

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

PROFILING OF SPECIALIZED METABOLITES IN GLANDULAR TRICHOMES OF
THE GENUS *SOLANUM* USING LIQUID CHROMATOGRAPHY AND MASS
SPECTROMETRY

By

Feng Shi

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Chemistry

2009

ABSTRACT

PROFILING OF SPECIALIZED METABOLITES IN GLANDULAR TRICHOMES OF THE GENUS *SOLANUM* USING LIQUID CHROMATOGRAPHY AND MASS SPECTROMETRY

By

Feng Shi

The plant kingdom synthesizes thousands of biologically active secondary metabolites, but our understanding of the functions of genes responsible for regulating their biosynthesis is far from complete. Existing knowledge of the diversity of plant chemistry is also remarkably limited, but an emerging approach known as metabolomics offers the potential for explosive growth in this area of research. Accelerating the pace of discoveries in this realm will require more rapid and powerful analytical tools capable of identifying and quantifying from hundreds to thousands of secondary metabolites using nontargeted metabolite profiling. Since the majority of plant metabolites are either polar or nonvolatile molecules, efforts to push the performance limits of coupling liquid chromatographic separations with mass spectrometry (LC/MS) are bringing about a new “Golden Age” of plant biochemistry.

This thesis describes an approach that takes advantage of fast spectrum acquisition available with time-of-flight mass spectrometers to extend nonselective collision induced dissociation (CID) by performing quasi-simultaneous acquisition of mass spectra using multiple collision conditions. This approach, termed multiplexed CID (muxCID), generates accurate mass measurements of molecular and fragment ion species with the additional benefit that the dependence of each ion's abundance upon collision energy is obtained throughout the entire course of an LC/MS analysis. The additional information obtained from ion breakdown curves, coupled with accurate mass measurements, aids

assignments of ions as molecular, fragment, noncovalent oligomer, or adducts. Coeluting metabolites are resolved using collision energy-selective fragmentation of individual metabolite classes.

This multiplexed CID approach has been adapted for use with fast ultraperformance fused-core LC separations to provide an analytical platform capable of measuring about 100 plant metabolites per minute of instrument time. Rapid profiling of secondary metabolites was performed for extracts of specialized epidermal structures called glandular trichomes for numerous relatives of tomato plus plant lines derived from tomato and a wild relative. This technology has made it possible to screen about 300 samples within 30 hours. Automated data processing strategies yielded measurements of more than 1500 analytical signals. Multivariate statistical analysis of the metabolite profiles revealed five specific phenotypes differing in abundances and profiles of bioactive compounds including acylsugars, glycoalkaloids, and flavonoids, including several novel metabolites. Mass spectra generated using collision induced dissociation also provided important information regarding metabolite structures that have opened windows into the metabolic diversity within the plant genus *Solanum*. This work was extended to use metabolite profiles to help guide efforts to breed tomato plants rich in anti-insect acylsugar metabolites.

The analytical methods and approaches described herein will serve to expand the potential of LC/MS for metabolite identification in metabolomics and lead the way to speeding up discoveries of gene functions in plants and other organisms.

ACKNOWLEDGEMENT

Many people deserve thanks and appreciation for this thesis. First and foremost, I own my sincerest gratitude to my advisor Dr. A. Daniel Jones for his continuous support of my Ph.D research. As greatest advisor and mentor in my life ever, his patience, knowledge, motivation and consideration guide and light my journey up to today. I could not imagine my research without him.

A special thank from me to my outstanding collaborator Dr. Rob Last, also the principle investigator of the project, for his invaluable advice and critiques, which would benefit not only the research but my feature career. I also would like to thank my other collaborators Dr. Gregg Howe, Dr. Martha Mutschler from Cornell University, Dr. Eran Pichersky from University of Michigan and Dr. David Gang from University of Arizona. It is an honor to learn from them and work with them. All the open minded discussions are memorable.

I also would like to acknowledge the postdocs Dr. Jin-ho Kang, Dr Tony Schillmiller in this project for being good colleagues with me. We have great collaborations to make this project going smoothly and create the foundation for this thesis.

It is a pleasure to thank my committee: Dr. Gavin Reid, Dr. Merlin Bruening and Dr. Kevin Walker, for their insightful comments and criticism, and encouragement.

Specially, I would like to express my thanks to my coworkers Mike, Ruth, Siobhan, Chao, Ramin, Behnaz, Bao, xiaoli from Jone's group and Jeongwoon, Amanda from Last's group for all the assistant and happiness they gave to me. I would also like to thank Lijun and Bev from mass facility for their help during these years.

Last but never the least, my deeply thank to my family: my husband Hongyang, Li for always supporting me and sharing the happiness and difficulties with me; my little daughter Sophia for being such a sweet baby and bringing so much joy into our life; my parents and my parents in law, for taking care of the baby and helping us with the housework. I am really grateful to have them around me. Their unconditional love and support accompany with me through these tough years and make this thesis going so well. I definitely could not work through my Ph.D. without them.

TABLE OF CONTENTS

List of Tables	x
List of Figures	xii
List of Abbreviations	xxiii
Chapter 1 Introduction	1
1.1 The Emergence of Metabolomics in Biological Research	3
1.2 Challenge of Metabolomics	6
1.3 Techniques for Metabolomics	7
1.4 Multivariate Statistic tools for Metabolomic Analyses	11
1.5 Metabolite Identification	16
1.6 Secondary Metabolites in <i>Solanum</i> Glandular Trichomes	18
1.7 Summary of Research Goals	21
Chapter 2 Discovery and deep profiling of metabolites using liquid chromatography/multiplexed collision-induced dissociation (muxCID)/mass spectrometry	23
2.1 Introduction	24
2.2 Experimental Section	26
2.2.1 Materials	26
2.2.2 Chemicals	26
2.2.3 Extraction Method	27
2.2.4 Liquid Chromatography	27
2.2.5 Mass Spectrometry	27
2.3 Results and Discussion	28
2.4 Conclusions	63
Chapter 3 Annotation of Acylsugar metabolites in glandular trichomes from accessions of the genus <i>Solanum</i> based on LC/TOF MS coupled with multiplexed CID	65
3.1 Introduction	66
3.2 Experimental Section	68
3.2.1 Materials	69
3.2.2 Chemicals	69
3.2.3 Extraction Method	69
3.2.4 High-Performance Liquid Chromatography	69

3.2.5 Mass Spectrometry	69
3.2.6 Characterization of Metabolites	70
3.2.7 Nomenclature	70
3.3 Results and Discussion	71
3.3.1 Structure annotation for glucose trimesters in <i>S. pennellii</i> LA0716	71
3.3.2. Structure annotation for glucose trimesters and sucrose trimesters in <i>S. pennellii</i> LA1522	90
3.3.3. Structure annotation of sucrose tetraesters in <i>S. habrochaites</i> LA1777	113
3.3.4 Structure annotation of sucrose triesters and tetraesters in <i>S. habrochaites</i> LA1353	127
3.4 Conclusions	134
Chapter 4 Annotation of polyphenol, glycoalkaloid, and terpenoid secondary metabolites in glandular trichomes from tomato, its wild relatives, and near isogenic lines, based on LC/TOF MS coupled with multiplexed CID	136
4.1 Introduction	137
4.2 Methods	141
4.2.1 Materials	141
4.2.2 Chemicals	141
4.2.3 Extraction Method	141
4.2.4 High-Performance Liquid Chromatography	142
4.2.5 Mass Spectrometry	142
4.3 Results	143
4.3.1 Structural annotation for polyphenols	143
4.3.2 Structural annotations for glycoalkaloids	163
4.3.3 Structural annotation of terpenoids	176
4.4 Conclusions	184
Chapter 5 Fast LC/time-of flight mass spectrometry for screening metabolic phenotypes for tomato chromosomal substitution lines	185
5.1 Introduction	186
5.2 Experimental Method	189
5.2.1 Plant Growth Conditions	189
5.2.2 Plant Extractions	190
5.2.3 Chromatography	190
5.2.4 Mass Spectrometry	190
5.2.5 Chemometric Data Analysis	191
5.3 Results and Discussion	191
5.3.1. Analysis of Chromosome Substitution Lines Using LC/TOF MS	191

5.3.1.1	ILs 1-3 and 1-4 are missing an acetyl group on major acyl sucrose metabolites	204
5.3.1.2	Two introgression lines with lower total acylsugars	207
5.3.1.3	ILs 8-1 and 8-1-1 causes a shift in acyl chain lengths without altering numbers of substitutions	210
5.3.1.4	Discovery of differences in accumulation of metabolites of lower abundance ILs 1-1 and 1-1-3 with novel glycoalkaloids	217
5.3.1.5	ILs 6-3 and 6-4 with higher ratio of tri- to dimethylmyricetin	223
5.3.2	F2 plants are screened to determine whether phenotypes dominant or recessive for recurrent parent M82	225
5.4.	Conclusions	228
Chapter 6	Qualitative and quantitative profiling of acylsugar metabolic phenotypes for tomato breeding lines using liquid chromatography/time-of-flight mass spectrometry	230
6.1	Introduction	231
6.2	Experimental Section	233
6.2.2	Chemicals	233
6.2.3	Preparation of internal standard (3-decanoyl-glucofuranose)	233
6.2.4	Extraction Method	234
6.2.5	High-Performance Liquid Chromatography	234
6.2.6	Mass Spectrometry	235
6.2.7	Gas Chromatography/Mass Spectrometry	235
6.2.8	Chemometric data analysis	236
6.2.9	Extraction and isolation of S 4:17	237
6.3	Results and Discussion	237
6.3.1	Annotation of acylsugar from conventional tomato <i>Solanum lycopersicum</i> M82	237
6.3.2	Compare total amounts of acylsugars and distributed amounts of acylsugars using LC/MS	246
6.3.3	Fatty acyl substituents of acylsugars among wild type <i>S. pennellii</i> LA0716, LA 1522, <i>S. habrochaites</i> LA1777 and cultivated tomato M82 using GC/MS	249
6.3.4	Compare total amounts of acylsugars and distributed amounts of glucose triesters, sucrose triesters and sucrose tetraesters, and fatty acyl substituents of acylsugars among Cornell breeding lines using LC/MS	251
6.4	Conclusions	271
Chapter 7	Concluding Remarks	273
Appendix		278

