.18B). The OPLS-DA model gave  $R^2Y$  and  $Q^2$  values of 0.996 and 0.8, respectively, achieving good fit and prediction devoid of overfitting with  $[R^2Y - Q^2] \leq 0.3$ . The variation between groups was 5.48% [ $R^2X(1)$ ] while the variation within groups was 16.3% [ $R^2X_0(1)$ ]. Therefore, the variation between the fractions within batches is higher than between fermentation batches. A loading S-plot (Figure 5.18C) was generated to visualize the variables responsible for the discrimination between groups. According to the end points of this plot, the fractions from the first batch contained structural features with chemical shifts mostly in the aliphatic and olefinic regions at less than 5 ppm, while the fractions from the second fractionation exhibited a mixture of signals distributed between 2.4 to 7.2 ppm with higher density of aromatic resonances. The model was validated by a permutation test (Figure 5.18D) with  $Q^2Y$  intercept at  $-1.09$ .



**Figure 5-18 MVDA analysis of NMR data of SBT92 fractions from two fermentation batches. (A) PCA-X, (B) OPLS-DA scores plot, and (C) corresponding loadings S-plot of SBT692 fractions from first and second fermentation batches. Features in the S-plot were coloured according to chemical shifts in ppm. (D) Permutation tests (100 permutations) to validate the OPLS-DA model of SBT692 differentiating the <sup>1</sup>H NMR spectral datasets between two fermentation batches**

## 5.3.2 Metabolite profile comparison between SBT692 and SBT687

### 5.3.2.1 HRMS DATA

PCA was performed on the HRMS datasets of two fermentation batches of SBT692 against SBT687 to further show the differences and/or similarities between the different bacterial strains. Using Pareto scaling, a  $R^2X$  value of 1 was achieved indicating good fit and  $Q^2$  of 0.998 which specified a good predictive ability of the model, with a difference between  $R^2X$  and  $Q^2$  at less than 0.3. Furthermore, the PCA scores plot exhibited similar outliers from the first fermentation batch of SBT692, as already shown in section 5.3.1.1, while SBT687 fractions overlapped with those of SBT692 from the second fermentation batch (Figure 5.19).

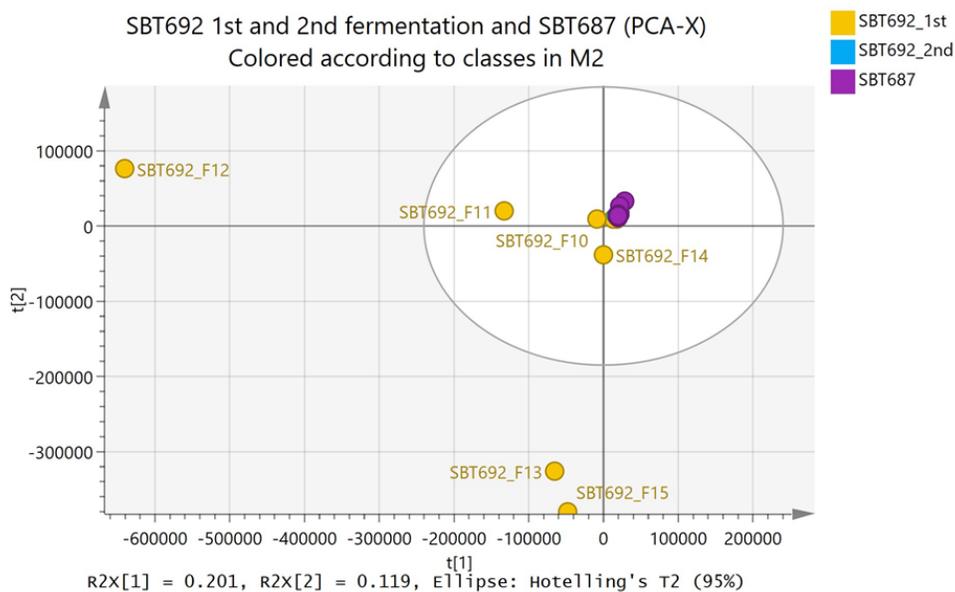
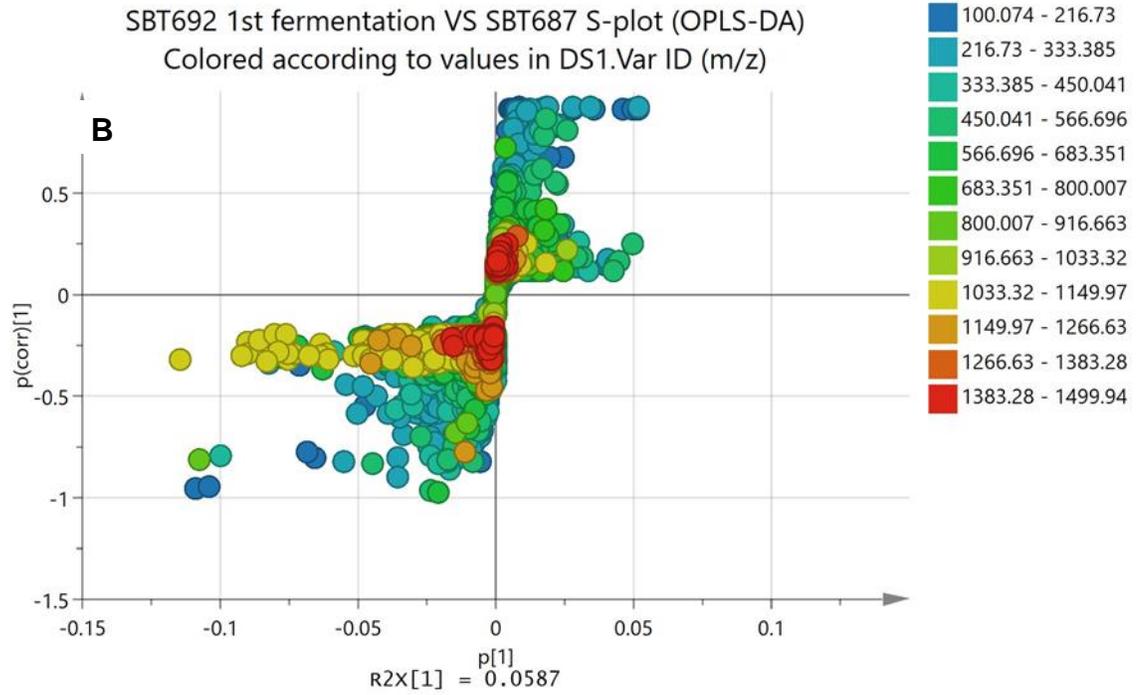
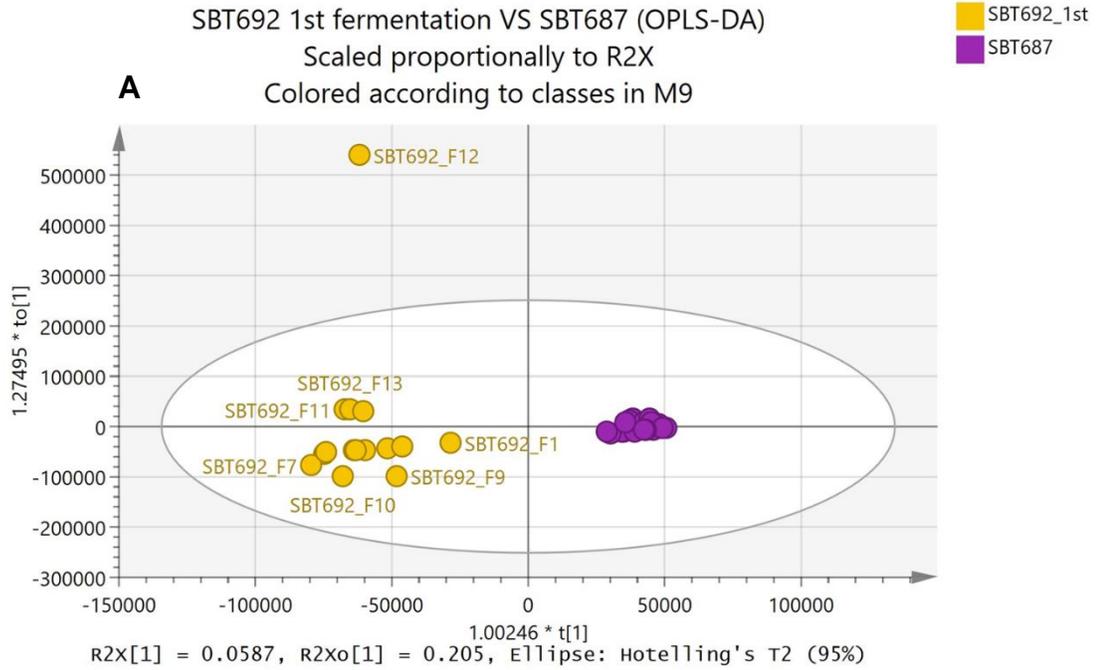
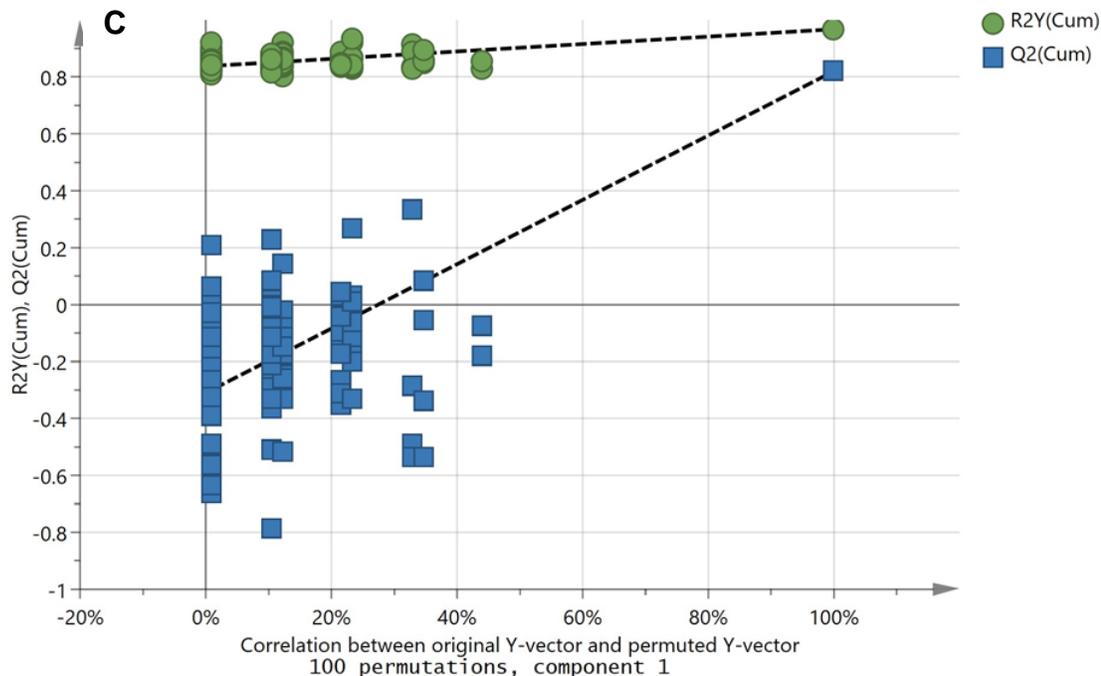


Figure 5-19 PCA-X scores plot of SBT687 and SBT692 fractions from first and second fermentation batches.

To predict the metabolites responsible for the differences between SBT687 fractions and the SBT692 fractions from first fermentation batch, an OPLS-DA model was generated (Figure 5.20A) achieving a  $R^2Y$  and  $Q^2$  of 0.967 and 0.82, respectively, indicating a model with both good fit and predictability. The difference between  $R^2Y$  and  $Q^2$  was less than 0.3 confirming that the model was not overfitted. Variations between groups and within group were exhibited at 5.87% [ $R^2X(1)$ ] and 20.5% [ $R^2X0(1)$ ], respectively. The low percent variation between groups demonstrated a small general difference in chemical profiles between strains. However, the presence of an outlier and the more dispersed distribution of the fractions from the first fermentation batch of SBT692 caused the percentage variation within group to be greater than the variation between the strains. Furthermore, to identify the metabolites with the greatest input to the segregation of SBT692 fractions from first fermentation batch versus those of SBT687, a loading S-plot was generated (Figure 5.20B). The end-point metabolites defined for the first fermentation batch of SBT 692 is obviously identical to those that differentiates it from second fermentation batch, which have been presented in the dereplication Table 5.10. The discriminating metabolites for SBT687 are the same bioactive features listed in Chapter 4, such as the manzamines, which is interesting as SBT687 is not only different from SBT692 simply because of the type of compounds it produces, but as well as differentiates its bioactivity from SBT692. The model was validated by a permutation test with  $Q^2Y$  intercept value of  $-0.31$  (Figure 5.20C).

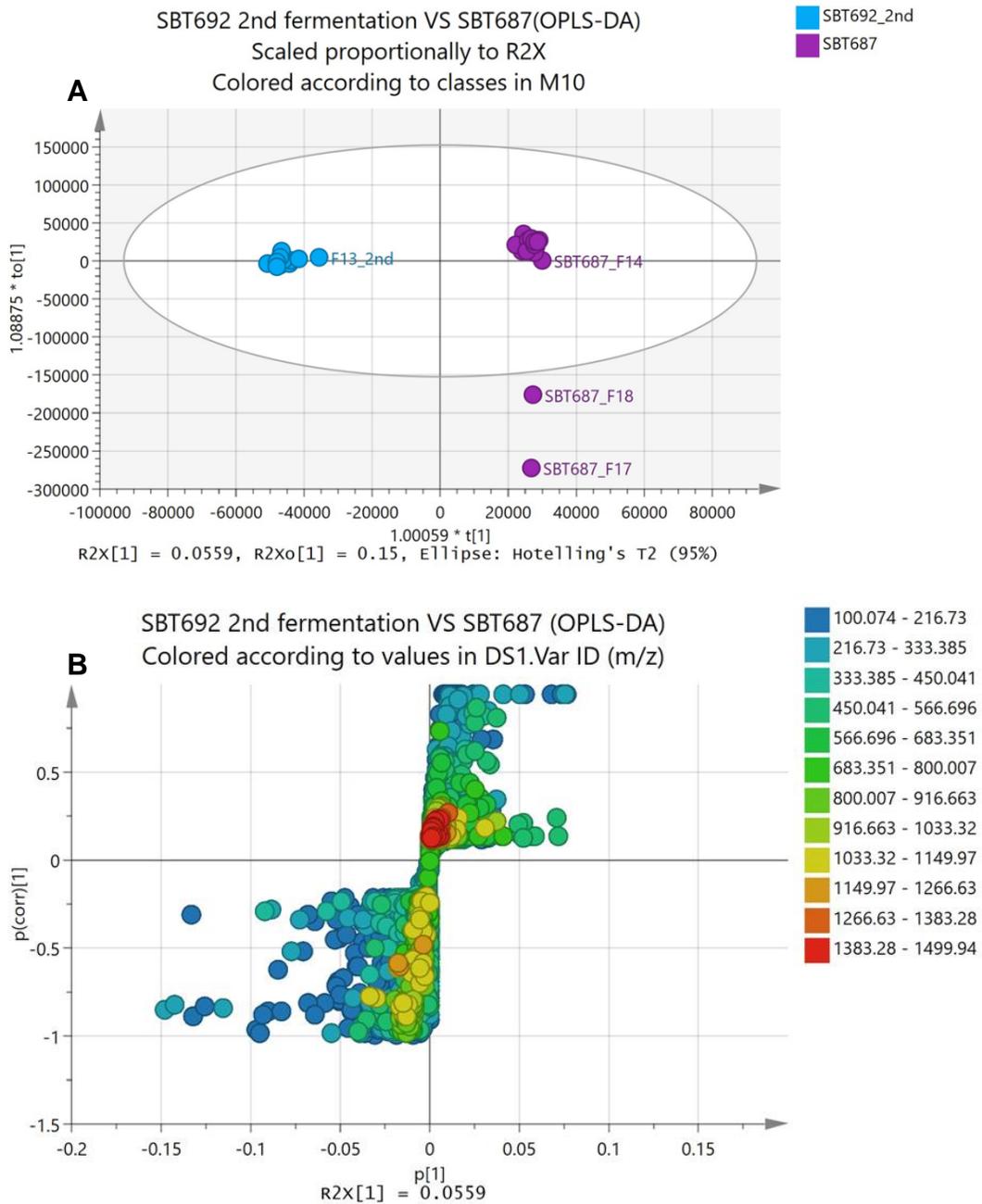


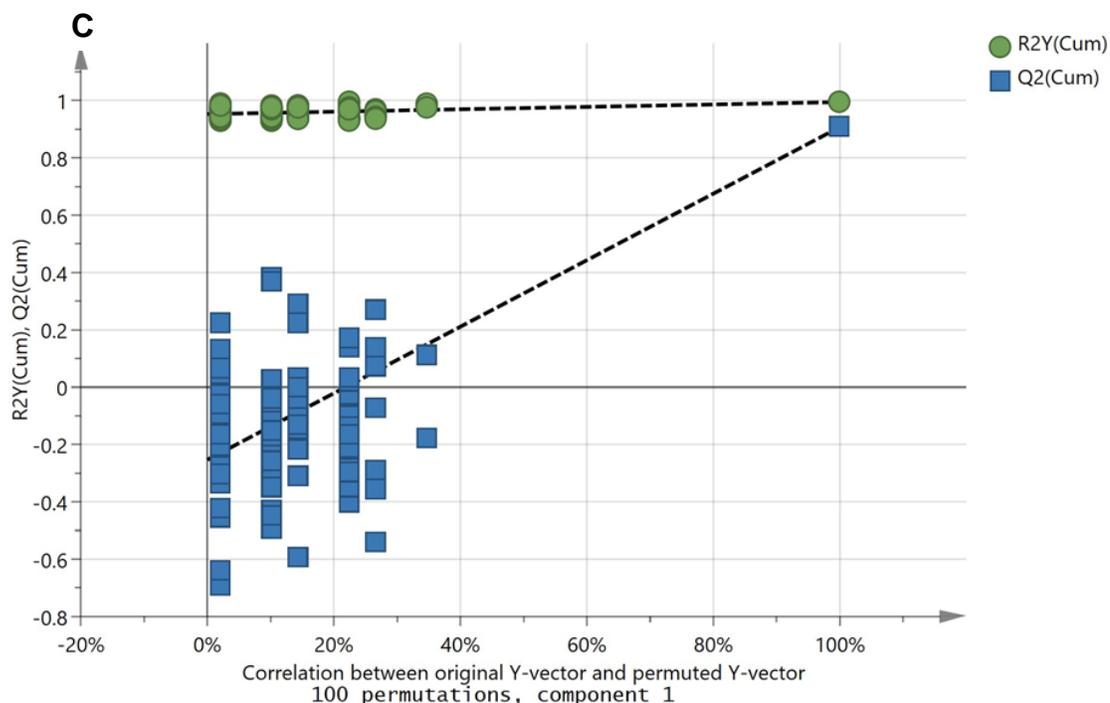


**Figure 5-20** MVDA analysis of mass spectral data of SBT92 fractions from first fermentation batch versus SBT687 fraction. (A) OPLS-DA scores plot and (B) corresponding loadings S-plot of SBT692 fractions from first fermentation batch versus SBT687 fractions. Features in the S-plot were coloured according to their  $m/z$  ranges. (C) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the mass spectral datasets between strains.

The HRMS dataset for SBT692 fractions from second fermentation batch was also compared with those of SBT687. An OPLS-DA model was generated (Figure 5.21A) achieving a  $R^2Y$  and  $Q^2$  values of 0.995 and 0.907, respectively, for a model with both good fit and predictability. The difference between  $R^2Y$  and  $Q^2$  was less than 0.3 indicating that the model was not overfitted. Variations between the two groups and within the groups were exhibited at 5.59% [ $R^2X(1)$ ] and 15% [ $R2X_0(1)$ ], respectively. Again, the variation between the two strains was lower than the diversity in chemical profiles between the fractions particularly those from SBT687. The percentage variation between strains was almost similar when SBT687 was compared to SBT692 fractions from the first fermentation batch. However, this time the higher variation within the groups was due to the outlying fractions F17 and F18 from SBT 687. The outlying fractions F17 and F18 of SBT687 were likewise the bioactive fractions described in Chapter 4. Furthermore, an S-

plot was generated (Figure 5.21B) and the end-point metabolites from SBT687 were the same as listed on Chapter 4. The permutation test gave a validation score with  $Q^2Y$  intercept value of -0.25 (Figure 5.21C).

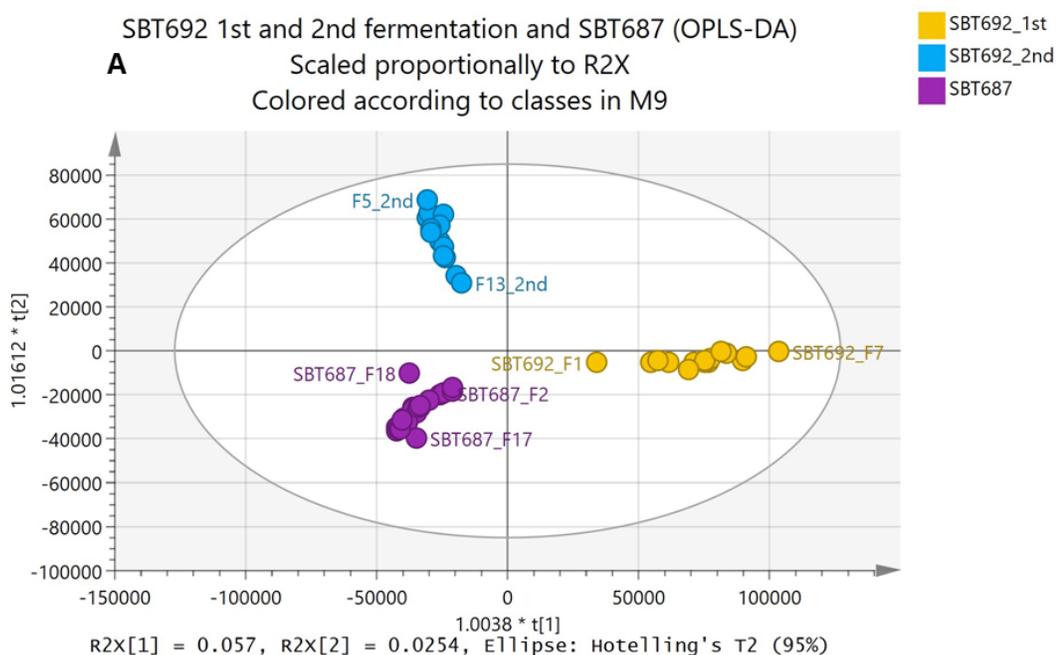


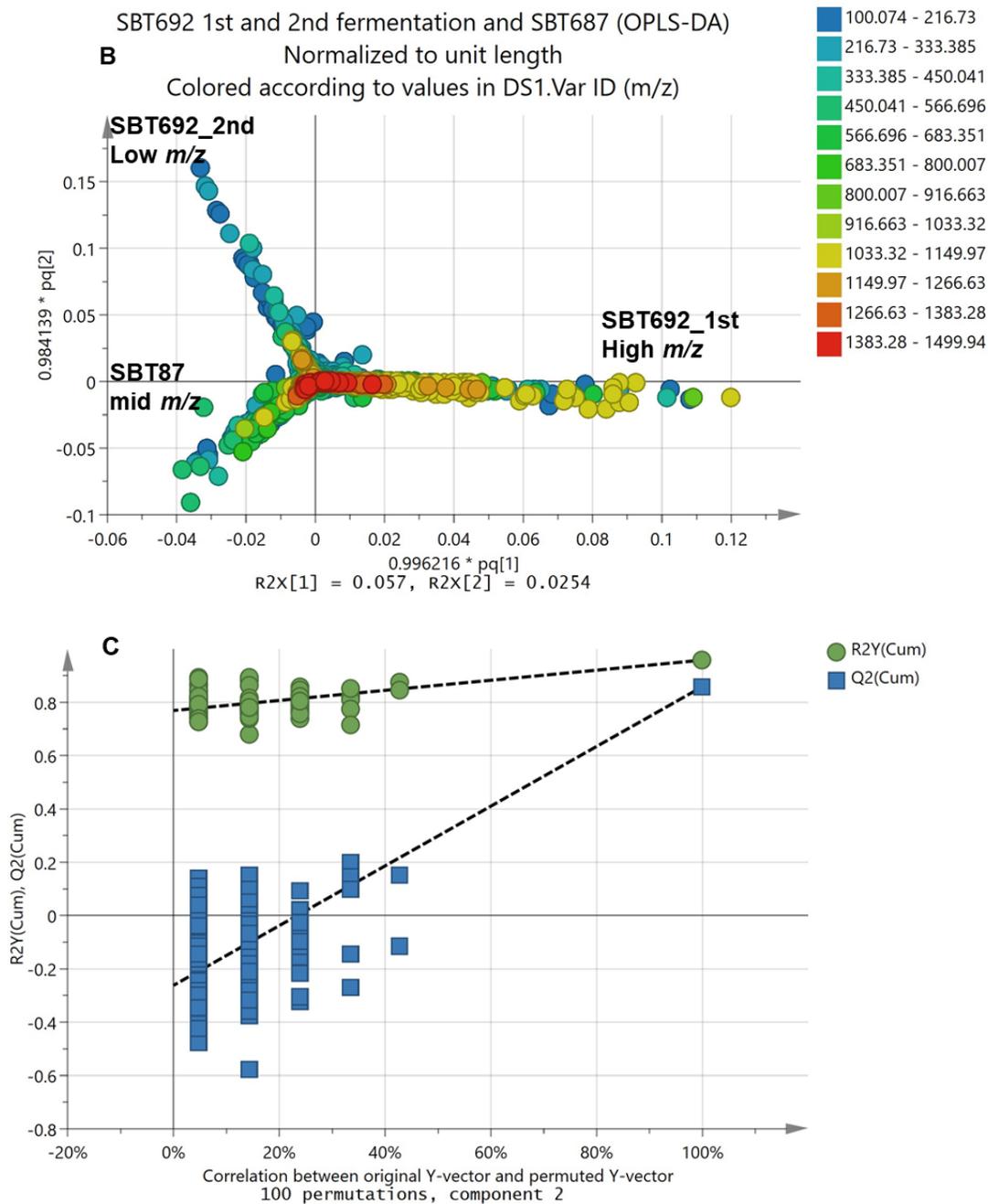


**Figure 5-21 MVDA analysis of mass spectral data of SBT92 fractions from second fermentation batch versus SBT687 fraction. (A) OPLS-DA scores plot and (B) corresponding loadings S-plot of SBT692 fractions from second fermentation batch versus SBT687 fractions. Features in the S-plot were coloured according to their  $m/z$  ranges. (C) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the mass spectral datasets between strains.**

To revalidate the OPLS-DA results between comparing just two groups at a time, the HRMS dataset for SBT692 fractions from both batches were also compared with those of SBT687. An OPLS-DA model was generated (Figure 5.22A) achieving a  $R^2Y$  and  $Q^2$  values of 0.952 and 0.841, respectively, for a model with both good fit and predictability. The difference between  $R^2Y$  and  $Q^2$  was less than 0.3 indicating that the model was not overfitted. The OPLS-DA demonstrated that SBT 687 remained distinct from both fermentation batches of SBT 692 as they were all positioned in three different quadrants. Variations between the groups and within the groups were exhibited at 5.70% [ $R^2X(1)$ ] and 2.54% [ $R^2X_0(1)$ ], respectively. The distinct separation between the strains and the

batches was reflected by the percent variation, which this time is 2x higher than the variation within groups, demonstrating greater chemical similarities within the groups without the occurrence of any outlying fractions. Although, the percentage variation between strains was similar to the previous analysis when only two groups were compared, the percentage variation within groups plummeted. However, after analysis of the loadings plot shown in Figure 5.22B it was possible to verify that the outlying fractions were still defined by the same sets of metabolites established by the previous S-plots. Compounds with mid-MW within the 300 to 600 amu were found in SBT687 fractions, compounds with mid- MWs between 200 to 450 and 1000 to 1150 amu were unique to the first batch while the second fermentation batch produced a higher density of low MW compounds at  $m/z$  100 to 450 amu. The permutation test confirmed validity of the model with a  $Q^2Y$  intercept value of -0.262 (Figure 5.22C).





**Figure 5-22** MVDA analysis of mass spectral data of SBT92 fractions from first and second fermentation batches versus SBT687 fractions. (A) OPLS-DA scores plot and (B) corresponding loadings S-plot of SBT692 fractions from first and second fermentation batches versus SBT687 fractions. Features in the S-plot were coloured according to their *m/z* ranges. (C) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the mass spectral datasets between strains.

### 5.3.2.2 NMR DATA

Multivariate analysis was also performed on the  $^1\text{H}$  NMR data of the fractions to detect the differences between the different fermentations and strains. PCA was performed on the  $^1\text{H}$  NMR data of SBT692 fifteen fractions from the first fermentation, thirteen fractions from second fractionation and twenty-one fractions from SBT687. The model applied Pareto scaling, achieving a  $R^2X$  value of 0.527, and a  $Q^2$  of 0.197, which specified a well fitted model at  $R^2X \geq 0.5$  for a metabolomics data set but prediction was not so good with low ( $Q^2 < 0.5$ ). The difference between  $R^2X$  and  $Q^2$  was 0.3, and the model was still not to be considered overfitted. The PCA scores plot (Figure 5.23) of fractions from both fermentation batches of SBT692 and those of SBT687 exhibited heterogeneous distribution among the four quadrants. Although most of the fractions of SBT 687 occurred in the quadrants between (-) PC1 and (-) PC2 or also described as the lower left quadrant.

