

ANALYTICAL STRATEGIES FOR PROFILING, ANNOTATION, AND STRUCTURE  
ELUCIDATION OF SPECIALIZED TERPENOID METABOLITES

By

Ekanayaka Appuhamilage Prabodha Ekanayaka

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Chemistry – Doctor of Philosophy

2014

## ABSTRACT

### ANALYTICAL STRATEGIES FOR PROFILING, ANNOTATION, AND STRUCTURE ELUCIDATION OF SPECIALIZED TERPENOID METABOLITES

By

Ekanayaka Appuhamilage Prabodha Ekanayaka

The main bottleneck in today's plant metabolomics lies with the identification of new metabolites. A number of plant metabolite databases that report liquid chromatography – mass spectrometry data have been constructed. However, more than 95% of compounds reported in these databases remain unannotated. The extensive range of unknown metabolites presents a significant challenge in interpreting metabolome data, and therefore developing methods that accelerate annotation and identification of previously unknown metabolites has great importance when metabolome data are used for functional genomic research.

The challenge of metabolite annotation was addressed by using relative mass defect (RMD) filtering of ion masses measured using liquid chromatography-mass spectrometry. Such calculated RMD values reflect the fractional hydrogen content of each detected ion, and reflect the biosynthetic precursors and transformations that generate metabolites *in vivo*. RMD filtering aids grouping of compounds of similar relative mass defect independent of absolute mass and chromatographic retention time. Therefore, metabolites and metabolite precursors are grouped together enabling potential associations among related metabolites to be developed. Furthermore, a systematic variation of RMD among the fragment/product ions observed in multiplexed collision-induced dissociation (CID) MS or liquid chromatography-tandem mass spectrometry (LC-MS/MS) data for compounds of interest allowed for the identification of terpene glycosides in complex matrices. However, once the metabolites are annotated in metabolomics data sets, establishing the structure of these compounds

requires the purification of the compound followed by *de novo* structure elucidation that relies heavily upon 1D and 2D NMR.

Chapter 2 of this dissertation discusses the application of RMD filtering based data analysis to both parent and fragment ions generated in LC-multiplexed CID MS metabolite profiles generated from wild tomato species *Solanum habrochaites* LA1777. This resulted in the discovery of over 24 novel sesquiterpene glycoside chemical formulas, with multiple isomers comprising a group of more than 200 sesquiterpenoid glycosides. Chapters 3 and 4 of this dissertation discuss the purification and *de novo* structure elucidation of seven example compounds from wild tomato glandular trichomes using NMR. The structures of the sesquiterpenoid cores established for these compounds are different from the structures of known volatile sesquiterpenoid compounds found in *S. habrochaites* LA1777 suggesting that the synthesis of these non-volatile terpenoids involves different biosynthetic enzymes from those involved in the synthesis of known volatile terpenoids.

Similarly, Chapter 5 of this dissertation discusses the application of these techniques to the analysis of metabolite profiles of the medicinal plant *Hoodia gordonii* generated using LC-multiplexed CID MS. This research led to the identification of 24 novel diterpene glycosides. These compounds are believed to share the diterpenoid cores found in some of the known diterpene glycosides from *Hoodia gordonii* and therefore these compounds are likely biosynthetic intermediates of the synthesis of some of the known diterpene glycosides.

Copyright by  
EKANAYAKA APPUHAMILAGE PRABODHA EKANAYAKA  
2014



## ACKNOWLEDGEMENTS

The research discussed in this dissertation would be impossible without the contribution of a number of people. Among all these people the most important person is my advisor Dr. A. Daniel Jones to whom I am most grateful for his guidance, influence and support. He has mentored me through both good and bad times during my career in graduate school and I consider myself lucky to have such a wonderful mentor who I will continue to benefit from for the rest of my life.

I also thank my collaborators Dr. Robert Last's research group and Dr. Cornelius Barry's research group from Michigan State University and Dr. Eran Pichersky's research group from the University of Michigan. It was an honor to learn from them and to work with them. Most of the research discussed here would have been impossible without the help of Dr. Daniel Holmes and Mr. Kermit Johnson of Max T. Rogers NMR facility at Michigan State University and so I am thankful to both of them for their help in generation and interpretation of NMR data presented in this dissertation.

I am also grateful to my committee members, Dr. Dana Spence, Dr. Kevin D. Walker, and Dr. Gary Blanchard for their support and comments on my work. I would like to extend my appreciation to many past and present members (Particularly, to Dr. Chao Li, Dr. Jiangyin Bao and Dr. Ramin Vismeh ) of Dr. Jones's research group for their engaging discussion and their friendship, which bring joy and happiness to my life at Michigan State University.

Completion of this dissertation would have been difficult without the moral and financial support from my loving wife Punsisi (Punsisi Upeka Ratnayake), my undergraduate research supervisor Dr. R.M. Rukmal Ratnayake (currently at Open University of Sri Lanka, Polgolla) and my parents (E.A

Gunatilaka and Mallika Kumarihami Harasgama). I would not have pursued my PhD at Michigan State University without the support of my undergraduate research supervisor Dr. R.M. Rukmal Ratnayake. I am very much grateful to her for her guidance with my work and life. Furthermore, I'd like to give a special thanks to my wife Punsisi, who always encouraged and supported me throughout my time at MSU. Finally, I am thankful to all my friends who made my stay at MSU enjoyable, encouraged and helped me with the completion of the research presented here.

## TABLE OF CONTENTS

LIST OF TABLES .....	ix
LIST OF FIGURES .....	x
Chapter 1 .....	1
1.1 Plants as sources of drugs, food and sustainable energy .....	1
1.2 Modern technology for using plants to support a sustainable human life .....	1
1.3 Growing interest in plant metabolite identification .....	3
1.4 Introduction to terpenes and trichomes .....	6
1.4.1 Trichomes in the genus <i>Solanum</i> and their chemical composition .....	6
1.4.2 Structures and biosynthesis of terpene metabolites in the genus <i>Solanum</i> .....	8
1.4.3 Biosynthesis of terpenes .....	9
1.4.4 Tomato and its terpenoid metabolites .....	13
1.4.5 Characteristic fragment ions for identifying terpenes from plant tissues using Gas Chromatography –Mass Spectrometry .....	15
1.5 Discovery and identification of specialized metabolites .....	16
1.6 Mass spectrometry-based metabolomics as a functional genomics tool .....	18
1.6.1 Mass Spectrometry .....	18
1.6.2 Collision induced dissociation (CID) and MS/MS .....	19
1.6.3 Multiplexed CID mass spectrometry .....	19
1.6.4 Mass spectrometry for discovery of novel metabolites .....	21
1.7 Summary of research .....	23
REFERENCES .....	25
Chapter 2: Strategies for rapid identification of sesquiterpene glycosides from complex matrices using relative mass defect filtering. ....	34
2.1 Introduction .....	34
2.2 Applying accurate mass and relative mass defect filtering for exploring plant metabolomes for the identification novel terpene glycosides .....	35
2.3 Materials and Methods .....	41
2.3.1 LC-MS and MS/MS experiments .....	41
2.3.2 Plant material .....	42
2.2.3 Data processing .....	42
2.4 Results and discussion .....	43
2.4.1 Recognition of sesquiterpene glycosides from ion relative mass defects .....	43
2.4.2 Discovery of sesquiterpene diol glycosides from <i>S. habrochaites</i> LA1777 .....	49
2.5 Conclusions .....	72
REFERENCES .....	74
Chapter 3: Purification and structure elucidation of campherenane diol glycosides from leaf glandular trichomes of <i>Solanum habrochaites</i> LA1777 .....	77
3.1 Introduction .....	77
3.2 Experimental methods .....	78
3.2.1 Plant material .....	78

3.2.2 Purification of campherenane diol glycosides.....	79
3.2.3 NMR Experiments.....	80
3.2.4 Acid hydrolysis, derivatization and GC-MS experiments.....	80
3.3 Results and discussion .....	81
3.3.1 Profiles of the campherenane diol diglycosides .....	81
3.3.2 Structure elucidation of sesquiterpene diol dihexoside malonate ester and establishing the structure of campherenane terpenoid core.....	86
3.3.3 Structure elucidation of sesquiterpene monohexoside malonate ester .....	87
3.3.4 Structure elucidation of sesquiterpene diol dihexoside .....	89
3.3.5 Structure elucidation of sesquiterpene diol dihexoside acetate ester .....	89
3.3.6 Using GC-MS data to support NMR based structure assignments of sugar moieties.....	90
3.4 Conclusions.....	91
APPENDIX.....	92
REFERENCES .....	122

Chapter 4: Purification and structure elucidation of sesquiterpene I and sesquiterpene II glycosides from leaf glandular trichomes of <i>Solanum habrochaites</i> LA1777 .....	126
4.1 Introduction.....	126
4.2 Plant material for metabolite purification .....	127
4.3 Purification of sesquiterpene I and sesquiterpene II glycosides .....	128
4.4 Structure elucidation of sesquiterpene I diol dihexoside malonate ester .....	129
4.5 Structure elucidation of Sesquiterpene II alcohol dihexoside acetate ester .....	131
4.6 Structure elucidation of Sesquiterpene II alcohol dihexoside.....	133
4.7 Conclusion .....	135
APPENDIX.....	136
REFERENCES .....	163

Chapter 5: Discovering terpene glycosides from <i>Hoodia gordonii</i> by mining multiplexed CID UHPLC-MS data and applying relative mass defect filtering .....	165
5.1 Introduction.....	165
5.2 Methods.....	167
5.2.1 LC-MS and LC-MS/MS experiments .....	167
5.2.2 Sample preparation for metabolite profiling .....	168
5.3 Results and discussion .....	168
5.3.1 Hoodia metabolite profiling .....	168
5.3.2 Data processing for and relative mass defect filtering for mining publically available data sets for the discovery of novel terpenoid metabolites .....	169
5.3.3 Variation of RMD of terpene glycosides and other non –terpenoid compounds.....	171
5.3.4 Putative assignments of novel steroidal glycosides from <i>Hoodia gordonii</i> .....	171
5.4 Conclusions.....	175
APPENDIX.....	176
REFERENCES .....	190

Chapter 6: Concluding remarks .....	192
-------------------------------------	-----

## LIST OF TABLES

Table 2.1: Characteristic fragment ions observed in negative ion mode MS/MS experiments for various sugar oligosaccharides and mono saccharides. The masses shown correspond to $[M-H]^-$ formed by each sugar group. The MS/MS spectra of candidate terpenoid compounds were examined for the presence of these fragment ions for identification of the presence of these oligosaccharides in the terpenoids. ....	39
Table 2.2: Compounds identified as terpene glycosides from <i>S. habrochaites</i> LA1777 based on RMD filtering of molecular and fragment ions. ....	56
Table 2.3: Compounds with greatest peak areas among the list of <i>S. habrochaites</i> LA1777 metabolites in the RMD range 440-636 ppm.....	59
Table 3.1: Chemical shifts (in ppm) of $^1H$ and $^{13}C$ resonances for (a) campherenane diol diglucoside malonate ester (peak 6a from Figure 3.1), (b) campherenane diol monoglucoside malonate ester (peak 6b), (c) campherenane diol diglucoside (peak 8c), and (d) campherenane diol diglucoside acetate ester (peak 5d). ....	93
Table 4.1: Chemical shift assignments to the carbons and protons of the three compounds purified from <i>S. habrochaites</i> LA1777.....	162

## LIST OF FIGURES

Figure 1.1 Chemical Structure of isoprene, the simplest terpene. ....	8
Figure 1.2 Mevalonate pathway of terpene biosynthesis. ....	11
Figure 1.3 Biosynthesis of isopentenyl pyrophosphate starting with pyruvic acid. ....	12
Figure 2.1 Complexity of a plant extract is evident from the number of peaks in a UHPLC-MS base peak chromatogram generated from a leaf dip extract of <i>S. habrochaites</i> LA1777. Analysis was performed using a 110 min chromatographic gradient and detected in negative ion mode. ....	43
Figure 2.2: MarkerLynx XS data processing results in a list of markers detected from the <i>S. habrochaites</i> LA1777 trichome leaf extracts. Markers detected are reported with their <i>m/z</i> ratio, retention time and their abundance in each sample being considered. ....	44
Figure 2.3. (A) Negative ion mode multiplexed CID mass spectrum of acylsugar S4:22, RMD = 492 ppm. (B) Negative ion mode multiplexed CID mass spectrum of acylsugar S4:23, RMD = 402 ppm. (C) Negative ion mode multiplexed CID mass spectrum of acylsugar S4:17, RMD = 440 ppm. Values for RMD of the major fragment ions are presented. All detected isomers displayed fragments of the same <i>m/z</i> values in negative ion mode spectra. All displayed negative ion mode multiplexed CID mass spectra were obtained using a collision potential of -60 V. ....	48
Figure 2.4 Negative-ion mode product ion MS/MS spectra of metabolites extracted from <i>S. habrochaites</i> LA1777. (A) MS/MS spectrum of <i>m/z</i> 609 from campherenane diol diglucoside; (B) MS/MS spectrum of <i>m/z</i> 811 from campherenane diol triglycoside malonate ester ( <i>m/z</i> 811.3587, RMD = 442 ppm). Values for RMD of the major fragment ions are presented. All chromatographically-resolved isomers (10 isomers of <i>m/z</i> 651 and 12 isomers for <i>m/z</i> 811) displayed fragments of the same <i>m/z</i> values in negative ion mode spectra. All negative ion mode MS/MS data were obtained using collision potential of -50 V. ....	51
Figure 2.5: Negative ion mode multiplexed CID mass spectrum of tomatine from <i>S. habrochaites</i> LA1777 obtained at 60 V collision potential. Note that the fragment ions all have RMD values greater than or equal to the [M-H] <sup>-</sup> ion at <i>m/z</i> 1032.5. ....	53
Figure 2.6: Variation of RMD of fragment ions as a function of ion <i>m/z</i> . Fragment ions were generated in negative ion mode MS/MS for some representative compounds (acyl sugars, sesquiterpene diglycoside malonate esters, triterpenoid glycoside and a triglycoside malonate ester) found in <i>S. habrochaites</i> LA1777 leaf dip extracts. ....	54
Figure 2.7: Relative mass defect filtering process used for discovering conjugated terpenoids from raw LC-MS data. ....	55
Figure 2.8. (A) Negative ion mode product ion MS/MS spectrum of products from [M-H] <sup>-</sup> ( <i>m/z</i> 649) for campherenan-2,12-diol malonate ester. (B) magnified region of <i>m/z</i> 228-255, showing the	

sesquiterpenoid core fragment at $m/z$ 239, which was too small to observe in A, where peaks are normalized to the base peak. ....	60
Figure 2.9. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 487) for campherenan-2,12-diol monoglucoside malonate ester. The lower spectrum displays the magnified region of $m/z$ 224-256, showing the sesquiterpenoid core fragment at $m/z$ 239.....	61
Figure 2.10. Negative ion mode product ion MS/MS spectrum of products from $[M+formate]^-$ ( $m/z$ 609) for campherenan-2,12-diol diglucoside. The lower spectrum displays the magnified region of $m/z$ 228-258, showing the sesquiterpenoid core fragment at $m/z$ 239 .....	62
Figure 2.11. Negative ion mode product ion MS/MS spectrum of products from $[M+formate]^-$ ( $m/z$ 651) for campherenan-2,12-diol diglucoside acetate ester. The lower spectrum displays the magnified region of $m/z$ 224-264, showing the sesquiterpenoid core fragment at $m/z$ 239.....	63
Figure 2.12. Negative ion mode product ion MS/MS spectrum of products from $[M+formate]^-$ ( $m/z$ 771) for a Campherenane-2,12-diol triglycoside (Compound 13 in Table 2.2) .....	64
Figure 2.13. Negative ion mode product ion MS/MS spectrum of products from $[M+formate]^-$ ( $m/z$ 591) for a sesquiterpene II dihexoside (Compound 1 in Table 2.2) .....	64
Figure 2.14. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 631) for a sesquiterpene II dihexoside malonate ester (Compound 3 in Table 2.2) .....	65
Figure 2.15. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 793) for a sesquiterpene II trihexoside malonate ester (Compound 5 in Table 2.2).....	65
Figure 2.16. Negative ion mode product ion MS/MS spectrum of products from $[M+HCOO]^-$ ( $m/z$ 633) for a sesquiterpene II dihexoside acetate ester (Compound 2 in Table 2.2).....	66
Figure 2.17. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 661) for a sesquiterpene I diglycoside malonate ester (Compound 22 in Table 2.2). ....	66
Figure 2.18. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 499) sesquiterpene I monoglycoside malonate ester (Compound 20 in Table 2.2). ....	67
Figure 2.19. Negative ion mode product ion MS/MS spectrum of $[M-H]^-$ ( $m/z$ 455) for Sesquiterpene I monoglycoside acetate ester (Compound 19 in Table 2.2). ....	67
Figure 2.20. Negative ion mode product ion MS/MS spectrum of $[M-H]^-$ ( $m/z$ 413) of Sesquiterpene I monoglycoside (Compound 18 in Table 2.2). ....	68
Figure 2.21. Negative ion mode product ion MS/MS spectrum of $[M+HCOO]^-$ ( $m/z$ 499) for sesquiterpene I monoglycoside malonate ester (Compound 20 in Table 2.2). ....	68

Figure 2.22. Negative ion mode product ion MS/MS spectrum of $[M-H]^-$ ( $m/z$ 497) of Sesquiterpene III monoglycoside malonate ester (Compound 17 in Table 2.2).	69
Figure 2.23. Negative ion mode product ion MS/MS spectrum of products of $[M-H]^-$ ( $m/z$ 411) of Sesquiterpene III monoglycoside (compound 15 in Table 2.2).	69
Figure 2.24. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 617) for Sesquiterpene I diglycoside acetate ester (compound 21).	70
Figure 2.25. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 823) for Sesquiterpene I triglycoside malonate ester (compound 23 in Table 2.2).	70
Figure 2.26. Negative ion mode product ion MS/MS spectrum of products from $[M-H]^-$ ( $m/z$ 985) for Sesquiterpene I tetraglycoside malonate ester (compound 24 in Table 2.2).	71
Figure 2.27. Negative ion mode product ion MS/MS spectrum of products from $[M+HCOO]^-$ ( $m/z$ 489) for Campherenane diol Monoglycoside acetate ester (compound 7 in Table 2.2).	71
Figure 2.28. Negative ion mode product ion MS/MS spectrum of products from $[M+HCOO]^-$ ( $m/z$ 753) Sesquiterpene II Triglycoside (compound 4 in Table 2.2).	72
Figure 3.1. Extracted ion UHPLC-MS chromatograms showing sesquiterpene glycoside metabolites identified in extracts from <i>Solanum habrochaites</i> LA1777. (a) $[M-H]^-$ ( $m/z$ 649) for campherenane diol diglucoside malonate esters, (b) $[M-H]^-$ ( $m/z$ 487) for campherenane diol monoglucoside malonate esters, (c) $[M+formate]^-$ ( $m/z$ 609) for campherenane diol diglucosides, (d) $[M+formate]^-$ ( $m/z$ 651) for campherenane diol diglucoside acetate esters, (e) $[M-H]^-$ ( $m/z$ 811) for campherenane diol triglucoside malonate esters, (f) $[M+formate]^-$ ( $m/z$ 771) for campherenane diol triglucosides, (g). $[M+formate]^-$ ( $m/z$ 446) for campherenane diol monoglucosides and (h) $[M+formate]^-$ ( $m/z$ 489) for campherenane diol monoglucosides acetate ester. Labels with larger font size in red designate the four purified metabolites.	85
Figure 3.2: Structures of (a) campherenane-2,12-diol diglucoside malonate ester (peak 6a from Figure 1), (b) campherenane-2,12-diol monoglucoside malonate ester (peak 6b), (c) campherenane-2,12-diol diglucoside (peak 8c), and (d) campherenane-2,12-diol diglucoside acetate ester (peak 5d) as determined using NMR and tandem mass spectrometry. Portions of the molecule corresponding to key fragment ions in negative ion MS/MS spectra are illustrated on each structure. Carbon atoms are numbered in accordance with NMR assignments in Table 3.1.	95
Figure 3.3. GC/MS total ion chromatograms of methoxime-trimethylsilyl derivatives of (a) products of acid hydrolysis of campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound 1; peak 6a from Figure 1), (b) glucose reference standard, (c) xylose reference standard, (d) galactose reference standard, (e) rhamnose reference standard.	96
Figure 3.4. $^1H$ and $^{13}C$ NMR spectra of the isolated campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound 1; peak 6a from Figure 3.1).	97



Figure 3.5. $^1\text{H}$ - $^1\text{H}$ COSY NMR spectrum of the campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound <b>1</b> ; peak 6a from Figure 3.1).	98
Figure 3.6. Multiplicity edited gHSQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectrum of campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound <b>1</b> ; peak 6a in Figure 1).	99
Figure 3.7. gHMBC NMR spectrum of campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound <b>1</b> ; peak 6a from Figure 1).	100
Figure 3.8. TOCSY NMR spectrum of campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound <b>1</b> ; peak 6a in Figure 1).	101
Figure 3.9. NOESY NMR spectrum of campherenane-2- <i>endo</i> -(6'-malonyl)glucoside)-12-glucoside (Compound <b>1</b> ; peak 6a in Figure 1).	102
Figure 3.10. $^1\text{H}$ and $^{13}\text{C}$ NMR spectra of the isolated 2- <i>endo</i> -campherenanol-12-(6'-malonyl)glucoside (Compound <b>2</b> ; peak 6b from Figure 1).	103
Figure 3.11. $^1\text{H}$ - $^1\text{H}$ COSY NMR spectrum of the 2- <i>endo</i> -campherenanol-12-(6'-malonyl)glucoside (Compound <b>2</b> ; peak 6b from Figure 1).	104
Figure 3.12. Multiplicity edited gHSQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectrum of 2- <i>endo</i> -campherenanol-12-(6'-malonyl)glucoside (Compound <b>2</b> ; peak 6b in Figure 1).	105
Figure 3.13. gHMBC NMR spectrum of 2- <i>endo</i> -campherenanol-12-(6'-malonyl)glucoside (Compound <b>2</b> ; peak 6b from Figure 1).	106
Figure 3.14. TOCSY NMR spectrum of 2- <i>endo</i> -campherenanol-12-(6'-malonyl)glucoside (Compound <b>2</b> ; peak 6b in Figure 1).	107
Figure 3.15. NOESY NMR spectrum of 2- <i>endo</i> -campherenanol-12-(6'-malonyl)glucoside (Compound <b>2</b> ; peak 6b in Figure 1).	108
Figure 3.16. $^1\text{H}$ and $^{13}\text{C}$ NMR spectra of the isolated campherenan-2- <i>endo</i> ,12-diglucoside (Compound <b>3</b> ; peak 8c from Figure 3.1).	109
Figure Figure 3.17. $^1\text{H}$ - $^1\text{H}$ COSY NMR spectrum of the campherenan-2- <i>endo</i> ,12-diglucoside (Compound <b>3</b> ; peak 8c from Figure 1).	110
Figure 3.18. Multiplicity edited gHSQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectrum of campherenan-2- <i>endo</i> ,12-diglucoside (Compound <b>3</b> ; peak 8c in Figure 1).	111
Figure 3.19. gHMBC NMR spectrum of campherenan-2- <i>endo</i> ,12-diglucoside (Compound <b>3</b> ; peak 8c from Figure 1).	112

Figure 3.20. TOCSY NMR spectrum of campherenan-2- <i>endo</i> ,12-diglucoside (Compound 3; peak 8c in Figure 1).....	113
Figure 3.21. 2D-NOESY NMR spectrum of campherenan-2- <i>endo</i> ,12-diglucoside (Compound 3; peak 8c in Figure 1).....	114
Figure 3.22. Proton NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 1).....	115
Figure 3.23. <sup>13</sup> C NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 3.1) .....	116
Figure 3.24. Multiplicity edited 2D gHSQC NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 1).....	117
Figure 3.25. 2D gdqCOSY NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 1).....	118
Figure 3.26. 2D HMBC NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 1).....	119
Figure 3.27. 2D TOCSY NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 1).....	120
Figure 3.28. 2D NOESY NMR spectrum of 2- <i>endo</i> -campherenanol-12-(2-(6"-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 1).....	121
Figure Figure 4.1: HPLC-MS extracted ion chromatogram profiles of the three compounds purified from <i>S. habrochaites</i> LA1777. a.) 10 isomers of Sesquiterpene I diol dihexoside malonate ester were separated (b) three isomers of [M+HCOO <sup>-</sup> ] of Sesquiterpene II alcohol dihexoside acetate ester were detected (c) 8 isomers of [M+HCOO <sup>-</sup> ] of Sesquiterpene II alcohol dihexoside were separated.....	137
Figure 4.2. Structures of (a) Sesquiterpene I diol dihexoside malonate ester (peak 8a from Figure 4.1), (b) Sesquiterpene II alcohol dihexoside acetate ester (peak 3b) and (c) Sesquiterpene II alcohol dihexoside (peak 6c), determined using NMR and tandem mass spectrometry. Portions of the molecule corresponding to key fragment ions in negative ion MS/MS spectra are illustrated on each structure. Carbon atoms are numbered in accordance with NMR assignments in Table 4.1 .....	138
Figure 4.3. 1D Proton NMR spectrum of compound a (peak 8a from Figure 4.1).....	139
Figure 4.4. 2D COSY spectrum of compound a (peak 8a from Figure 4.1).....	140
Figure 4.5. 2D HSQC spectrum of compound a (peak 8a from Figure 4.1).....	141
Figure 4.6. 2D HMBC spectrum of compound a (peak 8a from Figure 4.1).....	142

Figure 4.7. 2D TOCSY spectrum of compound a (peak 8a from Figure 4.1). .....	143
Figure 4.8. 2D NOESY spectrum of compound a (peak 8a from Figure 4.1). .....	144
Figure 4.9. 1D Proton NMR spectrum of compound b (peak 3b from Figure 4.1) .....	145
Figure 4.10. 2D COSY spectrum of compound b (peak 3b from Figure 4.1) .....	146
Figure 4.11. 2D HSQC spectrum of compound b (peak 3b from Figure 4.1) .....	147
Figure 4.12. 2D HMBC spectrum of compound b (peak 3b from Figure 4.1). .....	148
Figure 4.13. 2D TOCSY spectrum of compound b (peak 3b from Figure 4.1). .....	149
Figure 4.14. 1D NMR spectrum of compound c (peak 6c from Figure 4.1) .....	150
Figure 4.15. 2D COSY spectrum of compound c (peak 6c from Figure 4.1). .....	151
Figure 4.16. 2D HSQC spectrum of compound c (peak 6c from Figure 4.1). .....	152
Figure 4.17. 2D HMBC spectrum of compound c (peak 6c from Figure 4.1). .....	153
Figure 4.18. 2D TOCSY spectrum of compound c (peak 6c from Figure 4.1). .....	154
Figure 4.19. 2D NOESY spectrum of compound c (peak 6c from Figure 4.1). .....	155
Figure 4.20. GC-MS chromatogram and EI spectrum of most abundant peak from the acid hydrolysis product of compound b. ....	156
Figure 4.21. GC-MS chromatogram and EI spectrum of most abundant peak from the acid hydrolysis product of compound c. ....	157
Figure 4.22. HMBC correlations assigned for compound <b>a</b> .....	158
Figure 4.23. Important COSY correlations assigned for compound <b>a</b> . .....	158
Figure 4.24. NOE correlations assigned for compound <b>a</b> . .....	158
Figure 4.25. HMBC correlations assigned for compound <b>b</b> .....	159
Figure 4.26. COSY correlations assigned for compound <b>b</b> .....	159
Figure 4.27. NOE correlations assigned for compound <b>b</b> .....	160
Figure 4.28. HMBC correlations assigned for compound <b>c</b> .....	160

Figure 4.29. COSY correlations assigned for compound c .....	161
Figure 4.30. NOE correlations assigned for compound c .....	161
Figure 5.1: Five diterpenoid cores reported from <i>Hoodia</i> compounds (a-e) and the six different sugar groups that are found attached to the terpenoid core in current literature. ....	177
Figure 5.2: Some representative diterpenoids found in <i>Hoodia</i> . ....	178
Figure 5.3: Distribution of RMD of compounds detected in <i>Hoodia gordonii</i> from metabolite profiling in different RMD range categories. ....	179
Figure 5.4: Abundance of known and noval metabolites annotated from <i>Hoodia gordonii</i> .....	180
Figure. 5.5 Positive ion MS/MS product ion spectrum of Hoodia metabolite $m/z$ 803 $[M+H]^+$ .....	181
Figure 5.6 Positive ion MS/MS product ion spectrum of Hoodia metabolite $m/z$ 838 $[M+NH_4]^+$ ....	181
Figure 5.7 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 641.....	182
Figure 5.8 Positive ion product ion MS/MS spectrum of Hoodia metabolite of $m/z$ 1304. ....	183
Figure 5.9 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 1190.....	184
Figure 5.10 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 920.....	184
Figure 5.11 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 998.....	185
Figure. 5.12 Positive ion MS/MS product ion spectrum of $m/z$ 1316.....	185
Figure. 5.13 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 1206.....	186
Figure. 5.14 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 1222.....	186
Figure. 5.15. Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 1192.....	187
Figure. 5.16 Positive ion MS/MS product ion spectrum of Hoodia metabolite of $m/z$ 916.....	187
Figure. 5.17 Positive ion MS/MS product ion spectrum of (a) $m/z$ 983 (putative alkaloid), (b) Variation of RMD of parent ion and product ions in the MS/MS spectrum of an alkaloid compared to the diterpene glycoside (RMD of sugar fragments are not included). The diterpene glycoside displays a gradual increase of RMD from the parent ion to fragment ions whereas the alkaloid displays a constant RMD from both fragment ions and parent ions.....	188
Figure. 5.18 Positive ion MS/MS product ion spectra of <i>Hoodia</i> standards. ....	189

## **Chapter 1**

### **1.1 Plants as sources of drugs, food and sustainable energy**

Since prehistoric times, humans have depended on plants as sustainable sources of food and for treatments of illness [1]. However this dependence has been challenged by the increasing world population, now exceeding 7 billion, that places a greater stress on limited resources (land, water and clean air) to generate food and other necessities for human life. In addition, advancing standards of living in many parts of the world have changed styles of consumption, resulting in increased demand for fruits, vegetables, meat and dairy products [2]. Furthermore, increasing scarcity of fossil fuels has resulted in greater focus on biofuel usage and using plants as derived sources as industrial feedstocks [3,4]. All of these factors provide a driving force to better understand the genetic capacity of organisms to produce valuable chemicals.

Plants, in contrast to animals, accumulate a vast diversity of chemicals (phytochemicals). This makes them an invaluable source of compounds that can be difficult or expensive to produce via synthetic means. Plants possess this capability owing to their complicated genomes through mutations and gene duplications that have evolved to support sessile forms of life that depend on accumulation of chemicals for its survival (reproduction, defense, energy accumulation)[5,6].

### **1.2 Modern technology for using plants to support a sustainable human life**

This increased dependence of human lives on plants has led to a demand for plants that are more productive, more resistant to disease and drought, and plants that produce better-flavored products. As a result, genetic engineering and synthetic biology approaches have been employed to engineer plants/microbes with greater crop potential, stronger resistance to extreme conditions and more favorable characteristics [7]. However, successful application of synthetic biology

approaches requires understanding relevant biosynthetic pathways and having means of engineering microbes or plants to produce compounds of interest [7,8]. The main challenges in the process is that biology of living systems is often a result of synergetic interactions between multiple biosynthetic pathways, biological compartments and the external environment of the living system [8]. Therefore, in order to properly understand biological systems it is necessary to understand the expression of the genes (transcriptomics), expression of the proteins (proteomics) and the metabolites synthesized (metabolomics) leading to a thorough understanding of the biological system considered. A transcriptome represents the complete set of transcripts (mRNA, tRNA and rRNA) in a cell, an organ or organism, including information about the quantity of each, at a specific developmental stage or physiological condition [9]. Although transcriptomes provide information about levels of expression of individual genes, understanding the processes taking place within cells requires quantitative analysis of the entire complement of proteins, known as the proteome [10,11]. While these two approaches characterize transcription of genes and translation into proteins they represent potential cell functions, but often fail to address realized cellular functions. The suite of small molecules, or metabolites, in cells of tissues reflects the functions of cellular proteins, and the study of the whole spectrum of small molecules (molecules of mass less than 1500 Da are considered small molecules) is referred to as “metabolomics”[12]. Recent advances in biotechnology has resulted an accumulation of large amount of transcriptomic, proteomic, and metabolomics data from various biological systems. However, the main challenge has been understanding the relationship between these in order to understanding the function of those biological systems. This has been the theme of a number of investigations in recent years.

Such investigations begin from the central dogma of biology where one gene was believed to be associated with a single protein. Therefore, initial interest in understanding the genomes of organisms led to sequencing whole genomes (*e.g. Arabidopsis* genome, human genome project). However, it was soon realized that understanding the functions of genes is not accomplished as easily due to the presence of multiple gene sequences that display great sequence similarity. On the other hand, it was also discovered that slight differences in the sequences might lead to completely different functionalities [13]. This made assigning the functions of the genes challenging [13].

One main approach used in assigning gene functions in plants is called functional genomics, where correlations between mRNA expression level and accumulation levels of specific metabolites aid identification of a list of candidate genes that may be responsible for the synthesis of various metabolites. Reductions in cost of sequencing and quantifying expressed mRNA (Expressed Sequence Tags –ESTs) in a global approach, combined with quantification of metabolites from the same tissue pools to assign putative functions of the genes [14]. However, this approach provides only a set of candidate genes, and to confirm the identified candidate genes, further experimentation that involves “knockdown of expressed sequences” such as virus induced gene silencing (VIGS) or RNA interference (RNAi) is required. This makes mining the metabolome to identify the metabolites that are a crucial function to the functional genomics-based gene discovery process.

### **1.3 Growing interest in plant metabolite identification**

Plants display the presence of a vastly diverse set of small molecules that are classified as primary or secondary metabolites [15]. Primary metabolites are defined as metabolites that are common to all organisms and specialized (secondary) metabolites represent all other small

molecules that are found only in select groups of plants due to the specialized biosynthetic pathways present in those groups of plants [15]. The metabolic precursors of specialized (secondary) metabolites are often central metabolites or their activated forms (amino acids for peptides and proteins; isoprenyl-pyrophosphates for terpenes/terpenoids; acyl-CoA thioesters for polyketides)[16,17]. However, compared to primary metabolites, specialized metabolites display a vast diversity that is often restricted to a narrow range of genotypes[15]. This diversity is partly attributed to mutated enzyme functions that exhibit altered substrate selectivity. Some compounds undergo structural changes after biosynthesis of a core molecule (*e.g.* glycosylations, phosphorylations, conjugation with other secondary metabolites) that also contributes to metabolite structure diversity [18]. In comparison, some other compounds undergo regioselective conformational changes that makes them regiospecific toward target receptors[19]. Today a great interest is expressed toward metabolite identification both in plant sciences and in the drug discovery field. However, the identification of novel metabolites and establishing their structures remains one of the great bottlenecks that limits research progress in these areas.

In plant sciences, metabolite identification is a necessary step toward complete characterization of metabolic phenotypes. This allows for the identification of stress-induced responses in plants that is important in agricultural applications. For instance, plants exposed to drought accumulate different levels of compounds than those grown in normal conditions[20,21]. Similarly, exposure to insect attacks causes changes in the chemical composition of plants, most notably in levels of specialized metabolites[22]. Novel genome sequencing technologies have enabled the rapid generation of genome sequence information. Applications of phenotype characterization and co-expression gene analysis allows for accurate gene function annotation [23,24]. Such function



annotation of genes has led to genetic engineering of plants that resulted in the development of crops of better taste, plants with more insect resistance and drought resistance [25,26].

However, it is likely that the greatest beneficiary from natural product structure elucidation is the pharmaceutical industry. Natural products-based drug discovery benefits from metabolite identification for both discovering novel drugs from natural products as well as for characterization of biotransformation of new chemical entities [27,28,29]. Nature produces a diverse collection of molecules (both small and large molecules) that are complementary in structure to their targets, making some of them excellent therapeutics. This has made natural products obtained from plants (or other natural sources including fungi, animals, bacteria) an excellent source of candidate drugs. Low molecular weight neurotransmitter molecules (*e.g.* noradrenaline, adrenaline, serotonin, melatonin, histamines) are a good example of structural complementarity of natural products to their targets. Identifying and elucidating the structures of these compounds allowed for development of synthetic analogs that have become “blockbuster drugs” (*e.g.* antihistamines and variants of the terpene indole alkaloid camptothecin that are used in cancer treatments).

Theories have been proposed to explain why natural products display biological activity toward humans/animals. One proposal is that the long term co-evolution of humans/animals with the environment (plants) has led to the accumulation of compounds in nature that displays high structural fidelity toward enzymes/receptors in humans/animals and *vice versa* [1]. Another theory proposes that plants and animals originated from a common ancestor. The theory proposes that this common ancestor was capable of accumulating a large diversity of compounds that aided its survival under stressed environment in early evolutionary history. Later, as the plants and animals evolved in different paths, the animals lost their ability to synthesize these

compounds whereas plants maintained it. However, the animals still retained their ability to sense these compounds in plants that enabled their survival, while depending on plants as the carbon source. Therefore nature consists of a collection of molecules that display structural fidelity toward receptors in animals/humans [1,30].

## **1.4 Introduction to terpenes and trichomes**

### **1.4.1 Trichomes in the genus *Solanum* and their chemical composition**

Plants display “trichomes” that consist of specialized epidermal cells found on the surfaces of leaves, stems, and other tissues. Trichomes can be either secretory glandular or non-secretory epidermal cells. Secreting glandular trichomes (SGT) use the space between their gland cell walls and cuticle to accumulate large quantities of metabolic products that act as a first line of chemical defense to protect the plant from herbivory [31]. Some of the volatile compounds (such as terpenes) in trichomes produce unpleasant odors making the plants less attractive as a food source for insects [32]. Some other compounds such as acyl sugars accumulated in SGTs defend plants in a more indirect manner. The trichomes and their constituents are commonly consumed by lepidopteran larvae. However, acyl sugars contained in the trichomes release volatile odor compounds once they are subjected to alkaline hydrolysis in midguts of the larvae. These odor compounds attract the predators that attack the larvae and thereby protect the plant [33].

Comparatively, non-glandular trichomes act as physical or mechanical deterrents to insect oviposition and feeding. Some other non-glandular trichomes (hooked trichomes) are believed to be involved in plant defense by impaling insects [34]. About 30% of vascular plants possess secretory glandular trichomes [35]. Trichomes are found across the genus *Solanum* including tomato, tobacco, and potato[36,37,38]. Cultivated tomato (*Solanum lycopersicum*) and its wild

relatives display several different types of trichomes on hypocotyls, stems, leaves, floral organs, and immature fruit.

In 1943, Luckwill identified four distinct types of secretory glandular trichomes (SGT) among these based on their morphologies [39,40]. In Luckwill's classification, type I trichomes were distinguished from others by their multicellular base, a long (~2 mm) multicellular stalk, and a small glandular tip. In comparison, according to Luckwill, type IV trichomes have shorter stalk (~0.3 mm) and have a unicellular base, a multicellular stalk and a small glandular tip. Type VI trichomes consist of a four-celled glandular head and a short (~0.1 mm) multicellular stalk. Type VII trichomes consist of a yet shorter (<0.05 mm) unicellular stalk and an irregularly shaped 4- to 8-celled gland [40]. Type II, III and type V trichomes found on the surfaces of cultivated and wild species of tomato are non-glandular trichomes [40].

Among the secretory glandular trichomes in plants of the genus *Solanum*, not only do the trichomes differ in their morphology and the distribution, their chemical content also varies. For instance, type VI trichomes are selective in accumulating terpenoids metabolites whereas type I glands are the main sites of acylsugar accumulation in *S. habrochaites* [41,42]. However, recent research suggests that type I and IV identified by Luckwill, are closely similar in chemical content and gene expression[42]. One main objective of research on trichome chemistry is to discover the genetic basis for control of trichome chemistry.

Understanding the factors that control trichome chemistry is expected to lead to development of pest-resistant crops that can meet tomorrow's agricultural needs. It has been demonstrated that as

wild plants were domesticated, they lost their ability to synthesize some defense compounds observed in wild plants. As a result, most domesticated plants are susceptible to attack from pests and various diseases [42]. It is expected that reintroduction of the lost biochemical machinery for defense compound accumulation into crop species will generate crops that are more resistant to insect and pathogenic attack.

#### 1.4.2 Structures and biosynthesis of terpene metabolites in the genus *Solanum*

##### *Structures of terpenes*

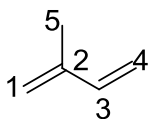


Figure 1.1 Chemical Structure of isoprene, the simplest terpene.

Terpene structure was first explained by *Ruzicka and Wallach* [43] based on 2-methylbutadiene (Figure 1.1), also called “isoprene”, structure. They proposed that terpenes are synthesized by the head-to-head, head-to-tail or tail-to-tail linkages between isoprene units. This proposal, now known as the isoprene rule, is adhered to by some terpenes (regular terpenes) but plenty of exceptions, known as irregular terpenes, occur in nature. The isoprene rule proposes that in natural terpenes, 1-1 (head-to-head) or 4-4 (tail-to-tail) linkages that connect the numbered carbon atoms (Figure 1.1) are not normal. Examples of regular terpenes include the monoterpenes (with 10 carbons, or two isoprenoid units) myrcene and limonene, and the diterpenoid (with 20 carbons) retinol. An example of an irregular terpene, formed by tail-to-tail linkage, is the tetraterpene  $\beta$ -carotene, which has 40 carbon atoms.

Among natural products, terpenes and terpenoids display a vast chemical diversity, and discovery of novel terpenoids remains an active area of research. Currently two main terpene biosynthetic pathways are known. These are mevalonate pathway (MVA pathway) and methylerythritol pyrophosphate pathway (MEP pathway), which are described in more detail below.

### **1.4.3 Biosynthesis of terpenes**

The biosynthesis of terpenes is explained using two main pathways that generate precursors of terpenes [39]. The mevalonate pathway (MVA pathway) takes place in the cytosol of the cell, and the non-mevalonate pathway (MEP pathway) takes place mainly in chloroplasts of higher plants. Accumulation of experimental results from various  $^{13}\text{C}$  labeling experiments that could not be explained by MVA pathway led scientists to research alternative biosynthetic pathways for terpene precursors. In the early–mid 1990s, the methylerythritol-4-phosphate pathway (MEP) pathway was first proposed. Further research has shown that plant plastids and most bacteria utilize the MEP pathway to accomplish terpene biosynthesis [38]. Both MEP and MVA pathways lead to the synthesis of the two main universal precursors of terpene synthesis, which are isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). It has been discovered that the two pathways do communicate between them to produce certain terpenes that have origins from both pathways despite them taking place in different locations of the cell [44].

#### *Mevalonate (MVA) pathway of terpene biosynthesis*

A series of enzymatic reactions in the mevalonate pathway leads to the formation of isopentenyl pyrophosphate and dimethylallyl diphosphate from three molecules of acetyl CoA (Figure 1.2). From the 5-carbon precursors DMAPP and IPP, other terpenes such as geranyldiphosphate (10

carbons, GPP), farnesyldiphosphate (15 carbons, FPP), and geranylgeranyldiphosphate (20 carbons, GGPP) are produced. The enzymes that catalyze the linking of C5 isoprenoid units are called prenyltransferases. Linear terpenoids consisting of 15 carbons are synthesized by connecting three isoprene groups. This is performed by either *cis*- or *trans*-isopentyl pyrophosphate synthases (IPPS) – *cis* and *trans*-FPP synthase. The stereochemistry specific nature of FPP synthase was established in 1959[45,46,47] and was followed up by the discovery of the first GPP synthase in 1964 [48]. Since then a number of prenyltransferase enzymes have been characterized from various sources ranging from microbes to higher plants and animals [46]. The presence of stereoselective prenyltransferases that catalyze the synthesis of terpenoids of different lengths has resulted in extensive diversity among terpenes. This is even further complicated by the presence of terpene synthases that can cyclize the linear precursors in various ways [49].

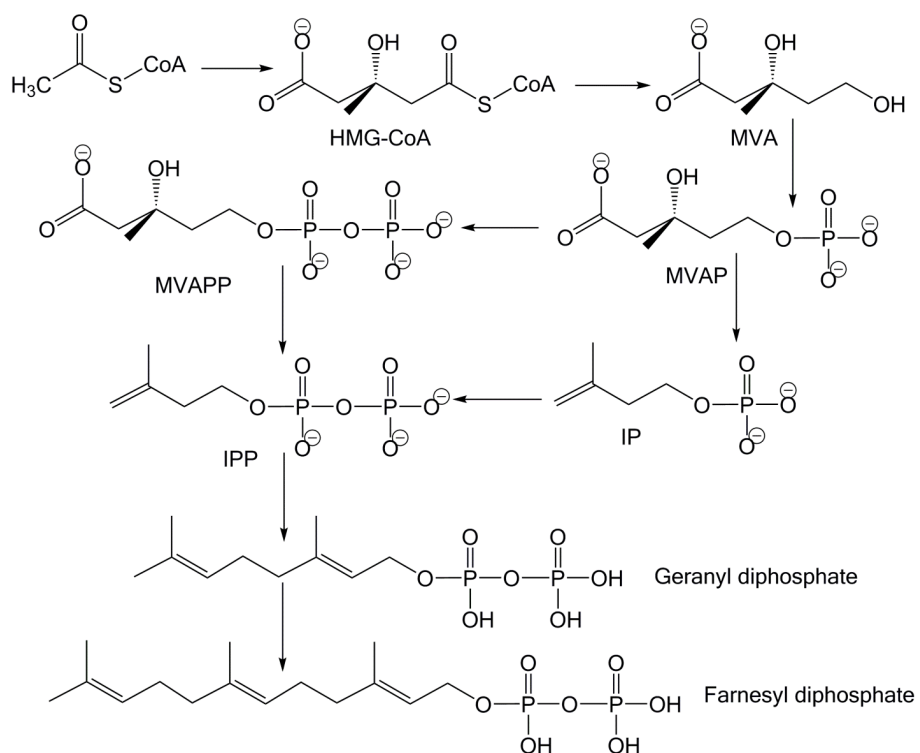


Figure 1.2. Mevalonate pathway of terpene biosynthesis.

### *Non-mevalonate pathway of terpene synthesis - MEP Pathway*

In the non-mevalonate pathway (Figure 1.3) methylerythritol phosphate (MEP) is produced starting with pyruvate condensation with glyceraldehyde diphosphate. MEP acts as the precursor of IPP.

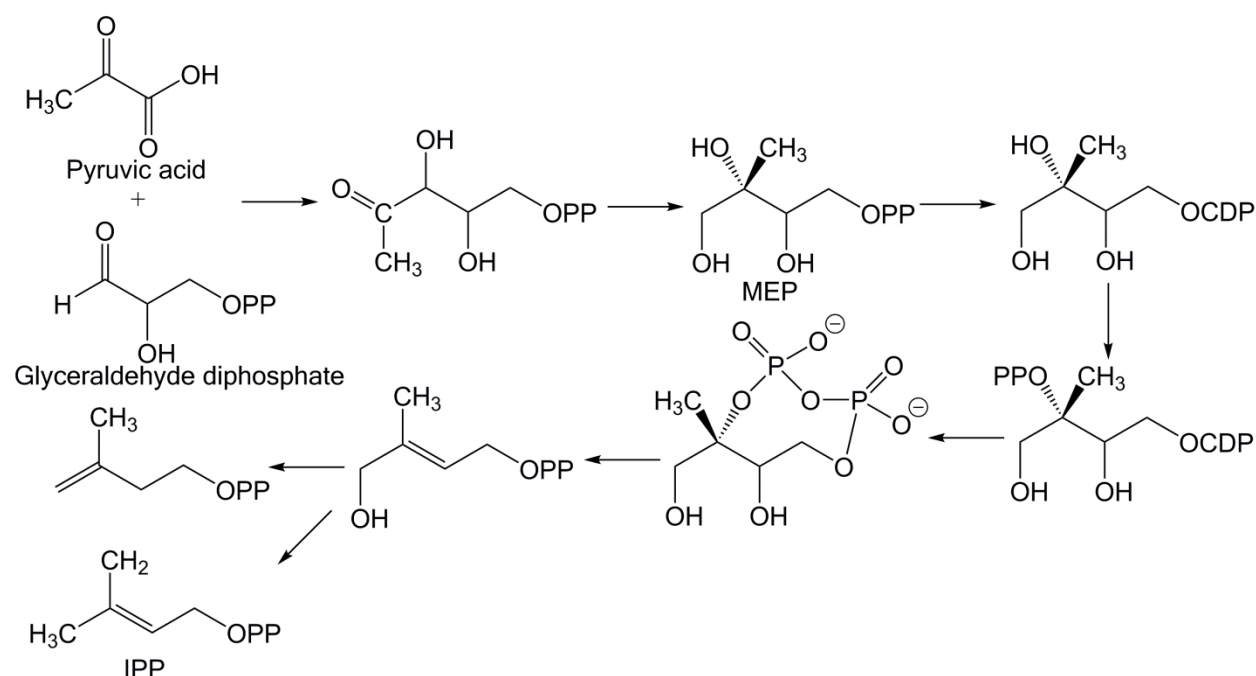


Figure 1.3 Biosynthesis of isopentenyl pyrophosphate starting with pyruvic acid.

Terpene synthase (TPS) enzymes catalyze the formation of hemiterpenes (C<sub>5</sub>), monoterpenes (C<sub>10</sub>), sesquiterpenes (C<sub>15</sub>) or diterpenes (C<sub>20</sub>) from DMAPP, GPP, FPP and GGPP. Despite understanding biological activity, mode of action, structures and biosynthetic pathways for some terpenes, our knowledge of biological roles of these compounds within plants has remained limited. As the functions of terpene biosynthetic genes are discovered, it becomes feasible to engineer plants and microbes that can produce these chemicals, either as enhanced chemical defenses for plants, or as biotechnology products with economic value. To achieve these advances, modern analytical chemistry and molecular biology tools must be integrated to probe the chemical diversity generated by plants and engineered microbes.

The diversity of terpene and terpenoid chemistry as well as the availability of genetic sequence information (both genomic DNA sequences and expressed sequence information) for both wild and domesticated tomato plants makes them an attractive model system for discovery of



enzymes involved in the synthesis and metabolic modification of these molecules. Among these, the wild tomato *Solanum habrochaites* LA1777 possesses a rich diversity of volatile terpenoids [50,51]. Plants of the wild tomato species *Solanum habrochaites* LA1777 exhibit lush production of glandular trichomes, and a rich diversity of terpenoids in its glandular trichomes that has yet to be fully explored [38,42]. *S. habrochaites* produces volatile monoterpenes and sesquiterpenes [52] as well as oxidized sesquiterpenes in the form of sesquiterpenoid acids [53].

#### **1.4.4 Tomato and its terpenoid metabolites**

Classical GC-MS based studies of tomato leaf have identified the presence of monoterpenes and sesquiterpenes in tomato [50,54]. However, these studies have largely been restricted to volatile terpenes since GC-MS is not applicable for the analysis of nonvolatile compounds unless they have been derivatized to increase their volatility. Such is not common practice in terpene analysis. Among the tomato plants studied for volatile terpenes, wild tomato *S. habrochaites* LA1777 has shown to be of particular interest due to its richness in terpenes[50].

Some of the common monoterpenes found in *S. habrochaites* LA1777 include pinene, terpinene, carene, limonene, and phellandrene[50]. However compared to monoterpenes, sesquiterpenes found in this plant display greater diversity. Two classes of volatile sesquiterpenes are known from *S. habrochaites* LA1777. Class I consists of germacrene,  $\alpha$ -humulene and  $\beta$ -caryophyllene. Class II consists of  $\alpha$ -santalene,  $\alpha$ -bergamotene, and  $\beta$ -bergamotene [55]. This classification of sesquiterpenes is based on the location of the terpene synthase genes in the chromosomes that correspond to the synthesis of each of these compound in the plants that commonly accumulate them – *Solanum habrochaites* and *Solanum lycopersicum* [55]. Genes coding for the terpene synthases responsible for the synthesis of Class I terpenes are located on chromosome 6 of both species. Similarly the genes coding for the terpene synthases that

synthesize class II terpenes are found on chromosome 8 of both species[55]. Despite earlier reports of volatile terpenes from *S. habrochaites*, the chemistry of nonvolatile terpenoids has remained largely unexplored. Santalenoic and bergamotenoic acids were the first sesquiterpene acids discovered and characterized from trichomes of *S. habrochaites* LA1777 [53], and they are sufficiently volatile to allow for gas chromatographic analysis. Their presence indicates the metabolic capacity to oxidize a side chain methyl group, and this provides a potential handle for other metabolic modifications, but intermediate alcohol or aldehyde products have yet to be reported. These sesquiterpene acids are notable in that they display insecticidal activity [56]. However, no conjugated terpenoids (*e.g.* sesquiterpene glycosides that are the main theme in this dissertation) have been reported from *S. habrochaites* before.

Sesquiterpene glycosides are found in other plants of the Solanaceae including potato (*Solanum tuberosum*) [57] and tobacco (*Nicotiana tabacum*) [58]. Some sesquiterpene glycosides possess biological activity including hypoglycemic effects[59] , inhibition of tissue factor activity by a nerolidol glycoside from *Eriobotrya japonica* [60] and immunomodulatory activity by sesquiterpenoid glycosides from *Dendrobium nobile* [61,62] that makes them potential pharmacophores and precursors of medicines. Therefore, exploring terpene glycosides from *S. habrochaites* was important for both investigating the chemical diversity and for discovering functions of genes that may influence plant terpenoid composition.

However, the main challenge in such studies lies with the limitations of conventional approaches for metabolite identification [1,63]. Identification of novel metabolites depends on first recognizing their presence, and this is often achieved through coupling of chromatography to mass spectrometry, since many sesquiterpenoids lack characteristic chromophores that signal their presence. A common strategy for metabolite annotation includes performing MS/MS

experiments on compounds of interest and using characteristic fragment ions to identify natural products that belong to a particular class of compounds. Flavonoid identification provides an example of this approach [64,65]. However, in the absence of MS/MS information or when characteristic fragment ions are not present, these methods pose limitations. The absence of a strategy to annotate unknown compounds in plant natural product profiles owing to the limited availability of standards has caused the pharmaceutical industry to move away from natural products-based drug discovery to synthetic library based drug discovery over the last decade [1].

In contrast to other secondary metabolite classes including flavonoids and alkaloids, terpenoids display a vast structural diversity owing to variation in backbone structure, stereochemistry, double bond position, and other sites of functionalization[66]. This diversity has rendered establishing characteristic fragment ions for the direct identification of non-volatile terpenoids such as terpene glycosides using mass spectrometry challenging. Therefore, recognition of the terpene glycosides in a metabolite profile has been based on the absence of fragment ions that represent other classes of molecules such as flavonoids [67]. While GC-MS has enabled the identification of volatile and semivolatile terpenes for many decades it is not amenable for nonvolatile terpenoids. As a result, conjugated terpenoid characterization has been limited mostly to the most abundant members of this class that have been purified.

#### **1.4.5 Characteristic fragment ions for identifying terpenes from plant tissues using Gas Chromatography –Mass Spectrometry**

Volatile terpenes are characterized by the presence of some characteristic ions in their mass spectra. For monoterpenes, these include the molecular ion ( $M^+$ ) at  $m/z$  136, and fragment 121, 93 and 69. For sesquiterpenes, characteristic fragment ions are  $m/z$  204 ( $M^+$ ), 161, 136, 121, 93 and 69 [68]. Furthermore, monoterpenes elute earlier (at lower temperatures in the temperature

program) than sesquiterpenes during GC-MS [68]. These characteristic fragment ions have been used in the detection of terpenes using GC-MS in a number of studies [68,69,70], but many terpenes yield similar mass spectra. As a result, values of chromatographic retention indices are often needed to provide additional evidence of terpene identities. In contrast, identification of glycosylated terpenoids presents a challenge due to their non-volatile nature, and LC-MS is a more appropriate strategy. However, glycosylated terpenoids often fail to yield characteristic fragment ions in MS/MS spectra that enable their conclusive identification.

### **1.5 Discovery and identification of specialized metabolites**

Identification of novel compounds usually requires their isolation from complex matrices. Separating components of interest from complex matrices has been performed for thousands of years. For instance, extraction of precious metals from ore was performed even in ancient civilizations [71]. Classical separation methods depended on differential physical properties including a substance's solubility and boiling point of the compound of interest relative to the matrix to achieve separation. "Distillation" the most commonly used method for separating essential oils from plant leaves even today, has been used since medieval times [72]. Similarly, liquid-liquid extraction and precipitation are used both in research and in industrial applications widely [73]. Regardless of the application, the ultimate objective of separation methods is to obtain a target compound/element in the maximum feasible purity.

For novel compound identifications, classical structural characterization was performed based on elemental analysis using combustion, physical parameters (melting point, boiling point, refractive index, optical rotation) and chemical reactivity. With the advent of ultraviolet absorption spectroscopy, infrared spectroscopy, x-ray crystallography, mass spectrometry, and nuclear magnetic resonance (NMR) spectroscopy, absolute structure assignment became possible

[74,75]. However, successful application of these methods toward the structure elucidation requires a pure compound. This poses a challenge since specialized metabolites in plants are present as low-abundance components of complex mixtures of compounds. Therefore, purification of compounds is of paramount importance for their identification.

Modern analytical chemistry often employs chromatographic methods for the purification of compounds. Earliest attempts of chromatography were performed by Mikhail Tswett [76,77]. Improvements in separation science were achieved with advancements in our understanding of physical chemistry of compound retention on solid particles [78,79]. Since then, a number of advances have been made in the field of chromatography, and the Nobel Prize in Chemistry awarded in 1952 to A.J.P Martin and R.L.M Synge for “inventing partition chromatography” stands out as a landmark [76]. Improvements in separation science were achieved with advances in our understanding of physical chemistry of compound partition onto solid particles [78,79]. This research led to the development of column based liquid chromatographic (LC) methods which was followed by phenomenal advances in instrumentation and column chemistries [80]. These advances led to fast high-performance LC separations, resulting in more efficient purification of natural products. A series of hyphenated methods, involving coupling of LC separations to a variety of detectors further accelerated use of liquid chromatography[81]. These include gas chromatography–mass spectrometry (GC-MS), liquid chromatography – mass spectrometry (LC-MS), liquid chromatography–NMR (LC-NMR), LC-UV, LC-UV-MS and LC-MS-NMR [74,82]. These methods propelled novel compound discovery and characterization by making structural information available at an early stage of analysis without the need to obtain a substantial amount of compound in pure form. Such explosive growth in structural information

has enabled more comprehensive analytical surveys of chemical composition that is described as the “omics” approach.

## **1.6 Mass spectrometry-based metabolomics as a functional genomics tool**

### **1.6.1 Mass Spectrometry**

Molecular mass is one of the single most useful measurements one can make to distinguish one substance from another. Mass spectrometry is an experimental technique that allows for the determination of mass-to-charge ratio ( $m/z$ ) of an ion in the gas phase. Therefore, fundamental requirements that need to be fulfilled for mass spectrometric characterization of a compound are a) it must be taken into the gas phase and b) it must form an ion (positive or negative) in the gas phase.

Some mass spectrometers (often called high resolution mass spectrometers) can measure the  $m/z$  of an ion with high accuracy (within 1 ppm of the theoretical mass). In contrast, some other mass spectrometers (low or unit resolution mass spectrometers) can measure  $m/z$  only with unit accuracy ( $\pm 1$  Da). The mass reported by high-resolution mass spectrometers is termed “accurate mass” since it is closer to the true mass of the compound measured. The mass reported by a unit resolution instrument is termed “nominal mass” of a given compound, and is a value usually rounded to an integer value. For instance, a compound might be detected at accurate mass of  $m/z$  649.3035 in negative ion mode using a high-resolution mass spectrometer. The decimal component of this reported mass (*i.e.* 0.3036) is termed “mass defect” of a compound. The same compound will be reported as the nominal mass  $m/z$  649 by a unit resolution mass spectrometer.

Some common high resolution mass spectrometers use time-of-flight (ToF) and Orbitrap technologies for the detection of ions. Unit resolution mass spectrometers are commonly based on the use of one or more quadrupoles for ion detection.

### **1.6.2 Collision induced dissociation (CID) and MS/MS**

International union of pure and applied chemistry's (IUPAC) has defined Collision induced dissociation as “an ion/neutral species interaction wherein the projectile ion is dissociated as a result of interaction with a target neutral species. This is brought about by conversion of part of the translational energy of the ion to internal energy in the ion during collision”. Therefore, the three key steps involved in CID are, interaction between a charged ion and a neutral species, energy transfer and dissociation. In modern mass spectrometers nitrogen gas is used as the neutral species and the ions are brought into collisions with the gas. The importance of CID is that it allows for the fragmentation of ions of a compound of interest and thereby yields further structural information about that compound. A common experiment used in this context is the MS/MS experiment.

MS/MS experiments are mass spectrometric experiments where all ions except for the ion of interest are filtered to isolate the ion of interest followed by colliding it with a neutral species (such as nitrogen) to generate fragment ions. Commonly, these experiments are performed using mass spectrometers equipped with one or more quadrupoles. Quadrupoles are used for filtering the ion of interest from others.

### **1.6.3 Multiplexed CID mass spectrometry**

While MS/MS experiments are useful for structural characterization of compounds usually it is applicable to only a few compounds of interest at a time. This poses a bottleneck for discovery

metabolomics where obtaining mass spectral information on both parent ions and their fragment ions for as many compounds as possible is gives the best possibility of discovering novel compounds. Multiplexed CID mass spectrometry allows for further advancement by multiplexing MS/MS experiments. Multiplexed CID MS is a method that enables the accurate mass information of both parent ions and fragment ions of a compound to be accumulated simultaneously using time-of-flight technology without isolating the ion of interest. Since no ion isolation is performed, this method allows for fragmentation information of all parent ions that reaches the mass spectrometer at any time to be obtained. One obvious complication here is that multiple compounds that enter the mass spectrometer will produce an assortment of fragment ions making spectral interpretation complicated. However, since this experiment is performed using a Time-of-flight instrument the accurate mass of both parent ions and fragment ions will be obtained, simplifying the spectral interpretation process. Furthermore, to simplify the data interpretation process commonly compounds in complicated mixtures (such as plant extracts) are separated using a chromatographic method (Ex: liquid chromatography) to minimize the number of compounds that reach the mass spectrometer at one time to make spectral interpretation easier.

Multiplexed CID MS functions by elevating the energy applied to the ions that enter the mass spectrometer during their movement from ion source to the ion detector. Furthermore, the pressure of the region of the mass spectrometer where the multiplexed CID MS fragmentation occurs is approximately  $2 \times 10^{-3}$  mbar which is about three orders of magnitude greater than pressure in the ion detector. This allows for enough collisions between ions and collision gas to take place enabling some of the parent ions to fragment while another proportion of them remain unfragmented. Thus, both fragment ion mass to charge ratio and the daughter ion mass to charge ratios can be measured simultaneously. Recent use of LC-multiplexed CID mass spectrometry



for rapid profiling of metabolites from wild tomato relative *Solanum habrochaites* accessions has been well documented [83,84].

#### **1.6.4 Mass spectrometry for discovery of novel metabolites**

*LC-MS opened gates to a world of compounds not observed before:* Gas chromatography – mass spectrometry (GC-MS) has emerged as the most powerful approach to analyze volatile components from plant extracts [85]. However, GC-MS is restricted to non-polar to low polar compounds that are volatile or compounds that can be derivatized to enhance their volatility. Therefore the number of compounds observed using GC-MS from plant extracts is restricted to a few hundred [85]. However, using liquid chromatography-mass spectrometry (LC-MS) it has been possible to extend detection and determination of molecular masses to non-volatile compounds that are more polar and have not been reported before [86]. This unraveled insights into a richer plant metabolome that were not observed using GC-MS analyses. Using LC-MS, both qualitative and quantitative assessments of the compounds in the metabolome can be made. However, owing to the diverse structures and dynamic range of metabolites found in plants, characterization of a given metabolome has required the use of a diverse set of tools to achieve comprehensive coverage of the metabolites [14,87].

The vast majority of plant metabolites are unknown. Thanks to the enhancements of instrumentation technologies the detection limits of mass spectrometers have become lower and as a result many more compounds that were not observed in plants before are becoming apparent (e.g. Medicinal plants consortium database - <http://medicinalplantgenomics.msu.edu/index.shtml> ). This enables greater amount of novel data to be accumulated, which in turn challenges the scientists to identify them. Therefore, for annotation of these metabolites that are not found in databases, a different approach is necessary.

*Metabolite identification is the main challenge ahead:* Identification of metabolites observed in metabolite profiles is the main bottleneck in today's metabolomics [88,89]. To address this challenge, most laboratories use metabolite databases including MassBank, HMDB, LipidMaps, METLIN and NIST databases [89]. The metabolite profiling data that are collected from a given sample are matched with these databases based on accurate mass to find matching compounds. Standards of metabolite annotation are used to enhance the accuracy and the quality of database available to make compound annotations based on them more accurate [90]. A second approach is to use authentic standards to verify the identity of compounds [91]. This involves obtaining the authentic standards and comparing their reported mass, fragment ions formed by collision induced dissociation (or other fragmentation method) and retention time with those found in unknown samples to find matching compounds. A commonly employed third approach is to apply mass defect (described under section 1.5.1) filters that would identify changes in accurate mass to assign compounds as derivatives of known compounds [92]. However, these approaches are useful only when searching for compounds that have already been annotated elsewhere and have been entered into the database or in the presence of authentic standards making novel strategies for metabolite identification a fundamental requirement.

In early days of metabolomics it was proposed that unknown metabolite annotation could be achieved based on accurate mass and database searches [70], perhaps under the naïve assumption that unknowns must bear similarity to known compounds. However, soon this perception was changed and it was apparent that accurate mass alone was not adequate for compound identification [93,94]. To enable accurate metabolite identification based on mass spectrometry, a combination of accurate mass information, isotopic abundance information and MS/MS data for the metabolite are used [89,93]. To obtain this information usually, first a mass spectrum is

obtained using a high-resolution instrument such as a ToF mass spectrometer that will allow the establishment of the accurate mass followed by performing the MS/MS experiments on selected masses of interest.

Despite all the information provided by such mass spectrometry experiments, these analyses alone do not allow the structure of an unknown metabolite to be confirmed. To achieve *de novo* structure elucidation, the compound of interest need to be either synthesized or purified from its matrix, followed by performing 1D and 2D NMR based structure elucidation [12].

### **1.7 Summary of research**

Among the secondary metabolites, glycosylated forms of terpenoids makes up a large proportion. These are hybrid molecules that consist of a terpene group synthesized from isoprenoid precursors and a glycoside component made up of various sugar moieties. Glycosylation is a common form of conjugation of metabolites that enhances their water solubility, which in turn, facilitates their intra- and inter-cellular movements within the plant. In addition, glycosylation and other conjugations stabilize metabolites by creating forms more readily sequestered in specific organelles (*e.g.* vacuoles) and allow them to be stored for longer periods of time because they are out of contact with enzymes that may transform them further. Steroidal glycosides are a class of compounds that is well characterized. In contrast sesquiterpene glycosides are less well known.