Bibliography	284
---------------------------	------------

Tal

LIST OF TABLES

- Table 1-1.** MarkerLynx software first detects peaks using specific mass data, aligns the peak according to retention time, integrates peaks over all samples, and assembles the results into a table.....13
- Table. 3-1.** Fragments and fatty acid constituents for detected glucose triesters from *S. pennellii* LA0716 in negative and positive mode (CID potential, 25 V).....88
- Table 3-2.** Exact mass measurements of [M+HCOO]⁻ ions of detected glucose triesters from *S. pennellii* LA0716 detected using multiplexed CID with collision potential (10 V).....89
- Table. 3-3.** Fragment ions and fatty acid constituents of detected glucose triesters in an extract of *S. pennellii* LA1522 leaf trichomes using negative and positive mode electrospray ionization (CID potential, 25 V)94
- Table.3-4.** Fragments and fatty acid constituents for detected sucrose triesters in *S. pennellii* LA1522 using ESI negative (CID potential, 55 V) and positive mode (CID potential, 40 V) 108
- Table 3-5.** Exact mass measurement of [M+HCOO]⁻ ions for detectable sucrose triesters in accession LA1522 detected using ESI negative mode under CID potential 10 V.....109
- Table 3-6.** Fragments and fatty acid constituents of detected sucrose tetraesters from *S. habrochaites* LA1777 using ESI positive mode (CID potential, 40 V) and negative (CID potential, 55 V)124
- Table 3-7.** Exact mass measurement of [M+HCOO]⁻ ions for detected sucrose tetraesters in *S. habrochaites* LA1777 detected using ESI negative mode under CID potential 10 V.....125
- Table 3-8.** Fragments and fatty acid constituents of detected acylsugars from *S. habrochaites* LA1353 using ESI positive mode (CID potential, 40 V) and negative (CID potential, 55 V).....129
- Table 5-1.** The detectable acylsugars for ILs were listed with retention time (min) and molecular mass for [M+HCOO]⁻208
- Table 6-1.** Structure of S 4:17 in *Solanum lycopersicum* M82 characterized by NMR (¹H, ¹³C, DEPT, gHMQC, gHMBC, gCOSY, and TOCSY). Chemical shifts of

Ta

Ta

Ta

Ta

Ta

Ta

Ta

proton and ¹³ C resonances are listed with proton-proton coupling constants.....	242
Table 6-2. Acyl sucrose metabolites detected in trichome extracts from <i>S. lycopersicum</i> M82 and their adduct ions and fragment ions observed using ESI negative (CID potential 55 V) and positive mode (40 V)	246
Table 6-3. Fatty acyl constituents for detected acylsugars in 8-1×0716 using ESI negative mode with CID potential at 55 V.....	259
Table 6-4. Fatty acyl constituents for detected acylsugars in 8-2×0716, 3-2×0716 and 7-5×0716 using ESI negative mode with CID potential at 55 V.....	260
Table 6-5. Fatty acid constituents of detected sucrose triesters from fixed lines using ESI negative (CID potential, 55 V)	269
Table 6-6. Fatty acid constituents of detected acylsugars from crosses between fixed lines and <i>S. pennellii</i> LA0716 sing ESI negative (CID potential, 55 V) ...	270
Table.A-1 Description of breeding lines from Cornell including introgression lines, fixed lines, crosses between introgression lines and <i>S. pennellii</i> LA0716, and new fixed lines.....	279
Table.A-2 Description of breeding lines from Cornell including <i>S. pennellii</i> LA0716, normal tomato lines, fixed lines and crosses between fixed lines and <i>S. pennellii</i> 0716.....	280

Fi

LIST OF FIGURES

- Figure 2-1** Schematic of LC/multiplexed CID method showing (A) application of multiple collision potentials to the ion transit lens to effect nonselective ion fragmentation, (B) mass spectra acquired using five different quasi-simultaneous CID conditions, showing molecular (★), noncovalent oligomer (○), and fragment ion species (□) using color-coded symbols to distinguish ions coeluting metabolites, and (C) breakdown curves showing common behavior for molecular ion species for two coeluting metabolites, a noncovalent dimer ion, and a fragment ion30
- Figure 2-2** Negative mode electrospray ionization mass spectra obtained using multiplexed collision induced dissociation using five aperture 1 voltages (from bottom to top: 10, 25, 40, 55 and 80 V) for two metabolites extracted from *S. pennellii* LA0716 by leaf dip.....33
- Figure 2-3** TIC (total ion chromatograph) for *S. habrochaites* LA1777 (top) and *S. pennellii* LA0716 (bottom) using ESI negative with CID potential at 10 V.....35
- Figure 2-4** Breakdown curves showing energy dependence of various ion abundances for metabolites extracted from leaf dips of *S. pennellii* LA0716 and *S. habrochaites* LA1777.....36
- Figure 2-5** A: Breakdown curves derived from metabolites detected during LC/negative ion ESI MS analyses of a leaf dip extract from *S. habrochaites* LA1777.....39
- Figure 2-6** Breakdown curves derived from LC/negative ion ESI MS analyses of a leaf dip extract from *S. pennellii* LA0716.....40
- Figure 2-7** Breakdown curves for the flavonoid glycoside rutin obtained using negative mode electrospray ionization following injections of 10 µl of solutions at five different concentrations.....42
- Figure 2-8** Multiplexed CID mass spectra (electrospray, negative mode) of coeluting metabolites from LC/TOF MS analysis of an extract of *S. pennellii* LA0716, showing spectra obtained from three of the five Aperture 1 potentials.....45
- Figure 2-9** Total ion chromatography (TIC) for IL1-1 (Bottom) and XIC of *m/z* 609, 1076 (elute at 2.46 min), 1094 and 593 (elute at 2.60min) under ESI negative mode.....49
- Figure 2-10** Averaged mass spectra across the LC/MS TIC peak eluting at 2.46 min using three CID voltages (10, 55 and 80 V)50

Figure 2-11 XICs for the ions labeled with \circ and \diamond from Figure 2-10 were generated for LC/MS analysis of an extract of IL 1-1	51
Figure 2-12 Negative mode mass spectra from an extract of IL-1-1 trichomes. Spectra were generated for three CID voltages (10, 55, and 80 V) by averaging spectra across the TIC peak eluting at 2.60 min	54
Figure 2-13 Breakdown curves showing energy dependence of various ion abundances for glycoalkaloid metabolites eluting at 2.46, 2.60 and 2.70min extracted from leaf dips of transgenic plant. Ion abundances are normalized to the total ion current for each individual ion as summed for all five collision potentials	55
Figure 2-14 Breakdown curves showing energy dependence of various ion abundances for flavonoid metabolites eluting at 2.46 and 2.60min extracted from leaf dips of transgenic plant	57
Figure 2-15 Concentration dependence of extracted ion chromatogram (XIC) peak areas for ions derived from rutin at different collision potentials using 10 μ l injections of rutin standard solution and negative mode electrospray ionization	60
Figure 2-16 (Top) TIC of <i>S. habrochaites</i> LA1777 and XIC of <i>m/z</i> 233 under highest CID voltage in ESI negative mode. (Bottom) Breakdown curves derived from LC/negative ion ESI MS analyses of a leaf dip extract from <i>S. habrochaites</i> LA1777 corresponding to three isomers A (\diamond), B (\blacksquare) and C (\blacktriangle) of deprotonated sesquiterpene acid ($[M-H]^-$ ions at <i>m/z</i> 233) showing isomer-dependence of signal upon collision potential	62
Figure 3-1 Generalized structure of triacylglucose metabolites from <i>S. pennellii</i> LA0716 as identified by Burke <i>et al.</i>	72
Figure 3-2 Extracted ion chromatograms (XICs) generated from an extract of leaf tissue from <i>S. pennellii</i> LA0716 of ions of <i>m/z</i> values corresponding to $[M+HCOO]^-$ for glucose triesters with total fatty acid carbon atoms ranging from 12 to 22	74
Figure 3-3 Proposed fragmentation pathway of the formate adduct of triacylglucose S3:18 extracted from <i>S. pennellii</i> LA0716 (RT = 21.11 min) using ESI negative mode with CID voltage 25 V	78
Figure 3-4 Electrospray ionization mass spectra of G 3:18 (RT = 27.11 min) extracted from <i>S. pennellii</i> LA0716 using an Aperture 1 CID potential of 25 V in negative (top) and 40 V in positive (bottom) mode	81

- Figure 3-5** Product ion MS/MS spectrum of m/z 519 ($[M+HCOO]^-$) from triacylglucose G 3:18 from *S. pennellii* LA0716 using collision energy 20 V.....82
- Figure 3-6** Top: Product ion MS/MS spectrum of m/z 492 ($[M+NH_4]^+$) from triacylglucose G 3:18 from *S. pennellii* LA0716. Bottom: product ion MS/MS spectrum for m/z 457 ($[M+H-H_2O]^+$) for G 3:18; m/z 457 was generated in source.....83
- Figure 3-7** (Top) Extracted ion chromatogram of m/z 547 (G 3:20) from LC/MS analysis of an extract of *S. pennellii* LA0716 under CID potential 10 V in ESI negative mode. Chromatographic peaks corresponding to four isomers detected as m/z 547 were resolved. (Bottom) ESI mass spectra obtained from these four isomers at 25 V in negative mode (1-4)85
- Figure 3-8** Fatty acyl groups from total acylsugars in an extract of *S. pennellii* LA0716 were transesterified to form fatty acid ethyl esters, and analyzed using GC/MS. Data for each fatty acid were calculated as a percentage of the total ion current chromatogram peak areas of fatty acids detected through GC/MS. The abbreviations “i” and “ai” refer to iso- and anteiso- branched isomers.....90
- Figure 3-9** Extracted ion chromatograms (XICs) generated from an extract of leaf tissue from *S. pennellii* LA1522 for ions corresponding to $[M+HCOO]^-$ for glucose triesters with total fatty acid carbon atoms ranging from 12 to 22 ·92
- Figure 3-10** Extracted ion chromatograms (XICs) generated from an extract of leaf tissue from *S. pennellii* LA1522 of ions of m/z values corresponding to $[M+HCOO]^-$ for sucrose triesters with total fatty acid carbon atoms ranging from 12 to 23.....96
- Figure 3-11** Breakdown curves derived from LC/negative ion ESI MS analyses of a leaf dip extract from *S. pennellii* LA1522 showing collision energy dependence of abundances of: (A) formate ion adducts of three triacylsucrose metabolites ($[M+HCOO]^-$ at m/z 737 S3:22 (▲); $[M-H]^-$ at m/z 691 (◇); C12 fatty acid at m/z 199 (■); C5 fatty acid (□).....99
- Figure 3-12** (TOP) ESI spectrum of triacylsucrose S3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using CID potential of 55 V. (Bottom) Product ion MS/MS spectrum for formate adduct of S3:22 (products of m/z 737.39 using collision potential of 30 V) 100
- Figure 3-13** Positive mode ESI spectrum of triacylsucrose S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using CID potential of 55 V.....103