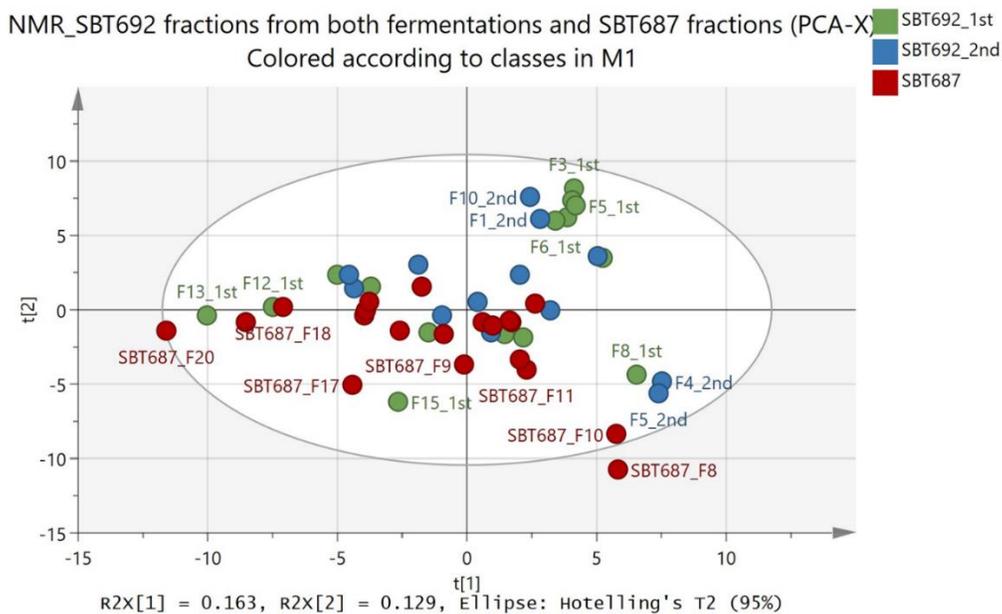


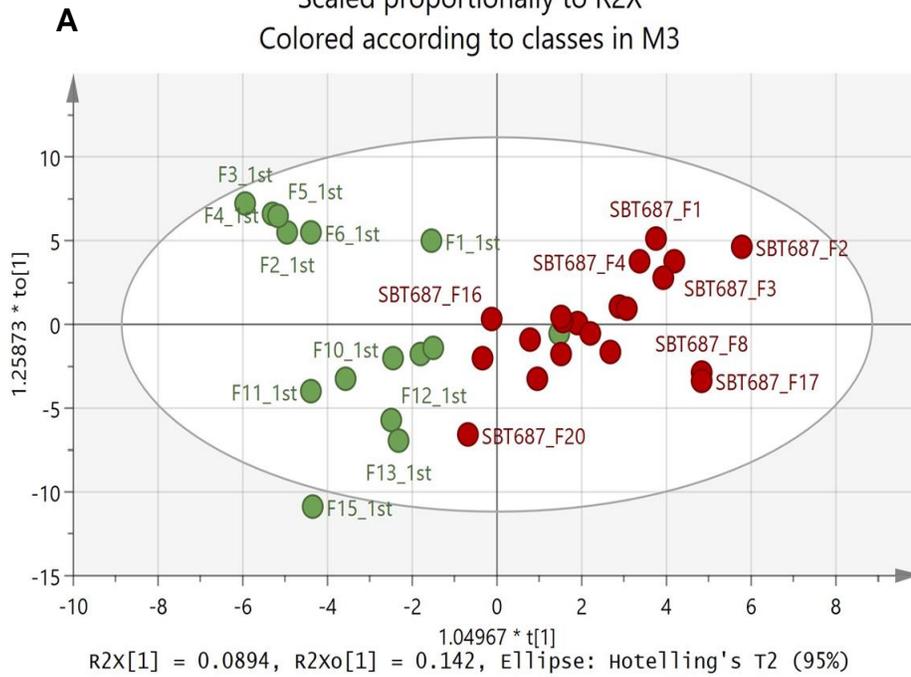
Figure 5-23 PCA-X score plot of  $^1\text{H}$  NMR data from SBT692 fractions from first and second fermentations and SBT687 fractions.

In contrast to PCA, an OPLS-DA model could identify the variables responsible to segregate the two strains. The OPLS-DA of the SBT692 fractions from first fermentation versus those of SBT687 achieved values of 0.701 and 0.242, respectively, with good fit but poor prediction. The difference between  $R^2Y$  and  $Q^2$  [ $R^2Y-Q^2 \gg 0.3$ ] was greater than 0.3, which indicated an overfitted model. The variation between groups was 8.94% [ $R^2X(1)$ ] while the variation within groups was 14.2% [ $R^2X_0(1)$ ]. The variation of the fractions within the respective groups was higher than between the different strains. The low percentage variation between strains was due to the presence of fractions crossing the respective quadrants assigned for each group. For instance, F11, F16 and F20 of SBT687 were on the left quadrant assigned for SBT 692, which indicated that these fractions were sharing more similarities with some of the SBT692 fractions of the first fermentation batch. On the other hand, F7 of SBT692 was overlapping with SBT687 fractions in the right quadrant (Figure 5.24A). The S-plot (Figure 5.24B) was again generated to determine the variables responsible for the discrimination between groups. According to the end points of the mentioned plot that were more noticeable for the SBT692 fractions, the presence of resonances lower than 2.5 ppm for these fractions, may indicate the presence of low molecular weight lipids. The model was validated by a permutation test (Figure 5.24C) with  $Q^2Y$  intercept value at -0.288.

NMR\_SBT692 1st fermentation VS SBT687 (OPLS-DA)

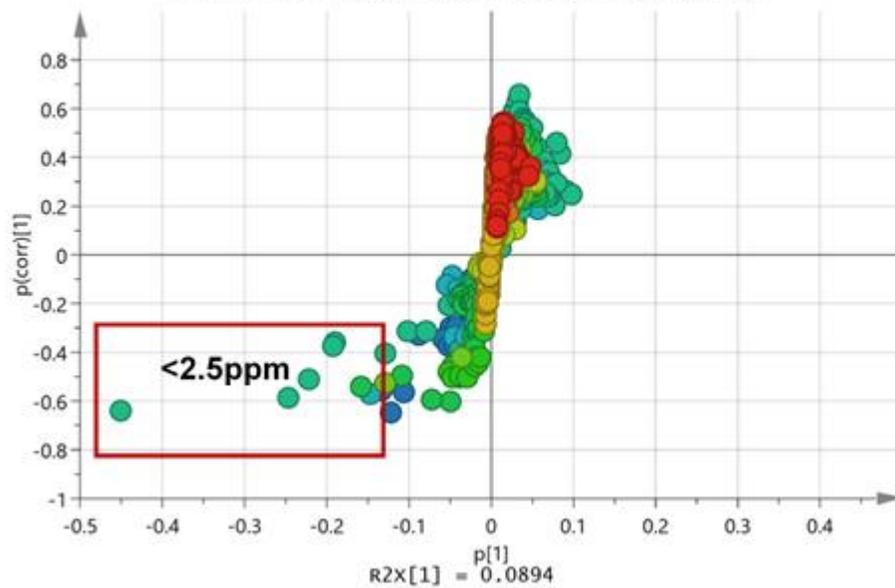
Scaled proportionally to R2X

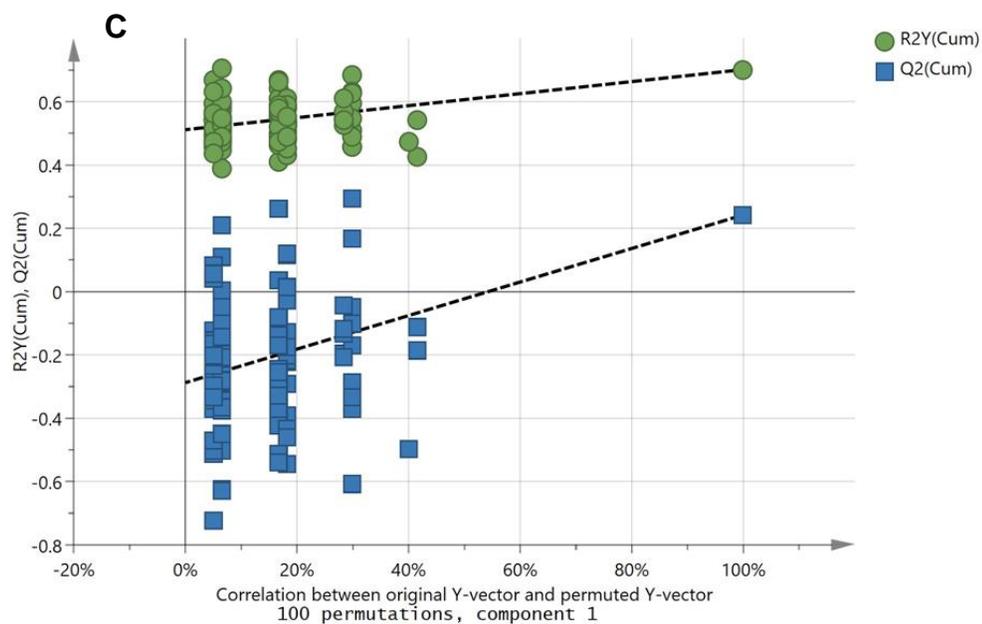
Colored according to classes in M3



NMR\_SBT692 1st fermentation VS SBT687 (OPLS-DA)

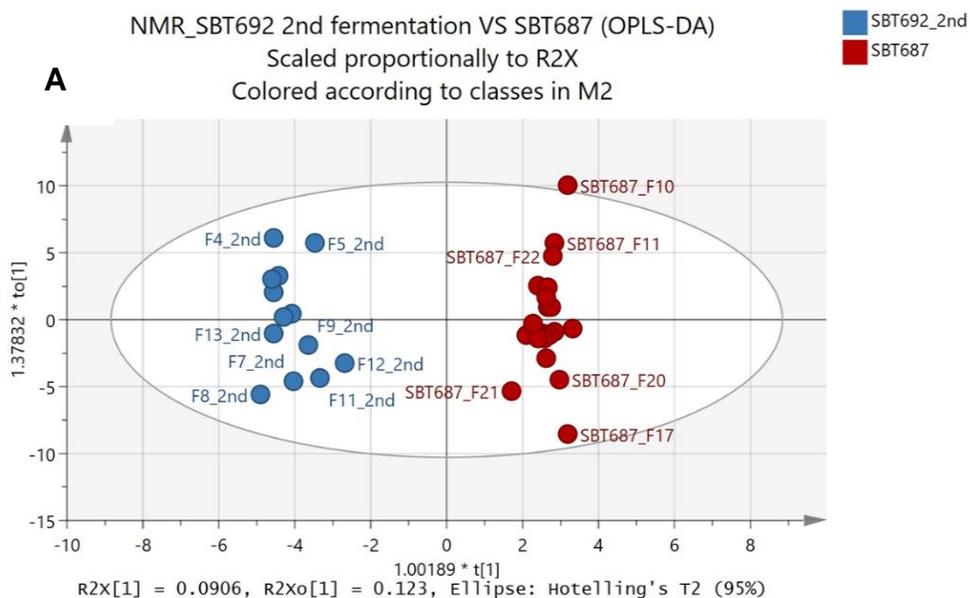
Colored according to values in DS1.Var ID (ppm)

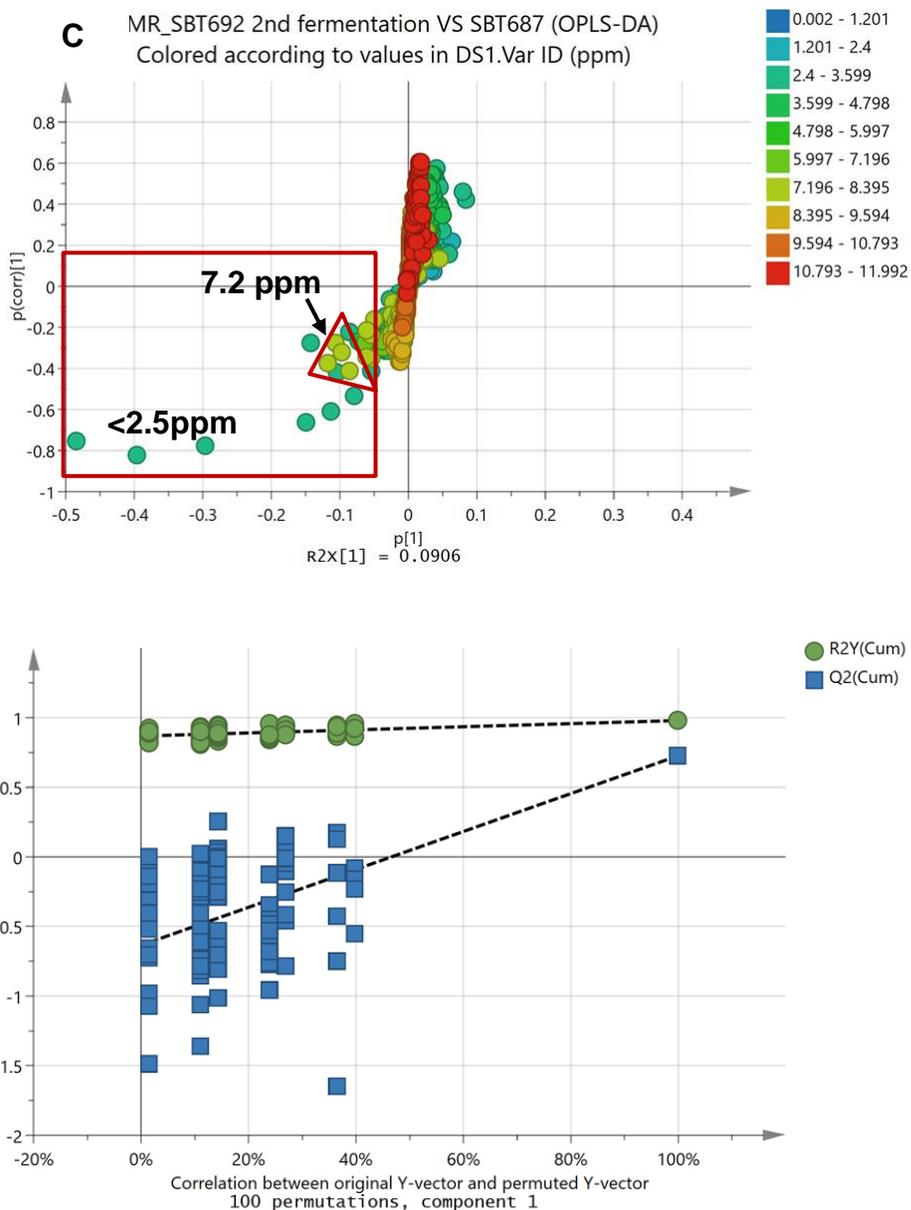




**Figure 5-24** MVDA analysis of  $^1\text{H}$ NMR spectral data of SBT92 fractions from first fermentation batch versus SBT687 fraction. (A) OPLS-DA scores plot and (B) corresponding loadings S-plot of SBT692 fractions from first fermentation batch versus SBT687 fractions. Features in the S-plot were coloured according to their chemical shifts in ppm. (C) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the  $^1\text{H}$ NMR spectral datasets between strains.

Moreover, a second OPLS-DA was performed between SBT692 fractions from the second fermentation batch versus those of SBT687.  $R^2Y$  and  $Q^2$  values of 0.979 and 0.727, respectively, were achieved, attaining a good fit and prediction with  $[R^2Y-Q^2] < 0.3$  indicating that the model was not overfitted. Both  $[R^2X(1)]$  and  $[R^2X_0(1)]$  variations were determined at 9.06 % and 12.3%, respectively, indicating that the percentage variation between strains and fractions within groups were quite comparable. The fractions from each strain studied under this section showed distinct clustering without any overlapping variables between the fractions of SBT687 on the right quadrant and those from the second fermentation batch of SBT692 on the left quadrant (Figure 5.25A). The fractions for each strain were equally dispersed resulting to a slight increase in variation within the respective groups. The loadings S-plot (Figure 5.25B) generated more distinguishable end point metabolites for SBT692 fractions of the second batch, these fractions contained structural features with chemical shifts mostly in aliphatic (red box) and aromatic region (red triangle), at less than 2.5 ppm and around 7.2 ppm, respectively. The model was validated by a permutation test (Figure 5.25C) with  $Q^2Y$  intercept at -0.635.

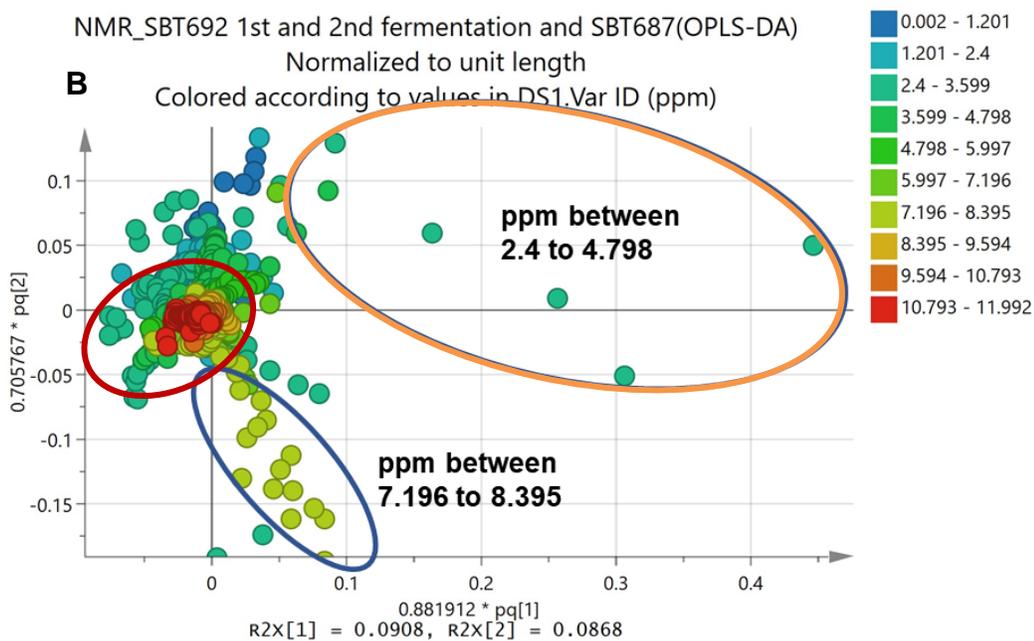
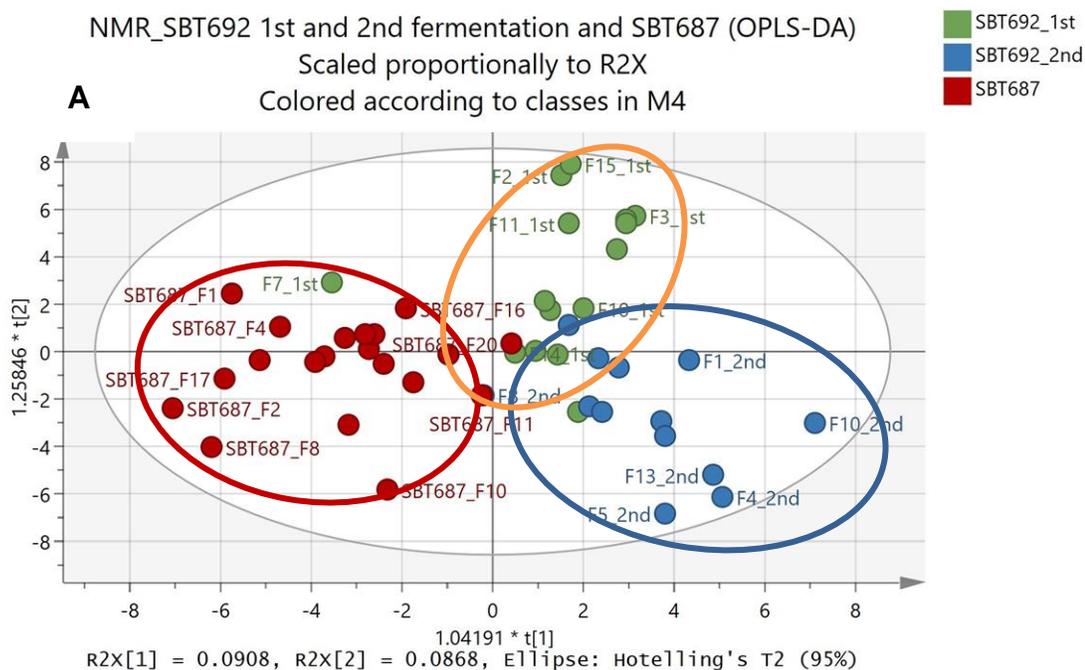


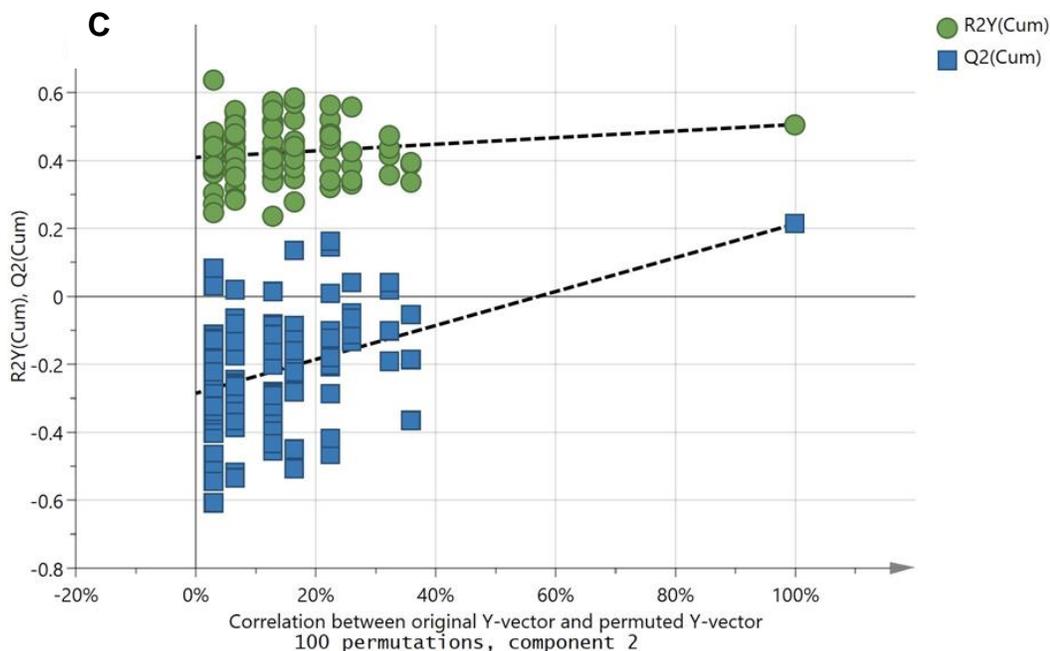


**Figure 5-25** MVDA analysis of  $^1\text{H}$ NMR spectral data of SBT92 fractions from second fermentation batch versus SBT687 fraction. (A) OPLS-DA scores plot and (B) corresponding loadings S-plot of SBT692 fractions from second fermentation batch versus SBT687 fractions. Features in the S-plot were coloured according to their chemical shifts in ppm. (C) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the  $^1\text{H}$ NMR spectral datasets between strains.

Similarly, as it was performed to the HRMS dataset, to revalidate the OPLS-DA results between comparing just two groups at a time, the NMR dataset for SBT692 fractions from both batches were also compared with those of SBT687. An OPLS-DA model was generated (Figure 5.26A) achieving a  $R^2Y$  and  $Q^2$  values of 0.606 and 0.379 respectively, for a model with both relatively a good fit and predictability for a metabolomics profiling study (Eriksson et al., 2012). The difference between  $R^2Y$  and  $Q^2$  was less than 0.3 indicating that the model was not overfitted. Variations between the three groups and within the groups were exhibited at 9.08% [ $R^2X(1)$ ] and 8.68% [ $R^2X_0(1)$ ], respectively. The variations between strains as well as fermentation batches and between fractions within the respective groups were quite similar. The OPLS-DA scores plot (Figure 5.26A) of fractions from both fermentation batches of SBT692 and those of SBT687 displayed heterogeneous distribution among the four quadrants with the presence of fractions crossing the respective quadrants assigned for each group. For instance, F20 of SBT687 was on the upper right quadrant assigned for first fermentation batch of SBT692 while, F7 of the first fermentation batch of SBT692 was found on the left quadrant assigned for SBT687 fractions. On the other hand, F8 from the first fermentation of SBT692 was overlapping with the fractions from the second fermentation of SBT692 at the lower right quadrant. Although most of the fractions of SBT 687 occurred in the quadrants on the lower and upper left quadrant, while, most the fractions of first and second fermentation batches were distributed on the upper and lower right quadrant, respectively. The loadings score plots obtained from the OPLS-DA model displayed the same set of functional groups described for the outlying fractions presented in the earlier sections (Figure 5.26B). SBT692 fractions from the first batch showed the presence of lower resonances from 2.4 to 4.8 ppm, whilst, SBT692 fractions of the second batch contained structural features with chemical shifts mostly aromatic region between 7.2 to 7.4 ppm. SBT687, positioned between the two fermentation batches of SBT 692, covered a wide range of resonances from the aliphatic to the olefinic and the aliphatic regions of the spectra, unique to SBT687 were resonances between 8 to 10 ppm that resembled those of the pyridine ring

and exchangeable *NH* protons of a manzamine alkaloid. The model was validated by a permutation test (Figure 5.26C) with  $Q^2Y$  intercept at -0.285





**Figure 5-26 MVDA analysis of <sup>1</sup>HNMR spectral data of SBT92 fractions from first and second fermentation batches versus SBT687 fractions.** (A) OPLS-DA scores plot and (B) corresponding OPLS-DA loadings plot SBT692 fractions from first and second fermentation batches versus SBT687 fractions, features were coloured according to their chemical shifts in ppm. (C) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the <sup>1</sup>HNMR spectral datasets between strains

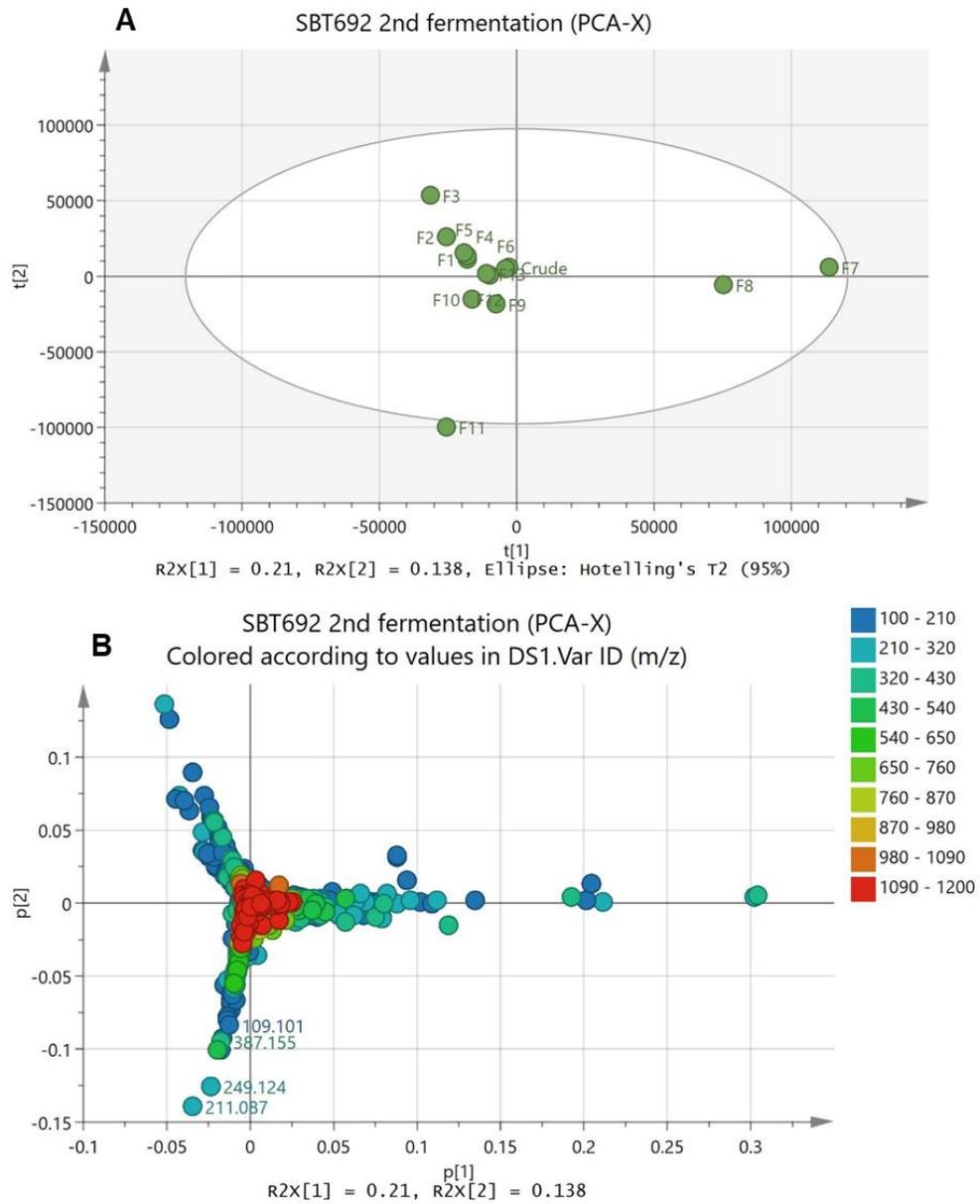
### 5.3.3 Multivariate analysis of slightly bioactive fractions obtained from the second fermentation batch of SBT692

From the second 15L scale-up, a slight inhibition effect of 40 to 50% was observed against TNF-alpha for F11 to F13 while F9 had slight bioactivity against MDA-MB-231 and Caco2 cancer cell lines. Even though the first fermentation batch seems to be more diverse in terms of its chemical profile, the second batch is likely to produce a new set of metabolites that is slightly more active than the latter. The base peak chromatogram in the negative ionisation mode of the crude extract from the second batch of scale-up indicated an increase in occurrence of acidic compounds, which may include short chain fatty acids, phenolics, polyketides, or anthraquinones. Polyketides and anthraquinones are

particularly known for their bioactivity. It will still be interesting to know whether these set of metabolites have already been earlier described in the literature with similar bioactivity as in this current study. Although SBT692 is unstable in terms of the production of its secondary metabolites, having the data on the type of compounds statistically probable to be responsible for the bioactivity will provide information to predict the biosynthetic pathway and hence, responsible gene cluster can be later explored in more stable or even industrially robust microorganisms.