Sesquiterpene glycosides consists of a terpenoid core synthesized by three isoprene groups and is conjugated to one or more carbohydrate moieties. Sesquiterpenes could be acyclic or consist of ring systems

Chapter 2, 3 and 4 of this dissertation describe the strategies for discovery, structure elucidation and exploration of chemical diversity of various sesquiterpene glycosides from various accessions of a wild tomato relative *Solanum habrochaites*. The genus *Solanum* is part of the plant family *Solanaceae* that includes a number of agriculturally important crops including tobacco, eggplant, pepper and petunia. This plant was selected for this research due to the availability of the genomic DNA sequence for tomato and EST (expressed sequence tag) sequences of wild tomato relatives. Also a number of tomato introgression lines developed are available [95]. Both these resources enable the identification of genes involved in the synthesis of compounds of interest. In addition, a rich history of characterization of compounds and genes from this family exists that is useful for this research. Thirdly, tomato is a plant with a fairly short lifecycle and a plant that is not too large in physical dimensions. This makes obtaining offspring easy allowing for easier manipulation of the plant. All these led to selection of this as a model plant for the characterization of metabolome in this project. Furthermore, this research has led to the development of a novel method for rapid classification of metabolites into various metabolite classes. Chapter 5 of this dissertation describes the application of this method for the analysis and discovery of compounds from metabolome of *Hoodia gordonii* and *Rosmarinus officinalis*, two plants studied in the development of the Medicinal Plants Consortium database. *H. gordonii* is a plant that has been extensively studied for the presence of dietary supplements. Similarly, *R. officinalis* is a medicinal plant that is used to prepare dietary supplements and is used as a spice in most culinary practices.

## **REFERENCES**

## REFERENCES

1. H.F. Ji, X.J. Li, H.Y. Zhang, Natural products and drug discovery. Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia?, *EMBO Rep* 10 (2009) 194-200.
2. Asia-Development-Bank, Food security and poverty in asia and the pacific. Key challenges and policy issues., (2012).
3. D. Graham-Rowe, Agriculture: Beyond food versus fuel, *Nature* 474 (2011) S6-8.
4. R.L. Last, A.D. Jones, Y. Shachar-Hill, Towards the plant metabolome and beyond, *Nat Rev Mol Cell Biol* 8 (2007) 167-174.
5. S.A. Rensing, D. Lang, A.D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, P.F. Perroud, E.A. Lindquist, Y. Kamisugi, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, T. Shin-I, Y. Kuroki, A. Toyoda, Y. Suzuki, S. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S. Aoki, N. Ashton, W.B. Barbazuk, E. Barker, J.L. Bennetzen, R. Blankenship, S.H. Cho, S.K. Dutcher, M. Estelle, J.A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K.A. Hicks, J. Hughes, M. Lohr, K. Mayer, A. Melkozernov, T. Murata, D.R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauts, P.J. Rushton, A. Sanderfoot, G. Schween, S.H. Shiu, K. Stueber, F.L. Theodoulou, H. Tu, Y. Van de Peer, P.J. Verrier, E. Waters, A. Wood, L.X. Yang, D. Cove, A.C. Cuming, M. Hasebe, S. Lucas, B.D. Mishler, R. Reski, I.V. Grigoriev, R.S. Quatrano, J.L. Boore, The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants, *Science* 319 (2008) 64-69.
6. M. Hurles, Gene duplication: The genomic trade in spare parts, *Plos Biology* 2 (2004) 900-904.
7. J.D. Keasling, Building with biology, *Nature* 492 (2012) 188-188.
8. D.I. Ellis, R. Goodacre, Metabolomics-assisted synthetic biology, *Curr Opin Biotechnol* 23 (2012) 22-28.
9. Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nature Reviews Genetics* 10 (2009) 57-63.
10. S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nature Biotechnology* 17 (1999) 994-999.
11. R. Aebersold, M. Mann, Mass spectrometry-based proteomics, *Nature* 422 (2003) 198-207.
12. O. Fiehn, Metabolomics - the link between genotypes and phenotypes, *Plant Molecular Biology* 48 (2002) 155-171.

13. D.G. Wang, J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, E.S. Lander, Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome, *Science* 280 (1998) 1077-1082.
14. L.W. Sumner, P. Mendes, R.A. Dixon, Plant metabolomics: large-scale phytochemistry in the functional genomics era, *Phytochemistry* 62 (2003) 817-836.
15. E. Pichersky, D.R. Gang, Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective, *Trends in Plant Science* 5 (2000) 439-445.
16. R. McDaniel, S. Ebertkhosla, D.A. Hopwood, C. Khosla, Engineered biosynthesis of novel polyketides, *Science* 262 (1993) 1546-1550.
17. J. Bohlmann, G. Meyer-Gauen, R. Croteau, Plant terpenoid synthases: Molecular biology and phylogenetic analysis, *Proceedings of the National Academy of Sciences of the United States of America* 95 (1998) 4126-4133.
18. C.M.M. Gachon, M. Langlois-Meurinne, P. Saindrenan, Plant secondary metabolism glycosyltransferases: the emerging functional analysis, *Trends in Plant Science* 10 (2005) 542-549.
19. M. Kujawa, H. Ebner, C. Leitner, B.M. Hallberg, M. Prongjit, J. Sucharitakul, R. Ludwig, U. Rudsander, C. Peterbauer, P. Chaiyen, D. Haltrich, C. Divne, Structural basis for substrate binding and regioselective oxidation of monosaccharides at C3 by pyranose 2-oxidase, *Journal of Biological Chemistry* 281 (2006) 35104-35115.
20. B. Hu, J. Simon, H. Rennenberg, Drought and air warming affect the species-specific levels of stress-related foliar metabolites of three oak species on acidic and calcareous soil, *Tree Physiology* 33 (2013) 489-504.
21. M.A.M. Khan, C. Ulrichs, I. Mewis, Water stress alters aphid-induced glucosinolate response in *Brassica oleracea* var. *italica* differently, *Chemoecology* 21 (2011) 235-242.
22. P.W. Pare, J.H. Tumlinson, Plant volatiles as a defense against insect herbivores, *Plant Physiology* 121 (1999) 325-331.
23. T. Tohge, A.R. Fernie, Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function, *Nature Protocols* 5 (2010) 1210-1227.
24. T. Tohge, A.R. Fernie, Annotation of Plant Gene Function via Combined Genomics, Metabolomics and Informatics, (2012) e3487.

25. J.A. Gatehouse, Biotechnological prospects for engineering insect-resistant plants, *Plant Physiology* 146 (2008) 881-887.
26. Z.G. Li, L.H. Yao, Y.W. Yang, A.D. Li, Transgenic approach to improve quality traits of melon fruit, *Scientia Horticulturae* 108 (2006) 268-277.
27. C. Prakash, C.L. Shaffer, A. Nedderman, Analytical strategies for identifying drug metabolites, *Mass Spectrom Rev* 26 (2007) 340-369.
28. A. Craney, S. Ahmed, J. Nodwell, Towards a new science of secondary metabolism, *Journal of Antibiotics* 66 (2013) 387-400.
29. A.R. Fernie, R.N. Trethewey, A.J. Krotzky, L. Willmitzer, Metabolite profiling: from diagnostics to systems biology, *Nat Rev Mol Cell Biol* 5 (2004) 763-769.
30. J. Clardy, C. Walsh, Lessons from natural molecules, *Nature* 432 (2004) 829-837.
31. G.J. Wagner, Secreting glandular trichomes - more than just hairs, *Plant Physiology* 96 (1991) 675-679.
32. J.-H. Kang, G. Liu, F. Shi, A.D. Jones, R.M. Beaudry, G.A. Howe, The Tomato odorless-2 Mutant Is Defective in Trichome-Based Production of Diverse Specialized Metabolites and Broad-Spectrum Resistance to Insect Herbivores, *Plant Physiology* 154 (2010) 262-272.
33. A. Weinhold, I.T. Baldwin, Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation, *Proceedings of the National Academy of Sciences* (2011).
34. D.A. Levin, The Role of Trichomes in Plant Defense., *The quarterly review of biology* 48 (1973) 3-15.
35. A. Fahn, Structure and function of secretory cells, *Advances in Botanical Research Incorporating Advances in Plant Pathology*, Vol 31 2000 31 (2000) 37-75.
36. A.L. Schillmiller, R.L. Last, E. Pichersky, Harnessing plant trichome biochemistry for the production of useful compounds, *Plant Journal* 54 (2008) 702-711.
37. R.W. Gibson, J.A. Pickett, Wild tomato repels aphids by release of aphid alarm pheromone., *Nature* 302 (1983) 608-609.
38. W.M. Tingey, S.A. Mehlenbacher, J.E. Laubengayer, Occurrence of glandular trichomes in wild *Solanum* species, *American Potato Journal* 58 (1981) 81-83.
39. L. Luckwill, The genus *Lycopersicon*: historical, biological, and taxonomic survey of the wild and cultivated tomatoes, Aberdeen University Press, Aberdeen, Scotland, 1943.



40. J.H. Kang, F. Shi, A.D. Jones, M.D. Marks, G.A. Howe, Distortion of trichome morphology by the hairless mutation of tomato affects leaf surface chemistry, *Journal of Experimental Botany* 61 (2010) 1053-1064.
41. A.L. Schillmiller, I. Schauvinhold, M. Larson, R. Xu, A.L. Charbonneau, A. Schmidt, C. Wilkerson, R.L. Last, E. Pichersky, Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate, *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009) 10865-10870.
42. E.T. McDowell, J. Kapteyn, A. Schmidt, C. Li, J.H. Kang, A. Descour, F. Shi, M. Larson, A. Schillmiller, L.L. An, A.D. Jones, E. Pichersky, C.A. Soderlund, D.R. Gang, Comparative Functional Genomic Analysis of Solanum Glandular Trichome Types, *Plant Physiology* 155 (2011) 524-539.
43. L. Ruzicka, The isoprene rule and the biogenesis of terpenic compounds, *Experientia* 9 (1953) 357-367.
44. C. Schuhr, T. Radykewicz, S. Sagner, C. Latzel, M. Zenk, D. Arigoni, A. Bacher, F. Rohdich, W. Eisenreich, Quantitative assessment of crosstalk between the two isoprenoid biosynthesis pathways in plants by NMR spectroscopy, *Phytochemistry Reviews* 2 (2003) 3-16.
45. G. Popjak, Cornfort.Jw, Substrate stereochemistry in squalene biosynthesis , *Biochemical Journal* 101 (1966) 553-&.
46. K.C. Wang, S. Ohnuma, Isoprenyl diphosphate synthases, *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids* 1529 (2000) 33-48.
47. F. Lynen, B.W. Agranoff, H. Eggerer, U. Henning, E.M. Möslin,  $\gamma,\gamma$ -Dimethyl-allyl-pyrophosphat und Geranyl-pyrophosphat, biologische Vorstufen des Squalens Zur Biosynthese der Terpene, VI1), *Angewandte Chemie* 71 (1959) 657-663.
48. A.A. Kandutsch, E. Levin, K. Bloch, H. Paulus, Purification of geranylgeranyl pyrophosphate synthase from *Micrococcus lysodeiticus*, *Journal of Biological Chemistry* 239 (1964) 2507-&.
49. B. Greenhagen, J. Chappell, Molecular scaffolds for chemical wizardry: Learning nature's rules for terpene cyclases, *Proceedings of the National Academy of Sciences of the United States of America* 98 (2001) 13479-13481.
50. E. Gonzales-Vigil, D.E. Hufnagel, J. Kim, R.L. Last, C.S. Barry, Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative Solanum habrochaites, *Plant Journal* 71 (2012) 921-935.

51. Y. Matsuba, T.T.H. Nguyen, K. Wiegert, V. Falara, E. Gonzales-Vigil, B. Leong, P. Schafer, D. Kudrna, R.A. Wing, A.M. Bolger, B. Usadel, A. Tissier, A.R. Fernie, C.S. Barry, E. Pichersky, Evolution of a Complex Locus for Terpene Biosynthesis in *Solanum*, *Plant Cell* 25 (2013) 2022-2036.
52. L. Lundgren, G. Norelius, G. Stenhagen, Leaf volatiles from some wild tomato species, *Nordic Journal of Botany* 5 (1985) 315-320.
53. R.M. Coates, J.F. Denissen, J.A. Juvik, B.A. Babka, Identification of  $\alpha$ -santalenoic and endo- $\beta$ -bergamotenoic acids as moth oviposition stimulants from wild tomato leaves, *Journal of Organic Chemistry* 53 (1988) 2186-2192.
54. M.A. Farag, P.W. Pare, C-6-green leaf volatiles trigger local and systemic VOC emissions in tomato, *Phytochemistry* 61 (2002) 545-554.
55. R.S. van der Hoeven, A.J. Monforte, D. Breeden, S.D. Tanksley, J.C. Steffens, Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*, *Plant Cell* 12 (2000) 2283-2294.
56. J.A. Juvik, M.J. Berlinger, T. Ben-David, J. Rudich, Resistance among accessions of the generally *lycopersicon* and *solanum* to four of the main insect pests of tomato in Israel, *Phytoparasitica* 10 (1982) 145-156.
57. H. Tazaki, N. Ohta, K. Nabeta, H. Okuyama, M. Okumura, Structure of sesquiterpene glucosides from potato leaves, *Phytochemistry* 34 (1993) 1067-1070.
58. X. Feng, J.S. Wang, J. Luo, L.Y. Kong, Two new sesquiterpene glucosides from the leaves of *Nicotiana tabacum*, *Journal of Asian Natural Products Research* 11 (2009) 675-680.
59. J. Chen, W.L. Li, J.L. Wu, B.R. Ren, H.Q. Zhang, Hypoglycemic effects of a sesquiterpene glycoside isolated from leaves of loquat (*Eriobotrya japonica* (Thunb.) Lindl.), *Phytomedicine* 15 (2008) 98-102.
60. M.H. Lee, K.S. Yeon, N.H. Yong, Tissue factor inhibitory sesquiterpene glycoside from *Eriobotrya japonica*, *Archives of Pharmacal Research* 27 (2004) 619-623.
61. Q. Ye, G. Qin, W. Zhao, Immunomodulatory sesquiterpene glycosides from *Dendrobium nobile*, *Phytochemistry* 61 (2002) 885-890.
62. W. Zhao, Q. Ye, X. Tan, H. Jiang, X. Li, K. Chen, A.D. Kinghorn, Three New Sesquiterpene Glycosides from *Dendrobium nobile* with Immunomodulatory Activity, *Journal of Natural Products* 64 (2001) 1196-1200.
63. B. Zhou, J.F. Xiao, L. Tuli, H.W. Ransom, LC-MS-based metabolomics, *Molecular Biosystems* 8 (2012) 470-481.

64. Y.L. Ma, Q.M. Li, H. VandenHeuvel, M. Claeys, Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry, *Rapid Communications in Mass Spectrometry* 11 (1997) 1357-1364.
65. E.A.P. Ekanayaka, C. Li, A.D. Jones, Sesquiterpenoid glycosides from glandular trichomes of the wild tomato relative *Solanum habrochaites*, *Phytochemistry*.
66. D.O. Kennedy, E.L. Wightman, Herbal Extracts and Phytochemicals: Plant Secondary Metabolites and the Enhancement of Human Brain Function, *Advances in Nutrition* 2 (2011) 32-50.
67. J.L. Ward, J.M. Baker, A.M. Llewellyn, N.D. Hawkins, M.H. Beale, Metabolomic analysis of *Arabidopsis* reveals hemiterpenoid glycosides as products of a nitrate ion-regulated, carbon flux overflow, *Proc Natl Acad Sci U S A* 108 (2011) 10762-10767.
68. G. Wang, L. Tian, N. Aziz, P. Broun, X. Dai, J. He, A. King, P.X. Zhao, R.A. Dixon, Terpene Biosynthesis in Glandular Trichomes of Hop, *Plant Physiology* 148 (2008) 1254-1266.
69. S.M. Colby, J. Crock, B. Dowdle-Rizzo, P.G. Lemaux, R. Croteau, Germacrene C synthase from *Lycopersicon esculentum* cv. VFNT Cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase, *Proceedings of the National Academy of Sciences of the United States of America* 95 (1998) 2216-2221.
70. O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry, *Analytical Chemistry* 72 (2000) 3573-3580.
71. D. Killick, King Croesus' gold: Excavations at Sardis and the history of gold refining, *American Journal of Archaeology* 106 (2002) 113-114.
72. V.D. Zheljazkov, T. Astatkie, B. O'Brocki, E. Jeliaskova, Essential Oil Composition and Yield of Anise from Different Distillation Times, *Hortscience* 48 (2013) 1393-1396.
73. M. Giuliatti, M.M. Seckler, S. Derenzo, M.I. Re, E. Cekinski, Industrial crystallization and precipitation from solutions: State of the technique, *Brazilian Journal of Chemical Engineering* 18 (2001) 423-440.
74. D.A. Dias, S. Urban, U. Roessner, A Historical Overview of Natural Products in Drug Discovery, *Metabolites* 2 (2012) 303-336.
75. G. Pataki, J. Borko, Some Recent Advances in Thin Layer Chromatography I. New Applications in Amino Acid Peptide und Nucleic Acid Chemistry, *Chromatographia* 1 (1968) 406-417.

76. L.S. Ettre, K.I. Sakodynskii, Tswett, M.S and the discovery of chromatography 1. Early work (1899-1903), *Chromatographia* 35 (1993) 223-231.
77. L.S. Ettre, K.I. Sakodynskii, Tswett, M.S and the discovery of chromatography 2. Completion of the developemnt of Chromatography (1903-1910), *Chromatographia* 35 (1993) 329-338.
78. E. Heftmann, *Chromatography, Analytical Chemistry* 36 (1964) 14R - 35R.
79. K.R. Williams, Colored Bands: History of Chromatography, *Journal of Chemical Education* 79 (2002) 922-923.
80. T.L. Chester, Recent Developments in High-Performance Liquid Chromatography Stationary Phases, *Analytical Chemistry* 85 (2013) 579-589.
81. G. Rajput, H. Patel, J. Patel, K. Patel, M. Patel, Introduction to hyphenated techniques and their applications in pharmacy.
82. E.H. Liu, L.W. Qi, J. Cao, P. Li, C.Y. Li, Y.B. Peng, Advances of modern chromatographic and electrophoretic methods in separation and analysis of flavonoids, *Molecules* 13 (2008) 2521-2544.
83. F. Shi, Profiling of specialized metabolites in glandualr trichomes of the genus *Solanum* using liquid chromatography and mass spectrometry., Chemistry, Michigan State University, East Lansing, 2009.
84. C. Li, Mass spectrometric profiling and localization of metabolites in biological samples., Chemistry, Michigan State University, East Lansing, 2011.
85. O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, Metabolite profiling for plant functional genomics, *Nat Biotechnol* 18 (2000) 1157-1161.
86. R.C. De Vos, S. Moco, A. Lommen, J.J. Keurentjes, R.J. Bino, R.D. Hall, Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry, *Nat Protoc* 2 (2007) 778-791.
87. K. Saito, F. Matsuda, Metabolomics for Functional Genomics, Systems Biology, and Biotechnology, in: S. Merchant, W.R. Briggs, D. Ort (Eds.), *Annual Review of Plant Biology*, Vol 61, 2010, pp. 463-489.
88. R.D. Hall, Plant metabolomics in a nutshell: potential and future challenges, *Annual Plant Reviews* 43 (2011) 1–24.

89. R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G.J. Patti, G. Siuzdak, An accelerated workflow for untargeted metabolomics using the METLIN database, *Nature Biotechnology* 30 (2012) 826–828.
90. F. Matsuda, Y. Shinbo, A. Oikawa, M.Y. Hirai, O. Fiehn, S. Kanaya, K. Saito, Assessment of Metabolome Annotation Quality: A Method for Evaluating the False Discovery Rate of Elemental Composition Searches, *Plos One* 4 (2009).
91. A.D. Hegeman, Plant metabolomics-meeting the analytical challenges of comprehensive metabolite analysis, *Briefings in Functional Genomics* 9 (2010) 139-148.
92. H.Y. Zhang, D.L. Zhang, K. Ray, M.S. Zhu, Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry, *Journal of Mass Spectrometry* 44 (2009) 999-1016.
93. T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm, *Bmc Bioinformatics* 7 (2006).
94. T. Kind, O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, *Bmc Bioinformatics* 8 (2007) 105.
95. D. Zamir, Improving plant breeding with exotic genetic libraries, *Nat Rev Genet* 2 (2001) 983-989.

## **Chapter 2: Strategies for rapid identification of sesquiterpene glycosides from complex matrices using relative mass defect filtering.**

The research findings described in this chapter have, in part, been published in the journal article: Sesquiterpenoid glycosides from glandular trichomes of the wild tomato relative *Solanum habrochaites*. *Phytochemistry* 98: 223-231

### **2.1 Introduction**

Terpenes represent the largest group of plant natural products, with over 55,000 known compound structures [1]. Their diversity stems from an assortment of isoprenoid metabolites that vary in length and configuration, and these undergo a wide variety of metabolic cyclizations and secondary transformations (*e.g.* hydroxylation, reduction, glycosylation, and acylation) [2]. However, isoprene remains the fundamental structural unit of all terpenes as discussed in Chapter 1. As a result, despite common secondary transformations, terpenoids almost always consist of a hydrocarbon core, but recognition of metabolites derived from terpenoid biosynthetic pathways has remained inefficient. Despite the vast diversity of terpenoids, a systematic approach that can distinguish metabolites that have a terpene hydrocarbon core from the rest of the molecules in a complex matrix can allow for the distinctive identification of terpenes among others. This information can guide research into the roles of specific enzymes in terpene biosynthesis and degradation. To exploit this characteristic, this Chapter presents a method for classification of compounds based on the concept of relative mass defect (described below) that is expected to accelerate identification of terpene glycosides from complicated matrices.

## **2.2 Applying accurate mass and relative mass defect filtering for exploring plant metabolomes for the identification novel terpene glycosides**

Mass spectrometry plays important roles as a structure elucidation tool in modern metabolomics [3]. The growing use of medium- to high-resolution mass spectrometers has provided greater mass measurement accuracy, often with low part-per-million (ppm) mass measurement errors. For metabolites of low molecular mass, such measurements often provide sufficient information to assign molecular formulas without ambiguity, but the number of possible formulas within a given error tolerance grows rapidly as a function of mass. For metabolites of higher (> 500 Da) molecular masses, multiple elemental formulas may be consistent with measured masses within measurement errors of about 1 ppm. Moreover, assignments of molecular formulas often fail to yield reliable assignments of metabolites to specific biosynthetic origins. This is because a given elemental formula can correspond to a number of different chemical structures. For instance, a “simple search” in the ChemSpider database ([www.chemspider.com](http://www.chemspider.com)) for the elemental formula of glucose  $C_6H_{12}O_6$  results in 256 chemical structures with identical monoisotopic mass!

Among the methods proposed for associating compounds with specific metabolite classes using mass spectrometry, relative mass defect (RMD) is a property that can be calculated for both molecular and fragment ions in mass spectra. RMD facilitates classification of compounds to various biosynthetic classes, since RMD is correlated with fractional hydrogen content [4].

Relative mass defect is ratio of the mass defect to accurate mass of a given compound, expressed in parts per million. Relative mass defect (RMD) is calculated in ppm, with  $RMD = (\text{mass defect} / \text{measured monoisotopic mass}) \times 10^6$ . Similar exploitation of absolute mass defect filtering has been helpful for identifying peptides [5]. The main drawback of using absolute mass defect filtering instead of relative mass defect filtering is that the former does not scale well with molecular mass, whereas normalizing mass defect to ion mass yields values that better reflect

biosynthetic origins. For example, monoterpene ( $C_{10}H_{16}$ ), diterpene ( $C_{20}H_{32}$ ), and triterpene ( $C_{30}H_{48}$ ) hydrocarbons would all have different absolute mass defects, reflecting differences in the number of hydrogen atoms in each molecule, but the relative mass defect, obtained by normalizing to molecular mass, would be the same, reflecting their common biosynthetic origins. For example, the monoisotopic mass of a monoterpene of formula  $C_{10}H_{16}$  is 136.1252 Da, a diterpene of formula  $C_{20}H_{32}$  has a monoisotopic mass of 272.2504 Da, and triterpene  $C_{30}H_{48}$  with monoisotopic mass of 408.3756. All three compounds have RMD of 920 ppm. These calculations demonstrate that regardless of the size of the terpenoid, hydrocarbons with constant proportions of hydrogen to carbon yield the same RMD value, even though their absolute mass defects are quite different.

A positive absolute mass defect usually reflects a large number of hydrogen atoms because the atomic mass of hydrogen (1.00783 Da) is slightly greater than the rounded-off integer value (by 7.83 mDa), and carbon (12.00000 Da) does not contribute to mass defect because its monoisotopic mass is exactly an integer value. Oxygen (15.99491 Da) has a small negative mass defect (-5.09 mDa) and nitrogen (14.00307 Da) has a positive defect (+3.07 mDa). Since most metabolites have much larger numbers of hydrogen atoms than oxygen or nitrogen, the RMD for most metabolites is largely governed by the fraction of the mass that is made up of hydrogen atoms. Compounds rich in hydrogen such as alkanes and fatty acids often display RMD greater than those of terpenoids because the hydrogen content of terpenoids is lower. For example, RMD values of stearic acid ( $C_{18}H_{36}O_2$ ; 12.76 %H; 955 ppm) and palmitic acid ( $C_{16}H_{32}O_2$ ; 12.58 %H; 937 ppm) are slightly higher than isoprene ( $C_5H_8$  and its oligomers  $(C_5H_8)_n$ , all of which have 11.84 %H and RMD = 919 ppm. However, it is common that terpenoids undergo metabolic oxidation resulting in a decrease of RMD value. For example, the RMD of farnesol ( $C_{15}H_{26}O$ ;



11.79 %H) equals 893 ppm. However, it is common that terpenoids undergo extensive metabolic oxidation or conjugation by oxygen-rich groups (*e.g.* carbohydrates, phosphates, and oxygen-rich esters), and the RMD values decrease as a result of increased fractional content of oxygen and decreased fractional hydrogen content. This difference in RMD reflects the extent of oxidation of the compound and enables a rapid assignment of ions to broad structural classes based on hydrogen content. Because conjugation decreases the RMD value of an intact terpenoid, this one value alone does not allow conclusive assignment of a compound as a terpenoid based on RMD of a molecular ion. For example, a common tetraacylsucrose found in glandular trichomes of tomato has an elemental formula of  $C_{29}H_{48}O_{15}$ , which gives an RMD value of 470 ppm. Further support annotation of metabolites as terpenoids, the RMDs of fragment ions in mass spectra that are generated using collision-induced dissociation need to be considered. For most conjugated terpenoids, the RMD of fragment ions remaining after the loss of conjugate groups is greater than the RMD of the molecular ion because the conjugate groups have lower fractional hydrogen content than the intact molecules.

For molecules that are rich in reduced carbon such as isoprene, which is the building block of terpenes, the RMD is 920 ppm, reflecting the 11.8 wt % hydrogen content of  $C_5H_8$ . This value does not change with oligomerization, as a sesquiterpene such as  $\alpha$ -santalene ( $C_{15}H_{24}$ ) also has 11.8% hydrogen. Oxidative transformations usually involve either increases in molecular mass without increasing hydrogen content, or decreases in hydrogen content. In both cases, the fraction of the molecular mass accounted for by hydrogen atoms decrease, and values of RMD also decrease. For example, oxidative conversion of  $\alpha$ -santalene ( $C_{15}H_{24}$ ) to  $\alpha$ -santalenol ( $C_{15}H_{24}O$ ) by addition of a single oxygen atom shifts RMD downward to 830 ppm, and insertion of a second oxygen atom ( $C_{15}H_{24}O_2$ ) yields RMD of 752 ppm. This information has particular

relevance when searches for terpenoids employ LC-MS for analysis because electrospray ionization usually requires molecules to have acidic or basic groups. For terpenoids, this means that one or more oxygen atoms must be in the molecule for it to be ionized. In many organisms, terpenoids may be conjugated to polar groups (e.g. glycosides or phosphates) that decrease a terpenoid's RMD value further because they decrease the fractional hydrogen content. For example, each glycosylation by a hexose adds  $C_6H_{10}O_5$  to a terpenoid, so the glycoside of a sesquiterpene alcohol ( $C_{21}H_{34}O_6$ ) would have  $RMD = 616$  ppm. Additional oxidation or conjugation by malonate (addition of  $C_3H_2O_3$ ) would decrease the RMD value of terpenoid metabolites further. The diterpene glycosides and their malonate esters [6] provide examples. The RMD of nicotianoside VII (a diterpenoid glycoside of formula  $C_{50}H_{78}O_{24}$  with two malonate esters) is 426 ppm and nicotianoside VI (one malonyl group less with elemental formulae of  $C_{47}H_{76}O_{24}$ ) has RMD of 461 ppm. Since conjugated terpenoids usually consist of a terpenoid core that is rich in reduced carbon and conjugate groups of low fractional hydrogen content, the RMD of terpenoids conjugated to sugar molecules vary from ~400 ppm to ~600 ppm. In contrast, polyphenolic metabolites have low hydrogen content, and their RMD is less than 300 ppm (230 ppm for the phenolic acid salicylic acid, 167 ppm for the flavonoid kaempferol).

Negative ion mode mass spectrometry aids identification of terpenoid conjugate functional groups based on neutral losses observed, [7] but negative ion tandem mass spectra often fail to exhibit product (fragment) ions characteristic of the unmodified terpenoid core. For instance, negative ion mode MS/MS spectra of common hexose derivatives (e.g. mono-, di- and tri-saccharides) display fragment ions (Table 2.1) characteristic of carbohydrate groups, as well as fragments derived from neutral mass losses that indicate losses of those carbohydrates from the pseudomolecular ions. Positive ion mode MS/MS spectra also allows for the annotation of mono-

Table 2.1: Characteristic fragment ions observed in negative ion mode MS/MS experiments for various sugar oligosaccharides and mono saccharides. The masses shown correspond to  $[M-H]^-$  formed by each sugar group. The MS/MS spectra of candidate terpenoid compounds were examined for the presence of these fragment ions for identification of the presence of these oligosaccharides in the terpenoids.

Negative ion mode fragment ion <i>m/z</i> (Theoretical exact mass)	Common sugar moiety
503.1618	Triglycoside(Hexose-Hexose-hexose; $C_{18}H_{31}O_{16}^-$ )
485.1512	Triglycoside - $H_2O$ ( $C_{18}H_{29}O_{15}^-$ )
589.1622	Triglycoside malonate ester ( $C_{21}H_{33}O_{19}^-$ )
571.1516	Triglycoside malonate ester - $H_2O$
341.1089	Diglycoside (hexose-hexose; $C_{12}H_{21}O_{11}^-$ )
323.0984	Diglycoside - $H_2O$ ( $C_{12}H_{19}O_{10}^-$ )
179.0561	Monoglycoside (Hexose; $C_6H_{11}O_6^-$ )
221.0667	Monoglycoside acetate ester ( $C_8H_{13}O_7^-$ )
161.0455	Monoglycoside - $H_2O$ ( $C_6H_9O_5^-$ )
101.0232	Fragment ion from hexoses ( $C_4H_5O_3^-$ )
113.0228	Fragment ion from hexoses
125.0244	Fragment ion from hexoses

, di- and tri-saccharides and their respective malonylated and acetylated forms based on fragment ions characteristic of the sugar residues and neutral losses characteristic of the sugar and malonate groups [8,9]. These results are similar to MS/MS characterization of flavonoid

glycosides from *Arabidopsis thaliana* using LC-MS/MS [9]. Based on the presence of characteristic fragment ions in these MS/MS spectra, the presence of disaccharides rhamnose – glucose, xylose-glucose and trisaccharides (rhamnose – glucose – glucose) was established [9]. Carbohydrate-derived fragment ions have RMD values around 300 ppm, and product ions in MS/MS spectra with RMD values that fall in this range are suggestive of glycosides. In the case of a terpenoid glycoside, one would expect RMD of fragments derived from carbohydrates to lie around 300 ppm, the molecular RMD to fall in the range of 400-600 ppm, and fragments derived from the terpenoid core (without carbohydrates) to approach 800 ppm. Such patterns of RMD values in MS/MS spectra aid annotation of unknown metabolites as candidate terpenoid glycosides.

Owing to the scarcity of information regarding non-volatile terpenoids in the genus *Solanum*, this study has explored LC-MS metabolite profiling data generated for extracts of the wild tomato *S. habrochaites* LA1777. To accelerate recognition of terpenoid metabolites, filtering of molecular and fragment masses using RMD was performed. The study is explained in detail below.

## 2.3 Materials and Methods

### 2.3.1 LC-MS and MS/MS experiments

All LC-MS and LC-MS/MS experiments were performed using a Waters Xevo G2-S QToF mass spectrometer equipped with the standard electrospray ionization ion source. This instrument was coupled to a Waters Acquity ultra-high pressure liquid chromatography system. Separations were performed using an Ascentis Express C-18 UHPLC column (2.1 mm × 100 mm, 2.7 μm; Supelco Co. USA). For negative ion mode experiments, solvents were 0.15% aqueous formic acid - pH 2.85 (A) and methanol (B). Solvent gradient (A:B) used for profiling of conjugated terpenoids was as follows: 0-1.00 min (99:1), 1.01-4.00 (linear ramp to 55:45), hold at (55:45) 4.01-9.00 min, step to (50:50) and hold at (50:50) over 9.01-14.00 min, step to (1:99) and hold over 14.01-17.00 min, step to (99:1) and hold over 17.01-20.00 min. Flow rate was 0.30 ml/min and column temperature was held at 35 °C. Mass spectra were acquired using negative-ion mode electrospray ionization and dynamic range extension over  $m/z$  50-1500, with mass resolution ( $M/\Delta M$ , full width-half maximum) approximately 20000. Five parallel collision energy functions were used, with 0.1 s per function. Collision cell potentials used for negative ion mode fragmentation for each function were 5, 15, 25, 35 and 45 V respectively. Other parameters include capillary voltage of 3.0 kV, desolvation temperature of 350 °C, source temperature of 100 °C, Cone gas (N<sub>2</sub>) at 0 L/hr and desolvation gas (N<sub>2</sub>) at 800 L/hr. Positive ion mode MS/MS experiments were performed using the same instrument specifications and LC method, except, for solvent A, 10 mM ammonium formate (pH 2.65) was used instead of 0.15% formic acid solution. In all analyses, leucine enkephalin (0.1 ng/μL) was used as a lock mass reference infused at a flow rate of 3.0 μL/min and lock mass spectra were acquired for 100 ms at 10 sec intervals. Capillary voltage for lock spray was 2.5 kV.

All MS/MS experiments were performed on the same instrument employing the same LC methods. For negative ion mode MS/MS experiments a collision potential of -50 V was used and for positive ion mode experiments +30 V was used.

### **2.3.2 Plant material**

Sample preparation for metabolite profiling: *Solanum habrochaites* LA1777 plants were grown in Michigan State University plant growth chambers (28 °C, 16:8 h day/night cycle, 150  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , 96% humidity) for six weeks using seeds obtained from the C. M. Rick Tomato Genetics Resource Center at the University of California-Davis. Ten leaflets were harvested from each plant (six weeks post germination) and were extracted by dipping in 5 mL of methanol: water (80:20 v/v) for about 30 sec. Three biological replicates were used for the profiling. Extracts were concentrated by drying under a stream of  $\text{N}_2$  gas at room temperature, and the residues were redissolved in 0.5 ml methanol: water (80:20 v/v). The extract was centrifuged (10000g for 10 min at 25 °C) to remove debris, the supernatant was transferred to an autosampler vial, and 2.5  $\mu\text{L}$  was introduced to the LC column. Automated peak detection, integration, and retention time alignment were performed using Waters MarkerLynx XS software, and lists of  $m/z$  values, retention times, and extracted ion chromatogram peak areas were exported as text files and processed further using Microsoft Excel software.

### **2.2.3 Data processing**

Automated chromatographic peak detection was performed using Waters MarkerLynx XS software. The lowest collision energy function (function 1) was used for peak detection, integration, retention time alignment, and deisotoping. The parameters used with MarkerLynx processing were as follows. Marker intensity threshold: 800 counts, mass window 0.05 Da,

retention time window: 0.25 min,  $m/z$  range: 100 – 1500, retention time range 0.5 – 20.0 min. No peak smoothing was applied.

## 2.4 Results and discussion

### 2.4.1 Recognition of sesquiterpene glycosides from ion relative mass defects

Analysis of leaf extracts of *S. habrochaites* LA1777 using LC-multiplexed CID MS in negative ion mode yielded evidence of complex mixtures of chemicals (Figure 2.1).

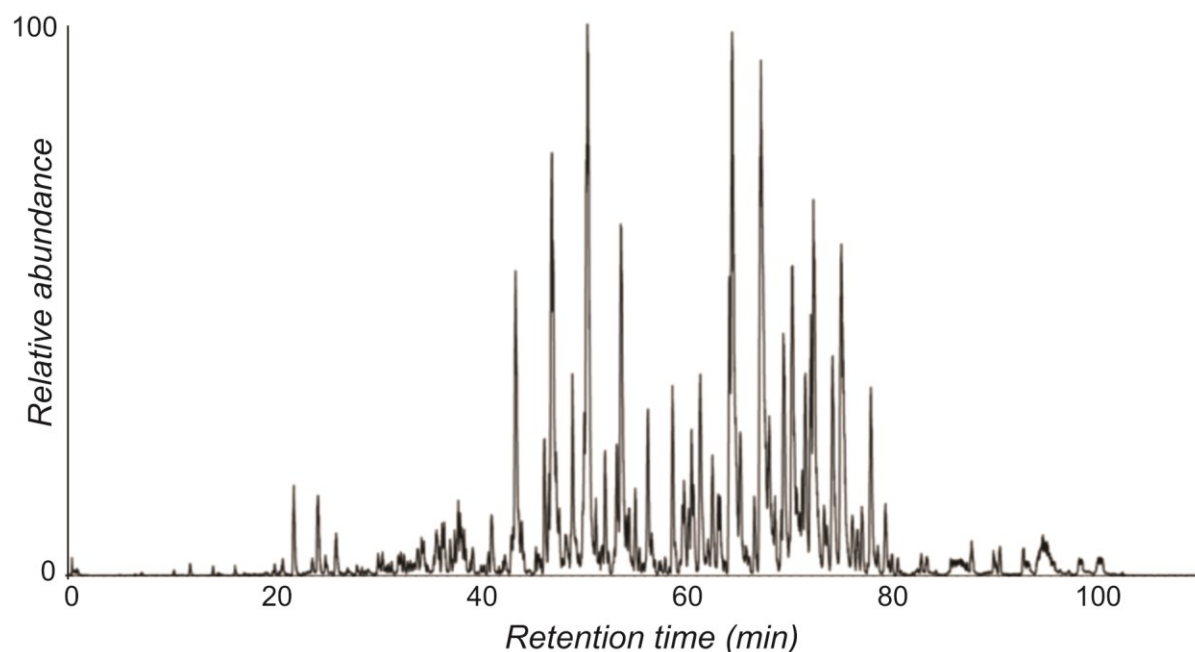


Figure 2.1 Complexity of a plant extract is evident from the number of peaks in a UHPLC-MS base peak chromatogram generated from a leaf dip extract of *S. habrochaites* LA1777. Analysis was performed using a 110 min chromatographic gradient and detected in negative ion mode.

Automated peak detection, deisotoping, integration, and retention time alignment using Waters MarkerLynx XS software yielded a total of 3280  $m/z$ -retention time pairs, estimated to arise from about 1000 distinct metabolites (Fig. 2.2).

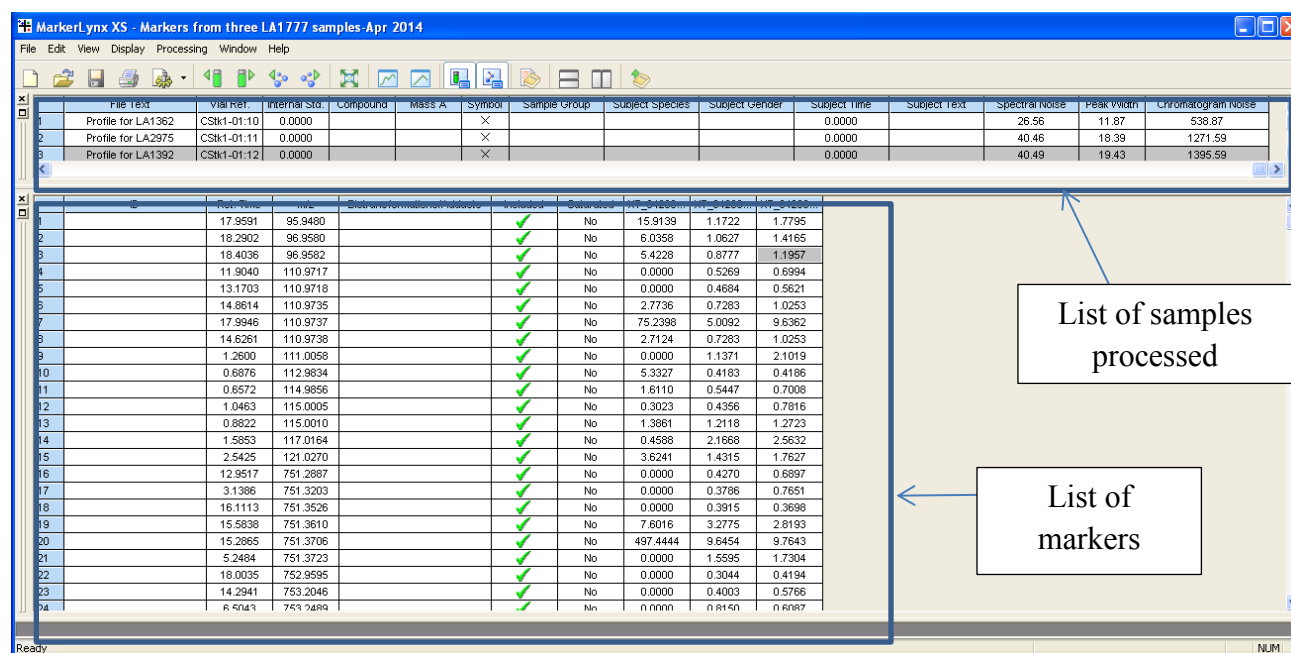


Figure 2.2: MarkerLynx XS data processing results in a list of markers detected from the *S. habrochaites* LA1777 trichome leaf extracts. Markers detected are reported with their  $m/z$  ratio, retention time and their abundance in each sample being considered.

Examination of the RMD values for the entire data set indicated that 3199 (98 %) of the ions detected in automated peak picking had positive absolute mass defects, with the remainder having negative absolute mass defects typical of inorganic salt cluster ions and instrument contaminants (*e.g.* trifluoroacetate,  $\text{NaHPO}_4^-$ ). Of the ions with positive mass defects, 1805 (55% of total) displayed RMD in the range of 400 to 650 ppm and 1177 (36% of total) possessed RMD from 200 to 400 ppm, the latter range being more typical of aromatic, probably polyphenolic metabolites. Since the objective of this exercise was identification of sesquiterpene glycosides from this data set, three theoretical boundary condition approximations were made: 1) based on the theoretical  $m/z$  of 383.2439 for  $[\text{M}-\text{H}]^-$  of farnesol monoglycoside ( $\text{C}_{21}\text{H}_{35}\text{O}_6^-$ ), the maximum RMD a sesquiterpene glycoside (with minimum of one hexose moiety) can display is 636 ppm;



2) based on the theoretical  $m/z$  of 869.4024 for  $[M-H]^-$  farnesol tetraglycoside, minimum RMD a sesquiterpene glycoside (with maximum of four hexose moieties) can display is 463 ppm; 3) based on the theoretical  $m/z$  of 383.2439 for  $[M-H]^-$  of farnesol monoglycoside, minimum nominal  $m/z$  of a sesquiterpenoid monoglycoside should be 383. It is notable that some terpenoid compounds display malonylation as evidenced by the malonylated diterpene glycosides of *Nicotiana attenuata* and malonylated sesquiterpenes of *Panax ginseng*. [6,10,11,12]. Similarly, acetylated terpenes and acetylated terpene glycosides have been isolated from *Iphiona scabra* and *Combretum imberbe* [13,14]. Dozens of such malonylated and acetylated terpenoids have been reported [15,16]. As a result, rule #2 described above is modified to account for such possibilities, setting the lower limit of RMD to 440 ppm to account for malonylated sesquiterpene compounds (RMD of a farnesol triglycoside malonate ester would be 441 ppm). Similarly, to account for minor mass detection errors the maximum RMD possible for sesquiterpene glycosides with minimum of one hexose moiety was set to 640 ppm.

Applying the first two conditions, the number of compounds with RMD 440 to 640 ppm detected is 1280 (38% of total markers). Applying the third condition to this set of 1280 markers results a list of 1074 markers (33% of total) that are putative terpene glycosides. Next, this list of markers was arranged in the descending order of peak area and the 200 most abundant markers were selected as candidate terpenoid glycosides. The first 38 compounds in this list are listed in Table 2.3.

*Setting the rules based on the requirement:* The boundary conditions set in rules discussed here were set based on the objective of identifying sesquiterpene glycosides. These need to be re-established if the objective was diterpene or triterpene glycoside identification. For instance for diterpene glycoside identification the rules will be established based on geranylgeranyl

monoglycoside since geranylgeranyl acts as the precursor of diterpenes. This will result in the following rules. 1.) Based on theoretical mass of geranylgeranyl monoglycoside the minimum  $m/z$  of  $[M-H]^-$  of a diterpene glycoside would be 451.3065 and the maximum RMD a diterpene glycoside can display is 679 ppm. 2.) Based on theoretical mass of geranylgeranyl tri glycoside the minimum  $m/z$  of  $[M-H]^-$  of a diterpene triglycoside would be 775.4122 and the minimum RMD a diterpene glycoside can display is 531 ppm. 3) based on the theoretical  $m/z$  of 451.3065 for  $[M-H]^-$  of geranylgeranyl monoglycoside, minimum nominal  $m/z$  of a diterpenoid monoglycoside should be 451. Therefore, compounds in RMD range 500 – 700 ppm will need to be considered if the objective was to discover diterpene glycosides.

*Distinguishing terpenoid glycosides from other compounds:* Applying the RMD criteria described allows for the elimination of most, if not all, compounds that are not terpene glycosides. However, the final list of compounds may possess a number of non-terpenoid compounds, and to distinguish these from terpenoids, calculation of RMD of the fragment ions generated using either MS/MS or multiplexed CID should be performed. Inclusion of non-terpenoid compounds in this list occurs because a given elemental formula (and  $m/z$  value) can correspond to a number of different chemical structures as discussed in introduction section. Therefore, in order to identify terpenoids, additional information Fragment ion RMD values distinguish terpenoid glycosides from other compounds since losses of all carbohydrate moieties will yield a fragment ion corresponding to a terpenoid core, with RMD values > 800 ppm. Furthermore, even if all sugars are not removed, the RMD of the fragment ions formed by terpenoid glycosides will be greater than that of the pseudomolecular ion because RMDs are about 325-350 ppm for hexoses and hexose fragments, values much less than for terpenoid cores. Therefore, terpenoid glycosides are characterized by fragment ions that display an increasing

RMD as their mass decreases due to the loss of sugar groups. This phenomenon is illustrated in Figure 2.3, which shows the MS/MS spectra of three of the five metabolites with pseudomolecular ion RMD falling in the 440 to 636 ppm range and with greatest peak areas. Among the fragment ions of these compounds, only the fragments  $m/z$  199 (RMD  $\sim$  840 ppm; Fig. 2.3.A and 2.3.B) display RMD close to that of isoprene (919 ppm), but,  $m/z$  199 has a mass too low to be a sesquiterpenoid core ( $C_{15}H_{24}$  would be 204 Da). In addition, there is no systematic increase of RMD with as fragment masses decrease among the fragment ions of in any of these compounds, suggesting the groups being lost have hydrogen contents similar to, or greater than, the intact molecule. These findings suggest the molecules are not terpenoid glycosides, and in fact, they are all acylsucroses. In contrast, the MS/MS spectra observed for glycosides of the sesquiterpene alcohol campherenane diol (discussed below) shown in Figure 2.3 shows a clear systematic increase in RMD as fragment masses decrease, consistent with a terpenoid glycoside.

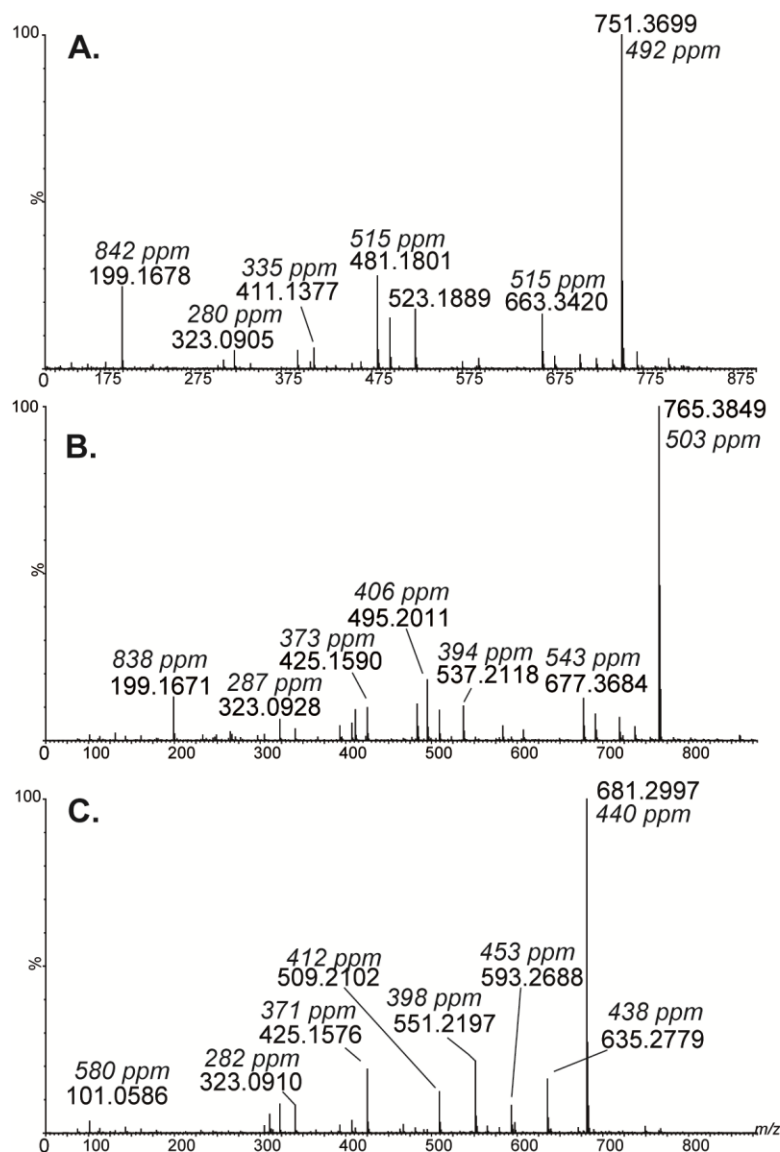


Figure 2.3. (A) Negative ion mode multiplexed CID mass spectrum of acylsugar S4:22, RMD = 492 ppm. (B) Negative ion mode multiplexed CID mass spectrum of acylsugar S4:23, RMD = 402 ppm. (C) Negative ion mode multiplexed CID mass spectrum of acylsugar S4:17, RMD = 440 ppm. Values for RMD of the major fragment ions are presented. All detected isomers displayed fragments of the same  $m/z$  values in negative ion mode spectra. All displayed negative ion mode multiplexed CID mass spectra were obtained using a collision potential of -60 V.

### 2.4.2 Discovery of sesquiterpene diol glycosides from *S. habrochaites* LA1777

In the list of the 200 *S. habrochaites* LA1777 metabolites with greatest peak areas within the RMD range 440-640 ppm (as discussed in Section 2.4.1), the metabolite  $m/z$  of 609 was ranked 17<sup>th</sup> in peak area. Negative-ion mode multiplexed CID mass spectra of this compound revealed a systematic increase of RMD with decreasing fragment masses, as is characteristic of terpenoid glycosides. In order to ensure that these fragment ions were derived from the proposed pseudomolecular ion, MS/MS spectra were generated.

The MS/MS spectrum of this compound (Figure 2.4A) revealed  $m/z$  563.3045 as a fragment ion. This was 46 Da less than the parent ion mass. Molecular mass of formic acid is also 46 Da. Since formic acid was in the mobile phase used for this experiment it was concluded  $m/z$  609 represented a formate adduct of  $m/z$  563. Such ionization behavior is typical of compounds that lack acidic functional groups, but can form pseudomolecular ions by anion adduction. The next most abundant fragment was  $m/z$  401, which was 162 Da less than  $m/z$  563. 162 Da neutral loss is a common hexose sugar neutral loss and therefore, 401 represents a form that has one less sugar moiety than 563. Note that the RMD of each of these fragments (539 ppm and 631 ppm for  $m/z$  563 and 401 respectively) increases as the  $m/z$  of fragment decreased (Fig. 2.4A). Since this is a characteristic feature of terpene glycosides, the observations are consistent with this metabolite being a member of this class. However, no prominent fragment ions with high RMD typical of terpenoid cores were observed with  $m/z < 401$ . An elemental formula calculation for  $m/z$  401.2532 yielded a formula of  $C_{21}H_{37}O_7^-$ , which has 6 more than the 15 carbon atoms that are usually found in sesquiterpenes. Therefore, potentially  $m/z$  401 represents the terpenoid core with other groups present in it. The low mass fragment ion of  $m/z$  323 is a characteristic fragment corresponding to [dihexose-H-H<sub>2</sub>O]<sup>-</sup> ( $C_{12}H_{19}O_{10}^-$ ) suggesting the presence of a diglucoside in the

compound. Based on the presence of this  $m/z$  323 it can be proposed that this compound has a minimum of two hexose groups attached to it, and that they are linked to one another. It is important to note that for further characterization of this compound required that it be purified and subjected to structure determination using 1D and 2D NMR (Chapter 3), since further fragmentation of the core was not observed in negative ion mode.

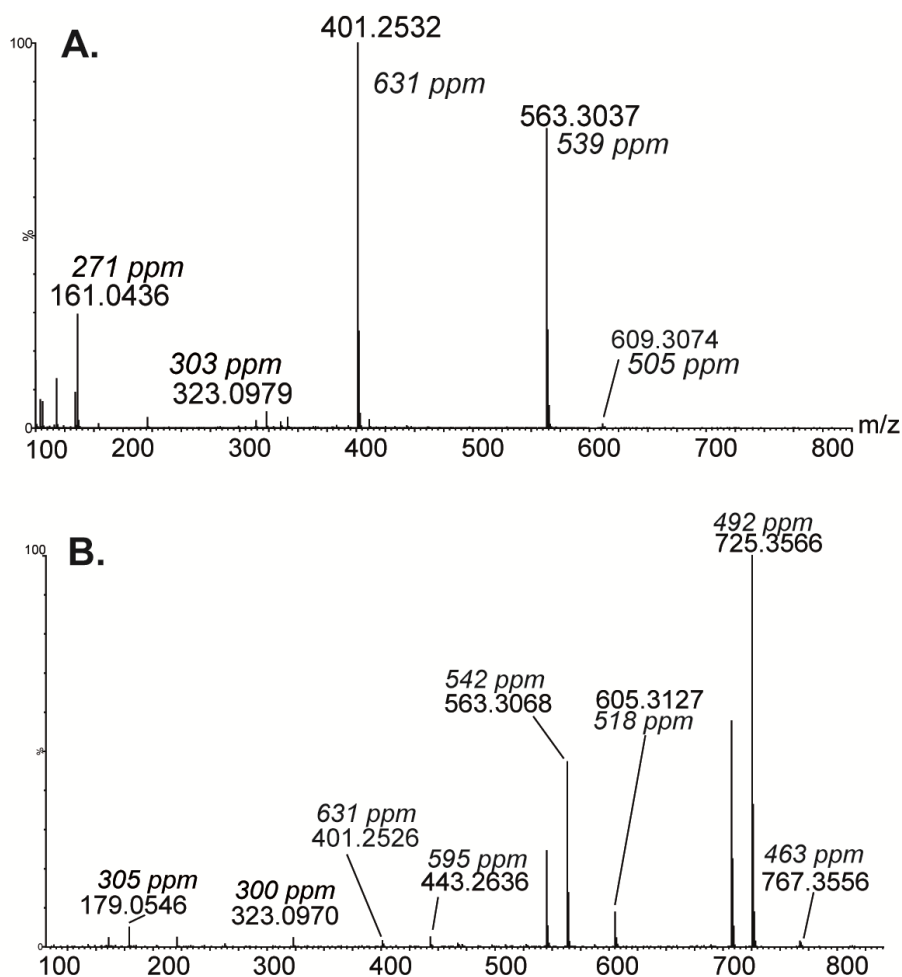


Figure 2.4 Negative-ion mode product ion MS/MS spectra of metabolites extracted from *S. habrochaites* LA1777. (A) MS/MS spectrum of  $m/z$  609 from campherenane diol diglucoside; (B) MS/MS spectrum of  $m/z$  811 from campherenane diol triglycoside malonate ester ( $m/z$  811.3587, RMD = 442 ppm). Values for RMD of the major fragment ions are presented. All chromatographically-resolved isomers (10 isomers of  $m/z$  651 and 12 isomers for  $m/z$  811) displayed fragments of the same  $m/z$  values in negative ion mode spectra. All negative ion mode MS/MS data were obtained using collision potential of -50 V.

Discovery of sesquiterpene diol triglycoside malonate ester from *S. habrochaites* LA1777 provides an example that involves the presence of additional groups other than sugar moieties. The exact mass of this molecule was  $m/z$  811.3587 in negative ion mode (RMD 442 ppm). In the MS/MS spectrum (Figure 2.3B) of the compound it fragments by loss of  $\text{CO}_2$  to give  $m/z$  767.3707 (RMD 483 ppm) followed by the loss of  $\text{C}_2\text{H}_2\text{O}$  to give  $m/z$  725.3587 (RMD 495 ppm). The loss of  $\text{CO}_2$  (44 Da) followed by the loss of another 42 Da ( $\text{C}_2\text{H}_2\text{O}$ ) is characteristic of malonate esters. Next, the loss of the first anhydroglucose (neutral loss of 162 Da) leads to fragment ion  $m/z$  563.3057 (RMD 543 ppm), and loss of a second anhydroglucose group resulted in  $m/z$  401.2516 (RMD 627 ppm). Loss of a third anhydroglucose results in fragment ion of  $m/z$  239.2017 (RMD 841 ppm). Fragment ion 239.2017 represents the sesquiterpenoid core ( $\text{C}_{15}\text{H}_{27}\text{O}_2^-$ ) of the molecule as it does not undergo any further fragmentation in negative ion mode. This compound also displays the characteristic feature of conjugated terpenoids – presence of fragments with increasing RMD value with decreasing  $m/z$ . Based on these observed characteristics this compound can be annotated as a sesquiterpenoid triglycoside malonate ester.

The described method can also be applied to other types of terpenoids. For instance tomatine, a triterpenoid glycoside also displays similar behavior. RMD of the  $[\text{M}+\text{formate}]^-$  ion of tomatine in negative ion mode is 520 ppm. The major fragments display a gradual increase of RMD from 520 ppm to 674 ppm, and correspond to losses of carbohydrate moieties (Fig. 2.5).



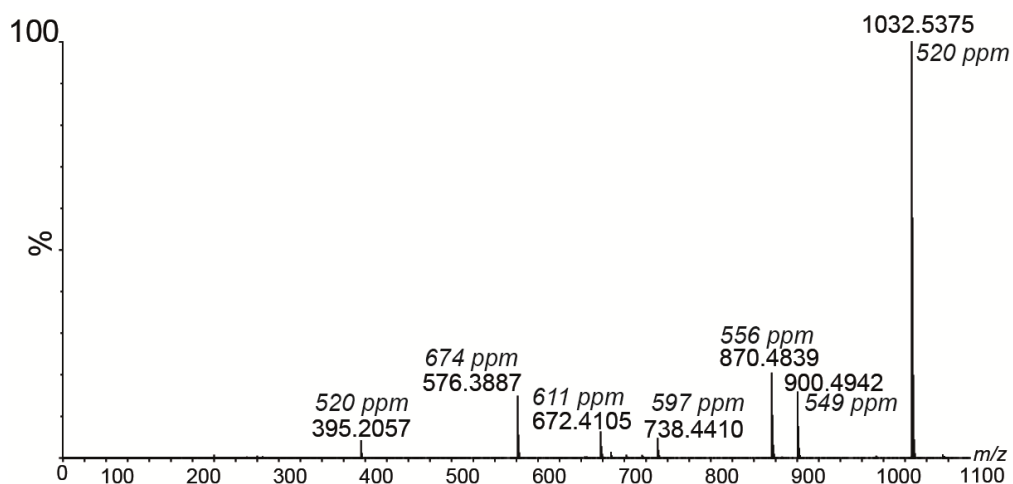


Figure 2.5: Negative ion mode multiplexed CID mass spectrum of tomatine from *S. habrochaites* LA1777 obtained at 60 V collision potential. Note that the fragment ions all have RMD values greater than or equal to the  $[M-H]^-$  ion at  $m/z$  1032.5.

Plotting RMD as a function of  $m/z$  for some example *S. habrochaites* metabolites provides a clear image of this difference of RMD variation among terpene glycoside fragment ions from the fragment ions of other compounds (Fig. 2.6).

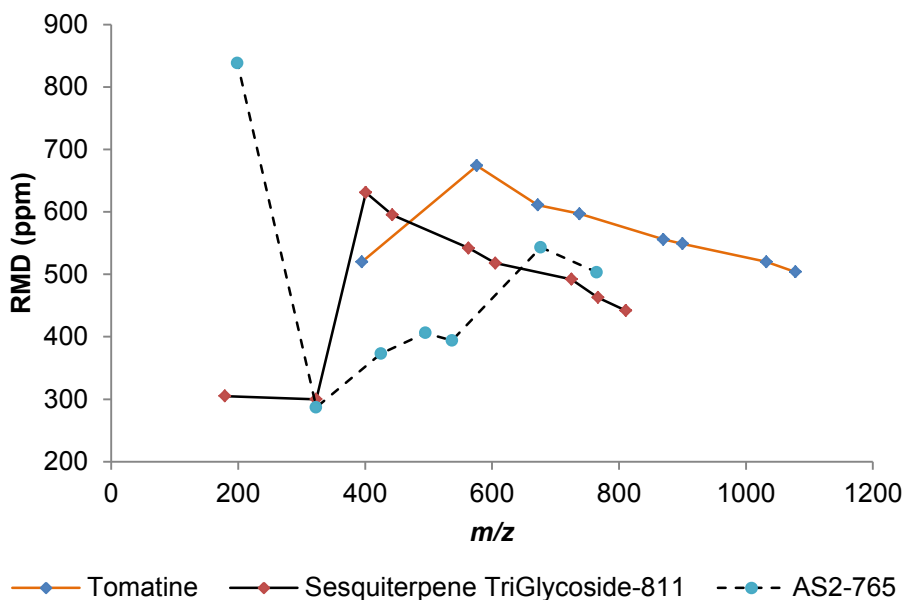


Figure 2.6: Variation of RMD of fragment ions as a function of ion  $m/z$ . Fragment ions were generated in negative ion mode MS/MS for some representative compounds (acyl sugars, triterpenoid glycoside and a triglycoside malonate ester) found in *S. habrochaites* LA1777 leaf dip extracts.

The proposed process of applying RMD filtering for the discovery of novel glycosylated terpenoids is described graphically in Figure 2.7. Applying this procedure to the most abundant 200 compounds in the list allowed for the identification of a total of 224 different terpene glycoside (including the isomers) from *S. habrochaites* plant extract (Table 2.2). Three different sesquiterpenes cores were established including the campherenane diol core. The structures of some of these compounds were established using NMR after purifying them. Structure elucidation of these compounds (including the sesquiterpenes I, II and III terpenoid cores listed in Table 2.2) is described in Chapter 3 of this dissertation. The MS/MS data generated for each of these compounds and the RMD of each fragment ion are presented in Figs. 2.8-2.28.

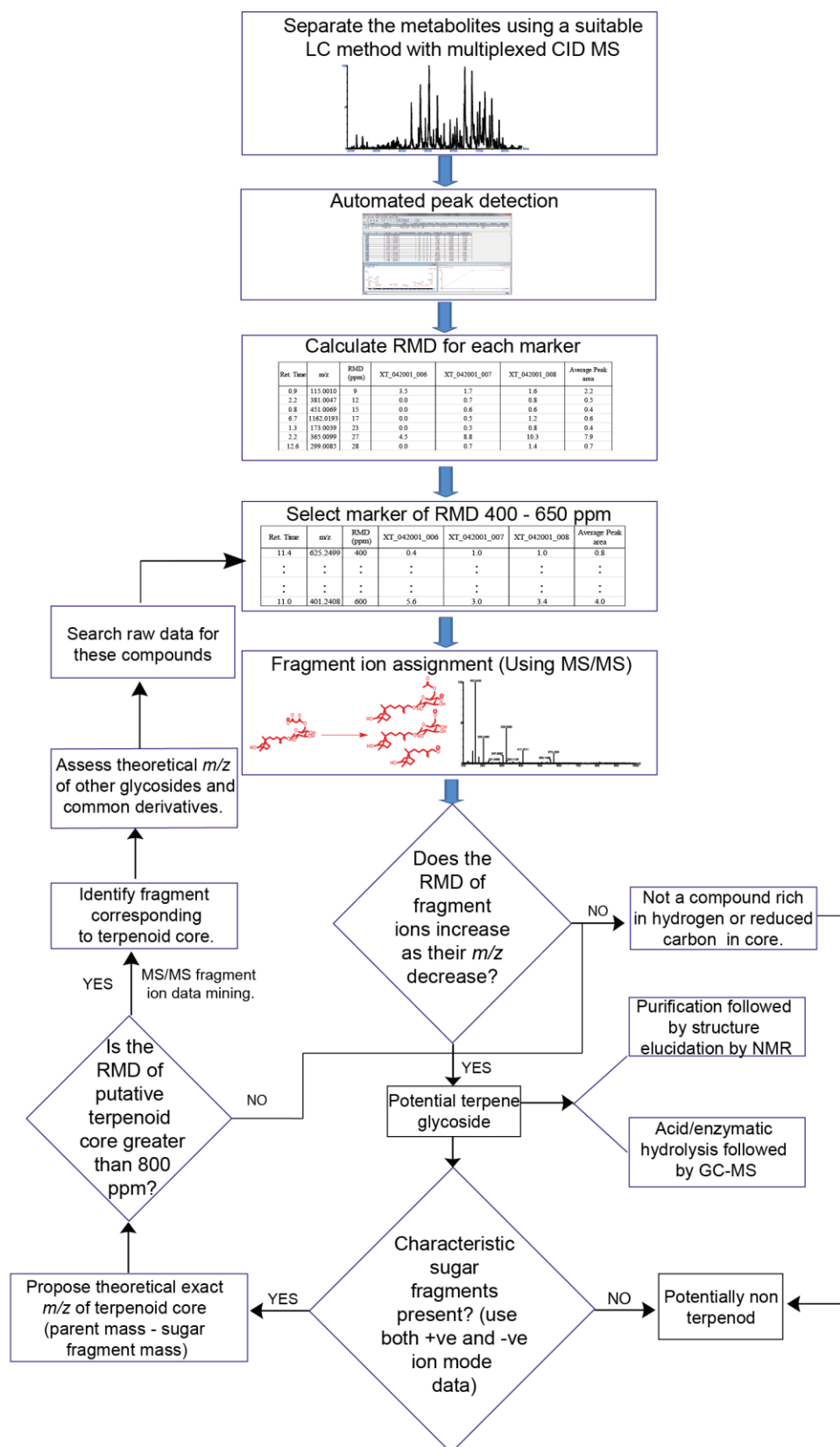


Figure 2.7: Relative mass defect filtering process used for discovering conjugated terpenoids from raw LC-MS data

Table 2.2: Compounds identified as terpene glycosides from *S. habrochaites* LA1777 based on RMD filtering of molecular and fragment ions.