- Figure 3-14** (Top) Product ion MS/MS spectra of (Top) $[M+NH_4]^+$ at m/z 710.43 and (Bottom) $[M+Na]^+$ at m/z 715.39 for S 3:22 extracted from *S. pennellii* LA1522.....103
- Figure 3-15** Proposed fragment pathway of formate adduct of S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using ESI negative mode with CID voltage 55 V. Specific substitution positions of individual fatty acids remain uncertain..... 104
- Figure 3-16** Proposed fragment pathway for sodium adduct of S 3:21 extracted from *S. pennellii* LA1522 based on MS/MS product ion spectrum. Specific substitution positions of individual fatty acids remain uncertain.....106
- Figure 3-17** Proposed fragment pathway of ammonium adduct for S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using ESI positive mode with CID voltage 40 V. Specific substitution positions of individual fatty acids remain uncertain..... 107
- Figure 3-18** Fatty acyl groups from total acylsugars in an extract of *S. pennellii* LA1522 were transesterified to form fatty acid ethyl esters, and analyzed using GC/MS. Data for each fatty acid were calculated as a percentage of the total ion current chromatogram peak areas of fatty acids detected through GC/MS.....111
- Figure 3-19** Multiplexed CID mass spectra (electrospray, negative mode) of coeluting metabolites from LC/TOF MS analysis of an extract of *S. pennellii* LA1522 showing spectra obtained from three of the five Aperture 1 potentials.....113
- Figure 3-20** (A) Extracted ion chromatograms (XICs) from negative ion mode LC/MS analyses of (A) *S. pennellii* LA1522 and (B) *S. habrochaites* LA1777. Selected m/z values correspond to $[M+HCOO]^-$ of sucrose triesters (highlighted with blue dashed lines) and sucrose tetraesters (red dashed line).....115
- Figure 3-21** (Top) Negative mode ESI spectrum of tetraacylsucrose S 4:21 (RT = 31.50 min) under CID potential 55 V and (Bottom) product ion MS/MS spectrum of formate adduct (m/z 737.35) for S 4:21 from *S. habrochaites* LA1777..... 117
- Figure 3-22** Positive mode ESI spectrum of tetraacylsucrose S 4:21 (RT = 31.50 min) from *S. habrochaites* LA1777 using CID potential 40 V..... 119
- Figure 3-23** (Top) Product ion MS/MS spectrum of $[M+NH_4]^+$ at m/z 710.40 and (Bottom) product ion MS/MS spectrum for m/z 317 ($[FRU+acyl]^+$) of S 4:21

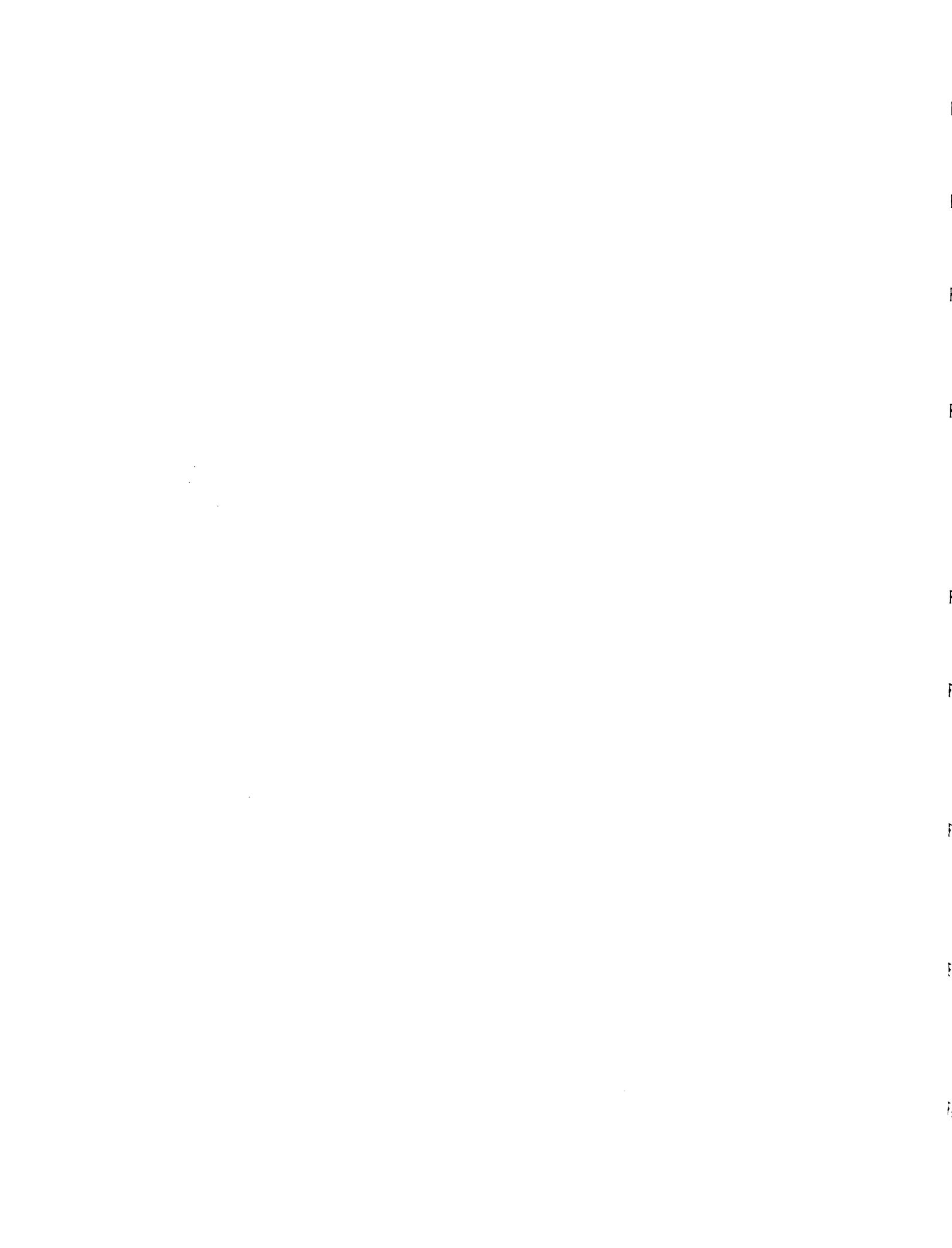
from *S. habrochaites* LA1777; m/z 317 was generated in source from tetraesters S4:21 from *S. pennellii* LA1777.....120

- Figure 3-24** Collision induced dissociation of $[M+NH_4]^+$ acylsugar ions yield a dominant acylfructofuranose fragment ion, and a less abundant acylglucopyranose fragment ion.....121
- Figure 3-25** Proposed fragment pathway upon CID of the formate adduct for S 4:21 (RT = 31.5 min) extracted from *S. habrochaites* LA1777 using ESI negative mode with CID voltage 55 V.....122
- Figure 3-26** Proposed fragment pathway of ammonium adduct for S 4:21 (RT = 31.5 min) extracted from *S. habrochaites* LA1777 using ESI positive mode with CID voltage 55 V.....123
- Figure 3-27** Fatty acyl groups from total acylsugars in an extract of *S. habrochaites* LA1777 were transesterified to form fatty acid ethyl esters, and analyzed using GC/MS. Levels for each fatty acid were calculated as a percentage of the total ion current chromatogram peak areas of fatty acids detected through GC/MS.....127
- Figure 3-28** (Top) Negative mode ESI spectra of S 5:25 (RT = 31.91 min) and (Bottom) S 4:26 (RT = 36.04 min) from *S. habrochaites* LA1353 using CID potential 55 V.....130
- Figure 3-29** Putative structures for detected acylsugars among *S. pennellii* LA0716 and LA1522 and *S. habrochaites* LA1777 and LA1353. Specific substitution positions of individual fatty acids remain uncertain for some metabolites.....132
- Figure 4-1** Generalized structure of flavonoid metabolites known in tomato and its relatives, where R groups can be hydrogen, hydroxyl, or other substituted oxygen-containing groups.....138
- Figure 4-2** Generalized structure of glycoalkaloid metabolites known in tomato and its relatives, where R groups are glycosides.....139
- Figure 4-3** Putative structures for several flavonoid derivatives from *Solanum* wild species.....145
- Figure 4-4** ESI mass spectra for quercetin-3-rutinoside from *S. pennellii* LA0716 using positive mode with CID 40 V.....146
- Figure 4-5** (Top) XIC for m/z 609 from *S. pennellii* LA0716 using ESI negative mode with CID voltage 10 V. (Bottom) ESI spectrum of quercetin-3-rutinoside (RT = 11.52 min) using ESI negative mode with CID voltage 55 V.....147