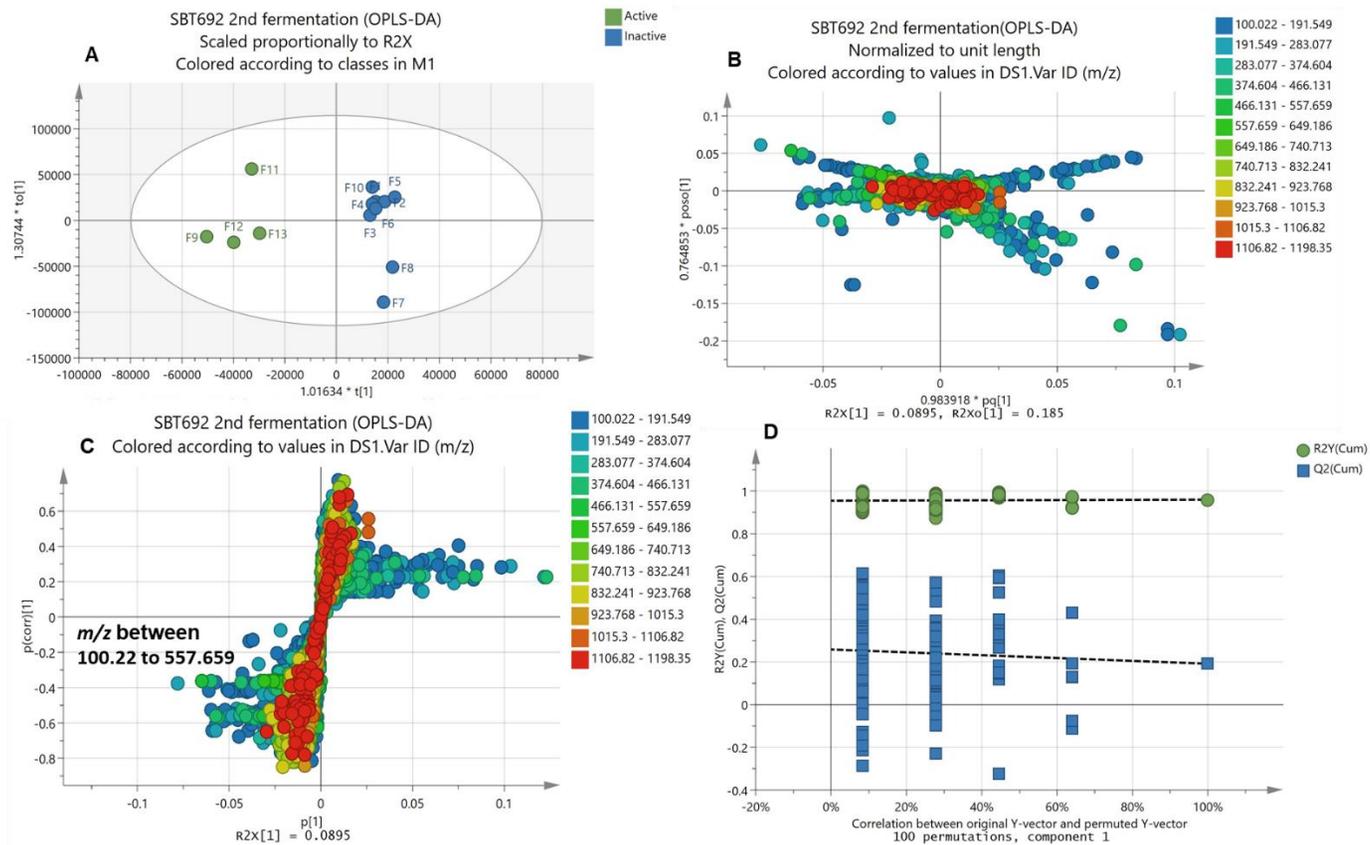
### 5.3.3.1 HRMS DATA

PCA was done on mass spectral data of all the fractions (Figure 5.27A). Using pareto scaling, PCA plots were generated achieving a R<sup>2</sup>X value of 1 and Q<sup>2</sup> of 1 after adding 13 components, which indicated a well fitted model and good prediction of variables, respectively. As observed from the PCA scores scatter plot, almost all fractions were clustering together on the mid-axis of the left quadrant indicating strong similarities of their chemical profiles, except for F7 and F8 that occurred on the right quadrant while F11 was an outlier (Figure 5.27A). Furthermore, the loadings scatter plot denoted the molecular ion peaks of the detected secondary metabolites of samples found on the corresponding quadrant position on the scores plot being possible to visualize the discriminating metabolites for the outlier fraction F11 (Figure 5.27B). The endpoint metabolites found for F11 were within 100 to 400 amu range with mass ion peaks at  $m/z$  109.101, 211.087, 249.124 and 387.155 (Table 5.11).



**Figure 5-27 PCA metabolic screening of SBT692 (2<sup>nd</sup> fermentation).** (A) PCA-X scores plot of SBT692 crude extract and fractions from second fermentations. (B) Loadings plot acquired from the PCA-X model with metabolites from the outlier fraction F11 labelled with respective *m/z*.

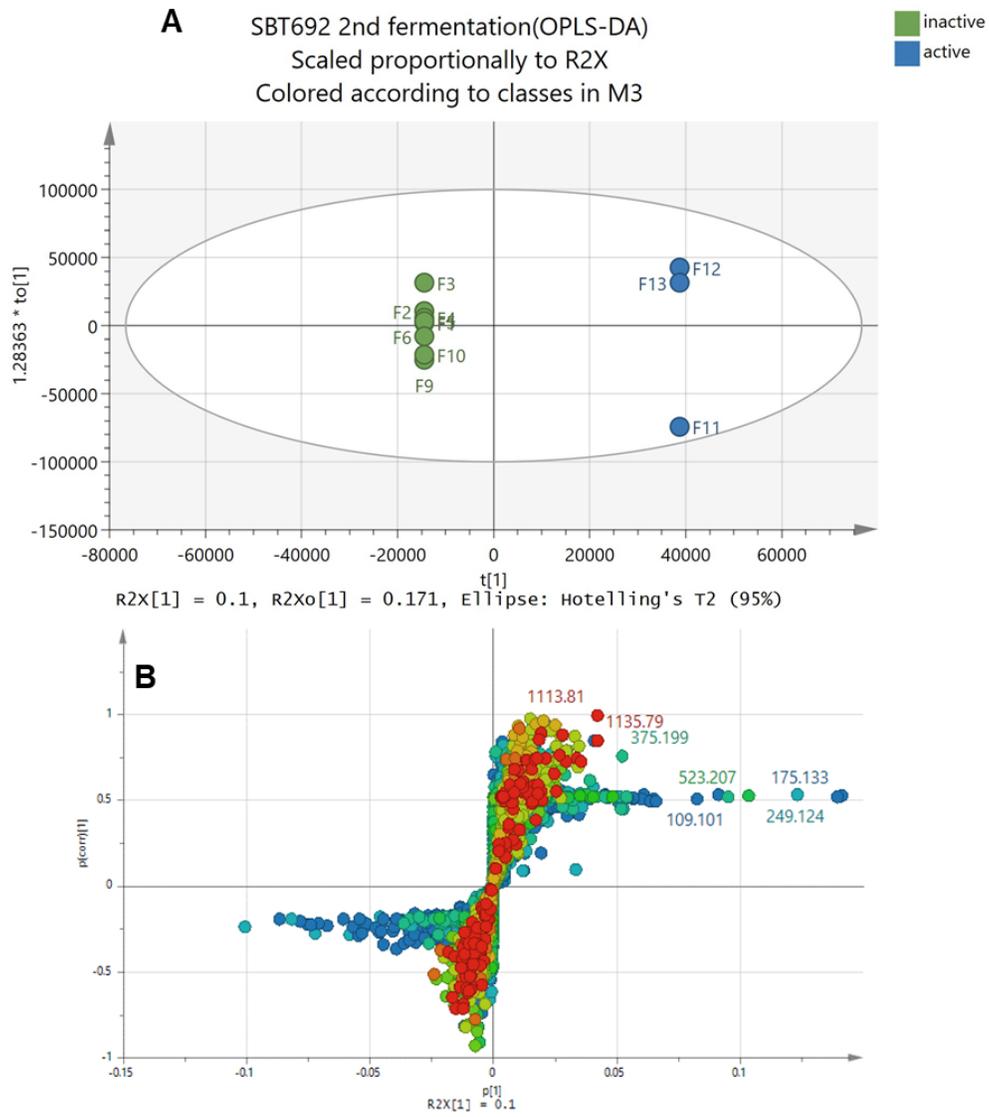
Additionally, an OPLS-DA was performed on the HRMS datasets of SBT692 fractions from the second fermentation to attempt to pinpoint compounds that could be responsible to this slight change in bioactivity of the fractions obtained from the second fermentation batch of SBT 692, fractions F9 and F11 to F13 were grouped together versus the considered inactive fractions (Figure 5.28A) Using Pareto scaling, a  $R^2Y$  value of 0.96 was achieved indicating good fit and  $Q^2$  of 0.192 which specified a weak predictive ability of the model, with a difference between  $R^2Y$  and  $Q^2$  at more than 0.3. The large difference between the  $R^2Y$  and  $Q^2$  values could indicate that the model was overfitted. The active fractions were distributed on the left side but in different quadrants, with F9, F12 and F13 located in the lower quadrant whilst F11 was on the upper quadrant, showing possible different chemical profiles between the slightly bioactive fractions. Moreover, F7 and F8 from the inactive fractions were distributed in the lower right quadrant while the remaining inactive fractions were clustering in the upper quadrant. Variations between groups and within group were exhibited at 8.95% [ $R^2X(1)$ ] and 18.5% [ $R^2X_0(1)$ ], respectively. The more dispersed distribution of the fractions from both groups of fractions, caused the percentage variation within group to be greater than the variation between the active versus inactive group. Moreover, a loading scatter plot (Figure 5.28B) and S-plot (Figure 5.28C) were performed to visualize the discriminating metabolites from the active fractions which were mostly metabolites yielding low to mid-MW with mass ion peaks at  $m/z$  100.220 to 557.659. The model was excluded as the permutation test afforded a  $Q^2Y$  intercept at 0.258 ( $Q^2Y > 0$ ) (Figure 5.28D).



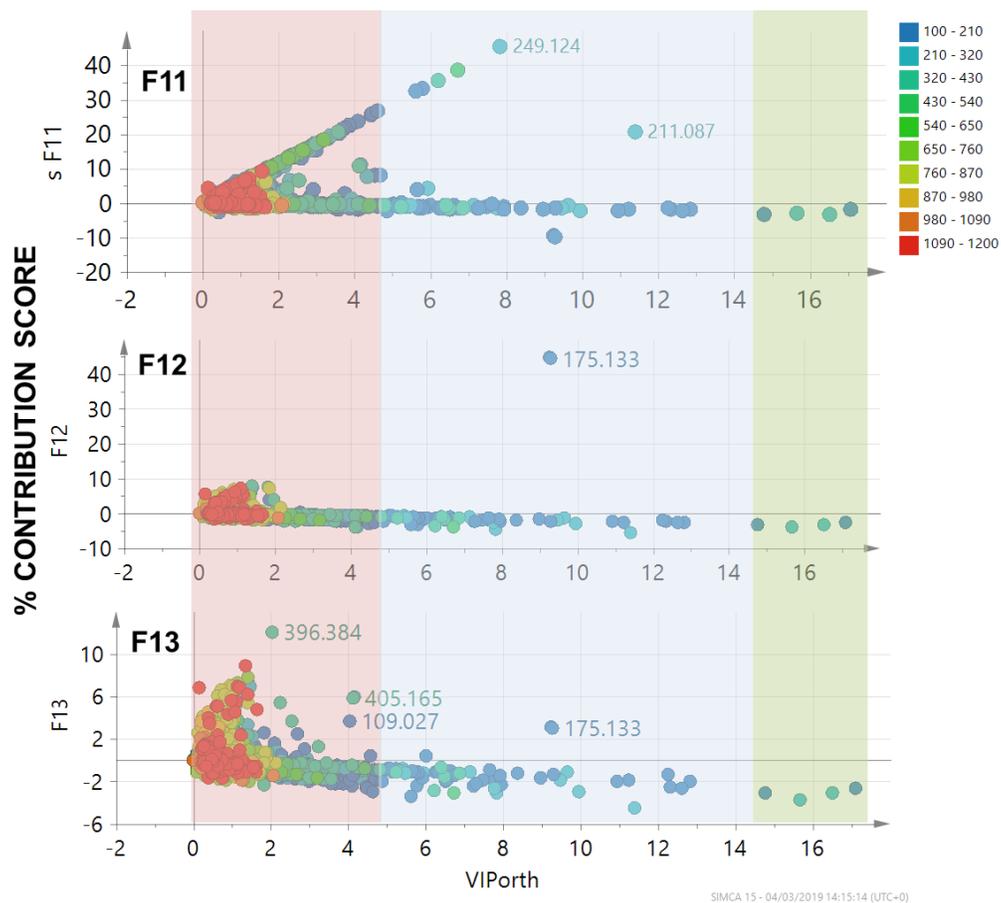
**Figure 5-28** MVDA first analysis of mass spectral data of SBT92 fractions from second fermentation batch according to bioactivity. (A) OPLS-DA scores plot of SBT692 active fractions from second fermentation batch versus inactive and (B) corresponding loadings plot and (C) S-plot of SBT692 OPLS-DA. Features in the loading plot and S-plot were coloured according to their *m/z* ranges. (D) Permutation tests (100 permutations) to validate the OPLS-DA model differentiating the mass spectral datasets between fractions.

Hence the first model was excluded because it cannot be validated by the permutation test, a new model was created with the active fractions against TNF-alpha, excluding the inactive outlying fractions F7 and F8 that could be affecting the model and F9 was set as inactive (Figure 5.29). The  $R^2Y$  and  $Q^2$  of 1 and 0.506, respectively, specified a model with excellent fit and good predictability. Once again, the difference between  $R^2Y$  and  $Q^2$  was greater than 0.3 indicating that the model could be overfitted. Variations between groups and within group were analogous to the first model and were at 10% [ $R^2X(1)$ ] and 17.1% [ $R^2X0(1)$ ], respectively. The more dispersed distribution of the active fractions caused the percentage variation within group to be greater than the variation between the active versus inactive group (Figure 5.29A). A loadings S-plot (Figure 5.29B) was generated to visualize the variables responsible for the discrimination between groups. According to the end points of these plots, the active fractions contained metabolites with variable range of MW with mass ion peaks from  $m/z$  100 to 1200 amu (Table 5.11). The P values were relatively good for 50% of the end point features that were found significant at  $P \leq 0.05$ . A permutation test was performed to test the validity of the model which afforded a positive  $Q^2Y$  intercept at 0.443 ( $Q^2Y > 0$ ) that concluded the model was not valid due to poor predictability.

Although the permutation test was not able to validate the OPLS model, IP and % contribution scores could be employed to statistically validate the predicted bioactive metabolites against TNF-alpha (Figure 5.29B) specifically those afforded by F11, F12, and F13 (Figure 5.30). As shown, the predicted bioactive features for F11 and F12 had high VIP and % contribution scores, greater than 7.0 and 20% or 40%, whereas those from F13 gave lower % contribution scores of 2 to 10% while the VIP scores were 2.0 to 10.0. The predicted targeted bioactive metabolites against TNF-alpha were dereplicated and details are listed in Table 5.11

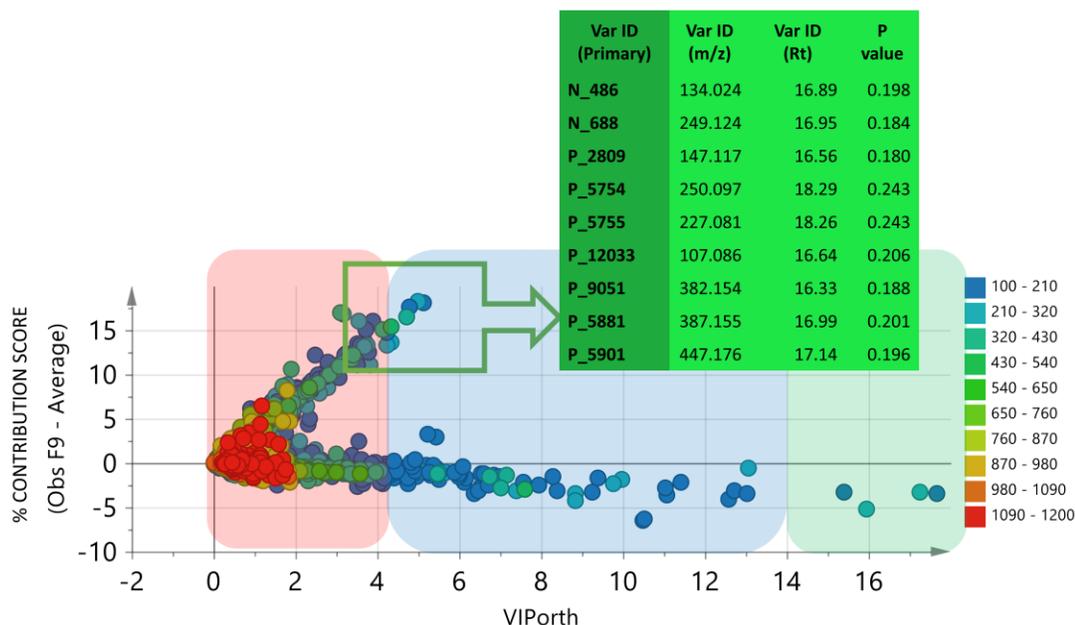


**Figure 5-29** MVDA second analysis of mass spectral data of SBT92 fractions from second fermentation batch according to bioactivity against TNF-alpha. (A) OPLS-DA scores plot of SBT692 active fractions against TNF-alpha from second fermentation batch versus inactive and (B) corresponding loadings S-plot of SBT692 OPLS-DA. Features in the and S-plot were coloured according to their *m/z* ranges and endpoint metabolites were labelled with respective *m/z*.



**Figure 5-30 Contribution scores plots for F11, F12 and F13 to indicate the orthogonal VIP (Variable Influence/Importance in Projection) scores on an OPLS model for the “end point” features for the respective fractions on the loadings S-plot (Figure 5.29B) as labelled here with their *m/z* ratios. Coloured regions designate increasing orthogonal VIP scores where red < 5, blue = 5 to 15, and green > 15.**

Similarly, a VIP and % contribution scores plot analysis was also done on F9, which afforded bioactivity against MDA-MB-231 and Caco2 cancer cell lines. As it was observed on the contribution score plot on Figure 5.31, the predicted bioactive features for F9 gave % contribution scores of 10 to 15% while the VIP scores were between 4.0 to 6.0. The predicted targeted bioactive metabolites with potential anti-cancer activity were dereplicated and details are listed in Table 5.11. Even though the P values were found greater than 0.05 (but less than 0.25), which were not statistically significant in terms of its bioactivity to the tested cell lines, the predicted bioactive metabolites were still found unique to F9 and hence predicted selective to MDA-MB-231 and Caco2 cancer cell lines.



**Figure 5-31 Contribution scores plots for F9 to indicate the orthogonal VIP (Variable Influence/Importance in Projection) scores on an OPLS model for the “end point” features for the respective fractions on the loadings S-plot (Figure 5.28C) as tabulated here with their *m/z* ratios. Coloured regions designate increasing orthogonal VIP scores where red < 5, blue = 5 to 15, and green > 15.**

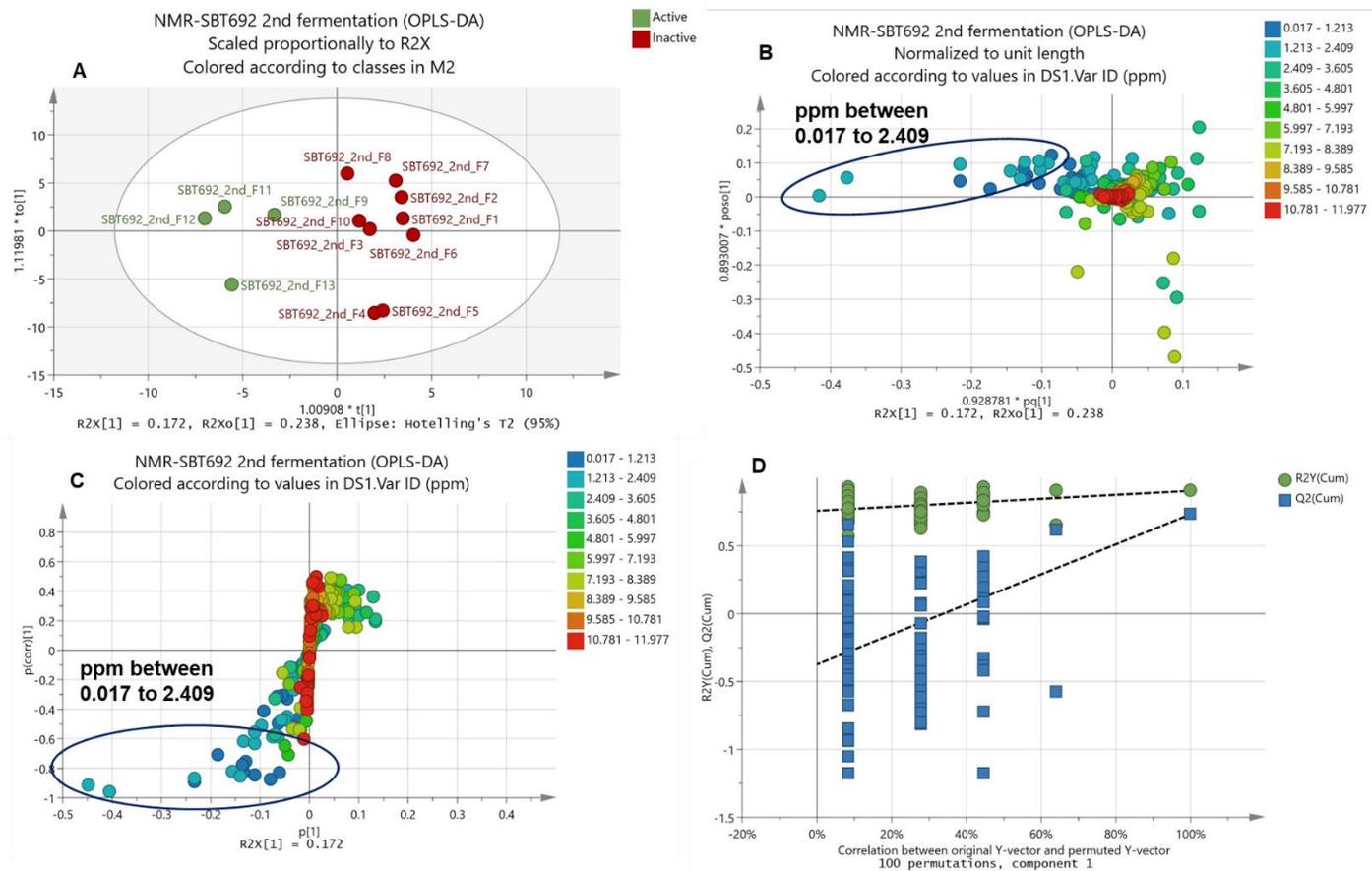
**Table 5-11 Dereplication table of the predicted bioactive end point metabolites from fraction F9, F11, F12 and F13 from SBT692 second fermentation** visualized in the S-plot of the OPLS-DA analysis. Highlighted metabolites were common between bioactive fractions. (NS = not significant,  $P>0.05$ ).

MzMineID	m/z	Rt (min)	Peak Area (highest)	MWt	Molecular Formual	Name	Source	Pvalue
<b>Predicted bioactive metabolites for F11, F12 and F13</b>								
<b>P_3375</b>	405.1652	13.8	7.50E+08	404.1579	C <sub>20</sub> H <sub>24</sub> N <sub>2</sub> O <sub>7</sub>	myxochelin-A	<i>Angiococcus disciformis an 30 (myxobacterales)</i>	2.7E-10
<b>P_13012</b>	175.1331	13.93	2.56E+05	174.1258	C <sub>9</sub> H <sub>18</sub> O <sub>3</sub>	No hits	No hits	0.10 NS
<b>P_1250</b>	109.0278	13.99	3.90E+06	108.0205	C <sub>6</sub> H <sub>4</sub> O <sub>2</sub>	benzoquinone	No hits	0.11 NS
<b>P_2321</b>	203.1794	18.41	3.41E+07	202.1722	C <sub>15</sub> H <sub>22</sub>	neolaurencyne	Rhodophyta <i>Laurencia okamurai</i>	3.9E-07
<b>P_3336</b>	135.1169	20.15	2.41E+07	134.1096	C <sub>10</sub> H <sub>14</sub>	No hits	No hits	0.01
<b>P_3196</b>	119.0857	20.17	2.33 E+07	118.0784	unpredicted	No hits	No hits	9.5E-06
<b>P_3108</b>	107.0859	20.2	2.98 E+07	106.0786	C <sub>8</sub> H <sub>10</sub>	No hits	No hits	6.0E-06
<b>P_8689</b>	149.1325	20.41	5.32E+07	148.1252	C <sub>11</sub> H <sub>16</sub>	dictyopterene B	<i>Ochrophyta Dictyopteris plagiogramma</i>	0.03
<b>P_1002</b>	211.0865	20.51	2.25 E+08	210.0793	C <sub>13</sub> H <sub>10</sub> N <sub>2</sub> O	1-vinyl-8-hydroxy- $\beta$ -carboline	Bryozoa <i>Cribricellina cribraria</i>	0.03
<b>P_8584</b>	297.0757	20.61	2.03 E+07	296.0684	C <sub>17</sub> H <sub>12</sub> O <sub>5</sub>	haloquinone	<i>Streptomyces venezuelae subsp. xanthophaeus Tue 2115</i>	0.01
<b>P_8691</b>	161.1324	20.65	2.01 E+07	160.1251	C <sub>12</sub> H <sub>16</sub>	inflatene	Cnidaria <i>Clavularia inflata</i>	0.10 NS
<b>P_3304</b>	147.1168	20.66	2.50 E+07	146.1095	C <sub>11</sub> H <sub>14</sub>	desmarestene	<i>Ochrophyta Desmarestia sp</i>	0.01
<b>P_184</b>	385.1407	21.06	2.29 E+07	386.1479	C <sub>20</sub> H <sub>22</sub> N <sub>2</sub> O <sub>6</sub>	No hits	No hits	0.10 NS
<b>N_82</b>	249.1243	21.08	1.13 E+08	250.1316	C <sub>13</sub> H <sub>18</sub> N <sub>2</sub> O <sub>3</sub>	nongidin A/ jenamidine A1/A2	<i>Streptomyces sp. GT 061115, Streptomyces sp. strain HKI0297</i>	0.09
<b>P_841</b>	523.207	21.09	8.16 E+07	522.2002	C <sub>28</sub> H <sub>30</sub> N <sub>2</sub> O <sub>8</sub>	No hits	No hits	0.10 NS
<b>P_844</b>	387.1548	21.09	6.97E+07	386.1476	C <sub>20</sub> H <sub>22</sub> N <sub>2</sub> O <sub>6</sub>	No hits	No hits	0.10 NS

<b>N_137</b>	134.0235	21.09	6.13 E+07	135.0308	C <sub>5</sub> H <sub>3</sub> N <sub>4</sub> O	No hits	No hits	0.09 NS
<b>P_2921</b>	375.1986	22.80	8.56E+06	374.1914	C <sub>15</sub> H <sub>22</sub> N <sub>10</sub> O <sub>2</sub>	No hits	No hits	0.01
<b>P_8585</b>	225.1022	23.07	1.94 E+07	224.0949	C <sub>14</sub> H <sub>12</sub> N <sub>2</sub> O	dihydrophenazin- monoacetate	[synthetic]	0.10 NS
<b>P_2463</b>	1113.811	31.83	3.1E+05	1112.804	C <sub>75</sub> H <sub>100</sub> N <sub>8</sub>	No hits	No hits	2.8E-10
<b>Predicted bioactive metabolites for F9</b>								
<b>P_9051</b>	382.1542	16.33	2.48E+07	381.1469	C <sub>24</sub> H <sub>19</sub> N <sub>3</sub> O <sub>2</sub>	benzomalvin A	<i>Penicillium sp.</i>	0.180 NS
<b>P_2809</b>	147.1168	16.56	2.50 E+07	146.1095	C <sub>11</sub> H <sub>14</sub>	desmarestene	Ochrophyta <i>Desmarestia</i> <i>sp</i>	0.188 NS
<b>P_12033</b>	107.0859	16.64	2.98 E+07	106.0786	C <sub>8</sub> H <sub>10</sub>	No hits	No hits	0.201 NS
<b>N_486</b>	134.0235	16.89	3.9 E+07	135.0308	C <sub>5</sub> H <sub>3</sub> N <sub>4</sub> O	No hits	No hits	0.198 NS
<b>N_688</b>	249.1243	16.95	1.13 E+08	250.1316	C <sub>13</sub> H <sub>18</sub> N <sub>2</sub> O <sub>3</sub>	nongidin A/ jenamidine A1/A2	<i>Streptomyces sp. GT</i> <i>061115, Streptomyces sp.</i> <i>strain HKI0297</i>	0.184 NS
<b>P_5881</b>	387.1548	16.99	3.15E+07	386.1476	C <sub>20</sub> H <sub>22</sub> N <sub>2</sub> O <sub>6</sub>	No hits	No hits	0.201 NS
<b>P_5901</b>	447.1759	17.14	2.67E+07	446.1686	C <sub>22</sub> H <sub>26</sub> N <sub>2</sub> O <sub>8</sub>	No hits	No hits	0.20 NS
<b>P_5755</b>	227.0184	18.26	2.47E+07	226.0741	C <sub>13</sub> H <sub>10</sub> N <sub>2</sub> O <sub>2</sub>	beta-carboline-3- carboxylic acid methyl ester 6-Methoxy-1- phenazinol	<i>Streptomyces sp. Ma5373</i>  <i>Streptomyces thioluteus</i>	0.24 NS
<b>P_5454</b>	111.0443	18.29	1.17E+07	110.0370	C <sub>6</sub> H <sub>6</sub> O <sub>2</sub>	No hits	No hits	0.24 NS

### 5.3.3.2 NMR DATA

The  $^1\text{H}$  NMR spectral data from SBT692 fractions from the second fermentation, was subjected to multivariate analysis similar to the mass spectral dataset (Figure 5.30) to explain and predict the structure family, functional groups and substituents responsible for the slight increase in bioactivity of the strain after a second batch of fermentation. Using Pareto scaling, an OPLS-DA was performed achieving a  $R^2Y$  value of 0.906, and a  $Q^2$  of 0.733, which specified a well fitted model ( $R^2X \geq 0.5$ ) with good prediction of variables with  $Q^2 > 0.5$ . The difference between  $R^2X$  and  $Q^2$  was less than 0.3, confirming complete absence of overfitting. The variation between groups (active versus inactive) was 17.2% [ $R^2X(1)$ ] while the variation within groups was 23.8% [ $R^2X_0(1)$ ]. Therefore, the variation within between fractions or within groups was higher than between the active versus inactive groups. From the OPLS-DA model, a dispersed distribution of the fractions within the quadrants of the respective groups could explain the high percentage of variation within groups (Figure 5.32A). Moreover, a loadings scatter plot (Figure 5.32B) was generated to visualize the outlying features from the active fractions, which exhibited signals lower than 2.5 ppm. The S-plot (Figure 5.32C) was created to visualize the variables responsible for the discrimination between the active versus the inactive fraction, which exhibited, similarly as the loading scatter plot, the presence of resonances lower than 2.5 ppm that may indicate the presence of low molecular weight lipids. The model was validated by a permutation test (Figure 5.32D) with  $Q^2Y$  intercept at  $-0.373$ .