Compound #	Experimental $m/z$	RMD (ppm)	Proposed elemental formula of neutral molecule	$\Delta m$ (ppm)	Compound type	Measured $m/z$ of terpenoid core fragment ion	Proposed elemental formula of terpenoid core fragment ion	$\Delta m$ (ppm)	RMD of terpenoid core fragment (ppm)	# of Isomers observed
1	545.2932	538	C <sub>27</sub> H <sub>46</sub> O <sub>11</sub>	-1	Sesquiterpene II Diglycoside	221.1527	C <sub>14</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>	-9	691	8
2	587.3049	519	C <sub>29</sub> H <sub>48</sub> O <sub>12</sub>	-3	Sesquiterpene II Diglycoside acetate ester	221.1529	C <sub>14</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>	-8	692	1
3	631.2929	464	C <sub>37</sub> H <sub>44</sub> O <sub>13</sub>	3	Sesquiterpene II Diglycoside malonate ester	221.1565	C <sub>14</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>	8	708	2
4	707.3112	440	C <sub>32</sub> H <sub>52</sub> O <sub>17</sub>	-3	Sesquiterpene II Triglycoside	221.1539	C <sub>14</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>	-4	696	4
5	793.3474	438	C <sub>36</sub> H <sub>58</sub> O <sub>19</sub>	1	Sesquiterpene II diol Triglycoside malonate ester	221.1527	C <sub>14</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>	-9	691	11
6	401.252	628	C <sub>21</sub> H <sub>38</sub> O <sub>7</sub>	-6	Campherenane diol Monoglycoside	239.1997	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-8	836	6

Table 2.2 (cont'd)

7	443.2632	594	C <sub>23</sub> H <sub>40</sub> O <sub>8</sub>	-4	Campherenane diol Monoglycoside acetate ester	239.1994	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-9	834	14
8	445.2421	544	C <sub>22</sub> H <sub>38</sub> O <sub>9</sub>	-4	Campherenane diol Monoglycoside derivative	239.1999	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-7	836	5
9	487.2536	520	C <sub>24</sub> H <sub>40</sub> O <sub>10</sub>	-3	Campherenane diol Monoglycoside malonate ester	239.1994	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-9	834	9
10	563.3055	542	C <sub>27</sub> H <sub>48</sub> O <sub>12</sub>	-3	Campherenane diol diglycoside	239.2048	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	13	857	7
11	605.3152	521	C <sub>29</sub> H <sub>50</sub> O <sub>13</sub>	-4	Campherenane diol diglycoside acetate ester	239.1993	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-10	834	9
12	649.3042	469	C <sub>30</sub> H <sub>50</sub> O <sub>15</sub>	-5	Campherenane diol diglycoside malonate ester	239.1993	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-10	834	14
13	725.3607	497	C <sub>33</sub> H <sub>58</sub> O <sub>17</sub>	2	Campherenane diol triglycoside	239.2007	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-4	840	6
14	811.3587	442	C <sub>36</sub> H <sub>60</sub> O <sub>20</sub>	-2	Campherenane diol triglycoside malonate ester	239.2006	C <sub>15</sub> H <sub>27</sub> O <sub>2</sub> <sup>-</sup>	-4	839	13
15	411.1986	483	C <sub>21</sub> H <sub>32</sub> O <sub>8</sub>	-9	Sesquiterpene III monoglycoside	249.1481	C <sub>15</sub> H <sub>21</sub> O <sub>3</sub> <sup>-</sup>	-4	595	13
16	453.2082	459	C <sub>23</sub> H <sub>34</sub> O <sub>9</sub>	-9	Sesquiterpene III monoglycoside acetate ester	249.1483	C <sub>15</sub> H <sub>21</sub> O <sub>3</sub> <sup>-</sup>	-3	596	14

Table 2.2 (cont'd)

17	497.2011	404	C <sub>24</sub> H <sub>34</sub> O <sub>11</sub>	-3	Sesquiterpene III monoglycoside malonate ester	249.1475	C <sub>15</sub> H <sub>21</sub> O <sub>3</sub> <sup>-</sup>	-6	592	11
18	413.2143	519	C <sub>21</sub> H <sub>34</sub> O <sub>8</sub>	-9	Sesquiterpene I monoglycoside	251.1634	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-8	651	17
19	455.227	499	C <sub>22</sub> H <sub>36</sub> O <sub>9</sub>	-4	Sesquiterpene I monoglycoside acetate ester	251.1638	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-6	653	21
20	499.2162	433	C <sub>24</sub> H <sub>36</sub> O <sub>12</sub>	-5	Sesquiterpene I monoglycoside malonate ester	251.1643	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-4	655	11
21	617.2795	453	C <sub>29</sub> H <sub>45</sub> O <sub>14</sub>	-2	Sesquiterpene I diglycoside acetate ester	251.1636	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-7	652	4
22	661.2693	407	C <sub>30</sub> H <sub>46</sub> O <sub>16</sub>	-3	Sesquiterpene I diglycoside malonate ester	251.1639	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-6	653	8
23	823.3245	394	C <sub>36</sub> H <sub>56</sub> O <sub>21</sub>	1	Sesquiterpene I triglycoside malonate ester	251.164	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-5	653	5
24	985.3742	380	C <sub>30</sub> H <sub>50</sub> O <sub>26</sub>	-3	Sesquiterpene I tetraglycoside malonate ester	251.1641	C <sub>15</sub> H <sub>23</sub> O <sub>3</sub> <sup>-</sup>	-5	654	11

---

Table 2.3: Compounds with greatest peak areas among the list of *S. habrochaites* LA1777

metabolites in the RMD range 440-636 ppm

Ret. Time	<i>m/z</i>	Average Peak area	RMD (ppm)
15.3	765.3836	219.44	501
15.3	751.3706	172.28	493
17.8	707.4408	170.61	623
17.5	707.4412	167.58	624
15.2	723.3353	140.38	464
15.1	681.3032	122.81	445
15.2	737.3490	120.67	473
15.0	871.4180	109.13	480
17.3	707.4419	98.18	625
14.8	873.3857	86.43	442
15.3	925.5072	64.32	548
11.0	975.4882	63.70	500
14.5	857.3901	63.41	455
15.3	911.5283	60.36	580
15.1	709.3192	59.31	450
15.4	967.5318	52.59	550
6.8	609.3073	47.11	504
15.1	901.4240	43.67	470
15.0	887.4166	40.09	469
12.4	609.2946	34.21	484
6.5	609.3055	31.96	501
13.7	413.2069	30.53	501
15.2	695.3402	29.89	489
15.3	723.3752	29.74	519
16.1	663.3840	29.15	579
14.2	455.2153	28.87	473
15.0	873.4442	28.52	509
7.3	1078.5147	27.86	477
13.2	455.2149	27.17	472
15.2	1011.5815	26.43	575
11.3	651.3034	25.33	466
10.7	455.2166	24.14	476
8.6	651.3154	23.74	484
7.4	1078.5139	23.15	476
15.0	885.4205	22.62	475

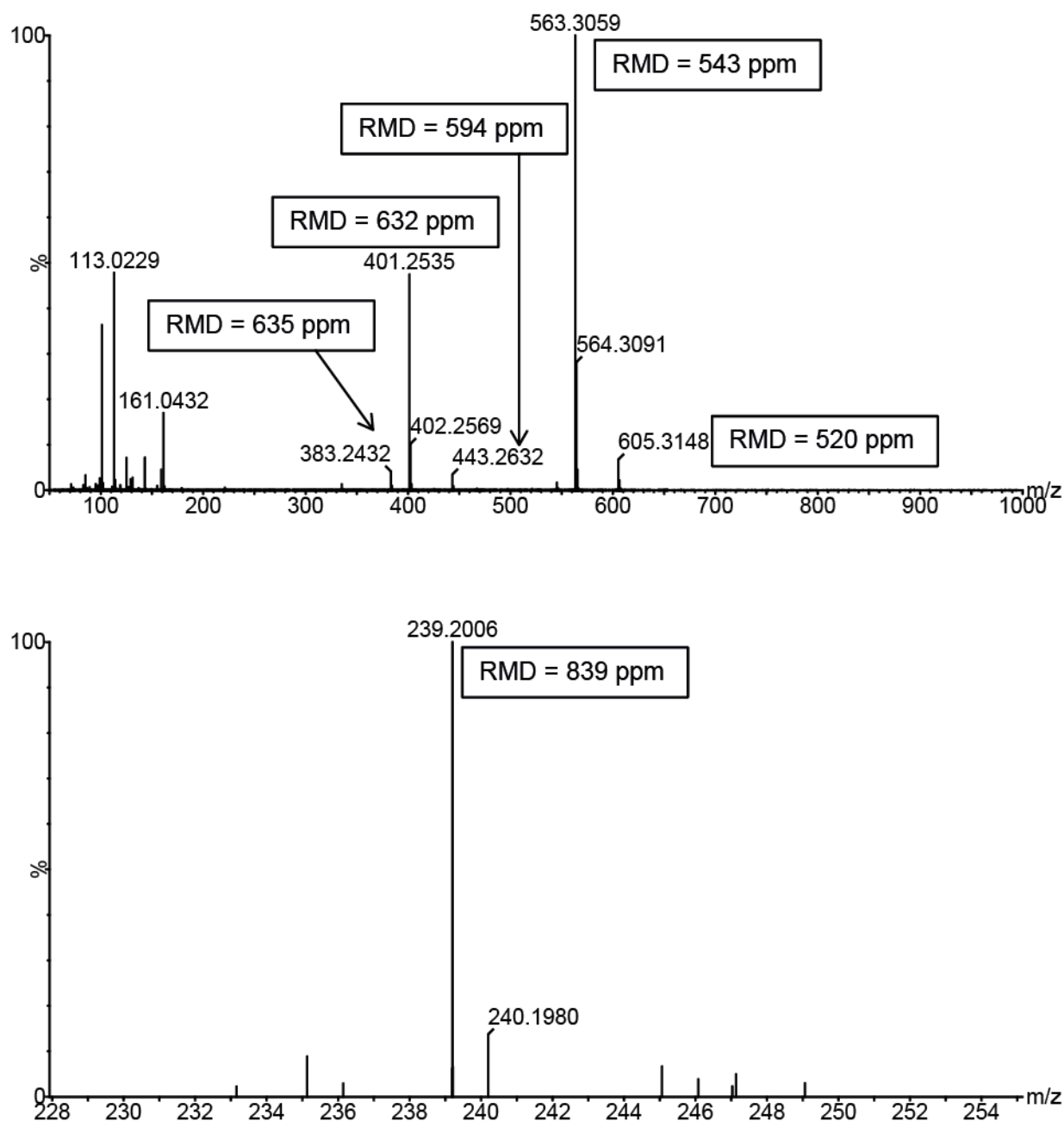


Figure 2.8. (A) Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  649) for campherenan-2,12-diol malonate ester. (B) magnified region of  $m/z$  228-255, showing the sesquiterpenoid core fragment at  $m/z$  239 which was too small to observe in A, where peaks are normalized to the base peak.



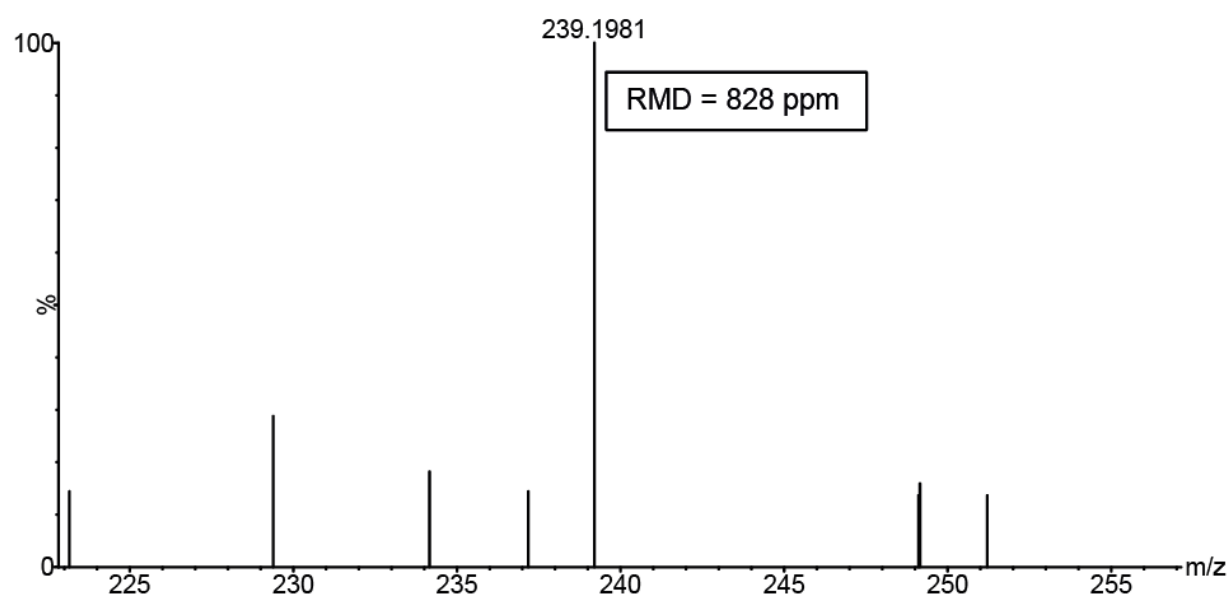
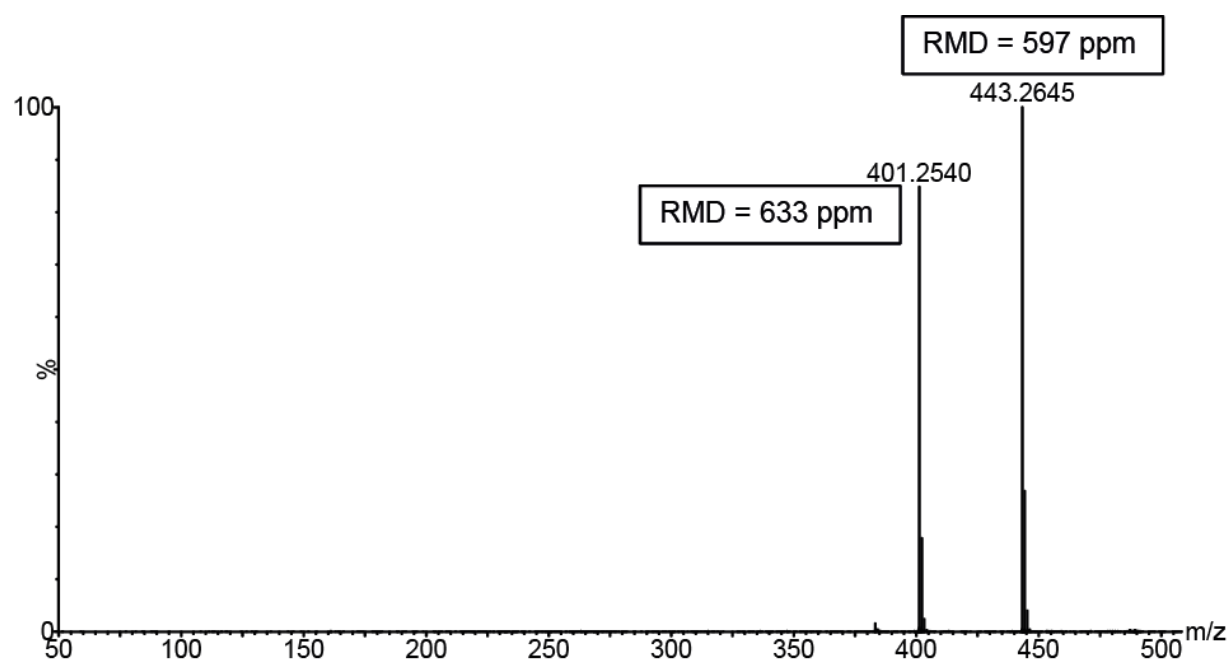


Figure 2.9. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  487) for campherenan-2,12-diol monoglucoside malonate ester. The lower spectrum displays the magnified region of  $m/z$  224-256, showing the sesquiterpenoid core fragment at  $m/z$  239.

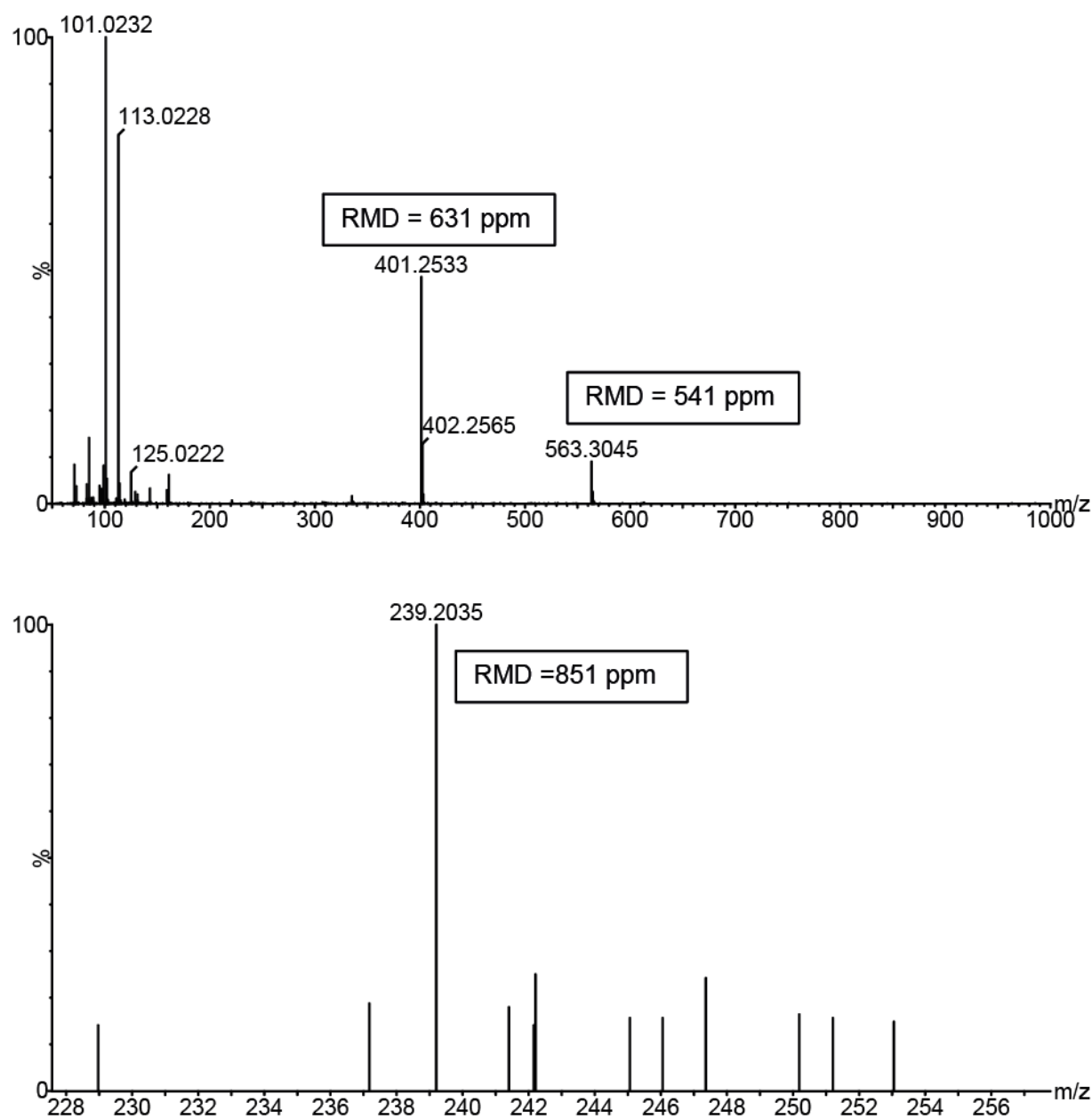


Figure 2.10. Negative ion mode product ion MS/MS spectrum of products from  $[M+\text{formate}]^-$  ( $m/z$  609) for campherenan-2,12-diol diglucoside. The lower spectrum displays the magnified region of  $m/z$  228-258, showing the sesquiterpenoid core fragment at  $m/z$  239.

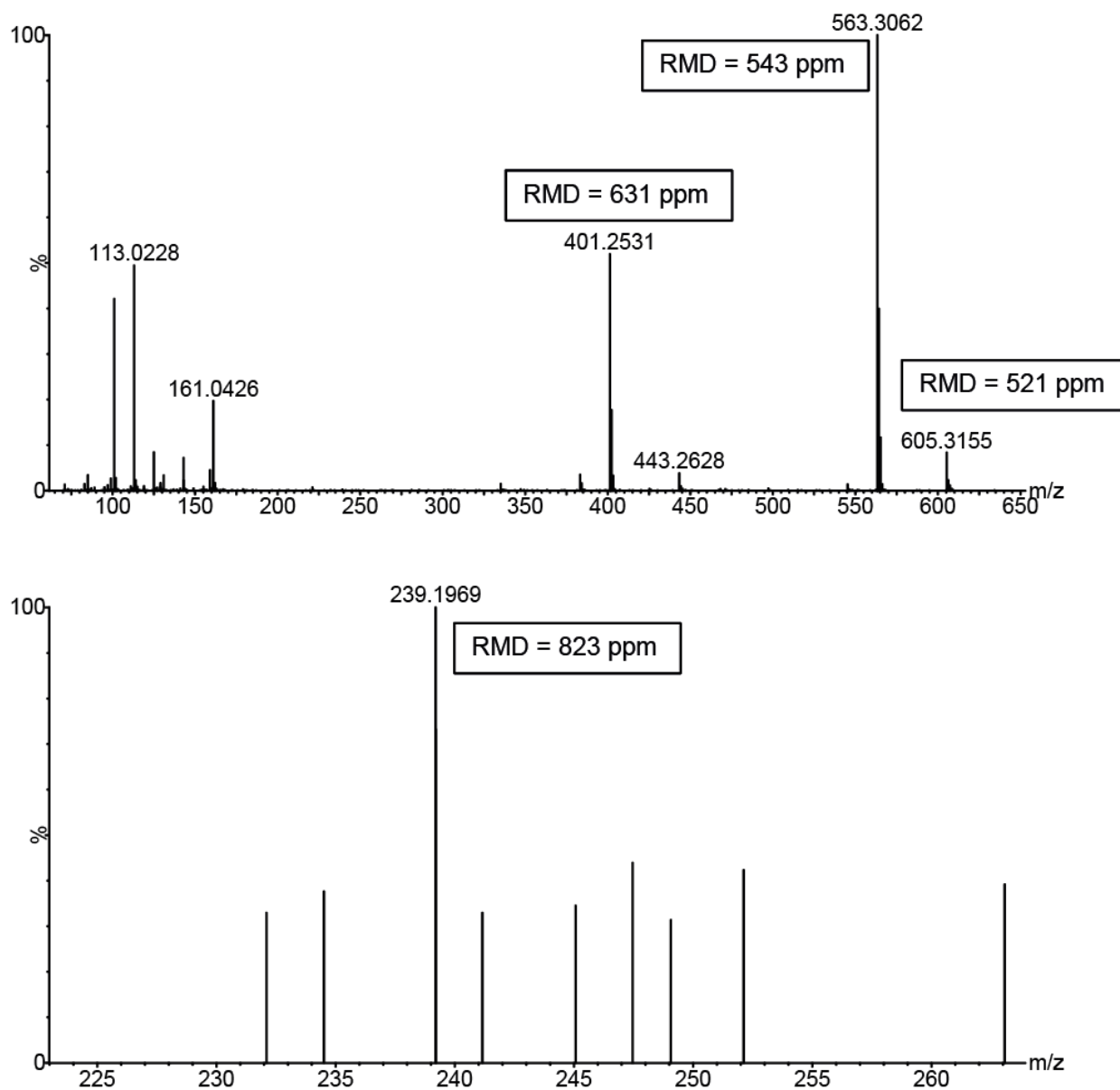


Figure 2.11. Negative ion mode product ion MS/MS spectrum of products from  $[M+\text{formate}]^-$  ( $m/z$  651) for campherenan-2,12-diol diglucoside acetate ester. The lower spectrum displays the magnified region of  $m/z$  224-264, showing the sesquiterpenoid core fragment at  $m/z$  239.

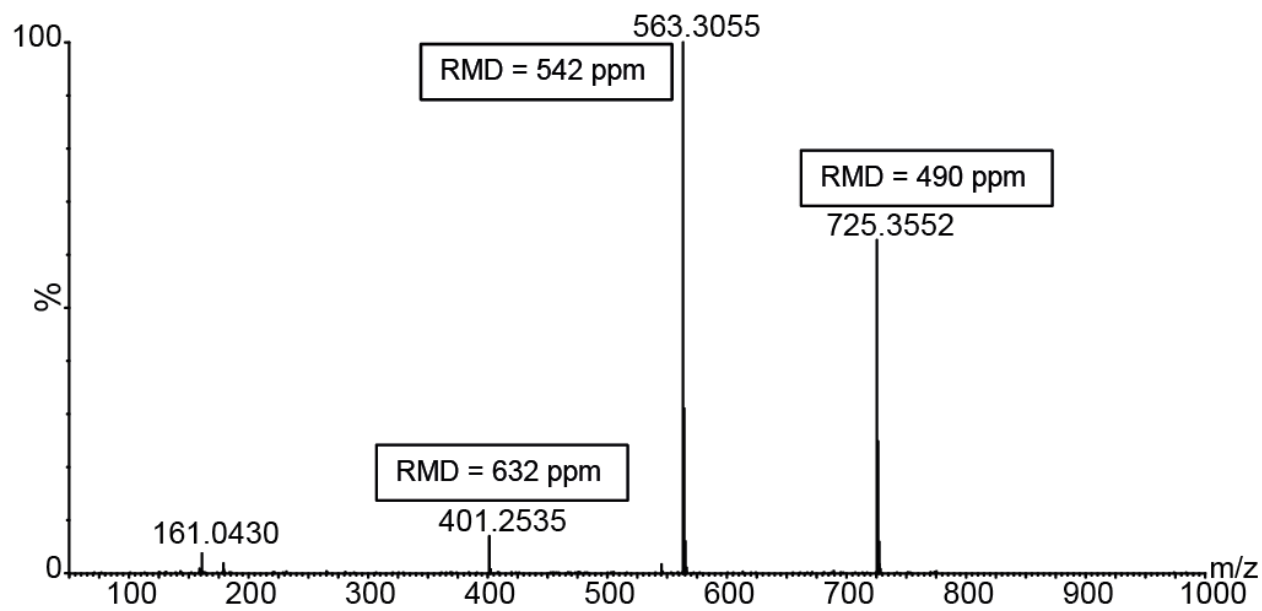


Figure 2.12. Negative ion mode product ion MS/MS spectrum of products from  $[M+\text{formate}]^-$  ( $m/z$  771) for a campherenane-2,12-diol triglycoside (Compound 13 in Table 2.2).

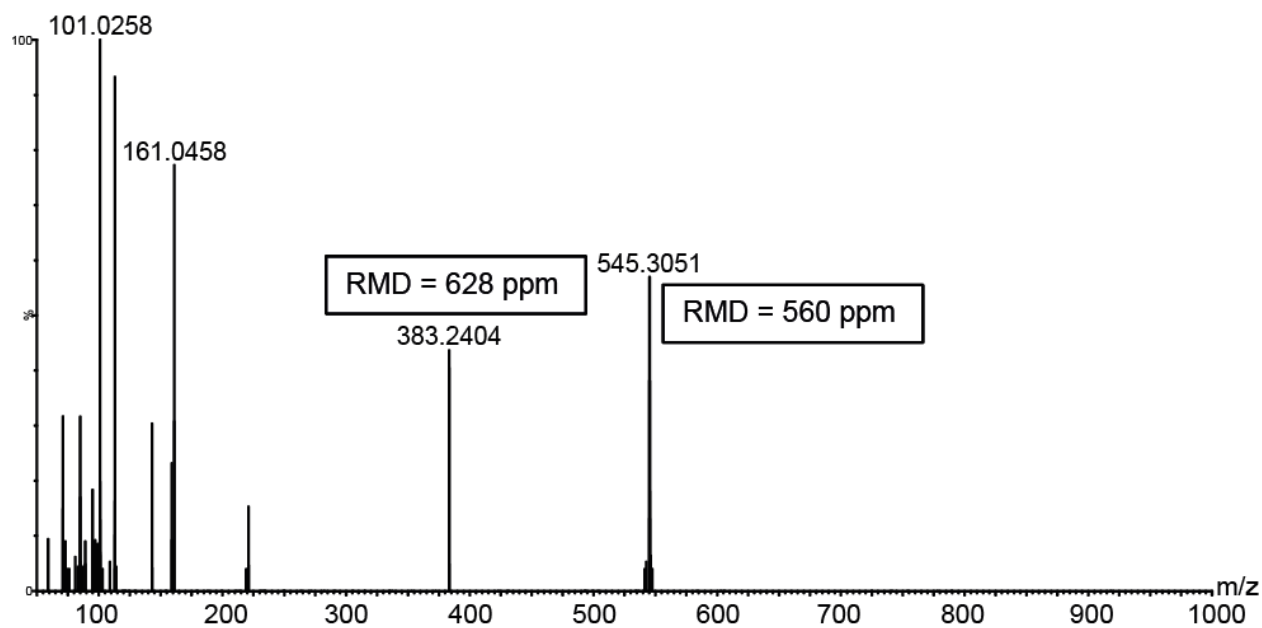


Figure 2.13. Negative ion mode product ion MS/MS spectrum of products from  $[M+\text{formate}]^-$  ( $m/z$  591) for a sesquiterpene II dihexoside (Compound 1 in Table 2.2).

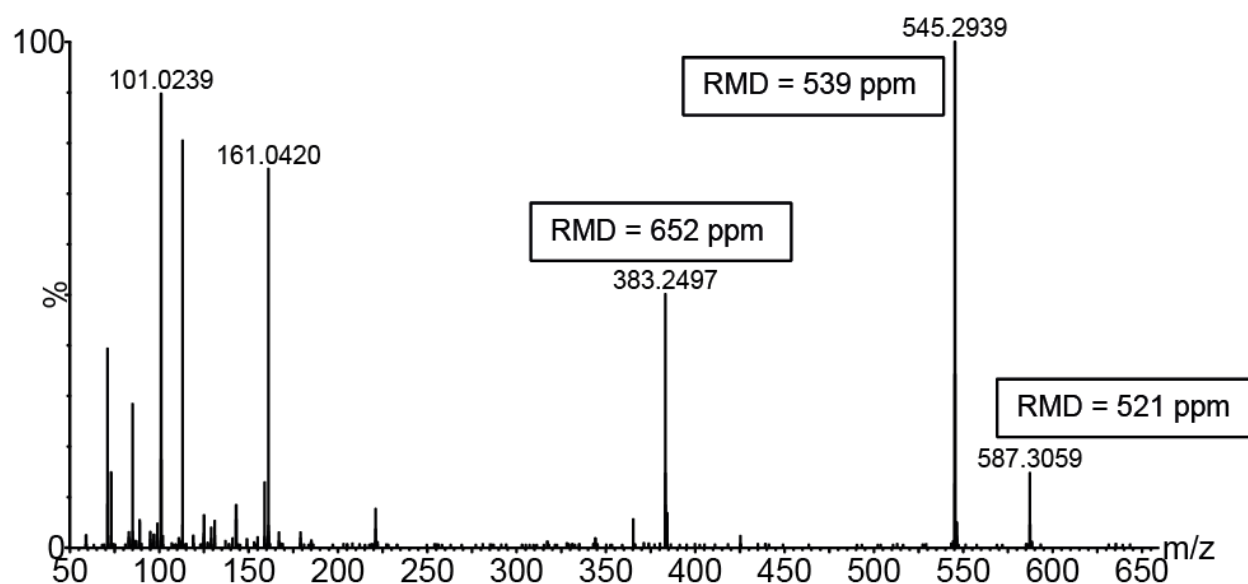


Figure 2.14. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  631) for a sesquiterpene II dihexoside malonate ester (Compound 3 in Table 2.2).

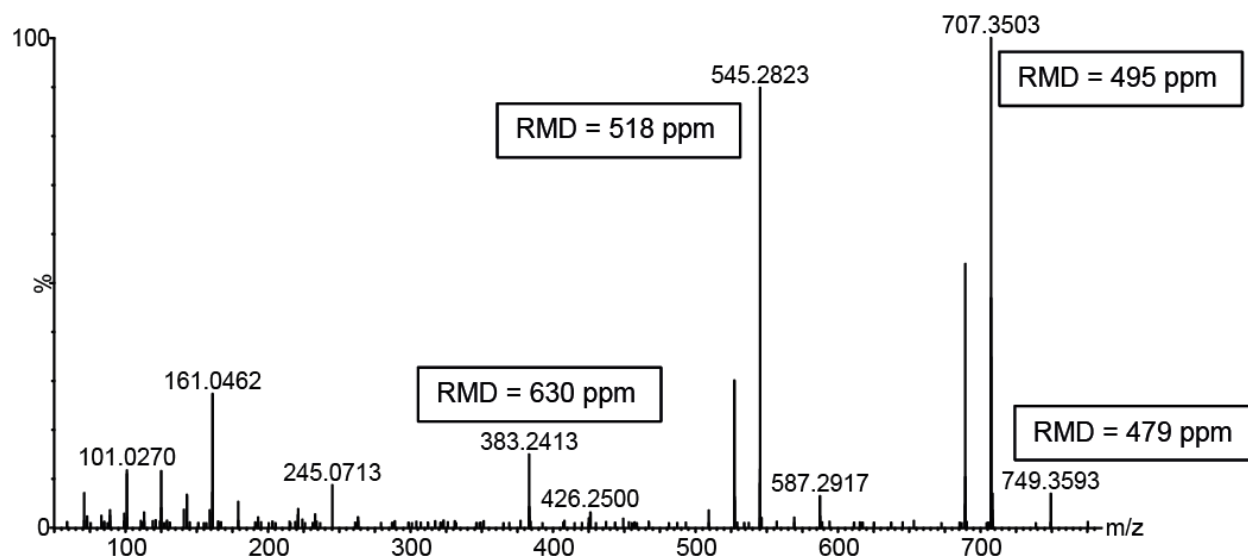


Figure 2.15. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  793) for a sesquiterpene II trihexoside malonate ester (Compound 5 in Table 2.2).

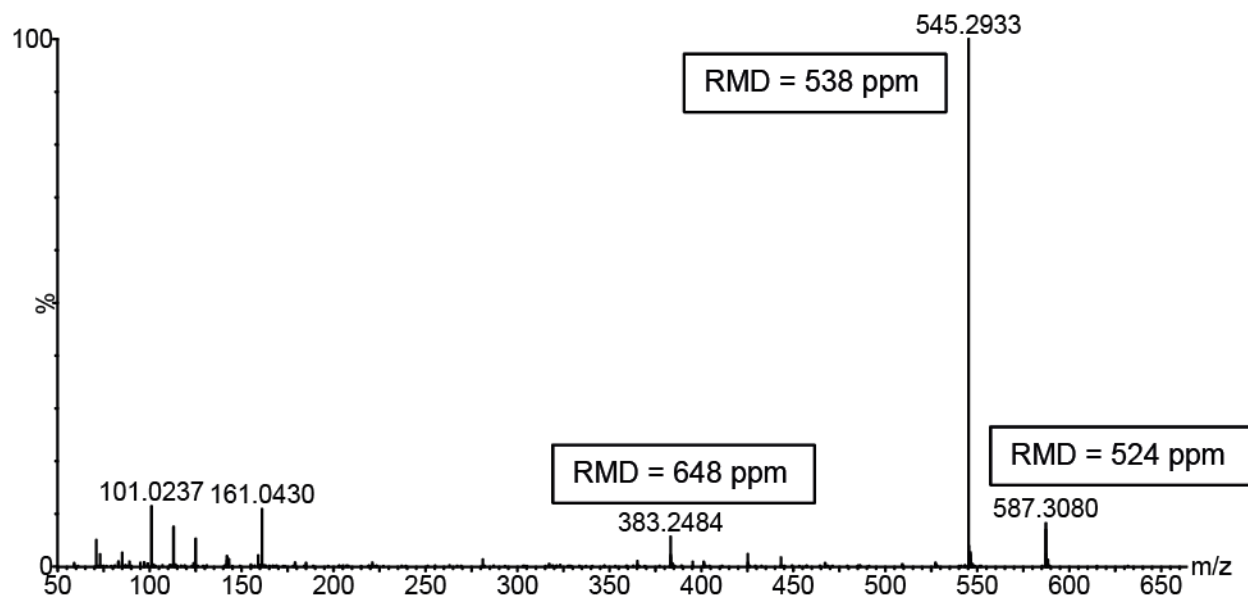


Figure 2.16. Negative ion mode product ion MS/MS spectrum of products from  $[M+HCOO]^-$  ( $m/z$  633) for a sesquiterpene II dihexoside acetate ester (Compound 2 in Table 2.2).

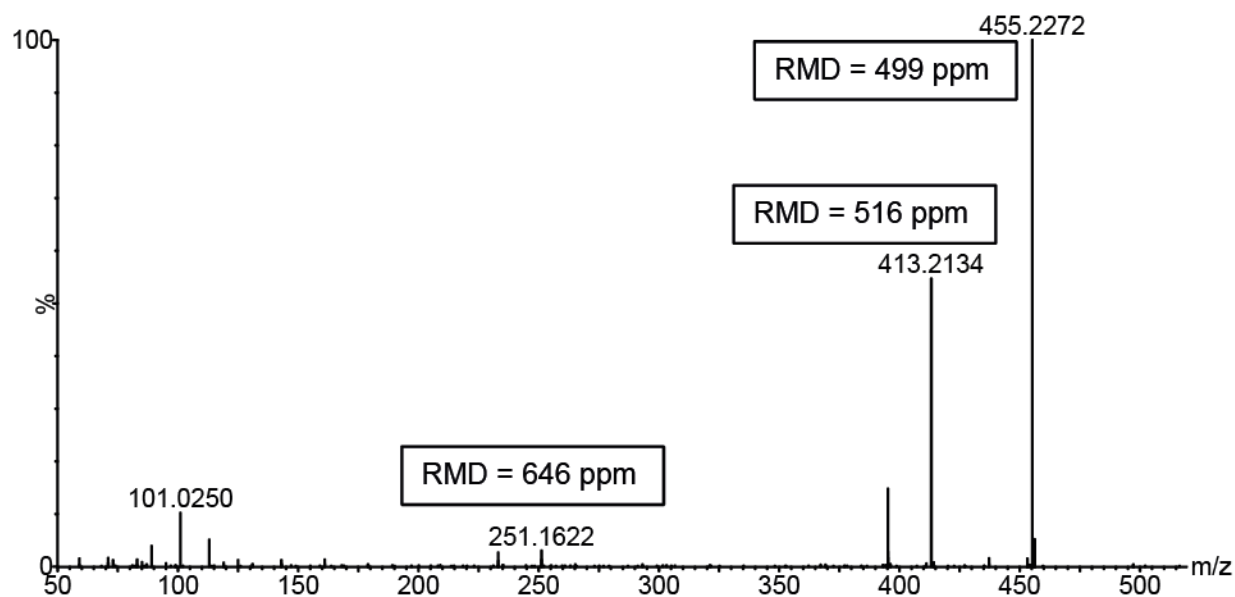


Figure 2.17. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  661) for a Sesquiterpene I diglycoside malonate ester (Compound 22 in Table 2.2).

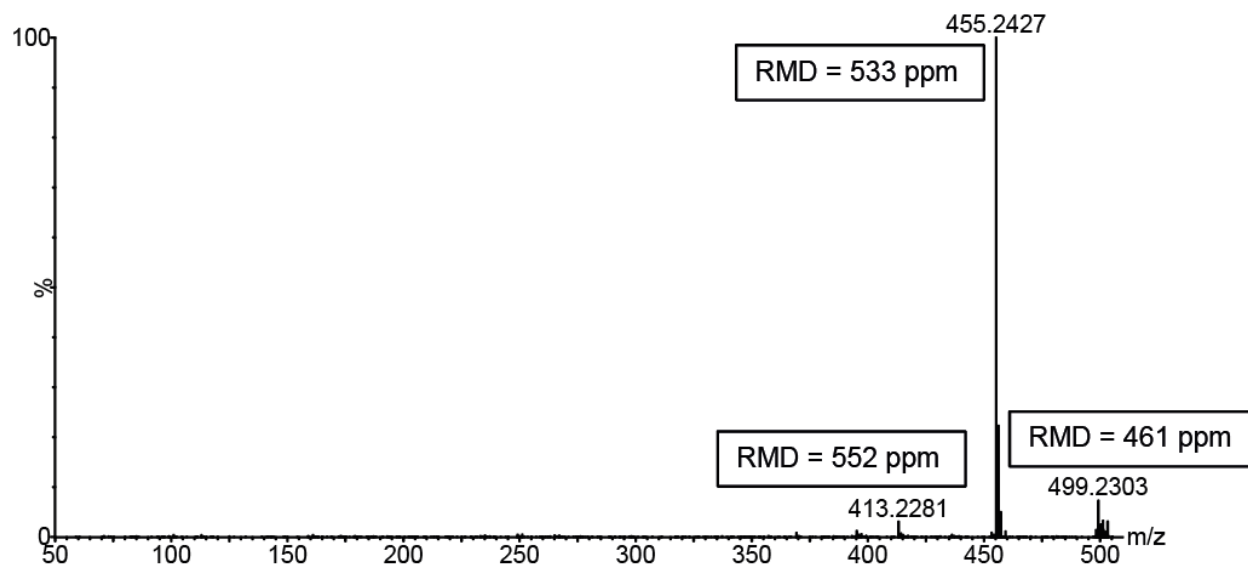


Figure 2.18. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  499) sesquiterpene I monoglycoside malonate ester (Compound 20 in Table 2.2).

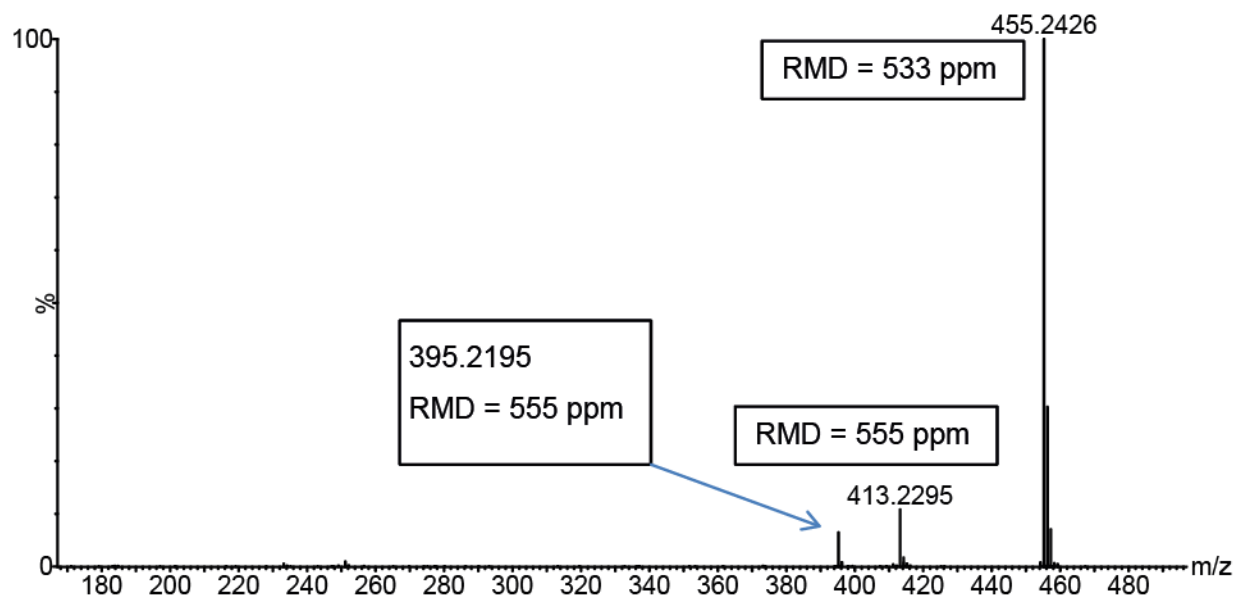


Figure 2.19. Negative ion mode product ion MS/MS spectrum of  $[M-H]^-$  ( $m/z$  455) for Sesquiterpene I monoglycoside acetate ester (Compound 19 in Table 2.2).

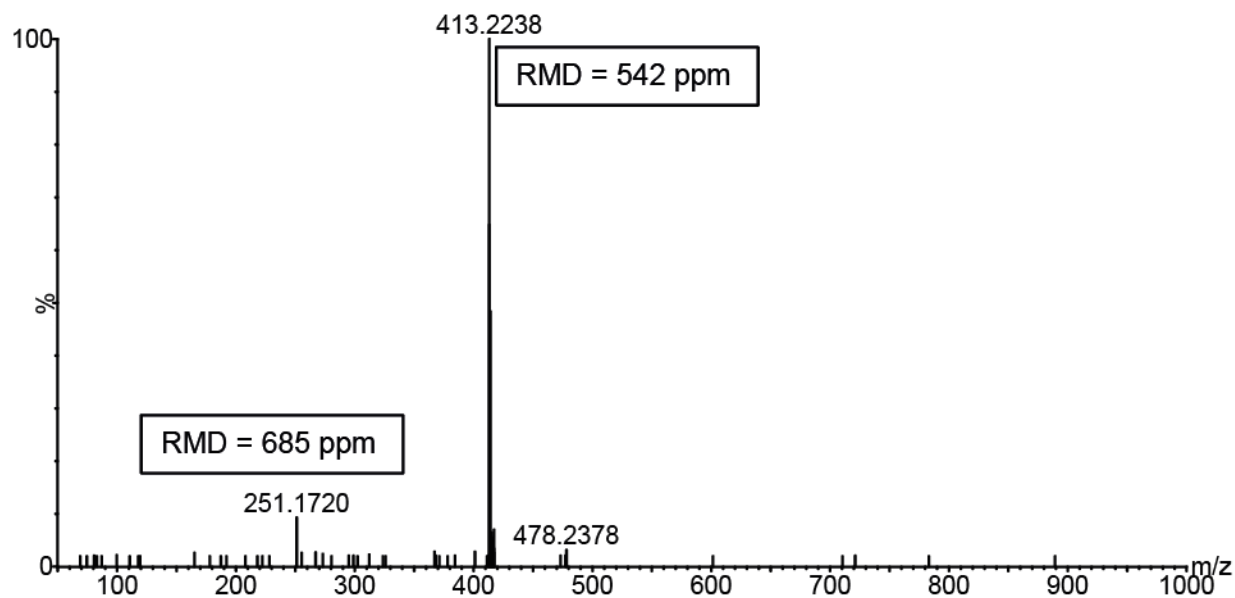


Figure 2.20. Negative ion mode product ion MS/MS spectrum of  $[M-H]^-$  ( $m/z$  413) of Sesquiterpene I monoglycoside (Compound 18 in Table 2.2).

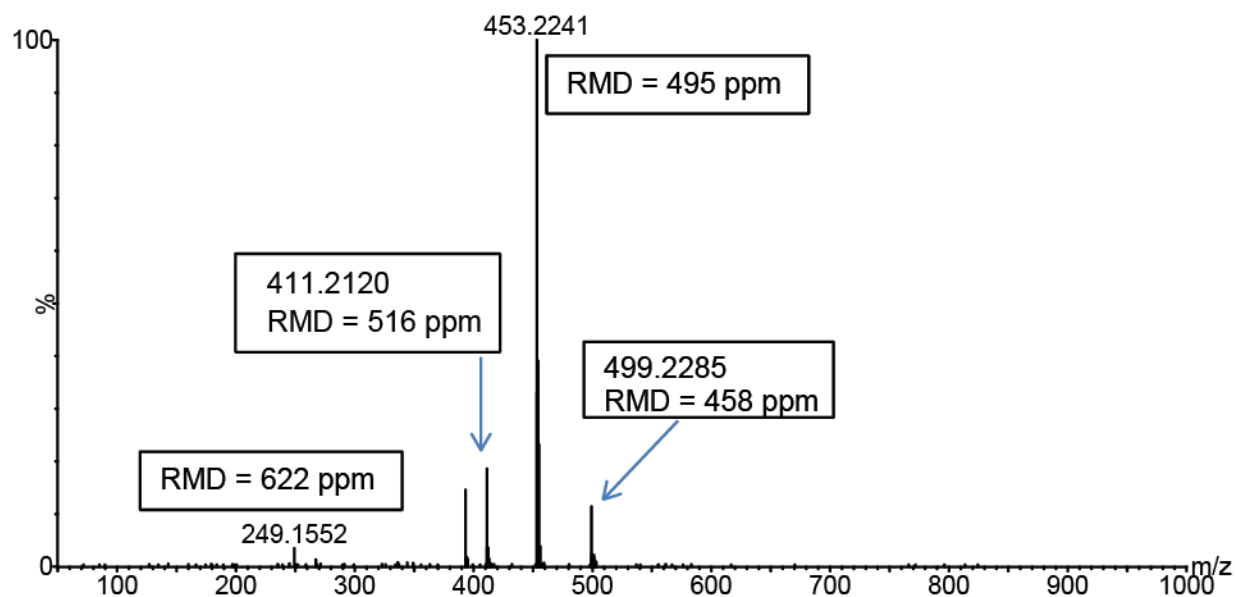


Figure 2.21. Negative ion mode product ion MS/MS spectrum of  $[M+HCOO]^-$  ( $m/z$  499) for sesquiterpene I monoglycoside malonate ester (Compound 20 in Table 2.2).



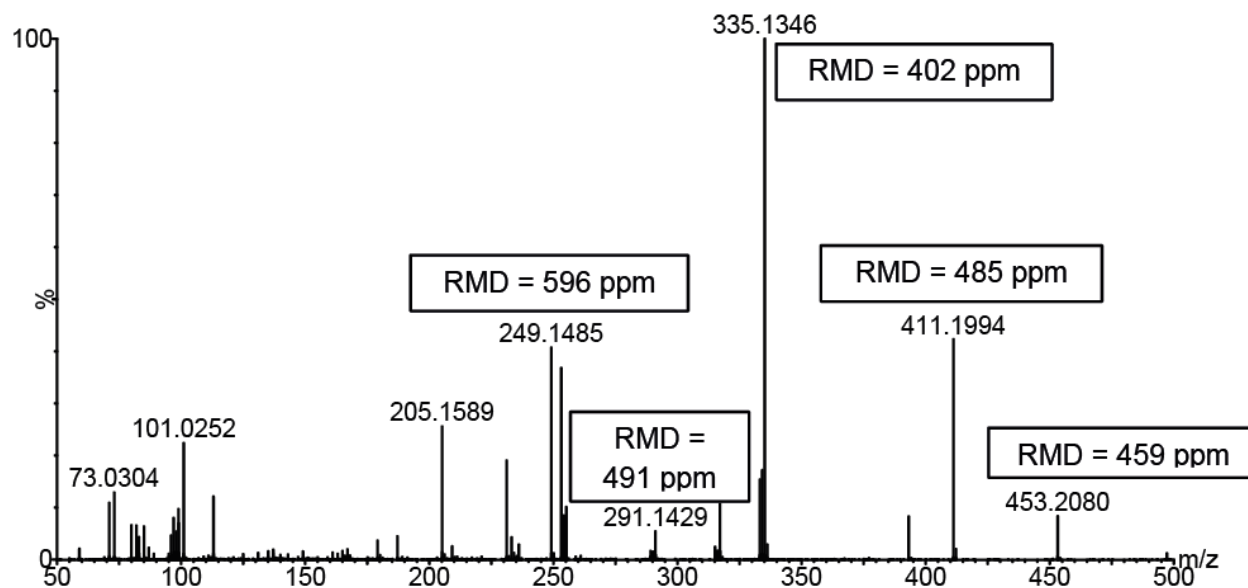


Figure 2.22. Negative ion mode product ion MS/MS spectrum of  $[M-H]^-$  ( $m/z$  497) of Sesquiterpene III monoglycoside malonate ester (Compound 17 in Table 2.2).

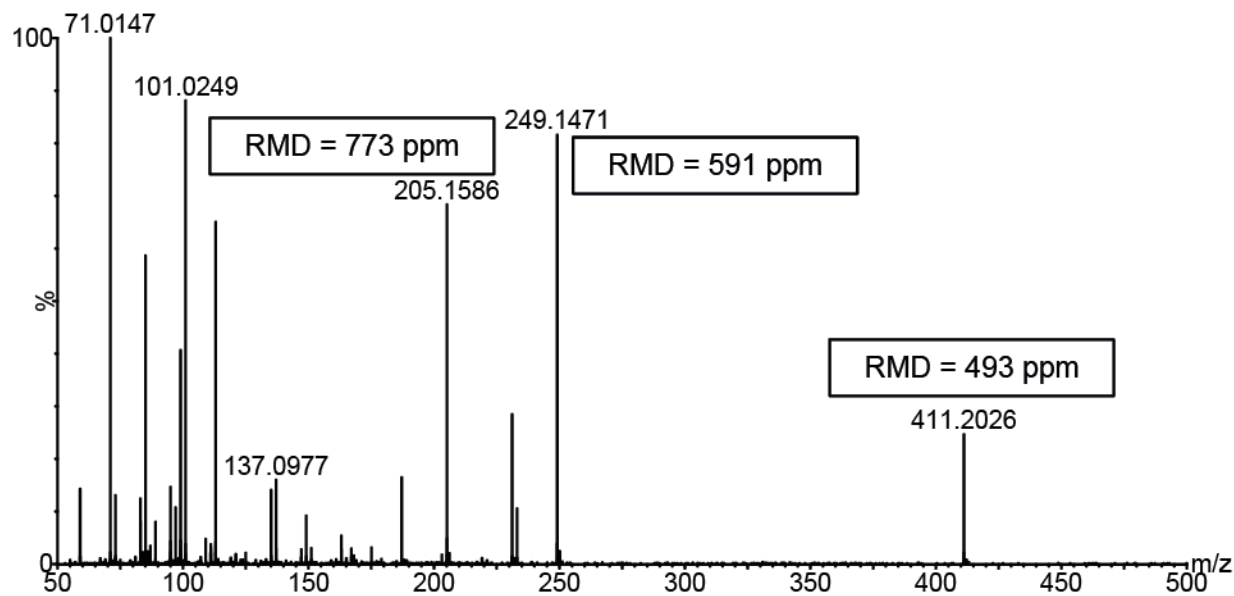


Figure 2.23. Negative ion mode product ion MS/MS spectrum of products of  $[M-H]^-$  ( $m/z$  411) of Sesquiterpene III monoglycoside (compound 15 in Table 2.2).

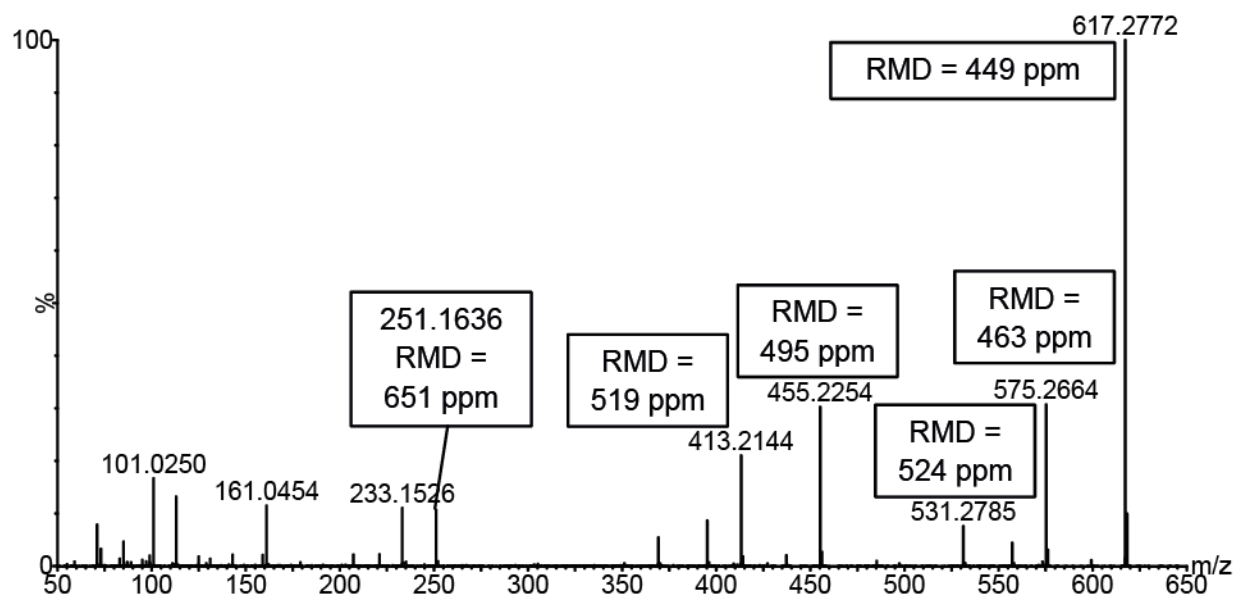


Figure 2.24. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  617) for Sesquiterpene I diglycoside acetate ester (compound 21). 27

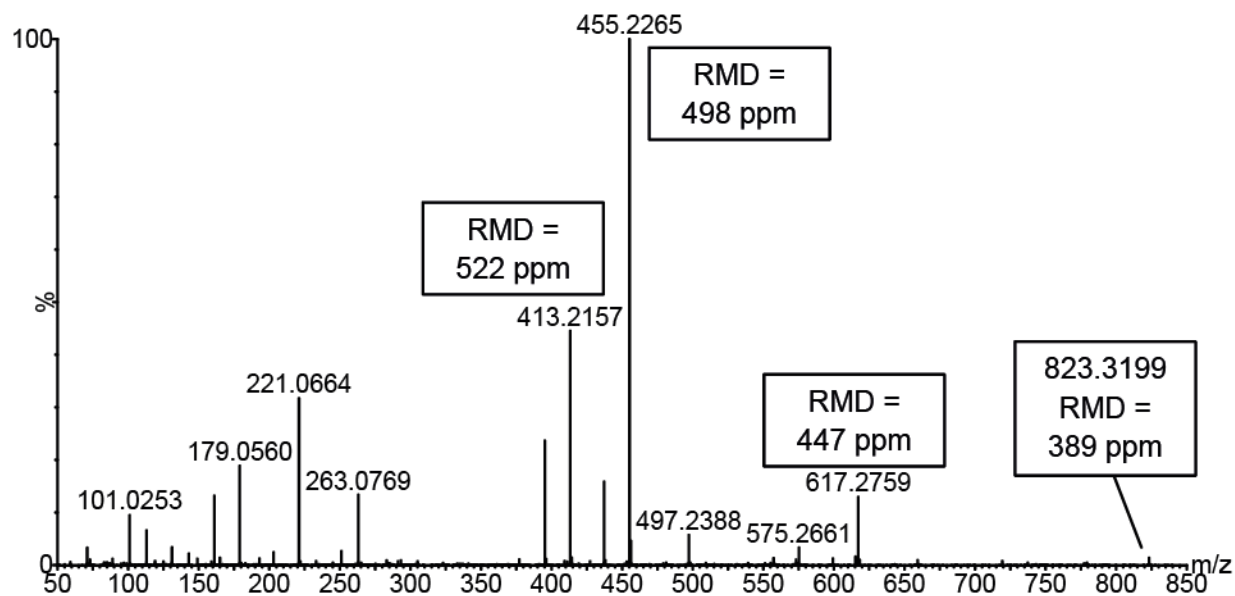


Figure 2.25. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  823) for Sesquiterpene I triglycoside malonate ester (compound 23 in Table 2.2).

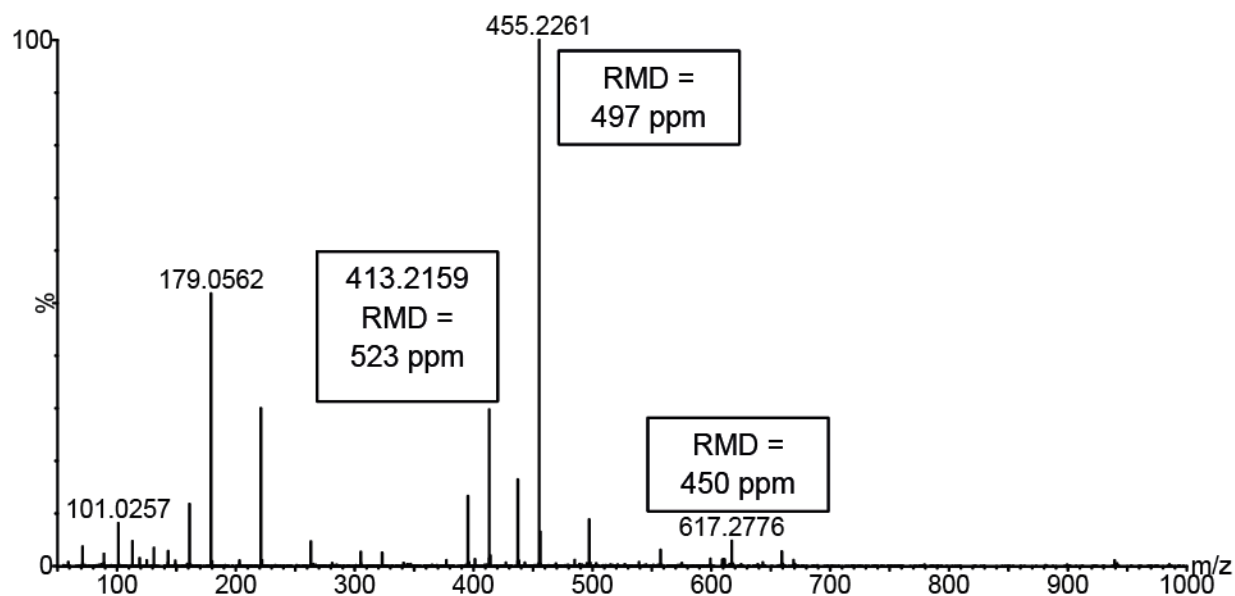


Figure 2.26. Negative ion mode product ion MS/MS spectrum of products from  $[M-H]^-$  ( $m/z$  985) for Sesquiterpene I tetraglycoside malonate ester (compound 24 in Table 2.2)

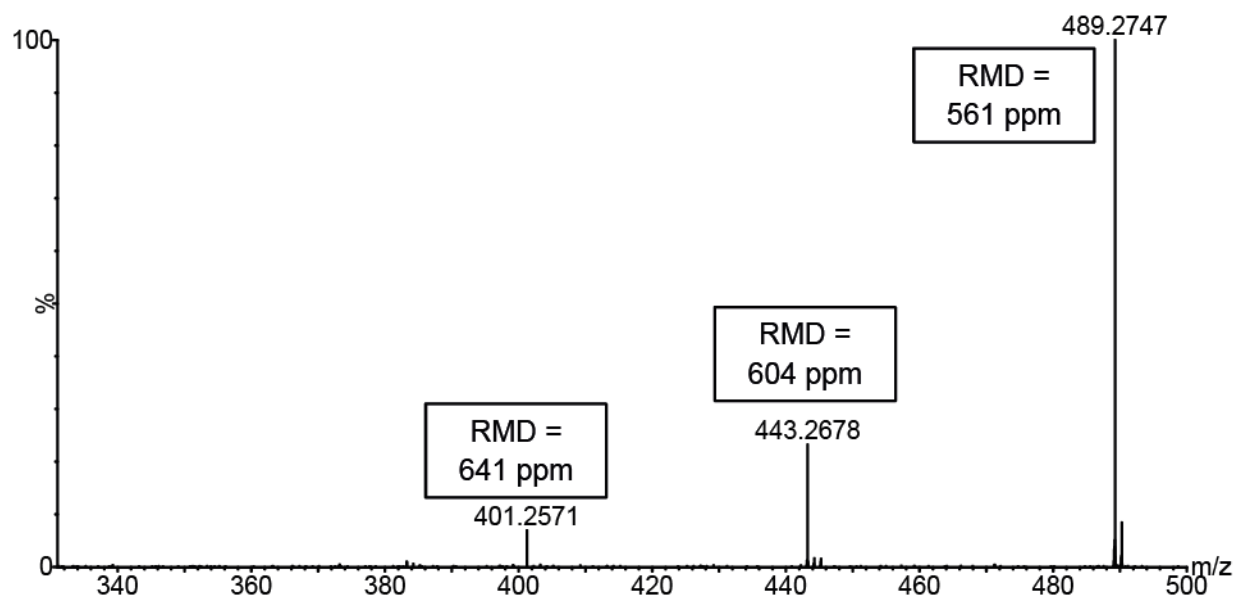


Figure 2.27. Negative ion mode product ion MS/MS spectrum of products from  $[M+HCOO]^-$  ( $m/z$  489) for Campherenane diol Monoglycoside acetate ester (compound 7 in Table 2.2).

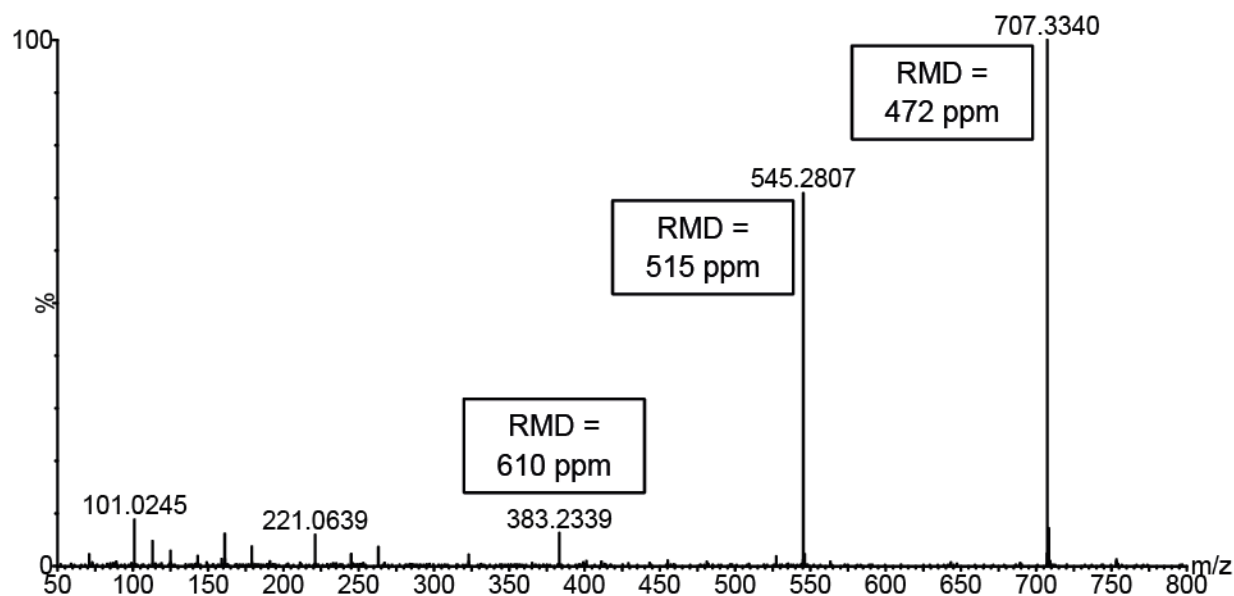


Figure 2.28. Negative ion mode product ion MS/MS spectrum of products from  $[M+HCOO]^-$  ( $m/z$  753) Sesquiterpene II Triglycoside (compound 4 in Table 2.2).

## 2.5 Conclusions

Analysis of the RMD variation among precursor ions and product/fragment ions generated information that accelerates the terpene glycoside discovery process by eliminating signals from the numerous compounds that do not give RMD values consistent with sesquiterpenoid glycosides. While the method may not exclusively represent terpenoids, it allows for the elimination of many non-terpenoid compounds from the complicated list of large number of metabolites detected in a given plant extract. The proposed method has allowed for the annotation of over 200 novel sesquiterpene glycosides from a plant system that has been studied for many decades. This is an indication that the method explained here will enable the rapid identification of terpene glycosides are largely understudied due to the lack of a systematic method for recognition of their presence.