Figure 4-6 Proposed fragmentation pathway for quercetin-3-rutinoside from <i>S. pennellii</i> LA0716 using ESI negative mode with CID potential 55 V.....	148
Figure 4-7 (Top) Two isomeric metabolites appear in the extracted ion chromatogram <i>m/z</i> 609 for an extract of <i>S. habrochaites</i> LA1777 using CID voltage 10 V. (Middle) ESI spectrum of first peak at 11.04 min for kaempferol-3-glucosylglucoside at CID voltage of 55 V. (Bottom) ESI spectrum of second peak (RT = 11.45 min) annotated as quercetin-3-rutinoside using ESI negative mode with CID voltage 55 V.....	150
Figure 4-8 Putative structures of four anthocyanins petunidin-3-O-(p-coumaroyl)-rutinoside-5-O-glucoside (<i>m/z</i> 933), malvidin-3-O-(p-coumaroyl)-rutinoside-5-O-glucoside(<i>m/z</i> 947), delphinidin-3-O-(p-coumaroyl)-rutinoside-5-O-glucoside (<i>m/z</i> 919) and petunidin-3-(caffeoyl)-rutinoside-5-O-glucoside (<i>m/z</i> 949)	152
Figure 4-9 ESI mass spectra from (top) anthocyanins at <i>m/z</i> 933 (RT = 18.69 min) and (bottom) <i>m/z</i> 947 (RT = 20.23 min) from <i>L. lycopersicum</i> M82 using positive mode with CID 40 V.....	154
Figure 4-10 Important steps in the biosynthesis of anthocyanins and flavonoids (adapted from B. Winkel-Shirley and K. Saito ^{30, 31})	155
Figure 4-11 LC/MS XICs for a trichome extract from <i>S. habrochaites</i> LA1777 using ESI positive mode with CID voltage 10 V showing: <i>m/z</i> 347 (dimethylmyricetin, 2M), <i>m/z</i> 361 (trimethylmyricetin, 3M), <i>m/z</i> 375 (tetramethylmyricetin, 4M) and <i>m/z</i> 389 (pentamethylmyricetin, 5M)	157
Figure 4-12 Putative structures of major methylated flavonoids in <i>Solanum</i> species LA1777. The exact positions of methyl groups on these flavonoids retain unknown.....	159
Figure 4-13 Nomenclature of fragment ions generated from flavonoids.....	160
Figure 4-14 (Top) Positive mode mass spectrum of trimethylmyricetin from <i>S. habrochaites</i> LA1777 using CID potential 80 V. (Bottom) Product ion MS/MS spectrum of [M+H] ⁺ , <i>m/z</i> 361 trimethylmyricetin (positive mode) using collision energy of 35 V.....	160
Figure 4-15 Proposed formation of fragment ions of trimethylmyricetin in ESI positive mode, obtained from extract of <i>S. habrochaites</i> LA1777.....	161
Figure 4-16 (Top) Positive mode mass spectrum of tetramethylmyricetin extracted from <i>S. habrochaites</i> LA1777 with CID potential 80 V. (Bottom) Positive mode product ion MS/MS spectrum of tetramethylmyricetin at <i>m/z</i> 361 using collision energy 40 V.....	162

Figure 4-17 Chemical structure of tomatine	163
Figure 4-18 (Top) LC/MS XICs of $[M+H]^+$ for α -tomatine (RT = 14.78 min) and (Bottom) dehydrotomatine (RT = 14.29 min) from a leaf dip extract of tomato (<i>S. lycopersicum</i> M82)	167
Figure 4-19 Mass spectra of α -tomatine (Top) and dehydrotomatine (Bottom) from LC/MS analysis of an extract from <i>S. lycopersicum</i> M82 using ESI positive mode with CID potential 80 V	168
Figure 4-20 A proposed fragmentation pathway for α -tomatine using ESI positive mode based on literature reports	169
Figure 4-21 LC/MS XICs of an extract of <i>S. pennellii</i> LA0716 using ESI in positive ion mode corresponding to protonated molecules of (Top) α -tomatine and (Bottom) dehydrotomatine	172
Figure 4-22 (Top) LC/MS XIC of m/z 1032.5 from IL1-1 using ESI positive mode and CID voltage 10 V. ESI spectrum (Middle) of first peak (RT = 2.73 min) corresponding to dehydrotomatine isomer and ESI spectrum (Bottom) of second peak (RT = 3.00 min) corresponding to dehydrotomatine (double bond in 5,6-position) using CID 80 V	173
Figure 4-23 A proposed fragmentation pathway for first eluting dehydrotomatine isomer peak (RT = 2.73 min) from IL 1-1 using ESI positive mode with CID voltage 80 V. Fragment ions at m/z 273 and 255 indicate that the position of the double bond is not on rings A-D	174
Figure 4-24 LC/MS XIC of m/z 1050 corresponding to a putative hydroxytomatine metabolite from IL1-1 using ESI positive mode. (Bottom) ESI spectrum for hydroxytomatine from IL1-1 using ESI positive mode with CID potential of 80 V	175
Figure 4-25 A proposed fragmentation pathway for hydroxytomatine from IL 1-1 using ESI positive mode	176
Figure 4-26 LC/MS XIC of sesquiterpene acid at m/z 233 from <i>S. habrochaites</i> using ESI negative mode	177
Figure 4-27 Product ion MS/MS spectrum of sesquiterpene acid at m/z 233 from <i>S. habrochaites</i> LA1777 using ESI negative mode with collision energy 30 V	178
Figure 4-28 A proposed fragmentation pathway for sesterpene acid at m/z 233 from <i>S. habrochaites</i> using ESI negative mode	179

- Figure 4-29** (Top) LC/MS XICs of diterpene acid at m/z 319 and (Bottom) hydroxysesquiterpene acid at m/z 249 from *S. habrochaites* LA1777 using ESI negative mode.....180
- Figure 4-30** (Top) LC/MS XICs of m/z 649 $[M-H]^-$ of sesterterpene metabolite using ESI negative mode. (Middle) Possible Elemental formulas based on accurate mass measurement. (Bottom) ESI spectrum of m/z 649 (RT = 16.83 min) from *S. habrochaites* LA1777 using ESI negative mode with CID potential 55 V.....183
- Figure 4-31** Constant neutral loss (86 Da) spectrum from an extract *S. habrochaites* LA1777 using flow injection analysis and ESI negative mode.....184
- Figure 5-1** *S. pennellii* introgression lines. (adapted from http://zamir.sgn.cornell.edu/Qtl/il_story.htm)188
- Figure 5-2** LC/MS total ion chromatogram (TIC) obtained from leaf extraction of introgression lines (ILs) 8-1-1 with 5 min gradient using Ascentis Express C18 fused core column, 2.1 × 50 mm; 2.7 μm (Top) compared to 43 min gradient with Thermo BetaBasic C18 column, 1 × 150 mm, 3 μm (Bottom) in ESI negative mode.....194
- Figure 5-3** Extracted ion chromatograms (XICs) of $[M+HCOO]^-$ for leaf dip extract of IL 8-1-1 showing peaks corresponding to sucrose triesters or tetraesters from m/z 639 to m/z 779 (homologs differing by 14 Da in molecular mass owing to additional methylene group in the acid moieties).....196
- Figure 5-4.** Markerlynx data for part of introgression line samples. The software first generates extracted ion chromatograms (XICs) for each mass, then detects and integrates the peak intensity, and aligns the peaks according to retention time. The metabolite ID was assigned at the first column for each peak..199
- Figure 5-5** Principal component analysis (PCA) on 66 ILS plus recurrent parent M82 (total 276 samples) obtained using pareto scaling with mean centering. PCA score plot; Symbols: ILS 1-3, 1-4 (●); 5-3, 11-3 (◆); 8-1, 8-1-1 (▲); Squares (■) represent recurrent parent M82 (n=31); Cross (*) represent other 60 ILS (n=3 or 4 for each IL)201
- Figure 5-6** (Left) PCA scores plot from LC/TOF MS metabolite data represent three clusters of ILS separated from recurrent parent M82 using Pareto scaling with mean centering. The clusters are 1-3, 1-4 (●); 5-3, 11-3 (◆); 8-1, 8-1-1 (▲) and M82 (■). (Right) PCA loading plot showed corresponding analytical signals (retention time-mass pairs) contributing to the difference among clusters in PCA scores plot.....202



- Figure 5-7** LC/MS total ion LC/MS chromatograms obtained from leaf dip of M82 (Top) and IL 1-3 (Bottom) using ESI negative mode. The peaks are labeled after identified using multiplexed CID MS method.....206
- Figure 5-8** (Top) Mass spectrum of peak at RT = 3.35 min in IL 1-3 and (Bottom) Mass spectrum of peak at RT = 3.42 min from *S. lycopersicum* M82 with CID potential 55 V.....206
- Figure 5-9** Total amount of acylsugars for 66 ILs and recurrent parent M82 based on Quanlynx software in Masslynx 4.1 (waters). Sum of all the peak areas of acylsugars were normalized to internal standard (Propyl-4-hydroxybenzoate) and dry leaf dip.....209
- Figure 5-10** A. Total ion chromatograms from LC-TOF MS analysis of acylsugars in leaf dips of M82 and IL8-1-1. Labeling nomenclature for acylsugars - S 3:22 (5,5,12) is an acylsucrose with 3 acyl chains having a total of 22 carbons and numbers in parentheses indicate the lengths of the individual acyl chains. B. Amounts of acylsugars showing differences in abundance between M82 and IL8-1-1 are shown as integrated peak areas normalized to the internal standard and the dry weight of the extracted leaflet.....212
- Figure 5-11** ESI negative LC/MS XICs of C4 *m/z* 87.04 and C5 *m/z* 101.06 fatty acyl groups cleaved from acylsugars with CID potential 55 V for A: *S. lycopersicum* M82 and B: *S. pennellii* LA0716. C: ILs 8-1.....214
- Figure 5-12** Side chains on acylsugars from leaf dip samples were transesterified to the corresponding ethyl esters and analyzed by GC/MS as described in chapter 6. The peak area % values for the base peak XIC for each fatty acid in M82 and IL8-1-1 are shown along with the standard error with *n* = 2 for IL 8-1-1 and *n* = 11 for M82.....217
- Figure 5-13** A. Extracted ion chromatograms of M82 and IL1-1 for *m/z* 1076.5 show an earlier eluting peak found only in the IL and not in M82. B. Schematic representation of chromosome 1 introgressions showing locus controlling the glycoalkaloid phenotype is located on bin 1-A or 1-B (IL1-1-2 was not analyzed in this study). C. Structures for dehydrotomatine isomers.....220
- Figure 5-14** OPLS-DA loading plot showed interclass difference between IL 1-1 (Right) and recurrent parent M82 (Left) using fused core column with 5 min gradient. The ions were labeled with retention time and *m/z*. The upper right corner and the lower left corner are corresponding to IL 1-1 (*n* = 4) and M82 (*n* = 4), respectively.....222
- Figure 5-15** Total amount of mono-, di-, trimethylmyricetin for 66 ILs and recurrent parent M82 were obtained based on Quanlynx software in Masslynx 4.1 (waters). The peak area of mono-, di-, and trimethylmyricetin is normalized