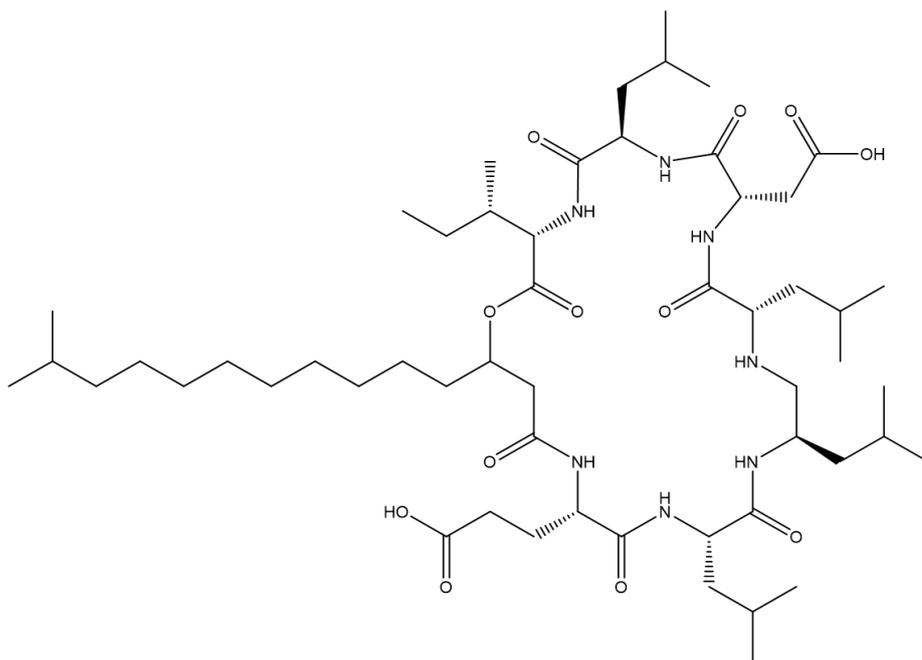
**Figure 5-32 MVDA analysis of NMR data of SBT92 fractions from second fermentation according to bioactivity . (A)** OPLS-DA scores plot of SBT692 active fractions from second fermentation batch versus inactive fractions and **(B)** corresponding loadings scores plot and **(C)** loadings S-plot of SBT692 OPLS-DA. Features in the loadings scatter plot and S-plot were coloured according to chemical shifts in ppm. **(D)** Permutation tests (100 permutations) to validate the OPLS-DA model of SBT692 differentiating the <sup>1</sup>H NMR spectral datasets between fractions

## 5.4 Discussion

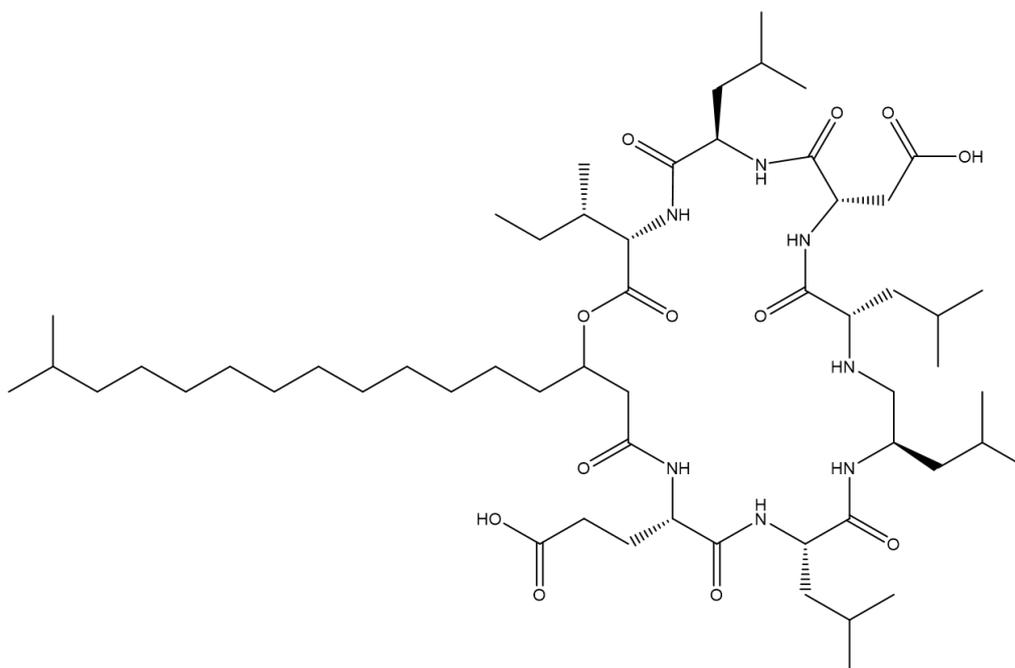
In the present work, besides the small 3L fermentation, two batches of 15L scale-up were performed on *Micromonospora sp.* N74 (SBT692) which were further subjected to extraction, fractionation, and metabolic profiling analysis. NMR and HRMS were used as analytical platforms to achieve the metabolomics analysis. The preliminary bioassay screening of SBT692 showed inhibition against several factors as well as against TNF-alpha-induced NF-kB. However, the fractions did not display any inhibition activity towards TNF-alpha, as well as cytotoxicity effect on NCTC cell line and cell viability effect for the tested cancer cell lines, MCF-7, MDA-MB- 231, T84, HT29 and CaCo2.

*Micromonospora sp.* is an important source of bioactive compounds with different therapeutically properties. The genus is known to exhibit a considerable physiological and biochemical diversity being as well a defined group in terms of morphology, phylogeny and chemotaxonomy (Hirsch and Valdés, 2010). The dereplication study of the most diverse fractions, afforded 50% of the compounds putatively identified while the other 50% did not give any match from the in-house macro database. The identified metabolites belonged to distinct classes, such as, diketopiperazines, lipopeptides, phenolics, quinones, lactones, guanidines alkaloids, macrolides, steroids, among others, showing the high diversity of secondary metabolites being produced by the *Micromonospora sp.* N74 (SBT692), as shown in Figure 5.33/34.





pumilacidin A



pumilacidin C

**Figure 5-34 (continued)** Major secondary metabolites identified from the dereplication of the outlying fractions being produced by the *Micromonospora* sp. N74 (SBT692) from the first fermentation batch.

The first fermentation batch of SBT692 produced a diverse set of metabolites that disappeared during the second batch of 15L scale-up. Below is a description of the discriminating compounds found in the first batch of scale-up as earlier reported in the literature.

Albonoursin is a diketopiperazine (3-benzylidene-6-isobutylidene-2,5-dioxypiperazine) that first reported in 1963 by Rosenfeld *et al.* (Fukushima *et al.*, 1973). This compound was produced by several actinomycetes, exhibiting known antibacterial activity (Fukushima *et al.*, 1973, Li *et al.*, 2014). Pumilacidin A and C were found in F15 of the first fermentation batch. Pumilacidins are cyclic lipopeptides with a cyclized heptapeptide attached to a  $\beta$ -hydroxy fatty acid derived from the surfactin family (Elleuch *et al.*, 2012). Pumilacidin has been described having antiviral activity against herpes simplex virus type 1 (HSV-1) as well a protective effect against gastric ulcers (Naruse *et al.*, 1990). Myxochelin A is a phenolic catechol siderophore with weak antimicrobial activity (Kunze *et al.*, 1989). However, this compound or its derivatives exhibited other biological activities like growth suppression of K-562 leukemia cells (Schieferdecker *et al.*, 2015) or inhibition of the 5-lipoxygenase (5-LO) pathway (Schieferdecker *et al.*, 2017). Cribrostatin 5 is an isoquinoline quinone showing antineoplastic effects and antibacterial and/or antifungal activities (Pettit *et al.*, 2000). The compound 19,23-dihydroxyprotylonolide is a 16-membered macrolide antibiotic reported for the first time in 1994 but also isolated from *Micromonospora sp.* YS-02930K (Yasumuro *et al.*, 1994). Dinordehydrobatzelladine B a polycyclic guanidine alkaloid, was identified in 2009 from a marine sponge *Monanchora arbuscula* and showed mild antitumor and antimalarial activity (Laville *et al.*, 2009, Mayer *et al.*, 2013). The 6-deoxyerythronolide B was described as a key precursor of the macrolide antibiotic erythromycin (Meng *et al.*, 2011). Echinasteroside C is a polyhydroxysteroidal glycoside commonly isolated from different species of starfish, of the phylum Echinodermata, class Asteroidea (Iorizzi *et al.*, 1993, Malyarenko *et al.*, 2016). This class of compounds have been reported to display a high range of biological activities including antiviral, antibacterial, antifungal, cytotoxic, hemolytic, antitumor, immunoregulatory, neuritogenic and anti-biofouling effects (November, 2018, Kicha *et al.*, 2008). Primycin was first described in the Nature in 1954 (Valyi-

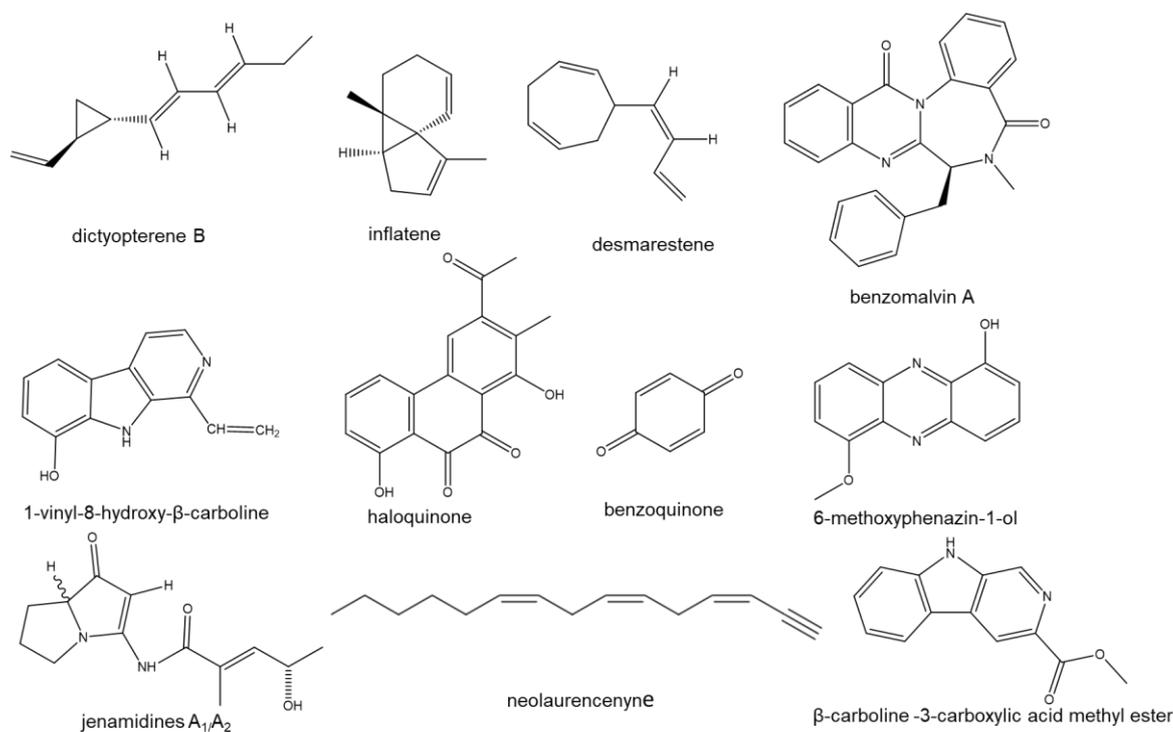
Nagy et al., 1954) Being classified as a guanidine type antibiotic with high efficacy against Gram-positive pathogens, primycin is only used topically in dermatological indications (Gesztelyi et al., 1980, Feiszt et al., 2014).

From the above-mentioned dereplicated compounds, only myxochelin A was described to have an inhibitory role in inflammation, specifically in the 5-LO pathway and not against TNF-alpha. The 5-LO is responsible for the catalyses of two initial steps in the biosynthesis of leukotrienes, known mediators of a variety of inflammatory and allergic reactions as well associated with cancer, where an increased formation of 5-LO products was observed (Haeggstrom and Funk, 2011, Radmark et al., 2015).

The second fermentation batch of SBT692 produced new set of metabolites that were slightly more active than the first fermentation batch. Figure 5.35 illustrates the discriminating compounds found in the second batch of scale-up as earlier reported in the literature, and below the description focusing on the ones with described anti-cancer activity.

Previous bioassays studies with the alkaloid 1-vinyl-8-hydroxy-beta-carboline (1-VHB-carboline) exhibited cytotoxicity (100 ng/mL, LC<sub>50</sub>) against mice lymphoma (P388D1, ATCC CCL-46) cells lines (Prinsep et al., 1991) and relatively potent cytotoxicity at 19 µM (LC<sub>50</sub>) against human NCI-60 tumour cell lines screening panel (Beutler et al., 1993). Haloquinone, a phenanthrenequinone derivative, synthetic derivatives halt aberrantly activated Wnt/Beta-Catenin signalling pathways (Halbedl et al., 2013) at 2.5 µM in human embryonic kidney (HEK293) cell lines. Notably, all derivatives were also able to block the expression of Axin-2 (AXIN2), a known inhibitor of the Wnt signalling pathway (Krishnamurthy and Kurzrock, 2018) in colorectal cancer cell lines (SW-480, ATCC, CCL-228) (Halbedl et al., 2013). Jenamidine A1/A2, a bicyclic alkaloid, inhibits the proliferation at 1.91 µg/mL (GI<sub>50</sub>) of human bone marrow-derived chronic myelogenous leukemia cell lines (K-562, ATCC, CCL-243) (Snider and Duvall, 2005, Duvall et al., 2006, Stevens et al., 2011). Benzoquinones and its numerous derivatives are well known secondary metabolites with reported anti-cancer properties (Lu et al., 2013). For instance, one benzoquinone

with growing interest as potential therapeutically agent for cancer is embelin (Lu et al., 2016). Embelin, among modulating other important pathways, may be responsible for inducing apoptotic cell death by supressing the NF-k $\beta$  and p53 pathways, holding several mechanisms of action that modulate diverse representative characteristic of tumour cells (Ko et al., 2018, Prabhu et al., 2018).



**Figure 5-35 Major secondary metabolites identified from the dereplication of the slightly bioactive fractions afforded by the *Micromonospora* sp. N74 (SBT692) from the second fermentation.**

The initial crude extract accounted for inhibition activity against TNF-alpha. However, after the 15 L scale-up fermentation and fractionation, the inhibition activity disappeared. Therefore, different metabolomic profiling approaches using both <sup>1</sup>H NMR and mass spectral data along with multivariate analysis of the fractions from both fermentation batches, allowed the validation of the lack of stability of *Micromonospora* sp. N74 (SBT692) in terms of the production of its secondary metabolites. Hence, the loss of bioactivity during the scale-up process. Furthermore, there was evidence of the unstable production of metabolites, with changes in chemical profile between the first and the second fermentation scale-up, as it was seen in the

data presented. Moreover, both *Micromonospora sp.* (SBT687 and SBT692) were isolated from the same sponge and identical geographical location, thus, it was expected that both strains were producing identical secondary metabolites. Nevertheless, it was clear that the SBT687 and SBT692 strains were producing different type of metabolites, as it was shown with the metabolomics profiling data analysis.

## Chapter 6

CONCLUSIONS AND FUTURE WORK

## 6.1 Conclusions

The present project aimed to identify and isolate the bioactive metabolites production in the thermophile strains of *Muricauda ruestringensis* (SBT531 and SBT587) isolated from geothermal intertidal pools in Iceland and in *Micromonospora sp.* (SBT687 and SBT692) isolated from the Mediterranean sponge *Phorbas tenacior* from the Santorini volcanic complex of Crete. A metabolomics-guided isolation approach was implemented, applying NMR and HR-LC/MS as analytical platforms to achieve the metabolomics analysis. Through the application of metabolomics tools, fermentation and growth parameters were optimised to increase the production yield of the respective bioactive target metabolites, accomplishing the purposed objectives.

### 6.1.1 *Muricauda ruestringensis* (SBT531 and SBT587) metabolomic profile and biological importance of isolated metabolites as potential targets for metabolic diseases

Since both *M. ruestringensis* strains, SBT531 and SBT587, were isolated from the same geographical location, was expected that both strains secondary metabolites were analogous, which was confirmed after scrutiny of their similar NMR and HR-LC/MS data. The fractionation of SBT531 afforded a series of bioactive alpha and beta-hydroxy acid derivatives. Aseanostatin P6 (13-methyltetradecanoic acid) was the major compound identified which exhibited inhibition activity against alpha-glucosidase (30 µg/mL) and PTP1B (30 and 10 µg/mL, Ki 3 µg/mL) with less than 40% inhibition ( $\leq 40\%$  threshold), followed by two derivatives, 3-hydroxy-13-methyltetradecanoic acid and 2-hydroxy-14-methylhexadecanoic acid. These compounds showed inhibition activity against PTP1B at a concentration of 30 µg/mL but not at 10 µg/mL. The comparison of the three isolated fatty acids in terms of their inhibition ability against PTP1B, suggested that the inhibition activity was structure-

dependent as it decreased or disappeared with the elongation and/or hydroxylation and methylation of the chain.

### **6.1.2 *Micromonospora* sp. N17 (SBT687) and N74 (SBT692) potential agrochemical and anti-cancer biological activity from isolated/dereplicated compounds and metabolomic profile**

Considering that both strains of *Micromonospora* sp. N17 and N74 (SBT687 and SBT692) were isolated from the same sponge and identical geographical location, was anticipated that both strains were producing identical secondary metabolites. Nevertheless, it was clear that the SBT687 and SBT692 strains were producing different classes of metabolites, as it was shown with the metabolomic data analysis. The initial insecticidal activity against parasite *Lepeophtheirus* sp. (sea lice) at 75  $\mu\text{g/mL}$  of the *Micromonospora* sp. N17 (SBT687) decreased from 100% inhibition to 72%, after fractionation, this could be explained that the activity is stronger in the crude extract where all compounds, predominantly the ones identified/mentioned, are organised with a synergistic effect. Application of multivariate analysis allowed to predict the most statistically probable bioactive metabolites of SBT687, which were targeted for isolation and elucidation work. The compounds were structurally identified as a combination of manzamines and 4-epi-manoyl oxide acid, from fractions 17 and 18, respectively. Even though the labdane type diterpene, 4-epi-manoyl oxide acid, is reported as a plant metabolite, it was isolated from a marine microorganism, that could suggest that the metabolite is not being produced by the plant but by the microorganism. Manzamines-yielding sponges were often isolated from Indian and Pacific oceans, hence that this class of compounds were identified from the Mediterranean sponge derivate *Micromonospora* sp. N17 (SBT687), suggested the hypothetical role of microbes in the biosynthesis of manzamines.

Different scale-up fermentations were performed on *Micromonospora* sp. N74 (SBT692). The preliminary bioassay screening of the initial crude extract of SBT692 showed inhibition against TNF-alpha TRPA1, TRPM8, PPARA as well as cell viability effect of human colorectal carcinoma cell line (HCT116). However, afterward the 15 L scale-up fermentation and fractionation, the inhibitory effects towards TNF-alpha, as well as cell viability effect for the tested breast and colon cancer cell lines, MCF-7, MDA-MB- 231, T84, HT29 and CaCo2, were missing. Subsequently, several metabolomic approaches using the HR-LCMS and <sup>1</sup>H NMR data permitted to acknowledge the unstable production of secondary metabolites by the *Micromonospora* sp. N74 (SBT692), hence the loss of bioactivity with further fermentations.

## 6.2 Future work

Moving forward over the traditional bioassay-guided isolation approach in natural products (NPs) drug discovery seems feasibly given today's wealth in cutting-edge technologies/platforms, software solutions and chemical libraries (Weller, 2012). Workflows with pre-fractionation, high-throughput screening (HTS) and computer-assisted structure elucidation are becoming the routine of choice (Harvey et al., 2015) of every NP researcher. Furthermore, the genome sequencing on *M. ruestringensis* provided support to identify and understand the possible biosynthetic gene cluster (BGC) encoding the biosynthetic pathway to produce specialized metabolites, associated with the genome mining approach by information extraction from genome sequences. Further research is needed about structure activity relationship on branched hydroxylated fatty acids derivatives from *M. ruestringensis* strains to evaluate the effect of the elongation, hydroxylation and/or methylation of the chain in terms of their inhibition ability against PTP1B. Additionally, is needed future exhaustive study and identification about the BGC responsible for the hydroxylation, elongation and methylation of the chain, with further laboratory gene cloning, knock-out and hydroxylation experiments to validate the theoretical predicted studies.

Regarding the *Micromonospora* sp. (SBT687) additional experiments could be performed to determine the interaction with terpenes and alkaloids extracted from this strain, to confirm the accurate presence of insecticidal activity, as well as SAR investigation concerning the insecticidal activity. Moreover, structure modifications of the beta-carboline moiety containing urea with manzamine congeners can be in future studies developed to evaluate the potential insecticidal activity against sea lice. Although, *Micromonospora* sp. (SBT692) evidenced unstable production of secondary metabolites with loss of bioactivity during the scale-up process, new small scale-up fermentations could be repeated using a variation of culture conditions to verify stability of the metabolites production.

## List of References

- ALI, M. Y., KIM, D. H., SEONG, S. H., KIM, H. R., JUNG, H. A. & CHOI, J. S. 2017. alpha-Glucosidase and Protein Tyrosine Phosphatase 1B Inhibitory Activity of Plastoquinones from Marine Brown Alga *Sargassum serratifolium*. *Mar Drugs*, 15.
- ALTARABEEN, M., DALETOS, G., EBRAHIM, W., MULLER, W. E., HARTMANN, R., LIN, W. & PROKSCH, P. 2015. Ircinal E, a New Manzamine Derivative from the Indonesian Marine Sponge *Acanthostrongylophora ingens*. *Nat Prod Commun*, 10, 1951-3.
- AMIRKIA, V. & HEINRICH, M. 2015. Natural products and drug discovery: a survey of stakeholders in industry and academia. *Front Pharmacol*, 6, 237.
- ANAND, S., PRASAD, M. V., YADAV, G., KUMAR, N., SHEHARA, J., ANSARI, M. Z. & MOHANTY, D. 2010. SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res*, 38, W487-96.
- ANDERSSON, R., GREF, R. & LUNDGREN, L. N. 1990. Manoyl oxide acid from resin of *Pinus sylvestris* needles. *Phytochemistry*, 29, 1320-1322.
- ANG, K. K., HOLMES, M. J., HIGA, T., HAMANN, M. T. & KARA, U. A. 2000. In vivo antimalarial activity of the beta-carboline alkaloid manzamine A. *Antimicrob Agents Chemother*, 44, 1645-9.
- ASHOK, P., GANGULY, S. & MURUGESAN, S. 2014. Manzamine alkaloids: isolation, cytotoxicity, antimalarial activity and SAR studies. *Drug Discov Today*, 19, 1781-1791.
- AWANG, K., ABDULLAH, N. H., HADI, A. H. A. & SU FONG, Y. 2012. Cardiovascular Activity of Labdane Diterpenes from *Andrographis paniculata* in Isolated Rat Hearts. *Journal of Biomedicine and Biotechnology*, 2012, 5.
- BACHMANN, B. O. 2010. Biosynthesis: is it time to go retro? *Nat Chem Biol*, 6, 390-3.
- BACHMANN, B. O., VAN LANEN, S. G. & BALTZ, R. H. 2014. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol*, 41, 175-84.
- BAKER, D. D., CHU, M., OZA, U. & RAJGARHIA, V. 2007. The value of natural products to future pharmaceutical discovery. *Nat Prod Rep*, 24, 1225-44.
- BAKER, G. C., GAFFAR, S., COWAN, D. A. & SUHARTO, A. R. 2001. Bacterial community analysis of Indonesian hot springs. *FEMS Microbiol Lett*, 200, 103-9.
- BANERJEE, P., EREHMAN, J., GOHLKE, B. O., WILHELM, T., PREISSNER, R. & DUNKEL, M. 2015. Super Natural II--a database of natural products. *Nucleic Acids Res*, 43, D935-9.
- BARAN, R., KOCHI, H., SAITO, N., SUEMATSU, M., SOGA, T., NISHIOKA, T., ROBERT, M. & TOMITA, M. 2006. MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics*, 7, 530.
- BARANASIC, D., ZUCKO, J., DIMINIC, J., GACESA, R., LONG, P. F., CULLUM, J., HRANUELI, D. & STARCEVIC, A. 2014. Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J Ind Microbiol Biotechnol*, 41, 461-7.
- BARDYSHEV, I. I., DEGTYARENKO, A. S. & PEKHK, T. I. 1982. 8,13-Epoxyabd-14-en-19-oic acid — A component of the needles of *Pinus sylvestris*. *Chemistry of Natural Compounds*, 18, 447-449.
- BECKER, S., KORTZ, L., HELMSCHRODT, C., THIERY, J. & CEGLAREK, U. 2012. LC-MS-based metabolomics in the clinical laboratory. *J Chromatogr B Analyt Technol Biomed Life Sci*, 883-884, 68-75.

- BENTO, A. P., GAULTON, A., HERSEY, A., BELLIS, L. J., CHAMBERS, J., DAVIES, M., KRUGER, F. A., LIGHT, Y., MAK, L., MCGLINCHEY, S., NOWOTKA, M., PAPADATOS, G., SANTOS, R. & OVERINGTON, J. P. 2014. The ChEMBL bioactivity database: an update. *Nucleic Acids Res*, 42, D1083-90.
- BERDY, J. 2005. Bioactive microbial metabolites. *J Antibiot (Tokyo)*, 58, 1-26.
- BERNARDET, J.-F. & NAKAGAWA, Y. 2006. An Introduction to the Family Flavobacteriaceae. In: DWORKIN, M., FALKOW, S., ROSENBERG, E., SCHLEIFER, K.-H. & STACKEBRANDT, E. (eds.) *The Prokaryotes: Volume 7: Proteobacteria: Delta, Epsilon Subclass*. New York, NY: Springer New York.
- BEUTLER, J. A., CARDELLINA, J. H., PRATHER, T., SHOEMAKER, R. H., BOYD, M. R. & SNADER, K. M. 1993. A Cytotoxic  $\beta$ -Carboline from the Bryozoan *Catenicella cribraria*. *Journal of Natural Products*, 56, 1825-1826.
- BHATNAGAR, I. & KIM, S. K. 2010. Immense essence of excellence: marine microbial bioactive compounds. *Mar Drugs*, 8, 2673-701.
- BLIN, K., MEDEMA, M. H., KAZEMPOUR, D., FISCHBACH, M. A., BREITLING, R., TAKANO, E. & WEBER, T. 2013. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*, 41, W204-12.
- BLUNT, J. & MUNRO, M. 2012. MarinLit database. *University of Canterbury*.
- BOARD, P. C. G. E. 2019. PDQ Genetics of Colorectal Cancer. *PDQ Cancer Information Summaries*. Bethesda, MD: National Cancer Institute.
- BOCHNER, B. R., SIRI, M., HUANG, R. H., NOBLE, S., LEI, X. H., CLEMONS, P. A. & WAGNER, B. K. 2011. Assay of the multiple energy-producing pathways of mammalian cells. *PLoS One*, 6, e18147.
- BODDY, C. N. 2014. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J Ind Microbiol Biotechnol*, 41, 443-50.
- BODE, H. B., BETHE, B., HOFES, R. & ZEECK, A. 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem*, 3, 619-27.
- BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-20.
- BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A. & JEMAL, A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68, 394-424.
- BRISTOW, M. R., GINSBURG, R., STROSBERG, A., MONTGOMERY, W. & MINOBE, W. 1984. Pharmacology and inotropic potential of forskolin in the human heart. *J Clin Invest*, 74, 212-23.
- BRUNS, A., ROHDE, M. & BERTHE-CORTI, L. 2001. *Muricauda ruestringensis* gen. nov., sp. nov., a facultatively anaerobic, appendaged bacterium from German North Sea intertidal sediment. *Int J Syst Evol Microbiol*, 51, 1997-2006.
- BUCKINGHAM, J. 1997. *Dictionary of Natural Products, Supplement 4*, CRC press.
- BUCKWALTER, B. L., BURFITT, I. R., NAGEL, A. A., WENKERT, E. & NÄF, F. 1975. <sup>13</sup>C-NMR. Spectroscopy of Naturally Occurring Substances. XXXV. Labdanic diterpenes. *Helvetica Chimica Acta*, 58, 1567-1573.
- BYLESJÖ, M., RANTALAINEN, M., CLOAREC, O., NICHOLSON, J. K., HOLMES, E. & TRYGG, J. 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics*, 20, 341-351.
- CABOCHE, S., PUPIN, M., LECLERE, V., FONTAINE, A., JACQUES, P. & KUCHEROV, G. 2008. NORINE: a database of nonribosomal peptides. *Nucleic Acids Res*, 36, D326-31.