Multiplexed CID MS/MS allows for the rapid examination of MS/MS spectra of compounds eliminating the requirement for separate MS/MS experiments for all markers detected using automates peak picking.

The research discussed here used the negative ion mode MS/MS for the annotation of terpene glycosides from complex matrices, and provides the first evidence for an extensive and diverse group of sesquiterpenoid glycosides in *S. habrochaites* LA1777. However, the proposed method is perceived as equally applicable to positive ion mode data sets as well. This is because the relative mass defect is an intrinsic property of the compound being considered and is independent of ionization mode being used for the detection.

## REFERENCES

## REFERENCES

1. T.J. Maimone, P.S. Baran, Modern synthetic efforts toward biologically active terpenes, *Nat Chem Biol* 3 (2007) 396-407.
2. D. Tholl, Terpene synthases and the regulation, diversity and biological roles of terpene metabolism, *Current Opinion in Plant Biology* 9 (2006) 297-304.
3. Z. Lei, D.V. Huhman, L.W. Sumner, Mass spectrometry strategies in metabolomics, *J Biol Chem* 286 (2011) 25435-25442.
4. M.C. Stagliano, J.G. DeKeyser, C.J. Omiecinski, A.D. Jones, Bioassay-directed fractionation for discovery of bioactive neutral lipids guided by relative mass defect filtering and multiplexed collision-induced dissociation, *Rapid Commun Mass Spectrom* 24 (2010) 3578-3584.
5. M.L. Toumi, H. Desaire, Improving mass defect filters for human proteins, *J Proteome Res* 9 (2010) 5492-5495.
6. S. Heiling, M.C. Schuman, M. Schoettner, P. Mukerjee, B. Berger, B. Schneider, A.R. Jassbi, I.T. Baldwin, Jasmonate and ppHsystemin regulate key Malonylation steps in the biosynthesis of 17-Hydroxygeranyllinalool Diterpene Glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*, *Plant Cell* 22 (2010) 273-292.
7. K. Levsen, H.M. Schiebel, J.K. Terlouw, K.J. Jobst, M. Elend, A. Preiss, H. Thiele, A. Ingendoh, Even-electron ions: a systematic study of the neutral species lost in the dissociation of quasi-molecular ions, *J Mass Spectrom* 42 (2007) 1024-1044.
8. V.A. Marinos, M.E. Tate, P.J. Williams, Protocol for FAB MS/MS Characterization of Terpene Disaccharides of Wine, *Journal of Agricultural and Food Chemistry* 42 (1994) 2486-2492.
9. P. Kachlicki, J. Einhorn, D. Muth, L. Kerhoas, M. Stobiecki, Evaluation of glycosylation and malonylation patterns in flavonoid glycosides during LC/MS/MS metabolite profiling, *J Mass Spectrom* 43 (2008) 572-586.
10. C.-C. Ruan, Z. Liu, X. Li, X. Liu, L.-J. Wang, H.-Y. Pan, Y.-N. Zheng, G.-Z. Sun, Y.-S. Zhang, L.-X. Zhang, Isolation and Characterization of a New Ginsenoside from the Fresh Root of *Panax Ginseng*, *Molecules* 15 (2010) 2319-2325.
11. S. Guangzhi, L. Zhi, L. Xianggao, Z. Yinan, W. Jiyan, Isolation and Identification of Two Malonyl-ginsenosides from the Fresh Root of *Panax Ginseng*, *Chinese Journal of Analytical Chemistry* 33 (2005) 1783-1786.

12. L.R. Sun, J. Yan, L. Zhou, Z.R. Li, M.H. Qiu, Two New Triterpene Glycosides with Monomethyl Malonate Groups from the Rhizome of *Cimifuga foetida* L, *Molecules* 16 (2011) 5701-5708
13. C.B. Rogers, Pentacyclic triterpenoid rhamnosides from *Combretum imberbe* leaves, *Phytochemistry* 27 (1988) 3217-3220
14. M.G. El-Ghazouly, N.A. El-Sebakhy, A.A.S. El-Din, C. Zdero, F. Bohlmann, Sesquiterpene xylosides from *Iphiona scabra*, *Phytochemistry* 26 (1987) 439-443.
15. H. Pfander, H. Stoll, Terpenoid glycosides, *Nat Prod Rep* 8 (1991) 69-95.
16. B.M. Fraga, Natural sesquiterpenoids, *Natural Product Reports* 30 (2013) 1226-1264.



## **Chapter 3: Purification and structure elucidation of campherenane diol glycosides from leaf glandular trichomes of *Solanum habrochaites* LA1777**

### **3.1 Introduction**

A systematic investigation of methanol leaf dip extracts of *Solanum habrochaites* LA1777 was performed in order to purify and elucidate the structures of sesquiterpenoid specialized metabolites that were annotated using relative mass defect filtering. This chapter discusses the purification and structure elucidation of the campherenane diol derivatives introduced in chapter two. The research described in this Chapter has already been published in the journal *Phytochemistry* [1]

Among the sesquiterpenes reported in the literature, campherenanes are rare[2], but bear some structural similarity to santalenes(Fig. 3.1 a. and b.) [3]. Campherenane derivatives have been isolated from *Santalum album* [4], *Santalum austrocaledonicum* [5] and *Illicium tsangii* [6] but not from any plants of the genus *Solanum* (tomato, potato). The biosynthetic pathway of santalenes in *Solanum habrochaites* LA1777 leaf trichomes have been proposed by Sallaud *et al* [7]. Furthermore, Sallaud *et al.* has proposed that depending on the stereochemistry of the precursor and the position of ring closure different types of sesquiterpenes can result. For instance, another sesquiterpene -bergamotene – also is synthesized starting with the same *Z,Z*-farnesyl diphosphate. Therefore, it is likely that the precursors of campherenane also are either *Z,Z*-FPP or *E,E*-FPP however, a group of enzymes that convert these precursors specifically into campherenane compounds must be present in these plants.

Glycosylated terpenes are seldom reported from genus *Solanum*. Some acyclic diterpene glycosides have been reported from tobacco, a member of genus *Solanum* [8]. Yet, glycosylated

sesquiterpenes are not known from this genus. Glycosylation results in changes of physical properties of terpenoids such as volatility[9]. Therefore, much interest has been placed on developing terpene glycosides as slow releasing aroma and flavor compounds[10,9]. This has resulted in development of methods to glycosylate terpenoids by chemical or enzymatic means [11]

The abundance of genetic information about tomato (*S. lycopersicum*) [12] combined with the availability of introgression lines derived from *S. habrochaites* and *S. lycopersicum* [13] facilitates discovery of sesquiterpenoid metabolism genes from metabolomic and transcriptomic data. An introgression line is an offspring that is genetically closely similar to parent plant with only a few genetic differences. Therefore, by comparing the chemical constituents of parent plant and the offspring plant it is possible to assign chemical traits to genes.

## **3.2 Experimental methods**

### **3.2.1 Plant material**

Five *Solanum habrochaites* LA1777 plants were grown using seeds obtained from Tomato Genetics Resource Center of University California, Davis, USA, in a plant growth chamber (28 °C, 16 hour day time, 96% humidity; 15  $\mu\text{mol m}^{-2} \text{s}^{-1}$  irradiance), and leaf and stem tissues were harvested and extracted in six-week intervals. Final mass of fresh plant material extracted was about 5 kg. Extraction was performed by dipping each branch in 1 L of 100% methanol for about 30 s. This approach aims for selective extraction of glandular trichomes. Extracts were subsequently combined, and solvent was evaporated under vacuum on a rotary evaporator without heating, and the residue was redissolved in about 10 ml of 100% methanol. This extract

was divided into approximately three equal portions, and a single compound was isolated from each portion.

### **3.2.2 Purification of campherenane diol glycosides**

Semipreparative HPLC was performed on an Acclaim 120 C18 column ( $4.6 \times 150$  mm, 5  $\mu$ m particles; Dionex Co. USA) using a Waters semipreparative HPLC system consisting of two Waters 510 pumps controlled using a Waters 610 gradient controller. The system was equipped with a Waters 717 plus autosampler and fractions were collected using a LKB BROMMA 221 Superrac fraction collector. For the semipreparative HPLC, flow rate was 2.0 ml/min and the injection volume for each injection was 150  $\mu$ L. The linear solvent gradient used for the purification employed water (solvent A) and methanol (solvent B) was as follows (A:B): 0-1.0 min (99%:1%), 4.00 min (60%:40%), 4.50 min (55%:45%), and held at this composition until 9.00 min, 9.50 min (50%:50%), and held until 14.00 min, 14.50 min (0%:100%), and held until 17.00 min, then returned to initial conditions at 17.01 min and held until 25.00 min (99%:1%). Forty fractions were collected (30 s intervals for the first 20 min), and each fraction was tested for the presence of compound of interest using a 5.50 min UHPLC-MS method in negative ion mode ESI using the same instrument configuration used for profiling. This gradient was (A = 0.15% aqueous formic acid; B = methanol (A:B): 0.0-2.0 min (95%:5%), step at 2.01 min (40%:60%), linear increase to 3.00 min (0%: 100%), and held at this condition until 5.0 min, then returned to initial condition at 5.01 min and held until 5.5 min. From these analyses, the fractions that were over 90% pure based on total ion chromatogram peak areas were identified and combined, and solvent was removed by evaporating to dryness at room temperature under a N<sub>2</sub> gas flow.

### 3.2.3 NMR Experiments

All NMR experiments were performed in a Bruker Avance 900 spectrometer, equipped with TCI triple-resonance inverse detection CryoProbe, at 900 MHz ( $^1\text{H}$  frequency) at 25 °C. NMR experiments were performed using  $\text{CD}_3\text{OD}$  matched NMR tubes (Shigemi Inc., Allison Park, PA USA). Once dried, each purified compound was redissolved in 99.8%  $\text{CD}_3\text{OD}$  (Sigma-Aldrich, St. Louis, MO, USA) and analyzed using NMR. All spectra were calibrated using residual non-deuterated solvent signal as  $\delta^1\text{H}$  ( $\text{CD}_3\text{OD}$ ) = 3.34 ppm and  $\delta^{13}\text{C}$  ( $\text{CD}_3\text{OD}$ ) = 49.08 ppm. Proton,  $^{13}\text{C}$ , gHSQC, gHMBC, DQF-COSY, TOCSY and NOESY spectra were acquired.

### 3.2.4 Acid hydrolysis, derivatization and GC-MS experiments

Acid hydrolysis was performed to identify the sugars attached to the terpenoid core. A 200  $\mu\text{L}$  aliquot of a 30-ml semipreparative HPLC fraction containing the purified compound (peak 6a in Fig. 3.1) was mixed with 200  $\mu\text{L}$  of 2.0 M sulfuric acid and incubated in a sealed vessel at 25 °C for 6 hrs. Next, 1.0 M  $\text{BaCl}_2$  was added dropwise until no further precipitation of  $\text{BaSO}_4$  was observed. The mixture was centrifuged (10,000xg) for 5 min, and the supernatant was evaporated to dryness on a SpeedVac. To the dried residue, 150  $\mu\text{L}$  of 20 mg/mL methoxyamine hydrochloride/pyridine was added and incubated at 25 °C for 2 hrs. Next, 150  $\mu\text{L}$  of MSTFA (N-methyl-N-trimethylsilyltrifluoroacetamide) was added to the mixture and incubated for another hour. Sugar standards (glucose, xylose, galactose, rhamnose purchased from Sigma-Aldrich, St. Louis, MO, USA) were derivatized using the same procedure, and these were analyzed using an Agilent 5975 mass spectrometer with 6890N GC and autosampler. The column was an Agilent DB-5 (10 m x 0.18 mm, 0.34  $\mu\text{m}$  film) using a temperature program as follows; Initial temperature 40 °C, hold for 1.00 min; increase at 30 °C/min to 90 °C, then 5 °C/min to 110 °C, 40 °C/min to 165 °C, 5 °C/min up to 180 °C, and 40 °C/min to 320 °C, and maintained at 320 °C for

one minute. Ionization employed 70 eV electron ionization. The same procedure was used to verify the identity of carbohydrates in fractions 6b and 8c.

### 3.3 Results and discussion

#### 3.3.1 Profiles of the campherenane diol diglycosides

The metabolite profiling performed in the experiments discussed in Chapter 2 allowed for the initial annotation of campherenane diol diglycosides from *Solanum habrochaites* LA1777 leaf extract. Negative ion mode UHPLC-MS profiling of leaf dip metabolites from wild tomato species *S. habrochaites* LA1777 showed more than 50 chromatographic peaks not evident in analyses of extracts from other *Solanum* species. Multiplexed non-selective collision-induced dissociation (CID) time of flight (TOF) mass spectrometry enabled the rapid annotation of these molecules based on accurate pseudomolecular and co-eluting fragment ion masses, as well as the collision energy-dependence of fragment ion abundances. Assignments of fragment ions in multiplexed nonselective CID spectra may be subject to artifactual false positives when metabolites co-elute, but in this study the origins of characteristic fragment ions were confirmed using MS/MS experiments that employed selection of specific precursor ions before CID.

Chapter 2 described the annotation of campherenane diol diglycoside detected in negative-ion mode LC/MS as  $m/z$  609 and campherenane diol triglycoside malonate ester detected as  $m/z$  811. At least 10 different isomers were identified for  $m/z$  609 from the leaf dip extract of LA1777 (Fig. 3.1c). Peak 8c (Fig. 3.1) was purified using HPLC and the structure was elucidated using NMR. Similarly, at least 8 isomers were identified for  $m/z$  811 from the same extract (Fig. 3.1e). The annotation of the other campherenane diol derivatives is described below.

One of the metabolite peaks ( $R_t = 9.0$  min; peak 6a in Fig. 3.1) showed  $[M-H]^-$  at  $m/z$  649 (experimental  $m/z$  649.3089; theoretical  $m/z$  for  $C_{30}H_{49}O_{15}^- = 649.3077$ ;  $\Delta m = +1.8$  ppm), and MS/MS revealed fragment ions corresponding to facile loss of  $CO_2$  (44 Da) and loss of an additional 42 Da ( $C_2H_2O$ ). These neutral losses suggested presence of a malonate ester. Additional neutral loss of 162.05 Da suggested the presence of a hexose group, and accurate mass measurements of the remaining fragment at  $m/z$  401 (experimental  $m/z$  401.2536; theoretical  $m/z$  for  $C_{21}H_{37}O_7^- = 401.2545$ ;  $\Delta m = -2.2$  ppm) yielded a relative mass defect of 640 ppm which, when combined with the observed resistance to further fragmentation, were consistent with fractional hydrogen content of a glycosylated sesquiterpenoid core. Though initial inspection of nonselective CID mass spectra suggested loss of only a single neutral carbohydrate fragment, product ion MS/MS spectra for  $m/z$  649 at higher collision energies yielded a fragment at  $m/z$  239 (experimental  $m/z$  239.2006; theoretical for  $C_{15}H_{27}O_2^- = 239.2017$ ;  $\Delta m = -4.6$  ppm) consistent with an additional loss of 162 Da, and led to annotation as a malonate ester of a sesquiterpene diol dihexoside. At least 9 additional metabolites were observed in the UHPLC-MS/MS profile that shared the same molecular mass and common fragment ions in the CID mass spectra, suggesting a diverse group of isomeric sesquiterpenoid dihexoside malonate esters.

Another peak in the total ion chromatogram (peak 6b in the extracted ion chromatogram in Figure 3.1b) exhibited a base peak ion at  $m/z$  487 (experimental  $m/z$  487.2539; theoretical  $m/z$  for  $C_{24}H_{39}O_{10}^- = 487.2549$ ;  $\Delta m = -2.0$  ppm) that was 162 Da less than 649. This molecule also underwent collision-induced dissociation to give  $m/z$  401 and 239 fragments. Since the molecular mass was 162 Da less than the malonate ester of sesquiterpene diol dihexoside, it was concluded

to be a malonate ester of a sesquiterpene diol monohexoside. At least six isomers of this molecule were detected in UHPLC-MS/MS profiles from *S. habrochaites* LA1777 (Fig. 3.1b).

Another metabolite (peak 5d in Fig. 3.1d) was observed in the UHPLC-MS analyses at  $m/z$  651 (experimental  $m/z$  651.3233; theoretical  $m/z$  for  $C_{30}H_{51}O_{15}^-$  = 651.3210;  $\Delta m$  = +3.5 ppm). The MS/MS spectrum yielded  $[M-H]^-$  at  $m/z$  605 (experimental  $m/z$  of  $[M-H]^-$  = 605.3164; theoretical  $m/z$  605.3179;  $\Delta m$  = -2.5 ppm), which was 44 Da lower in mass than the sesquiterpene diol dihexoside malonate (peak 6a in Fig. 3.1), and 42 Da greater than the sesquiterpene diol dihexoside, consistent with an acetate ester in place of the malonate ester. Therefore, this metabolite was annotated as an acetate ester of a sesquiterpene diol dihexoside. At least 10 chromatographically-resolved isomers of this could be annotated from UHPLC-MS/MS profiles of *S. habrochaites* LA1777 (Fig. 3.1d) The MS/MS spectra of all isomers showed the loss of the acetyl group (neutral loss of 42 Da) followed by the sequential losses of two anhydrohexoses (162 Da) to yield  $m/z$  401 and aglycone core of  $m/z$  239.

A non-malonylated form of sesquiterpene diol trihexoside described in Chapter 2 also was detected (Fig. 3.1f, peak 6f) as  $m/z$  771 (experimental  $m/z$  771.3621; theoretical  $m/z$  for  $C_{34}H_{59}O_{19}^-$  = 771.3656;  $\Delta m$  = -4.5 ppm; annotated as  $[M+formate]^-$ ). The MS/MS product ion spectrum exhibited  $[M-H]^-$  at  $m/z$  725 and fragments at  $m/z$  563, 401 and 239 that corresponded to three sequential neutral losses of 162 Da indicating the presence of the three hexoses. Eight isomers of this sesquiterpene diol trihexoside were evident in UHPLC-MS profiles (Fig. 3.1f) and yielded the same fragment ion masses.

Further exploration of the UHPLC-MS profile yielded recognition of a set of metabolites yielding base peaks at  $m/z$  447 (experimental  $m/z$  447.2585 for peak 2g in Figure 3.1g;

theoretical  $m/z$  for  $C_{22}H_{37}O_9^- = 447.2600$ ;  $\Delta m = -3.3$  ppm), consistent with  $[M+\text{formate}]^-$  of the sesquiterpenoid monoglycoside described above (peak 6b) but without the malonate group. Formate adducts exhibited a characteristic neutral loss of formic acid (46 Da) under collision-induced dissociation conditions. The  $[M-H]^-$  ion co-eluting with the above formate adduct is observed in MS/MS spectrum of products from  $m/z$  447 at  $m/z$  401 (experimental  $m/z$  401.2545; theoretical  $m/z$  for  $C_{21}H_{37}O_7^- = 401.2520$ ;  $\Delta m = -6.2$  ppm), and 239, corresponding to the loss of anhydrohexose ( $C_6H_{10}O_5$ ). Therefore, this metabolite was annotated as a sesquiterpene diol monohexoside. Eight isomers with matching pseudomolecular and fragment masses were observed in extracts from *S. habrochaites* LA1777 (Fig. 3.1g).

Similarly, another compound detected at  $m/z$  489 (experimental  $m/z$  489.2687 for peak 4h in Figure 3.1h; theoretical  $m/z$  for  $C_{24}H_{41}O_{10}^- = 489.2705$ ;  $\Delta m = -3.6$  ppm), consistent with  $[M+\text{formate}]^-$  of a sesquiterpenoid monoglycoside acetate ester. Formate adducts exhibited a characteristic neutral loss of formic acid (46 Da) under collision-induced dissociation conditions. Like the formate adduct of sesquiterpenoid monoglycoside described above, the  $[M-H]^-$  ion co-eluting with the above formate adduct is observed in MS/MS spectrum of products from  $m/z$  489 at  $m/z$  443 (experimental  $m/z$  443.2632; theoretical  $m/z$  for  $C_{23}H_{40}O_8^- = 443.2650$ ;  $\Delta m = -4.0$  ppm), 401 and 239, corresponding to the sequential loss of the acetate group (42 Da) followed by the loss of anhydrohexose ( $C_6H_{10}O_5$ ). Therefore, this metabolite was annotated as a sesquiterpene diol monohexoside acetate ester. Four isomers with matching pseudomolecular and fragment masses were observed in extracts from *S. habrochaites* LA1777 (Fig. 3.1h).



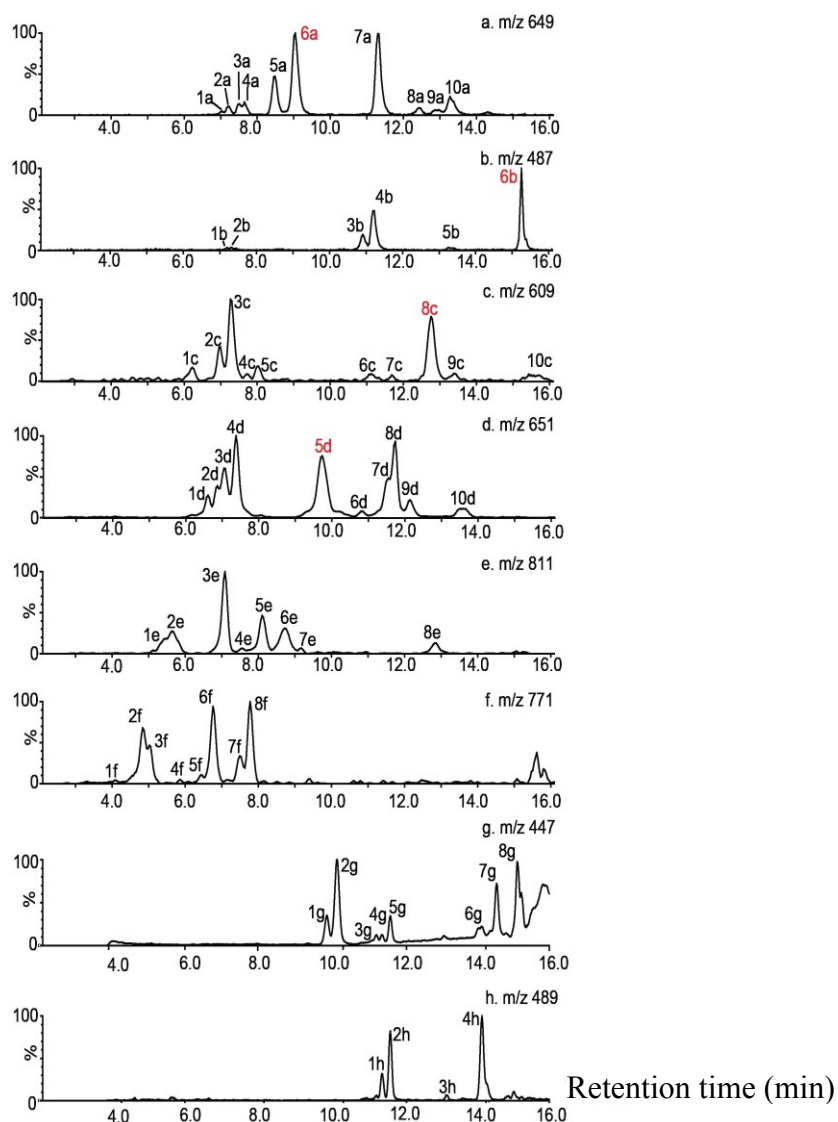


Figure 3.1. Extracted ion UHPLC-MS chromatograms showing sesquiterpene glycoside metabolites identified in extracts from *Solanum habrochaites* LA1777. (a)  $[M-H]^-$  ( $m/z$  649) for camphenenane diol diglucoside malonate esters, (b)  $[M-H]^-$  ( $m/z$  487) for camphenenane diol monoglucoside malonate esters, (c)  $[M+\text{formate}]^-$  ( $m/z$  609) for camphenenane diol diglucosides, (d)  $[M+\text{formate}]^-$  ( $m/z$  651) for camphenenane diol diglucoside acetate esters, (e)  $[M-H]^-$  ( $m/z$  811) for camphenenane diol triglucoside malonate esters, (f)  $[M+\text{formate}]^-$  ( $m/z$  771) for camphenenane diol triglucosides, (g).  $[M+\text{formate}]^-$  ( $m/z$  447) for camphenenane diol monoglucosides and (h)  $[M+\text{formate}]^-$  ( $m/z$  489) for camphenenane diol monoglucosides acetate ester. Labels with larger font size in red designate the four purified metabolites.

### 3.3.2 Structure elucidation of sesquiterpene diol dihexoside malonate ester and establishing the structure of campherenane terpenoid core

One of the major isomers of sesquiterpene diol dihexoside malonate ester (Fig. 1a, peak 6a.  $R_t = 9.0$  min) was purified from extracts of *S. habrochaites* LA1777, and its structure was elucidated using 1D and 2D NMR. The structure shown in Fig. 3.2a was assigned to this molecule based upon NMR and mass spectrometry information. In the proton NMR spectrum (Fig. 3.4 in appendix) the anomeric protons corresponding to the two glucose moieties could be identified at proton resonances  $\delta_H$  4.63 ppm ( $J = 8.52$  Hz),  $\delta_H = 4.28$  ppm ( $J = 7.97$  Hz) as two doublets. This supported and anchored the annotation of two hexose moieties. Coupling constants of each anomeric proton suggested that the glycosides are bound by beta-substitution. Three methyl group resonances could be identified. Two of these were singlets and one was a doublet ( $\delta_H$  0.99 ppm ( $J = 7.42$  Hz)). The doublet was indicative that one of the three methyl groups was connected to a tertiary carbon atom with a single proton. Other protons were not well resolved in the proton spectrum, and therefore to assign these, the multiplicity edited gHSQC spectrum was used (Fig. 3.6) , and all protons were readily identifiable along with their corresponding carbon connectivity. To identify neighboring protons, COSY, gHMBC and TOCSY spectra (Figs. 3.5, 3.7, 3.8 in appendix) were used. Starting from the anomeric protons,  $^1H$  and  $^{13}C$  resonances of the two hexose moieties were assigned. Next, the structure of terpenoid core was assigned using the same approach. In the terpenoid core, protons located on primary and secondary carbons with glycoside attachments were distinctive, with methylene protons at  $\delta_H$  3.36 ppm and  $\delta_H$  3.43 ppm for the primary carbon position (C12) and  $\delta_H$  4.00 ppm at the secondary carbon position (C2). The correlations observed in gHMBC spectra between these protons and anomeric carbons of the hexose moieties were important in identifying the position of each hexose attachment. Similarly,

protons on C2 carbon atom correlated in gHMBC spectra with the C22 carbon resonance and vice versa. The gHMBC spectrum was important in assigning the quaternary carbon atoms also. The two quaternary carbon atoms ( $\delta_C$  51.14 ppm and  $\delta_C$  52.06 ppm) displayed correlations in gHMBC spectra with the neighboring protons (Fig. 3a) and this enables their correct assignments. Next, the methylene protons in malonate were weakly visible in the gHSQC spectrum ( $\delta_H$  4.43 ppm and  $\delta_H$  4.33 ppm). The position of the malonate ester attachment was assigned using correlations observed in both gHMBC and NOESY (Fig. 3.9) spectra. Positioning of malonate could be inferred based on HMBC correlations between the hexose moiety protons ( $\delta_H$  3.71 ppm) and the carbonyl carbon ( $\delta_C$  172.45 ppm) (Fig. 3.4). In the NOESY spectrum, malonate protons showed correlations with protons in the hexose attached to the secondary position of the terpenoid core. ( $\delta_H$  4.32 ppm and  $\delta_H$  4.00 ppm of terpenoid core, and  $\delta_H$  4.32 and  $\delta_H$  3.18 ppm of hexose). Furthermore, in the NOESY spectrum, the proton at the secondary alcohol position (C2) shows correlation with protons on methyl attached to C7 position of terpenoid core ( $\delta_H$  4.00 ppm and  $\delta_H$  0.91 ppm terpenoid core) confirming the position of the secondary alcohol group. This information helps distinguish this terpenoid core structure from the previously reported camphenenane type cores [4]. Based on these observations the final structure shown in Fig. 3.2a was assigned and the carbon and proton resonances for each carbon and proton in the molecule are shown in Table 3.1.

### 3.3.3 Structure elucidation of sesquiterpene monohexoside malonate ester

Three main isomers of sesquiterpene monohexoside malonate molecule are observed in extracts of LA1777 (Fig. 1b  $m/z$  487). The major isomer (Fig. 3.1b peak 6b,  $R_t$  = 15.3 min) of this compound was purified and analyzed using 1D and 2D NMR spectroscopy. Based on NMR and mass spectrometry experiments the structure presented in Fig. 3.2b was assigned to this

compound. Proton and gHSQC NMR spectra of the compound (Fig. 3.10 and 3.12) were suggestive of the presence of three methyl groups corresponding to the sesquiterpene core and a single anomeric proton from the hexose moiety on the molecule ( $\delta_{\text{H}}$  4.25, d ( $J=8.04$  Hz)). Based on coupling constant of the anomeric proton the glycosidic bond that connected the hexose moiety to the terpenoid core was identified to be a beta bond. One of the resonances corresponding to a methyl was a doublet ( $\delta_{\text{H}}$  0.99 ppm ( $J=7.09$  Hz)) suggesting its located attached to a tertiary carbon atom. Proton – carbon connectivity of each CH, CH<sub>2</sub> and CH<sub>3</sub> was assigned using the multiplicity edited gHSQC data (Fig. 3.12). To assign neighboring carbons and protons, COSY and gHMBC spectra were used. To ascertain the positions of the quaternary carbons the gHMBC spectrum was used (Fig. 3.13). Starting from the anomeric proton, the resonances corresponding to the hexose moiety were assigned using gHSQC, COSY and gHMBC. Next, the structure of the terpenoid core was assembled using the same approach. In the terpenoid core, the protons on the primary alcohol and secondary carbon where the hexose group is attached (C2 and C12 respectively) could be identified ( $\delta_{\text{H}}$  3.72 ppm and  $\delta_{\text{H}}$  3.42 ppm for C2 and  $\delta_{\text{H}}$  3.98 ppm for secondary carbon (C12). To confirm the position of the primary alcohol group, the NOESY spectrum (Fig. 3.15) also was important. In the NOESY spectrum, the proton at C2 position correlates with protons on methyl group attached to C7 position ( $\delta_{\text{H}}$  3.98 ppm and  $\delta_{\text{H}}$  0.90 ppm of terpenoid core). The position of hexose was inferred based on the correlations observed in gHMBC spectra between these protons and the anomeric carbon atoms. The anomeric proton correlated with C12 carbon atom and the protons of C12 carbon atom correlated with anomeric carbon (C16) atom. Using this, the position of hexose moiety was determined. Position of the malonate group was inferred based on NOESY and gHMBC experiments. One of the methylene protons of malonate ( $\delta_{\text{H}}$  4.40 ppm) displayed NOEs with

hexose moiety protons ( $\delta_{\text{H}}$  3.36 ppm) indicating it is located proximal to hexose in space. The exact position of malonate was assigned using correlations observed in gHMBC between the hexose group proton ( $\delta_{\text{H}}$  3.87 ppm) and carbonyl carbon. The resonances assigned to each proton and carbon atom are shown in Table 3.1.

### 3.3.4 Structure elucidation of sesquiterpene diol dihexoside

A third novel molecule isolated from *S. habrochaites* LA1777 was the sesquiterpene dihexoside. At least ten isomers of this molecule were observed from UHPLC-MS profiling (Fig. 3.1c). This molecule displayed the presence of two hexose moieties but no sign of malonate was detected by NMR or mass spectrometry, and the observation of  $[\text{M}+\text{formate}]^-$  rather than  $[\text{M}-\text{H}]^-$  supports the notion that this metabolite lacks acidic groups. Two major isomers of this molecule were detected along with a number of less abundant isomers from LA1777, and the more abundant isomer (Fig. 1c Isomer 8c at  $R_t = 12.8$  min) was purified and analyzed using NMR (Fig. 3.2c and Table 3.1). The  $^1\text{H}$  NMR spectrum (Fig. 3.16) shows the presence of two anomeric protons ( $\delta_{\text{H}}$  4.60 ppm ( $J=7.84$  Hz) and  $\delta_{\text{H}}$  4.37 ppm ( $J=7.82$  Hz)). Coupling constants of the anomeric protons suggest beta glycosidic bonds that connect the hexoses. Positions of attachment for the hexose moieties to the terpenoid core were inferred based on HMBC couplings of anomeric protons and anomeric carbons with the terpenoid core (Fig. 3.19). Spectra suggest that all three compounds share a common sesquiterpene diol (campherenane diol) core. (Fig. 3.16 to Fig. 3.21)

### 3.3.5 Structure elucidation of sesquiterpene diol dihexoside acetate ester

Sesquiterpene dihexoside acetate ester was detected from *S. habrochaites* LA1777 at  $m/z$  651 as  $[\text{M}+\text{formate}]^-$  which yielded a  $[\text{M}-\text{H}]^-$  at  $m/z$  605 upon collisional activation. The presence of two hexose moieties and an acetate group in this molecule was judged based on neutral losses observed during collisional activation of  $[\text{M}+\text{formate}]^-$  ion. One of the most abundant isomers,

(Fig. 3.1. 5d) was purified using preparative HPLC and NMR spectral analysis was performed. The NMR spectra showed the presence of two anomeric protons ( $\delta_H$  4.66 ppm ( $J=8.2$  Hz) and  $\delta_H$  4.43 ppm ( $J=7.68$  Hz)) confirming the presence of two hexoses (Fig. 3.22). The coupling constants of the two anomeric protons suggest beta glycosidic bonds that connect the hexoses. HMBC spectra confirmed that the two hexose moieties are connected to one another. Also the connectivity of the hexoses to the terpenoid core was confirmed using HMBC. HSQC and proton displayed the presence of four methyl groups in the molecule. Three of these were attributed to the terpenoid core and the last one was inferred to be present in the acetate group. Spectra suggest that all four isolated compounds share a common sesquiterpene diol (2,12-camphenene diol) core with identical relative stereochemical configuration at C7 (Fig. 3.22 to Fig. 3.28).

### **3.3.6 Using GC-MS data to support NMR based structure assignments of sugar moieties**

Despite the usefulness of NMR in *de novo* structure assignments, assigning NMR chemical shifts for the accurate assignment of sugar structures is challenging since different sugars (ex: glucose, galactose) can show similar chemical shifts. This can be particularly challenging in the presence of multiple sugars in a given molecule due to overlapping chemical shifts. However, the gas chromatographic retention times of MSTFA derivatives of different sugars are characteristic, and therefore offer a reliable alternative to NMR for assigning the sugars.

The sugar composition of these glycoconjugates was examined using acid hydrolysis followed by derivatization and GC-MS, which distinguishes sugars based on their retention times and mass spectra. The GC-MS retention times and spectra were compared with results from reference monosaccharide standards, and the only sugar detected in any of these analyses was glucose.

Results of these analyses are presented in Fig. 3.3.

### 3.4 Conclusions

Compared to the santalenes, campherenanes identifying in this study do not contain any double bonds or the three membered ring structures. Additionally, two hydroxyl groups are found in the campherenane diol core. These structural differences indicate that campherenanes potentially undergo more extensive redox reactions during the biosynthesis compared to santalenes.

Furthermore, the additional glycosylation, acetylation and malonylation are indicative of the involvement of a number of additional secondary reactions that transforms the basic terpenoid core into a non-volatile glycosylated terpene form. Such extensive transformation of sesquiterpene precursors into glycosylated sesquiterpenoid form is an indication of the involvement of a large number of different enzymes in the biosynthetic pathway that are still to be explored. Thus, these findings open the doors to a novel class of compounds that have been hidden from researchers studying *Solanum habrochaites* LA1777 for many decades.

## **APPENDIX**



Table 3.1: Chemical shifts (in ppm) of  $^1\text{H}$  and  $^{13}\text{C}$  resonances for (a) campherenane diol diglucoside malonate ester (peak 6a from Figure 3.1), (b) campherenane diol monoglucoside malonate ester (peak 6b), (c) campherenane diol diglucoside (peak 8c), and (d) campherenane diol diglucoside acetate ester (peak 5d).

	<b>a</b>			<b>c</b>			<b>c</b>			<b>d</b>		
<b>Position</b>	$\delta\text{ C}$	$\delta\text{ H}$		$\delta\text{ C}$	$\delta\text{ H}$		$\delta\text{ C}$	$\delta\text{ H}$		$\delta\text{ C}$	$\delta\text{ H}$	
1	51.1			51.2			51.2			52.0		
2	78.7	4.00		78.8	3.98		78.0	3.95		78.5	4.00	
3	39.6	0.99	2.22	39.7	0.99	2.22	39.0	0.95	2.19	39.7	0.97	2.21
4	43.9			43.7	1.78		43.0	1.77		43.5	1.77	
5	44.9	1.41	1.77	29.4	1.25	1.66	29.0	1.21	1.65	29.6	1.27	1.67
6	27.6	2.00	1.29	27.6	2.00	1.29	27.0	1.95	1.28	27.5	2.01	1.30
7	52.1			51.6			51.6			51.9		
8	35.0	1.48	1.01	34.8	1.48	1.02	34.5	1.45	0.96	34.9	1.48	1.00
9	43.1	1.42	1.13	31.1	1.32	1.32	31.4	1.24	1.31	22.8	1.48	1.23
10	36.2	1.50	1.14	36.3	1.51	1.14	35.7	1.45	1.1	36.0	1.49	1.13
11	35.2	1.81		35.1	1.80		34.8	1.77		35.2	1.80	
12	77.0	3.36	3.43	76.9	3.72	3.42	76.9	3.70	3.45	76.8	3.75	3.48
13	18.3	0.99 (3H,d)		17.8	0.99 (3H,d)		17.8	0.94 (3H,d)		17.7	0.99 (3H,d)	
14	16.0	0.91 (3H,s)		16.1	0.9 (3H,s)		16.1	0.87 (3H,s)		16.5	0.91 (3H,s)	
15	14.0	0.86 (3H,s)		14.2	0.86 (3H,s)		13.8	0.82 (3H,s)		14.5	0.89 (3H,s)	
16	105.6	4.28		105.8	4.25		103.5	4.37		103.9	4.43	
17	75.9	3.21		76.2	3.23		83.4	3.48		83.2	3.50	
18	78.6	3.39		71.9	3.40		78.6	3.58		78.6	3.60	
19	72.7	3.31		72.3	3.31		78.8	3.38		71.95	3.34	

Table 3.1 (cont'd)

20	75.8	3.52		78.8	3.35		76.9	3.27		78.73	3.31	
21	63.6	3.88	3.69	63.2	3.67	3.87	63.7	3.68	3.88	63.0	3.89	3.72
22	106.3	4.63		170.9			105.8	4.63		106.00	4.66	
23	83.7	3.43		65.0	4.4	4.35	76.1	3.24		76.6	3.28	
24	82.7	3.52		170.9			78.9	3.41		78.6	3.44	
25	79.0	3.58					72.6	3.84		72.3	3.32	
26	72.8	3.18					72.3	3.34		78.7	3.31	
27	63.6	3.71	3.88				63.6	3.90	3.68	63.0	3.89	3.72
28	172.5									172.6		
29	66.3	4.43	4.33							32.3	1.37	
30	170.4											

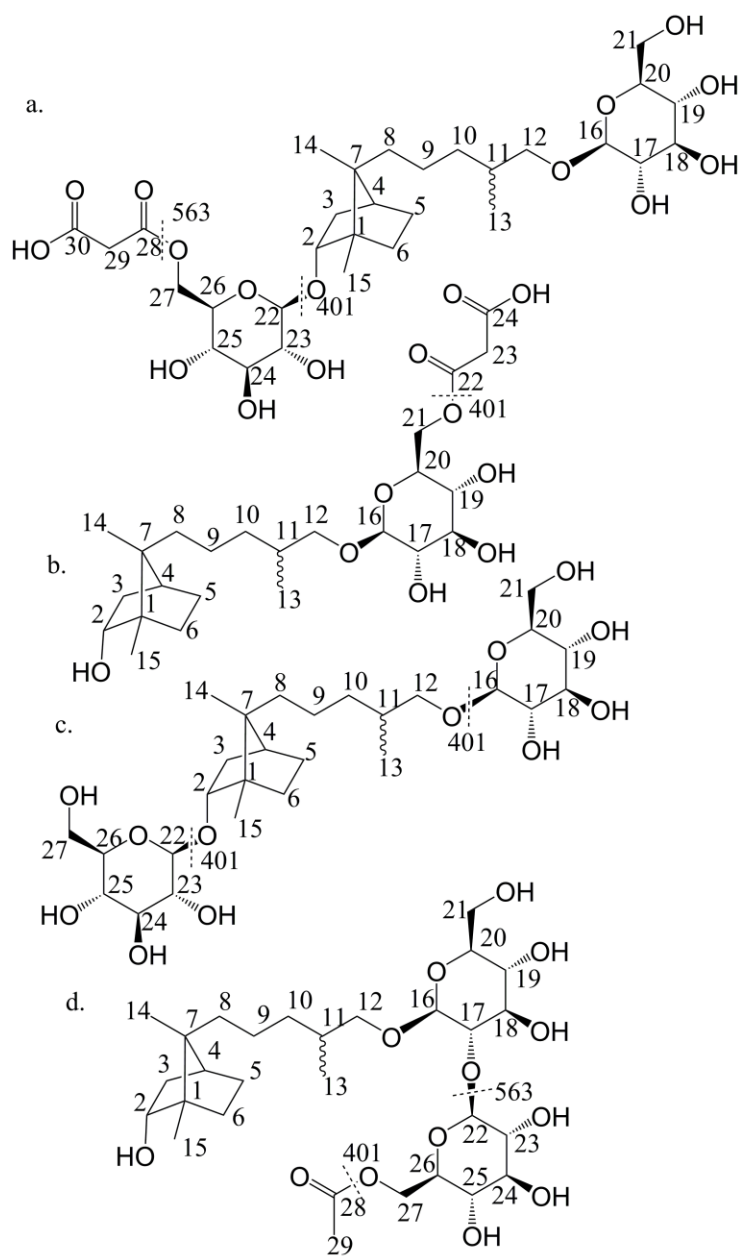


Figure 3.2: Structures of (a) campherenane-2,12-diol diglucoside malonate ester (peak 6a from Figure 1), (b) campherenane-2,12-diol monoglucoside malonate ester (peak 6b), (c) campherenane-2,12-diol diglucoside (peak 8c), and (d) campherenane-2,12-diol diglucoside acetate ester (peak 5d) as determined using NMR and tandem mass spectrometry. Portions of the molecule corresponding to key fragment ions in negative ion MS/MS spectra are illustrated on each structure. Carbon atoms are numbered in accordance with NMR assignments in Table 3.1.

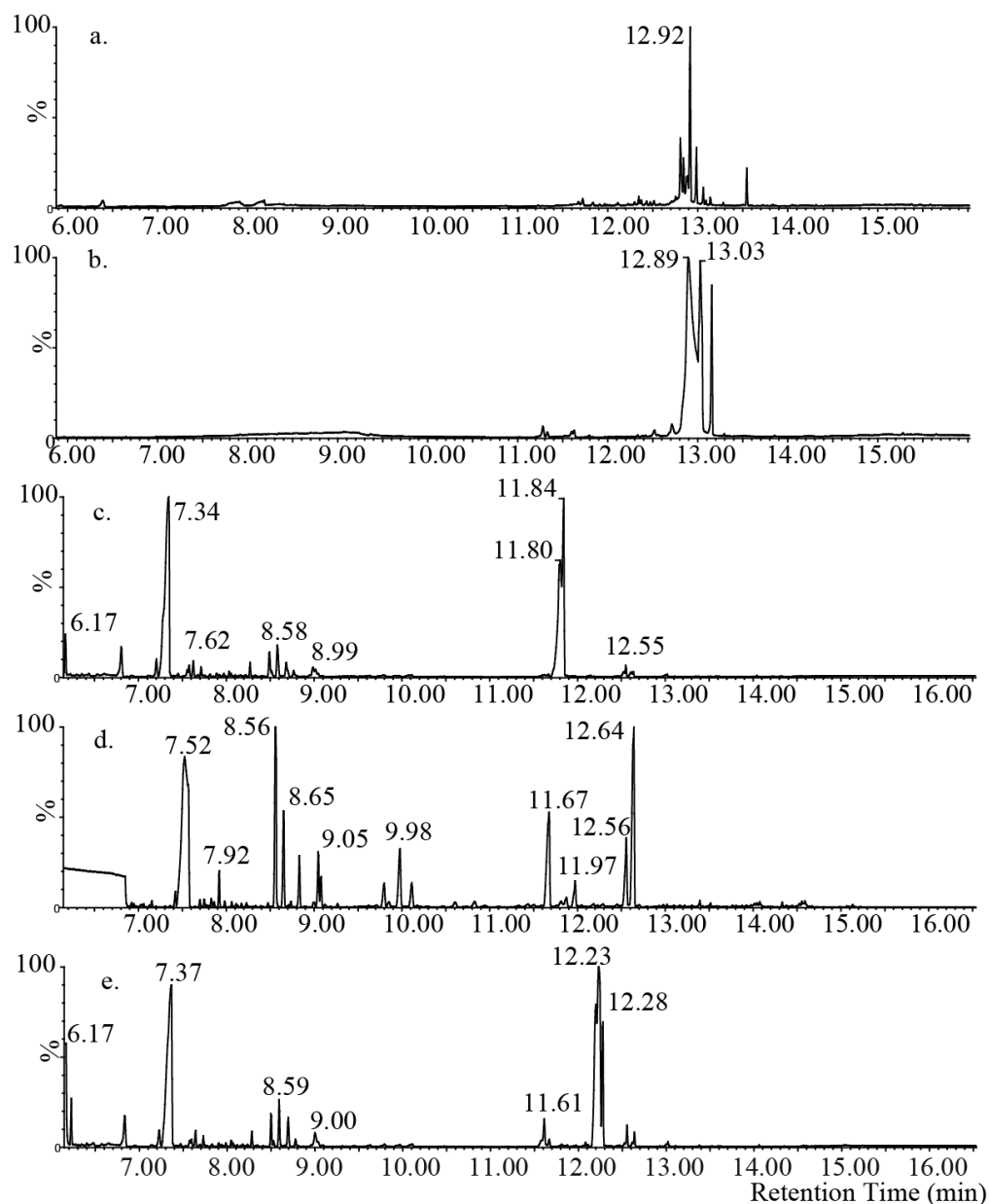


Figure 3.3. GC/MS total ion chromatograms of methoxime-trimethylsilyl derivatives of (a) products of acid hydrolysis of campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound **1**; peak 6a from Figure 1), (b) glucose reference standard, (c) xylose reference standard, (d) galactose reference standard, (e) rhamnose reference standard.

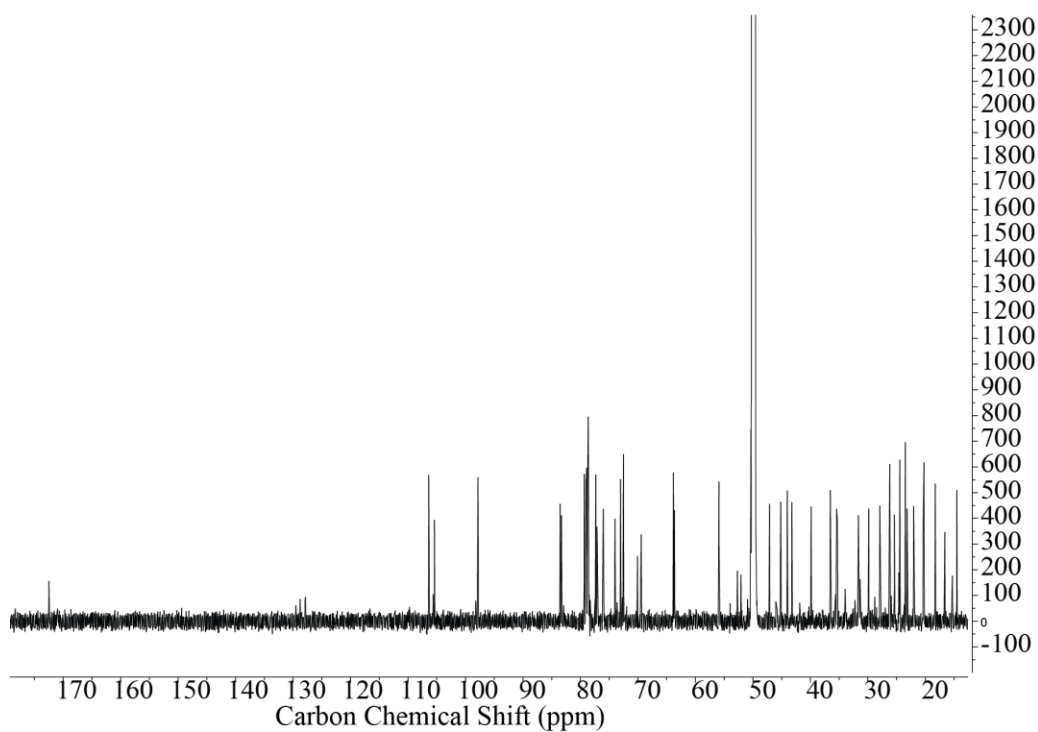
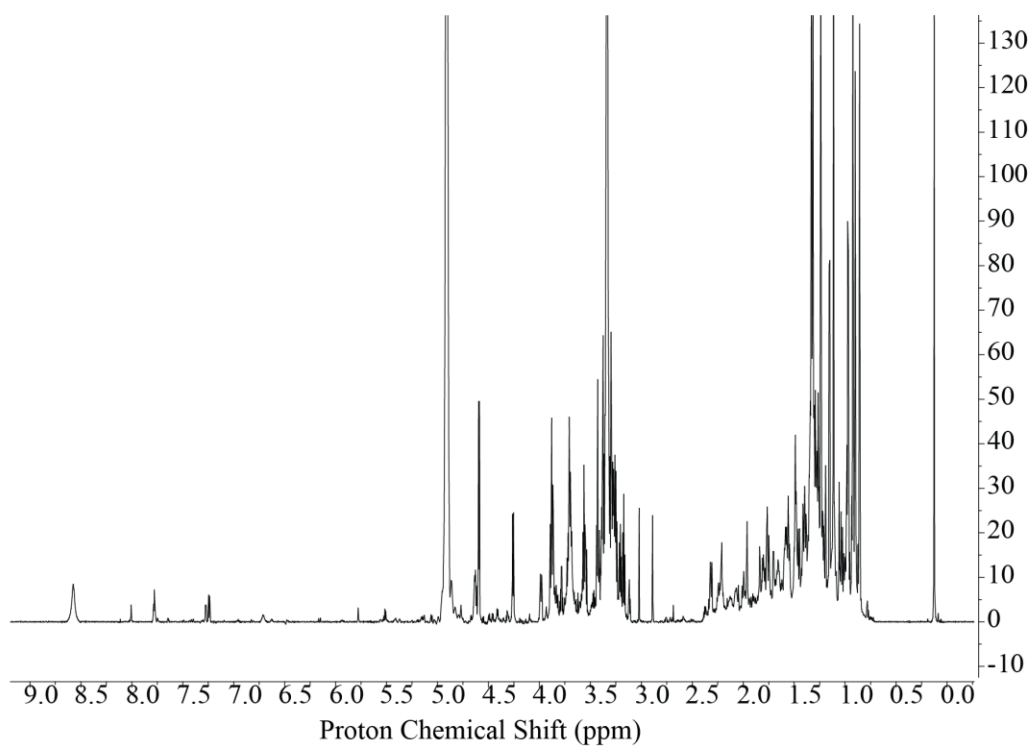


Figure 3.4.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of the isolated campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound **1**; peak 6a from Figure 3.1).

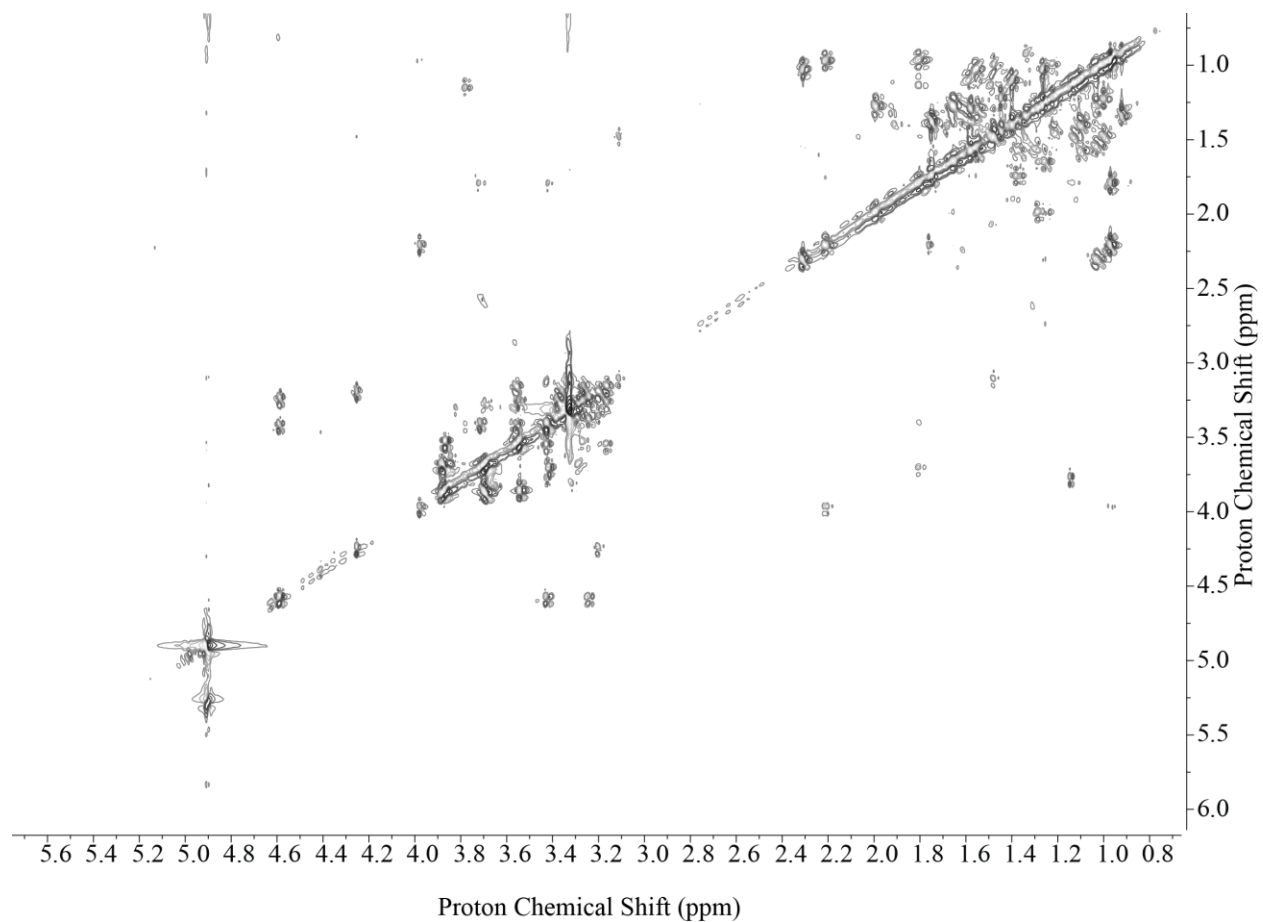


Figure 3.5.  $^1\text{H}$ - $^1\text{H}$  COSY NMR spectrum of the campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound **1**; peak 6a from Figure 3.1).

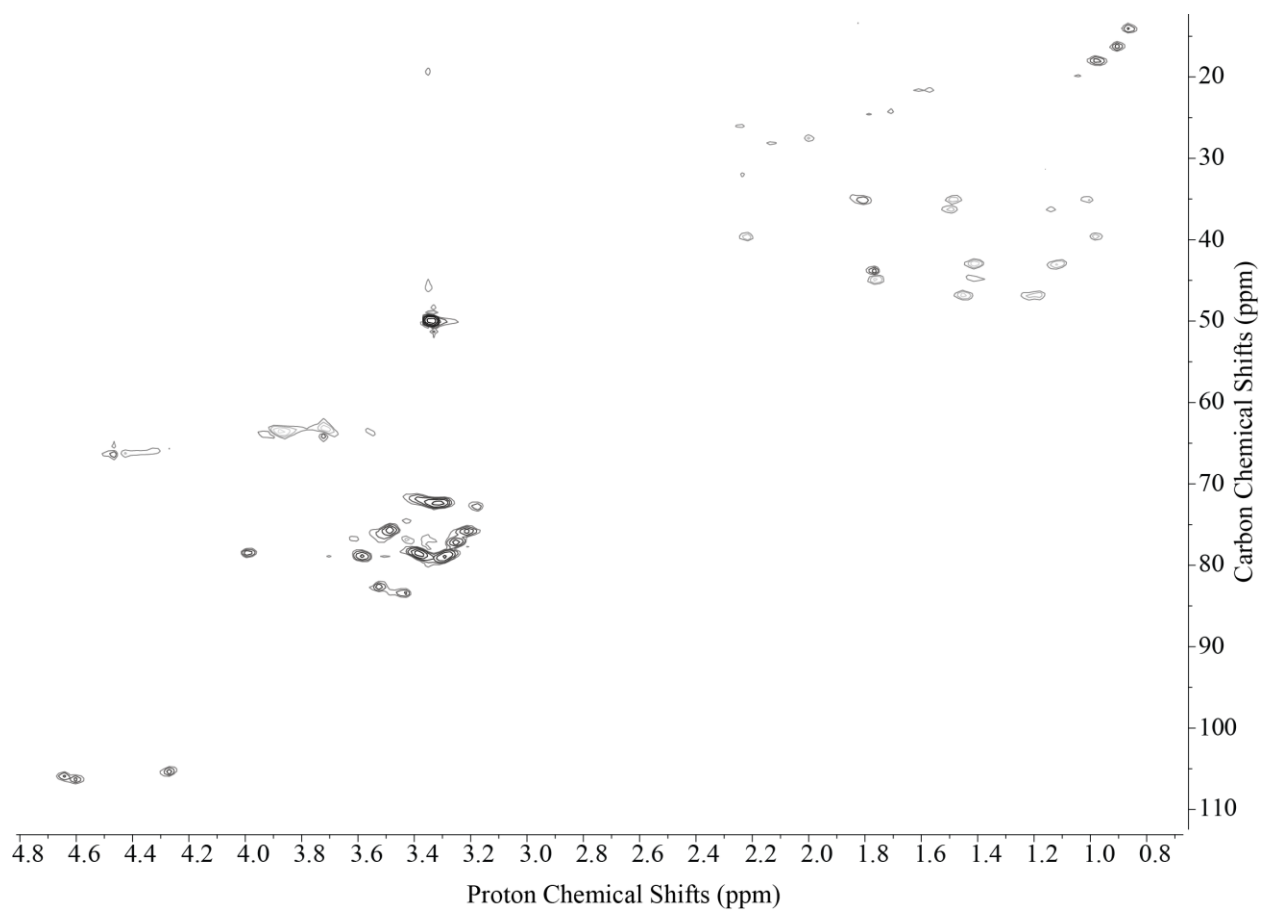


Figure 3.6. Multiplicity edited gHSQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectrum of campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound **1**; peak 6a in Figure 1).

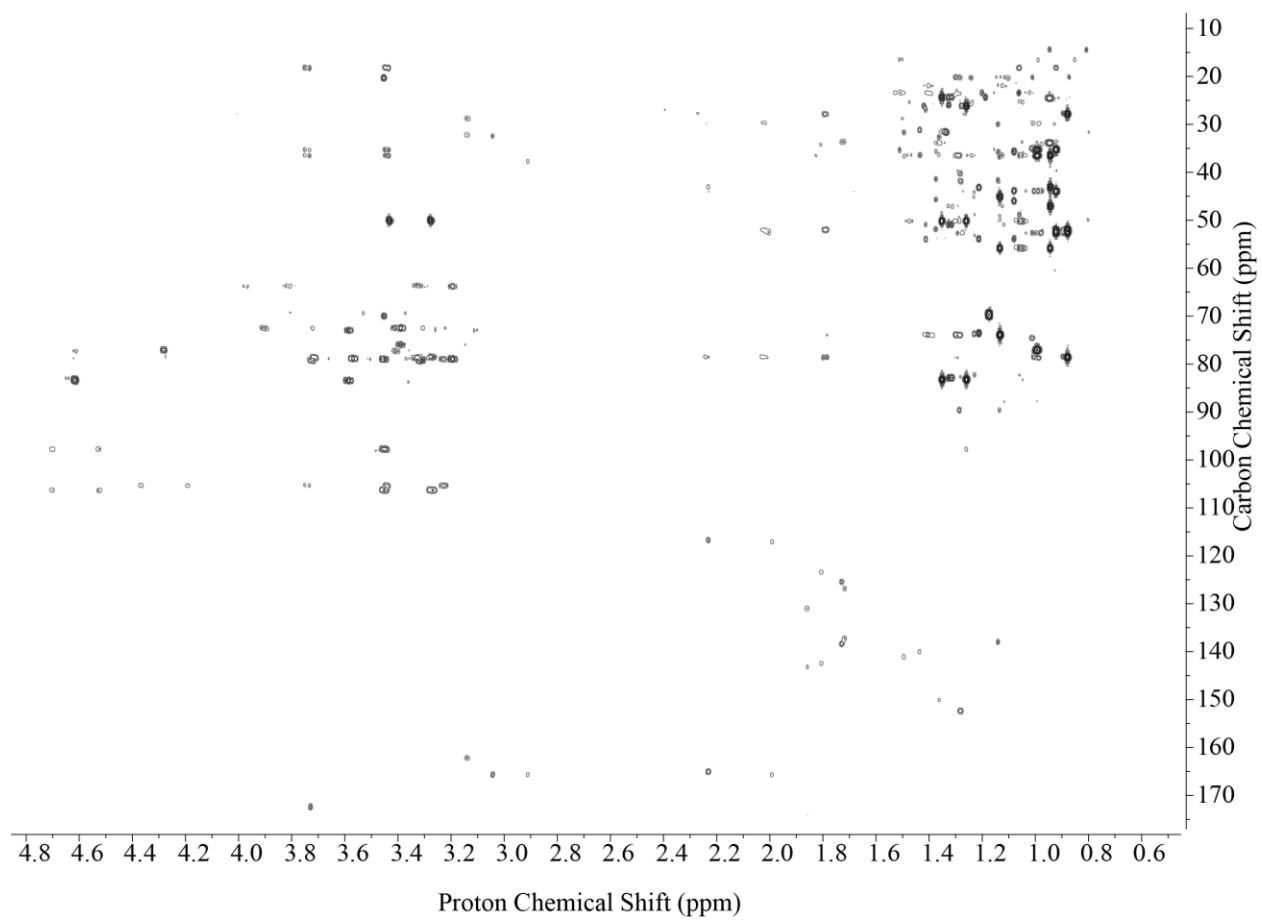


Figure 3.7. gHMBC NMR spectrum of campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound **1**; peak 6a from Figure 1).



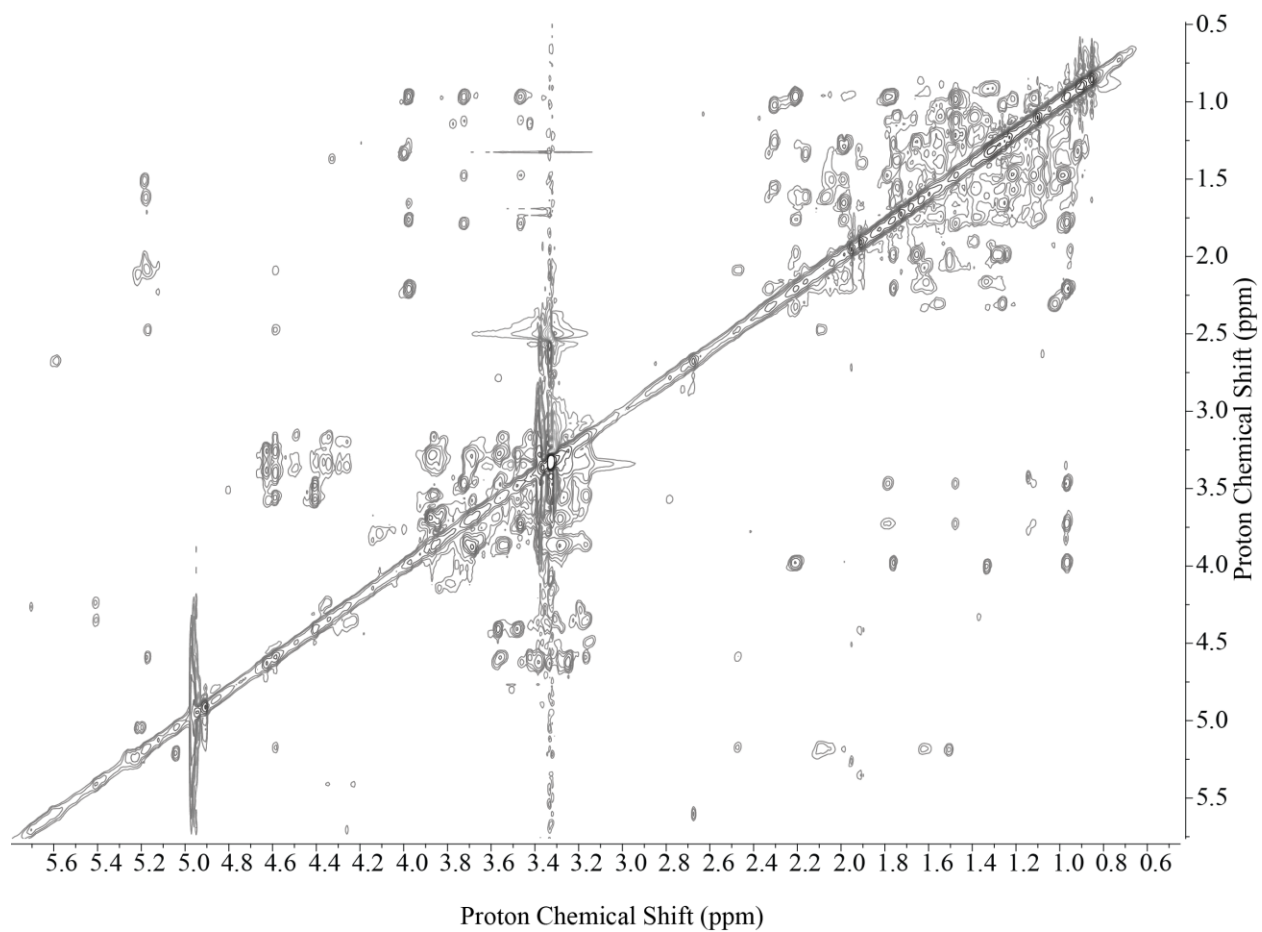


Figure 3.8. TOCSY NMR spectrum of campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound 1; peak 6a in Figure 1).