to internal standard (propyl-4-hydroxybenzoate) and dry leaf dip individually for all the samples.....	224
Figure 5-16 Total amount of sucrose triesters and tetraesters were obtained based on Quanlynx software in Masslynx 4.1 (Waters) for 18 F2 plants of 1-3 and 22 F2 plants of 1-4 and 10 control plants M82.....	226
Figure 6-1 Structure of synthesized internal standard 3-decanoylglucofuranose.....	234
Figure 6-2 XICs of formate adducts from S 4:15 to S 4:24 (highlighted with red dash line) and S 3:15, S 3:20 to S 3:22 (highlighted with blue dash line) from <i>S. lycopersicum</i> M82.....	239
Figure 6-3 Electrospray ionization mass spectra of S 4:17 (RT = 3.11 min) extracted from <i>S. lycopersicum</i> M82 using CID potential of 55 V (top) in negative and 40 V (bottom) in positive mode.....	245
Figure 6-4 Distributed amounts of glucose triesters, sucrose triesters, and sucrose tetraesters of <i>S. pennellii</i> LA0716, LA1522, <i>S. habrochaites</i> LA1777 and <i>S. lycopersicum</i> M82 based on sum of total peak areas of formate adducts for detected acylsugars normalized to internal standard and dry leaf weight for each acylsugar.....	248
Figure 6-5 Distribution of fatty acyl groups for total acylsugars among <i>S. pennellii</i> 0716, 1522, <i>S. habrochaites</i> 1777 and <i>S. lycopersicum</i> M82 obtained from GC/MS.....	251
Figure 6-6 Total amounts of acylsugars in Cornell tomato breeding lines. Reported levels represent the sum of XIC peak areas of acylsugar formate adducts normalized to internal standard and dry leaf weight.....	254
Figure 6-7. Distribution of glucose triesters, sucrose triesters and sucrose tetraesters among all the breeding lines.....	255
Figure 6-8 XICs of deprotonated ion $[M-H]^-$ for C4 and C5 fatty acids using CID potential 55 V, corresponding to m/z 87 and 101, from crosslines 8-1×0716 and 8-2×0716.....	258
Figure 6-9 Amounts of acylsugars in <i>S. pennellii</i> , normal tomato lines, fixed-lines and crosslines based on sum of peak areas, then normalized to internal standard and dry leaf weight (Top). Composition is broken down into glucose triesters, sucrose triesters and sucrose tetraesters were among all the breeding lines (Bottom)	263
Figure 6-10 TIC for LA0716 (top), 071026×LA0716 (middle) and 071026 (bottom) using ESI negative mode with CID potential at 10 V.....	266

Figure 6-11 ESI negative XICs of C9 *m/z* 157.13 and C10 *m/z* 185.16 fatty acyl groups using CID potential 55 V for fixed line 071026 and crosses between 071026 and *S. pennellii* LA0716.....267

Figure 6-12 Distribution of amounts of individual fatty acyl groups among fixed-lines, crosslines and *S. pennellii* LA0716 based on GC/MS profiles. 077044-6, 077048, 077049-6, 077051-6, 077058-6 are fixed lines; 077082-6 to 077089-6 are crosslines.....268

LIST OF ABBREVIATIONS

APCI	atmospheric pressure chemical ionization
CE-MS	capillary electrophoresis-mass spectrometry
CID	collision induced dissociation
DDA	data dependent analysis
DFI-MS	direct flow injection-mass spectrometry
ESI	electrospray ionization
FID	flame ionization detection
FRU	fructofuranose
FT-ICR-MS	Fourier-transformation ion cyclotron resonance mass spectrometer
GC-MS	gas chromatography-mass spectrometry
GLU	glucopyranose
GST_s	glandular secreting trichomes
HPLC	high-performance liquid chromatography
IL	introgression line
MS	mass spectrometry
OPLS-DA	orthogonal projection to latent structures-discriminant analysis
PCA	principal component analysis
QTL	quantitative trait locus
RT	retention time
SPE	solid phase extraction
TIC	total ion chromatogram

TOF

time-of-flight

XIC

extracted ion chromatogram

UPLC

ultra-performance liquid chromatography

UV-vis

ultraviolet-visible spectrometry

Chapter 1. Introduction

The rapid growing world population will have an astounding impact on food and energy resources.¹ In order to meet the needs for quick expanding human population, large amounts of energy are devoted to irrigation and production of fertilizers and pesticides, to maintain agricultural productivity. The world needs more productive agriculture to feed a growing population and provide renewable feedstocks for chemical production and energy. One way to achieve these goals is through metabolic engineering of plants and microbes. Genetic modifications of plants offers improvements in production of diverse natural products (secondary metabolites), that can protect plants against insect attack, pathogen stress and animal herbivores and increase agricultural productivity.^{2,3} Many plant metabolites also have important value as pharmaceuticals, fragrances and food additives.⁴ In addition, engineering of plant biomass through genetic modification is an area of active research that aims to develop renewable sources of energy. However, many genes involved in regulation of important metabolites remain largely unknown, and those involved in biosynthesis of key metabolites are not completely understood. Metabolite levels can be viewed as the responses of biological systems to environmental or genetic manipulation,⁵ thus, identification and quantification chemical metabolites among transgenic plants can aid assignment of functions to genes.⁶

Glandular trichomes were chosen as a targeted model in the Solanum Trichome Project to investigate gene functions important in regulation of specialized natural products. Trichomes are highly specialized epidermal cells common to many plants and are involved in the synthesis, storage, and secretion of secondary metabolites.⁷ They are easily removed from plant tissue surfaces since they protrude from the epidermis.⁸ Therefore, RNA, proteins and small molecules that they contain are easily sampled.⁸

Information about DNA sequences, metabolite identities and levels, and gene expression obtained from trichomes have great utility for identifying gene and enzyme functions⁹ when specialized type of trichomes are isolated from other cell types. First, metabolite profiling can be performed for isolated trichomes, and transcripts (mRNA) are extracted for construction of cDNA libraries and EST (Expressed Sequence Tag) database. Gang *et al.*¹⁰ and Iijima *et al.*^{11,12} isolated glandular trichomes from three different cultivars, prepared EST databases from them, and profiled metabolites for mint and basil. Through annotation of correlations between transcript and metabolite levels or from altered metabolite profiles among different cultivars, functions of specific genes that regulate production of methylated phenylpropenes were discovered.⁹ So far this approach has provided important information about secondary metabolism,¹³⁻¹⁵ and has accomplished this more efficiently than early studies that explored gene functions through enzyme assays.¹⁶ Since altered metabolite profiles may be associated to specific genes to support annotations of gene function, this approach is also used in the research described below for trichomes that are prolific in synthesizing specialized metabolites. This dissertation describes metabolite profiling of several wild type species in the genus *Solanum* and hundreds of transgenic plants. My role in this research has focused on development of analytical methods for metabolite profiling and rapid screening of metabolic phenotypes screening to guide discoveries of gene functions important for metabolic engineering.

1.1 The Emergence of Metabolomics in Biological Research

Metabolites are often classified as primary or secondary metabolites.¹⁷ Primary metabolites are important to plant growth, development¹⁸ whereas secondary metabolites

are often connected cell signaling, interspecies communication and responses to biotic and abiotic stress.¹⁹

Broad assessment of the entire suite of metabolite produced by a biological system is named “metabolomics”.^{5,20-23} The general idea of ‘metabolomics’ was first brought up several years ago in the field of microbiology,²⁴ and its importance in plant science was discovered soon after. The variety of metabolites, as intermediates and end products of biochemical pathways, is regulated by many structural and regulatory genes in addition to environmental influences.²⁵ Plants are sources of numbers of metabolites whose structures, functions and utility have been largely unknown. At this stage, we have only a vague idea of how large the complement of metabolites within a particular species is; estimates from 5000-25000, and are comparable in order of magnitude to the number of genes.²⁶ These numbers considerably exceed estimates of the number of metabolites synthesized by prokaryotic microorganisms (~1500) and animals (~2500).²⁷ Although more than 100,000 plant secondary metabolites have already been identified, in fact these may represent only 10% of the actual total synthesized by the plant kingdom.²⁸

The metabolomics approach strives for an overview of whole-cell metabolic patterns.²⁹ Depending on the number of compounds they analyze, the level of structure information they obtain, and their sensitivity, the analysis of metabolites is divided into target analysis, metabolic profiling, metabolomics, or metabolic fingerprinting.³⁰ Target analysis aims at quantitative analysis of known substrate and/or product of targeted protein.²³ The most common approach, metabolite profiling, a term first brought up in the 1970s to refer to the qualitative and quantitative analysis of complex mixtures of physiological origin.^{31,32} One of the early pioneers in this field, Professor Charles

Sweeley, was one of the first to demonstrate such an approach, using gas chromatography, GC/MS, and computerized data systems and library searching for profiling of metabolites from assorted classes.³³ Much of this work was conducted at Michigan State University. At the other extreme, metabolic fingerprinting detects many compounds without requiring identification of their structures.³⁴ This technology focuses on pattern recognition of metabolic phenotype to produce a broad picture,³⁰ rather than cataloguing specific metabolites.