- CACHO, R. A., TANG, Y. & CHOOI, Y. H. 2014. Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Front Microbiol*, 5, 774.
- CAI, Q., HUANG, H., QIAN, D., CHEN, K., LUO, J., TIAN, Y., LIN, T. & LIN, T. 2013. 13-methyltetradecanoic acid exhibits anti-tumor activity on T-cell lymphomas in vitro and in vivo by down-regulating p-AKT and activating caspase-3. *PLoS One*, 8, e65308.
- CANGANELLA, F. & WIEGEL, J. 2014. Anaerobic thermophiles. *Life (Basel)*, 4, 77-104.
- CAPRIOLI, J. & SEARS, M. 1983. Forskolin lowers intraocular pressure in rabbits, monkeys, and man. *Lancet*, 1, 958-60.
- CARBALLEIRA, N. M., CRUZ, H. & AYALA, N. L. 2002. Total synthesis of 2-methoxy-14-methylpentadecanoic acid and the novel 2-methoxy-14-methylhexadecanoic acid identified in the sponge *Agelas dispar*. *Lipids*, 37, 1033-7.
- CARMONA-ANTOÑANZAS, G., HUMBLE, J. L., CARMICHAEL, S. N., HEUMANN, J., CHRISTIE, H. R. L., GREEN, D. M., BASSETT, D. I., BRON, J. E. & STURM, A. 2016. Time-to-response toxicity analysis as a method for drug susceptibility assessment in salmon lice. *Aquaculture*, 464, 570-575.
- CARRO, L., NOUIOUI, I., SANGAL, V., MEIER-KOLTHOFF, J. P., TRUJILLO, M. E., MONTERO-CALASANZ, M. D. C., SAHIN, N., SMITH, D. L., KIM, K. E., PELUSO, P., DESHPANDE, S., WOYKE, T., SHAPIRO, N., KYRPIDES, N. C., KLENK, H.-P., GÖKER, M. & GOODFELLOW, M. 2018. Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential. *Scientific Reports*, 8, 525.
- CEBULA, R. E., BLANCHARD, J. L., BOISCLAIR, M. D., PAL, K. & BOCKOVICH, N. J. 1997. Synthesis and phosphatase inhibitory activity of analogs of sulfircin. *Bioorganic & Medicinal Chemistry Letters*, 7, 2015-2020.
- CHAI, Y., WENG, G., SHEN, S., SUN, C. & PAN, Y. 2015. The Protonation Site of para-Dimethylaminobenzoic Acid Using Atmospheric Pressure Ionization Methods. *Journal of The American Society for Mass Spectrometry*, 26, 668-676.
- CHARAN, R. D., SCHLINGMANN, G., JANSO, J., BERNAN, V., FENG, X. & CARTER, G. T. 2004. Diazepinomicin, a new antimicrobial alkaloid from a marine *Micromonospora* sp. *J Nat Prod*, 67, 1431-3.
- CHATWICHEN, J., BASU, S., MURPHY, M. E., HAMANN, M. T. & WINKLER, J. D. 2015. Design, Synthesis and Biological Evaluation of  $\beta$ -Carboline Dimers Based on the Structure of Neokauluamine. *Tetrahedron letters*, 56, 3515-3517.
- CHEN, C. Y. 2011. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One*, 6, e15939.
- CHEN, J., LI, W., YAO, H. & XU, J. 2015. Insights into drug discovery from natural products through structural modification. *Fitoterapia*, 103, 231-41.
- CIMERMANCIC, P., MEDEMA, M. H., CLAESEN, J., KURITA, K., WIELAND BROWN, L. C., MAVROMMATIS, K., PATI, A., GODFREY, P. A., KOEHRSEN, M., CLARDY, J., BIRREN, B. W., TAKANO, E., SALI, A., LININGTON, R. G. & FISCHBACH, M. A. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158, 412-421.
- CONWAY, K. R. & BODDY, C. N. 2013. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res*, 41, D402-7.
- CRAGG, G. M. & NEWMAN, D. J. 2013. Natural products: a continuing source of novel drug leads. *Biochimica et biophysica acta*, 1830, 3670-3695.
- CREYDT, M. & FISCHER, M. 2017. Plant Metabolomics: Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Nontargeted Mass Spectrometry

- in Positive and Negative Electrospray Ionization Mode. *Analytical Chemistry*, 89, 10474-10486.
- CUMMINS, C. S. & MOSS, C. W. 1990. Fatty acid composition of *Propionibacterium propionicum* (*Arachnia propionica*). *Int J Syst Bacteriol*, 40, 307-8.
- DALMASO, G. Z., FERREIRA, D. & VERMELHO, A. B. 2015. Marine extremophiles: a source of hydrolases for biotechnological applications. *Mar Drugs*, 13, 1925-65.
- DANILENKO, M., WANG, Q., WANG, X., LEVY, J., SHARONI, Y. & STUDZINSKI, G. P. 2003. Carnosic acid potentiates the antioxidant and prodifferentiation effects of 1 $\alpha$ ,25-dihydroxyvitamin D<sub>3</sub> in leukemia cells but does not promote elevation of basal levels of intracellular calcium. *Cancer Res*, 63, 1325-32.
- DAS, S., WARD, L. R. & BURKE, C. 2008. Prospects of using marine actinobacteria as probiotics in aquaculture. *Appl Microbiol Biotechnol*, 81, 419-29.
- DE CARVALHO, C. C. & FERNANDES, P. 2010. Production of metabolites as bacterial responses to the marine environment. *Mar Drugs*, 8, 705-27.
- DE JONG, A., VAN HEEL, A. J., KOK, J. & KUIPERS, O. P. 2010. BAGEL2: mining for bacteriocins in genomic data. *Nucleic Acids Res*, 38, W647-51.
- DE JONG, A., VAN HIJUM, S. A., BIJLSMA, J. J., KOK, J. & KUIPERS, O. P. 2006. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res*, 34, W273-9.
- DE ROSA, M., PACE, U., REGA, D., COSTABILE, V., DURATURO, F., IZZO, P. & DELRIO, P. 2015. Genetics, diagnosis and management of colorectal cancer (Review). *Oncol Rep*, 34, 1087-96.
- DE SOUZA, N. J., DOHADWALLA, A. N. & REDEN, J. 1983. Forskolin: a labdane diterpenoid with antihypertensive, positive inotropic, platelet aggregation inhibitory, and adenylate cyclase activating properties. *Med Res Rev*, 3, 201-19.
- DIAS, D. A., URBAN, S. & ROESSNER, U. 2012. A historical overview of natural products in drug discovery. *Metabolites*, 2, 303-36.
- DICKSCHAT, J. S., BODE, H. B., KROPPESTEDT, R. M., MULLER, R. & SCHULZ, S. 2005. Biosynthesis of iso-fatty acids in myxobacteria. *Org Biomol Chem*, 3, 2824-31.
- DIMASI, J. A., FELDMAN, L., SECKLER, A. & WILSON, A. 2010. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*, 87, 272-7.
- DIMINIC, J., ZUCKO, J., RUZIC, I. T., GACESA, R., HRANUELI, D., LONG, P. F., CULLUM, J. & STARCEVIC, A. 2013. Databases of the thiotemplate modular systems (CSDB) and their in silico recombinants (r-CSDB). *J Ind Microbiol Biotechnol*, 40, 653-9.
- DONA, A. C., KYRIAKIDES, M., SCOTT, F., SHEPHARD, E. A., VARSHAVI, D., VESELKOV, K. & EVERETT, J. R. 2016. A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Comput Struct Biotechnol J*, 14, 135-53.
- DONKOR, E. S. 2013. Sequencing of bacterial genomes: principles and insights into pathogenesis and development of antibiotics. *Genes*, 4, 556-572.
- DUAN, H., TAKAISHI, Y., MOMOTA, H., OHMOTO, Y., TAKI, T., JIA, Y. & LI, D. 1999. Immunosuppressive Diterpenoids from *Tripterygium wilfordii*. *Journal of Natural Products*, 62, 1522-1525.
- DUMAS, A. 2016. Approaches for controlling sea lice infestation in global salmon farming: What is applicable in Canadian aquaculture industry AU - Yossa, Rodrigue. *Journal of Applied Aquaculture*, 28, 314-329.
- DUNN, W. B., BAILEY, N. J. & JOHNSON, H. E. 2005. Measuring the metabolome: current analytical technologies. *Analyst*, 130, 606-25.

- DUNN, W. B., ERBAN, A., WEBER, R. J. M., CREEK, D. J., BROWN, M., BREITLING, R., HANKEMEIER, T., GOODACRE, R., NEUMANN, S., KOPKA, J. & VIANT, M. R. 2013. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, 9, 44-66.
- DUPONT, S., CARRÉ-MLOUKA, A., DESCARREGA, F., ERESKOVSKY, A., LONGEON, A., MOURAY, E., FLORENT, I. & BOURGUET-KONDRACKI, M. L. 2014. Diversity and biological activities of the bacterial community associated with the marine sponge *Phorbas tenacior* (Porifera, Demospongiae). *Letters in Applied Microbiology*, 58, 42-52.
- DUVALL, J. R., WU, F. & SNIDER, B. B. 2006. Structure reassignment and synthesis of Jenamidines A1/A2, synthesis of (+)-NP25302, and formal synthesis of SB-311009 analogues. *J Org Chem*, 71, 8579-90.
- EDER, J. & HERRLING, P. L. 2016. Trends in Modern Drug Discovery. *Handb Exp Pharmacol*, 232, 3-22.
- EDRADA, R. A., PROKSCH, P., WRAY, V., WITTE, L., MULLER, W. E. & VAN SOEST, R. W. 1996. Four new bioactive manzamine-type alkaloids from the Philippine marine sponge *Xestospongia ashmorica*. *J Nat Prod*, 59, 1056-60.
- EGAN, L. J. & TORUNER, M. 2006. NF-kappaB signaling: pros and cons of altering NF-kappaB as a therapeutic approach. *Ann N Y Acad Sci*, 1072, 114-22.
- EL-DEREDY, W. 1997. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. *NMR Biomed*, 10, 99-124.
- EL SAYED, K. A., KELLY, M., KARA, U. A. K., ANG, K. K. H., KATSUYAMA, I., DUNBAR, D. C., KHAN, A. A. & HAMANN, M. T. 2001. New Manzamine Alkaloids with Potent Activity against Infectious Diseases. *Journal of the American Chemical Society*, 123, 1804-1808.
- ELCHEBLY, M., PAYETTE, P., MICHALISZYN, E., CROMLISH, W., COLLINS, S., LOY, A. L., NORMANDIN, D., CHENG, A., HIMMS-HAGEN, J., CHAN, C. C., RAMACHANDRAN, C., GRESSER, M. J., TREMBLAY, M. L. & KENNEDY, B. P. 1999. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science*, 283, 1544-8.
- ELDRIDGE, G. R., VERVOORT, H. C., LEE, C. M., CREMIN, P. A., WILLIAMS, C. T., HART, S. M., GOERING, M. G., O'NEIL-JOHNSON, M. & ZENG, L. 2002. High-Throughput Method for the Production and Analysis of Large Natural Product Libraries for Drug Discovery. *Analytical Chemistry*, 74, 3963-3971.
- ELLEUCH, L., SHAABAN, K., ABELAZIZ, M., CHAKCHOUK-MTIBAA, A., NAGIA, M., MELLOULI, L. & SHAABAN, M. 2012. *Cyclic lipopeptides and other bioactive secondary metabolites from a new terrestrial Streptomyces sp. TN272*.
- ELLIGER, C. A., ZINKEL, D. F., CHAN, B. G. & WAISS, A. C., JR. 1976. Diterpene acids as larval growth inhibitors. *Experientia*, 32, 1364-6.
- ERIKSSON, L., ROSEN, J., JOHANSSON, E. & TRYGG, J. 2012. Orthogonal PLS (OPLS) Modeling for Improved Analysis and Interpretation in Drug Design. *Mol Inform*, 31, 414-9.
- EVA-MARIA UNSIN, C., RAJSKI, S. & SHEN, B. 2013. The Role of Genetic Engineering in Natural Product-Based Anticancer Drug Discovery.
- EXARCHOU, V., KRUCKER, M., VAN BEEK, T. A., VERVOORT, J., GEROTHANASSIS, I. P. & ALBERT, K. 2005. LC-NMR coupling technology: recent advancements and applications in natural products analysis. *Magn Reson Chem*, 43, 681-7.
- FAUNG, S. T., CHIU, L. & WANG, C. T. 1996. Platelet lysis and functional perturbation by 13-methyl myristate. The major fatty acid in *Flavobacterium ranacida*. *Thromb Res*, 81, 91-100.

- FEHMIDA, B., MUHAMMAD, F., ESAM, I. A., MUHAMMAD, Y., SANA, A. A., MOHAMMAD, A. K., IKRAM, U. & MUHAMMAD, I. N. 2017. Bacteria From Marine Sponges: A Source of New Drugs. *Current Drug Metabolism*, 18, 11-15.
- FEISZT, P., MESTYAN, G., KERENYI, M., DOBAY, O., SZABO, J., DOMBRADI, Z., URBAN, E. & EMODY, L. 2014. Re-evaluation of in vitro activity of primycin against prevalent multiresistant bacteria. *Int J Med Microbiol*, 304, 1077-85.
- FITZMAURICE, C., ALLEN, C., BARBER, R. M., BARREGARD, L., BHUTTA, Z. A., BRENNER, H., DICKER, D. J., CHIMED-ORCHIR, O., DANDONA, R., DANDONA, L., FLEMING, T., FOROUZANFAR, M. H., HANCOCK, J., HAY, R. J., HUNTER-MERRILL, R., HUYNH, C., HOSGOOD, H. D., JOHNSON, C. O., JONAS, J. B., KHUBCHANDANI, J., KUMAR, G. A., KUTZ, M., LAN, Q., LARSON, H. J., LIANG, X., LIM, S. S., LOPEZ, A. D., MACINTYRE, M. F., MARCZAK, L., MARQUEZ, N., MOKDAD, A. H., PINHO, C., POURMALEK, F., SALOMON, J. A., SANABRIA, J. R., SANDAR, L., SARTORIUS, B., SCHWARTZ, S. M., SHACKELFORD, K. A., SHIBUYA, K., STANAWAY, J., STEINER, C., SUN, J., TAKAHASHI, K., VOLLSET, S. E., VOS, T., WAGNER, J. A., WANG, H., WESTERMAN, R., ZEEB, H., ZOECKLER, L., ABD-ALLAH, F., AHMED, M. B., ALABED, S., ALAM, N. K., ALDHAHRI, S. F., ALEM, G., ALEMAYOHU, M. A., ALI, R., AL-RADDADI, R., AMARE, A., AMOAKO, Y., ARTAMAN, A., ASAYESH, H., ATNAFU, N., AWASTHI, A., SALEEM, H. B., BARAC, A., BEDI, N., BENSENOR, I., BERHANE, A., BERNABE, E., BETSU, B., BINAGWAHO, A., BONEYA, D., CAMPOS-NONATO, I., CASTANEDA-ORJUOLA, C., CATALA-LOPEZ, F., CHIANG, P., CHIBUEZE, C., CHITHEER, A., CHOI, J. Y., COWIE, B., DAMTEW, S., DAS NEVES, J., DEY, S., DHARMARATNE, S., DHILLON, P., DING, E., DRISCOLL, T., EKWUEME, D., ENDRIES, A. Y., FARVID, M., FARZADFAR, F., FERNANDES, J., FISCHER, F., TT, G. H., GEBRU, A., GOPALANI, S., HAILU, A., et al. 2017. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol*, 3, 524-548.
- FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & ET AL. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269, 496-512.
- FRAGOSO-SERRANO, M., GONZALEZ-CHIMEO, E. & PEREDA-MIRANDA, R. 1999. Novel labdane diterpenes from the insecticidal plant *hyptis spicigera*1. *J Nat Prod*, 62, 45-50.
- FUKUSHIMA, K., YAZAWA, K. & ARAI, T. 1973. Biological activities of albonoursin. *J Antibiot (Tokyo)*, 26, 175-6.
- FURUSATO, A., KATO, H., NEHIRA, T., EGUCHI, K., KAWABATA, T., FUJIWARA, Y., LOSUNG, F., MANGINDAAN, R. E. P., DE VOOGD, N. J., TAKEYA, M., YOKOSAWA, H. & TSUKAMOTO, S. 2014. Acanthomanzamines A–E with New Manzamine Frameworks from the Marine Sponge *Acanthostrongylophora ingens*. *Organic Letters*, 16, 3888-3891.
- GALM, U. & SHEN, B. 2007. Natural product drug discovery: the times have never been better. *Chem Biol*, 14, 1098-104.
- GAYA, M., REPETTO, V., TONEATTO, J., ANESINI, C., PIWIEN-PILIPUK, G. & MORENO, S. 2013. Antiadipogenic effect of carnosic acid, a natural compound present in *Rosmarinus officinalis*, is exerted through the C/EBPs and PPARgamma

- pathways at the onset of the differentiation program. *Biochim Biophys Acta*, 1830, 3796-806.
- GERWICK, W. H. & MOORE, B. S. 2012. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem Biol*, 19, 85-98.
- GESZTELYI, I., KONYA, L. & KOVER, A. 1980. Effect of primycin on some electric properties of the frog skeletal muscle. *Acta Physiol Acad Sci Hung*, 55, 1-11.
- GHANI, U. 2015. Re-exploring promising alpha-glucosidase inhibitors for potential development into oral anti-diabetic drugs: Finding needle in the haystack. *Eur J Med Chem*, 103, 133-62.
- GONZÁLEZ, A. G., ARTEAGA, J. M., BRETÓN, J. L. & FRAGA, B. M. 1977. Five new labdane diterpene oxides from *Eupatorium jhanii*. *Phytochemistry*, 16, 107-110.
- GRANT, A. N. 2002. Medicines for sea lice. *Pest Manag Sci*, 58, 521-7.
- GRIFFITHS, W. J., KOAL, T., WANG, Y., KOHL, M., ENOT, D. P. & DEIGNER, H. P. 2010. Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl*, 49, 5426-45.
- GU, J., GUI, Y., CHEN, L., YUAN, G., LU, H. Z. & XU, X. 2013. Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS One*, 8, e62839.
- GULDER, T. A. & MOORE, B. S. 2009. Chasing the treasures of the sea - bacterial marine natural products. *Curr Opin Microbiol*, 12, 252-60.
- GUO, D. X., XIANG, F., WANG, X. N., YUAN, H. Q., XI, G. M., WANG, Y. Y., YU, W. T. & LOU, H. X. 2010. Labdane diterpenoids and highly methoxylated bibenzyls from the liverwort *Frullania inouei*. *Phytochemistry*, 71, 1573-8.
- GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics*, 29, 1072-5.
- GUZMÁN, E. A., JOHNSON, J. D., LINLEY, P. A., GUNASEKERA, S. E. & WRIGHT, A. E. 2011. A novel activity from an old compound: Manzamine A reduces the metastatic potential of AsPC-1 pancreatic cancer cells and sensitizes them to TRAIL-induced apoptosis. *Investigational new drugs*, 29, 777-785.
- HAEGGSTROM, J. Z. & FUNK, C. D. 2011. Lipxygenase and leukotriene pathways: biochemistry, biology, and roles in disease. *Chem Rev*, 111, 5866-98.
- HALBEDL, S., KRATZER, M. C., RAHM, K., CROSTA, N., MASTERS, K. S., ZIPPERT, J., BRASE, S. & GRADL, D. 2013. Synthesis of novel inhibitors blocking Wnt signaling downstream of beta-catenin. *FEBS Lett*, 587, 522-7.
- HANSEN, R. P., SHORLAND, F. B. & COOKE, N. J. 1953. The branched-chain fatty acids of mutton fat. II. The isolation of (+)-12-methyltetradecanoic acid and the 13-methyltetradecanoic acid. *Biochem J*, 53, 374-8.
- HARRISON, K. A. 1989. Maternal mortality in developing countries. *Br J Obstet Gynaecol*, 96, 1119-23.
- HARVEY, A. L. 2008. Natural products in drug discovery. *Drug Discov Today*, 13, 894-901.
- HARVEY, A. L., EDRADA-EBEL, R. & QUINN, R. J. 2015. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov*, 14, 111-29.
- HASHIZUME, H., HIROSAWA, S., SAWA, R., MURAOKA, Y., IKEDA, D., NAGANAWA, H. & IGARASHI, M. 2004. Tripropeptins, novel antimicrobial agents produced by *Lysobacter* sp. *J Antibiot (Tokyo)*, 57, 52-8.
- HASHIZUME, H., IGARASHI, M., HATTORI, S., HORI, M., HAMADA, M. & TAKEUCHI, T. 2001. Tripropeptins, novel antimicrobial agents produced by *Lysobacter* sp. I. Taxonomy, isolation and biological activities. *J Antibiot (Tokyo)*, 54, 1054-9.
- HASTINGS, J., DE MATOS, P., DEKKER, A., ENNIS, M., HARSHA, B., KALE, N., MUTHUKRISHNAN, V., OWEN, G., TURNER, S., WILLIAMS, M. &

- STEINBECK, C. 2013. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res*, 41, D456-63.
- HAYS, T., RUSYN, I., BURNS, A. M., KENNETT, M. J., WARD, J. M., GONZALEZ, F. J. & PETERS, J. M. 2005. Role of peroxisome proliferator-activated receptor-alpha (PPARalpha) in bezafibrate-induced hepatocarcinogenesis and cholestasis. *Carcinogenesis*, 26, 219-27.
- HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107, 1-8.
- HEFFNER, S. A., BOVEY, F. A., VERGE, L. M., MIRAU, P. A. & TONELLI, A. 1986. Two-dimensional proton and carbon-13 NMR spectroscopy of styrene-methyl methacrylate copolymers. *Macromolecules*, 19, 1628-1634.
- HELGESEN, K. O. & HORSBERG, T. E. 2013. Single-dose field bioassay for sensitivity testing in sea lice, *Lepeophtheirus salmonis*: development of a rapid diagnostic tool. *J Fish Dis*, 36, 261-72.
- HENEBERG, P. 2009. Use of protein tyrosine phosphatase inhibitors as promising targeted therapeutic drugs. *Curr Med Chem*, 16, 706-33.
- HILL, R. T. B., MD, (US), PERAUD, OLIVIER (BALTIMORE, MD, US), HAMANN, MARK T. (OXFORD, MS, US), KASANA, NOER (UNIVERSITY, MS, US). 2005. *Manzamine-producing actinomycetes*. United States patent application 20050244938.
- HIRAMOTO, S., KINOSHITA, N., HATANAKA, S. & SETO, H. 1996. Stimulation of apolipoprotein E secretion in human hepatoma Hep G2 cells by a cyclic acylpeptide, N-4909. *J Antibiot (Tokyo)*, 49, 949-52.
- HIRSCH, A. M. & VALDÉS, M. 2010. Micromonospora: An important microbe for biomedicine and potentially for biocontrol and biofuels. *Soil Biology and Biochemistry*, 42, 536-542.
- HO, R. & SHI, Q. H. 1982. Forskolol as a novel lipolytic agent. *Biochem Biophys Res Commun*, 107, 157-64.
- HOLMES, E., NICHOLLS, A. W., LINDON, J. C., CONNOR, S. C., CONNELLY, J. C., HASELDEN, J. N., DAMMENT, S. J. P., SPRAUL, M., NEIDIG, P. & NICHOLSON, J. K. 2000. Chemometric Models for Toxicity Classification Based on NMR Spectra of Biofluids. *Chemical Research in Toxicology*, 13, 471-478.
- HON, J. D., SINGH, B., SAHIN, A., DU, G., WANG, J., WANG, V. Y., DENG, F. M., ZHANG, D. Y., MONACO, M. E. & LEE, P. 2016. Breast cancer molecular subtypes: from TNBC to QNBC. *Am J Cancer Res*, 6, 1864-1872.
- HORIKOSHI, K. 1998. Barophiles: deep-sea microorganisms adapted to an extreme environment. *Curr Opin Microbiol*, 1, 291-5.
- HU, X., TAO, N., WANG, X., XIAO, J. & WANG, M. 2016. Marine-derived bioactive compounds with anti-obesity effect: A review. *Journal of Functional Foods*, 21, 372-387.
- HUANG, Y., GUO, Z., SONG, H., LIU, Y., WANG, L. & WANG, Q. 2018. Design, Synthesis, and Biological Activity of beta-Carboline Analogues Containing Hydantoin, Thiohydantoin, and Urea Moieties. *J Agric Food Chem*, 66, 8253-8261.
- HUERTA, M., URZÚA, Z., TRUJILLO, X., GONZÁLEZ-SÁNCHEZ, R. & TRUJILLO-HERNÁNDEZ, B. 2010. Forskolol Compared with Beclomethasone for Prevention of Asthma Attacks: A Single-Blind Clinical Trial. *Journal of International Medical Research*, 38, 661-668.
- HUGHES, J. P., REES, S., KALINDJIAN, S. B. & PHILPOTT, K. L. 2011. Principles of early drug discovery. *Br J Pharmacol*, 162, 1239-49.
- HUNTEMANN, M., TESHIMA, H., LAPIDUS, A., NOLAN, M., LUCAS, S., HAMMON, N., DESHPANDE, S., CHENG, J. F., TAPIA, R., GOODWIN, L. A., PITLUCK, S.,

- LIOLIOS, K., PAGANI, I., IVANOVA, N., MAVROMATIS, K., MIKHAILOVA, N., PATI, A., CHEN, A., PALANIAPPAN, K., LAND, M., HAUSER, L., PAN, C., BRAMBILLA, E. M., ROHDE, M., SPRING, S., GOKER, M., DETTER, J. C., BRISTOW, J., EISEN, J. A., MARKOWITZ, V., HUGENHOLTZ, P., KYRPIDES, N. C., KLENK, H. P. & WOYKE, T. 2012. Complete genome sequence of the facultatively anaerobic, appendaged bacterium *Muricauda ruestringensis* type strain (B1(T)). *Stand Genomic Sci*, 6, 185-93.
- IBRAHIM, M. A., SHILABIN, A. G., PRASANNA, S., JACOB, M., KHAN, S. I., DOERKSEN, R. J. & HAMANN, M. T. 2008. 2-N-Methyl modifications and SAR studies of manzamine A. *Bioorganic & medicinal chemistry*, 16, 6702-6706.
- ICHIBA, T., CORGIAT, J. M., SCHEUER, P. J. & KELLY-BORGES, M. 1994. 8-Hydroxymanzamine A, a  $\beta$ -Carboline Alkaloid from a Sponge, *Pachypellina* sp. *Journal of Natural Products*, 57, 168-170.
- ICHIBA, T., SAKAI, R., KOHMOTO, S., SAUCY, G. & HIGA, T. 1988. New manzamine alkaloids from a sponge of the genus *xestospongia*. *Tetrahedron Letters*, 29, 3083-3086.
- ICHIKAWA, N., SASAGAWA, M., YAMAMOTO, M., KOMAKI, H., YOSHIDA, Y., YAMAZAKI, S. & FUJITA, N. 2013. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res*, 41, D408-14.
- IGARASHI, Y., TRUJILLO, M. E., MARTINEZ-MOLINA, E., YANASE, S., MIYANAGA, S., OBATA, T., SAKURAI, H., SAIKI, I., FUJITA, T. & FURUMAI, T. 2007. Antitumor anthraquinones from an endophytic actinomycete *Micromonospora lupini* sp. nov. *Bioorg Med Chem Lett*, 17, 3702-5.
- IGNEA, C., IOANNOU, E., GEORGANTEA, P., TRIKKA, F. A., ATHANASAKOGLU, A., LOUPASSAKI, S., ROUSSIS, V., MAKRIS, A. M. & KAMPRANIS, S. C. 2016. Production of the forskolin precursor 11 $\beta$ -hydroxy-manoyl oxide in yeast using surrogate enzymatic activities. *Microbial Cell Factories*, 15, 46.
- IMHOFF, J. F., LABES, A. & WIESE, J. 2011. Bio-mining the microbial treasures of the ocean: new natural products. *Biotechnol Adv*, 29, 468-82.
- IORIZZI, M., DE RICCARDIS, F., MINALE, L. & RICCIO, R. 1993. Starfish Saponins, 52. Chemical Constituents from the Starfish *Echinaster brasiliensis*. *Journal of Natural Products*, 56, 2149-2162.
- ISMET, A., VIKINESWARY, S., PARAMASWARI, S., WONG, W. H., WARD, A., SEKI, T., FIEDLER, H. P. & GOODFELLOW, M. 2004. Production and Chemical Characterization of Antifungal Metabolites From *Micromonospora* sp. M39 Isolated From Mangrove Rhizosphere Soil. *World Journal of Microbiology and Biotechnology*, 20, 523-528.
- IWAI, K., AISAKA, K. & SUZUKI, M. 2010. *Friedmanniella luteola* sp. nov., *Friedmanniella lucida* sp. nov., *Friedmanniella okinawensis* sp. nov. and *Friedmanniella sagamiharensis* sp. nov., isolated from spiders. *Int J Syst Evol Microbiol*, 60, 113-20.
- J BRANSON, E., S RØNSBERG, S. & RITCHIE, G. 2000. *Efficacy of teflubenzuron (CalicideR) for the treatment of sea lice, Lepeophtheirus salmonis (Kroyer 1838), infestations of farmed Atlantic salmon (Salmo salar L.)*.
- JANG, K. H., JEON, J. E., RYU, S., LEE, H. S., OH, K. B. & SHIN, J. 2008. Polyoxygenated diterpenes from the sponge *Phorbas* sp. *J Nat Prod*, 71, 1701-7.
- JENSEN, P. R., MINCER, T. J., WILLIAMS, P. G. & FENICAL, W. 2005. Marine actinomycete diversity and natural product discovery. *Antonie Van Leeuwenhoek*, 87, 43-8.
- JIANG, C. S., LIANG, L. F. & GUO, Y. W. 2012. Natural products possessing protein tyrosine phosphatase 1B (PTP1B) inhibitory activity found in the last decades. *Acta Pharmacol Sin*, 33, 1217-45.

- JIMENEZ, D. F., REVIE, C. W., HARDY, S. P., JANSEN, P. A. & GETTINBY, G. 2013. Multivariate evaluation of the effectiveness of treatment efficacy of cypermethrin against sea lice (*Lepeophtheirus salmonis*) in Atlantic salmon (*Salmo salar*). *BMC Vet Res*, 9, 258.
- KANEDA, T. 1991. Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev*, 55, 288-302.
- KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28, 27-30.
- KATAJAMAA, M. & ORESIC, M. 2005. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics*, 6, 179.
- KHALDI, N., SEIFUDDIN, F. T., TURNER, G., HAFT, D., NIERMAN, W. C., WOLFE, K. H. & FEDOROVA, N. D. 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol*, 47, 736-41.
- KHARE, S. & VERMA, M. 2012. Epigenetics of colon cancer. *Methods Mol Biol*, 863, 177-85.
- KHATER, S., ANAND, S. & MOHANTY, D. 2016. In silico methods for linking genes and secondary metabolites: The way forward. *Synth Syst Biotechnol*, 1, 80-88.
- KHAYATT, B. I., OVERMARS, L., SIEZEN, R. J. & FRANCKE, C. 2013. Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. *PLoS One*, 8, e62136.
- KICHA, A. A., KAPUSTINA, I. I., IVANCHINA, N. V., KALINOVSKY, A. I., DMITRENOK, P. S., STONIK, V. A., PAL'YANOVA, N. V., PANKOVA, T. M. & STAROSTINA, M. V. 2008. Polyhydroxylated steroid compounds from the Far Eastern starfish *Distolasterias nipon*. *Russian Journal of Bioorganic Chemistry*, 34, 118-124.
- KIM, C.-K., RISWANTO, R., WON, T. H., KIM, H., ELYA, B., SIM, C. J., OH, D.-C., OH, K.-B. & SHIN, J. 2017a. Manzamine Alkaloids from an *Acanthostrongylophora* sp. Sponge. *Journal of Natural Products*, 80, 1575-1583.
- KIM, J. M., JIN, H. M. & JEON, C. O. 2013. *Muricaudataeanensis* sp. nov., isolated from a marine tidal flat. *International journal of systematic and evolutionary microbiology*, 63, 2672-2677.
- KIM, N. R., JEONG, D. W., KO, D. S. & SHIM, J. H. 2017b. Characterization of novel thermophilic alpha-glucosidase from *Bifidobacterium longum*. *Int J Biol Macromol*, 99, 594-599.
- KLAMAN, L. D., BOSS, O., PERONI, O. D., KIM, J. K., MARTINO, J. L., ZABOLOTNY, J. M., MOGHAL, N., LUBKIN, M., KIM, Y. B., SHARPE, A. H., STRICKER-KRONGRAD, A., SHULMAN, G. I., NEEL, B. G. & KAHN, B. B. 2000. Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Mol Cell Biol*, 20, 5479-89.
- KO, J. H., LEE, S. G., YANG, W. M., UM, J. Y., SETHI, G., MISHRA, S., SHANMUGAM, M. K. & AHN, K. S. 2018. The Application of Embelin for Cancer Prevention and Therapy. *Molecules*, 23.
- KOBAYASHI, J. I., TSUDA, M., KAWASAKI, N., SASAKI, T. & MIKAMI, Y. 1994. 6-Hydroxymanzamine A and 3,4-Dihydromanzamine A, New Alkaloids from the Okinawan Marine Sponge *Amphimedon* sp. *Journal of Natural Products*, 57, 1737-1740.
- KOBAYASHI, M., CHEN, Y.-J., AOKI, S., IN, Y., ISHIDA, T. & KITAGAWA, I. 1995. Four new  $\beta$ -carboline alkaloids isolated from two Okinawan marine sponges of *Xestospongia* sp. and *Haliclona* sp.). *Tetrahedron*, 51, 3727-3736.