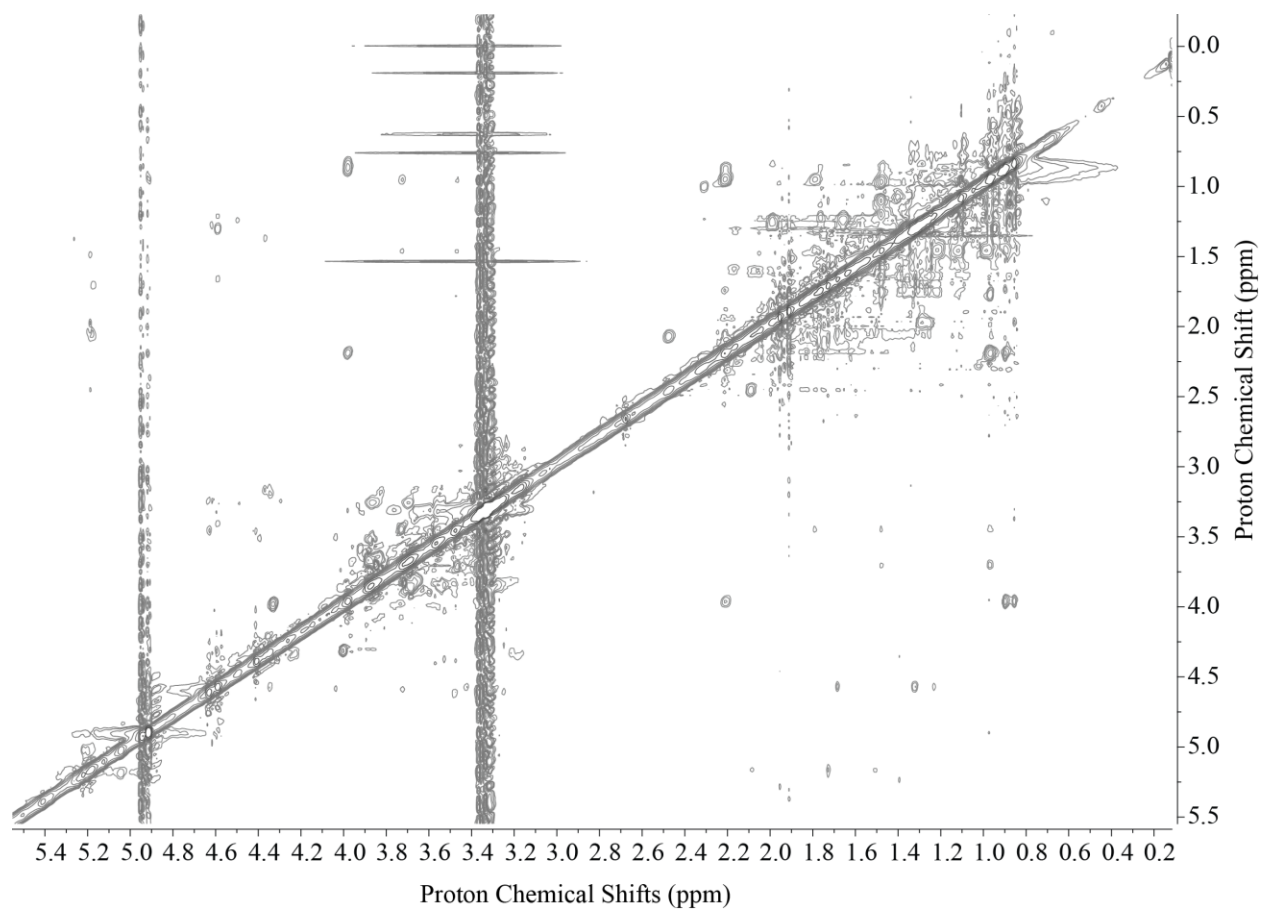


Figure 3.9. NOESY NMR spectrum of campherenane-2-*endo*-(6'-malonyl)glucoside)-12-glucoside (Compound **1**; peak 6a in Figure 1).

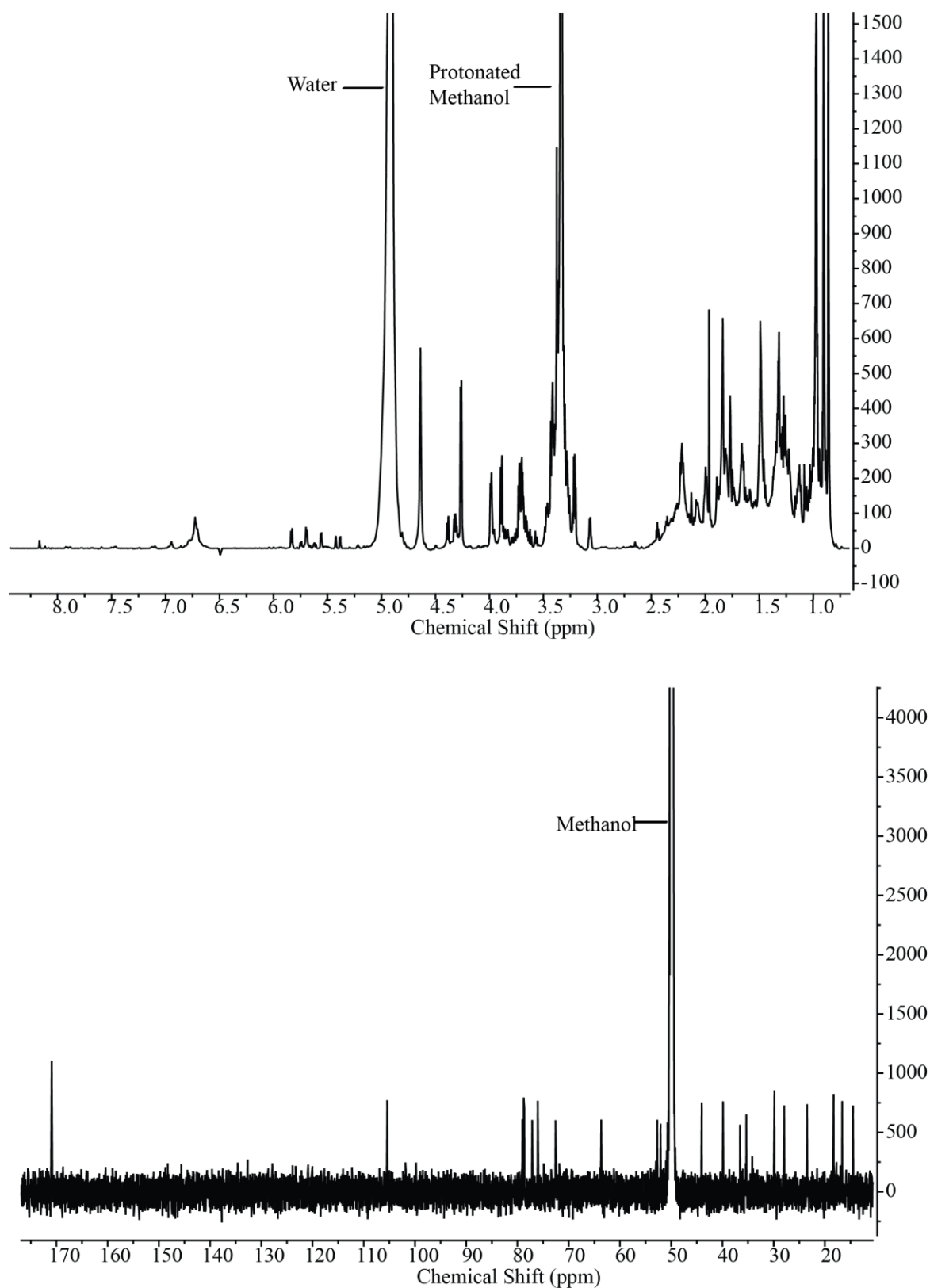


Figure 3.10.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of the isolated 2-*endo*-campherenanol-12-(6'-malonyl)glucoside (Compound **2**; peak 6b from Figure 1).

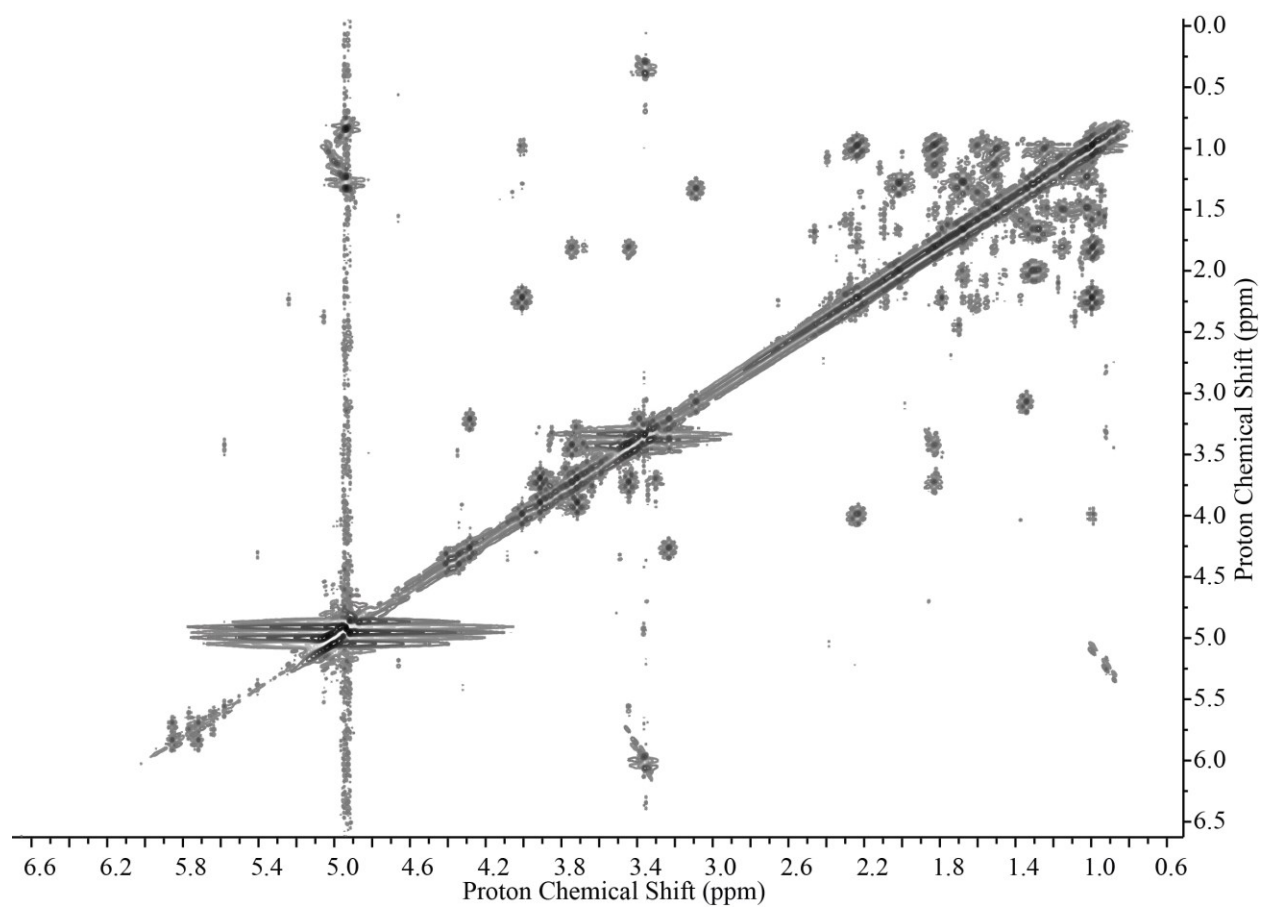


Figure 3.11.  $^1\text{H}$ - $^1\text{H}$  COSY NMR spectrum of the 2-*endo*-campherenanol-12-(6'-malonyl)glucoside (Compound **2**; peak 6b from Figure 1).

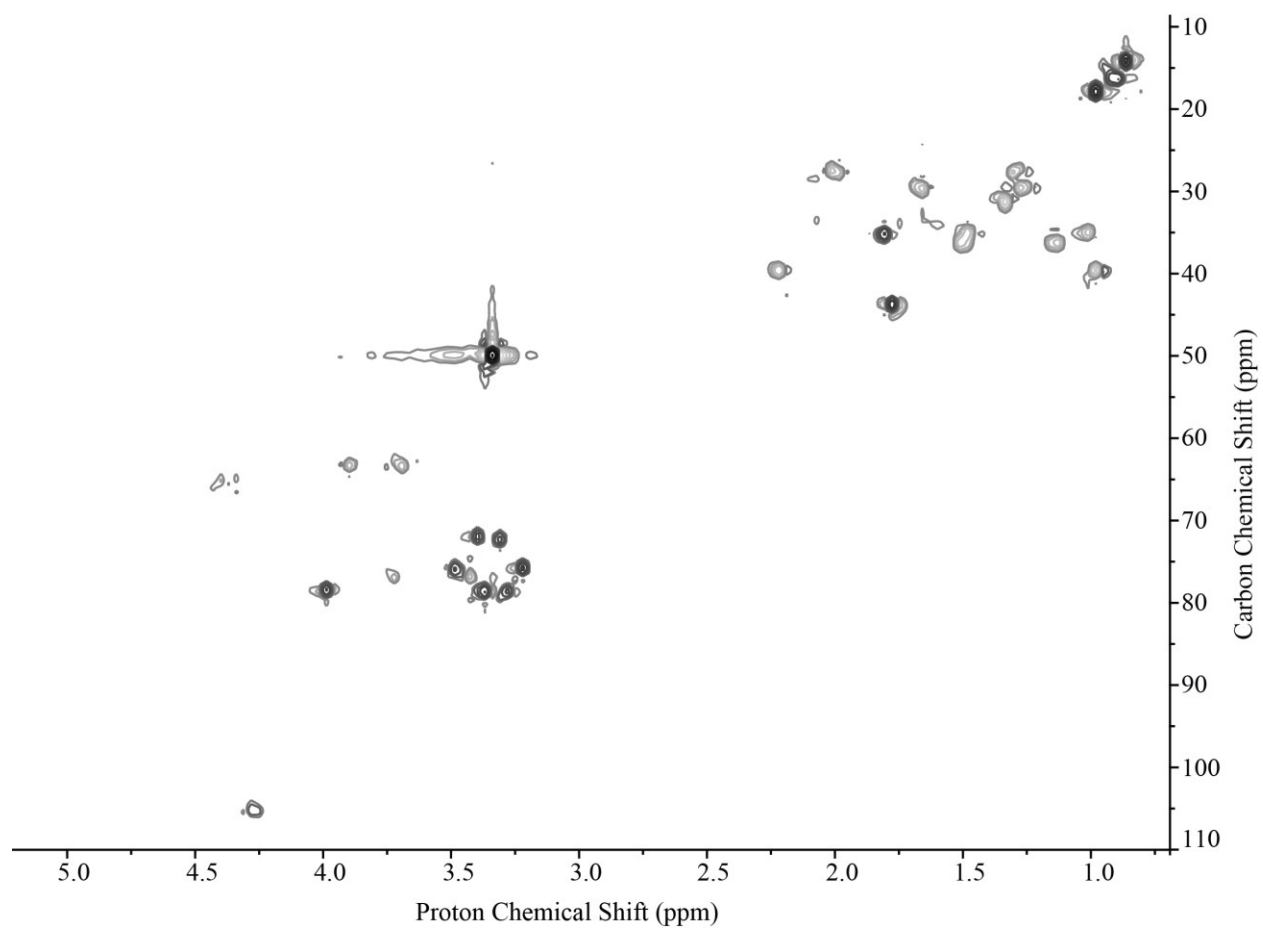


Figure 3.12. Multiplicity edited gHSQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectrum of 2-*endo*-campherenanol-12-(6'-malonyl)glucoside (Compound **2**; peak 6b in Figure 1).

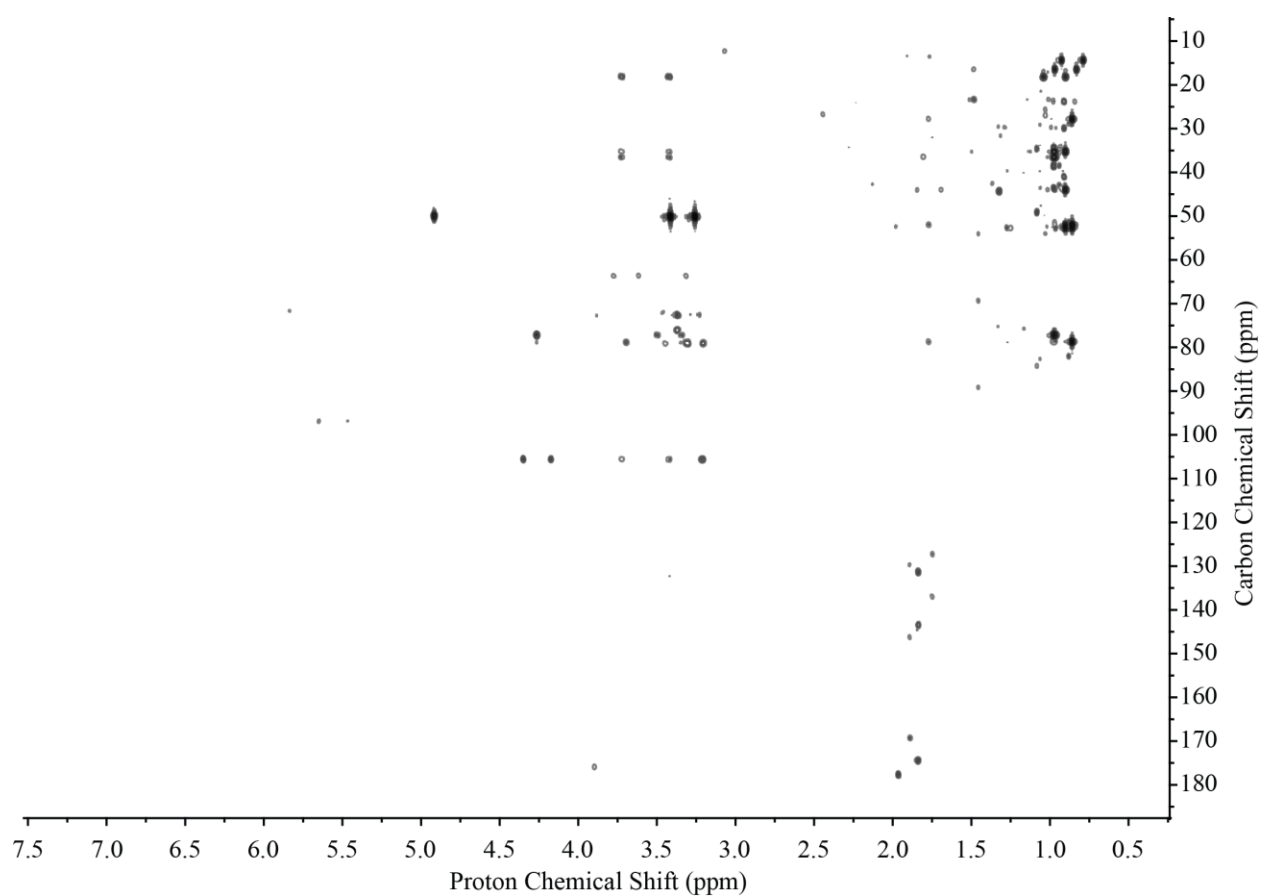


Figure 3.13. gHMBC NMR spectrum of 2-*endo*-campherenanol-12-(6'-malonyl)glucoside (Compound **2**; peak 6b from Figure 1).

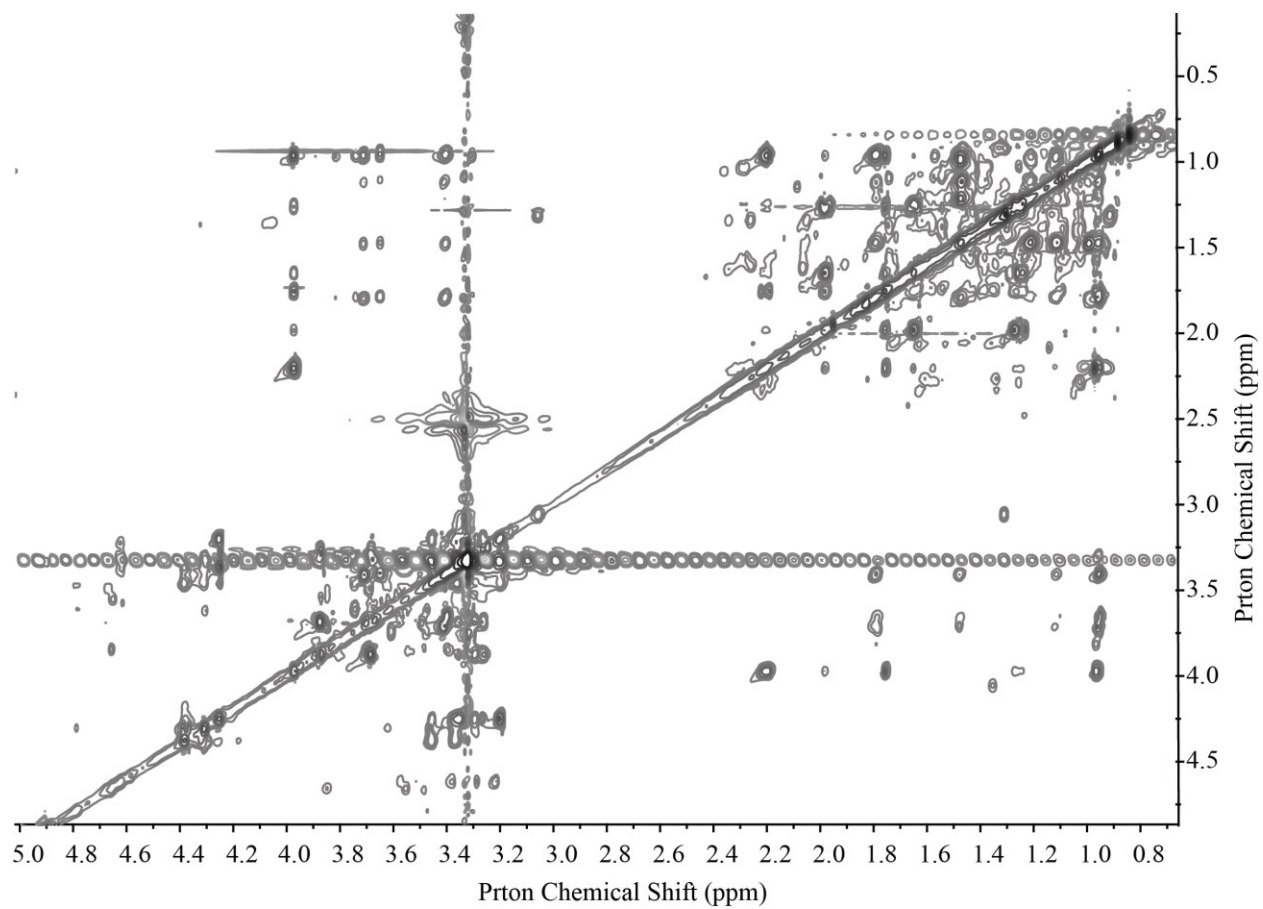


Figure 3.14. TOCSY NMR spectrum of 2-*endo*-campherenanol-12-(6'-malonyl)glucoside (Compound **2**; peak 6b in Figure 1).

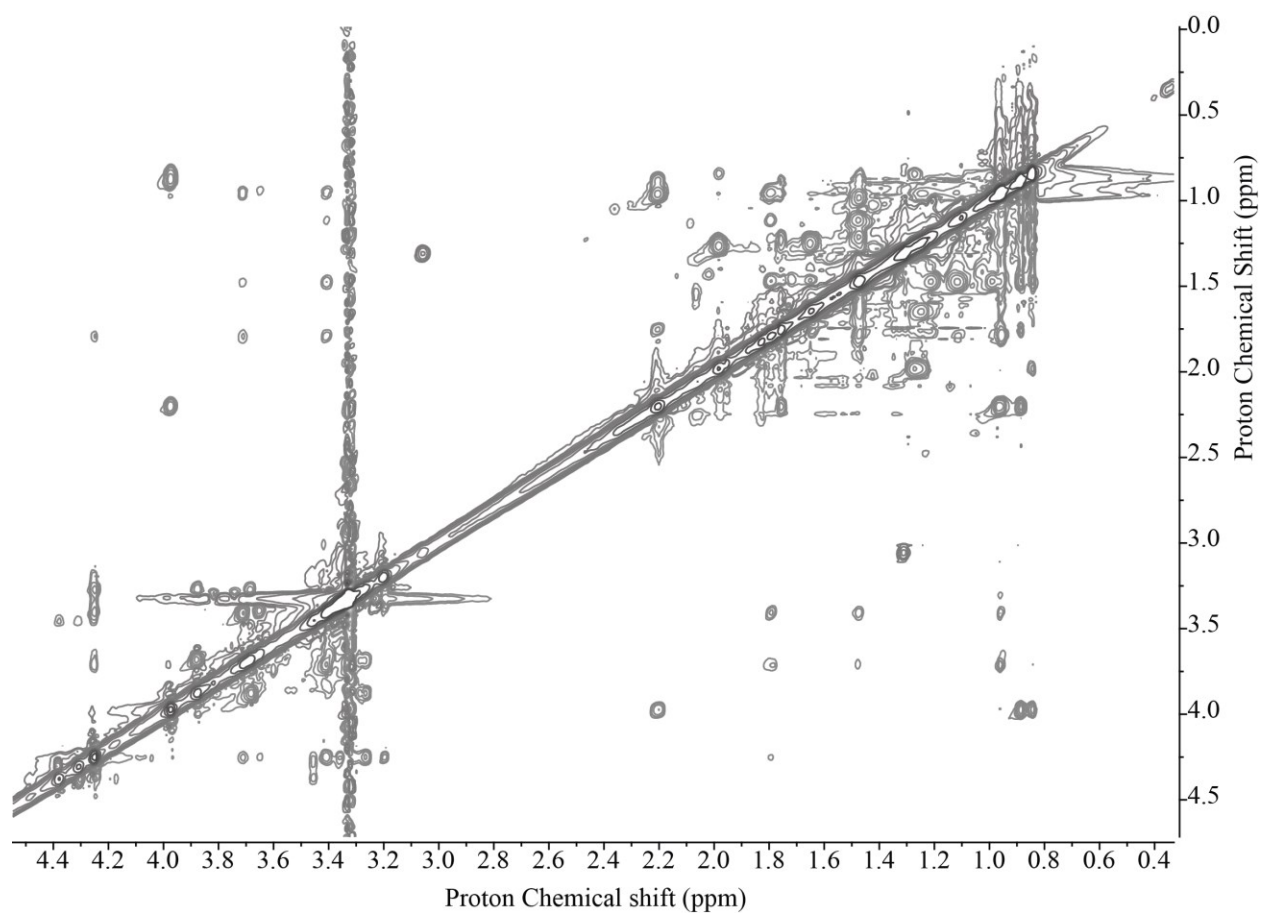


Figure 3.15. NOESY NMR spectrum of 2-*endo*-campherenanol-12-(6'-malonyl)glucoside (Compound **2**; peak 6b in Figure 1).



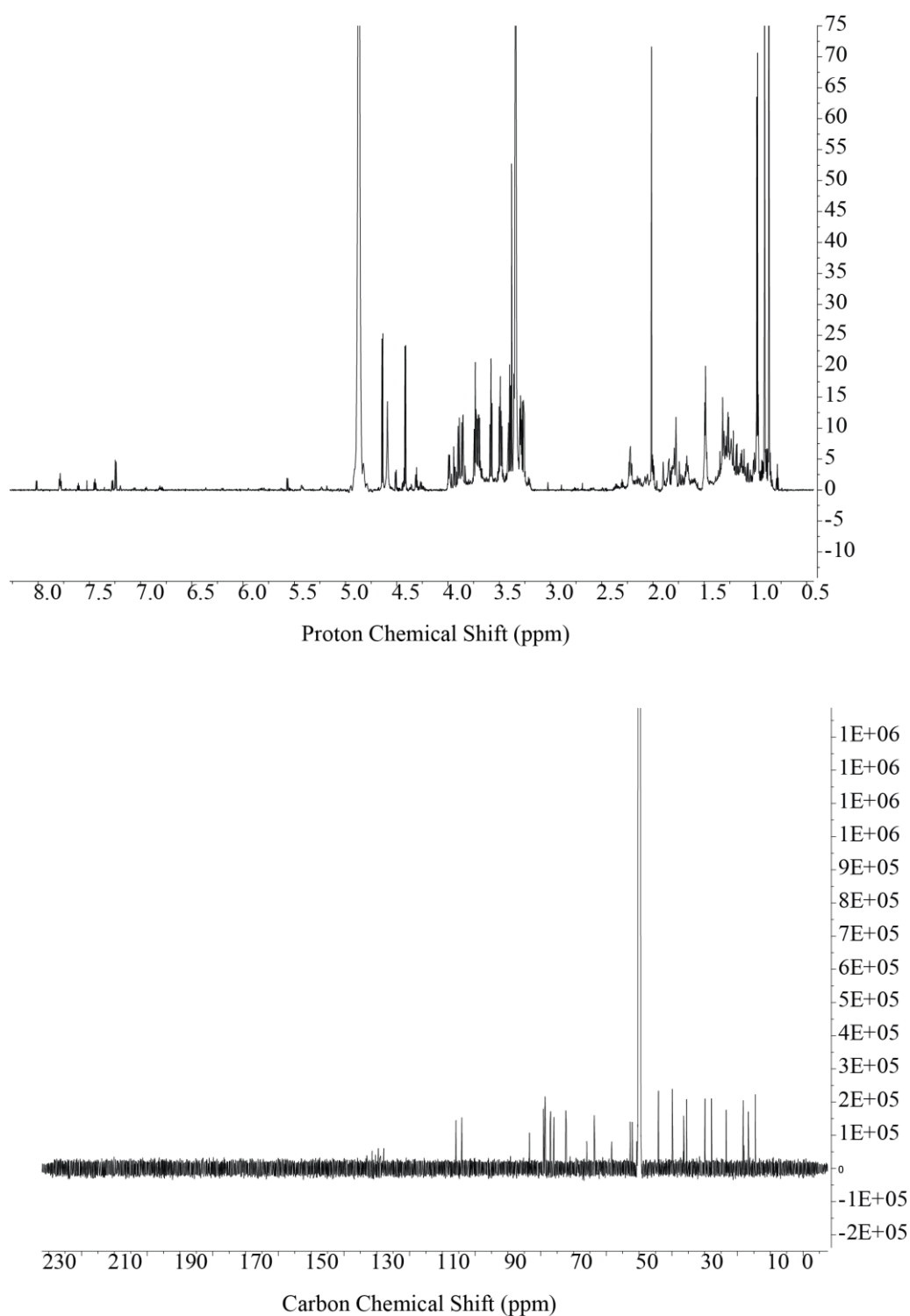


Figure 3.16.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of the isolated campherenan-2-*endo*,12-diglucoside (Compound 3; peak 8c from Figure 3.1).

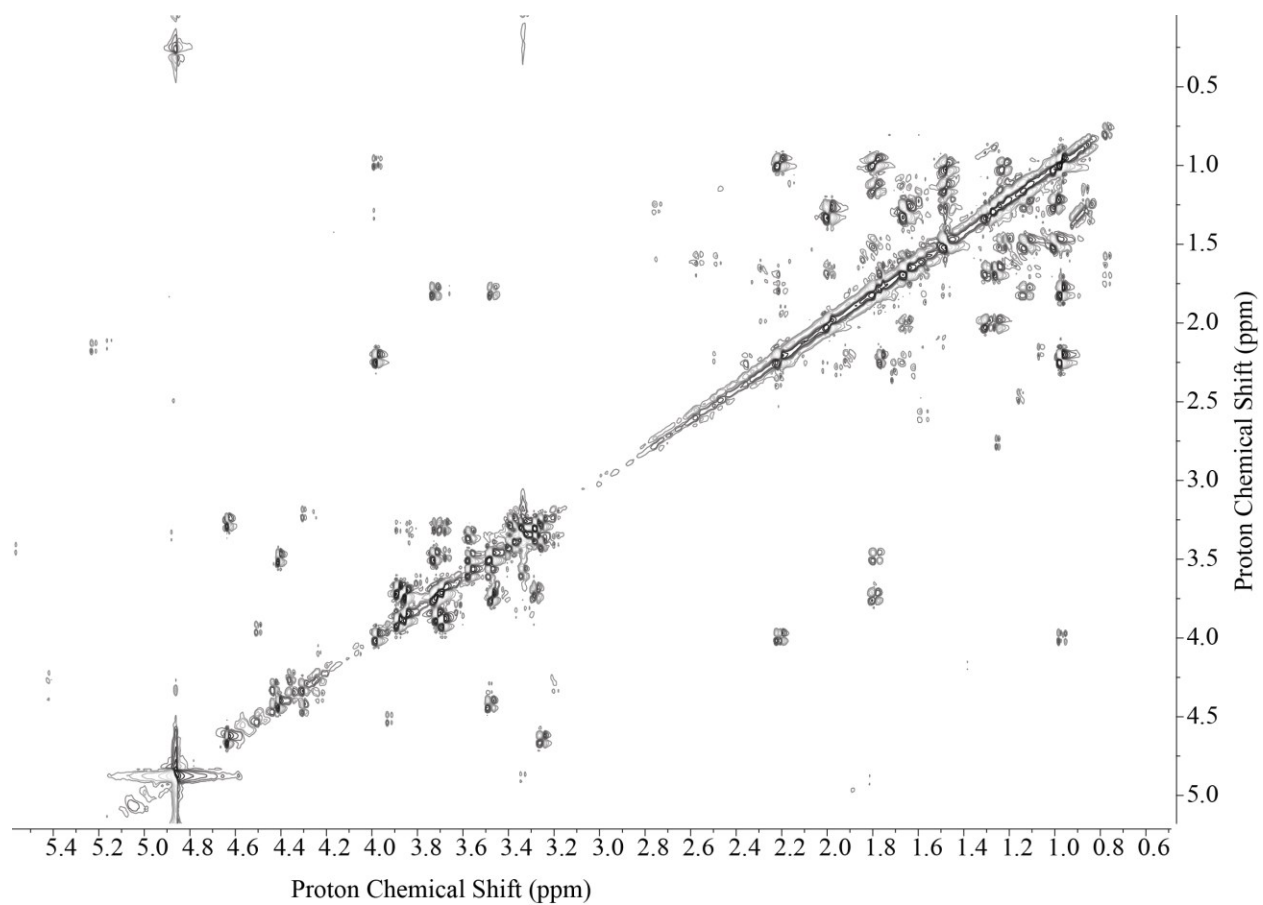


Figure Figure 3.17.  $^1\text{H}$ - $^1\text{H}$  COSY NMR spectrum of the campherenan-2-*endo*,12-diglucoside (Compound 3; peak 8c from Figure 1).

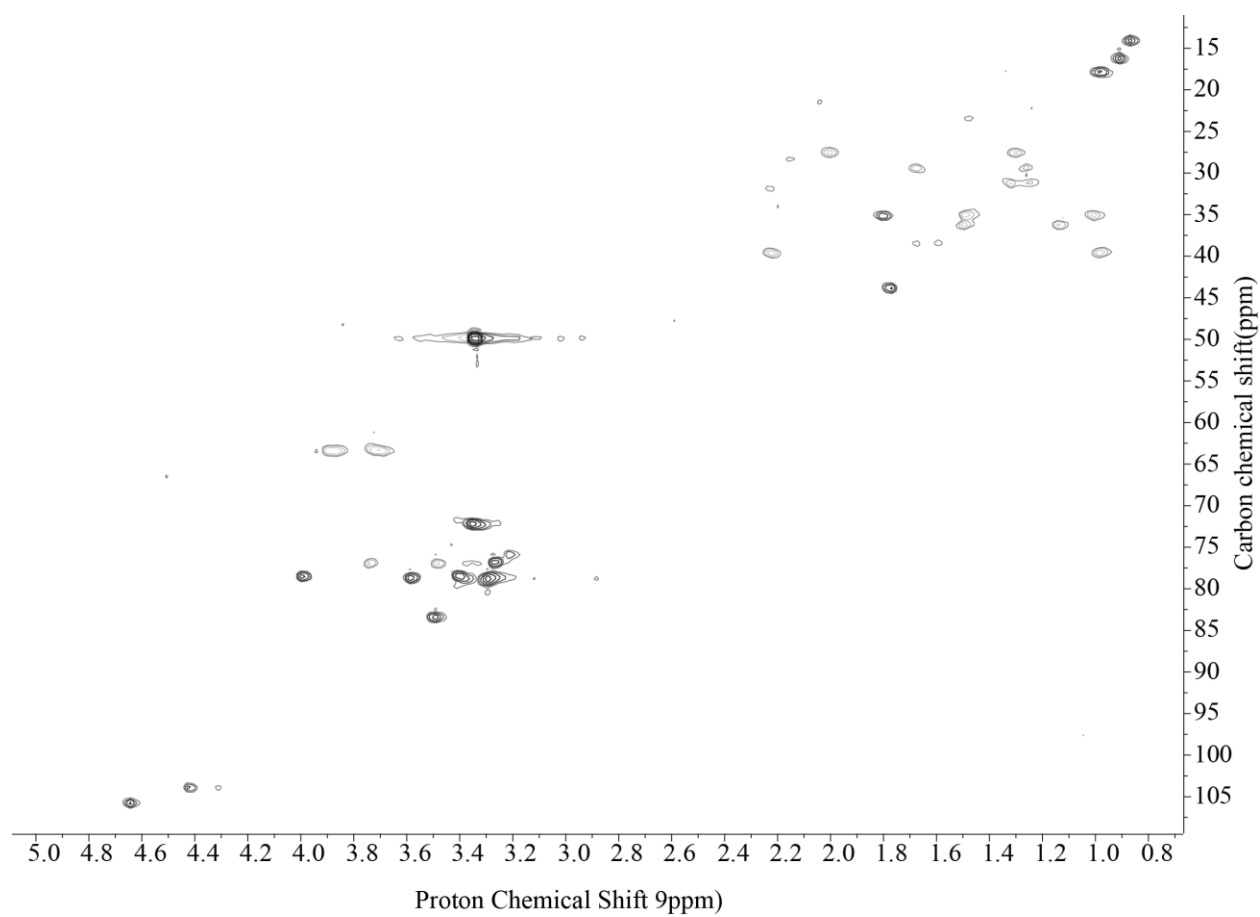


Figure 3.18. Multiplicity edited gHSQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectrum of campherenan-2-*endo*,12-diglucoside (Compound 3; peak 8c in Figure 1).

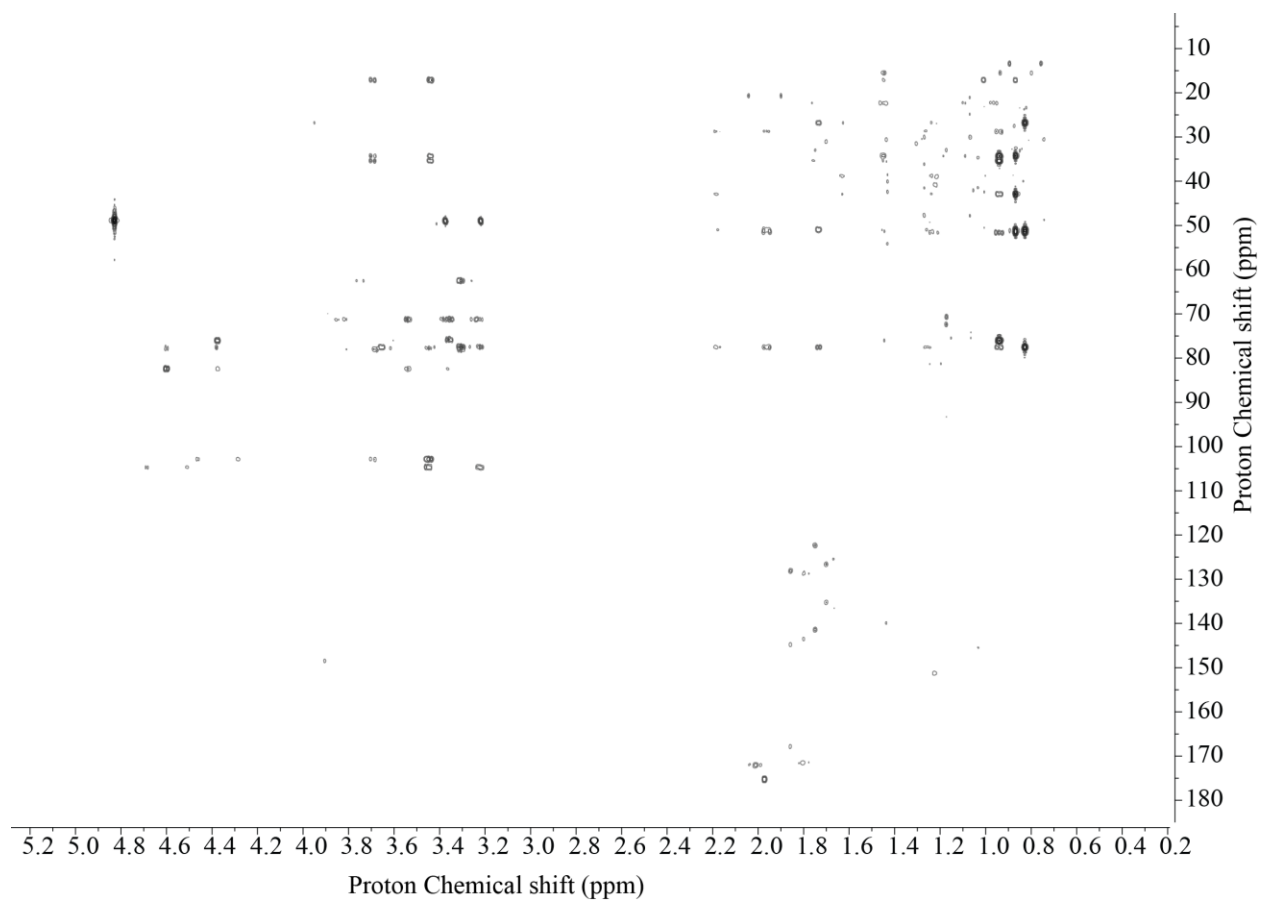


Figure 3.19. gHMBC NMR spectrum of campherenan-2-*endo*,12-diglucoside (Compound **3**; peak 8c from Figure 1).

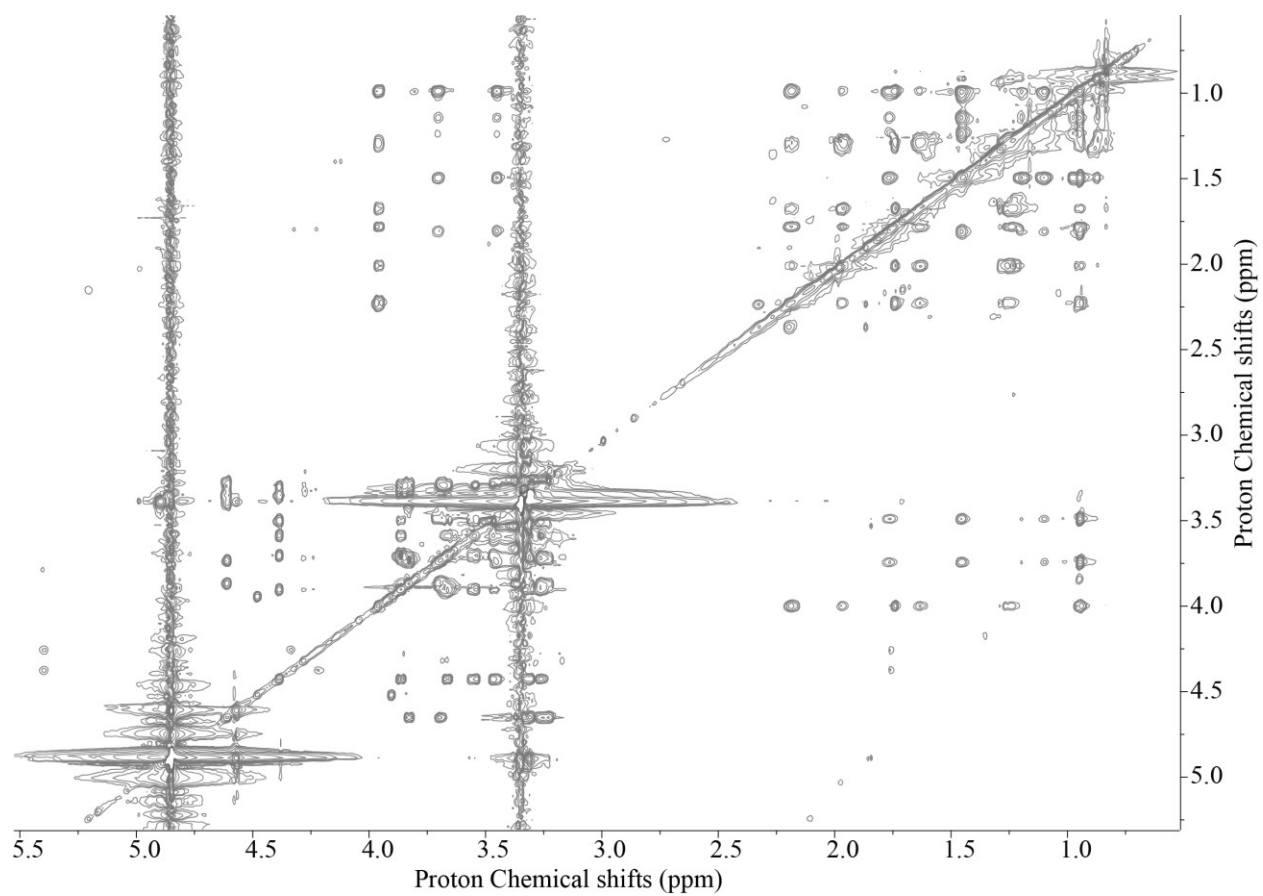


Figure 3.20. TOCSY NMR spectrum of campherenan-2-*endo*,12-diglucoside (Compound **3**; peak 8c in Figure 1).

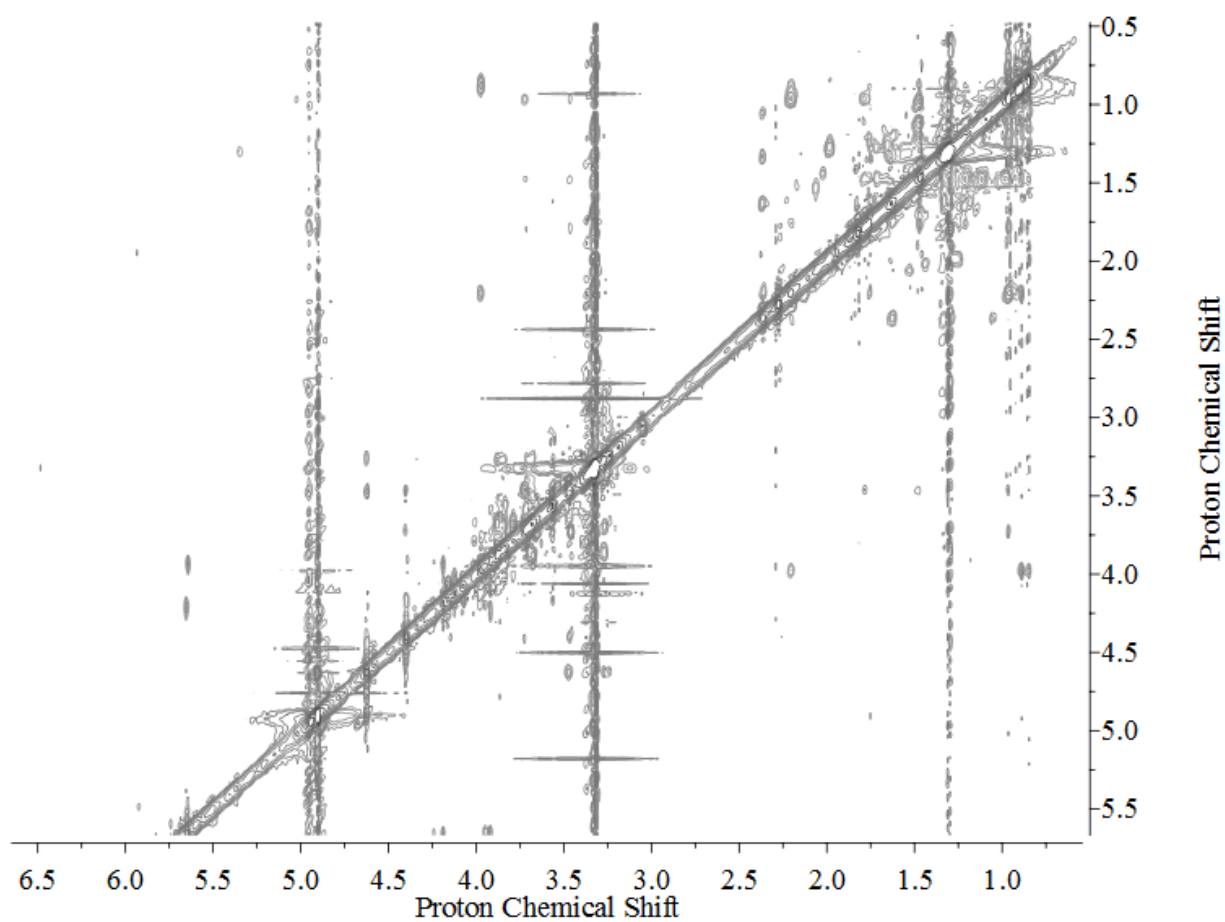


Figure 3.21. 2D-NOESY NMR spectrum of campherenan-2-*endo*,12-diglucoside (Compound **3**; peak 8c in Figure 1).

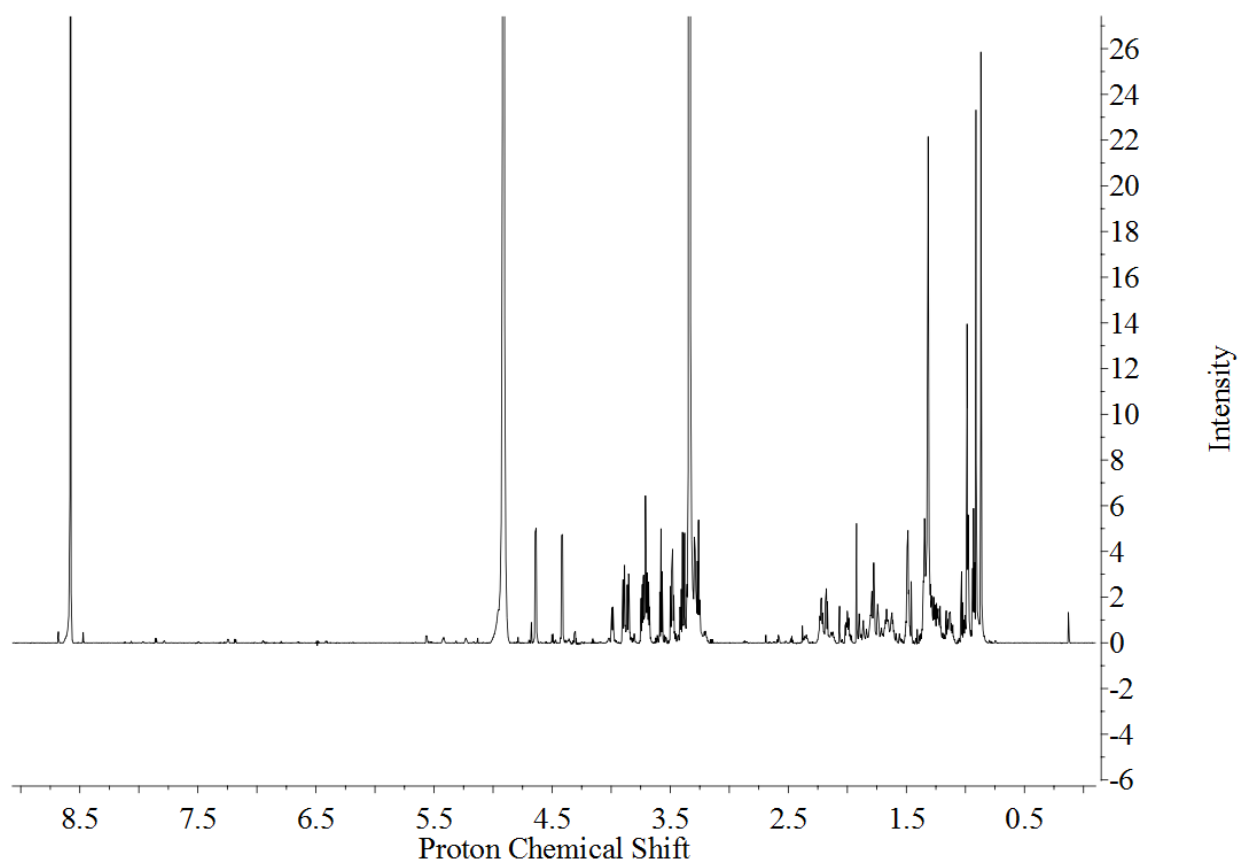


Figure 3.22. Proton NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound **4**; peak 5d in Figure 1).

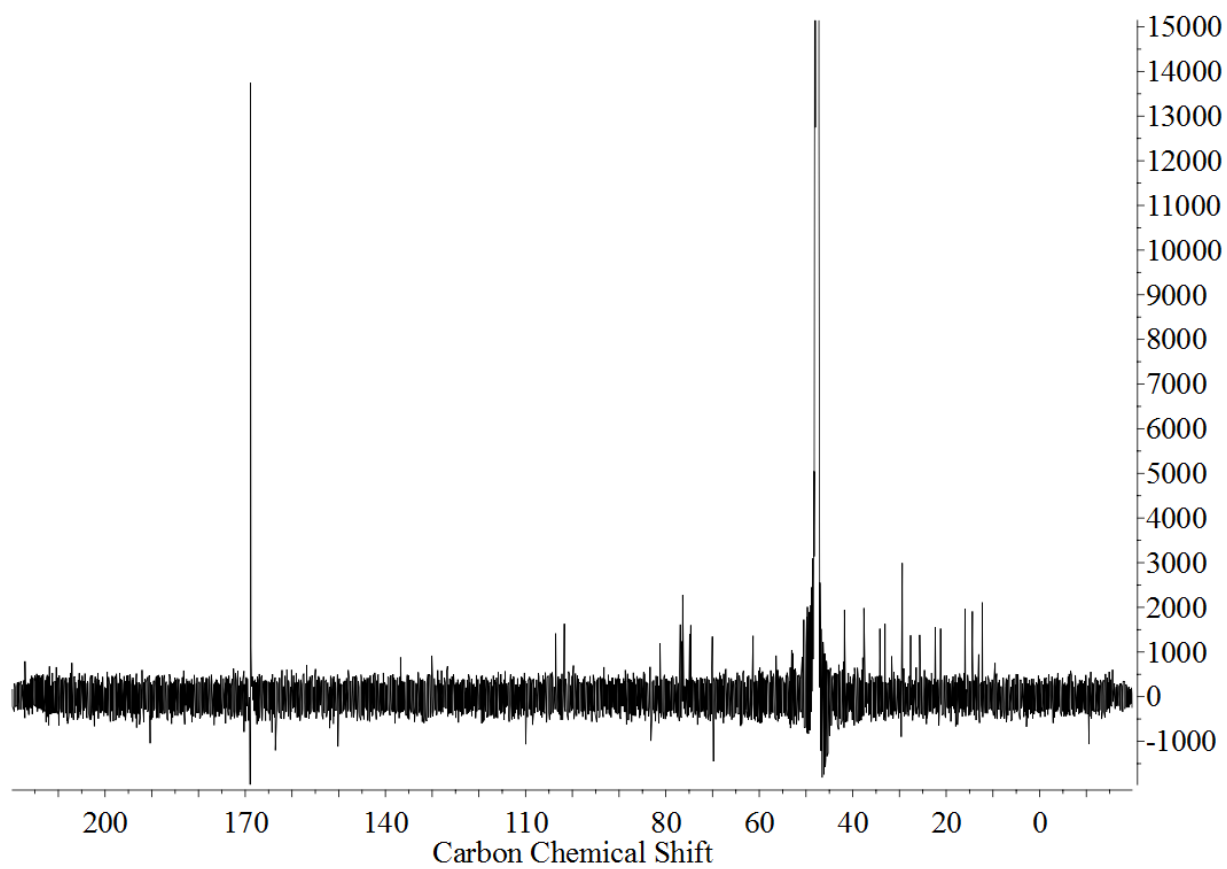


Figure 3.23. <sup>13</sup>C NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound 4; peak 5d in Figure 3.1)



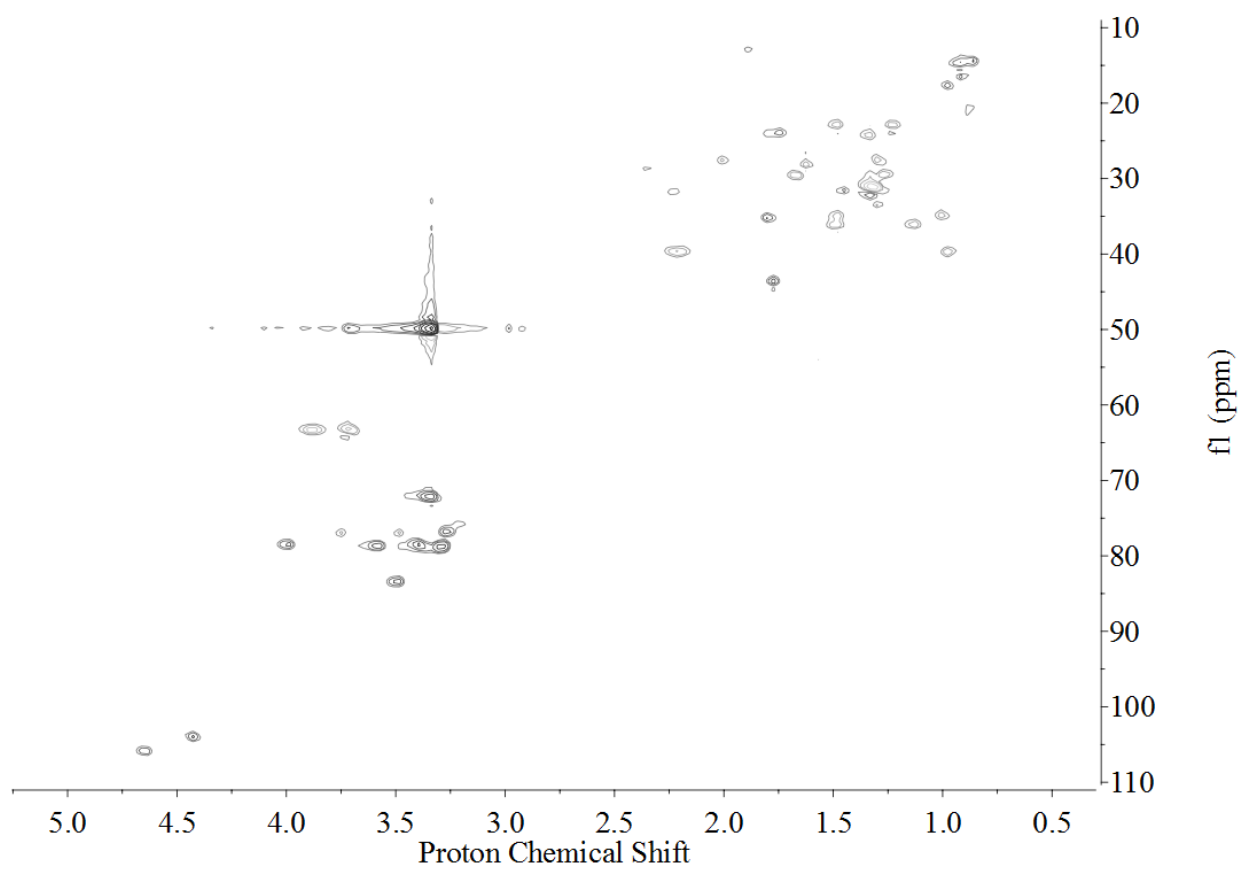


Figure 3.24. Multiplicity edited 2D gHSQC NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound **4**; peak 5d in Figure 1).

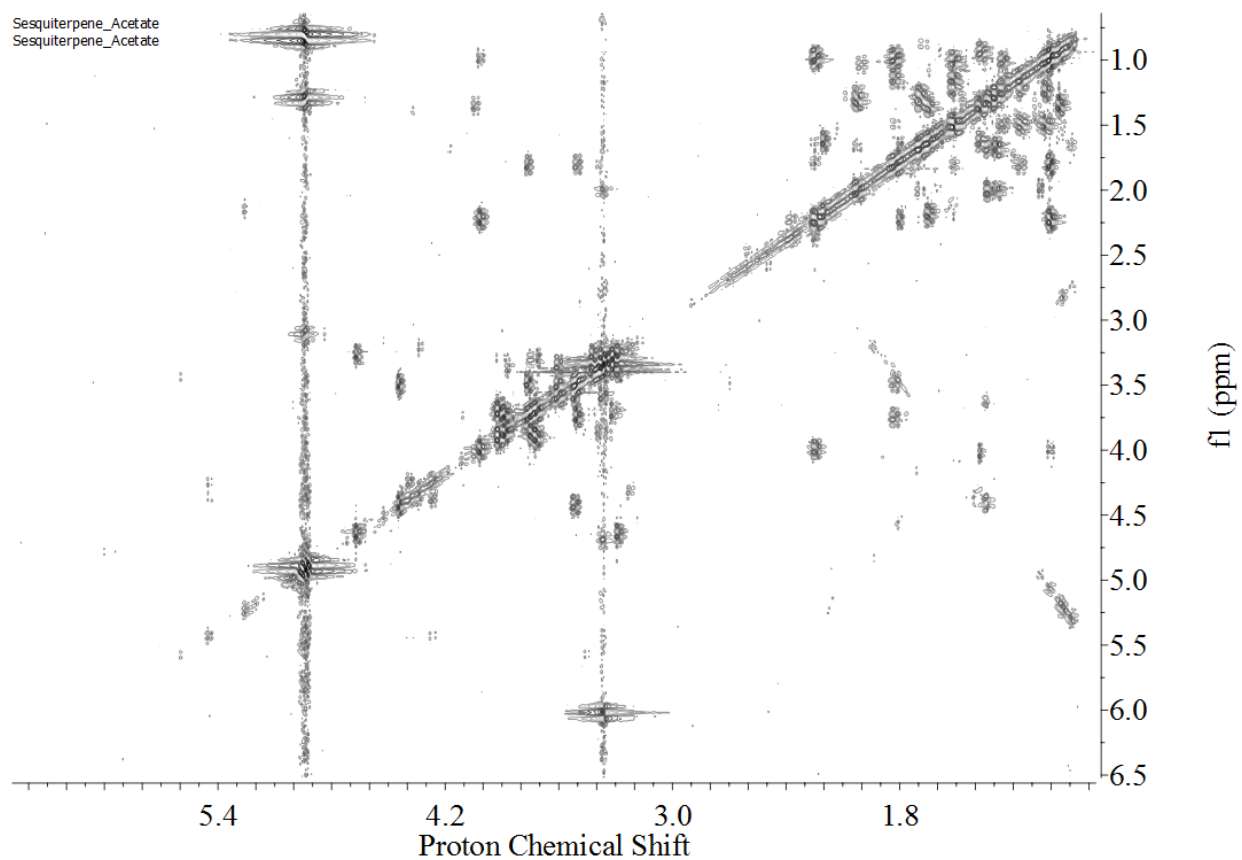


Figure 3.25. 2D gdqCOSY NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound **4**; peak 5d in Figure 1).

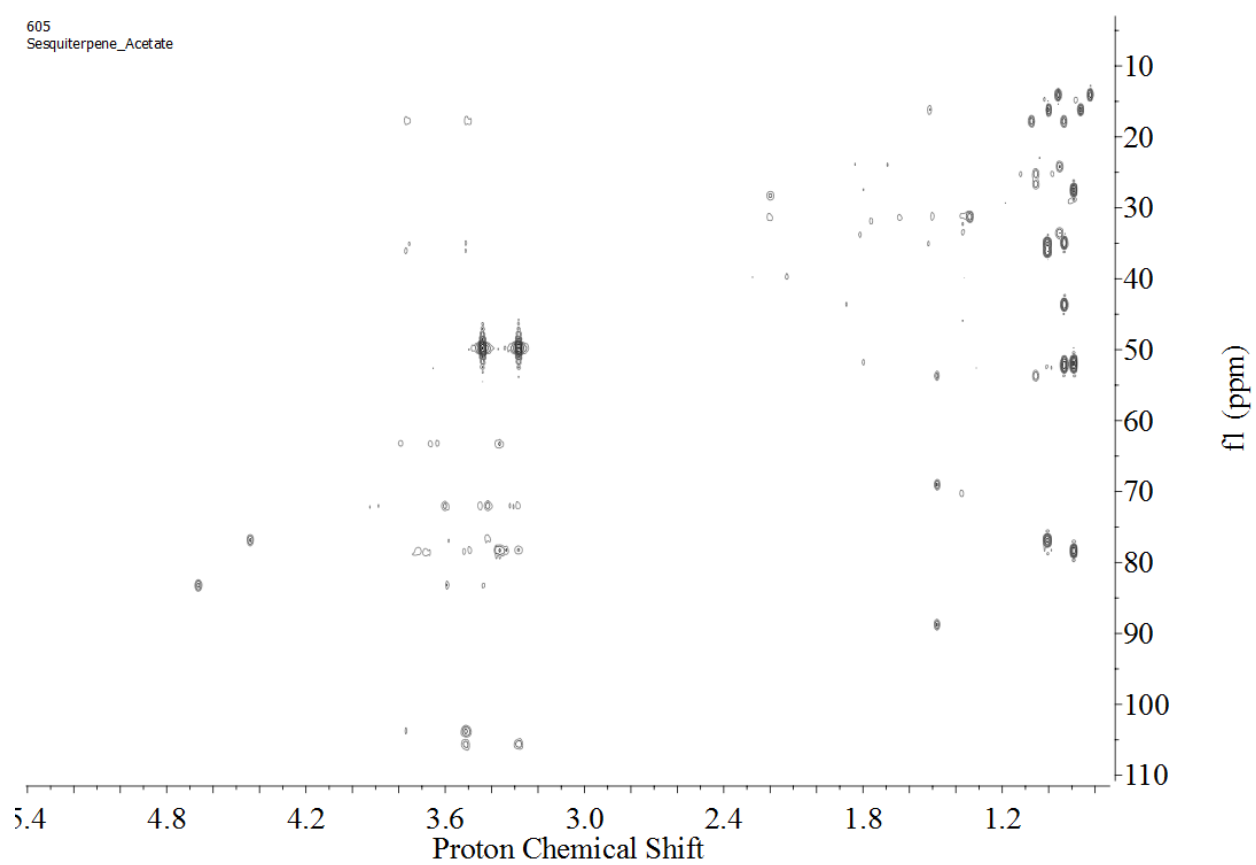


Figure 3.26. 2D HMBC NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound **4**; peak 5d in Figure 1).