In order to find out the functions of metabolites and corresponding enzymes, two basic strategies can be performed, termed forward genetics and reverse genetics.³⁰ In the forward genetic approach, naturally occurring mutants and mutagenized plants with interesting phenotypes can be used to guide identification and characterization of genes correlated with one or more phenotypes. In reverse genetic approaches, analysis begins with a specific cloned gene, a protein sequence, or a specific genetic mutation and seeks to define the possible phenotypes derived from a specific gene.³⁰ No matter which approach is applied, our understanding of phenotype is based on what can be observed about the characteristic behavior of an organism, and this can often be observed at the metabolite level.⁵ Since the metabolites are the end products of gene expression,³⁰ efficient and comprehensive approaches to genomic analysis must include measurements of chemical constituents—the metabolites—of individual cells or tissues of the organism. By coupling metabolite analysis with the information provided by genomics, we create a science of metabolic genomics, which accelerates completely deciphering the complex inner workings of living systems.

1.2 Challenge of Metabolomics

The major challenge in metabolomics lies in the diversity and complexity of the plant metabolome and these factors provide the impetus for pursuing cutting-edge technological developments.³⁵ Metabolome analyses aim to determine all metabolites in a plant extract; however, no single routine technology for metabolomics, such as DNA sequencing for genome analyses or DNA arrays for transcriptome analyses, is available.²⁷ The genome and the transcriptome are constructed from four nucleotides with highly similar chemical properties, thus facilitating high throughput analytical approaches.²³ Unfortunately, metabolites have a much greater variability in the arrangements of atoms and functional groups compared to the linear 4-letter codes for genes or the linear 20-letter codes for proteins.⁵ Most metabolomics methods rely on GC/MS, LC/MS to identify metabolites through comparison against commercially available standards.²³ Structural elucidation of metabolites is a major obstacle since considerable amounts of high quality purified materials were needed to perform multiple dimensional NMR. Although tandem mass spectrometry requires smaller amounts of material, complete structures of unknown metabolites only can be identified at some limited level by it alone. The alternative approach for metabolite identification is to build a library of standard profiles, created with purified metabolites.³⁶ Although the libraries for NMR, IR and Mass Spectrometry have existed for decades, they are still not complete for the nearly 200,000 known natural products.³⁷ In addition, the range of chemical properties of metabolites covers ionic inorganic species, hydrophilic carbohydrates, hydrophobic lipids, and complex natural products.³⁸ No single analytical platform yields efficient

identification and quantification of all metabolites, therefore, multiple technologies are generally employed.³⁹

Dynamic range of quantitative analysis is also a major challenge for metabolomics. The dynamic range of the most powerful analytical techniques lies between 4 and 6 orders of magnitude for individual components.²³ However, the dynamic range of many techniques is often limited by sample matrix or the presence of other compounds that can interfere with measurement of specific analytes. For instance, oligosaccharide profiling by liquid chromatography-mass spectrometry is compromised when amino acids and peptides are present in the sample because these latter substances are preferentially ionized and can suppress ionization of oligosaccharides.²³

Another difference for metabolomics compared to other “omics” is that it is difficult or impossible to establish a direct link between individual metabolites and genes as mRNA and proteins.³⁵ The number of metabolites is far beyond the number of genes in one species involved in the biosynthesis since multiple metabolites were produced from one gene due to several mechanisms.⁴⁰ Therefore, metabolome analysis need to cover the identification and quantification of all intracellular and extracellular metabolites to study gene function, which creates more challenge for metabolomics.

1.3 Techniques for Metabolomics

Advancing analytical technologies offer great opportunities for viewing global changes of metabolites in a system caused by perturbation of genes in order to associate metabolic phenotypes with gene functions of an organism.³⁴ Due to the complexity and diversity of chemical and physical properties of plant metabolites, multiple analytical

approaches must be combined to achieve this goal, including photodiode array detection (UV/Vis), GC, NMR and MS.

Photodiode array detection (UV/Vis) coupled with various separation techniques has found only limited application due to the low extinction coefficients of many metabolites and lack of selectivity in analyte detection.⁴¹ It is typically limited to select classes of metabolites containing strong and distinct chromophores, such as the anthocyanin pigments in plants.^{42,43} Therefore, UV/Vis detection is primarily used for target analysis and metabolic profiling,^{44,45} but not for global metabolome characterization. Gas chromatographic separation coupled to flame ionization detection (FID) has found wider use in metabolite analysis.^{46,47} The chromatographic resolution of GC provides can separate hundreds of metabolites in a single analysis and allows for high throughput analyses, but flame ionization detection provides minimal information regarding the identities of the analytes beyond chromatographic retention times. For that reason, GC is also limited to target analysis and metabolic profiling where the only compounds of interest have been identified and are available as standards.

NMR serves as an alternative approach for determining chemical structures of metabolites and measuring their abundances. Although ¹H NMR provides comprehensive measurements of hydrogen-containing metabolites, it is generally low sensitivity and suffers from overlapping signals.⁴⁸ Recently, NMR has been interfaced with HPLC to yield multidimensional data useful for metabolite structure elucidation.⁴⁹ The primary disadvantage of current HPLC/NMR technology lies in its low duty cycle and elevated expense due to the need for deuterated mobile phase.²³

Recent metabolome studies have expanded to a variety of body fluids and tissues and now employ a range of instruments including FT-ICR-MS, GC-MS, GC-GC-MS, HPLC-

MS, UPLC-MS and capillary electrophoresis-MS.^{50,51} FT-ICR-MS offers the highest mass resolution (100,000-1,000,000) and mass accuracy (0.1-1 mDa), detection limits in the attomole to femtomole range and MSⁿ capability,⁵² and these features make it a powerful tool for metabolite fingerprinting investigation.⁵³ Despite these features, the application of mass spectrometry methods, including FT-ICR-MS, often cannot differentiate isomeric metabolites without resorting to an additional dimension of analysis such as physical separation or isomer-selective fragmentation. In addition, the high **capital** and operating costs of high-performance FT-ICR-MS instrumentation serve as a **barrier** to the adoption of this technology as the analytical technique of choice for **metabolome** analyses.

Both targeted and nontargeted metabolite analyses are often performed using GC/MS, **which** has proven capability for profiling from several hundred to slightly more than a **thousand** components.^{24,54-56} GC/MS can contribute high separation efficiencies with up **to 250,000** theoretical plates.²³ Furthermore, GC/MS offers the advantage of having **extensive** libraries of 70 eV electron ionization mass spectra that aid metabolite identification. However, one drawback of using GC/MS derives from the requirement that metabolites must be volatile and not decompose at elevated temperatures used for chromatographic separation. Since many metabolites are ionic or polar compounds of low volatility, chemical derivatization must be performed to convert them to more volatile forms. Because many metabolites must be derivatized, extracts must often be evaporated to dryness before derivatization. Numerous metabolites, including secondary metabolites such as multiply glycosylated conjugates and betaines with permanently charged groups, still cannot be detected using GC/MS even after derivatization owing to their decomposition during heating and vaporation. In addition, identification of unknown



metabolites following derivatization and GC/MS analysis has frequently failed because libraries of metabolite mass spectra are incomplete. This is illustrated in one of the more successful metabolomic studies to date. In a collaboration between the laboratories of Oliver Fiehn and Michigan State University's Michael Thomashow, the effects of cold stress on the metabolome of the model plant *Arabidopsis* were explored using derivatization and GC/MS analysis.⁵⁷ In this study, only 26% (30 of 114) of metabolites exhibiting greater than 5-fold change after cold treatment were identified. Such results **highlight** how the incomplete success of metabolite identification in metabolomic **investigations** results from inadequacies in existing libraries of metabolite mass spectra **and** the challenges of interpreting spectra. Therefore, alternative analytical techniques are **needed** to identify and quantify a broader range of specialized secondary metabolites.

The combination of liquid chromatography with mass spectrometry (LC/MS) has **emerged** as a rapid, efficient and powerful approach to complement GC/MS for analysis **of a** broad range of metabolites including polar and semi-polar compounds.⁵⁸⁻⁶⁰ LC/MS **removes** the need for sample derivatization and therefore samples can be analyzed with minimal sample processing or even directly from the tissue source.^{61,62} Most semi-polar metabolites (phenolic acids, flavonoids, alkaloids and other glycosylated species) can be separated using reverse-phase columns. Polar compounds can be measured⁵⁸ (sugars, amino sugars, amino acids, vitamins, carboxylic acids and nucleotides) using hydrophilic and ion-exchange separations. Recent developments termed ultra performance liquid chromatography (UPLC), provide higher separation efficiency and resolution through use of sub-2 μm particles.⁶³⁻⁶⁶ Short UPLC provide separations of efficiency equal to longer conventional HPLC columns, and this allows for analysis times to be shortened relative to conventional methods. UPLC separations are beginning to find application in

nontargeted metabolite analysis^{63,67,68} or targeted profiling,^{66,69,70} in view of their performance in terms of speed, robustness, and high sample throughput.

Technologies that couple liquid separations to mass spectrometry have also improved in recent years, following the development of the soft atmospheric pressure ionization methods atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI). APCI is a less "soft" ionization technique than ESI,⁷¹ i.e. it generates more fragment ions relative to pseudomolecular ions, and is more frequently applied for ionization of less polar and less thermally labile compounds. ESI has provided a major breakthrough for the analysis of biological samples by mass spectrometry because this technique is applied directly to the analytes in solution,^{72,73} requires no derivatization, provides low detection limits, and generates reproducible results. Furthermore, recent advances in electronics and computing have hastened the development of mass spectrometers. Among this new generation, time-of-flight (TOF) is the most prominent because it requires fast processing of signals, usually at rates faster than 1 GHz. TOF mass analysis offers high mass accuracy, medium to high mass resolution, adequate dynamic range and fast spectrum acquisition capability.^{74,75} These features have made TOF analyses highly valued and broadly applied for targeted and nontargeted metabolite profiling.^{70,76-78} Therefore, development of technologies in separation and mass spectrometers make LC/MS become a powerful high-throughput approach for metabolomics.