- KOBAYASHI, M. & MITSUHASHI, H. 1982. Marine sterols. XIII. Isolation and synthesis of 1 beta,3 beta,5,6 beta-tetrahydroxy-5 alpha-androstan-17-one from the soft coral *Sarcophyton glaucum*. *Steroids*, 40, 673-7.
- KOHAMA, Y., IIDA, K., SEMBA, T., MIMURA, T., INADA, A., TANAKA, K. & NAKANISHI, T. 1992. Studies on Thermophile Products. IV. Structural Elucidation of Cytotoxic Substance, BS-1, Derived from *Bacillus stearothermophilus*. *CHEMICAL & PHARMACEUTICAL BULLETIN*, 40, 2210-2211.
- KOHDA, H., MIZUTANI, K. & TANAKA, O. 1983. Chemical Constituents of Flowers of *Stevia rebaudiana* Bertoni AU - Darise, Muchsin. *Agricultural and Biological Chemistry*, 47, 133-135.
- KOHLI, P., NAYYAR, N., SHARMA, A., SINGH, A. K. & LAL, R. 2016. *Algoriphagus roseus* sp. nov., isolated from a hexachlorocyclohexane-contaminated dumpsite. *Int J Syst Evol Microbiol*, 66, 3558-3565.
- KONDO, K., SHIGEMORI, H., KIKUCHI, Y., ISHIBASHI, M., SASAKI, T. & KOBAYASHI, J. 1992. Ircinal A and B from the Okinawan marine sponge *Ircinia* sp.: plausible biogenetic precursors of manzamine alkaloids. *The Journal of Organic Chemistry*, 57, 2480-2483.
- KRISHNAMURTHY, N. & KURZROCK, R. 2018. Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. *Cancer Treat Rev*, 62, 50-60.
- KUBOTA, T., NAKAMURA, K., KURIMOTO, S.-I., SAKAI, K., FROMONT, J., GONOI, T. & KOBAYASHI, J. I. 2017. Zamamidine D, a Manzamine Alkaloid from an Okinawan Amphimedon sp. Marine Sponge. *Journal of Natural Products*, 80, 1196-1199.
- KUMAR, A., TAKADA, Y., BORIEK, A. M. & AGGARWAL, B. B. 2004. Nuclear factor-kappaB: its role in health and disease. *J Mol Med (Berl)*, 82, 434-48.
- KUMAR, C. G., HIMABINDU, M. & JETTY, A. 2008. Microbial biosynthesis and applications of gentamicin: a critical appraisal. *Crit Rev Biotechnol*, 28, 173-212.
- KUMAR, S. V., SARAVANAN, D., KUMAR, B. & JAYAKUMAR, A. 2014. An update on prodrugs from natural products. *Asian Pac J Trop Med*, 7s1, S54-9.
- KUNZE, B., BEDORF, N., KOHL, W., HOFLE, G. & REICHENBACH, H. 1989. Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (Myxobacterales). Production, isolation, physico-chemical and biological properties. *J Antibiot (Tokyo)*, 42, 14-7.
- KWONG, J. C., MCCALLUM, N., SINTCHENKO, V. & HOWDEN, B. P. 2015. Whole genome sequencing in clinical and public health microbiology. *Pathology*, 47, 199-210.
- LAM, K. S. 2007. New aspects of natural products in drug discovery. *Trends Microbiol*, 15, 279-89.
- LAVILLE, R., THOMAS, O. P., BERRUÉ, F., MARQUEZ, D., VACELET, J. & AMADE, P. 2009. Bioactive Guanidine Alkaloids from Two Caribbean Marine Sponges. *Journal of Natural Products*, 72, 1589-1594.
- LEVITON, A. & HARGROVE, R. E. 1952. Microbiological Synthesis of Vitamin B12 by Propionic Acid Bacteria. *Industrial & Engineering Chemistry*, 44, 2651-2655.
- LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-60.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-9.
- LI, Y., LAI, Y.-M., LU, Y., YANG, Y.-L. & CHEN, S. 2014. Analysis of the biosynthesis of antibacterial cyclic dipeptides in *Nocardiosis alba*. *Archives of Microbiology*, 196, 765-774.

- LI, Y., SONG, L. M., GUO, M. W., WANG, L. F. & LIANG, W. X. 2016. *Sphingobacterium populi* sp. nov., isolated from bark of *Populus x euramericana*. *Int J Syst Evol Microbiol*, 66, 3456-3462.
- LIN, T., YIN, X., CAI, Q., FAN, X., XU, K., HUANG, L., LUO, J., ZHENG, J. & HUANG, J. 2012. 13-Methyltetradecanoic acid induces mitochondrial-mediated apoptosis in human bladder cancer cells. *Urol Oncol*, 30, 339-45.
- LINDGREN, F., HANSEN, B., KARCHER, W., SJÖSTRÖM, M. & ERIKSSON, L. 1996. Model validation by permutation tests: Applications to variable selection. *Journal of Chemometrics*, 10, 521-532.
- LINDNER, E., DOHADWALLA, A. N. & BHATTACHARYA, B. K. 1978. Positive inotropic and blood pressure lowering activity of a diterpene derivative isolated from *Coleus forskohli*: Forskolol. *Arzneimittelforschung*, 28, 284-9.
- LINDON, J. C., NICHOLSON, J. K. & WILSON, I. D. 2000. Directly coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *J Chromatogr B Biomed Sci Appl*, 748, 233-58.
- LIU, K., LIU, Y., WANG, N., GU, Z., SHEN, L., XU, B., ZHOU, Y., LIU, H. & JIAO, N. 2016. *Hymenobacter glaciecola* sp. nov., isolated from glacier ice. *Int J Syst Evol Microbiol*, 66, 3793-3798.
- LOMMEN, A. 2009. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal Chem*, 81, 3079-86.
- LOPES, A. S., CRUZ, E. C., SUSSULINI, A. & KLASSEN, A. 2017. Metabolomic Strategies Involving Mass Spectrometry Combined with Liquid and Gas Chromatography. *Adv Exp Med Biol*, 965, 77-98.
- LOVE, D. N., BAILEY, G. D., COLLINGS, S. & BRISCOE, D. A. 1992. Description of *Porphyromonas circumdentaria* sp. nov. and reassignment of *Bacteroides salivus* (Love, Johnson, Jones, and Calverley 1987) as *Porphyromonas* (Shah and Collins 1988) *salivosa* comb. nov. *Int J Syst Bacteriol*, 42, 434-8.
- LOVE, D. N., KARJALAINEN, J., KANERVO, A., FORSBLUM, B., SARKIALA, E., BAILEY, G. D., WIGNEY, D. I. & JOUSIMIES-SOMER, H. 1994. *Porphyromonas canoris* sp. nov., an asaccharolytic, black-pigmented species from the gingival sulcus of dogs. *Int J Syst Bacteriol*, 44, 204-8.
- LU, H., WANG, J., WANG, Y., QIAO, L. & ZHOU, Y. 2016. Embelin and Its Role in Chronic Diseases. *Adv Exp Med Biol*, 928, 397-418.
- LU, J. J., BAO, J. L., WU, G. S., XU, W. S., HUANG, M. Q., CHEN, X. P. & WANG, Y. T. 2013. Quinones derived from plant secondary metabolites as anti-cancer agents. *Anticancer Agents Med Chem*, 13, 456-63.
- LU, X., MALUMBRES, R., SHIELDS, B., JIANG, X., SAROSIEK, K. A., NATKUNAM, Y., TIGANIS, T. & LOSSOS, I. S. 2008. PTP1B is a negative regulator of interleukin 4-induced STAT6 signaling. *Blood*, 112, 4098-108.
- LUCAS, X., SENGER, C., ERXLEBEN, A., GRUNING, B. A., DORING, K., MOSCH, J., FLEMMING, S. & GUNTHER, S. 2013. StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res*, 41, D1130-6.
- LUDWICZUK, A., SKALICKA-WOŹNIAK, K. & GEORGIEV, M. I. 2017. Chapter 11 - Terpenoids. In: BADAL, S. & DELGODA, R. (eds.) *Pharmacognosy*. Boston: Academic Press.
- MACINTYRE, L., ZHANG, T., VIEGELMANN, C., MARTINEZ, I. J., CHENG, C., DOWDELLS, C., ABDELMOHSEN, U. R., GERNERT, C., HENTSCHEL, U. & EDRADE-EBEL, R. 2014. Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. *Mar Drugs*, 12, 3416-48.
- MALYARENKO, T. V., KHARCHENKO, S. D., KICHA, A. A., IVANCHINA, N. V., DMITRENOK, P. S., CHINGIZOVA, E. A., PISLYAGIN, E. A., EVTUSHENKO,

- E. V., ANTOKHINA, T. I., MINH, C. V. & STONIK, V. A. 2016. Anthenosides L-U, Steroidal Glycosides with Unusual Structural Features from the Starfish *Anthenea aspera*. *J Nat Prod*, 79, 3047-3056.
- MARINELLI, F., GENILLOU, O., FEDORENKO, V. & RON, E. Z. 2015. Specialized Bioactive Microbial Metabolites: From Gene to Product. *Biomed Res Int*, 2015, 276964.
- MARKOWITZ, V. M., CHEN, I. M., PALANIAPPAN, K., CHU, K., SZETO, E., PILLAY, M., RATNER, A., HUANG, J., WOYKE, T., HUNTEMANN, M., ANDERSON, I., BILLIS, K., VARGHESE, N., MAVROMATIS, K., PATI, A., IVANOVA, N. N. & KYRPIDES, N. C. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res*, 42, D560-7.
- MARQUINA, S., PARRA, J. L., GONZÁLEZ, M., ZAMILPA, A., ESCALANTE, J., TREJO-HERNÁNDEZ, M. R. & ÁLVAREZ, L. 2009. Hydroxylation of the diterpenes ent-kaur-16-en-19-oic and ent-beyer-15-en-19-oic acids by the fungus *Aspergillus niger*. *Phytochemistry*, 70, 2017-2022.
- MARTÍN, A. S., ROVIROSA, J., BECKER, R. & CASTILLO, M. 1980. Diterpenoids from *Baccharis tola*. *Phytochemistry*, 19, 1985-1987.
- MARTINS, J. N., FIGUEIREDO, F. S., MARTINS, G. R., LEITÃO, G. G. & COSTA, F. N. 2017. Diterpenes and a new benzaldehyde from the mangrove plant *Rhizophora mangle*. *Revista Brasileira de Farmacognosia*, 27, 175-178.
- MASZENAN, A. M., JIANG, H. L., TAY, J. H., SCHUMANN, P., KROPPESTEDT, R. M. & TAY, S. T. 2007. *Granulicoccus phenolivorans* gen. nov., sp. nov., a Gram-positive, phenol-degrading coccus isolated from phenol-degrading aerobic granules. *Int J Syst Evol Microbiol*, 57, 730-7.
- MATHES, E., O'DEA, E. L., HOFFMANN, A. & GHOSH, G. 2008. NF-kappaB dictates the degradation pathway of IkappaBalpha. *Embo j*, 27, 1357-67.
- MAYER, A. M., GLASER, K. B., CUEVAS, C., JACOBS, R. S., KEM, W., LITTLE, R. D., MCINTOSH, J. M., NEWMAN, D. J., POTTS, B. C. & SHUSTER, D. E. 2010. The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol Sci*, 31, 255-65.
- MAYER, A. M., HALL, M. L., LYNCH, S. M., GUNASEKERA, S. P., SENNETT, S. H. & POMPONI, S. A. 2005. Differential modulation of microglia superoxide anion and thromboxane B2 generation by the marine manzamines. *BMC Pharmacology*, 5, 6.
- MAYER, A. M. S., RODRÍGUEZ, A. D., TAGLIALATELA-SCAFATI, O. & FUSESTANI, N. 2013. Marine Pharmacology in 2009–2011: Marine Compounds with Antibacterial, Antidiabetic, Antifungal, Anti-Inflammatory, Antiprotozoal, Antituberculosis, and Antiviral Activities; Affecting the Immune and Nervous Systems, and other Miscellaneous Mechanisms of Action. *Marine Drugs*, 11, 2510.
- MCCHESENEY, J. D., VENKATARAMAN, S. K. & HENRI, J. T. 2007. Plant natural products: back to the future or into extinction? *Phytochemistry*, 68, 2015-22.
- MCMURRAY, J., SEIDELIN, P. H. & STRUTHERS, A. D. 1989. Evidence for a proximal and distal nephron action of atrial natriuretic factor in man. *Nephron*, 51, 39-43.
- MCNALLY, M. & CAPON, R. J. 2001. Phorbacin B and C: novel diterpenes from a southern Australian marine sponge, *Phorbacin* species. *J Nat Prod*, 64, 645-7.
- MEDEMA, M. H., BLIN, K., CIMERMANCIC, P., DE JAGER, V., ZAKRZEWSKI, P., FISCHBACH, M. A., WEBER, T., TAKANO, E. & BREITLING, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*, 39, W339-46.
- MEDEMA, M. H. & FISCHBACH, M. A. 2015. Computational approaches to natural product discovery. *Nat Chem Biol*, 11, 639-48.

- MEDEMA, M. H., KOTTMANN, R., YILMAZ, P., CUMMINGS, M., BIGGINS, J. B., BLIN, K., DE BRUIJN, I., CHOOI, Y. H., CLAESEN, J., COATES, R. C., CRUZ-MORALES, P., DUDELA, S., DUSTERHUS, S., EDWARDS, D. J., FEWER, D. P., GARG, N., GEIGER, C., GOMEZ-ESCRIBANO, J. P., GREULE, A., HADJITHOMAS, M., HAINES, A. S., HELFRICH, E. J., HILLWIG, M. L., ISHIDA, K., JONES, A. C., JONES, C. S., JUNGSMANN, K., KEGLER, C., KIM, H. U., KOTTER, P., KRUG, D., MASSCHELEIN, J., MELNIK, A. V., MANTOVANI, S. M., MONROE, E. A., MOORE, M., MOSS, N., NUTZMANN, H. W., PAN, G., PATI, A., PETRAS, D., REEN, F. J., ROSCONI, F., RUI, Z., TIAN, Z., TOBIAS, N. J., TSUNEMATSU, Y., WIEMANN, P., WYCKOFF, E., YAN, X., YIM, G., YU, F., XIE, Y., AIGLE, B., APEL, A. K., BALIBAR, C. J., BALSUS, E. P., BARONA-GOMEZ, F., BECHTHOLD, A., BODE, H. B., BORRIS, R., BRADY, S. F., BRAKHAGE, A. A., CAFFREY, P., CHENG, Y. Q., CLARDY, J., COX, R. J., DE MOT, R., DONADIO, S., DONIA, M. S., VAN DER DONK, W. A., DORRESTEIN, P. C., DOYLE, S., DRIESSEN, A. J., EHLING-SCHULZ, M., ENTIAN, K. D., FISCHBACH, M. A., GERWICK, L., GERWICK, W. H., GROSS, H., GUST, B., HERTWECK, C., HOFTE, M., JENSEN, S. E., JU, J., KATZ, L., KAYSSER, L., KLASSEN, J. L., KELLER, N. P., KORMANEC, J., KUIPERS, O. P., KUZUYAMA, T., KYRPIDES, N. C., KWON, H. J., LAUTRU, S., LAVIGNE, R., LEE, C. Y., LINQUAN, B., LIU, X., LIU, W., et al. 2015. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol*, 11, 625-31.
- MEDEMA, M. H., PAALVAST, Y., NGUYEN, D. D., MELNIK, A., DORRESTEIN, P. C., TAKANO, E. & BREITLING, R. 2014. Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput Biol*, 10, e1003822.
- MEDEMA, M. H., TAKANO, E. & BREITLING, R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol*, 30, 1218-23.
- MENG, H., LU, Z., WANG, Y., WANG, X. & ZHANG, S. 2011. In silico improvement of heterologous biosynthesis of erythromycin precursor 6-deoxyerythronolide B in *Escherichia coli*. *Biotechnology and Bioengineering*, 16, 445-456.
- MIESS, H., VAN TRAPPEN, S., CLEENWERCK, I., DE VOS, P. & GROSS, H. 2016. Reclassification of *Pseudomonas* sp. PB-6250T as *Lysobacter firmicutimachus* sp. nov. *Int J Syst Evol Microbiol*, 66, 4162-4166.
- MINGMA, R., TANAKA, K., OMURA, S., TAKAHASHI, Y. & MATSUMOTO, A. 2016. *Actinorhabdospora filicis* gen. nov., sp. nov., a new member of the family Micromonosporaceae. *Int J Syst Evol Microbiol*, 66, 3071-7.
- MIYAZAWA, M., SHIMAMURA, H., NAKAMURA, S.-I. & KAMEOKA, H. 1995. Antimutagenic Activity of (+)-Polyalthic Acid from *Vitex rotundifolia*. *Journal of Agricultural and Food Chemistry*, 43, 3012-3015.
- MOHIMANI, H., KERSTEN, R. D., LIU, W. T., WANG, M., PURVINE, S. O., WU, S., BREWER, H. M., PASA-TOLIC, L., BANDEIRA, N., MOORE, B. S., PEVZNER, P. A. & DORRESTEIN, P. C. 2014a. Automated genome mining of ribosomal peptide natural products. *ACS Chem Biol*, 9, 1545-51.
- MOHIMANI, H., LIU, W. T., KERSTEN, R. D., MOORE, B. S., DORRESTEIN, P. C. & PEVZNER, P. A. 2014b. NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery. *J Nat Prod*, 77, 1902-9.
- MOHIMANI, H., LIU, W. T., MYLNE, J. S., POTH, A. G., COLGRAVE, M. L., TRAN, D., SELSTED, M. E., DORRESTEIN, P. C. & PEVZNER, P. A. 2011. Cycloquest: identification of cyclopeptides via database search of their mass spectra against genome databases. *J Proteome Res*, 10, 4505-12.
- MOHS, R. C. & GREIG, N. H. 2017. Drug discovery and development: Role of basic biological research. *Alzheimers Dement (N Y)*, 3, 651-657.

- MORGAN, S. G., FISHER, J. L., MILLER, S. H., MCAFEE, S. T. & LARGIER, J. L. 2009. Nearshore larval retention in a region of strong upwelling and recruitment limitation. *Ecology*, 90, 3489-502.
- MORIDI FARIMANI, M. & MIRAN, M. 2014. Labdane diterpenoids from *Salvia reuterana*. *Phytochemistry*, 108, 264-269.
- MORINAKA, B. I., MASUNO, M. N., PAWLIK, J. R. & MOLINSKI, T. F. 2007. Amaranzole A, a new N-imidazolyl steroid from *Phorbas amaranthus*. *Org Lett*, 9, 5219-22.
- MORINAKA, B. I., PAWLIK, J. R. & MOLINSKI, T. F. 2009. Amarofoxocanes A and B: sulfated dimeric sterols defend the Caribbean coral reef sponge *Phorbas amaranthus* from fish predators. *J Nat Prod*, 72, 259-64.
- MORLOCK, G. E., RISTIVOJEVIC, P. & CHERNETSOVA, E. S. 2014. Combined multivariate data analysis of high-performance thin-layer chromatography fingerprints and direct analysis in real time mass spectra for profiling of natural products like propolis. *J Chromatogr A*, 1328, 104-12.
- MUIR, P., LI, S., LOU, S., WANG, D., SPAKOWICZ, D. J., SALICHOS, L., ZHANG, J., WEINSTOCK, G. M., ISAACS, F., ROZOWSKY, J. & GERSTEIN, M. 2016. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol*, 17, 53.
- MUKHERJEE, S., STAMATIS, D., BERTSCH, J., OVCHINNIKOVA, G., VEREZEMSKA, O., ISBANDI, M., THOMAS, A. D., ALI, R., SHARMA, K., KYRPIDES, N. C. & REDDY, T. B. 2017. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res*, 45, D446-d456.
- MULLOWNEY, M. W., E, O. H., TANOUYE, U., BURDETTE, J. E., PHAM, V. C. & MURPHY, B. T. 2015. A Pimarane Diterpene and Cytotoxic Angucyclines from a Marine-Derived Micromonospora sp. in Vietnam's East Sea. *Mar Drugs*, 13, 5815-27.
- MURRAY, A. G. 2016. Increased frequency and changed methods in the treatment of sea lice (*Lepeophtheirus salmonis*) in Scottish salmon farms 2005-2011. *Pest Manag Sci*, 72, 322-6.
- NAKA, T., FUJIWARA, N., YANO, I., MAEDA, S., DOE, M., MINAMINO, M., IKEDA, N., KATO, Y., WATABE, K., KUMAZAWA, Y., TOMIYASU, I. & KOBAYASHI, K. 2003. Structural analysis of sphingophospholipids derived from *Sphingobacterium spiritivorum*, the type species of genus *Sphingobacterium*. *Biochim Biophys Acta*, 1635, 83-92.
- NAKAMURA, Y., AFENDI, F. M., PARVIN, A. K., ONO, N., TANAKA, K., HIRAI MORITA, A., SATO, T., SUGIURA, T., ALTAF-UL-AMIN, M. & KANAYA, S. 2014. KNApSAcK Metabolite Activity Database for retrieving the relationships between metabolites and biological activities. *Plant Cell Physiol*, 55, e7.
- NARUSE, N., TENMYO, O., KOBARU, S., KAMEI, H., MIYAKI, T., KONISHI, M. & OKI, T. 1990. Pumilacidin, a complex of new antiviral antibiotics. Production, isolation, chemical properties, structure and biological activity. *J Antibiot (Tokyo)*, 43, 267-80.
- NAZARENKO, E. L., CRAWFORD, R. J. & IVANOVA, E. P. 2011. The structural diversity of carbohydrate antigens of selected gram-negative marine bacteria. *Mar Drugs*, 9, 1914-54.
- NEJADGHOLI, I. & BOLIC, M. 2015. A comparative study of PCA, SIMCA and Cole model for classification of bioimpedance spectroscopy measurements. *Comput Biol Med*, 63, 42-51.
- NEWMAN, D. J. 2008. Natural products as leads to potential drugs: an old process or the new hope for drug discovery? *J Med Chem*, 51, 2589-99.
- NOVEMBER, J. 2018. More than Moore's Mores: Computers, Genomics, and the Embrace of Innovation. *J Hist Biol*.

- NOYER, L., GROLEZ, G. P., PREVARSKAYA, N., GKIKI, D. & LEMONNIER, L. 2018. TRPM8 and prostate: a cold case? *Pflugers Arch.*
- O'DONNELL, A. G., MINNIKIN, D. E., GOODFELLOW, M., PARLETT, J. H., SCHOFIELD, G. M. & SCHAAL, K. P. 1985. Lipid and wall amino acid composition in the classification and identification of *Arachnia propionica*. *Zentralbl Bakteriol Mikrobiol Hyg A*, 260, 300-10.
- O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V. S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D. & PRUITT, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44, D733-45.
- OECKINGHAUS, A., HAYDEN, M. S. & GHOSH, S. 2011. Crosstalk in NF-kappaB signaling pathways. *Nat Immunol*, 12, 695-708.
- OHTANI, I. I., ICHIBA, T., ISOBE, M., KELLY-BORGES, M. & SCHEUER, P. J. 1995. Kauluamine, an unprecedented manzamine dimer from an Indonesian marine sponge, *Prianos* sp. *Journal of the American Chemical Society*, 117, 10743-10744.
- OJIMA, I. 2008. Modern natural products chemistry and drug discovery. *J Med Chem*, 51, 2587-8.
- OLSVIK, P., SAMUELSEN, O., ERDAL, A., HOLMELID, B. & LUNESTAD, B. 2013. *Toxicological assessment of the anti-salmon lice drug diflubenzuron on Atlantic cod Gadus morhua*.
- OLSVIK, P. A., ORNSRUD, R., LUNESTAD, B. T., STEINE, N. & FREDRIKSEN, B. N. 2014. Transcriptional responses in Atlantic salmon (*Salmo salar*) exposed to deltamethrin, alone or in combination with azamethiphos. *Comp Biochem Physiol C Toxicol Pharmacol*, 162, 23-33.
- ONAKA, H., MORI, Y., IGARASHI, Y. & FURUMAI, T. 2011. Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl Environ Microbiol*, 77, 400-6.
- OVERBEEK, R., BEGLEY, T., BUTLER, R. M., CHOUDHURI, J. V., CHUANG, H. Y., COHOON, M., DE CRECY-LAGARD, V., DIAZ, N., DISZ, T., EDWARDS, R., FONSTEIN, M., FRANK, E. D., GERDES, S., GLASS, E. M., GOESMANN, A., HANSON, A., IWATA-REUYL, D., JENSEN, R., JAMSHIDI, N., KRAUSE, L., KUBAL, M., LARSEN, N., LINKE, B., MCHARDY, A. C., MEYER, F., NEUWEGER, H., OLSEN, G., OLSON, R., OSTERMAN, A., PORTNOY, V., PUSCH, G. D., RODIONOV, D. A., RUCKERT, C., STEINER, J., STEVENS, R., THIELE, I., VASSIEVA, O., YE, Y., ZAGNITKO, O. & VONSTEIN, V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*, 33, 5691-702.
- OVERBEEK, R., OLSON, R., PUSCH, G. D., OLSEN, G. J., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., PARRELLO, B., SHUKLA, M., VONSTEIN, V., WATTAM, A. R., XIA, F. & STEVENS, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*, 42, D206-14.

- PARK, M. S., JUNG, S. R., LEE, K. H., LEE, M. S., DO, J. O., KIM, S. B. & BAE, K. S. 2006. *Chryseobacterium soldanellicola* sp. nov. and *Chryseobacterium taeanense* sp. nov., isolated from roots of sand-dune plants. *Int J Syst Evol Microbiol*, 56, 433-8.
- PARK, S., JUNG, Y. T., WON, S. M. & YOON, J. H. 2016. *Maribacter litorisediminis* sp. nov., isolated from a tidal flat. *Int J Syst Evol Microbiol*, 66, 4236-4242.
- PATRIDGE, E., GAREISS, P., KINCH, M. S. & HOYER, D. 2016. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov Today*, 21, 204-7.
- PATTI, G. J., YANES, O. & SIUZDAK, G. 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*, 13, 263-9.
- PEI, Z., LIU, G., LUBBEN, T. H. & SZCZEPANKIEWICZ, B. G. 2004. Inhibition of protein tyrosine phosphatase 1B as a potential treatment of diabetes and obesity. *Curr Pharm Des*, 10, 3481-504.
- PENCE, H. E. & WILLIAMS, A. 2010. ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education*, 87, 1123-1124.
- PENG, J., HU, J.-F., KAZI, A. B., LI, Z., AVERY, M., PERAUD, O., HILL, R. T., FRANZBLAU, S. G., ZHANG, F., SCHINAZI, R. F., WIRTZ, S. S., THARNISH, P., KELLY, M., WAHYUONO, S. & HAMANN, M. T. 2003. Manadomanzamines A and B: A Novel Alkaloid Ring System with Potent Activity against Mycobacteria and HIV-1. *Journal of the American Chemical Society*, 125, 13382-13386.
- PEREIRA, F. & AIRES-DE-SOUSA, J. 2018. Computational Methodologies in the Exploration of Marine Natural Product Leads. *Mar Drugs*, 16.
- PERKINS, N. D. 2007. Integrating cell-signalling pathways with NF-kappaB and IKK function. *Nat Rev Mol Cell Biol*, 8, 49-62.
- PETERS, R. J. 2010. Two rings in them all: the labdane-related diterpenoids. *Nat Prod Rep*, 27, 1521-30.
- PETTIT, G. R., HERALD, C. L., ODE, R. H., BROWN, P., GUST, D. J. & MICHEL, C. 1980. The isolation of loliolide from an Indian Ocean opisthobranch mollusc. *J Nat Prod*, 43, 752-5.
- PETTIT, G. R., KNIGHT, J. C., COLLINS, J. C., HERALD, D. L., PETTIT, R. K., BOYD, M. R. & YOUNG, V. G. 2000. Antineoplastic agents 430. Isolation and structure of cribrostatins 3, 4, and 5 from the republic of maldives cribrochalina species. *J Nat Prod*, 63, 793-8.
- PICKENS, L. B., TANG, Y. & CHOOI, Y. H. 2011. Metabolic engineering for the production of natural products. *Annu Rev Chem Biomol Eng*, 2, 211-36.
- PRABHU, K. S., ACHKAR, I. W., KUTTIKRISHNAN, S., AKHTAR, S., KHAN, A. Q., SIVEEN, K. S. & UDDIN, S. 2018. Embelin: a benzoquinone possesses therapeutic potential for the treatment of human cancer. *Future Med Chem*, 10, 961-976.
- PRABHU, S., REKHA, P. D. & ARUN, A. B. 2014. Zeaxanthin biosynthesis by members of the genus *Muricauda*. *Pol J Microbiol*, 63, 115-9.
- PRIETO, C., GARCIA-ESTRADA, C., LORENZANA, D. & MARTIN, J. F. 2012. NRPSp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics*, 28, 426-7.
- PRINSEP, M. R., BLUNT, J. W. & MUNRO, M. H. 1991. New cytotoxic beta-carboline alkaloids from the marine bryozoan, *Cribricellina cribraria*. *J Nat Prod*, 54, 1068-76.
- PURRINGTON, C. B. 2003. SECONDARY PRODUCTS | Antifeedant Substances in Plants. In: THOMAS, B. (ed.) *Encyclopedia of Applied Plant Sciences*. Oxford: Elsevier.
- QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-2.
- RADMARK, O., WERZ, O., STEINHILBER, D. & SAMUELSSON, B. 2015. 5-Lipoxygenase, a key enzyme for leukotriene biosynthesis in health and disease. *Biochim Biophys Acta*, 1851, 331-9.

- RADWAN, M., HANORA, A., KHALIFA, S. & ABOU-EL-ELA, S. H. 2012. Manzamines: a potential for novel cures. *Cell Cycle*, 11, 1765-72.
- RAKHA, E. A. & GREEN, A. R. 2017. Molecular classification of breast cancer: what the pathologist needs to know. *Pathology*, 49, 111-119.
- RAO, K. V., DONIA, M. S., PENG, J., GARCIA-PALOMERO, E., ALONSO, D., MARTINEZ, A., MEDINA, M., FRANZBLAU, S. G., TEKWANI, B. L., KHAN, S. I., WAHYUONO, S., WILLETT, K. L. & HAMANN, M. T. 2006. Manzamine B and E and Ircinal A Related Alkaloids from an Indonesian Acanthostrongylophora Sponge and Their Activity against Infectious, Tropical Parasitic, and Alzheimer's Diseases. *Journal of Natural Products*, 69, 1034-1040.
- RAO, K. V., SANTARSIERO, B. D., MESECAR, A. D., SCHINAZI, R. F., TEKWANI, B. L. & HAMANN, M. T. 2003. New manzamine alkaloids with activity against infectious and tropical parasitic diseases from an Indonesian sponge. *Journal of natural products*, 66, 823-828.
- RATUR, A., GYANESHWAR, P., SINGH, S., TAK, N. & GEHLOT, H. 2013. Bacterial Endophytes and Their Significance in the Sustainable Production of Food in Non-Legumes.
- RAUSCH, C., WEBER, T., KOHLBACHER, O., WOHLLEBEN, W. & HUSON, D. H. 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res*, 33, 5799-808.
- RECH, J. C., ECKERT, W. A., MAHER, M. P., BANKE, T., BHATTACHARYA, A. & WICKENDEN, A. D. 2010. Recent advances in the biology and medicinal chemistry of TRPA1. *Future Med Chem*, 2, 843-58.
- REDDY, B. V., MILSHTEYN, A., CHARLOP-POWERS, Z. & BRADY, S. F. 2014. eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chem Biol*, 21, 1023-33.
- REITZ, Z. L., SANDY, M. & BUTLER, A. 2017. Biosynthetic considerations of triscatechol siderophores framed on serine and threonine macrolactone scaffolds. *Metallomics*, 9, 824-839.
- REN, H., WANG, B. & ZHAO, H. 2017. Breaking the silence: new strategies for discovering novel natural products. *Curr Opin Biotechnol*, 48, 21-27.
- RIESCO, R., CARRO, L., ROMAN-PONCE, B., PRIETO, C., BLOM, J., KLENK, H. P., NORMAND, P. & TRUJILLO, M. E. 2018. Defining the Species *Micromonospora saelicesensis* and *Micromonospora noduli* Under the Framework of Genomics. *Front Microbiol*, 9, 1360.
- RIOS, J. L., FRANCINI, F. & SCHINELLA, G. R. 2015. Natural Products for the Treatment of Type 2 Diabetes Mellitus. *Planta Med*, 81, 975-94.
- ROBERTSON, A. L., HOLMES, G. R., BOJARCZUK, A. N., BURGON, J., LOYNES, C. A., CHIMEN, M., SAWTELL, A. K., HAMZA, B., WILLSON, J., WALMSLEY, S. R., ANDERSON, S. R., COLES, M. C., FARROW, S. N., SOLARI, R., JONES, S., PRINCE, L. R., IRIMIA, D., RAINGER, G. E., KADIRKAMANATHAN, V., WHYTE, M. K. & RENSHAW, S. A. 2014. A zebrafish compound screen reveals modulation of neutrophil reverse migration as an anti-inflammatory mechanism. *Sci Transl Med*, 6, 225ra29.
- ROBOTTI, E. & MARENGO, E. 2016. Chemometric Multivariate Tools for Candidate Biomarker Identification: LDA, PLS-DA, SIMCA, Ranking-PCA. *Methods Mol Biol*, 1384, 237-67.
- ROMANO, G., COSTANTINI, M., SANSONE, C., LAURITANO, C., RUOCCO, N. & IANORA, A. 2017. Marine microorganisms as a promising and sustainable source of bioactive molecules. *Mar Environ Res*, 128, 58-69.