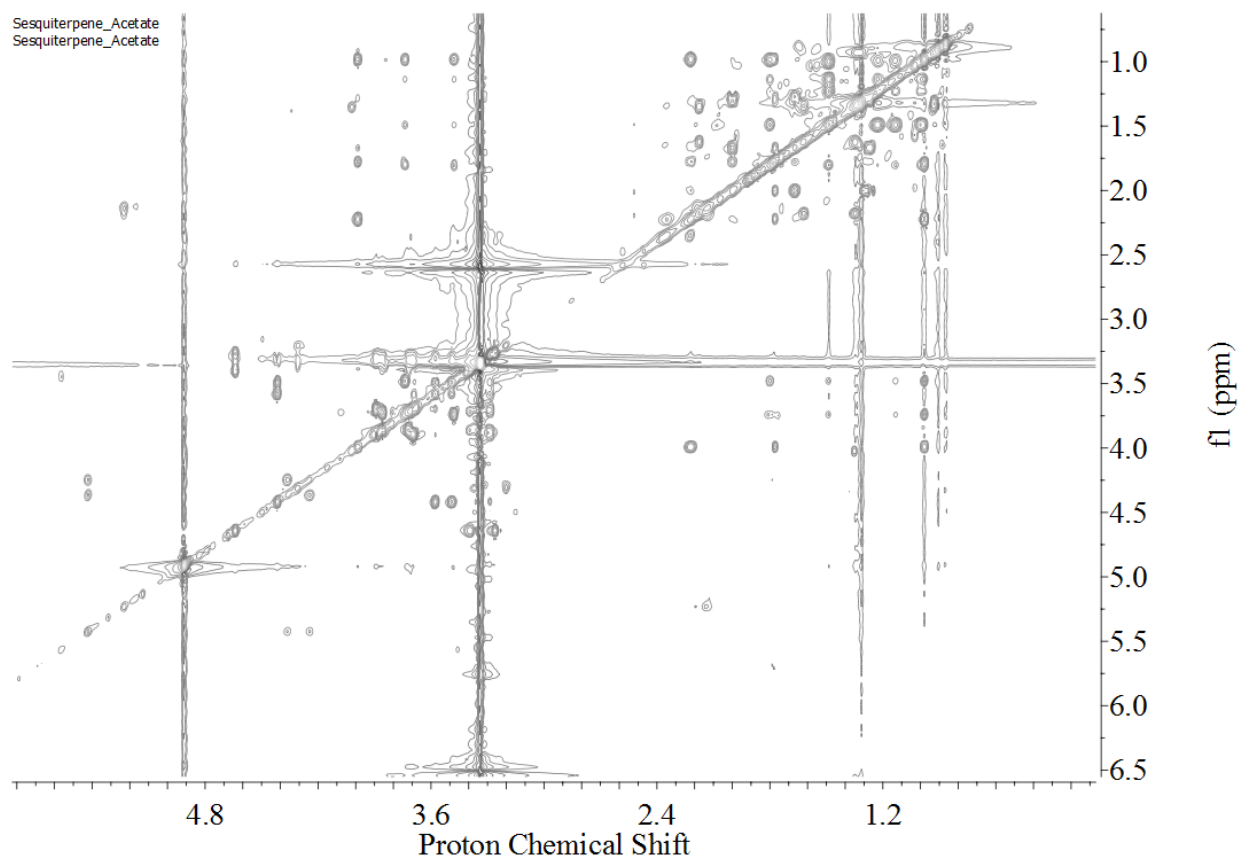


Figure 3.27. 2D TOCSY NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound **4**; peak 5d in Figure 1).

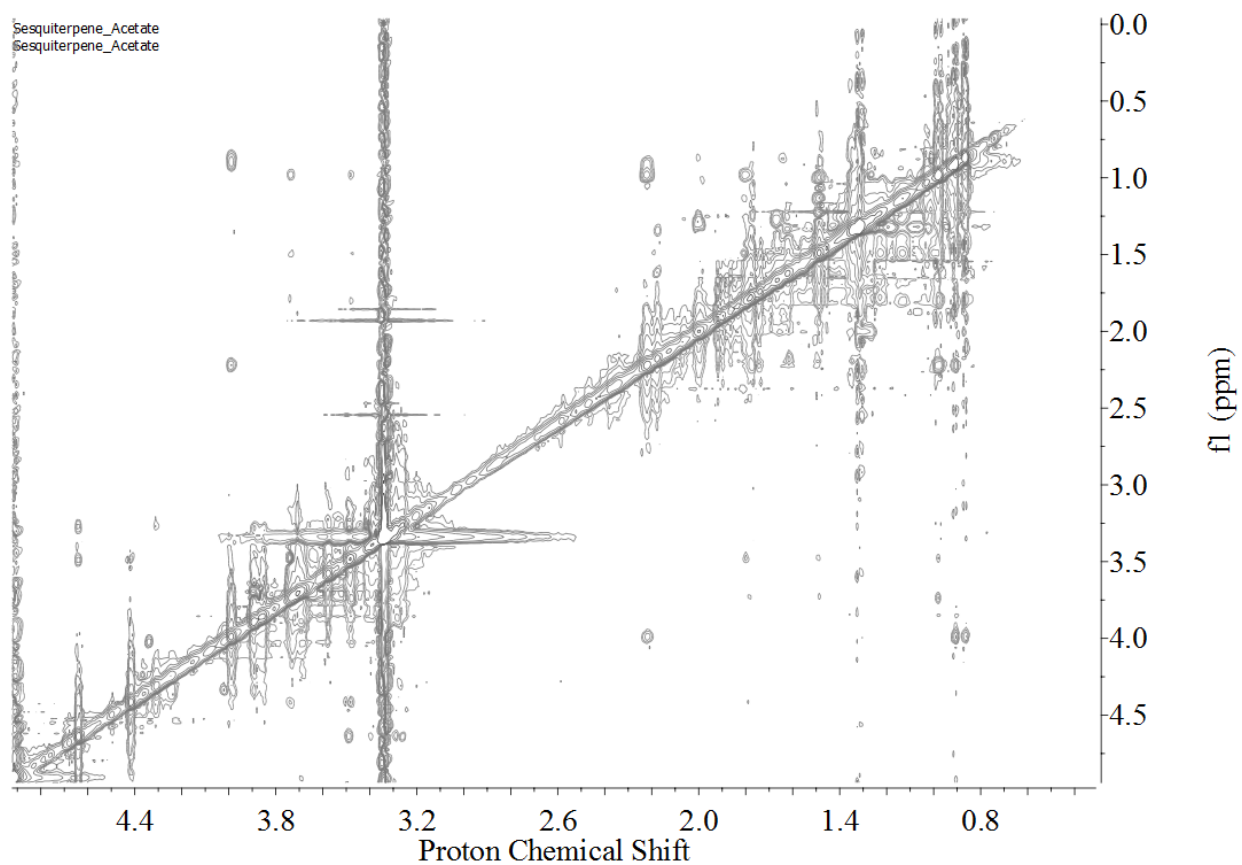


Figure 3.28. 2D NOESY NMR spectrum of 2-*endo*-campherenanol-12-(2-(6''-acetyl)glucosyl)glucoside (Compound **4**; peak 5d in Figure 1).

## REFERENCES

## REFERENCES

1. Ekanayaka EAP, Li C, Jones AD (2014) Sesquiterpenoid glycosides from glandular trichomes of the wild tomato relative *Solanum habrochaites*. *Phytochemistry* 98:223-231. doi:10.1016/j.phytochem.2013.11.011
2. Liu Y-N, Su X-H, Huo C-H, Zhang X-P, Shi Q-W, Gu Y-C (2009) Chemical Constituents of Plants from the Genus *Illicium*. *Chemistry & Biodiversity* 6 (7):963-989. doi:10.1002/cbdv.200700433
3. Coates RM, Denissen JF, Juvik JA, Babka BA (1988) Identification of alpha-santalenoic acid and endo-beta-bergamotenoic acids as moth oviposition stimulants from wild tomato leaves. *Journal of Organic Chemistry* 53 (10):2186-2192. doi:10.1021/jo00245a012
4. Kim TH, Ito H, Hatano T, Takayasu J, Tokuda H, Nishino H, Machiguchi T, Yoshida T (2006) New antitumor sesquiterpenoids from *Santalum album* of Indian origin. *Tetrahedron* 62 (29):6981-6989. doi:http://dx.doi.org/10.1016/j.tet.2006.04.072
5. Alpha T, Raharivelomanana P, Bianchini J-P, Faure R, Cambon A (1997) A sesquiterpenoid from *Santalum austrocaledonicum* var. *austrocaledonicum*. *Phytochemistry* 46 (7):1237-1239. doi:http://dx.doi.org/10.1016/S0031-9422(97)80018-0
6. Ngo K-S, Brown GD (1999) Santalane and isocampherenane sesquiterpenoids from *Illicium tsangii*. *Phytochemistry* 50 (7):1213-1218. doi:http://dx.doi.org/10.1016/S0031-9422(98)00652-9
7. Sallaud C, Rontein D, Onillon S, Jabes F, Duffe P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, Causse M, Tissier A (2009) A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaites*. *Plant Cell* 21 (1):301-317. doi:10.1105/tpc.107.057885
8. Jassbi AR, Zamanizadehnajari S, Kessler D, Baldwin IT (2006) A new acyclic diterpene glycoside from *Nicotiana attenuata* with a mild deterrent effect on feeding *Manduca sexta* larvae. *Zeitschrift Fur Naturforschung Section B-a Journal of Chemical Sciences* 61 (9):1138-1142
9. Herrmann A (2007) Controlled release of volatiles under mild reaction conditions: From nature to everyday products. *Angewandte Chemie-International Edition* 46 (31):5836-5863. doi:10.1002/anie.200700264
10. Goldstein H, Murakami AA, Navarro AL, Ryder DS (2004) Use of glycosides extracted from hop plant parts to flavor malt beverages. Google Patents,

11. Caputi L, Lim EK, Bowles DJ (2008) Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. *Chemistry-A European Journal* 14 (22):6656-6662. doi:10.1002/chem.200800548
  
12. Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, Egholm M, Knight J, Bogden R, Li CB, Shuang Y, Xu X, Pan SK, Cheng SF, Liu X, Ren YY, Wang J, Albiero A, Dal Pero F, Todesco S, Van Eck J, Buels RM, Bombarely A, Gosselin JR, Huang MY, Leto JA, Menda N, Strickler S, Mao LY, Gao S, Tecle IY, York T, Zheng Y, Vrebalov JT, Lee J, Zhong SL, Mueller LA, Stiekema WJ, Ribeca P, Alioto T, Yang WC, Huang SW, Du YC, Zhang ZH, Gao JC, Guo YM, Wang XX, Li Y, He J, Li CY, Cheng ZK, Zuo JR, Ren JF, Zhao JH, Yan LH, Jiang HL, Wang B, Li HS, Li ZJ, Fu FY, Chen BT, Han B, Feng Q, Fan DL, Wang Y, Ling HQ, Xue YBA, Ware D, McCombie WR, Lippman ZB, Chia JM, Jiang K, Pasternak S, Gelley L, Kramer M, Anderson LK, Chang SB, Royer SM, Shearer LA, Stack SM, Rose JKC, Xu YM, Eannetta N, Matas AJ, McQuinn R, Tanksley SD, Camara F, Guigo R, Rombauts S, Fawcett J, Van de Peer Y, Zamir D, Liang CB, Spannagl M, Gundlach H, Bruggmann R, Mayer K, Jia ZQ, Zhang JH, Ye ZBA, Bishop GJ, Butcher S, Lopez-Cobollo R, Buchan D, Filippis I, Abbott J, Dixit R, Singh M, Singh A, Pal JK, Pandit A, Singh PK, Mahato AK, Dogra V, Gaikwad K, Sharma TR, Mohapatra T, Singh NK, Causse M, Rothan C, Schiex T, Noirot C, Bellec A, Klopp C, Delalande C, Berges H, Mariette J, Frasse P, Vautrin S, Zouine M, Latche A, Rousseau C, Regad F, Pech JC, Philippot M, Bouzayen M, Pericard P, Osorio S, del Carmen AF, Monforte A, Granell A, Fernandez-Munoz R, Conte M, Lichtenstein G, Carrari F, De Bellis G, Fuligni F, Peano C, Grandillo S, Termolino P, Pietrella M, Fantini E, Falcone G, Fiore A, Giuliano G, Lopez L, Facella P, Perrotta G, Daddiego L, Bryan G, Orozco M, Pastor X, Torrents D, van Schriek K, Feron RMC, van Oeveren J, de Heer P, daPonte L, Jacobs-Oomen S, Cariaso M, Prins M, van Eijk MJT, Janssen A, van Haaren MJJ, Jo SH, Kim J, Kwon SY, Kim S, Koo DH, Lee S, Hur CG, Clouser C, Rico A, Hallab A, Gebhardt C, Klee K, Jocker A, Warfsmann J, Gobel U, Kawamura S, Yano K, Sherman JD, Fukuoka H, Negoro S, Bhutty S, Chowdhury P, Chattopadhyay D, Datema E, Smit S, Schijlen EWM, van de Belt J, van Haarst JC, Peters SA, van Staveren MJ, Henkens MHC, Mooyman PJW, Hesselink T, van Ham R, Jiang GY, Droege M, Choi D, Kang BC, Kim BD, Park M, Kim S, Yeom SI, Lee YH, Choi YD, Li GC, Gao JW, Liu YS, Huang SX, Fernandez-Pedrosa V, Collado C, Zuniga S, Wang GP, Cade R, Dietrich RA, Rogers J, Knapp S, Fei ZJ, White RA, Thannhauser TW, Giovannoni JJ, Botella MA, Gilbert L, Gonzalez R, Goicoechea JL, Yu Y, Kudrna D, Collura K, Wissotski M, Wing R, Schoof H, Meyers BC, Gurazada AB, Green PJ, Mathur S, Vyas S, Solanke AU, Kumar R, Gupta V, Sharma AK, Khurana P, Khurana JP, Tyagi AK, Dalmay T, Mohorianu I, Walts B, Chamala S, Barbazuk WB, Li JP, Guo H, Lee TH, Wang YP, Zhang D, Paterson AH, Wang XY, Tang HB, Barone A, Chiusano ML, Ercolano MR, D'Agostino N, Di Filippo M, Traini A, Sanseverino W, Frusciante L, Seymour GB, Elharam M, Fu Y, Hua A, Kenton S, Lewis J, Lin SP, Najjar F, Lai HS, Qin BF, Qu CM, Shi RH, White D, White J, Xing YB, Yang KQ, Yi J, Yao ZY, Zhou LP, Roe BA, Vezzi A, D'Angelo M, Zimbello R, Schiavon R, Caniato E, Rigobello C, Campagna D, Vitulo N, Valle G, Nelson DR, De Paoli E, Szinay D, de Jong HH, Bai YL, Visser RGF, Lankhorst RMK, Beasley H, McLaren K, Nicholson C, Riddle C, Gianese G, Tomato Genome C (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485 (7400):635-641. doi:10.1038/nature11119



13. Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome* 43 (5):803-813. doi:10.1139/g00-043

## **Chapter 4: Purification and structure elucidation of sesquiterpene I and sesquiterpene II glycosides from leaf glandular trichomes of *Solanum habrochaites* LA1777**

### **4.1 Introduction**

A systematic investigation of methanol leaf dip extracts of *Solanum habrochaites* LA1777 plant was performed in order to purify and elucidate structures of the sesquiterpene I and II glycosides that were annotated using relative mass defect filtering in Chapter 2. Assigning the structure of these compounds is an important first step in the road to the discovery of, biosynthetic precursors, enzymes involved in biosynthesis and the genes regulating the expression of the enzymes. This becomes an even more important endeavor for compounds that display structural modifications such as cyclizations, shifts in methyl group positions, glycosylations, and acylations. This is because such modifications in structure are suggestive that synthesis of these compounds involves a number of additional reaction steps involving additional enzymes, and thus exploring these compounds can lead to the discovery of novel biosynthetic pathways or chemical precursors. In establishing the structures of compounds, identifying the stereochemistry (relative or absolute) of the chiral centers as much as possible is important for establishing the identity of the precursors. As discussed in chapter 3 of this dissertation, it has been established that depending on the stereochemistry of the precursor, the end product also varies [1]. Furthermore, the biological activity of two enantiomers may not be the same due to enzyme specificity toward one enantiomer than the other[2].

Establishing the structure of these compounds requires the purification and obtaining the NMR data for establishing the structure. The purification of these compounds involved the use of different chromatographic conditions than the campherenane derivative purification. Two

purification cycles had to be performed at two different pH conditions until pure compound could be obtained.

Linear diterpene glycosides with mild deterrent effect on feeding of *Manduca sexta* larvae have been reported from *Nicotiana attenuata* [3]. Similarly, an acyclic sesquiterpene glycoside is reported from *Sapindus trifolzatus* [4] and a diglycosylated farnesol is reported from *Guioa crenulata* [5]. The compound reported from *Guioa crenulata* has acyl groups on the sugar moieties. Recently, four acyclic sesquiterpene glycosides with inhibitory effects on tumor necrosis factor- $\alpha$ -induced cytotoxicity were discovered from pericarps of *Sapindus rarak* [6]. Three bicyclic sesquiterpene glycosides are reported from *Elsholtzia bodinieri* [7]. Two of the three compounds reported herewith are bicyclic sesquiterpenes and the other is an acyclic sesquiterpene glycoside. All compounds carry two sugar groups attached and in the acyclic compound a malonate ester is identified. In contrast with the compounds with malonylation observed so far, this compound has the malonylation on a hydroxyl group attached to the terpenoid core of the molecule. In previously reported compounds, the malonylation or acetylation is observed on sugar group.

#### **4.2 Plant material for metabolite purification**

Five *Solanum habrochaites* LA1777 plants were grown in a plant growth chamber (28 °C, 16 hour day time, 96% humidity, lighting 150  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ), and leaf and stem tissues were harvested and extracted in six week intervals. Leaf trichomes were extracted by dipping each branch in 1 L of 100% methanol for about 30 s. After extraction, plant material was dried at 60 °C and dry mass was determined to be 183 g. Extracts were subsequently combined, and solvent was evaporated under vacuum on a rotary evaporator without heating, and the residue was

redissolved in about 10 ml of 100% methanol. Next, the three compounds were purified from this concentrated extract.

#### **4.3 Purification of sesquiterpene I and sesquiterpene II glycosides**

Semipreparative HPLC was performed on an Acclaim 120 C18 column ( $4.6 \times 150$  mm, 5  $\mu$ m particles; Dionex Co. USA) using a Waters 2795 semipreparative HPLC system. Fractions eluted from the column were collected using a LKB BROMMA 221 Superrac fraction collector. For the semipreparative HPLC, flow rate was 2.0 ml/min and the injection volume for each injection was 150  $\mu$ L. Purification of these three compounds required employing two different gradients in two chromatographic cycles. The first linear solvent gradient used for the purification employed water with 0.15% formic acid of pH 2.90 (solvent A) and acetonitrile (solvent B) was as follows (A:B): initial (99%:1%), 1.00 min (70%:30%), 31.00 min (53%:47%), 31.01 min (0%:100%), 33.00 min (0%:100%), 33.01 min (99%:1%), 36.00 min (99%:1%). Twenty fractions were collected (1-min intervals), and each fraction was tested for the presence of compound of interest using a 5.50 min UHPLC-MS method in negative ion mode ESI using a Waters LCT Premier TOF mass spectrometer. Instrument parameters were: capillary voltage of 3.0 kV, desolvation temperature of 350 °C, source temperature of 100 °C, Cone gas (N<sub>2</sub>) at 40 L/hr and desolvation gas (N<sub>2</sub>) at 350 L/hr. This gradient was (A = 0.15% aqueous formic acid; B = methanol (%A: %B): 0.0-2.0 min (95:5), step at 2.01 min (40:60), linear increase to 3.00 min (0: 100), and held at this condition until 5.0 min, then returned to initial condition at 5.01 min and held until 5.5 min. From these analyses, the fractions that contained the compound of interest were combined and concentrated to about 5 ml volume using Speedvac without heating. The concentrated sample was again chromatographed using the same gradient, however now water with 10 mM ammonium formate pH 7.50 (solvent A) and acetonitrile (solvent B) were used as solvents.

Again 20 fractions were collected and the fractions were analyzed using the 5.5 min LC-MS procedure above described on TOF mass spectrometer. From these analyses, the fractions that were over 90% pure based on total ion chromatogram peak areas were identified and combined, and solvent was removed by evaporating to dryness under a N<sub>2</sub> gas flow. All NMR experiments were performed as explained before [8]. Also the acid hydrolysis experiments and the chemical derivatizations of the sugars obtained from the hydrolysis was performed on compounds purified from peaks 8a, 3b and 6c of Fig. 4.1 as explained before [8].

#### 4.4 Structure elucidation of sesquiterpene I diol dihexoside malonate ester

This compound is observed at  $m/z$  661 (observed  $m/z$  661.2693; theoretical  $m/z$  for [C<sub>30</sub>H<sub>45</sub>O<sub>16</sub>]<sup>-</sup> 661.2713;  $\Delta m$  = -3.0 ppm ) as [M-H]<sup>-</sup>. The MS/MS product ion spectrum of this compound displays a neutral loss of 44 Da (corresponding to CO<sub>2</sub>) followed by a neutral loss of 42 Da (corresponding to C<sub>2</sub>H<sub>2</sub>O) giving fragment ions  $m/z$  617 (observed  $m/z$  617.2791; theoretical  $m/z$  for [C<sub>29</sub>H<sub>45</sub>O<sub>14</sub>]<sup>-</sup> 661.2815;  $\Delta m$  = -3.9 ppm) and 575 (observed  $m/z$  575.2739; theoretical  $m/z$  for [C<sub>27</sub>H<sub>43</sub>O<sub>13</sub>]<sup>-</sup> 575.2709;  $\Delta m$  = -5.2 ppm ). This is followed by the two sequential neutral losses of 162 Da (corresponding to two anhydroglucoses) to give fragment ions  $m/z$  413 (observed  $m/z$  413.2166; theoretical  $m/z$  for [C<sub>21</sub>H<sub>33</sub>O<sub>8</sub>]<sup>-</sup> 413.2181;  $\Delta m$  = -3.6 ppm) and 251 (observed  $m/z$  251.1650; theoretical  $m/z$  [C<sub>15</sub>H<sub>23</sub>O<sub>3</sub>]<sup>-</sup> 251.1653;  $\Delta m$  = -1.2 ppm ). Based on accurate mass of the putative terpenoid core elemental formula of C<sub>15</sub>H<sub>23</sub>O<sub>3</sub><sup>-</sup> was assigned. To further characterize the structure of this compound it was purified using semi-preparative HPLC followed by 1D and 2D NMR spectroscopic analysis (Fig. 4.3 to 4.8). Structure shown in Fig. 4.2a was assigned.

About 8 isomers of this compound were separated from *S. habrochaites* LA1777 and one of the major isomers (Fig. 4.1, peak 8a. R<sub>t</sub> = 7.1 min) was purified and structure was elucidated using

1D and 2D NMR. Final structure shown in Fig. 4.2a was assigned to this molecule using NMR and mass spectrometry information. In the proton NMR spectrum the anomeric protons corresponding to the two glucose moieties could be identified at proton resonances  $\delta_{\text{H}}$  5.56 ppm ( $J = 7.81$  Hz),  $\delta_{\text{H}} = 4.31$  ppm ( $J = 7.80$  Hz) as two doublets. This confirmed the presence of two hexose moieties and the coupling constants of each anomeric proton suggest that the glycosides are bound by beta-substitutions. Based on the correlations observed in the HMBC spectrum, it was concluded that the two sugar groups are connected by a 1-4 glycosidic bond. Three methyl groups were identified as singlets from proton NMR spectra and were confirmed by HSQC spectra. This indicated that the methyl groups were attached to quaternary carbons. COSY and HMBC spectra were used to identify the neighboring protons. Next, starting from the anomeric protons, the two hexose moieties were assigned followed by the assignment of the structure of terpenoid core. In the terpenoid core, the protons located on alkene groups and the two primary alcohol groups were distinctive. The alkene protons were detected at  $\delta_{\text{H}}$  5.41 ppm and  $\delta_{\text{H}}$  6.93 ppm. The two primary alcohol positions hosts two protons each. ( $\delta_{\text{H}}$  4.22 and 4.33) at C12 and ( $\delta_{\text{H}}$  3.99 and 3.99) ppm at C1. The correlations observed in the HMBC spectrum between the C12 protons and the anomeric carbon (C16) of the sugar moiety aided assignment of the position of the sugar moiety. Similarly the correlations of protons on C2 and carbonyl carbon at C28 enabled the identification of the position of the malonyl group. Furthermore, the correlations observed between the protons of methyl groups and the alkene carbons enabled the correct assignment of the positions of the methyl groups.

Based on the correlations observed in the NOE spectrum, the relative stereochemistry of methyl groups was established. NOE spectrum showed correlations between the proton at C11 and

methyl group at C15 (attached to C10) indicating these two groups are located in close proximity in space. Therefore these are located in a *cis* configuration at the C-C double bond between C10 and C11. Similar correlations were observed in the NOE spectrum between the proton at C8 and the methyl group attached to C7 indicating these are also located in a *cis* configuration. Additionally the two methyl groups (methyl group attached to C7 and C5 respectively) also displayed correlations in the NOE spectrum. However, the proton at C4 and the methyl group at C5 showed no NOE correlation, indicating that these two are located in a *trans* configuration at the double bond between C4 and C5.

#### 4.5 Structure elucidation of Sesquiterpene II alcohol dihexoside acetate ester

This compound is observed at  $m/z$  633 (observed  $m/z$  633.3115; theoretical  $m/z$  for  $[C_{29}H_{48}O_{12} + HCOO^-]$  633.3128;  $\Delta m = 2.0$  ppm) as  $[M + HCOO]^-$ . The compound displays a neutral loss of 42 Da (corresponding to  $C_2H_2O$ ) giving fragment ion  $m/z$  545 (observed  $m/z$  545.2923; theoretical  $m/z$  for  $C_{27}H_{45}O_{11}$  545.2967;  $\Delta m = 8.0$  ppm) which is followed a neutral loss of 162 Da (corresponding to an anhydroglucose) to give fragment ion  $m/z$  383 (observed  $m/z$  383.2429; theoretical  $m/z$  for  $[C_{21}H_{35}O_6]^-$  383.2439;  $\Delta m = 2.6$  ppm). To further characterize the structure of this compound it was purified using semi-preparative HPLC followed by 1D and 2D NMR spectroscopic analysis (Fig. 4.9 to 4.13). Proposed structure is shown in Fig. 4.2b.

Four isomers of this compound were separated from *S. habrochaites* LA1777 and one of the major isomers (Fig. 4.1, peak 3b.  $R_t = 14.9$  min) was purified and structure was elucidated using 1D and 2D NMR. Final structure shown in Fig. 4.2b was assigned to this molecule using NMR and mass spectrometry information. In the proton NMR spectrum two anomeric protons corresponding to the two glucose moieties could be identified at proton resonances  $\delta_H$  4.65 ppm

( $J = 7.96$  Hz),  $\delta_{\text{H}} = 4.67$  ppm ( $J = 7.69$  Hz) as two doublets. This supported the presence of two hexose moieties (in contrast with the LC-MS/MS spectra that showed one neutral loss of 162 Da corresponding to only one anhydroglucose). In contrast to Sesquiterpene I diol dihexoside malonate ester discussed above, correlations observed in the HMBC spectrum of Sesquiterpene II alcohol dihexoside acetate ester indicated that the two sugar groups are attached through a 1-6 glycosidic bond. The coupling constants of each anomeric proton suggest that the glycosides are bound by beta-substitutions. Four methyl groups were identified from the proton NMR spectrum and were confirmed by HSQC. In the terpenoid core, the proton located on alkene group (detected at  $\delta_{\text{H}}$  6.76 ppm) was distinctive. COSY and HMBC spectra were used to identify the neighboring protons. The HMBC correlations between the alkene carbon and the tertiary alcohol carbon were instrumental in establishing the structure of the terpenoid core. Next, starting from the anomeric protons, the two hexose moieties were assigned. The structure of the terpenoid core was assigned starting from the alkene proton at C2. Using the H-H correlations observed in the COSY spectrum most of the terpenoid core could be established. (Fig. 4.2b) This assignment was confirmed by the presence of distinctly identifiable proton spin systems in the TOCSY spectrum, which revealed longer range H-H correlations. Therefore, this serves as supporting evidence to confirm structural assignments made using COSY. In addition, the correlations observed in HMBC supported the assigned terpenoid core structure.

Relative stereochemistry was assigned using NOE spectra. The methyl group and the proton attached to C1 and C2 positions display correlations in NOE spectra indicating they are at a *cis* configuration. One of the two protons at C6 also shows NOE correlations with methyl at C1 and the proton at C5 show correlations with proton at C6. Indicating that these groups (protons at C6,



C5, C2 and methyl at C1) are in same plane. However, no NOE correlations were observed between the protons at C5 and C4 and it was concluded that these are located on opposite planes.

To confirm the  $m/z$  of the terpenoid core, acid hydrolysis was performed on the purified compound followed by GC-MS analysis. The GC-MS indicated a peak appearing to be a  $[M-H_2O]^+$ . The  $m/z$  of this fragment ion was 204 (observed  $m/z$  204.1862; theoretical  $m/z$  for  $[C_{14}H_{20}O]^+$  204.1878;  $\Delta m = 7.8$  ppm) (Fig. 4.20)

#### 4.6 Structure elucidation of Sesquiterpene II alcohol dihexoside

This compound is observed at  $m/z$  591 (observed  $m/z$  591.3013; theoretical  $m/z$  for  $[C_{27}H_{46}O_{11} + HCOO]^-$  591.3022;  $\Delta m = 1.5$  ppm) as  $[M+HCOO]^-$ . The MS/MS spectra of the compound displays  $[M-H]^-$  ion at  $m/z$  545 (observed  $m/z$  545.2947; theoretical  $m/z$  591.2967;  $\Delta m = 3.7$  ppm), which is followed by a neutral losses of 162 Da (corresponding to two anhydroglucoses) to give fragment ions  $m/z$  383. To further characterize the structure of this compound it was purified using semi-preparative HPLC followed by 1D and 2D NMR spectroscopic analysis and the structure was established (Fig. 4.14 to 4.19 and 4.2c).

Eight isomers of this compound were separated from *S. habrochaites* LA1777 and one of the major isomers (Fig. 4.1, peak 6c.  $R_t = 10.7$  min) was purified and structure was elucidated using 1D and 2D NMR. Final structure shown in Fig. 4.2c was assigned to this molecule using NMR and mass spectrometry information. In the proton NMR spectra two anomeric protons corresponding to the two glucose moieties could be identified at  $\delta_H$  4.66 ppm (doublet with  $J = 8.10$  Hz) and  $\delta_H = 4.63$  ppm (triplet of  $J = 7.92$  Hz). This supported the presence of two hexose moieties. Four methyl groups were identified from proton NMR spectra based on integrals and

were confirmed by HSQC spectra. The proton at alkene group was detected at  $\delta_{\text{H}} = 5.89$  ppm (triplet of  $J = 3.13$  Hz). COSY and HMBC spectra were used to identify the protons and carbon atoms neighboring the methyl groups and the alkene proton in the structure. Also the HMBC correlations between the tertiary carbon hosting the  $-\text{OH}$  and the neighboring protons were important in elucidating the structure. Starting from the anomeric protons, the two hexose moieties were assigned. Similar to the sesquiterpene alcohol dihexoside acetate ester the structure of the terpenoid core was assigned using COSY, TOCSY, HMBC, HSQC and NOE spectra. Based on this NMR evidence, the structure shown in Fig. 4.2c was assigned to this compound.

Similar to the sesquiterpene alcohol dihexoside acetate ester, to confirm the  $m/z$  of the terpenoid core, acid hydrolysis was performed on the purified compound followed by GC-MS analysis. The GC-MS indicated a peak at  $m/z$  204 (observed  $m/z$  204.1862; theoretical  $m/z$  for  $[\text{C}_{14}\text{H}_{20}\text{O}]^+$  204.1878;  $\Delta m = 7.8$  ppm) corresponding to  $[\text{M}-\text{H}_2\text{O}]^{++}$  (Fig. 4.21). This observation further supported the structure established using NMR.

## 4.7 Conclusion

Our efforts have resulted in the discovery of another two novel terpenoid cores that were not reported from *Solanum habrochaites* before. While these terpenoid compounds possess structural similarity to farnesol and naphthalenol, their presence in *S. habrochaites* LA1777 presents the opportunity for discovering the genes involved in their synthesis. Furthermore, the GC-MS profiling of volatile terpenoids from *S. habrochaites* reported in literature does not report of any compound that similar in structure to these. This is an important implication in that the non-volatile terpenoids in these plants presents very different structural characteristics from the volatile terpenoids. The proposed structures indicate that these compounds possess a number of chiral centers. Therefore, it can be expected that a number of stereoisomers of these compounds would be present. Our metabolite profiling work reported here (Fig. 4.1) and chapter 2 of this dissertation indicated that for all of these compounds multiple isomeric forms are detected. It can be perceived that some of the isomers could be the result of stereochemical differences in the terpenoid core.

## **APPENDIX**

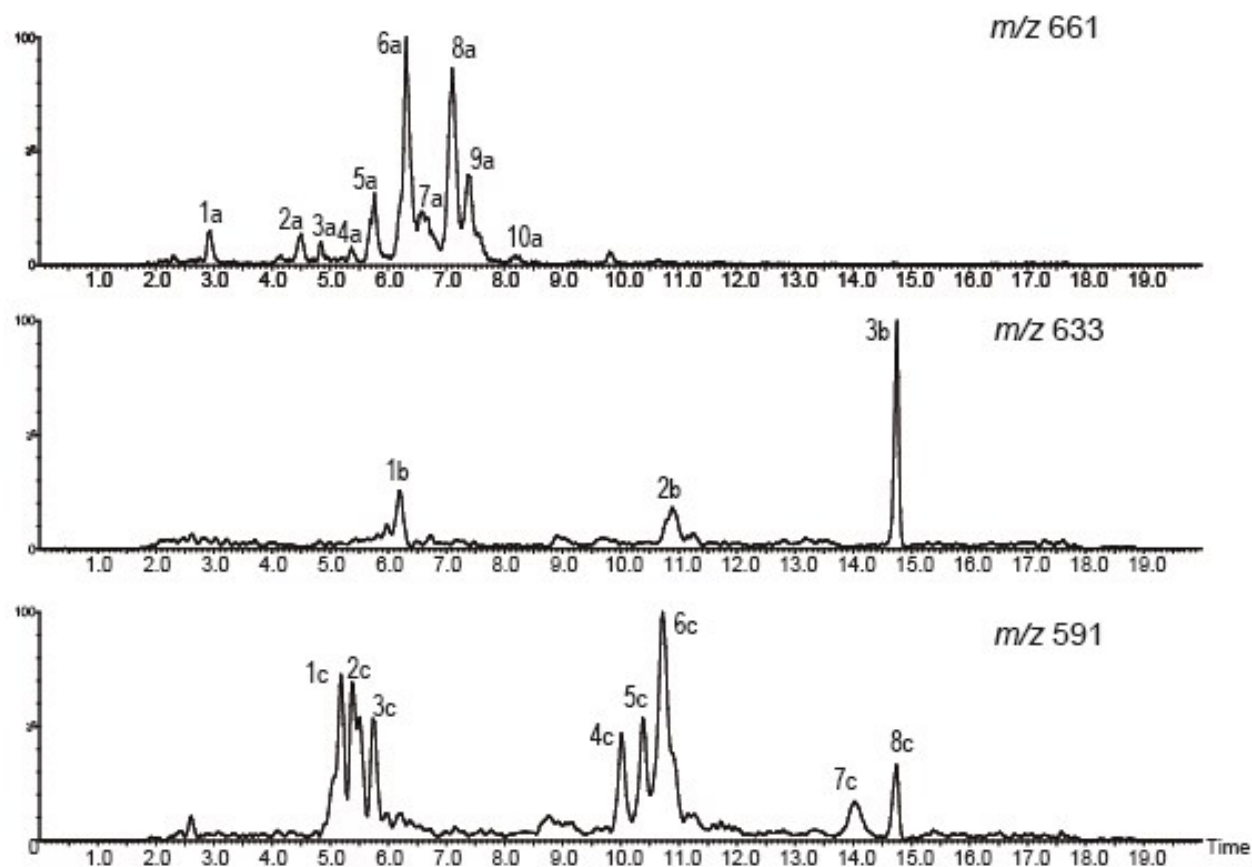


Figure 4.1: HPLC-MS extracted ion chromatogram profiles of the three compounds purified from *S. habrochaites* LA1777. a.) 10 isomers of Sesquiterpene I diol dihexoside malonate ester were separated (b) three isomers of  $[M+\text{formate}]^-$  of Sesquiterpene II alcohol dihexoside acetate ester were detected (c) 8 isomers of  $[M+\text{formate}]^-$  of Sesquiterpene II alcohol dihexoside were separated.

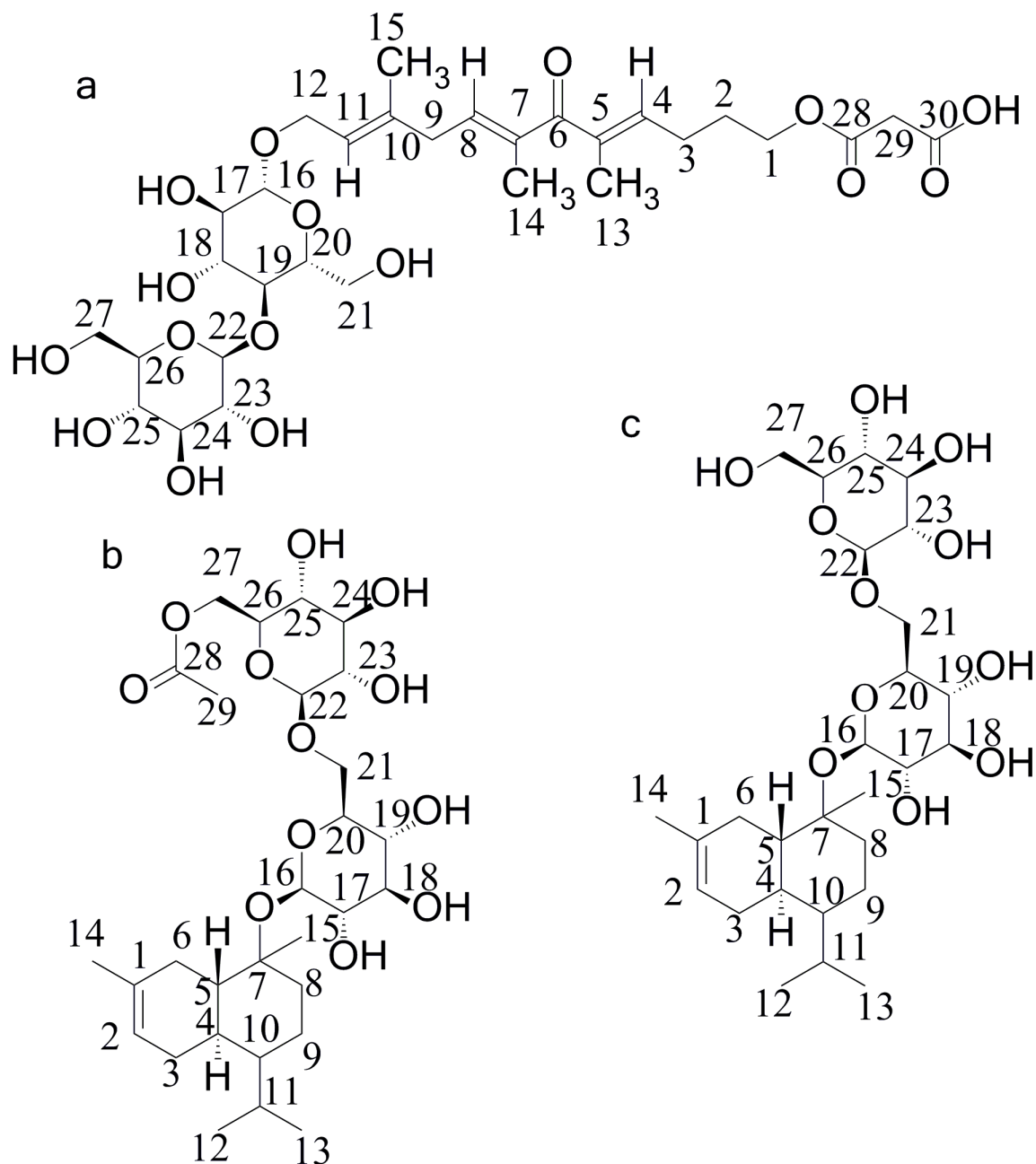


Figure 4.2. Structures of (a) Sesquiterpene I diol dihexoside malonate ester (peak 8a from Figure 4.1), (b) Sesquiterpene II alcohol dihexoside acetate ester (peak 3b) and (c) Sesquiterpene II alcohol dihexoside (peak 6c), determined using NMR and tandem mass spectrometry. Portions of the molecule corresponding to key fragment ions in negative ion MS/MS spectra are illustrated on each structure. Carbon atoms are numbered in accordance with NMR assignments in Table 4.1.

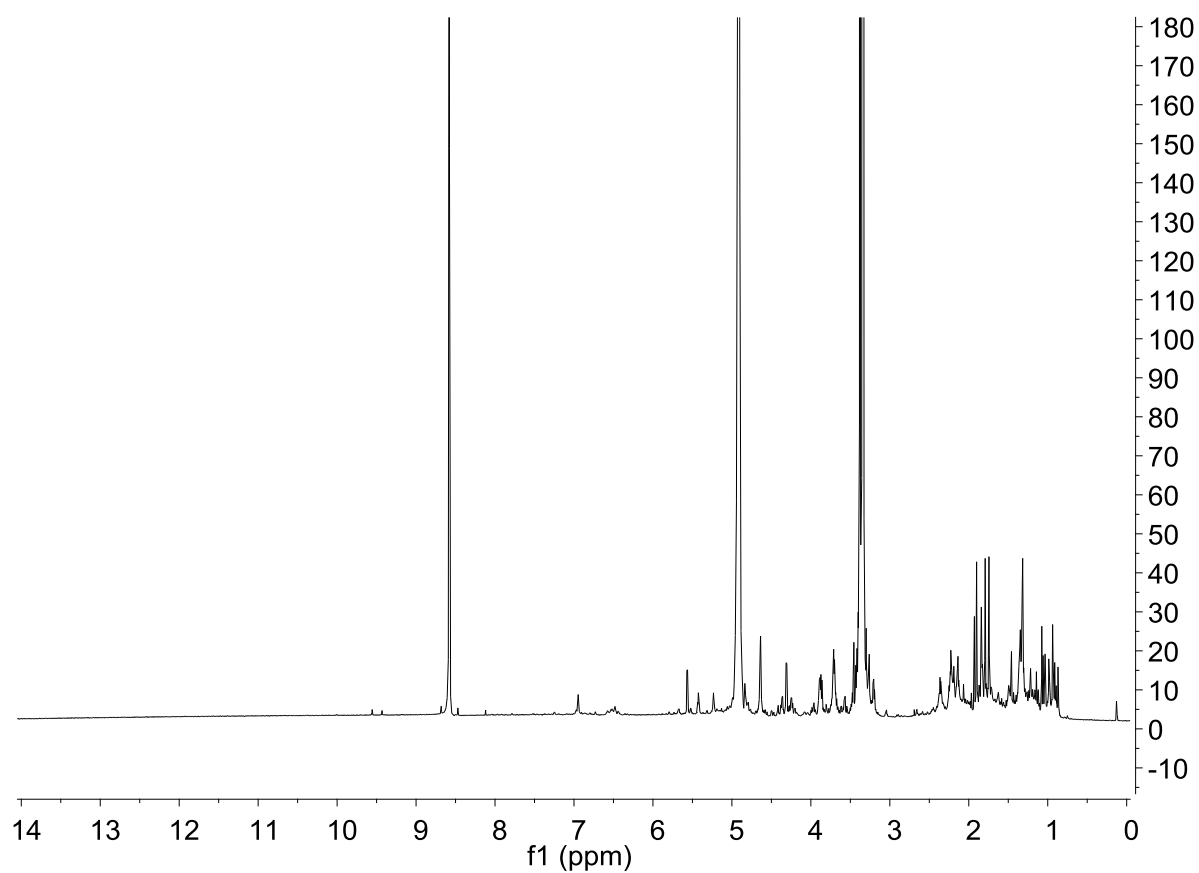


Figure 4.3. 1D Proton NMR spectra of compound a (peak 8a from Figure 4.1).

Figure 4.4. 2D COSY spectrum of compound a (peak 8a from Figure 4.1).

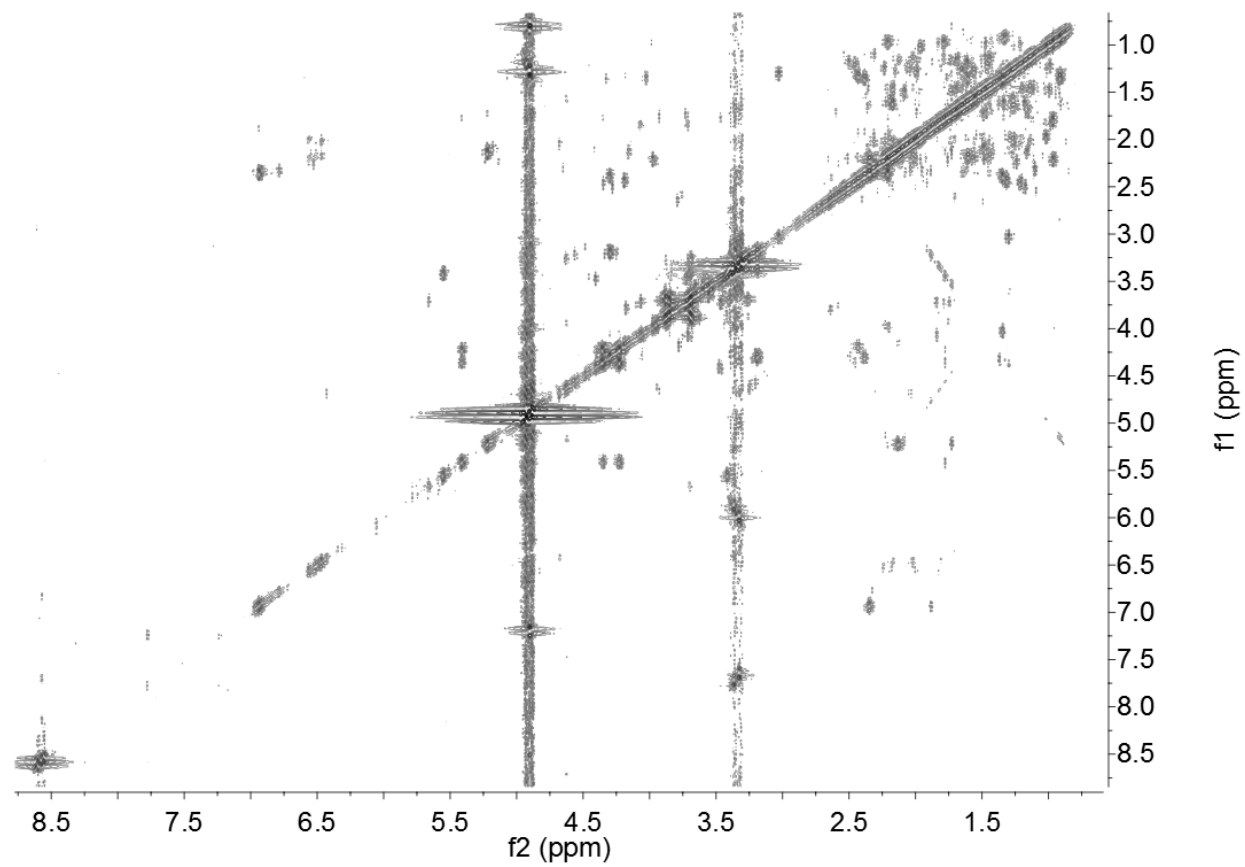


Figure 4.4. 2D COSY spectrum of compound a (peak 8a from Figure 4.1).



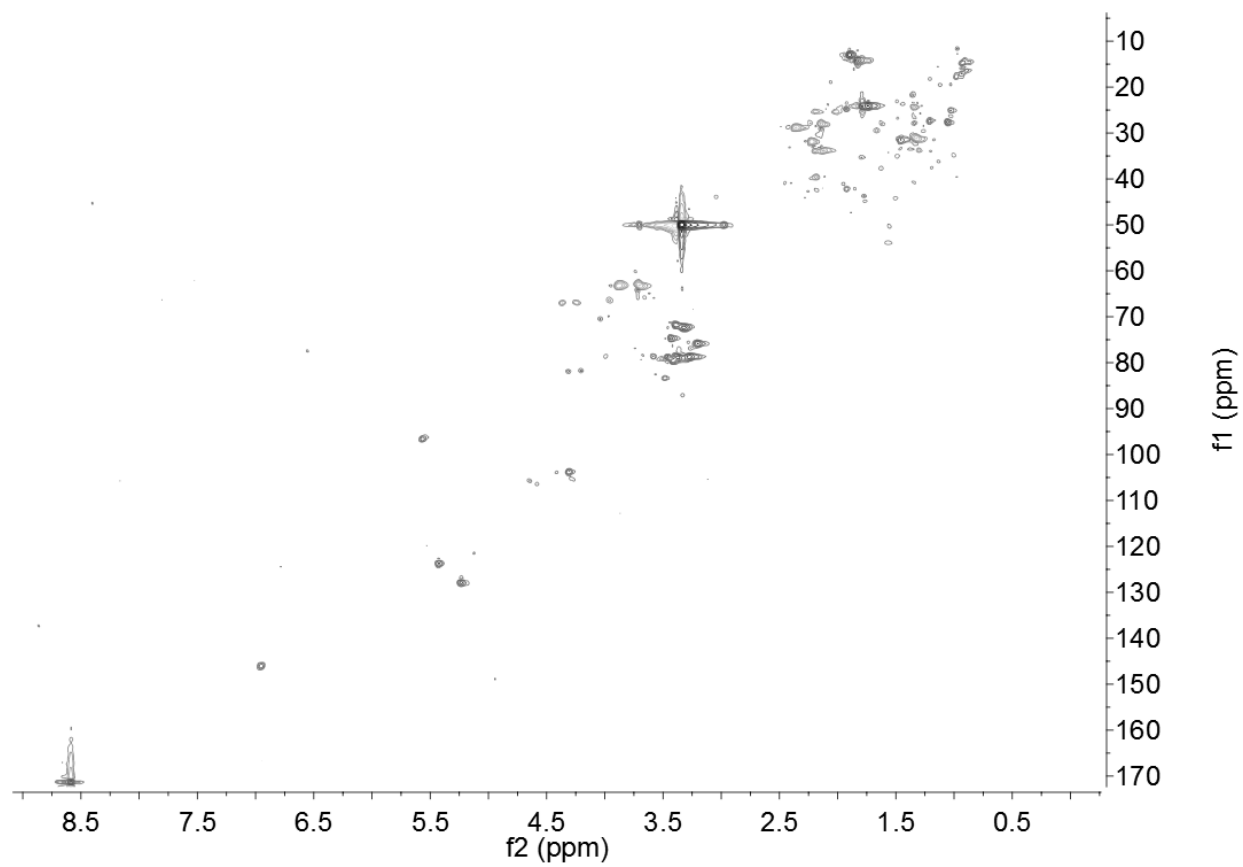


Figure 4.5. 2D HSQC spectrum of compound a (peak 8a from Figure 4.1).

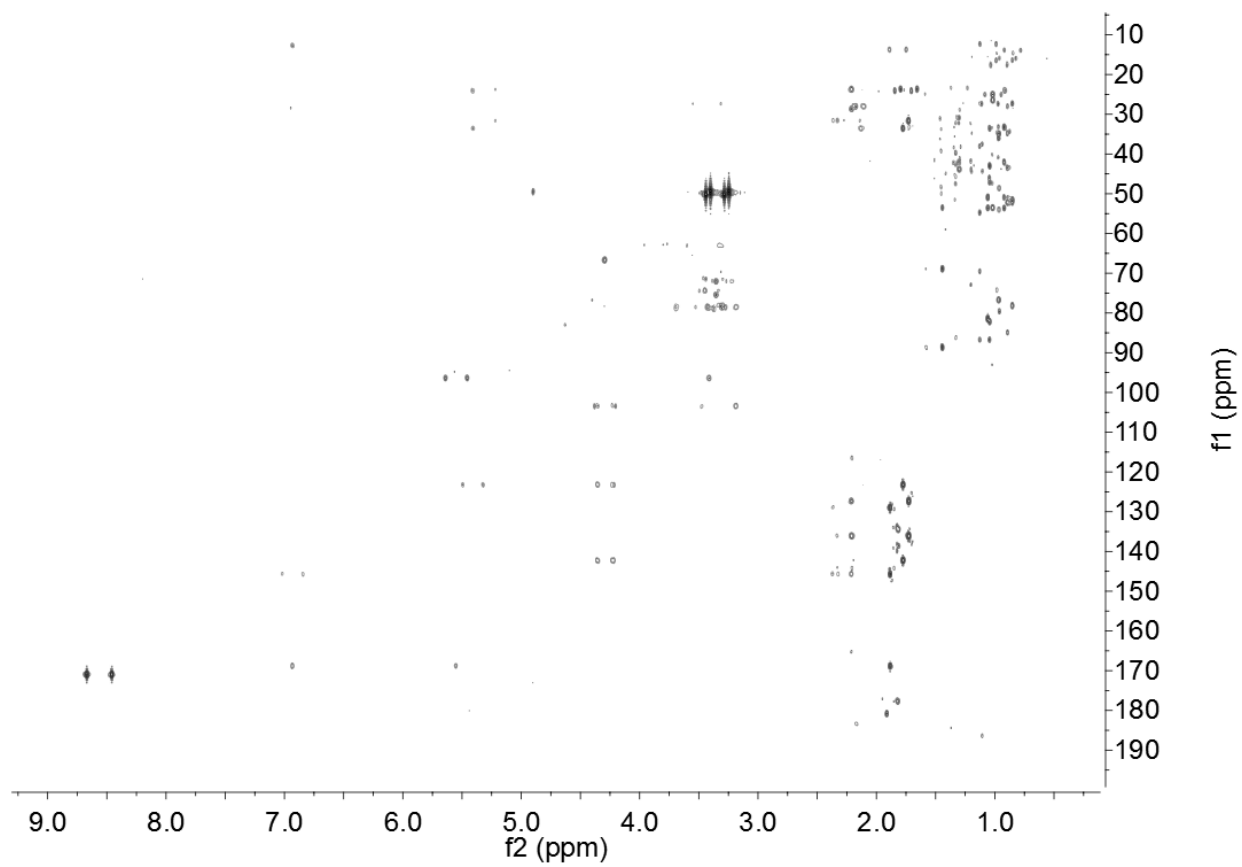


Figure 4.6. 2D HMBC spectrum of compound a (peak 8a from Figure 4.1).

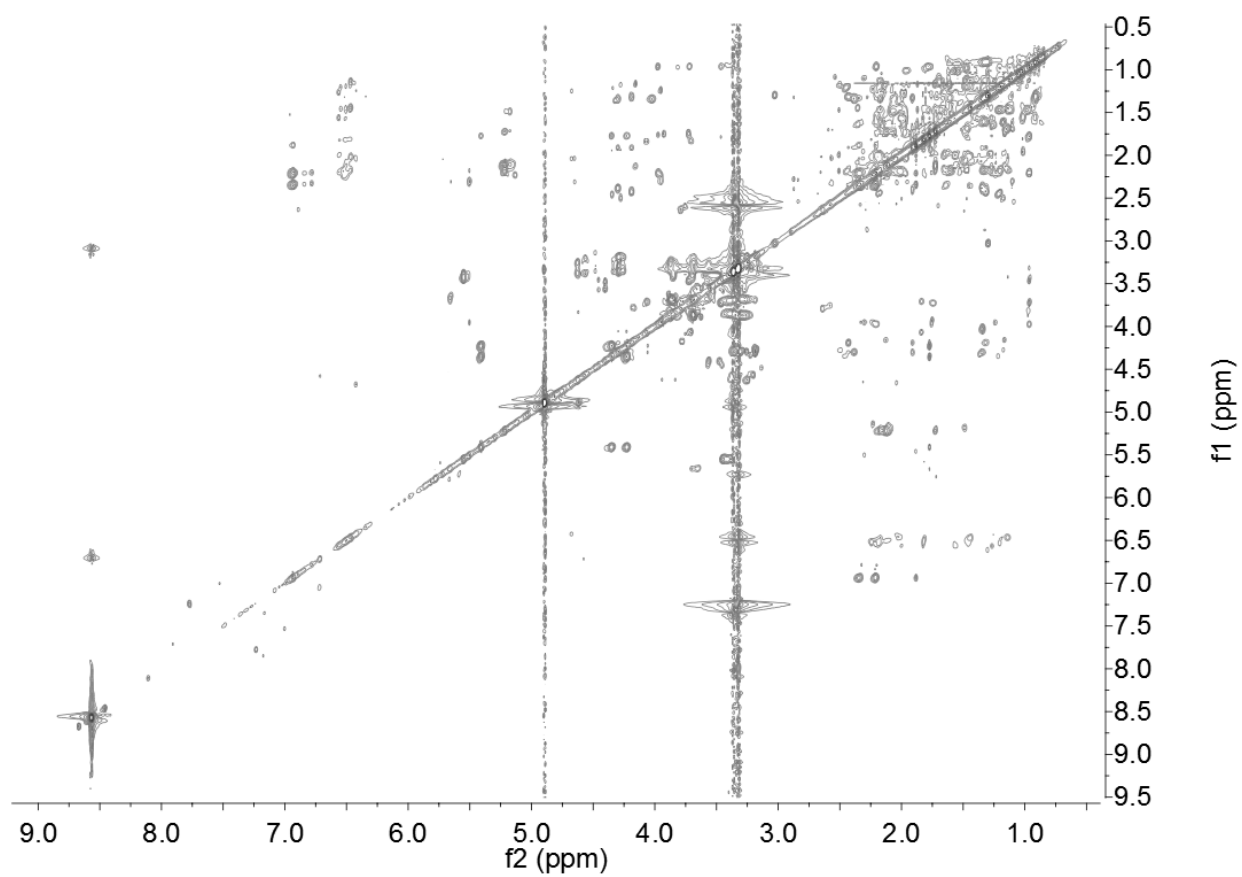


Figure 4.7. 2D TOCSY spectrum of compound a (peak 8a from Figure 4.1).

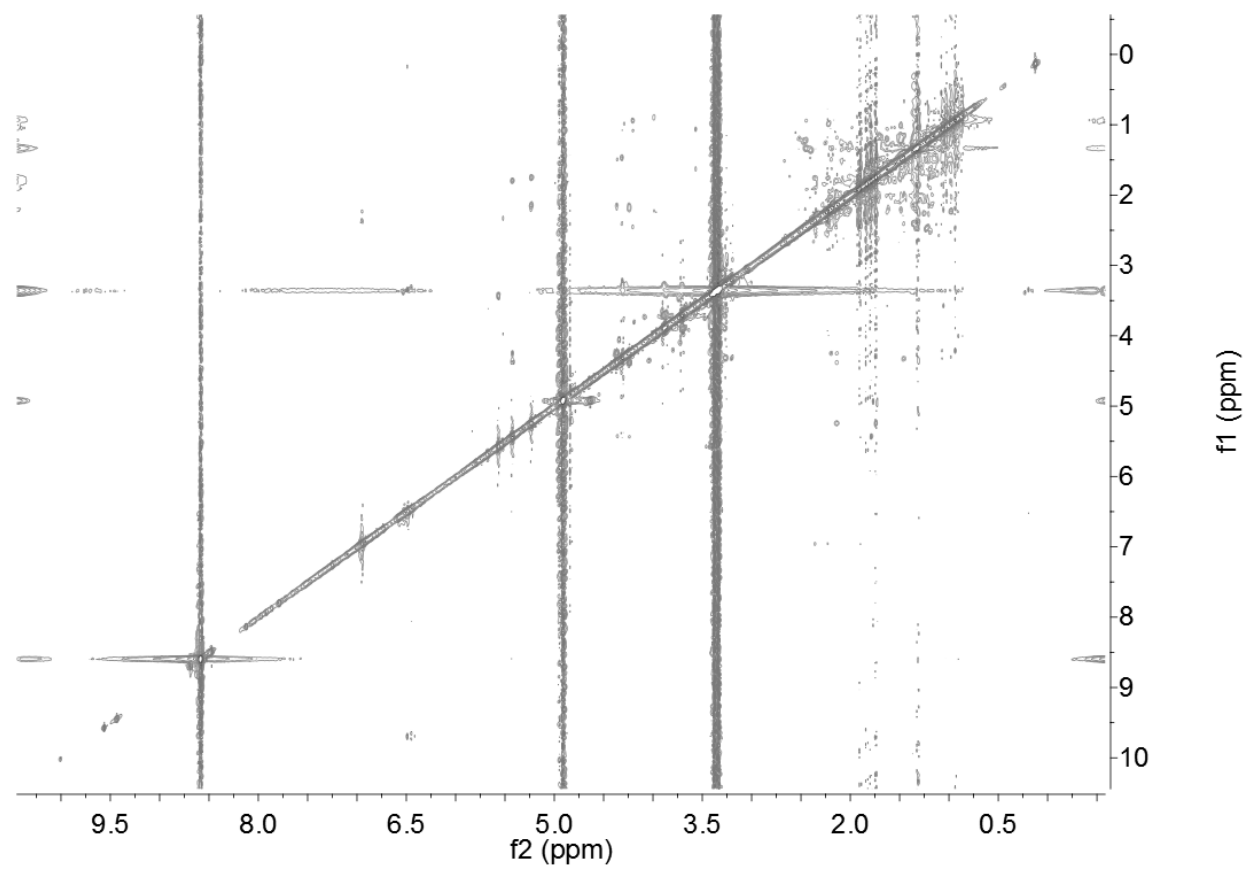


Figure 4.8. 2D NOESY spectrum of compound a (peak 8a from Figure 4.1).

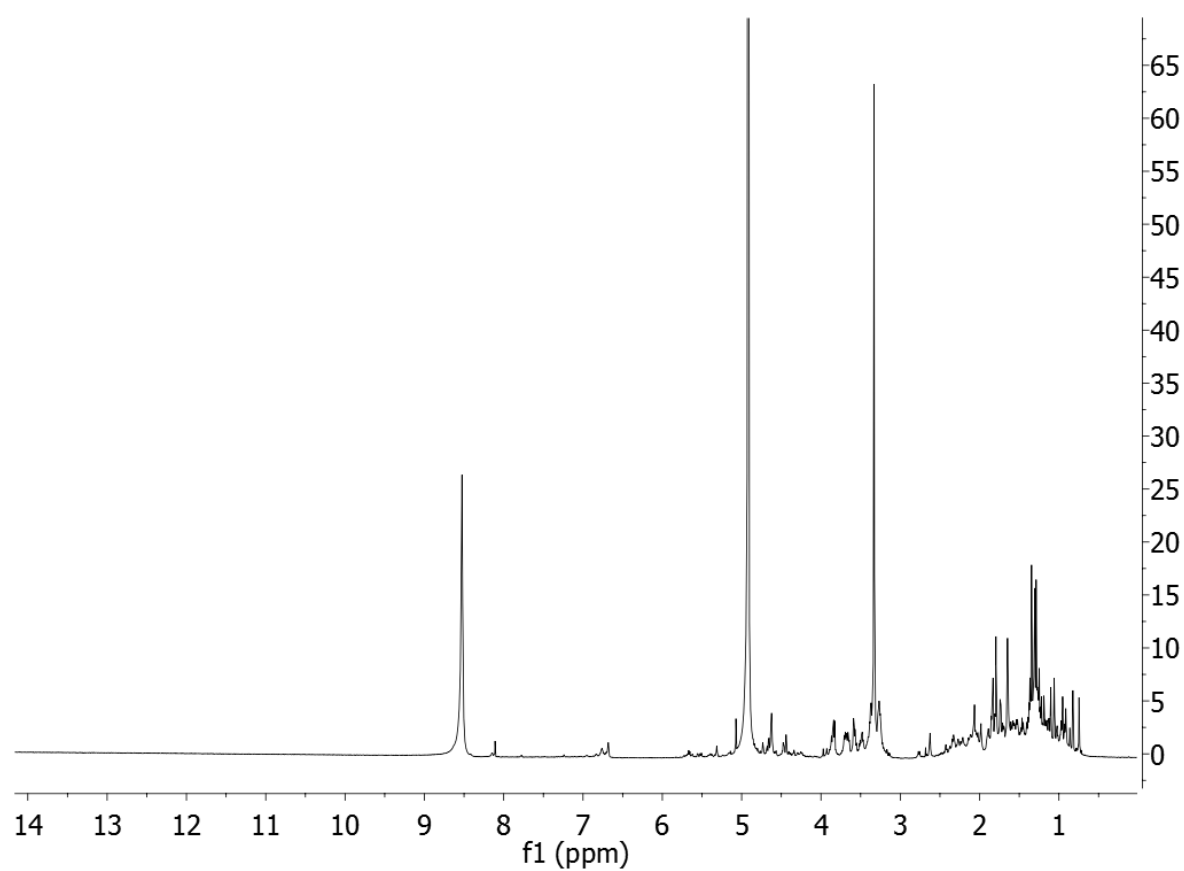


Figure 4.9. 1D Proton NMR spectrum of compound b (peak 3b from Figure 4.1).

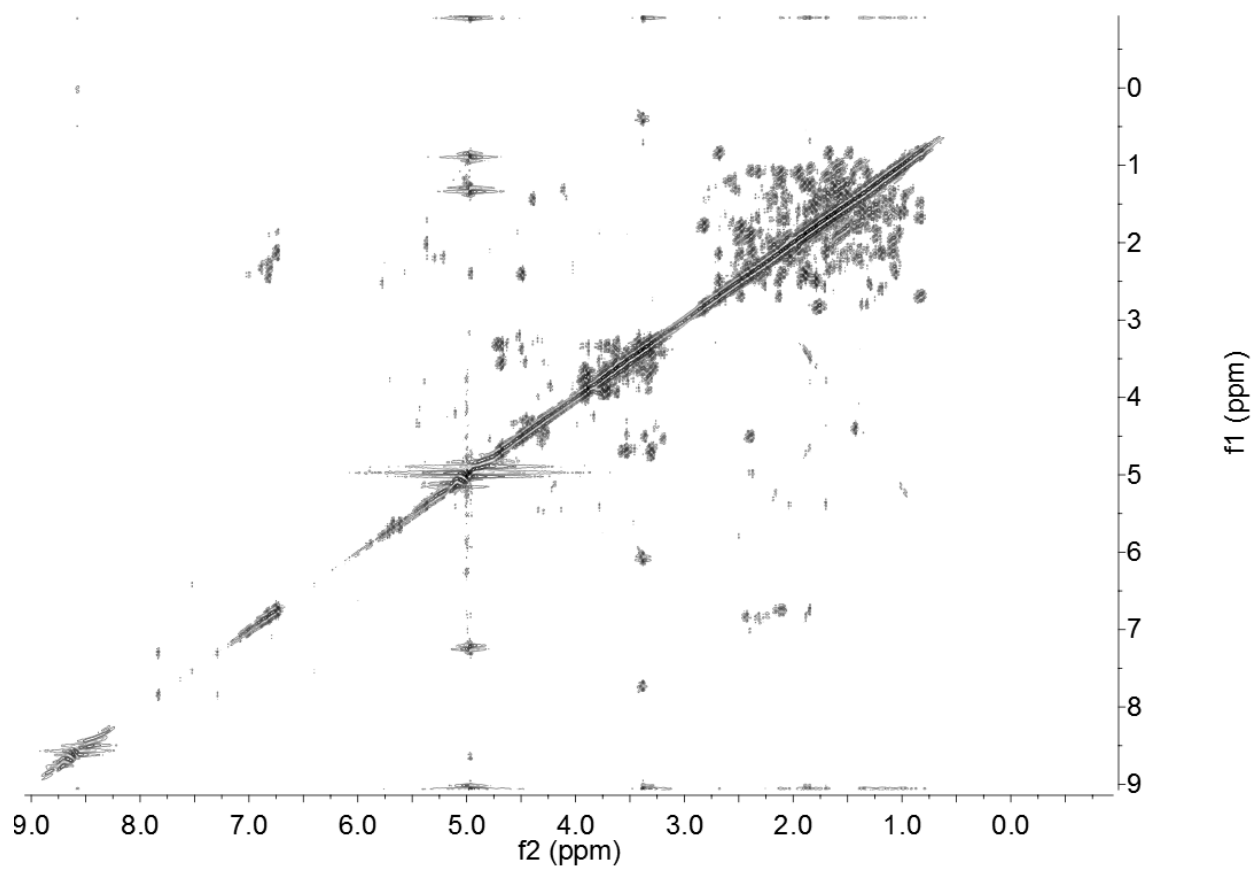


Figure 4.10. 2D COSY spectrum of compound b (peak 3b from Figure 4.1).