1.4 Multivariate statistical tools for metabolomic analyses

Compared to traditional targeted analysis, one of the major advantages for metabolomics lies in the measurement of changes in metabolism resulting from environmental, genetic or developmental changes.^{79,80} Instead of tracking a few metabolites, a single GC-MS profile can measure 300-500 distinct compounds.²³ Although a wealth of information can be obtained from this, the challenges lie more in data processing than data acquisition. Large scale metabolomic studies usually involve screening of thousands of samples varying in genotype, treatment, or time in a single **study**. The large volume of generated data accentuates the need to apply and develop **data** processing software and multivariate analysis tools.

Recent developments in processing software or unbiased mass peak extraction and **alignment** of LC/MS data, such as Waters MarkerLynx, now offer improved workflows **for** untargeted metabolomics approaches, which intend to gather information on as many **metabolites** as possible for each extract analyzed.⁸¹⁻⁸³ MarkerLynx software extracts data **from** multiple LC/MS analyses, then detects, integrates, and aligns chromatographic **peaks**. A list of intensities of the peaks detected was generated for the first sample, using retention time (RT) and m/z data pairs as the identifier of each peak. An arbitrary number was assigned to each of these RT- m/z pairs in the order of their elution for data alignment.⁶⁹ Ions of different samples were considered to be the same ion when they demonstrated the same RT and m/z within limited mass and retention time windows. After such alignment, peak area data for each RT- m/z pair are assembled into a table. (Figure 1-1). If a peak is not detected in the sample, a zero is recorded as peak area in the data table. This process is repeated for each sample, and data from each run in the batch sorted so that the correct peak area data for each RT- m/z pair are aligned in a final data table.⁸⁴ Integrated peak areas within a sample are then normalized to the sum of the peak

intensities for that sample, as this facilitates certain statistical comparisons of sample data. After alignment and normalization, data from a batch of samples were ready for appropriate statistic and multivariate analysis tools.

Table 1-1 MarkerLynx software first detects peaks using specific mass data, aligns the peak according to retention time, integrates peaks over all samples, and assembles the results into a table.

Mass m/z	Ret. Time (min)	Ion intensity (sample 1)	Ion intensity (sample 2)	Ion intensity (sample 3)	Ion intensity (sample 4)
681.3215	23.4517	38873.9219	18603.3223	40866.7187	23265.209
519.3077	30.6518	0	0	58.5839	41.1124
723.412	33.3373	450.8516	241.5029	386.5996	172.6767
533.3221	33.0887	0	0	119.8434	73.4247
709.3905	32.1405	4977.9126	3300.7446	5445.8208	2131.6414
519.3061	31.7391	138.9787	0	164.2645	95.6386
765.4271	34.5681	533.9459	284.1887	430.8036	164.2941
449.2242	19.5197	58.3577	65.1068	114.182	90.7419
435.2096	17.4071	0	0	0	0
547.3348	34.7608	70.6505	25.1308	42.1335	0
575.3652	37.4728	40.1302	0	0	0
435.2106	16.7895	0	0	0	0
737.4205	35.3425	16415.6641	13558.0674	17265.7578	11048.0996
477.2637	24.6339	0	0	0	0
561.3495	35.6978	135.8656	39.7652	50.622	32.1877
853.4323	19.5759	0	0	0	0
993.5841	31.7724	0	0	0	0
561.3501	36.5467	25.7131	0	28.8188	0
519.3071	31.1841	0	0	21.776	0
695.3716	29.9849	0	0	0	0
533.3224	33.577	0	0	49.091	31.9687
463.2445	22.2556	0	0	0	0
547.3354	34.2186	0	25.1308	0	26.4573
519.3069	28.8797	0	0	0	0

Principal component analysis (PCA) is one of the oldest and most widely used unsupervised multivariate techniques.²³ PCA can reduce the dimensions of a complex dataset and provides a rapid visualization of similarity and difference among samples. The similar samples group together into one cluster and different ones are far from each other based on components they contain. This is a mathematical way to present the correlation among sample sets. The most important use of PCA is to represent a multivariate data table as a low-dimensional plane, usually consisting of 2 to 5 **dimensions**,⁸⁵ with each dimension representing an orthogonal component of the data **variance**. Selection of two dimensions, often chosen to represent a substantial fraction of **data** variance, leads to construction of a two-dimensional plane, and each of the sample **points** was projected into this plane to visualize all the samples. The coordinates of the **observation** in this plane are called scores, and visualization of those scores is termed a **score** plot. Figure 1-1 (Left) displays a scores plot, which presents a projection of **metabolite** data from three different plant genotypes, with each point representing a **single** sample. The samples belonging to one accession cluster together and different accessions are far from each other. Examinations of scores plots often reveal groups of observation, trends or outliers.

Once the relation among the samples is recognized, it is necessary to identify the variables responsible for the clustering. Such information is provided by a loadings plot, in which each point represents a specific variable (e.g. peak area for a specific mass-retention time pair). The loadings plot identifies variables most responsible for clustering of the data. In the case of metabolome data, this corresponds to signals from specific metabolites. An important aspect is that directions in the scores plot correspond to

directions in the loadings plot. PCA analysis provides a powerful tool for understanding the underlying patterns in multivariate data,⁶⁹ and finds frequent application for visualization of metabolomics data and discovery of samples that are metabolic outliers.⁸⁶⁻⁸⁸

PCA often fails to reveal contributions to sample discrimination by metabolite of minor abundance, and is often followed by OPLS-DA (orthogonal projection to latent structures-discriminant analysis). OPLS-DA is a supervised statistical tool to display **difference** in signal intensity between samples and potentially correspond to **biomarkers**.^{84,89,90} OPLS-DA is used to take advantage of additional knowledge about **each** the class to which each sample belongs (e.g. wild type vs. mutant), and this **information** is organized in what is known as the Y matrix.^{79,91} OPLS is to separate the **systematic** variation in X into two groups, one that is linearly correlated to Y and one that **is orthogonal** to Y.^{79, 92} Same as PCA, OPLS-DA also has scores plot, which clearly **displays** the differences of the two sample groups in the horizontal as well as the **differences** within the same sample groups along the vertical axis. In the S-plot, each point represents a pair with exact mass and retention time. As a result, the largest and reliable magnitudes variables in the upper and lower outer the S-plot. Thus, OPLS-DA is applied for modeling two classes of data to increase the class separation and find potential markers^{90,93} In fact, for assessment of metabolomics study, PCA is suggested as a starting point for analyzing multivariate data and rapidly present overview information hidden in the data.⁸⁵ Then OPLS-DA was followed to obtain a clear and more straightforward interpretation, especially, the interclass variation can be obtained from it.⁸⁵

Minor genetic variations can lead to subtle changes in complex plant chemistry, and recognition of these metabolic differences from mass spectrometry data poses substantial challenges. Owing to the potential for several hundred metabolites to be present in plant tissue extracts, identification of every metabolite is usually impractical and unnecessary for assigning plant gene functions. A more efficient approach should first aim to recognize metabolite signals that distinguish genotypes through application of multivariate statistical analysis of MS data, and the statistical tools described above provide important functions to aid recognition of metabolites that distinguish samples. Efforts to assign chemical structures can then be focused on the outlier metabolites.

1.5 Metabolite Identification

Chemical identification of metabolites in metabolomics is fundamental for the extraction of biological perspective from the data.⁶² The basic strategy behind metabolite identification today is similar to the foundations laid by the pioneers of organic mass spectrometry including John Beynon, Fred McLafferty and Klaus Biemann.⁸⁴ First, providing high mass accuracies (<5 ppm) for the ion of interest, the range of possibilities of molecular formulas is narrowed down, and in the ideal case, leads to a short list of possible molecular formulas for each ion.⁹⁴ However, there are still more possible mathematical combination of elements fit certain molecular mass than a number molecular formulas that exist chemically, even the instrument can provide the accurate measurement for molecular mass, especially for mass of ion above 400 amu.⁶² Several methods can be used to exclude impossibility, such as nitrogen rules, used for evaluation of the presence or absence of N atoms in a molecular or ion.⁹⁴ Another tool for narrowing

the number of molecular formulas is to make use of the isotopic pattern for specific mass.^{95,96} According to Kind and Fiehn,⁹⁶ this strategy can remove more than 95% of false positives.

Following assignment of likely elemental compositions to a molecular ion, the ion is converted into fragment ions using one of several ion activation methods and tandem or multistage mass spectrometry. Isolating one ion and performing MS/MS can track functional groups and connectivity of fragments for elucidating the structures of metabolites.⁹⁴ In addition, an in-source collision induced dissociation (CID) approach, named MS^E,⁹⁷⁻⁹⁹ produces accurate mass measurement for fragments, not just molecular ions. The advantage for accurate mass measurement of fragment ions is association of fragments with molecular ions to figure out the functional group corresponding to mass loss between them. For example, the mass loss 44.02 Da (CO₂) obtained from CID in negative mode indicates the ion contains one or more mono carboxylic acid groups. MS/MS spectra can provide an indication of putative structures of metabolites using database and/or manual interpretation of the fragments.^{100,101}

In addition, a term “mass defect” is also very important to classify the metabolites. Among elements commonly found in metabolites, ¹H and ¹⁴N make positive contributions to mass defect (+7.8 and +3.1 mDa per nucleus, respectively), heavier elements (¹⁶O, ³¹P and ³²S) make negative contributions to mass defect (-5.1, -26.2 and -27.9 mDa respectively), and ¹²C contributes zero to mass defect. Owing to the high number abundance of hydrogen and low frequency of phosphorus and sulfur in most metabolites, the mass defect for molecular, dimer, and fragment ions largely reflects the number of hydrogen atoms. Divided the mass defect by the molecular mass will give a

measurement of %H. This value might facilitate mining certain compounds. For examples, if the ion locates within 200-300 ppm in tomato trichomes, it often belongs to polyphenols.