- ROTHSCHILD, L. J. & MANCINELLI, R. L. 2001. Life in extreme environments. *Nature*, 409, 1092-1101.
- ROTTIG, M., MEDEMA, M. H., BLIN, K., WEBER, T., RAUSCH, C. & KOHLBACHER, O. 2011. NRSPredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res*, 39, W362-7.
- ROY, A. 2018. Early Probe and Drug Discovery in Academia: A Minireview. *High Throughput*, 7.
- RUDI, A., STEIN, Z., GREEN, S., GOLDBERG, I., KASHMAN, Y., BENAYAHU, Y. & SCHLEYER, M. 1994. Phorbazoles A • D, novel chlorinated phenylpyrrolyloxazoles from the marine sponge phorbasp aff. clathrata. *Tetrahedron Letters*, 35, 2589-2592.
- RUTLEDGE, P. J. & CHALLIS, G. L. 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol*, 13, 509-23.
- SAHA, G., SAHA ROY, S. & GHOSH, S. 1990. Biyclo[2.2.1]heptane as cyclopentane precursor. Part 41 Stereocontrolled synthesis of a potential intermediate to chromophycane dolastane and clavularane. *Tetrahedron*, 46, 8229-8236.
- SAKAI, R., HIGA, T., JEFFORD, C. W. & BERNARDINELLI, G. 1986. Manzamine A, a novel antitumor alkaloid from a sponge. *Journal of the American Chemical Society*, 108, 6404-6405.
- SAKAI, R., KOHMOTO, S., HIGA, T., JEFFORD, C. W. & BERNARDINELLI, G. 1987. Manzamine B and C, two novel alkaloids from the sponge haliclona sp. *Tetrahedron Letters*, 28, 5493-5496.
- SANTOS, C. M. M., FREITAS, M. & FERNANDES, E. 2018. A comprehensive review on xanthone derivatives as alpha-glucosidase inhibitors. *Eur J Med Chem*, 157, 1460-1479.
- SARMIENTO QUINTERO, F., ARIZA, A. J., BARBOZA GARCIA, F., CANAL DE MOLANO, N., CASTRO BENAVIDES, M., CRUCHET MUNOZ, S., DELGADO CARBAJAL, L., DEWAELE OLIVERA, M. R., FERNANDEZ, A., HELLER, S., LADINO MELENDEZ, L., MARTINEZ, S. M., MAYOR OXILIA, R., MEJIA CASTRO, M., MONTERO BRENES, C., SANABRIA, M. C., TARAZONA COTE, M. C. & VERA CHAMORRO, J. F. 2016. [Overweight and obesity: review and update]. *Acta Gastroenterol Latinoam*, 46, 131-59.
- SARPE, V. A. & KULKARNI, S. S. 2011. Synthesis of maradolipid. *J Org Chem*, 76, 6866-70.
- SATOYOSHI, E. 1992. Therapeutic trials on progressive muscular dystrophy. *Intern Med*, 31, 841-6.
- SCHAECHTER, M. 2015. A brief history of bacterial growth physiology. *Front Microbiol*, 6, 289.
- SCHIEFERDECKER, S., KÖNIG, S., KOEBERLE, A., DAHSE, H.-M., WERZ, O. & NETT, M. 2015. Myxochelins Target Human 5-Lipoxygenase. *Journal of Natural Products*, 78, 335-338.
- SCHIEFERDECKER, S., KONIG, S., PACE, S., WERZ, O. & NETT, M. 2017. Myxochelin-Inspired 5-Lipoxygenase Inhibitors: Synthesis and Biological Evaluation. *ChemMedChem*, 12, 23-27.
- SEARLE, P. A. & MOLINSKI, T. F. 1995. 4-Dehydroecdysterone, a new ecdysteroid from the zoanthid Parazoanthus sp. *J Nat Prod*, 58, 264-8.
- SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30, 2068-9.
- SHAH, M. A., KEACH, J. E. & PANICHAYUPAKARANANT, P. 2018. Antidiabetic Naphthoquinones and Their Plant Resources in Thailand. *Chem Pharm Bull (Tokyo)*, 66, 483-492.

- SHAO, M., SHAN, B., LIU, Y., DENG, Y., YAN, C., WU, Y., MAO, T., QIU, Y., ZHOU, Y., JIANG, S., JIA, W., LI, J., LI, J., RUI, L., YANG, L. & LIU, Y. 2014. Hepatic IRE1alpha regulates fasting-induced metabolic adaptive programs through the XBP1s-PPARalpha axis signalling. *Nat Commun*, 5, 3528.
- SHEEJA, K., SHIHAB, P. K. & KUTTAN, G. 2006. Antioxidant and anti-inflammatory activities of the plant *Andrographis paniculata* nees. *Immunopharmacology and Immunotoxicology*, 28, 129-140.
- SHEN, B. 2015. A New Golden Age of Natural Products Drug Discovery. *Cell*, 163, 1297-300.
- SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol*, 26, 1135-45.
- SHEU, S. Y., CHEN, Y. L. & CHEN, W. M. 2016. *Flavobacterium verecundum* sp. nov., isolated from a freshwater river. *Int J Syst Evol Microbiol*, 66, 3337-3344.
- SIEGEL, J. P., SMITH, A. R. & NOVAK, R. J. 1997. Comparison of the cellular fatty acid composition of a bacterium isolated from a human and alleged to be *Bacillus sphaericus* with that of *Bacillus sphaericus* isolated from a mosquito larvicide. *Appl Environ Microbiol*, 63, 1006-10.
- SIMITHY, J., FUANTA, N. R., ALTURKI, M., HOBRATH, J. V., WAHBA, A. E., PINA, I., RATH, J., HAMANN, M. T., DERUITER, J., GOODWIN, D. C. & CALDERON, A. I. 2018. Slow-Binding Inhibition of *Mycobacterium tuberculosis* Shikimate Kinase by Manzamine Alkaloids. *Biochemistry*, 57, 4923-4933.
- SINGH, B. & SHARMA, R. A. 2015. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech*, 5, 129-151.
- SMANSKI, M. J., PETERSON, R. M., HUANG, S.-X. & SHEN, B. 2012. Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Current Opinion in Chemical Biology*, 16, 132-141.
- SMITH, C. A., WANT, E. J., O'MAILLE, G., ABAGYAN, R. & SIUZDAK, G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*, 78, 779-87.
- SNIDER, B. B. & DUVALL, J. R. 2005. Synthesis of jenamidines A1/A2. *Org Lett*, 7, 4519-22.
- SOHN, J. H., LEE, Y. R., LEE, D. S., KIM, Y. C. & OH, H. 2013. PTP1B inhibitory secondary metabolites from marine-derived fungal strains *Penicillium* spp. and *Eurotium* sp. *J Microbiol Biotechnol*, 23, 1206-11.
- SONG, H. J., LIU, Y. X., LIU, Y. X., HUANG, Y. Q., LI, Y. Q. & WANG, Q. M. 2014. Design, synthesis, anti-TMV, fungicidal, and insecticidal activity evaluation of 1,2,3,4-tetrahydro-beta-carboline-3-carboxylic acid derivatives based on virus inhibitors of plant sources. *Bioorg Med Chem Lett*, 24, 5228-33.
- SONGSUMANUS, A., TANASUPAWAT, S., IGARASHI, Y. & KUDO, T. 2013. *Micromonospora maritima* sp. nov., isolated from mangrove soil. *Int J Syst Evol Microbiol*, 63, 554-9.
- SOUSA, T. D. S., JIMENEZ, P. C., FERREIRA, E. G., SILVEIRA, E. R., BRAZ-FILHO, R., PESSOA, O. D. L. & COSTA-LOTUFO, L. V. 2012. Anthracyclines from *Micromonospora* sp. *Journal of natural products*, 75, 489-493.
- SPICER, R., SALEK, R. M., MORENO, P., CANUETO, D. & STEINBECK, C. 2017. Navigating freely-available software tools for metabolomics analysis. *Metabolomics*, 13, 106.
- STARCEVIC, A., ZUCKO, J., SIMUNKOVIC, J., LONG, P. F., CULLUM, J. & HRANUELI, D. 2008. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res*, 36, 6882-92.

- STEIN, L. D. 2010. The case for cloud computing in genome informatics. *Genome Biol*, 11, 207.
- STEVENS, K., TYRRELL, A. J., SKERRATT, S. & ROBERTSON, J. 2011. Synthesis of NP25302. *Org Lett*, 13, 5964-7.
- SU, Z., NING, B., FANG, H., HONG, H., PERKINS, R., TONG, W. & SHI, L. 2011. Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev Mol Diagn*, 11, 333-43.
- SULISTIYANI, MEGA, S. & YOANA PUSPITA, S. 2016. Inhibition of  $\alpha$ -glucosidase activity by ethanolic extract of *Melia azedarach* L. leaves. *IOP Conference Series: Earth and Environmental Science*, 31, 012025.
- SUNG, A. A., GROMEK, S. M. & BALUNAS, M. J. 2017. Upregulation and Identification of Antibiotic Activity of a Marine-Derived *Streptomyces* sp. via Co-Cultures with Human Pathogens. *Mar Drugs*, 15.
- SUPONG, K., SURIYACHADKUN, C., TANASUPAWAT, S., SUWANBORIRUX, K., PITTAYAKHAJONWUT, P., KUDO, T. & THAWAI, C. 2013. *Micromonospora sediminicola* sp. nov., isolated from marine sediment. *Int J Syst Evol Microbiol*, 63, 570-5.
- TAE, H., SOHNG, J. K. & PARK, K. 2009. MapiDB: an integrated web database for type I polyketide synthases. *Bioprocess Biosyst Eng*, 32, 723-7.
- TAMBO, E., KHATER, E. I., CHEN, J. H., BERGQUIST, R. & ZHOU, X. N. 2015. Nobel prize for the artemisinin and ivermectin discoveries: a great boost towards elimination of the global infectious diseases of poverty. *Infect Dis Poverty*, 4, 58.
- TAMRAKAR, A. K., MAURYA, C. K. & RAI, A. K. 2014. PTP1B inhibitors for type 2 diabetes treatment: a patent review (2011 - 2014). *Expert Opin Ther Pat*, 24, 1101-15.
- TANG, K., SU, Y., BRACKMAN, G., CUI, F., ZHANG, Y., SHI, X., COENYE, T. & ZHANG, X. H. 2015. MomL, a novel marine-derived N-acyl homoserine lactonase from *Muricauda olearia*. *Appl Environ Microbiol*, 81, 774-82.
- TAWFIKE, A. F., VIEGELMANN, C. & EDRADA-EBEL, R. 2013. Metabolomics and dereplication strategies in natural products. *Methods Mol Biol*, 1055, 227-44.
- TAYLOR-WELLS, J., GROSS, A. D., JIANG, S., DEMARES, F., CLEMENTS, J. S., CARLIER, P. R. & BLOOMQUIST, J. R. 2018. Toxicity, mode of action, and synergist potential of flonicamid against mosquitoes. *Pestic Biochem Physiol*, 151, 3-9.
- TAYLOR, M. W., RADAX, R., STEGER, D. & WAGNER, M. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev*, 71, 295-347.
- THAREJA, S., AGGARWAL, S., BHARDWAJ, T. R. & KUMAR, M. 2012. Protein tyrosine phosphatase 1B inhibitors: a molecular level legitimate approach for the management of diabetes mellitus. *Med Res Rev*, 32, 459-517.
- TONELLI, A. E. & SCHILLING, F. C. 1981. Carbon-13 NMR chemical shifts and the microstructure of polymers. *Accounts of Chemical Research*, 14, 233-238.
- TRAN, Q. T. N., WONG, W. S. F. & CHAI, C. L. L. 2017. Labdane diterpenoids as potential anti-inflammatory agents. *Pharmacological Research*, 124, 43-63.
- TRIVELLA, D. B. B. & DE FELICIO, R. 2018. The Tripod for Bacterial Natural Product Discovery: Genome Mining, Silent Pathway Induction, and Mass Spectrometry-Based Molecular Networking. *mSystems*, 3.
- TRUJILLO, M. E., HONG, K. & GENILLOU, O. 2014. The Family *Micromonosporaceae*. In: ROSENBERG, E., DELONG, E. F., LORY, S., STACKEBRANDT, E. & THOMPSON, F. (eds.) *The Prokaryotes: Actinobacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg.

- TSICHRITZIS, F. & JAKUPOVIC, J. 1991. Diterpenes from *Leyssera gnaphaloides*. *Phytochemistry*, 30, 211-213.
- TSUDA, M., INABA, K., KAWASAKI, N., HONMA, K. & KOBAYASHI, J. I. 1996. Chiral resolution of ( $\pm$ )-keramaphidin B and isolation of manzamine L, a new  $\beta$ -carboline alkaloid from a sponge *Amphimedon* sp. *Tetrahedron*, 52, 2319-2324.
- TSUGAWA, H., CAJKA, T., KIND, T., MA, Y., HIGGINS, B., IKEDA, K., KANAZAWA, M., VANDERGHEYNST, J., FIEHN, O. & ARITA, M. 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*, 12, 523-6.
- URBIETA, M. S., DONATI, E. R., CHAN, K. G., SHAHAR, S., SIN, L. L. & GOH, K. M. 2015. Thermophiles in the genomic era: Biodiversity, science, and applications. *Biotechnol Adv*, 33, 633-47.
- VALLIAPPAN, K., SUN, W. & LI, Z. 2014. Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Appl Microbiol Biotechnol*, 98, 7365-77.
- VALYI-NAGY, T., ÚRI, J. & SZILAGYI, I. 1954. Primycin, a New Antibiotic. *Nature*, 174, 1105-1106.
- VAN DEN BURG, B. 2003. Extremophiles as a source for novel enzymes. *Curr Opin Microbiol*, 6, 213-8.
- VAN HEEL, A. J., DE JONG, A., MONTALBAN-LOPEZ, M., KOK, J. & KUIPERS, O. P. 2013. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res*, 41, W448-53.
- VERA, L. M. & MIGAUD, H. 2016. Hydrogen peroxide treatment in Atlantic salmon induces stress and detoxification response in a daily manner. *Chronobiol Int*, 33, 530-42.
- VIANT, M. R., ROSENBLUM, E. S. & TIEERDEMA, R. S. 2003. NMR-based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health. *Environ Sci Technol*, 37, 4982-9.
- VILLAS-BOAS, S. G., MAS, S., AKESSON, M., SMEDSGAARD, J. & NIELSEN, J. 2005. Mass spectrometry in metabolome analysis. *Mass Spectrom Rev*, 24, 613-46.
- VLAD, P. F., RUSSO, A. G. & LAZUREVSKII, G. V. 1966. Synthesis of manoyl oxide and 13-epimanoyl oxide. *Chemistry of Natural Compounds*, 2, 139-141.
- VOLPE, M., MIRALTO, M., GUSTINCICH, S. & SANGES, R. 2018. ClusterScan: simple and generalistic identification of genomic clusters. *Bioinformatics*, 34, 3921-3923.
- WAGNER, G. N., FAST, M. D. & JOHNSON, S. C. 2008. Physiology and immunology of *Lepeophtheirus salmonis* infections of salmonids. *Trends Parasitol*, 24, 176-83.
- WALSH, C. T. & FISCHBACH, M. A. 2010. Natural products version 2.0: connecting genes to molecules. *Journal of the American Chemical Society*, 132, 2469-2493.
- WALSH, C. T. & TANG, Y. 2017. *Natural Product Biosynthesis*, Royal Society of Chemistry.
- WANG, M., CARVER, J. J., PHELAN, V. V., SANCHEZ, L. M., GARG, N., PENG, Y., NGUYEN, D. D., WATROUS, J., KAPONO, C. A., LUZZATTO-KNAAN, T., PORTO, C., BOUSLIMANI, A., MELNIK, A. V., MEEHAN, M. J., LIU, W. T., CRUSEMANN, M., BOUDREAU, P. D., ESQUENAZI, E., SANDOVAL-CALDERON, M., KERSTEN, R. D., PACE, L. A., QUINN, R. A., DUNCAN, K. R., HSU, C. C., FLOROS, D. J., GAVILAN, R. G., KLEIGREWE, K., NORTHEN, T., DUTTON, R. J., PARROT, D., CARLSON, E. E., AIGLE, B., MICHELSEN, C. F., JELSBAK, L., SOHLENKAMP, C., PEVZNER, P., EDLUND, A., MCLEAN, J., PIEL, J., MURPHY, B. T., GERWICK, L., LIAW, C. C., YANG, Y. L., HUMPF, H. U., MAANSSON, M., KEYZERS, R. A., SIMS, A. C., JOHNSON, A. R., SIDEBOTTOM, A. M., SEDIO, B. E., KLITGAARD, A., LARSON, C. B., P, C. A. B., TORRES-MENDOZA, D., GONZALEZ, D. J., SILVA, D. B., MARQUES, L. M., DEMARQUE, D. P., POCIUTE, E., O'NEILL, E. C., BRIAND, E., HELFRICH,

- E. J. N., GRANATOSKY, E. A., GLUKHOV, E., RYFFEL, F., HOUSON, H., MOHIMANI, H., KHARBUSH, J. J., ZENG, Y., VORHOLT, J. A., KURITA, K. L., CHARUSANTI, P., MCPHAIL, K. L., NIELSEN, K. F., VUONG, L., ELFEKI, M., TRAXLER, M. F., ENGENE, N., KOYAMA, N., VINING, O. B., BARIC, R., SILVA, R. R., MASCUCH, S. J., TOMASI, S., JENKINS, S., MACHERLA, V., HOFFMAN, T., AGARWAL, V., WILLIAMS, P. G., DAI, J., NEUPANE, R., GURR, J., RODRIGUEZ, A. M. C., LAMSA, A., ZHANG, C., DORRESTEIN, K., DUGGAN, B. M., ALMALITI, J., ALLARD, P. M., PHAPALE, P., et al. 2016a. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol*, 34, 828-837.
- WANG, Y., MING, H., GUO, W., CHEN, H. & ZHOU, C. 2016b. *Aquimarina aggregata* sp. nov., isolated from seawater. *Int J Syst Evol Microbiol*, 66, 3406-3412.
- WANG, Y., XIAO, J., SUZEK, T. O., ZHANG, J., WANG, J. & BRYANT, S. H. 2009. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*, 37, W623-33.
- WAŚKIEWICZ, A. & IRZYKOWSKA, L. 2014. *Flavobacterium* spp.—Characteristics, Occurrence, and Toxicity.
- WATANABE, D., TSUDA, M. & KOBAYASHI, J. 1998. Three new manzamine congeners from amphimedon sponge. *J Nat Prod*, 61, 689-92.
- WATERS, A. L., PERAUD, O., KASANA, N., SIMS, J. W., KOTHALAWALA, N., ANDERSON, M. A., ABBAS, S. H., RAO, K. V., JUPALLY, V. R., KELLY, M., DASS, A., HILL, R. T. & HAMANN, M. T. 2014. An analysis of the sponge *Acanthostrongylophora igens*' microbiome yields an actinomycete that produces the natural product manzamine A. *Frontiers in Marine Science*, 1.
- WEBER, T., BLIN, K., DUDELA, S., KRUG, D., KIM, H. U., BRUCCOLERI, R., LEE, S. Y., FISCHBACH, M. A., MULLER, R., WOHLLEBEN, W., BREITLING, R., TAKANO, E. & MEDEMA, M. H. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*, 43, W237-43.
- WEBER, T. & KIM, H. U. 2016. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol*, 1, 69-79.
- WELLER, M. G. 2012. A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques. *Sensors (Basel)*, 12, 9181-209.
- WETTERSTRAND, K. A. 2018. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP).
- WILSON, M. C. & PIEL, J. 2013. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol*, 20, 636-47.
- WILSON, Z. E. & BRIMBLE, M. A. 2009. Molecules derived from the extremes of life. *Nat Prod Rep*, 26, 44-71.
- WISHART, D. S. 2016. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov*, 15, 473-84.
- WOLD, S., ESBENSEN, K. & GELADI, P. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.
- WONGTANGTINTHARN, S., OKU, H., IWASAKI, H., INAFUKU, M., TODA, T. & YANAGITA, T. 2005. Incorporation of branched-chain fatty acid into cellular lipids and caspase-independent apoptosis in human breast cancer cell line, SKBR-3. *Lipids Health Dis*, 4, 29.
- WOOD, D. E. & SALZBERG, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*, 15, R46.

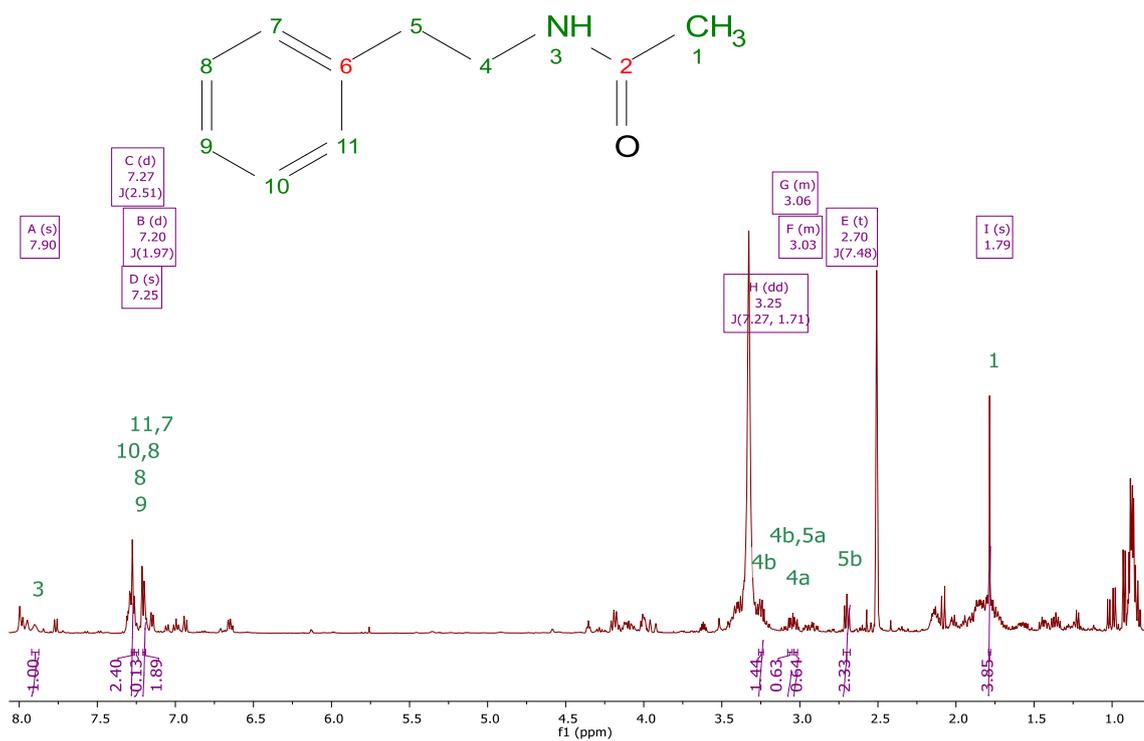
- WRIGHT, D. W., OPPELAL, F. & DEMPSTER, T. 2016. Early-stage sea lice recruits on Atlantic salmon are freshwater sensitive. *J Fish Dis*, 39, 1179-86.
- XIE, P., MA, M., RATEB, M. E., SHAABAN, K. A., YU, Z., HUANG, S.-X., ZHAO, L.-X., ZHU, X., YAN, Y., PETERSON, R. M., LOHMAN, J. R., YANG, D., YIN, M., RUDOLF, J. D., JIANG, Y., DUAN, Y. & SHEN, B. 2014. Biosynthetic Potential-Based Strain Prioritization for Natural Product Discovery: A Showcase for Diterpenoid-Producing Actinomycetes. *Journal of Natural Products*, 77, 377-387.
- XU, Z. 2011. Modernization: One step at a time. *Nature*, 480, S90.
- YADAV, G., GOKHALE, R. S. & MOHANTY, D. 2003. SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res*, 31, 3654-8.
- YAMADA, M., TAKAHASHI, Y., KUBOTA, T., FROMONT, J., ISHIYAMA, A., OTOGURO, K., YAMADA, H., ÔMURA, S. & KOBAYASHI, J. I. 2009. Zamamidine C, 3,4-dihydro-6-hydroxy-10,11-epoxymanzamine A, and 3,4-dihydromanzamine J N-oxide, new manzamine alkaloids from sponge Amphimedon sp. *Tetrahedron*, 65, 2313-2317.
- YAMAMOTO, R., YAMAMOTO, Y., IMAI, S., FUKUTOMI, R., OZAWA, Y., ABE, M., MATUO, Y. & SAITO, K. 2014. Effects of Various Phytochemicals on Indoleamine 2,3-Dioxygenase 1 Activity: Galanal Is a Novel, Competitive Inhibitor of the Enzyme. *PLOS ONE*, 9, e88789.
- YAMANAKA, S., FUDO, R., KAWAGUCHI, A. & KOMAGATA, K. 1988. AXONOMIC SIGNIFICANCE OF HYDROXY FATTY ACIDS IN MYXOBACTERIA WITH SPECIAL REFERENCE TO 2-HYDROXY FATTY ACIDS IN PHOSPHOLIPIDS. *The Journal of General and Applied Microbiology*, 34, 57-66.
- YANAI, M. & HIRAMOTO, S. 1999. First total synthesis of N-4909 and its diastereomer; a stimulant of apolipoprotein E secretion in human hepatoma Hep G2 cells. *J Antibiot (Tokyo)*, 52, 150-9.
- YANAI, M., SUZUKI, M., KAWAMURA, K., OSHIDA, N., HIRAMOTO, S., YASUDA, O., SHINGAI, A. & SUGURO, T. 2000. Synthesis of N-4909 analogs. Part I. A stimulant of apolipoprotein E secretion in human hepatoma G2 cells. *J Antibiot (Tokyo)*, 53, 1385-96.
- YANG, C., LI, Q. & LI, Y. 2014. Targeting nuclear receptors with marine natural products. *Mar Drugs*, 12, 601-35.
- YANG, Z., LIU, S., CHEN, X., CHEN, H., HUANG, M. & ZHENG, J. 2000. Induction of apoptotic cell death and in vivo growth inhibition of human cancer cells by a saturated branched-chain fatty acid, 13-methyltetradecanoic acid. *Cancer Res*, 60, 505-9.
- YANO, I., FURUKAWA, Y. & KUSUNOSE, M. 1969. Occurrence of  $\alpha$ -hydroxy fatty acids in Actinomycetales. *FEBS Letters*, 4, 96-98.
- YANO, I., FURUKAWA, Y. & KUSUNOSE, M. 1970. 2-hydroxy fatty acid-containing phospholipid of *Arthrobacter simplex*. *Biochim Biophys Acta*, 210, 105-15.
- YANO, I., OHNO, Y., MASUI, M., KATO, K., YABUUCHI, E. & OHYAMA, A. 1976. Occurrence of 2- and 3-hydroxy fatty acids in high concentrations in the extractable and bound lipids of *Flavobacterium meningosepticum* and *Flavobacterium Iib*. *Lipids*, 11, 685-688.
- YAO, W., WANG, J., ZHANG, P., PENG, J. & ZHANG, Y. 2004. Insecticidal activity of alcohol extracts from *Peganum harmala*. *Acta Botanica Boreali-Occidentalia Sinica*, 24, 1096-1099.
- YASUMURO, K., SHIBAZAKI, M., SASAKI, T., IMAI, H., YAMAGUCHI, H., SUZUKI, K., MORIOKA, M. & TAKEBAYASI, Y. 1994. 16-membered lactone compounds from izenamicins-producing microorganism. *J Antibiot (Tokyo)*, 47, 250-2.

- YU, J., YANG, L. N., WU, Y. Y., LI, B. H., WENG, S. M., HU, C. L. & HAN, Y. L. 2016. 13-Methyltetradecanoic acid mitigates cerebral ischemia/reperfusion injury. *Neural Regen Res*, 11, 1431-1437.
- ZAMORA, L., VELA, A. I., SÁNCHEZ-PORRO, C., PALACIOS, M. A., DOMÍNGUEZ, L., MOORE, E. R. B., VENTOSA, A. & FERNÁNDEZ-GARAYZÁBAL, J. F. 2013. Characterization of flavobacteria possibly associated with fish and fish farm environment. Description of three novel *Flavobacterium* species: *Flavobacterium collinsii* sp. nov., *Flavobacterium branchiarum* sp. nov., and *Flavobacterium branchiicola* sp. nov. *Aquaculture*, 416-417, 346-353.
- ZENG, Y., ZHANG, Y., WENG, Q., HU, M. & ZHONG, G. 2010. Cytotoxic and insecticidal activities of derivatives of harmine, a natural insecticidal component isolated from *Peganum harmala*. *Molecules*, 15, 7775-91.
- ZENGLER, K. & ZARAMELA, L. S. 2018. The social network of microorganisms — how auxotrophies shape complex communities. *Nature Reviews Microbiology*, 16, 383-390.
- ZHANG, H., MAJOR, J. M., LEWIS, R. J. & CAPON, R. J. 2008a. Phorbasins G-K: new cytotoxic diterpenes from a southern Australian marine sponge, *Phorbas* sp. *Org Biomol Chem*, 6, 3811-5.
- ZHANG, J., CHIODINI, R., BADR, A. & ZHANG, G. 2011. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, 38, 95-109.
- ZHANG, L. & BARRITT, G. J. 2004. Evidence that TRPM8 Is an Androgen-Dependent Ca<sup>2+</sup> Channel Required for the Survival of Prostate Cancer Cells. *Cancer Research*, 64, 8365.
- ZHANG, L., XI, L., RUAN, J. & HUANG, Y. 2012. *Micromonospora yangpuensis* sp. nov., isolated from a sponge. *International journal of systematic and evolutionary microbiology*, 62, 272-278.
- ZHANG, X., TIAN, J., LI, J., HUANG, L., WU, S., LIANG, W., ZHONG, L., YE, J. & YE, F. 2016a. A novel protein tyrosine phosphatase 1B inhibitor with therapeutic potential for insulin resistance. *Br J Pharmacol*, 173, 1939-49.
- ZHANG, Y., ADNANI, N., BRAUN, D. R., ELLIS, G. A., BARNES, K. J., PARKER-NANCE, S., GUZEI, I. A. & BUGNI, T. S. 2016b. *Micromonohalimanes A and B*: Antibacterial Halimane-Type Diterpenoids from a Marine *Micromonospora* Species. *J Nat Prod*, 79, 2968-2972.
- ZHANG, Y., GAEKWAD, J., WOLFERT, M. A. & BOONS, G. J. 2008b. Synthetic tetra-acylated derivatives of lipid A from *Porphyromonas gingivalis* are antagonists of human TLR4. *Org Biomol Chem*, 6, 3371-81.
- ZHAO, C., WU, Y., YANG, C., LIU, B. & HUANG, Y. 2015. Hypotensive, hypoglycaemic and hypolipidaemic effects of bioactive compounds from microalgae and marine micro-organisms. *International Journal of Food Science & Technology*, 50, 1705-1717.
- ZHAO, H., KASSAMA, Y., YOUNG, M., KELL, D. B. & GOODACRE, R. 2004. Differentiation of *Micromonospora* isolates from a coastal sediment in Wales on the basis of Fourier transform infrared spectroscopy, 16S rRNA sequence analysis, and the amplified fragment length polymorphism technique. *Appl Environ Microbiol*, 70, 6619-27.
- ZHOU, B.-N., SLEBODNICK, C., JOHNSON, R. K., MATTERN, M. R. & KINGSTON, D. G. I. 2000. New Cytotoxic Manzamine Alkaloids from a Palaun Sponge. *Tetrahedron*, 56, 5781-5784.
- ZHOU, Y., ZHANG, W., LIU, X., YU, H., LU, X. & JIAO, B. 2017. Inhibitors of Protein Tyrosine Phosphatase 1B from Marine Natural Products. *Chem Biodivers*, 14.