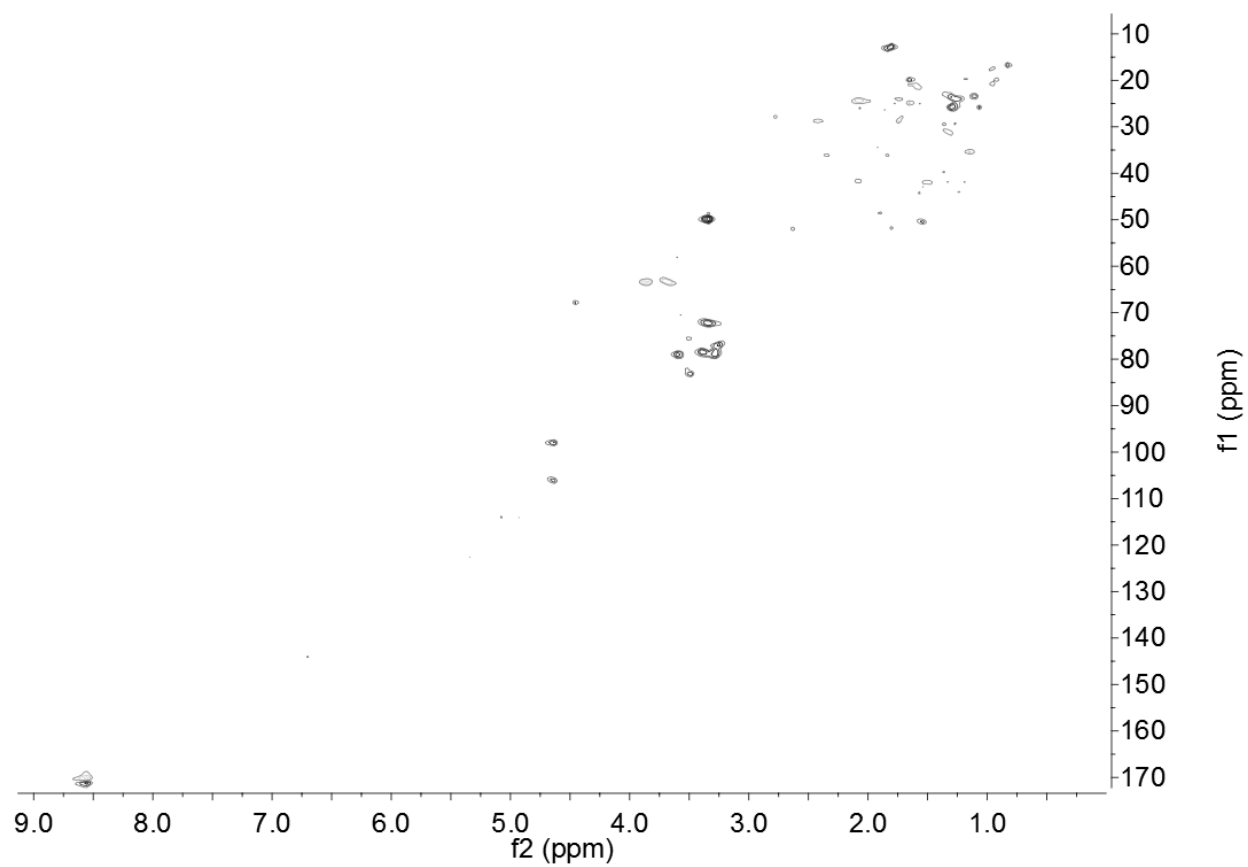


Figure 4.11. 2D HSQC spectrum of compound b (peak 3b from Figure 4.1).

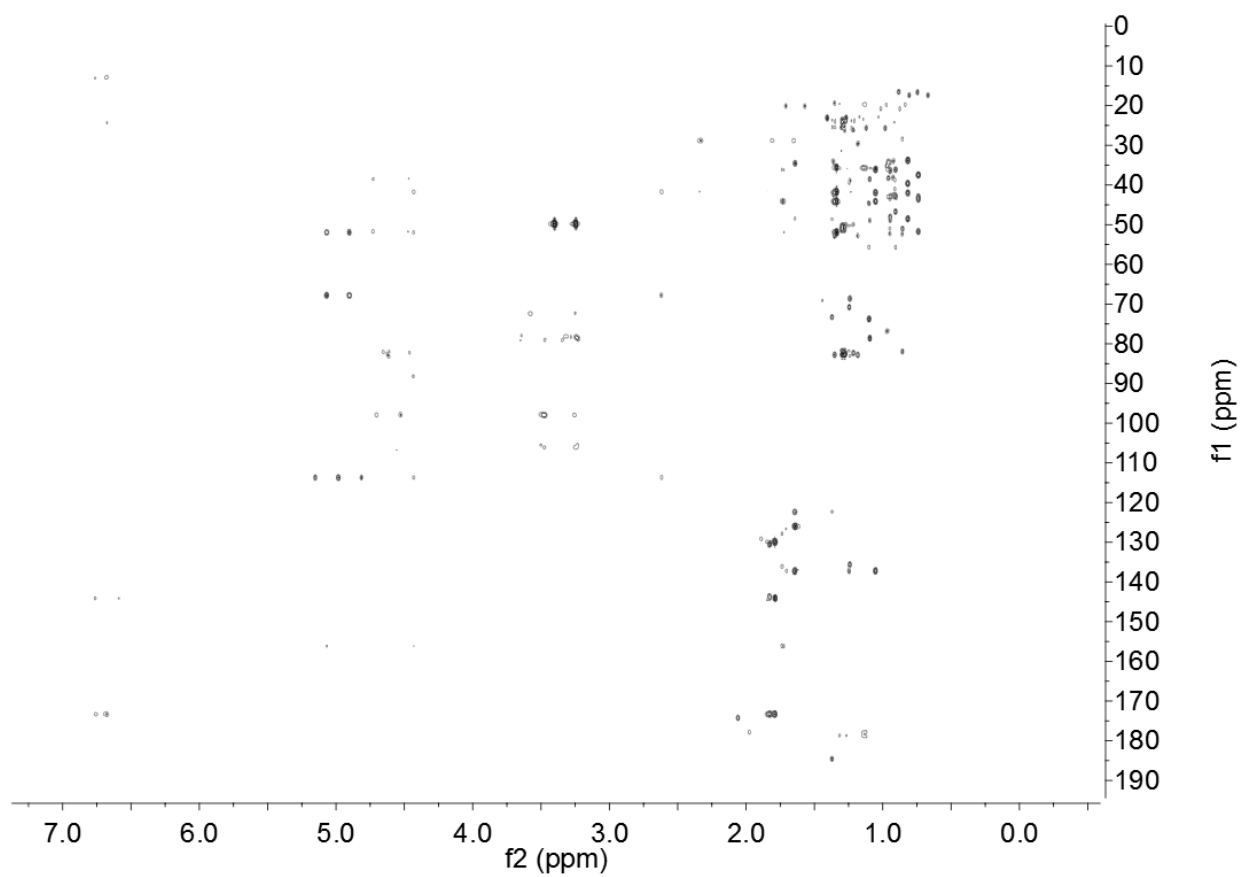


Figure 4.12. 2D HMBC spectrum of compound b (peak 3b from Figure 4.1).



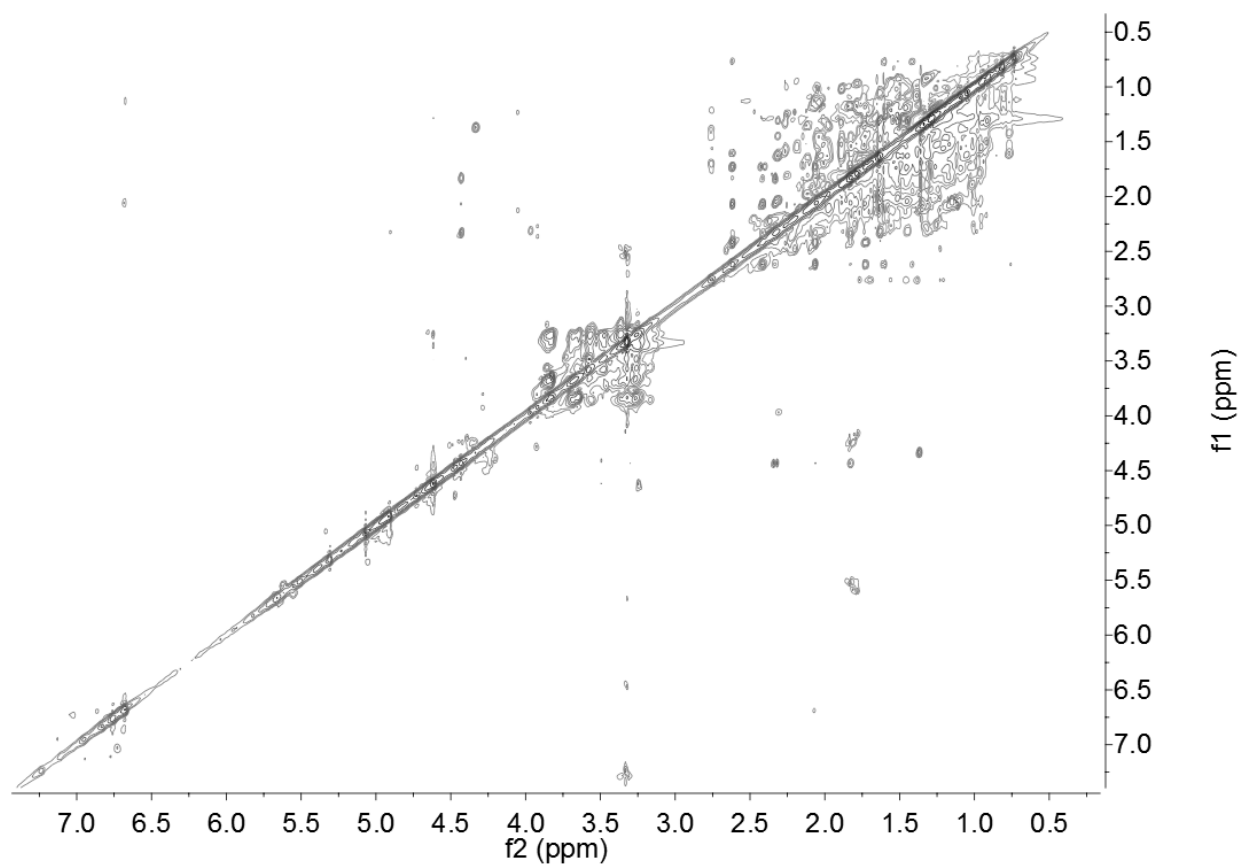


Figure 4.13. 2D TOCSY spectrum of compound b (peak 3b from Figure 4.1).

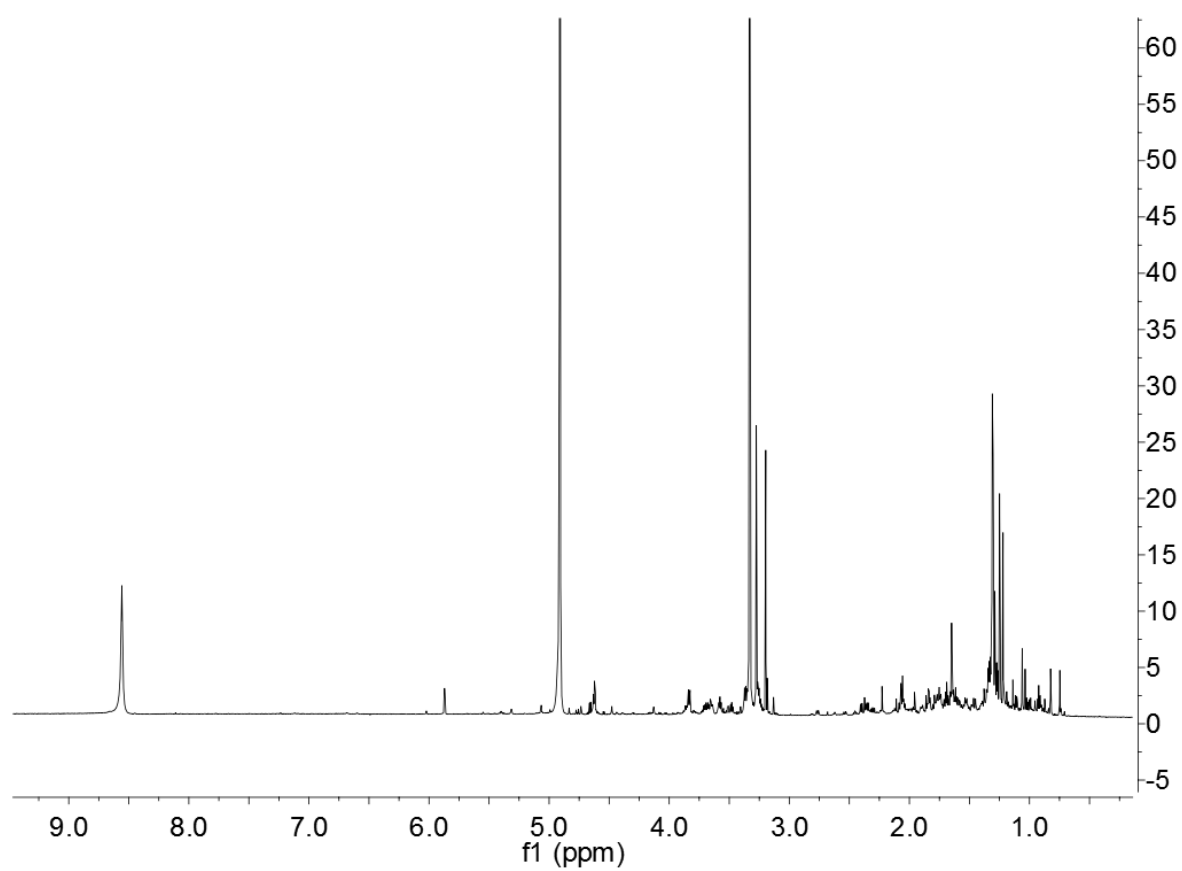


Figure 4.14. 1D  $^1\text{H}$  NMR spectrum of compound c (peak 6c from Figure 4.1).

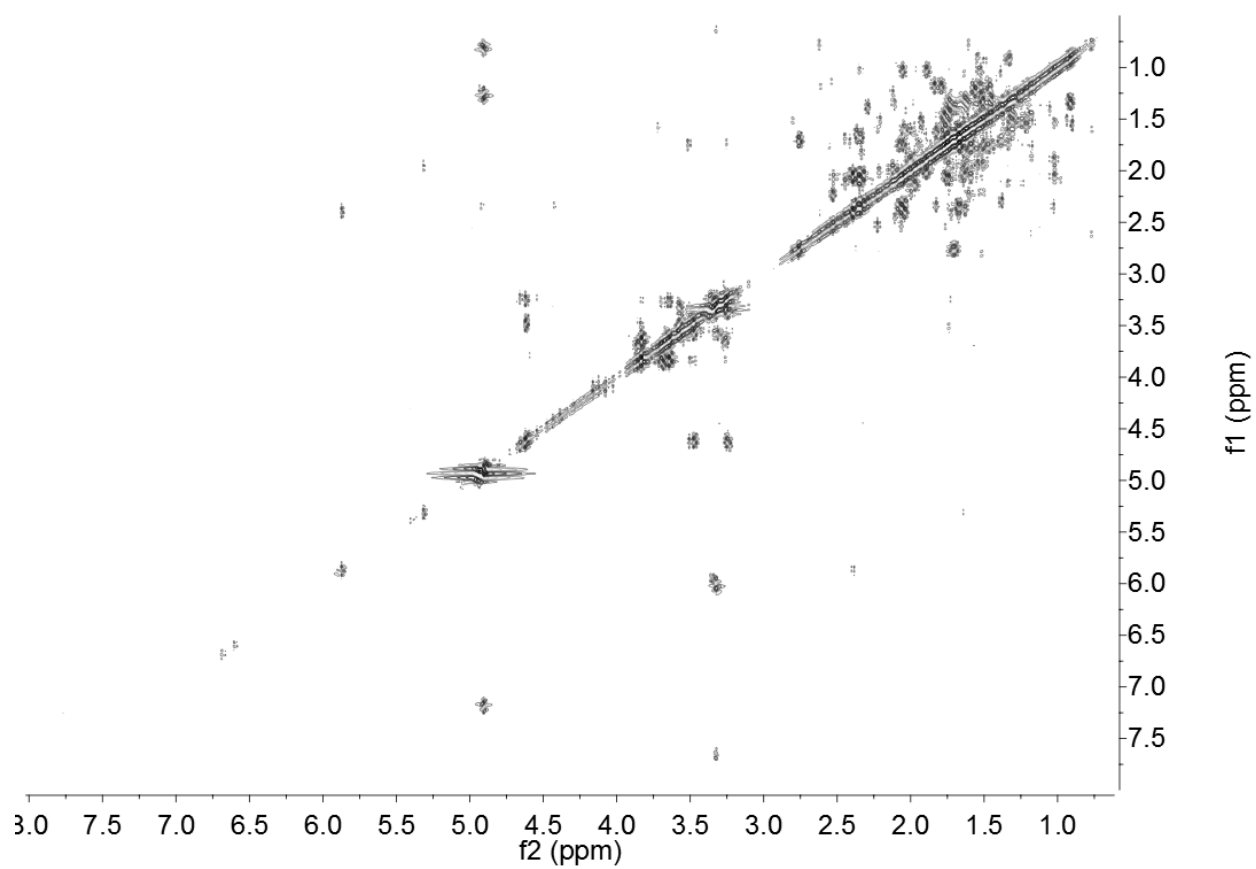


Figure 4.15. 2D <sup>1</sup>H COSY spectrum of compound c (peak 6c from Figure 4.1).

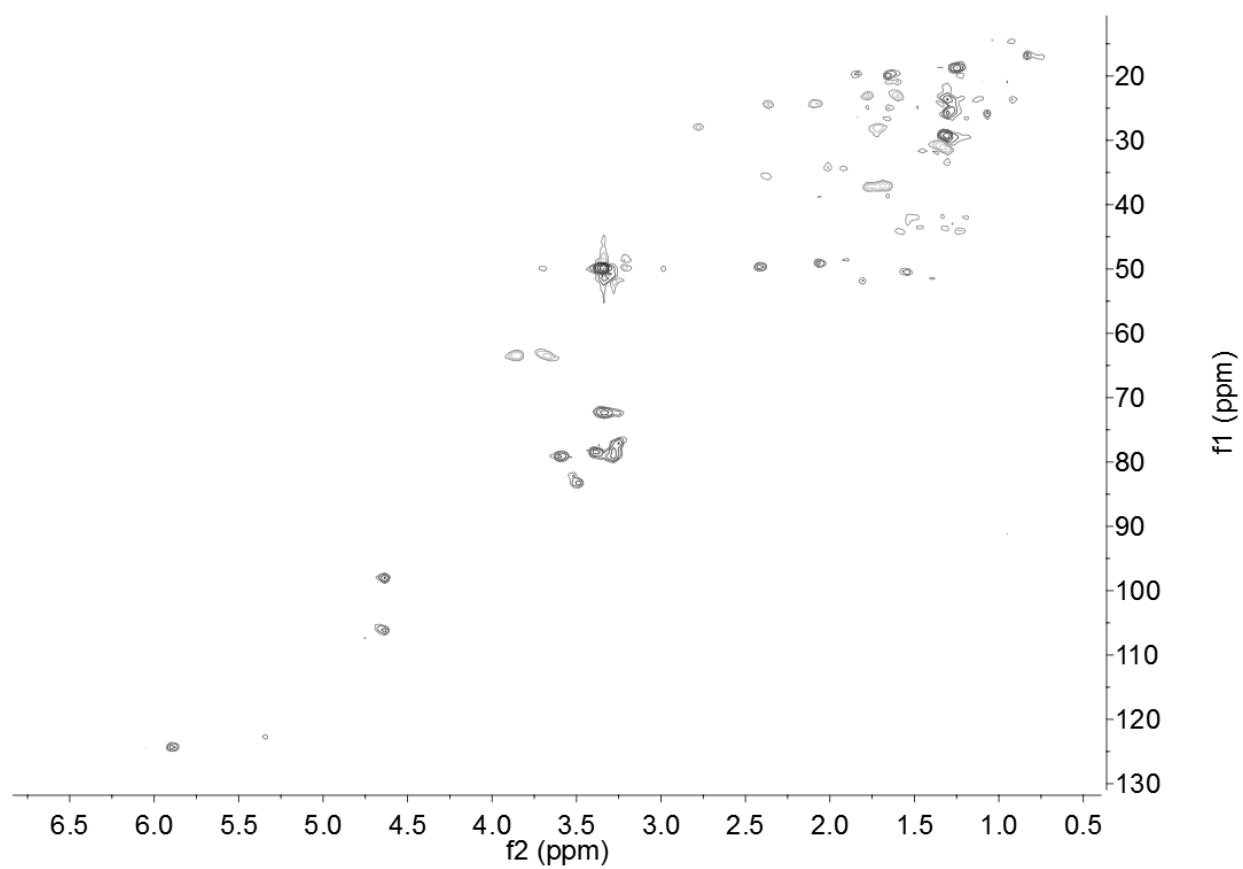


Figure 4.16. 2D HSQC spectrum of compound c (peak 6c from Figure 4.1).

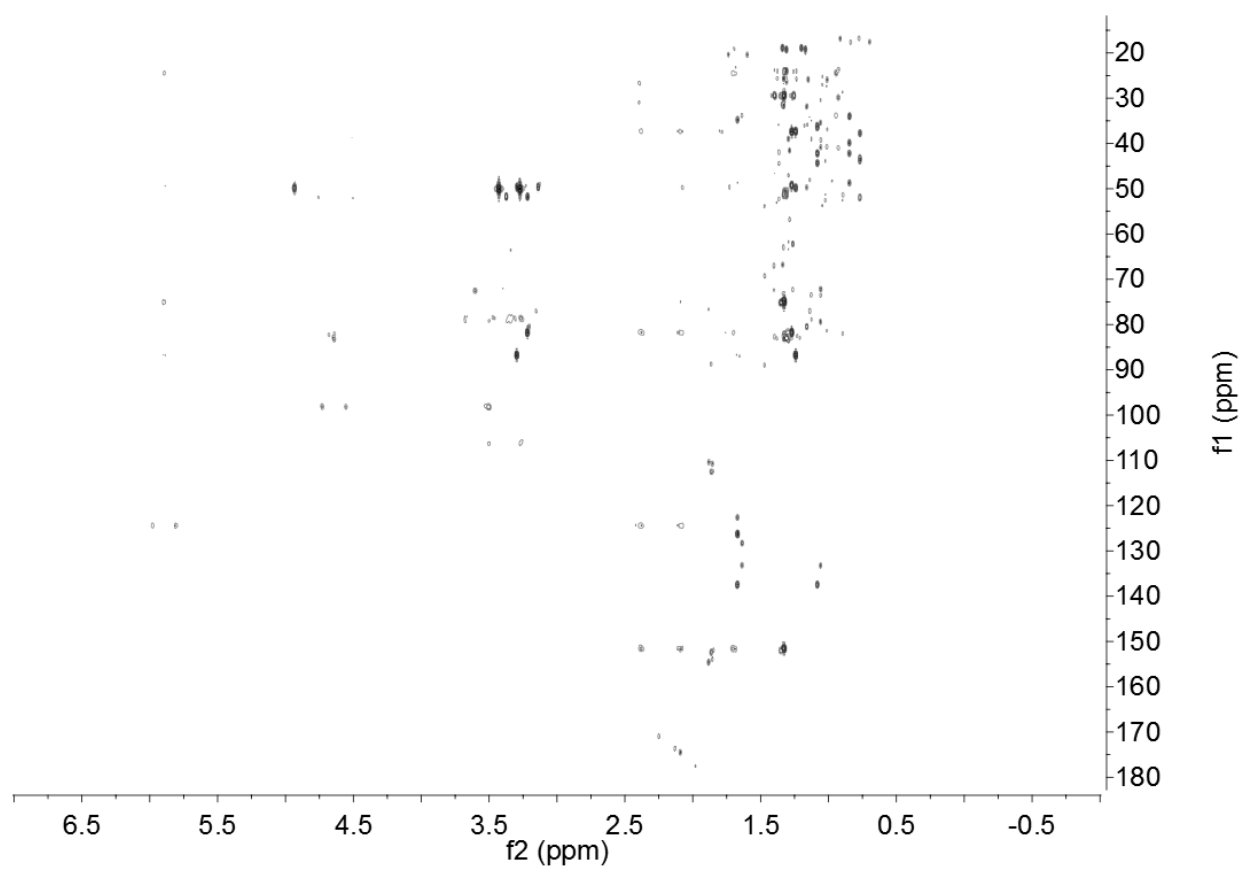


Figure 4.17. 2D HMBC spectrum of compound c (peak 6c from Figure 4.1).

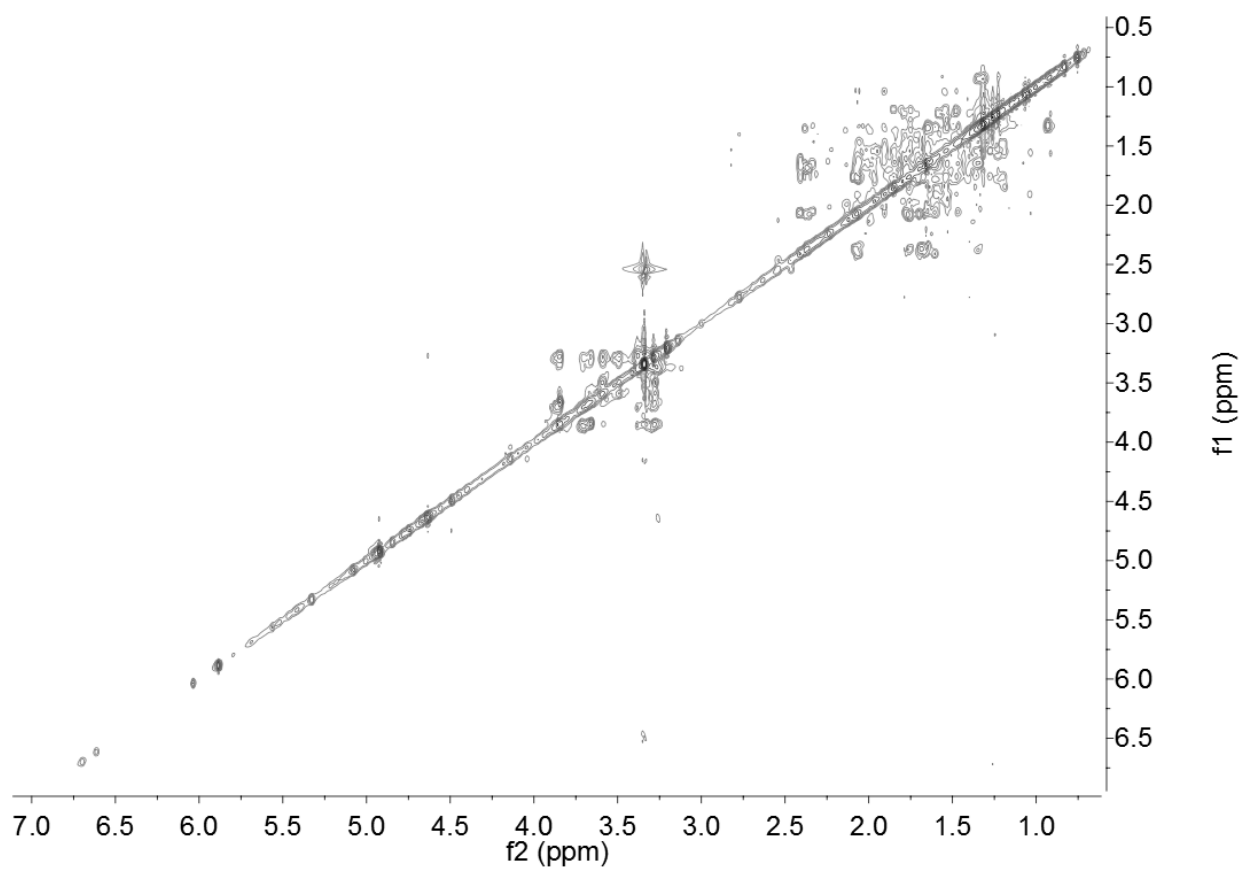


Figure 4.18. 2D TOCSY spectrum of compound c (peak 6c from Figure 4.1).

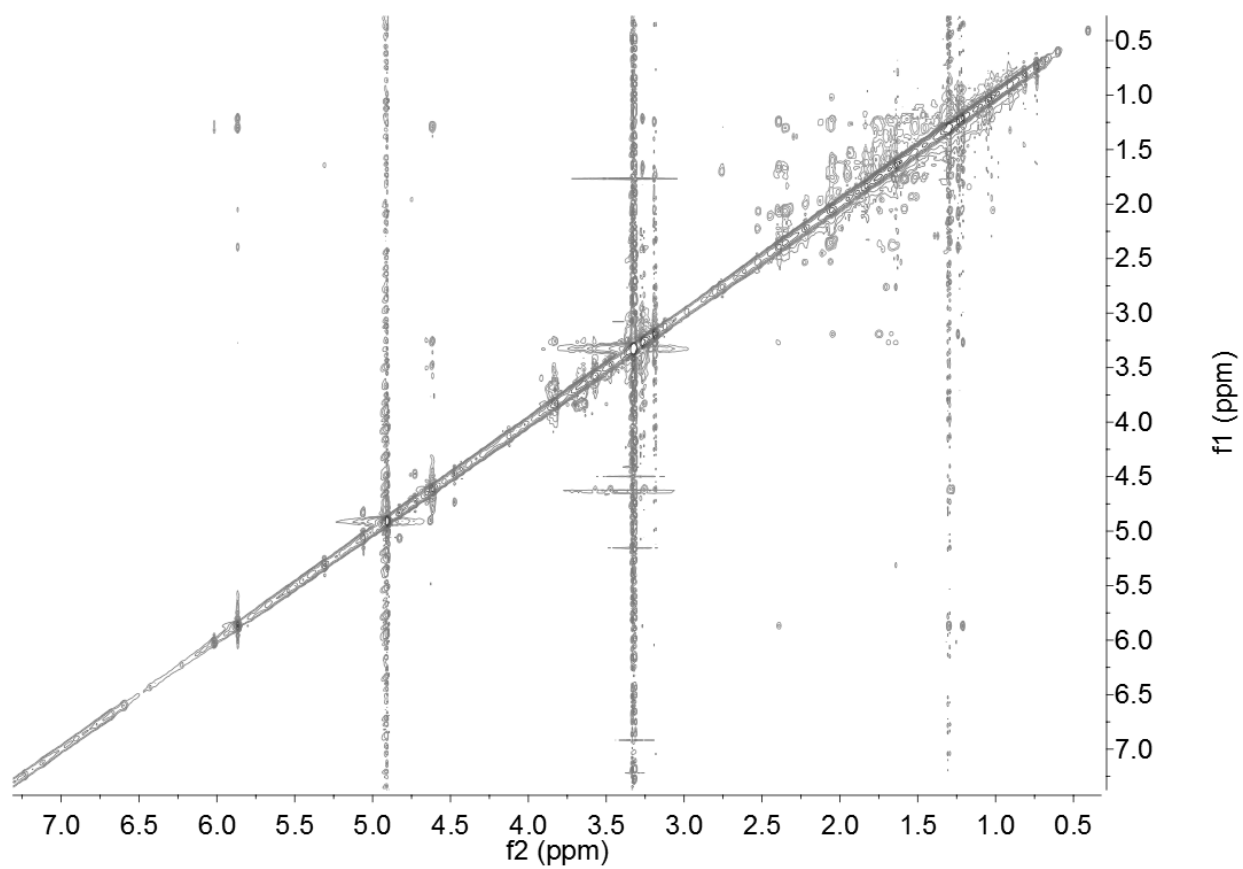


Figure 4.19. 2D NOESY spectrum of compound c (peak 6c from Figure 4.1).

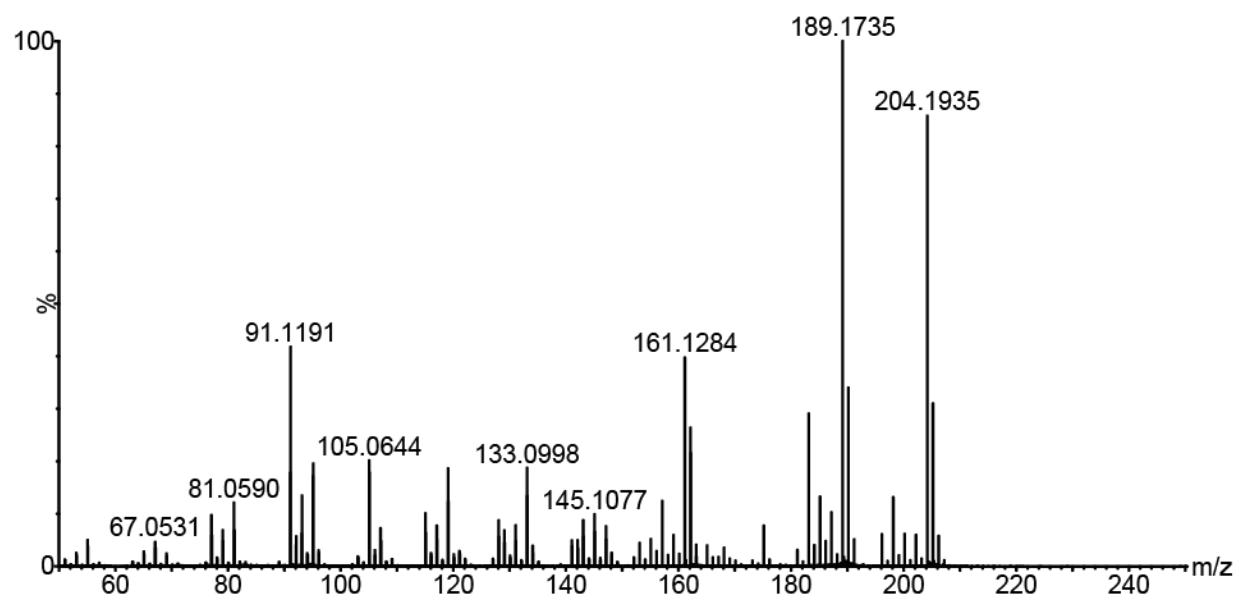
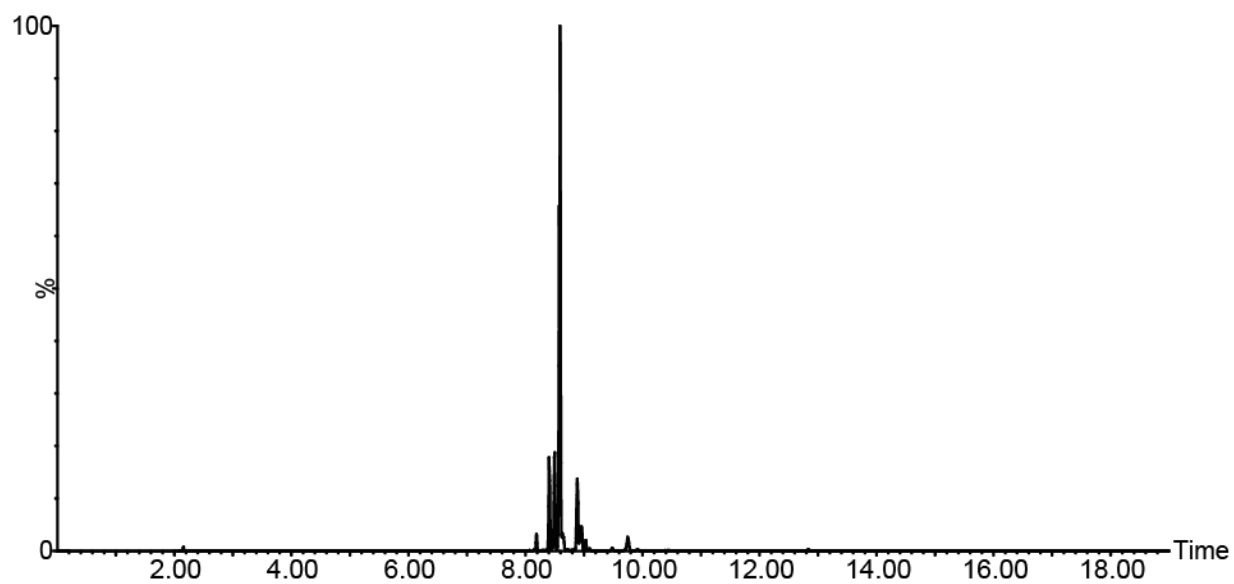


Figure 4.20. GC-MS chromatogram and EI spectrum of most abundant peak from the acid hydrolysis product of compound b.



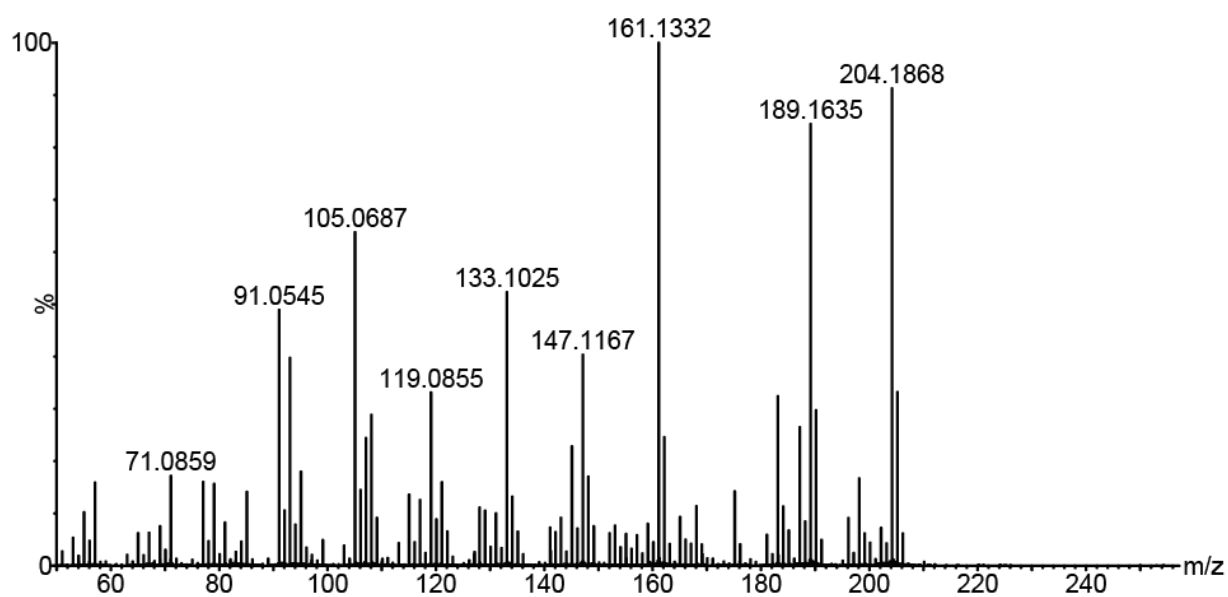
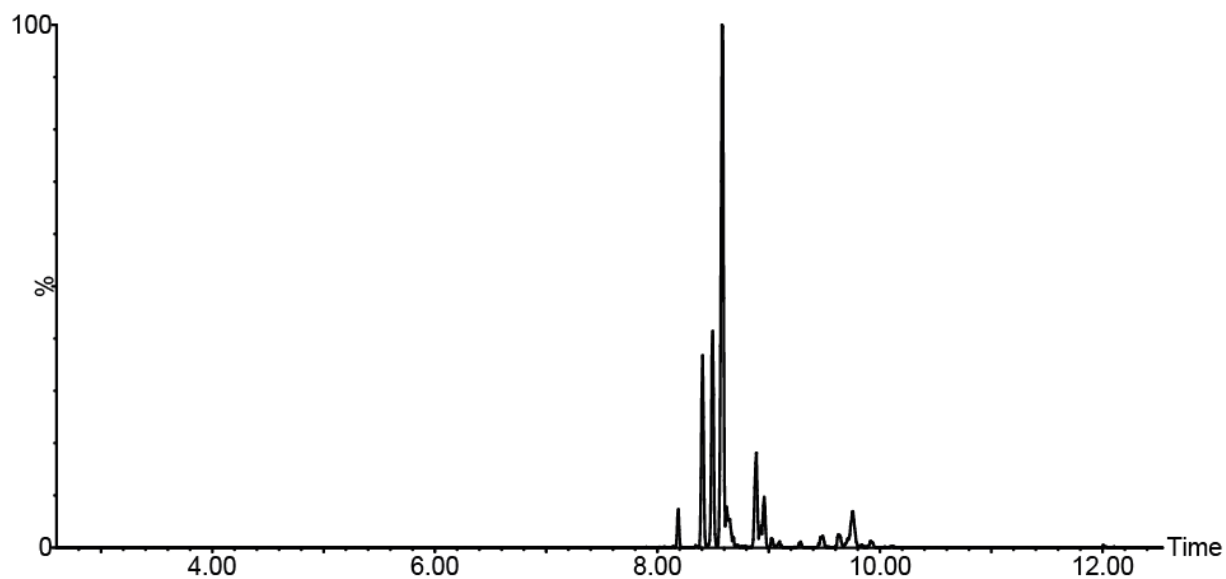


Figure 4.21. GC-MS chromatogram and EI spectrum of most abundant peak from the acid hydrolysis product of compound c.

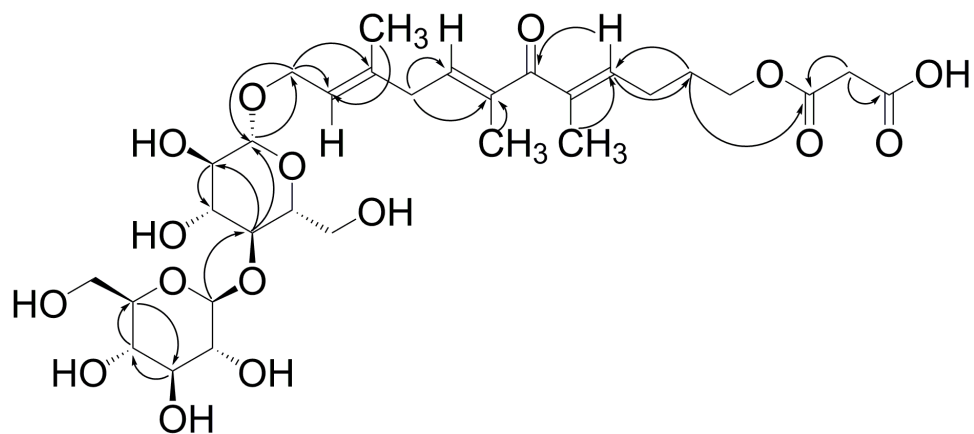


Figure 4.22. HMBC correlations assigned for compound **a**

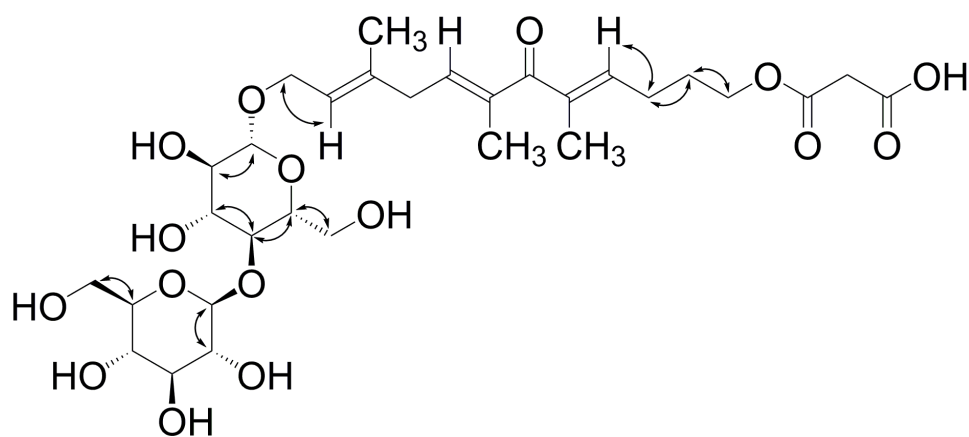


Figure 4.23. Important COSY correlations assigned for compound **a**

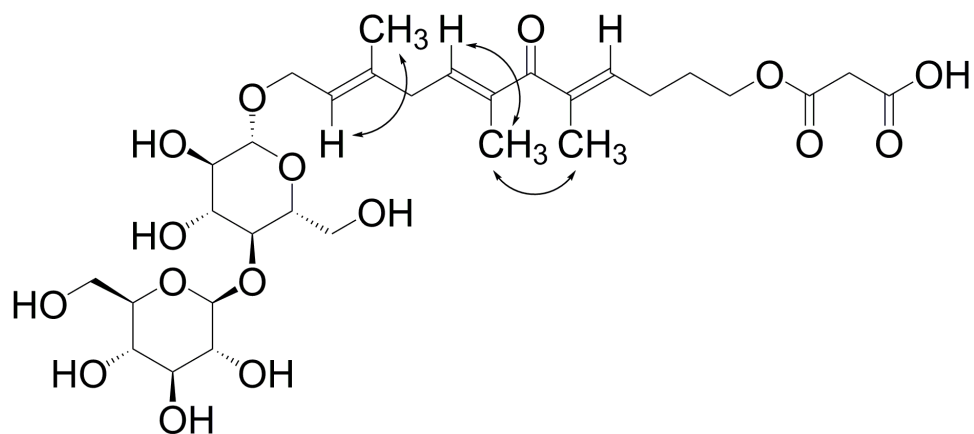


Figure 4.24. NOE correlations assigned for compound **a**

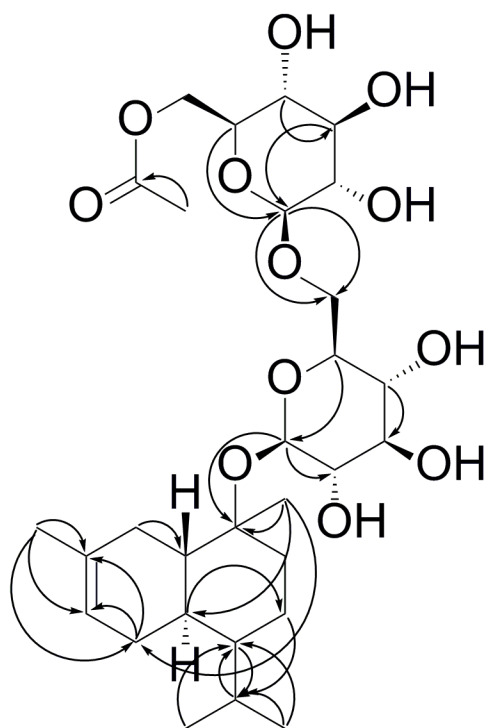


Figure 4.25. HMBC correlations assigned for compound **b**

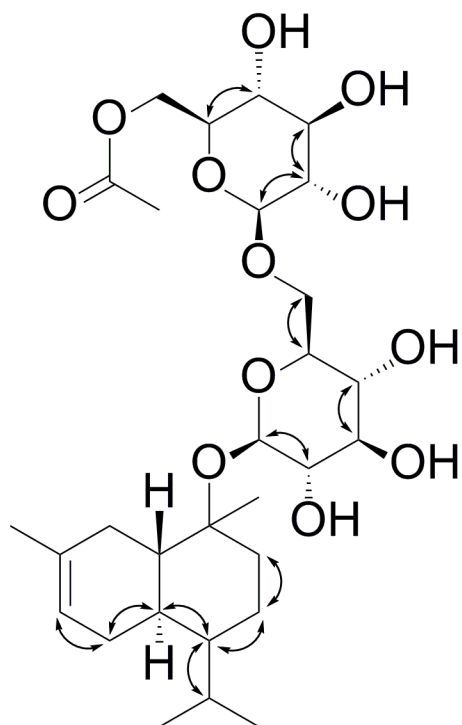


Figure 4.26. COSY correlations assigned for compound **b**

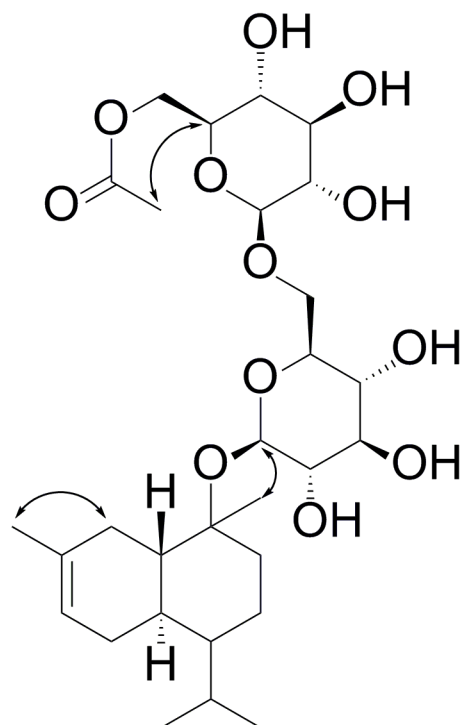


Figure 4.27. NOE correlations assigned for compound **b**

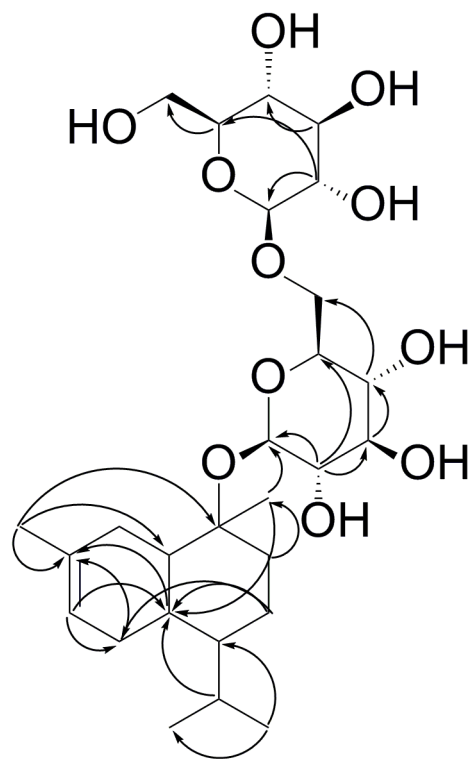


Figure 4.28. HMBC correlations assigned for compound **c**

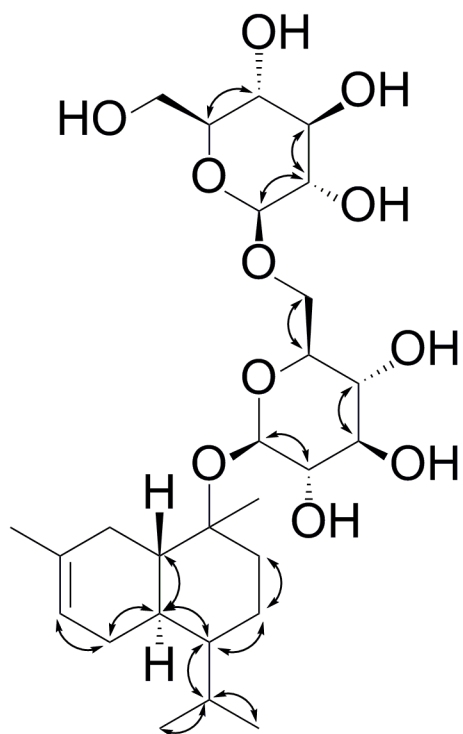


Figure 4.29. COSY correlations assigned for compound **c**

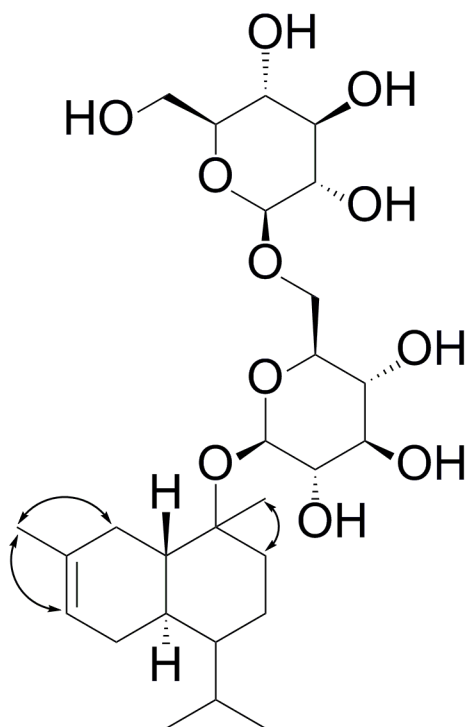


Figure 4.30. NOE correlations assigned for compound **c**

Table 4.1: Chemical shift assignments to the carbons and protons of the three compounds purified from *S. habrochaites* LA1777.

Position	Compound a			Compound b			Compound c		
	$\delta\text{C}$	$\delta\text{H}$		$\delta\text{C}$	$\delta\text{H}$		$\delta\text{C}$	$\delta\text{H}$	
<b>1</b>	66.30	3.99	3.99	173.30			151.30		
<b>2</b>	21.70	2.22	2.22	144.20	6.76		124.20	5.89	
<b>3</b>	28.70	2.36	2.36	36.10	1.82	2.34	24.36	2.06	2.36
<b>4</b>	146.20	6.93		51.7	1.80		49.10	2.06	
<b>5</b>	129.00			48.60	1.90		51.60	1.80	
<b>6</b>	169.10			24.90	1.69	1.77	37.20	1.67	1.77
<b>7</b>	136.00			82.90			81.70		
<b>8</b>	127.70	5.20		27.50	1.72	2.77	23.14	1.60	1.77
<b>9</b>	33.74	2.10	2.18	28.65	2.41	1.74	27.90	1.71	2.78
<b>10</b>	142.80			19.91	1.84		50.50	1.54	
<b>11</b>	123.80	5.41		50.30	1.55		19.96	1.64	
<b>12</b>	66.96	4.33	4.22	20.25	1.30		18.50	1.24	
<b>13</b>	12.70	1.86		25.90	1.28		29.20	1.30	
<b>14</b>	24.00	1.71		19.64	1.63		25.40	1.31	
<b>15</b>	23.95	1.78		23.50	1.25		23.40	1.27	
<b>16</b>	103.70	4.31		98.10	4.67		106.20	4.66	
<b>17</b>	75.80	3.18		83.10	3.54		77.00	3.23	
<b>18</b>	78.80	3.31		78.90	3.60		79.00	3.65	
<b>19</b>	83.15	3.49		72.50	3.26		72.40	3.32	
<b>20</b>	78.70	3.61		78.50	3.38		82.00	3.57	
<b>21</b>	63.00	3.87	3.69	63.50	3.88	3.66	63.50	3.86	3.69
<b>22</b>	96.70	5.56		106.40	4.65		98.00	4.63	
<b>23</b>	74.70	3.41		76.90	3.28		83.40	3.50	
<b>24</b>	64.50	3.72		78.70	3.42		79.00	3.60	
<b>25</b>	71.60	3.25		88.30	3.26		72.50	3.26	
<b>26</b>	74.50	3.45		68.00	4.44		78.50	3.38	
<b>27</b>	65.58	3.67	3.57	65.50	3.86	3.69	63.50	3.86	3.69
<b>28</b>	165.35			178.40					
<b>29</b>	25.40	2.17	2.01	23.60	1.11				
<b>30</b>	183.55								

## REFERENCES

## REFERENCES

1. Jassbi AR, Zamanizadehnajari S, Kessler D, Baldwin IT (2006) A new acyclic diterpene glycoside from *Nicotiana attenuata* with a mild deterrent effect on feeding *Manduca sexta* larvae. *Zeitschrift Fur Naturforschung Section B-a Journal of Chemical Sciences* 61 (9):1138-1142
2. Kasai R, Nishi M, Mizutani K, Miyahara I, Moriya T, Miyahara K, Tanaka O (1988) Trifolioside-II , an acyclic sesquiterpene oligoglycoside from pericarps of *Sapindus-trifoliatus*. *Phytochemistry* 27 (7):2209-2211
3. Magid AA, Voutquenne-Nazabadioko L, Litaudon M, Lavaud C (2005) Acylated farnesyl diglycosides from *Guioa crenulata*. *Phytochemistry* 66 (23):2714-2718. doi:10.1016/j.phytochem.2005.09.009
4. Morikawa T, Xie YY, Ninomiya K, Okamoto M, Muraoka O, Yuan D, Yoshikawa M, Hayakawa T (2010) Inhibitory Effects of Acylated Acyclic Sesquiterpene Oligoglycosides from the Pericarps of *Sapindus rarak* on Tumor Necrosis Factor-alpha-Induced Cytotoxicity. *Chemical & Pharmaceutical Bulletin* 58 (9):1276-1280
5. Hu HB, Jian YF, Zheng XD, Cao H (2007) Three sesquiterpene glycosides from *Elsholtzia bodinieri*. *Bulletin of the Korean Chemical Society* 28 (3):467-470
6. Ekanayaka EAP, Li C, Jones AD (2013) Sesquiterpenoid Glycosides from Glandular Trichomes of the Wild Tomato Relative *Solanum habrochaites*. *Phytochemistry*



## Chapter 5: Discovering terpene glycosides from *Hoodia gordonii* by mining multiplexed CID UHPLC-MS data and applying relative mass defect filtering

### 5.1 Introduction

*Hoodia gordonii* is a plant considered to be rich in appetite suppressants and therefore is used in inducing weight loss [1], though scientific evidence of its effectiveness has been lacking.

Extracts of *Hoodia* contain an assortment of diterpenoid glycosides known as Hoodigosides, and it has been proposed that one of these, termed P57AS, is the main active ingredient [2]. The appetite suppressant activity of P57AS has been attributed to its being able to increase the ATP content in the hypothalamus neurons that regulate food intake of humans [2]. Similarly, another appetite suppressant compound was isolated from *Hoodia* recently and was tested for its appetite suppressing activity using mouse models [3]. In addition to the known Hoodigosides with appetite suppressant activity, eleven biologically active (antioxidant and cytotoxic) oxypregnane glycosides have been recently reported from *Hoodia gordonii* [4]. Furthermore, *H. gordonii* has also been patented for its antidiabetic activity [5] and as a protectant against acetaminophen-induced gastrointestinal damage based on *in vitro* bioactivity assay [6].

*Hoodia* is rich in various diterpene glycoside compounds [7-9]. Five main terpenoid cores are observed in *Hoodia* [7]. These are Hoodigogenin A/Gordonoside A, Calogenin, Isoramanone, Hoodistanal and Dehydrohoodistanal (Fig. 5.1a-e) [7]. Diterpene glycosides found in *Hoodia* consist of sugars that are not commonly found in glycosides. Six main types of sugar groups found in diterpenoid glycosides from *Hoodia* are cymarose, oleandrose, digitoxose, thevetose, 3-*O*-methyl-6-deoxyallose and glucose (Fig. 5.1f-k) [2]. These sugar groups get attached to the terpenoid core to produce some of the most “decorated” terpene glycosides in nature (Fig. 5.2).

These findings suggest that *Hoodia gordonii* is a plant that possesses an assortment of biologically important compounds, but the mechanisms responsible for their formation have not been established. Therefore, investigating the chemical composition of this plant using a systematic study that involves deep profiling of specialized metabolites of different tissues of this plant is an important endeavor. Moreover, discovering the genes involved in regulation of these compounds will aid the development of means for engineering plants or microbes for the synthesis of these compounds at larger scale[10].

A systematic study was performed to analyze *Hoodia* stem tissues (primary step pith and spine tissues) using UHPLC-multiplexed CID MS. Relative mass defect (RMD) filtering was applied to the data generated from this analysis to aid recognition of terpenoid metabolites, including low-abundance compounds that are potential metabolic intermediates. The same RMD filtering method can be extended to analyzing metabolite data from *Hoodia gordonii* available online at the Medicinal Plants Consortium database (available at <http://metnetdb.org/PMR/>). The RMD filtering method correctly recognizes all 29 compounds that are annotated in database as terpene glycosides. Among the 100 most abundant metabolites in this category, only 18 have been annotated as known compounds. Using RMD filtering along with MS/MS analysis led to the annotation of 24 additional novel steroidal glycoside compounds among these unknown molecules, as described below.

## 5.2 Methods

### 5.2.1 LC-MS and LC-MS/MS experiments

All LC-MS experiments were performed using a Waters LCT Premier mass spectrometer using the standard electrospray ion source. This instrument was coupled to a Shimadzu Prominence Liquid Chromatography system (LC-20AD pumps with high pressure mixing). Separations were performed using an Ascentis Express C-18 UHPLC column (2.1 mm × 100 mm; 2.7 μm fused core particles, Supelco Co. USA). Aqueous 10 mM ammonium formate adjusted to pH 2.55 was used as solvent A and 100% acetonitrile was used as solvent B for the separation of the compounds. The LC gradient started with 1% B/ 99% A, held for one minute, increased to 30% B at three minutes, then 70% B at 31 minutes, then 100% B at 31.01 minutes, held at 100 % B until 33 minutes, decreased to 1% B at 33.01 minutes, held at 1% B. Total analysis time was 35 minutes. Total flow rate was 0.30 ml/min and column temperature was held at 50 °C. Mass spectra were acquired using positive-ion mode electrospray ionization, centroid peak detection and dynamic range extension over  $m/z$  50-1400, with mass resolution ( $M/\Delta M$ , full width-half maximum) approximately 10000. Five parallel collision energy functions were used, with 0.1 s per function. Collision cell potentials used for negative ion mode fragmentation for each function were 5, 20, 40, 60 and 80 V respectively.

All LC-MS/MS analyses were performed using a Waters Xevo G2-S QToF mass spectrometer coupled to a Waters Acquity ultra-high performance liquid chromatography system. The LC gradient used here was the same as the one used for LC-MS profiling (see above). Other parameters include capillary voltage of 3.0 kV, desolvation temperature of 350 °C, source temperature of 100 °C, cone gas (N<sub>2</sub>) at 0 L/hr and desolvation gas (N<sub>2</sub>) at 800 L/hr. In all analyses, leucine enkephalin (0.1 ng/μL) was used as a lock mass reference infused at a flow rate

of 3.0  $\mu\text{L}/\text{min}$ , and lock mass spectra were acquired for 100 ms at 10 sec intervals. Capillary voltage for lock spray was 2.5 kV.

All MS/MS experiments were performed on the same instrument employing the same LC method, and argon as collision gas. For positive ion mode MS/MS experiments a collision potential +30 V was used.

### **5.2.2 Sample preparation for metabolite profiling**

Aliquots (100  $\mu\text{L}$ ) of *Hoodia gordonii* extracts (chloroform: methanol 1:1 v/v) delivered to Michigan State University from the laboratory of Professor Joe Chappell of the University of Kentucky were transferred into 200  $\mu\text{L}$  glass inserts, and the inserts were placed into microcentrifuge tubes. Solvent was evaporated under vacuum at room temperature in a SpeedVac, and residues were reconstituted in 100  $\mu\text{L}$  of 20/80 v/v methanol/water. The inserts were placed in 2 mL HPLC autosampler vials and briefly sonicated. The samples were spiked with 5  $\mu\text{L}$  of a 10  $\mu\text{M}$  solution of telmisartan (achieving 0.5  $\mu\text{M}$  final concentration). The HPLC autosampler vials were labeled with MSU barcodes. The order of analysis was randomized using Microsoft Excel.

## **5.3 Results and discussion**

### **5.3.1 Hoodia metabolite profiling**

*Hoodia gordonii* metabolite profiling was performed using the samples prepared above (5.2.2) and using the LC-MS method described in 5.2.1. The data generated were subjected to automated peak picking using Waters MarkerLynx peak picking algorithm. Peak picking parameters used were as follows. Peak Width at 5% Height = 18 s, masses per retention time 15, minimum intensity as a percentage of the base peak 0.02%, mass window 0.05 Da, retention time

window 0.4 min. Peak picking using these parameters resulted in a list of 1535 signals organized as  $m/z$  – retention time pairs.

### **5.3.2 Data processing for and relative mass defect filtering for mining publically available data sets for the discovery of novel terpenoid metabolites**

The objective of this exercise was to mine the *Hoodia gordonii* LC-MS data for steroidal glycosides (which may be diterpene glycosides) and therefore the three boundary conditions discussed in Chapter 2 were re-established. As discussed in Chapter 1, the simplest form of diterpene is geranylgeranyl. Therefore, rules can be established based on theoretical derivatives of geranylgeranyl. 1). Based on theoretical mass of geranylgeranyl mono glycoside, the minimum  $m/z$  of  $[M+H]^+$  of a diterpene monoglycoside would be 453.3211 and the maximum RMD a diterpene glycoside can display is 708 ppm. 2). Similarly, the molecular mass of Hoodigoside S molecule which reported from *H. gordonii*, has a diterpenoid core that is extensively “decorated” with sugars and sugar derivatives was established as the maximum mass a diterpene glycoside from this plant can exhibit. Therefore 1479.7730 was considered as the upper limit of  $m/z$  for  $[M+H]^+$  and 522 ppm was established as the lower limit of relative mass defect of diterpene glycosides from this plant. The distribution of RMD among compounds is shown in Fig. 5.3. From this analysis it is clear that a majority of the compounds detected from *Hoodia* consists of RMD in the range 500 – 700 ppm.

Therefore, for exploring diterpene glycosides from *Hoodia gordonii*, compounds of the RMD range 500 – 710 ppm were selected as candidate compounds for diterpene discovery. Next, this list of markers was organized in the descending order of peak area reported in the database and higher collision energy fragments observed in the multiplexed CID MS data was examined for

each marker in this list to establish their terpenoid identities.

Examining higher energy functions of multiplexed CID MS data enabled annotation of those compounds that generated fragment ions with increasing RMD compared to the precursor ion, and these were selected as candidate terpenoids. Next, the MS/MS spectra were generated for these compounds to confirm the fragment ions observed in multiplexed CID MS. Furthermore, since a number of diterpene glycosides are known from *H. gordonii*, MS/MS spectra were generated for some of these standards and the MS/MS spectra from the candidate terpenoid compounds selected in the previous step were compared with these data.

The Medicinal Plants Consortium database annotates 29 compounds as steroidal glycosides all of which display a RMD between 500 - 710 ppm. Among the 100 most abundant compounds with RMD values in this range, only 16 compounds are known. RMD filtering and comparing the MS/MS data for unknown compounds with those from standards led to the discovery of 21 novel steroidal glycoside compounds that are putative derivatives of known compounds (based on the presence of common fragment ions assigned based on the accurate mass) and five alkaloid compounds among the 100 most abundant markers with RMD range 500 – 710 ppm. The abundance of these novel compounds and the compounds annotated in the database are shown in Fig. 5.4.

### 5.3.3 Variation of RMD of terpene glycosides and other non –terpenoid compounds.

In addition to all the terpene glycosides that appeared in the RMD range of 500 – 710 ppm, the second most abundant compound in this list ( $m/z$  983.5523) appeared to be an alkaloid. The MS/MS spectrum (Fig. 5.17 a) of this compound revealed a number of even mass fragment ions ( $m/z$  520.3010, 492.3056, 466.2877, 310.1608, 282.1655) that suggested this compound contained an odd number of nitrogen atoms, suggesting a putative alkaloid.

Alkaloids can be distinguished from terpenoids based on the presence of even mass fragment ions (indicative of nitrogen) in the MS/MS spectra of alkaloids in positive ion mode. In addition, product ions generated from alkaloids also display a different variation in RMD compared to those generated from terpenoids (Fig. 5.17 b). For diterpene glycosides the characteristic gradual increase of RMD with decreasing mass of fragment ions can be observed. The decreasing RMD of the fragment ions was observable among fragment ions of this compound. Minimum of two isomeric forms were separated and detected for this compound. Based on accurate mass information, a glycosylated form of this compound also was assigned from the metabolite profile data.

### 5.3.4 Putative assignments of novel steroidal glycosides from *Hoodia gordonii*

A number of terpenoid glycosides that are potentially modified structures of the known compounds were assigned based on accurate mass information and the observation of product ions in MS/MS spectra that are common to standards. These compounds were wither Calogenin or Hoodigogenin A derivatives.