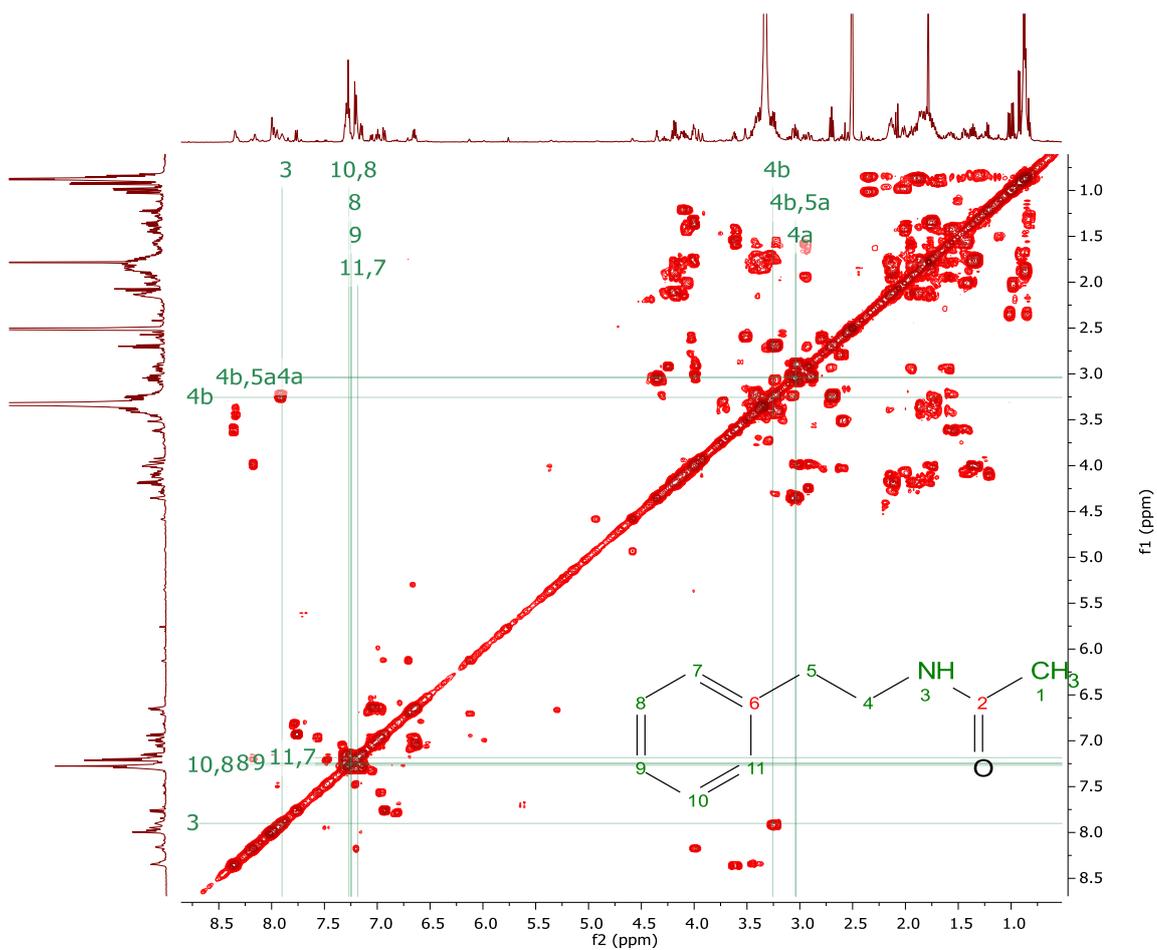
- ZIEMERT, N., PODELL, S., PENN, K., BADGER, J. H., ALLEN, E. & JENSEN, P. R. 2012. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One*, 7, e34064.
- ZINKEL, D. F. & CLARKE, W. B. 1985. Resin acids of *Pinus resinosa* needles. *Phytochemistry*, 24, 1267-1271.
- ZINKEL, D. F., MAGEE, T. V. & WALTER, J. 1985. Major resin acids of *Pinus nigra* needles. *Phytochemistry*, 24, 1273-1277.
- ZWENGER, S. & BASU, C. 2008. *Plant terpenoids: Applications and future potentials*.

## **Appendices & Supplemental Material**

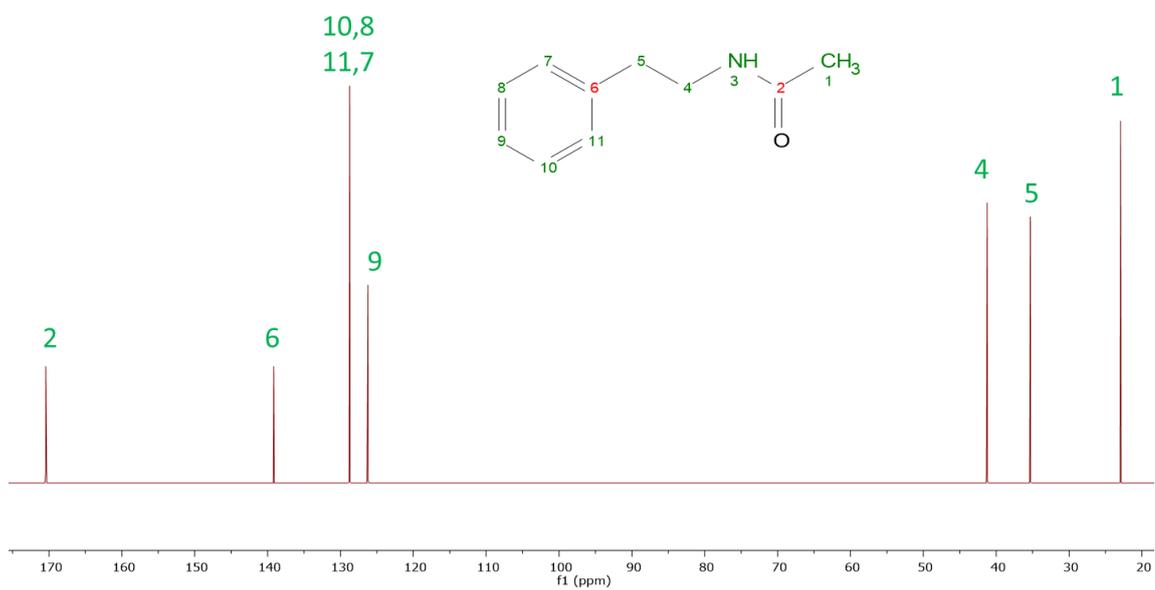
## Appendix I: NMR data of *N*-phenethylacetamide



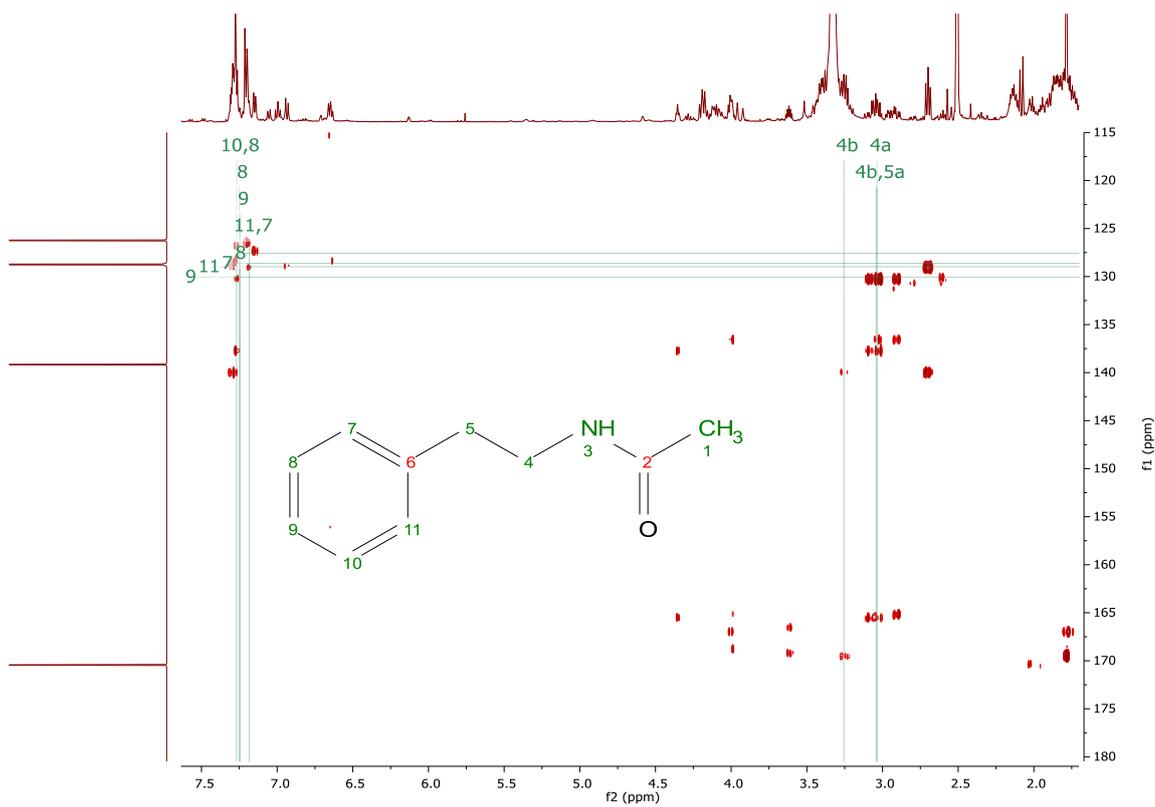
App 1  $^1\text{H}$  NMR (500 MHz) spectrum for *N*-phenethylacetamide (spectrum with a mixture of compounds), measured in DMSO- $d_6$ .



App 2  $^1\text{H}$ - $^1\text{H}$  COSY NMR (500 MHz) spectrum for N-phenethylacetamide, measured in DMSO- $d_6$ .



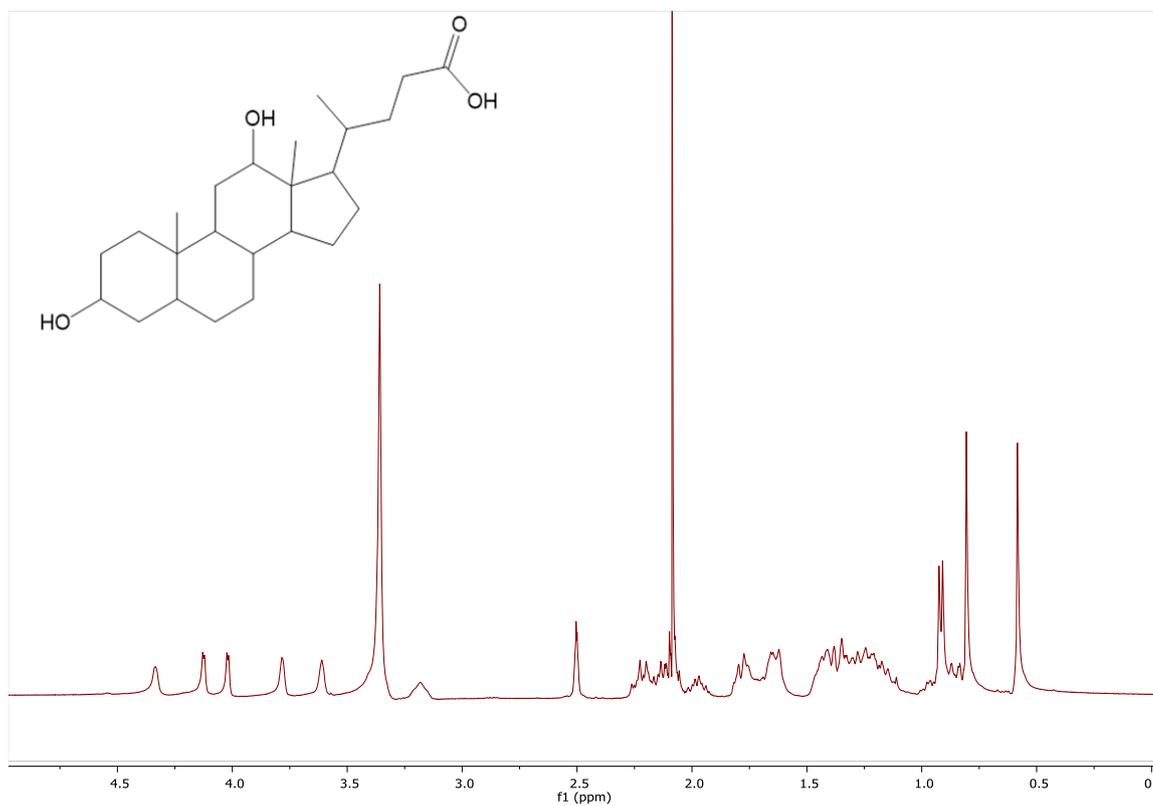
App 3 <sup>13</sup>C NMR (500 MHz) spectrum for N-phenethylacetamide, measured in DMSO-d<sub>6</sub>.



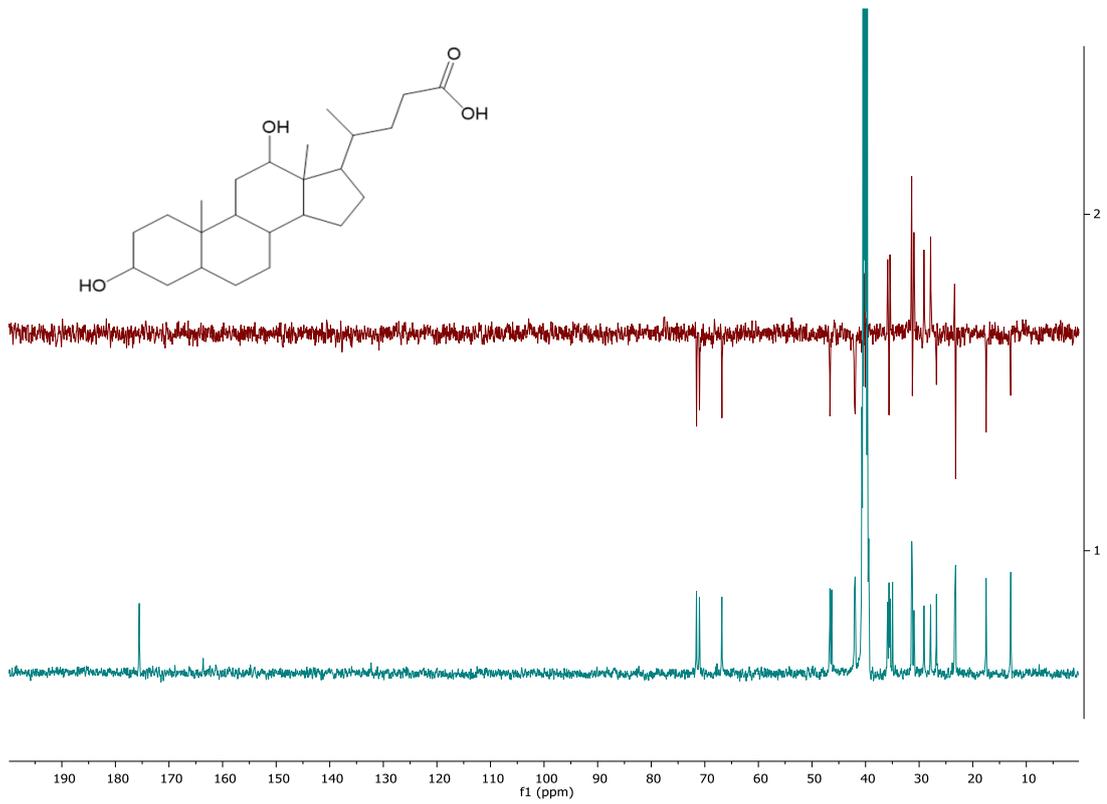
App 4  $^1\text{H}$ - $^{13}\text{C}$  HMBC NMR (500 MHz) spectrum for N-phenethylacetamide, measured in DMSO- $d_6$ .



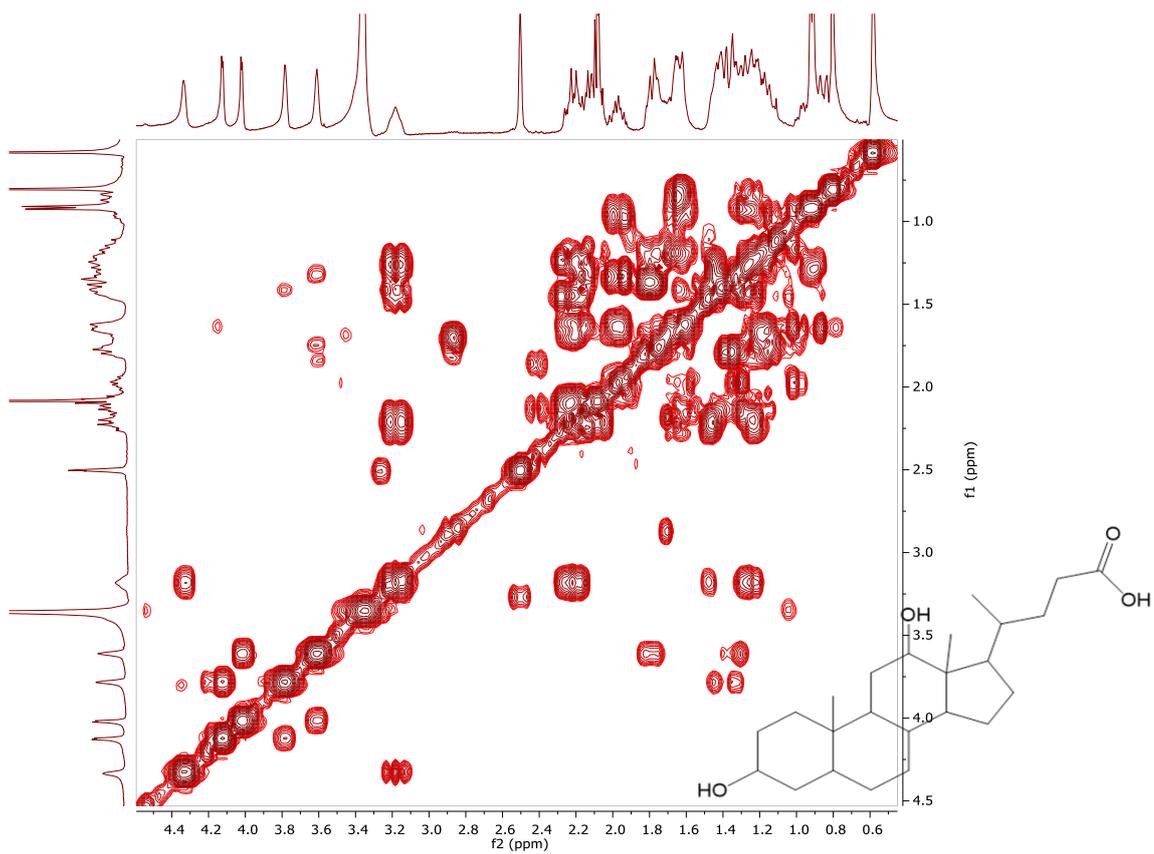
## Appendix II: NMR data of a cholic acid derivative



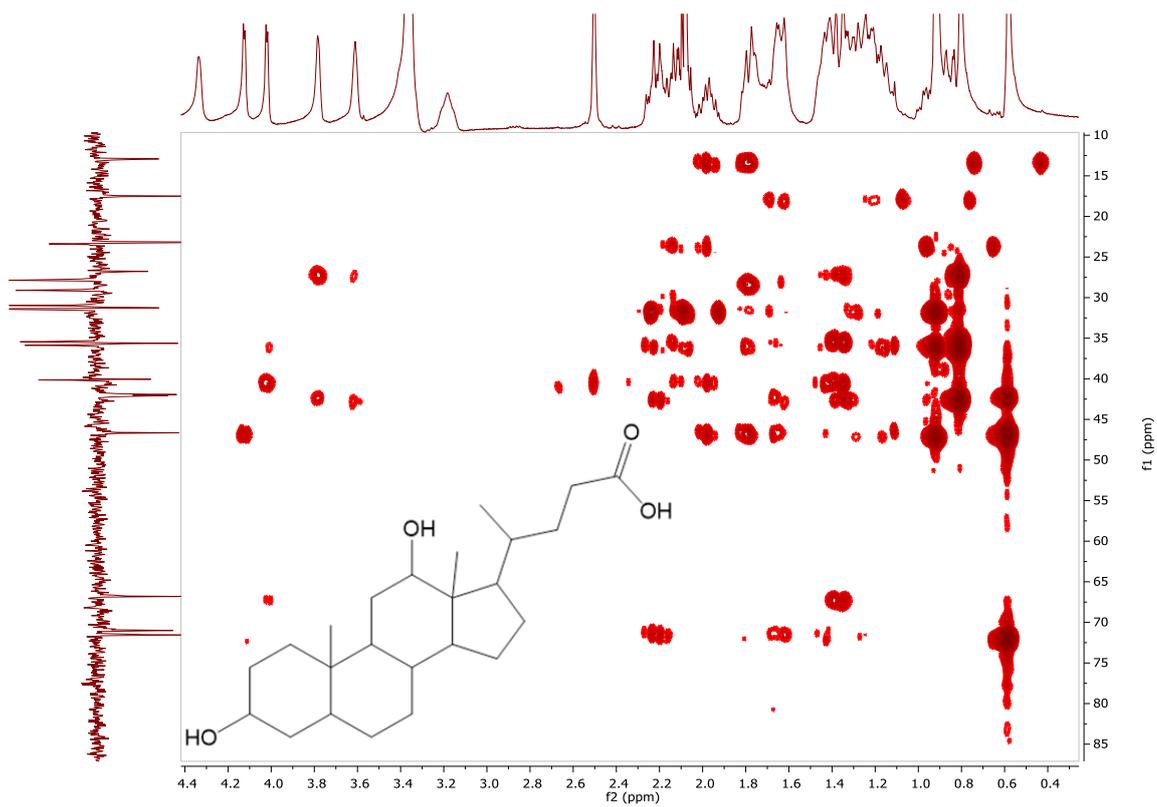
App 6  $^1\text{H}$  NMR (400 MHz) spectrum for cholic acid derivative, measured in  $\text{DMSO-d}_6$ .



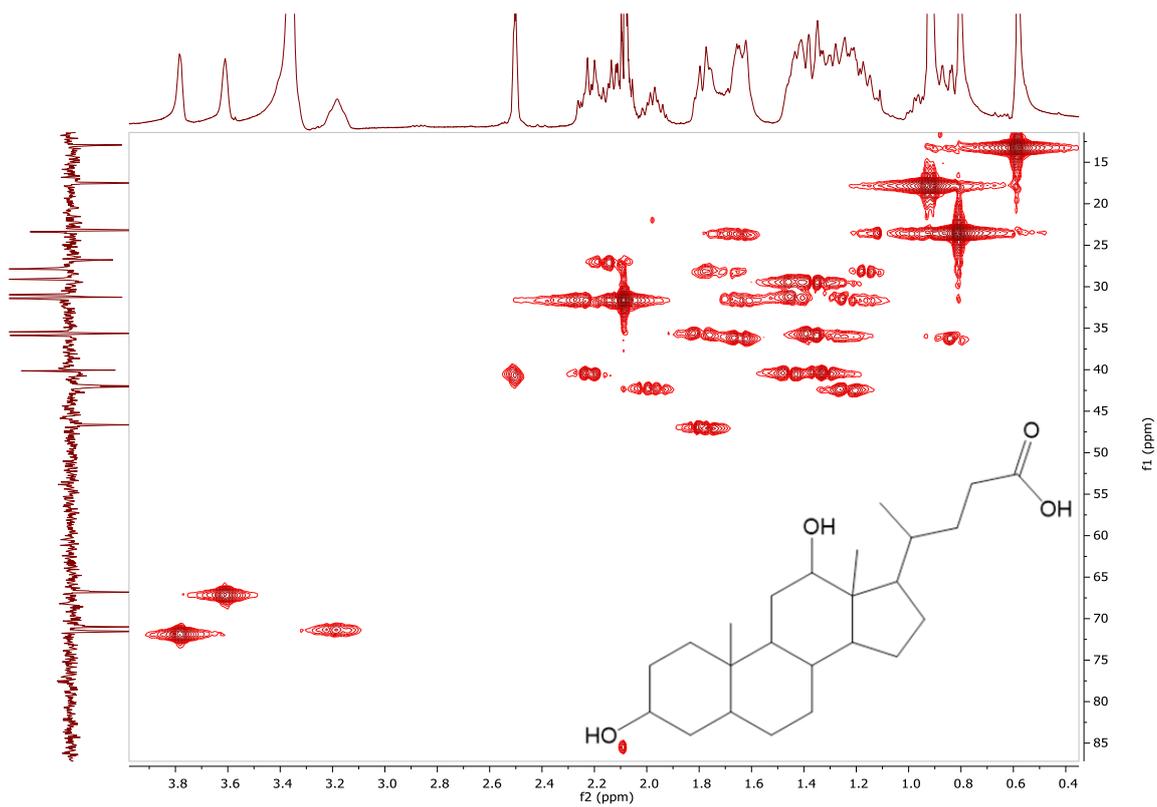
**App 7 Stacked <sup>13</sup>C (1) and DEPT (2) NMR (100 MHz) spectra for cholic acid derivative measured in DMSO-d<sub>6</sub>.**



App 8  $^1\text{H}$ - $^1\text{H}$  COSY NMR (400 MHz) spectrum for cholic acid derivative, measured in  $\text{DMSO-d}_6$ .

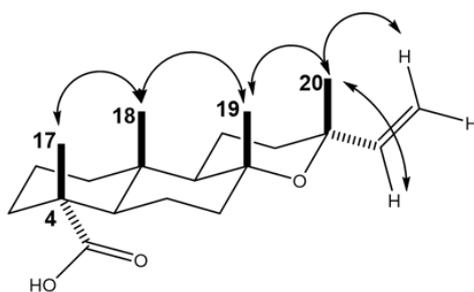
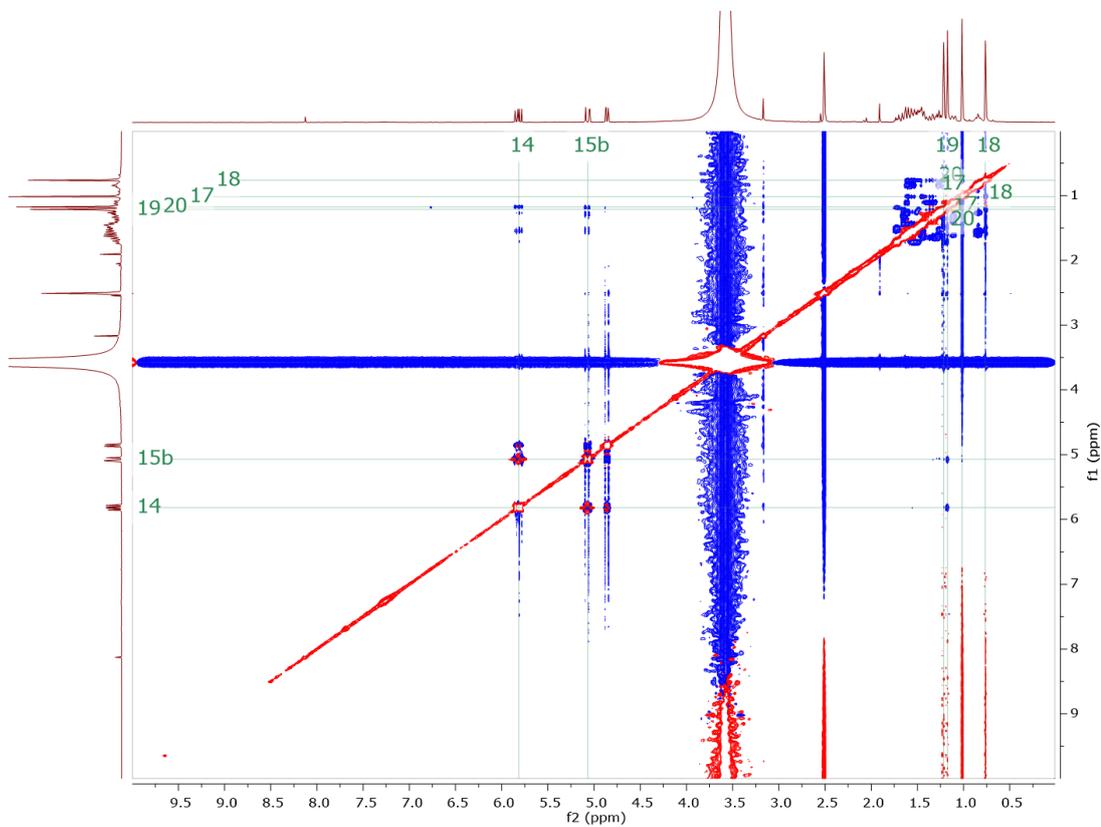


App 9  $^1\text{H}$ - $^{13}\text{C}$  HMBC NMR (400 MHz) spectrum for cholic acid derivative, measured in  $\text{DMSO-d}_6$ .



App 10  $^1\text{H}$ - $^{13}\text{C}$  HMQC NMR (400 MHz) spectrum for cholic acid derivative, measured in  $\text{DMSO-d}_6$ .

### Appendix III: NMR data of 4-epi-manoyl oxide acid



App.11  $^1\text{H}$ - $^1\text{H}$  NOESY (400 MHz) spectrum in  $\text{DMSO-d}_6$  indicated through space correlation of protons in manoyl oxide acid to establish the relative stereochemistry.