*Anhydrocalogenin triglycoside*: This compound was detected as  $[M+H]^+$  at  $m/z$  803.4136 (theoretical  $m/z$  803.4060 with elemental formula  $C_{39}H_{63}O_{17}$  for  $[M+H]^+$  ;  $\Delta m = 9.5$  ppm). The

MS/MS spectrum of this compound (Fig. 5.5) showed the presence of  $m/z$  487.1651 representing anhydro triglucose (theoretical  $m/z$  487.1657 with elemental formula  $C_{18}H_{31}O_{15}$  for  $[M-H_2O+H]^+$ ;  $\Delta m = 1.2$  ppm).  $m/z$  641.3565, 479.3042 and 299.2383 that represent diglycosylated, mono glycosylated and aglycone form of anhydrocalogenin core is formed as fragment ions after the sequential loss of the sugar moieties. (theoretical  $m/z$  299.2369 for calogenin core;  $\Delta m = 4.6$  ppm, 479.3042 and 641.3532 for calogenin mono and diglycoside;  $\Delta m = 8.1$  and 5.1 ppm respectively). Six isomeric forms of this compound were separated chromatographically.

Fragment ions of similar  $m/z$  was observed in MS/MS spectra of Hoodigoside L, Hoodigoside M, Hoodigoside R, Hoodigoside P, Hoodigoside Q, and Hoodigoside H, all of which are calogenin derivatives, further supporting that this compound represents anhydrocalogenin triglycoside (Figs 5.18 and 5.6). A related compound (a hydrated form representing of calogenin triglycoside) was observed as  $[M+NH_4]^+$  at  $m/z$  838 (experimental  $m/z$  838.4514; theoretical 838.4431 for elemental formula  $C_{39}H_{64}O_{18} + NH_4^+$ ;  $\Delta m = 9.8$  ppm) that corresponded to calogenin triglycoside. This compound also gave  $m/z$  479, 461, 325 and 299 as fragment ions in MS/MS spectra (Fig. 5.5) that were also observed in anhydrocalogenin triglycoside. Similarly a compound that was 162 Da less in mass than anhydrocalogenin triglycoside was observed at  $m/z$  641  $[M+H]^+$  (experimental  $m/z$  641.3590; theoretical 641.3532 for elemental formula  $C_{33}H_{53}O_{12} + H^+$ ;  $\Delta m = 9.0$  ppm). The MS/MS spectrum of this compound (Fig. 5.6) also displayed the presence of  $m/z$  299. However, no  $m/z$  325 was observed in this spectrum indicating that the two glucose moieties exist separately in the molecule instead of as a disaccharide.

A glycosylated form of Hoodigoside M was detected as a  $[M+NH_4]^+$  at  $m/z$  1304 (experimental  $m/z$  1304.6605; theoretical 1304.6481 for elemental formula  $C_{59}H_{98}O_{30} + NH_4^+$ ;  $\Delta m = 9.5$  ppm). This compound displayed the presence of two isomeric forms that were chromatographically



resolved. The MS/MS spectra of this compound and Hoodigoside M standard (Fig. 5.8 a,b, and c respectively) displayed fragments of  $m/z$  641.3552 corresponding to protonated form of diglycosylated anhydrocalogenin (theoretical  $m/z$  for elemental formula of  $C_{33}H_{53}O_{12}^+$  is 641.3532;  $\Delta m = 3.1$  ppm). However, in isomeric form 1, a fragment ion of  $m/z$  485 (experimental  $m/z$  485.2275; theoretical  $m/z$  for  $C_{20}H_{37}O_{13}^+$  485.2229) is observed. This corresponds to the trihexose group formed by the attachment of an additional glucose group to the dihexoside formed by D-thevetose and D-oleandrose of Hoodigoside M. In isomeric form 2,  $m/z$  485 is not observed and instead  $m/z$  649 (experimental  $m/z$  649.2223, theoretical  $m/z$  649.2186;  $\Delta m = 5.7$  ppm) occurs. This fragment corresponds to a tetraglycoside group indicating that the additional sugar group is attached to the triglucoside of Hoodigoside M. Additionally, the fragments corresponding to triglycoside ( $m/z$  487) and diglycoside ( $m/z$  325) groups are also observed indicating the presence of the sugar groups in both isomeric forms. Fragments corresponding to the calogenin terpenoid core  $m/z$  317 and 299 are observed in both the Hoodigoside M standard and in both isomeric forms.

Another calogenin derivative was observed as  $[M+NH_4]^+$  at  $m/z$  1190 (experimental  $m/z$  1190.6805; theoretical 1190.6681 for elemental formula  $C_{59}H_{96}O_{23} + NH_4^+$ ;  $\Delta m = 10.4$  ppm). The MS/MS spectrum of this compound is shown in Figure 5.7. This compound yields a product ion at  $m/z$  479, which corresponds to the glycosylated calogenin as a prominent fragment ion. In addition it displays  $m/z$  317 and 299 (theoretical  $m/z$  317.2475;  $\Delta m = 2.2$  ppm and  $m/z$  299.2369;  $\Delta m = 6.3$  ppm) as fragment ions that represent the presence of the calogenin core. The base peak observed in MS/MS of this compound  $m/z$  227 (experimental  $m/z$  227.1296; Theoretical  $m/z$  227.1278;  $\Delta m = 7.9$  ppm) corresponds to an anhydrocymarose with the aliphatic

2-methylbutenoate chain attached to it.

Calogenin forms another abundant compound detected as  $[M+NH_4]^+$  at  $m/z$  920 (experimental  $m/z$  920.4928; theoretical 920.4850 for elemental formula  $C_{44}H_{70}O_{19} + NH_4^+$ ;  $\Delta m = 8.4$  ppm). The MS/MS spectrum of this compound (Figure 5.10) indicates the presence of a disaccharide group (detected as fragment ion with  $m/z$  325.1157 (Theoretical  $m/z$  325.1129;  $\Delta m = 7.4$  ppm) and the glycosylated calogenin core detected as  $m/z$  461. Similar compound is observed at  $m/z$  998 as a  $[M+NH_4]^+$  (experimental  $m/z$  998.5272). The MS/MS of this compound (Figure 5.11) also displays fragment ions corresponding to a disaccharide group (detected as fragment ion with  $m/z$  325.1105 (Theoretical  $m/z$  325.1129;  $\Delta m = -7.0$  ppm) and glycosylated calogenin core (detected as fragment ion with  $m/z$  461.2925 (theoretical  $m/z$  461.2898;  $\Delta m = 5.5$  ppm) indicating it is a calogenin derivative. Other common calogenin fragment ions  $m/z$  299 and 281 (Experimental  $m/z$  299.2365 and 281.2285 respectively; theoretical  $m/z$  299.2369 and 281.2264;  $\Delta m = 1.3$  ppm and 7.4 ppm respectively) are observed further supporting this conclusion.

Hoodigogenin A derivatives: Aula *et al* proposed a fragmentation pathway for P57 that explains the formation fragment ions  $m/z$  457, 313 and 295 [1]. These same fragment ions are observed in the MS/MS fragmentation of 1316 (Fig. 5.12), 1206 (Fig. 5.13), 1222 (Fig. 5.14), 1192 (Fig. 5.15) and 916 (Fig. 5.16), indicating these compounds share a similar terpenoid core as P57 (Hoodigogenin A) and glycosylation. Based on accurate mass it can be inferred that  $m/z$  1316 represents a  $[M+NH_4]^+$  form of a P57 modified with one additional oleandrose and two digitoxose moieties. All these compounds form fragment ions  $m/z$  457 (experimental  $m/z$  457.2986; theoretical 457.2949 for elemental formula  $C_{28}H_{41}O_5^+$ ;  $\Delta m = 8.0$  ppm), 313

(experimental  $m/z$  313.2177; theoretical 313.2162 for elemental formula  $C_{21}H_{29}O_2^+$ ;  $\Delta m = 4.8$  ppm) and 295 (experimental  $m/z$  295.2074; theoretical 295.2056 for elemental formula  $C_{21}H_{27}O^+$ ;  $\Delta m = 6.1$  ppm).

## 5.4 Conclusions

Our investigation has disclosed the presence of an assortment of novel diterpene glycosidic derivatives of known compounds, their isomers and novel alkaloids. This is the first report to our knowledge that reports of alkaloids from *Hoodia gordonii*. This is an important finding since a number of alkaloids are associated with manipulation of neuronal activity [7]. Therefore, potentially these alkaloids also might be responsible for the biological effects of *Hoodia* extracts. Further investigation is required for establishing the structure of these compounds and their biological activity.

Applying RMD filtering to metabolite profiles of *Hoodia* has resulted in the discovery of a number of novel terpene glycosides. It is likely that some of these are biosynthetic intermediates of the diterpene glycosides observed in *Hoodia*. For instance the calogenin derivatives such as anhydrocalogenin tri glycoside and monoglycoside can be expected to be biosynthetic intermediates involved in the synthesis of larger calogenin compounds. Therefore, these compounds may hold the key to understanding the diterpene biosynthesis in *Hoodia*.

The discovery of a number of novel diterpene glycosides is an indication that the chemical diversity of terpenoid glycosides in *Hoodia* is even greater than currently reported. The multiple types of glycosylations and the presence of a number of isomeric forms of a given compound add yet further complexity to the chemical diversity of *Hoodia*.

## **APPENDIX**

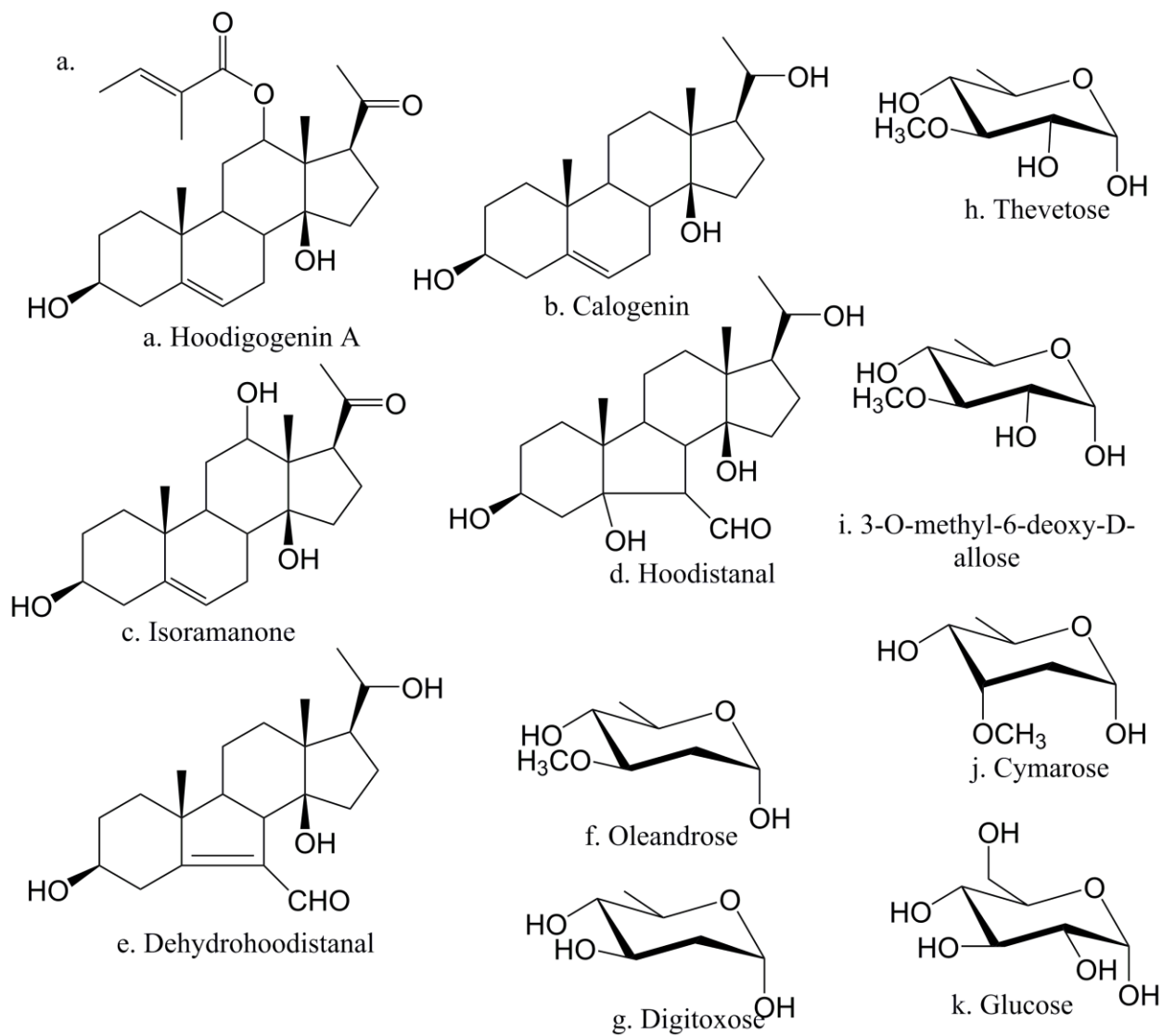


Figure 5.1: Five diterpenoid cores reported from *Hoodia* compounds (a-e) and the six different sugar groups that are found attached to the terpenoid core in current literature.

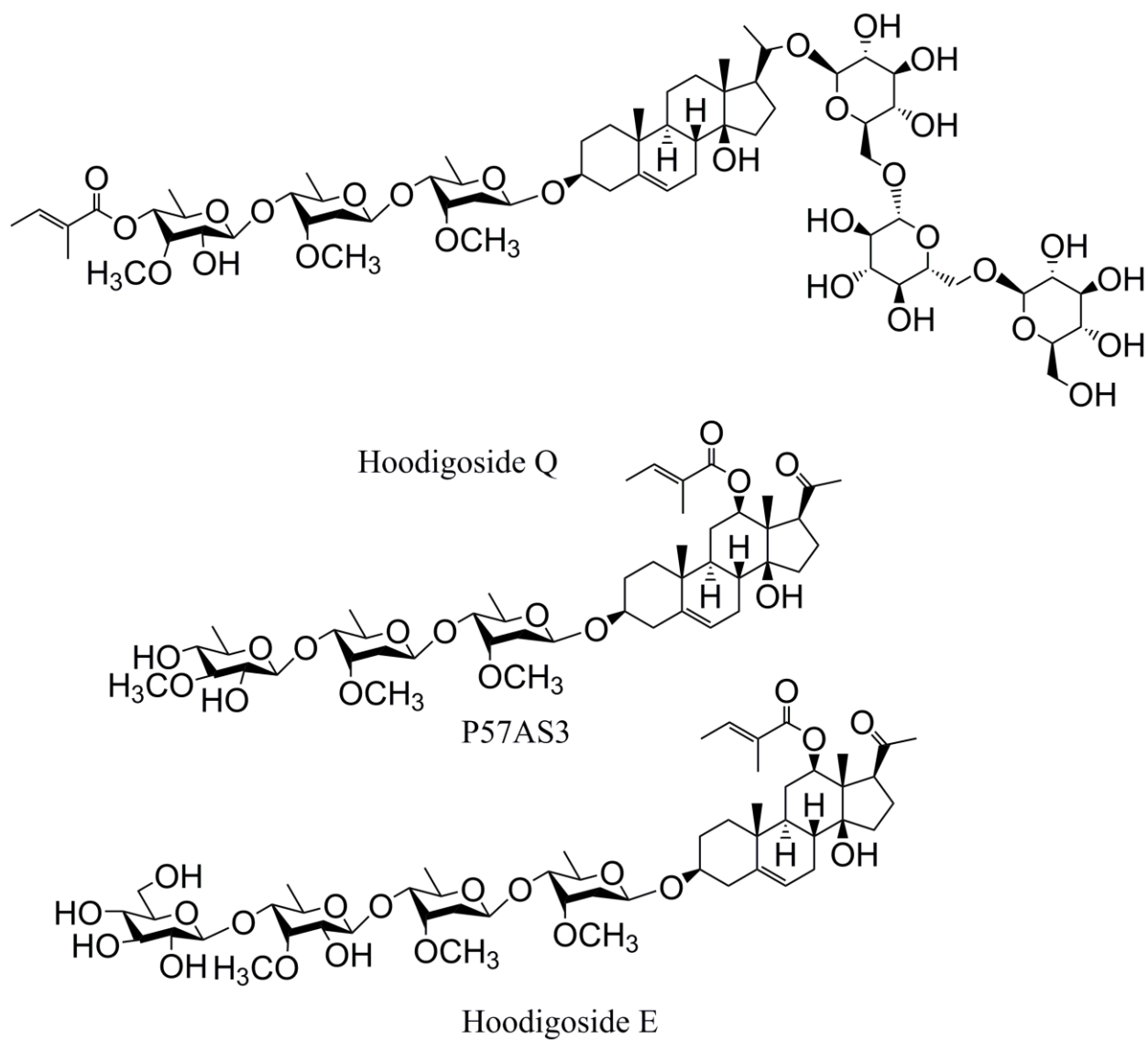


Figure 5.2: Some representative diterpenoids found in *Hoodia*.

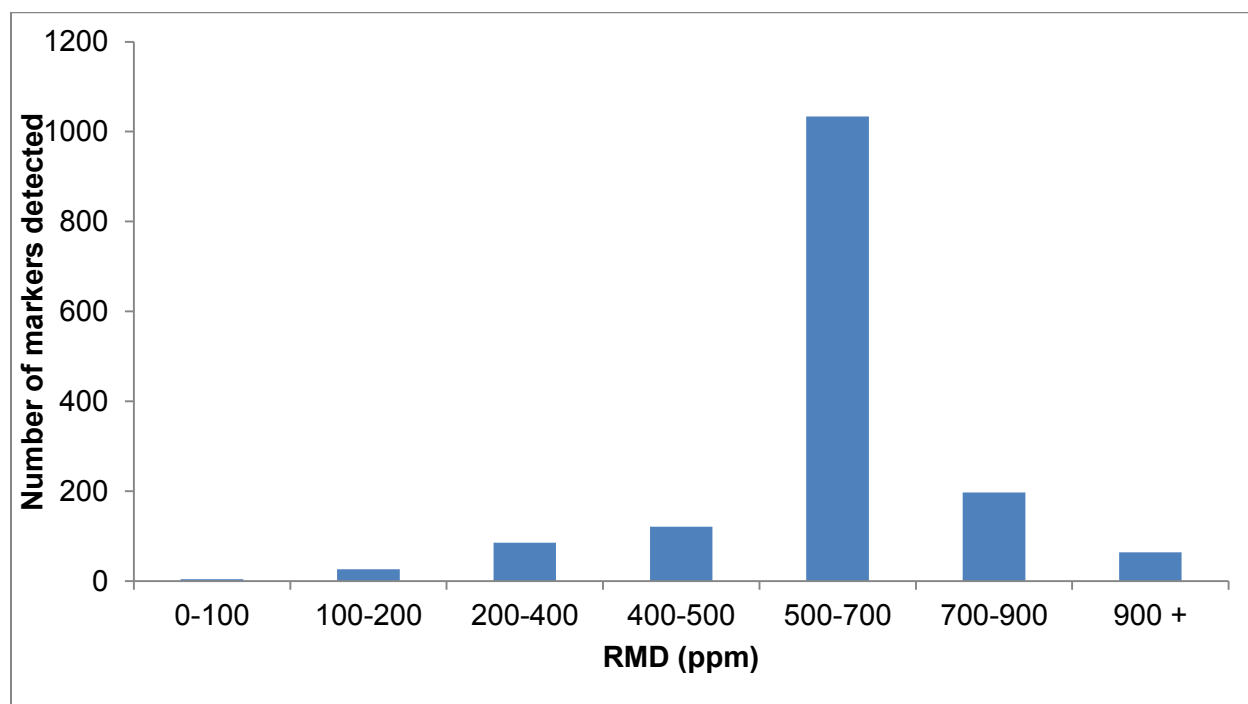


Figure 5.3: Distribution of RMD of ion detected in *Hoodia gordonii* from metabolite profiling in different RMD range categories.

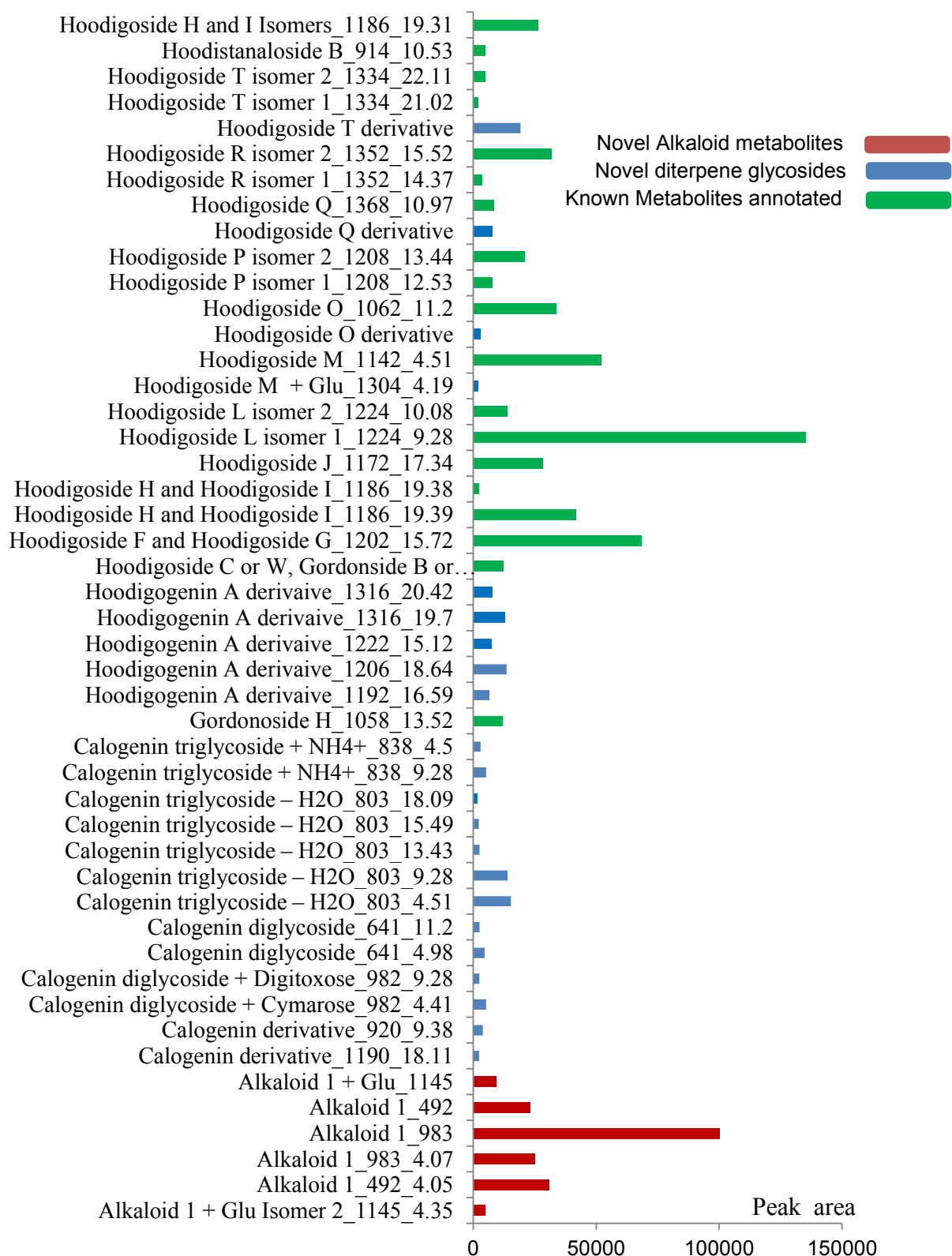


Figure 5.4: Abundance of known and novel metabolites annotated from *Hoodia gordonii*



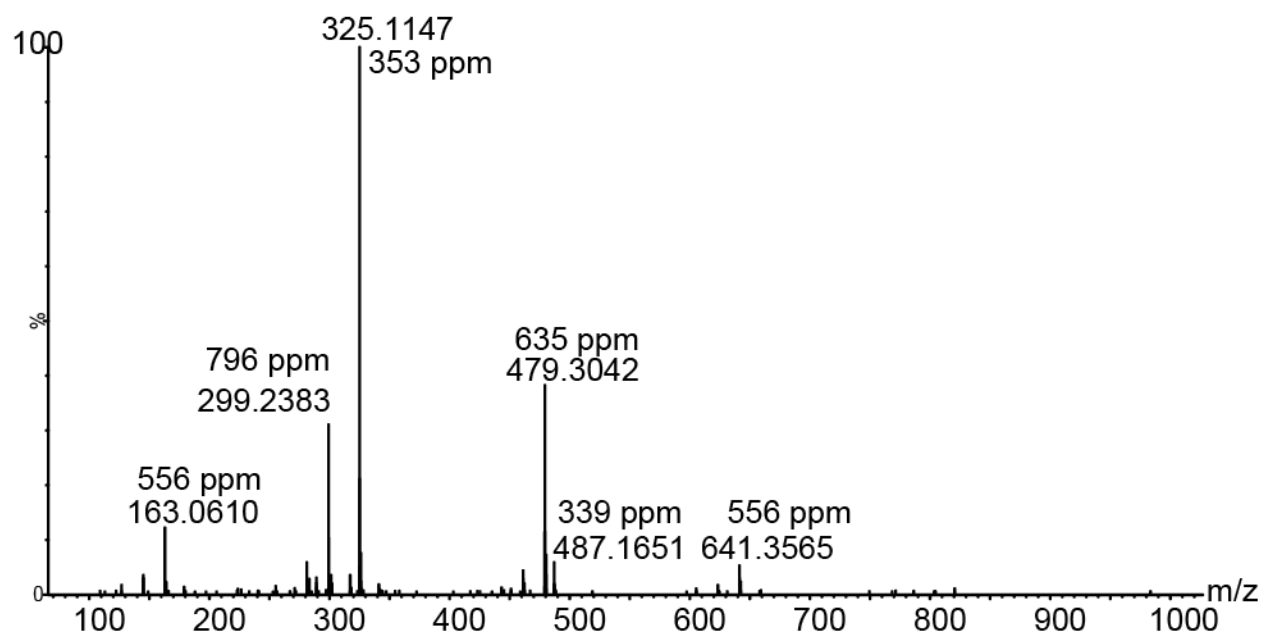


Figure. 5.5 Positive ion MS/MS product ion spectrum of Hoodia metabolite  $m/z$  803  $[M+H]^+$

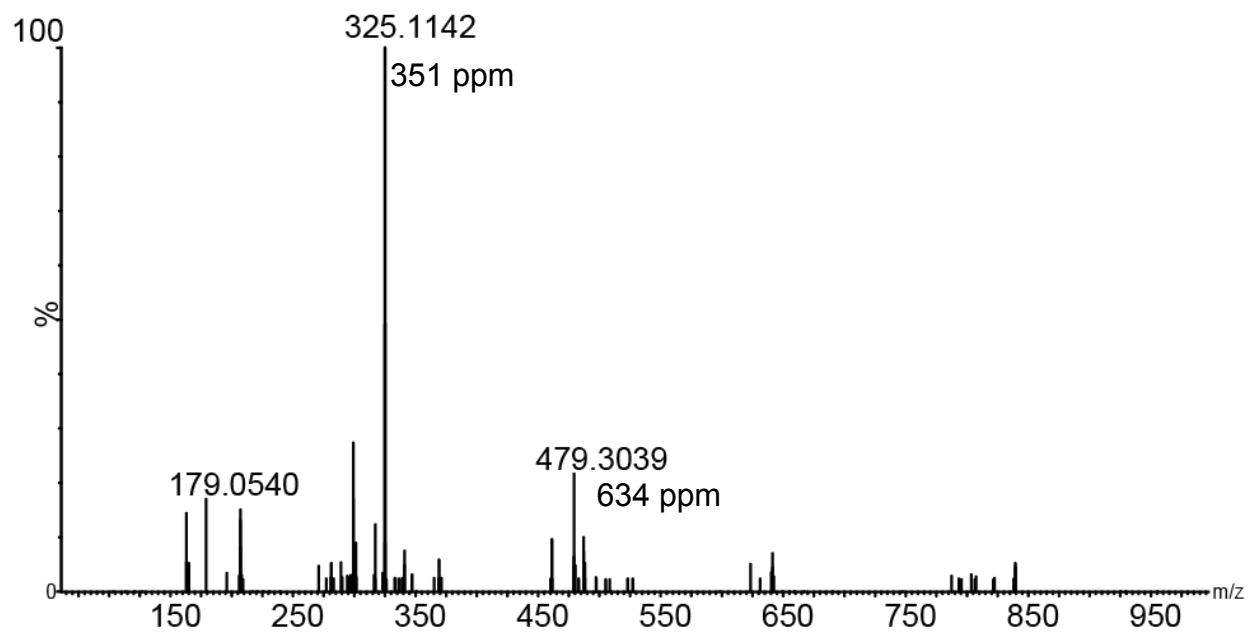


Figure 5.6 Positive ion MS/MS product ion spectrum of Hoodia metabolite  $m/z$  838  $[M+NH_4]^+$

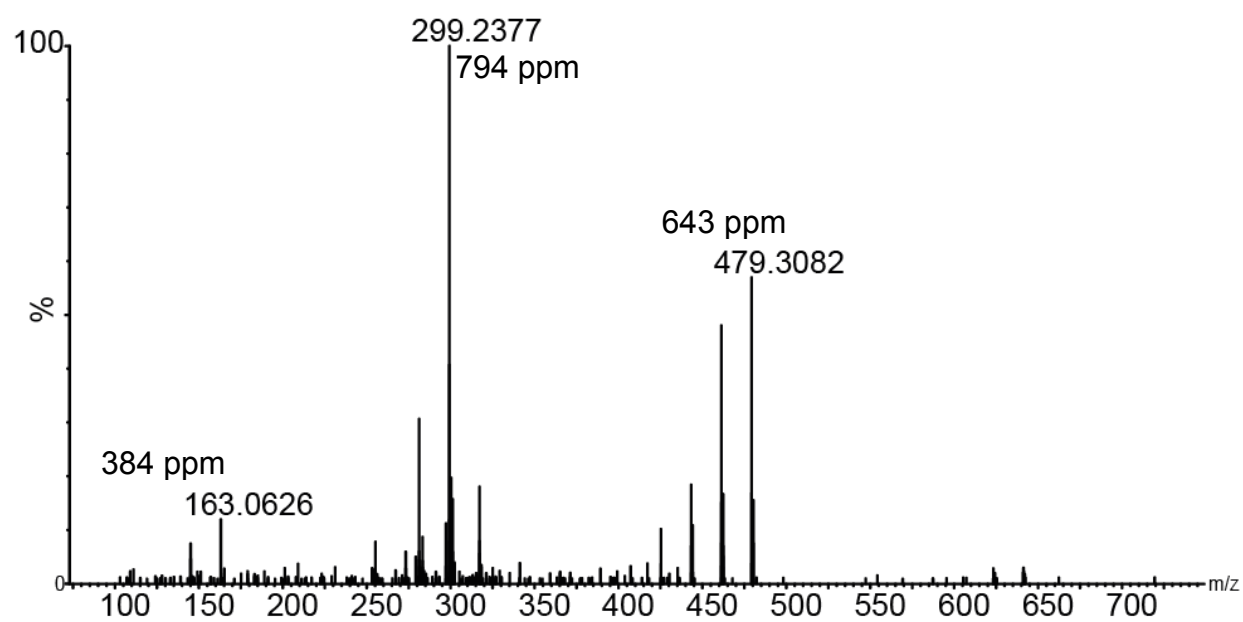


Figure 5.7 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  641.

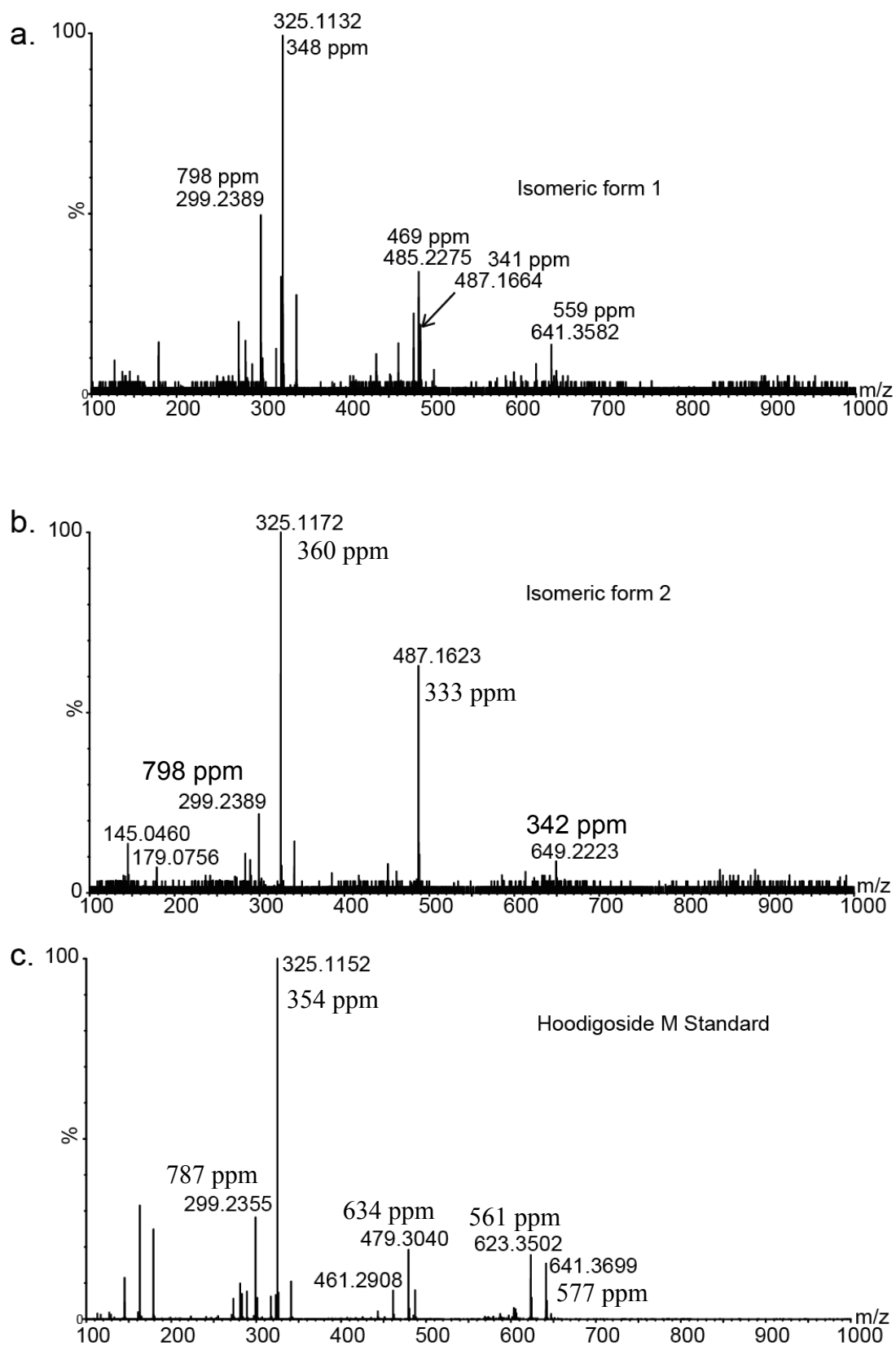


Figure 5.8 Positive ion product ion MS/MS spectrum of Hoodia metabolite of  $m/z$  1304.

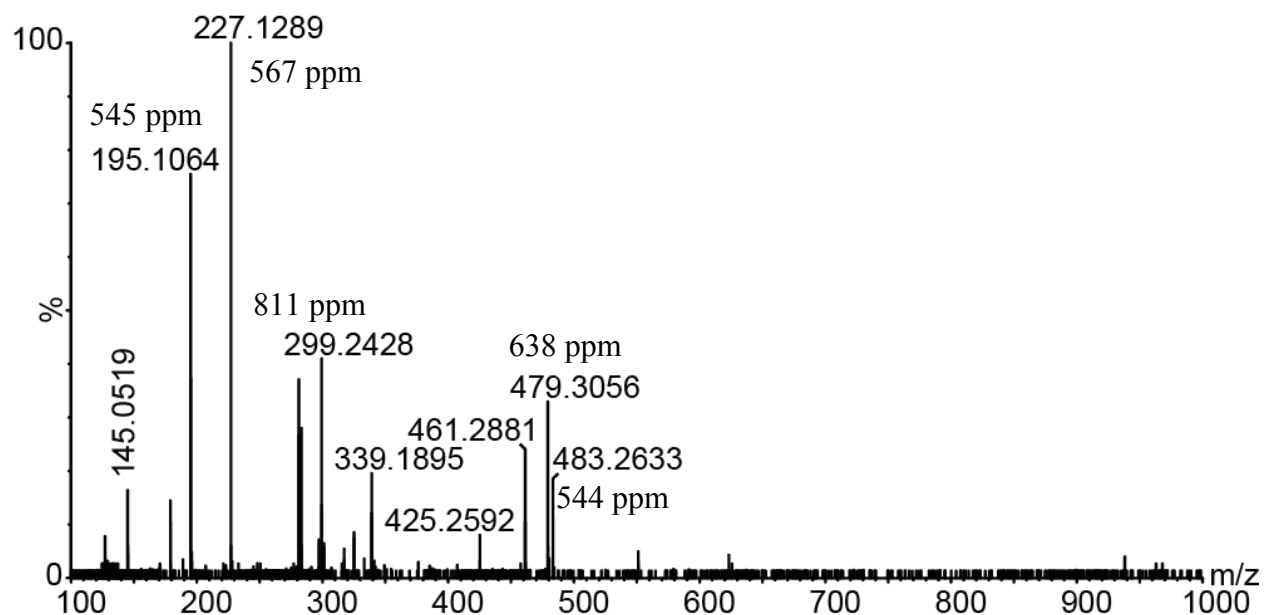


Figure 5.9 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  1190.98

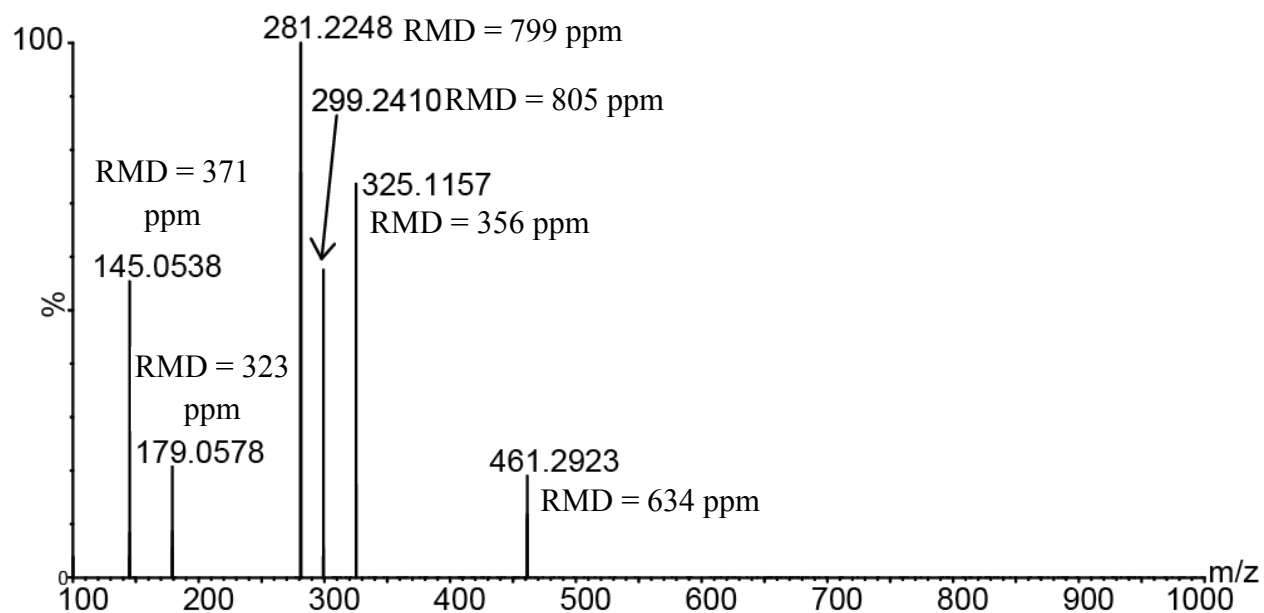


Figure 5.10 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  920

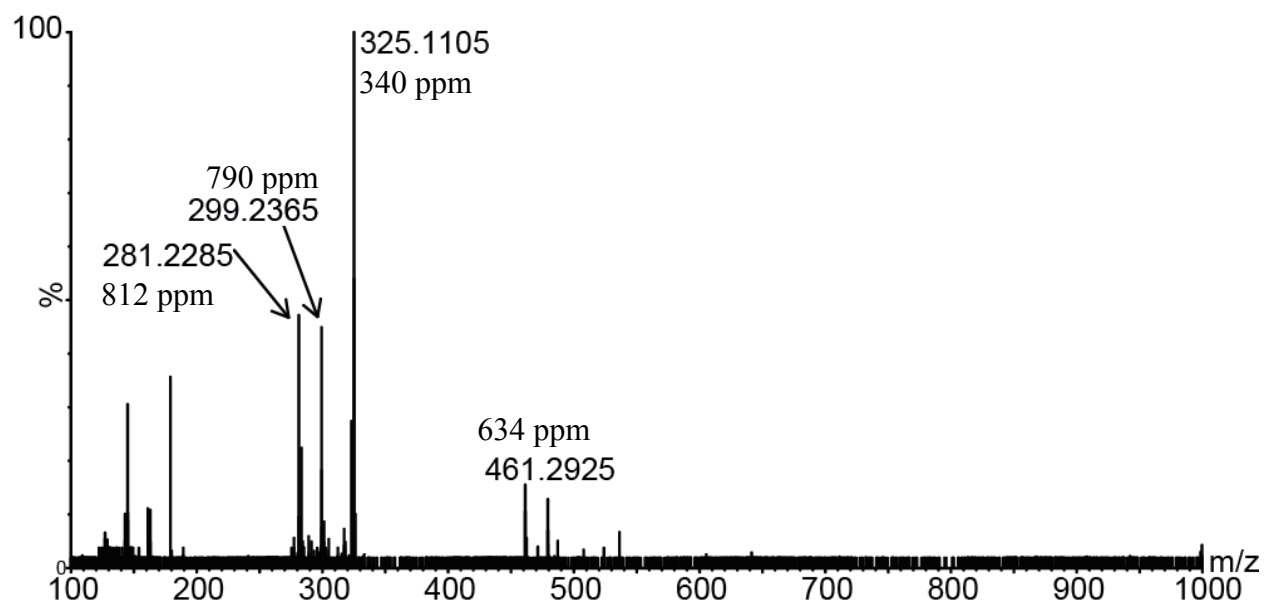


Figure 5.11 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  998 100

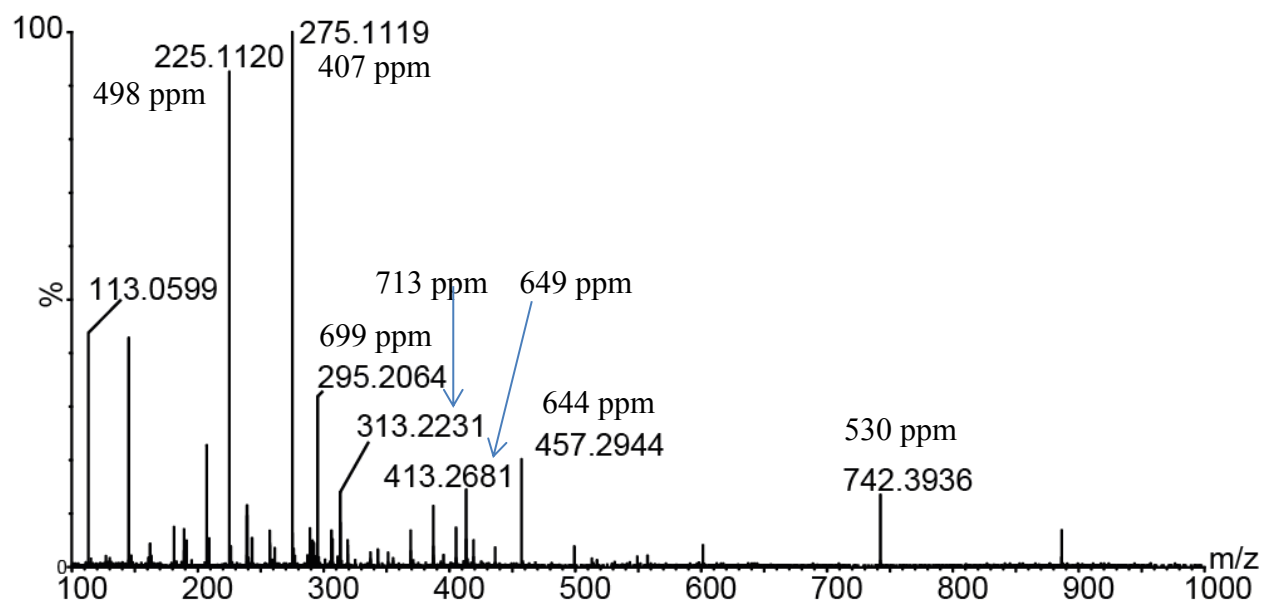


Figure. 5.12 Positive ion MS/MS product ion spectrum of  $m/z$  1316.

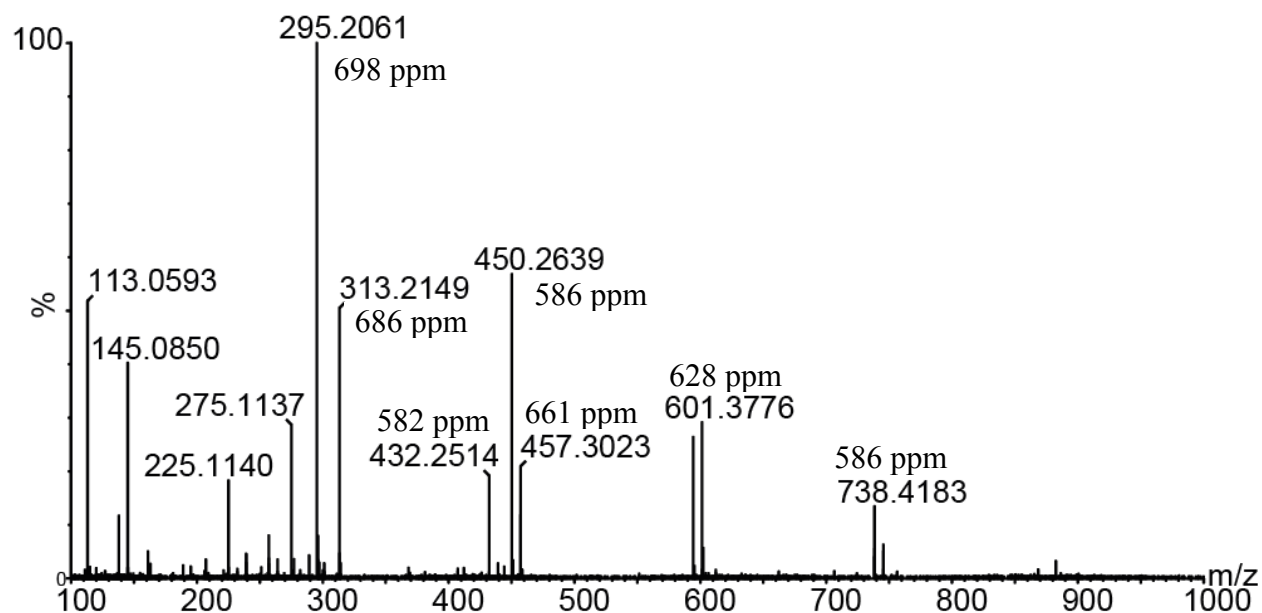


Figure. 5.13 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  1206

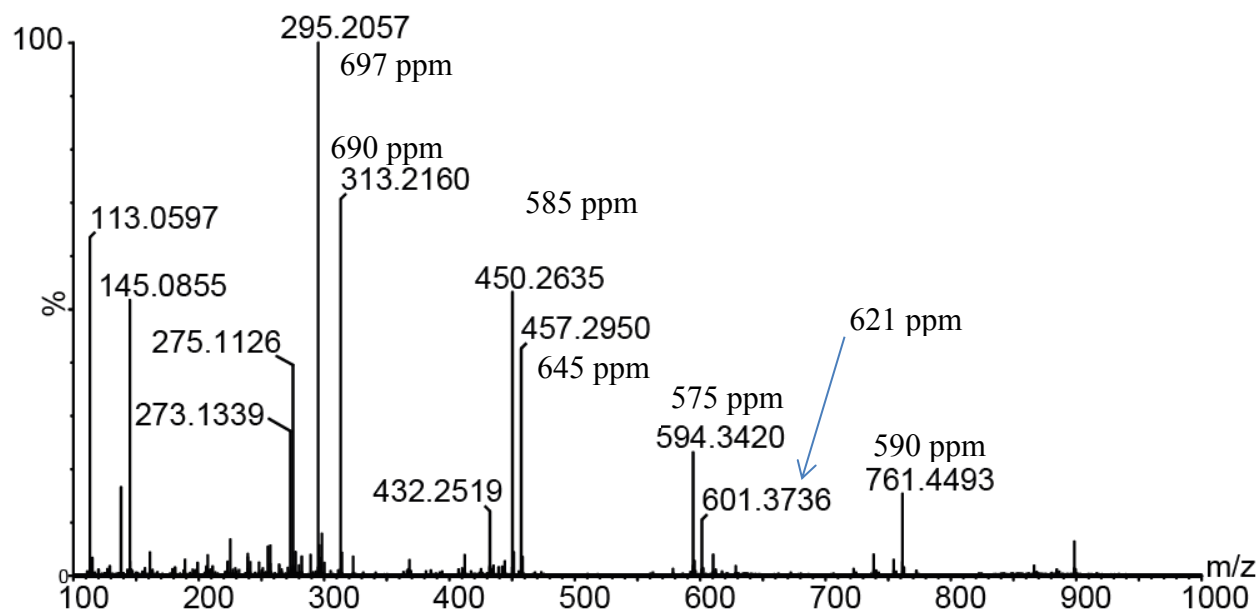


Figure. 5.14 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  1222

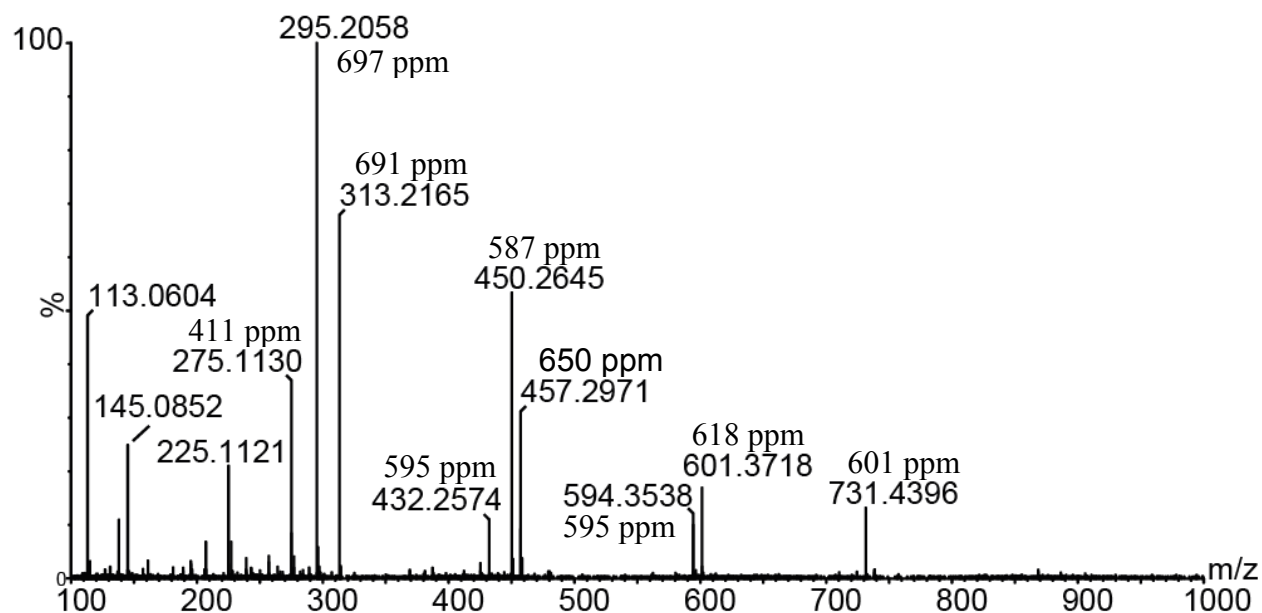


Figure. 5.15. Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  1192

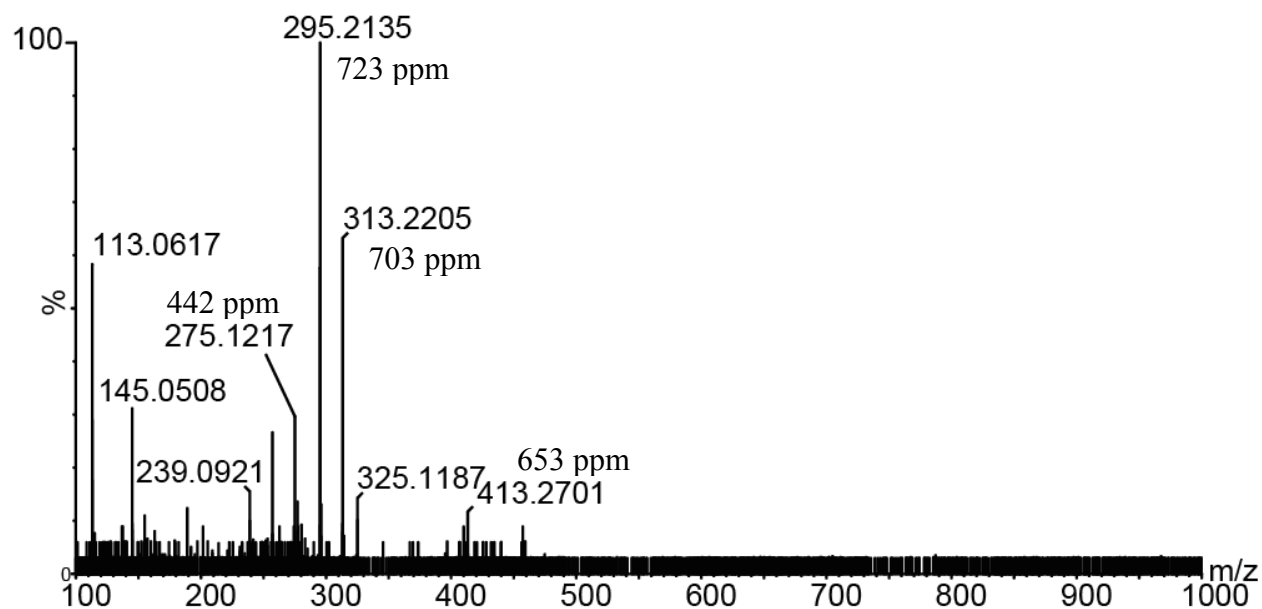


Figure. 5.16 Positive ion MS/MS product ion spectrum of Hoodia metabolite of  $m/z$  916

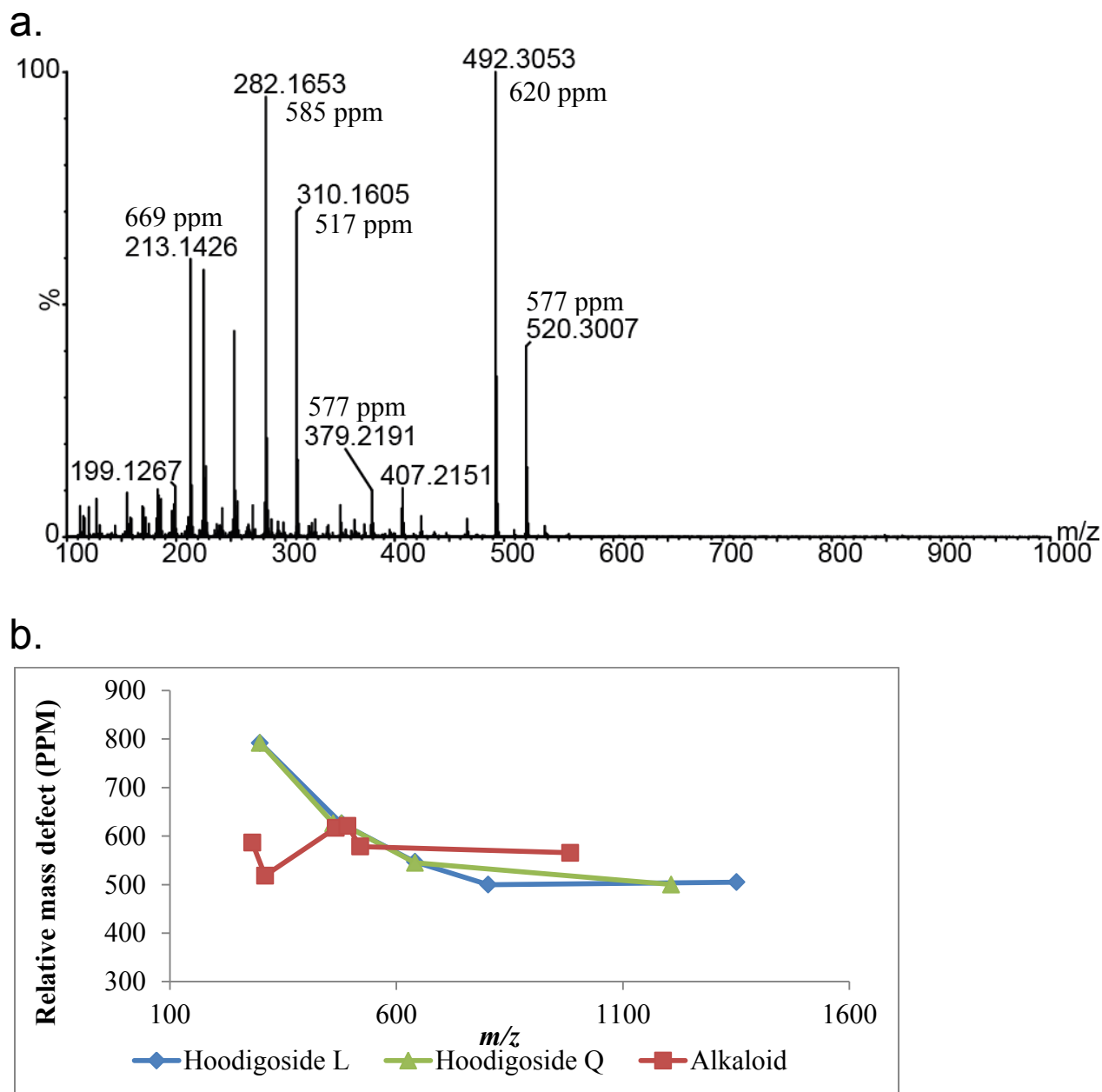


Figure. 5.17 Positive ion MS/MS product ion spectrum of (a)  $m/z$  983 (putative alkaloid), (b) Variation of RMD of parent ion and product ions in the MS/MS spectrum of an alkaloid compared to the diterpene glycoside (RMD of sugar fragments are not included). The diterpene glycoside displays a gradual increase of RMD from the parent ion to fragment ions whereas the alkaloid displays a constant RMD from both fragment ions and parent ions.



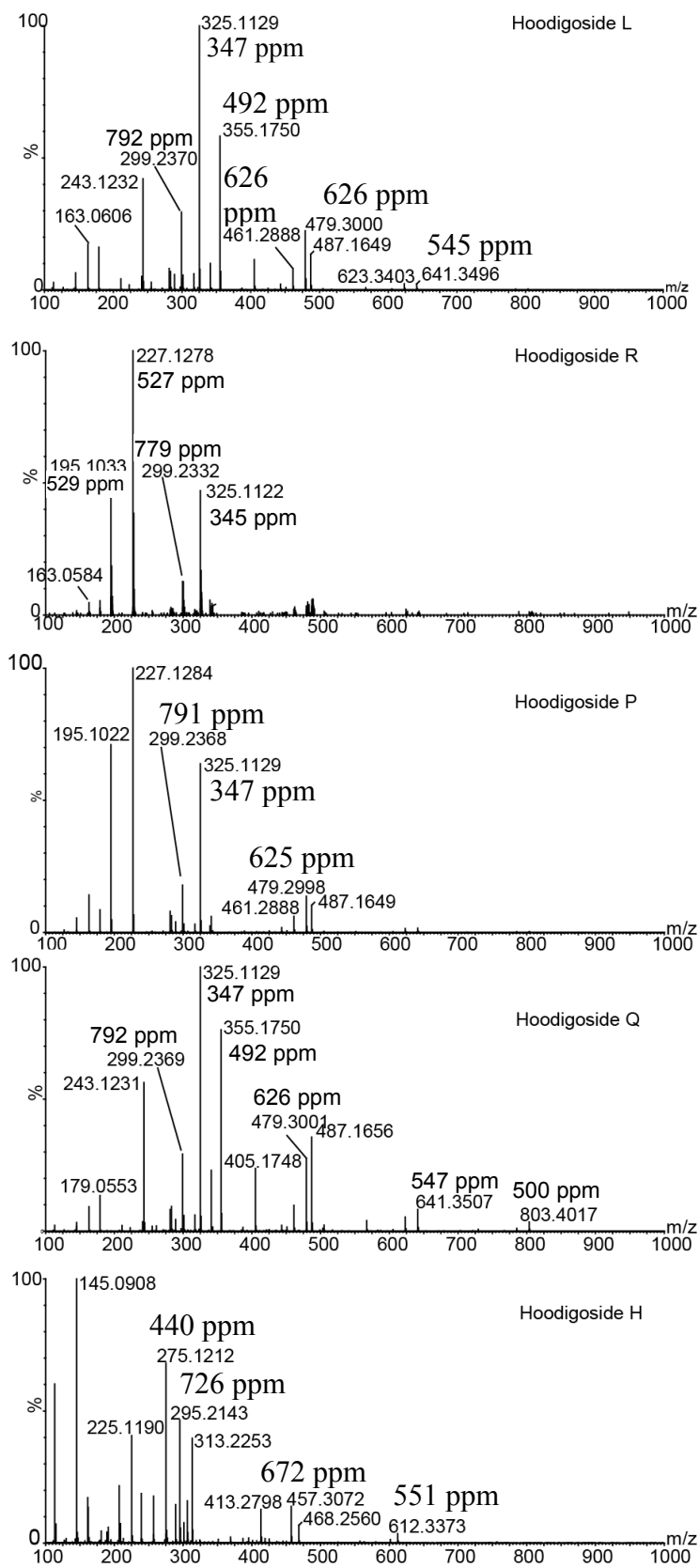


Figure. 5.18 Positive ion MS/MS product ion spectra of *Hoodia* standards.

## REFERENCES

## REFERENCES

1. Avula, B., et al., Identification and structural characterization of steroidal glycosides in *Hoodia gordonii* by ion-trap tandem mass spectrometry and liquid chromatography coupled with electrospray ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*, 2008. 22(16): p. 2587-96.
2. van Heerden, F.R., *Hoodia gordonii*: A natural appetite suppressant. *Journal of Ethnopharmacology*, 2008. 119(3): p. 434-437.
3. van Heerden, F.R., et al., An appetite suppressant from *Hoodia* species. *Phytochemistry*, 2007. 68(20): p. 2545-2553.
4. Pawar, R.S., et al., New oxypregnane glycosides from appetite suppressant herbal supplement *Hoodia gordonii*. *Steroids*, 2007. 72(6-7): p. 524-534.
5. Bindra, J., M. Cawthorne, and I. Rubin, Extracts, compounds & pharmaceutical compositions having anti-diabetic activity and their use, 2002, Google Patents.
6. Hakkinen, J., R.M. Horak, and V. Maharaj, Steroidal glycosides or plant extracts for treatment of gastric acid secretion damage. 2001: patent; EP1099444.
7. Vermaak, I., J.H. Hamman, and A.M. Viljoen, *Hoodia gordonii*: An Up-to-Date Review of a Commercially Important Anti-Obesity Plant. *Planta Medica*, 2011. 77(11): p. 1149-1160.
8. Shukla, Y.J., et al., Pregnane glycosides from *Hoodia gordonii*. *Phytochemistry*, 2009. 70(5): p. 675-683.
9. Dall'Acqua, S. and G. Innocenti, Steroidal glycosides from *Hoodia gordonii*. *Steroids*, 2007. 72(6-7): p. 559-568.
10. Schillmiller, A.L., R.L. Last, and E. Pichersky, Harnessing plant trichome biochemistry for the production of useful compounds. *Plant Journal*, 2008. 54(4): p. 702-711.

## **Chapter 6: Concluding remarks**

Increasing world population, climate change, depletion of non-renewable energy sources such as fossil fuels has made discovering novel sources of chemical feed stocks an important endeavor for the survival of mankind. This requirement has resulted in increased dependence on plants as a food and energy source of humans. The field of plant sciences has approached this challenge by genetically modifying plants to enhance production of food and renewable substances. This has prompted the requirement for discovering the genes involved in plant specialized metabolism in order to understand their functionality and to manipulate their behavior. Plant metabolomics has contributed to this by discovering the compounds that are important for humans from plants and providing information of their structure, quantity and location.

Research discussed in Chapter 2 of this dissertation resulted in the development of a novel strategy for the discovery of glycosylated terpenoids, a class of compounds for which annotation has posed a significant challenge for many years owing to their low levels of abundance and limitations of methods for their annotation in mixtures. We addressed this challenge by applying Relative Mass Defect (RMD) filtering which allowed for separating a substantial amount of unrelated compounds from glycosylated terpenoids. RMD filtering is applicable for the discovery of any kind of conjugated terpenoid as demonstrated in the application of the method for the discovery of sesquiterpene glycosides and diterpene glycosides (Chapter 5).

However, the discovery of these compounds has opened up a number of novel research challenges. First, what are the structural features that yield multiple isomers of each group of metabolites? RMD filtering allows for putative annotation of glycosylated terpenoids, but does not provide the final word regarding metabolite structure. The presence of numerous isomers of so many terpenoid glycosides presents a substantial challenge in purifying sufficient quantities

for NMR analysis. This issue has been addressed to some extent by purifying and elucidating the structures of several terpenoid metabolites (Chapter 3 and 4), but absolute configurations of some chiral centers remain to be determined. Owing to the lack of plant material, the compounds discovered from *Hoodia gordonii* (Chapter 5) are annotated only based on mass spectrometric evidence, and elucidating their structures using NMR still remains to be performed.

All the profiling and RMD filtering work reported here has resulted in the identification of a number of isomeric forms for each compound. Therefore an important question that needs answering is what are the structural differences among these isomers? Some of the isomerism may arise from the differences in the carbohydrate linkages, and these are often not apparent from mass spectrometry data alone. Also, the positions of substitution of additional groups such as malonyl or acetate esters may be different from one isomer to another. To establish these differences, more isomeric forms of each compound need to be purified and their structures elucidated using NMR and x-ray crystallography.

A third challenging questions is, do different accessions of the same plant species accumulate the same compounds or different compounds? For instance in the work presented in this dissertation, only *Solanum habrochaites* LA1777 leaf trichomes have been discussed. Preliminary data (not presented here) in metabolite profiling from other plant accessions of tomato in addition to LA1777 (These were LA1624, LA2975, LA1362, LA2107, LA1392, LA2098, LA2409, LA2101 LA2144, LA1223, LA1927 LA2106 and LA2204) has provided evidence that the isomer profile of a given compound (e.g. campherenane-2,12-diol diglycoside malonate ester) is different between different accessions. These data need further exploration to develop comparisons between different accessions and to explore these nonvolatile conjugated terpenoids further.

In summary our research presented here has opened up a number of opportunities for future scientists to explore plant terpenoid chemistry that has not been known for many decades and this can potentially result in the discovery of future drugs, chemical feed stock for industries, methods to develop better crops, food additives, and flavor compounds .