Retention time is another parameter which can assess the polarity of metabolites when separation is coupled with mass spectrometer.⁹⁴ And the advanced techniques make retention time reproducible, allowing direct comparison of chromatograms and construct metabolite database.^{102,103} The structure can be determined from exact mass measurement, isotopic pattern, fragmentation pattern or search the existed libraries or database. A number of databases give lists of known compounds for particular elemental compositions or exact mass values,⁸⁴ including SciFinder Scholar (<http://www.cas.org/products/facad/fsflash.html>), Metlin (<http://metlin.scripps.edu/metabosearch.hhp>), ChemSpider (<http://www.chemspider.com/>) and many others.⁵¹ The final proof for identification of a compound is through comparison of LC retention time and MS/MS spectrum with that of an authentic standard.⁶² However, in most cases, there are few economically available standard compounds for such comparison, especially in the secondary metabolism field.⁹⁴ Thus, for identification of novel unknown structures or non-novel compounds without standards,⁸ we can use traditionally acceptable physicochemical properties, such as accurate mass, elemental analysis, UV/IR spectra, and/or NMR spectra.^{62,94}

1.6 Secondary Metabolites in *Solanum* Glandular Trichomes

Most plants have hairs on their aerial surfaces, called trichomes, that serve a number of functions ranging from protection against insects to heat and moisture conservation.¹⁰⁴

Trichomes range in size, shape, number of cells and morphology, as well as composition across species and accessions.^{8,105} There are two general categories of trichomes: glandular and nonglandular in *Solanum* species. Glandular trichomes contain or secrete a mixture of secondary metabolites at their tips, and many have significant commercial value as drugs, fragrances, food additives, and natural pesticides.^{4,106} Glandular trichomes are accessible cell types that are easily sampled, and for this reason they provide a useful model system for investigation of secondary metabolism.

Several classes of chemical compounds have been studied identified in the glandular trichomes. Acylsugars are one of them and widely distributed within the trichomes. These compounds have glucose and sucrose as center, acylated by short- to medium-chain length fatty acids (C2-C12) analogs, some of which are branched and derived in part from amino acids.¹⁰⁷⁻¹¹³ Acylsugars are secreted from the gland exudates and they cause the plant surface to become sticky and provides a strong deterrent to insects⁸. Except the possible taste different, no significant variation in biological activity has been attributed to structural differences among acylsugars.¹¹⁴ Therefore, why the trichomes secrete the diverse and complex acylsugars across species and accessions^{113,115-118} is still unknown. A more complete understanding of mechanism of acylsugar synthesis and regulation of these secondary metabolites is necessary for their engineering of plants for optimal acylsugar accumulation.¹¹⁹

Another metabolite class, the terpenoids, is recognized as the most abundant and structurally diverse,¹²⁰ consisting of the C10 mono-, C15 sesqui-, and C20 diterpenoids, derived from the MEP (2-methyl-D-erythritol-4-phosphate) biosynthetic pathway.¹²¹ Terpenoids are essential to the plant development and growth¹²² and those volatile

compounds can attract natural enemies of herbivores, either predators or parasitoids.¹²³ These compounds are widely distributed in the plant kingdom, that includes the mints, sages, and basil.⁹

In addition, flavonoids are commonly found in trichomes. Trichomes of *Phyllirea latifolia*⁸ and *Betula*¹²⁴ accumulate flavonoid glycosides. Flavonoid aglycones are found on the surface of many plant species¹²⁵ and present in the glands of peppermint leaves.¹²⁶ Alkaloids are another important class among secondary metabolites found in gland. Toxins nicotine and related alkaloids exist in the roots, leaves and trichomes¹²⁷ Alkaloids are reported to have a variety of pharmacological and nutritional properties in animal and humans, thus, we need a better understanding of the role of these compounds both in the plant and in the diet.^{127,128} In summary, the metabolic pathways involved in the biosynthesis of most secondary metabolites are not well understood at this stage. Thus, my efforts have been part of a larger *Solanum* Trichome Project which aims to understand the evolution of glandular trichome chemistry and function.

Tomato and related species in the family Solanaceae (e.g. peppers, eggplant, tobacco, and petunia) were chosen as target plants since a variety of trichomes are present on the surfaces of leaves, stems, and reproductive structures of those plants.^{129,130} Plus, a rich history of research on tomato genetics, has led to a strong set of tools based upon mutagenesis and introgression breeding with wild relatives of tomato¹³¹. The ability to get fertile offspring from crosses of tomato with related wild species provides important tools for a genetic dissection of glandular trichome development and metabolism since accumulation of secondary metabolites is often species- or accession-specific.^{128,132} Therefore, many functional genomics tools for tomato and related species will assist our

research on plant secondary metabolism. For example, after analyzing the introgression lines for changes in trichome metabolites and comparing back to the recurrent parent (*S. lycopersicum* cv. M82), the region of the genome controlling any observed changes can be narrowed down to a particular locus based on the metabolic phenotype of ILs with overlapping introgressions in our study.

1.7 Summary of Research Goals

In order to provide comprehensive information and speed up metabolite profiling, a new approach, described in chapter 2, named multiplexed CID was developed, which can expand the information yielded from LC/MS metabolite screening. Accurate mass measurement for fragment ions and molecular ions are obtained in a single analysis, and this information accelerates structural elucidation. The technique also generates information about the dependence of ion abundances on an electric potential used to induce ion-molecule collisions. The shapes of breakdown curves constructed based on multiple CID potentials provide the possibility to assign unknown ions as oligomers, fragments and molecular ions, with reduced matrix effects and enhanced quantitative dynamic range. Hundreds of metabolites are annotated and classified using this approach in Chapters 3 and 4, such as acylsugars, flavonoids, alkaloids and terpenoids. The annotation of these metabolites broadens our knowledge of metabolites in *Solanum* species, and especially opens the eyes of biochemists to study the biosynthetic pathway for these secondary metabolites and their function to herbivore resistance. In addition, a high-throughput approach was developed based on multiplexed CID for phenotype screening of introgression lines to accelerate gene discovery in Chapter 5. Three hundred

samples were screened within two days and five phenotypes were pointed out from 65 introgression lines derived from cultivated tomato M82 and wild type *S. pennellii* LA0716, which accelerate the discovery of gene functions. Finally, in Chapter 6, this approach has been applied to qualitative and quantitative analysis of glucose and sucrose esters pattern using LC/TOF MS, and compares branched- or straight-chain fatty acids of different breeding lines by GC/MS for hundreds of Cornell breeding lines. The results indicate this approach has great potential to support plant breeders in efforts to develop desirable insect-resistant acylsugar breeding lines of tomato in economically important crops.

Chapter 2. Discovery and deep profiling of metabolites using liquid chromatography/multiplexed collision-induced dissociation (muxCID)/mass spectrometry

2.1 Introduction

Plants and fungi produce a remarkable diversity of specialized secondary metabolites, many of which have biological activities important in medicine, industry, and agriculture. Often these are synthesized in specialized cell types such as glandular trichomes,¹⁰⁶ which produce chemicals with important pharmacological, flavor, and fragrance properties and in plant defenses against insects and disease. Despite valiant research efforts, progress in identifying and exploiting the genes responsible for metabolite biosynthesis has moved slowly owing to a lack of “-omics” resources that reach beyond standard model organisms. Recent developments in high-throughput DNA sequencing offer prospects for rapid gene discoveries, but association of these genes with metabolite biosynthesis will require efficient identification and deep profiling of metabolites that can keep pace with transcript profiling based on next-generation sequencing.^{29,133}

Levels of specialized metabolites are dynamic, dependent on tissue and cell type, and are sensitive to influences of both genetics and environment.³⁰ Non-targeted and deep profiling of specialized metabolites presents substantial technical challenges, and modern mass spectrometry provides a powerful tool for metabolite profiling.¹³⁴ More common approaches for identification and quantification of specialized metabolites require generation of mass spectra that can provide masses of molecular and fragment ions. While this can be achieved through real-time data-dependent selection of ions that undergo subsequent fragmentation¹³⁵ as is common in proteomic analyses,¹³⁶⁻¹³⁸ this approach is best performed using slow gradient elution and often fails to detect low abundance analytes. DDA results in a loss of data in the MS mode when MS/MS data are

being acquired, especially for low abundant metabolites which are below the predefined threshold, thus making it less than ideal for non-targeted metabolite profiling.^{139, 140} Furthermore, the data-dependent switching between MS and MS/MS experiments dictates that the analytical method varies from sample to sample, thus posing potential for compromising quantitative comparisons of metabolite levels.

The energy dependence of ion fragmentation has been shown to distinguish the presence of specific functional groups in peptides¹⁴⁰ and oligosaccharides,^{141,142} drug metabolites,^{143,144} and carotenoid metabolites,¹⁵ but the latter two investigations used slower scanning quadrupole mass analyzers. One promising option for enhancing information content of LC/MS analyses involves alternating of spectrum acquisition between gentle and energetic fragmentation conditions without precursor ion selection, a process recently termed MS^E acquisition.^{97-99,145-151} This approach yields molecular and fragment information in one analysis with a single and reproducible method of execution, often fast time-of-flight mass analyzers capable of recording signals for ions across typical *m/z* ranges used for LC/MS analyses on the time frame of <100 μ s. Algorithms have been developed to separate matrix interference from analyte signals.¹⁵¹ These efforts have employed two collision conditions, fragmenting and non-fragmenting.^{152, 153} To our knowledge few reports have generated spectra using more than two CID conditions in a single analysis, but the results were typically used to choose two conditions, fragmenting and non-fragmenting, used for sample analyses.¹⁵⁴ This report also suggested that a single high collision energy CID condition often did not yield desired fragmentation for all metabolites, but did not explore use of the energy-dependence of fragmentation beyond generation of structurally useful fragment ions for structure confirmation.

Deep profiling of complex mixtures of specialized metabolites poses the challenge of characterizing more than 1000 metabolites in a single LC/MS analysis. Further discoveries of the genes involved in metabolite biosynthesis will require large-scale profiling efforts of numerous genotypes, tissues, treatments, and time points for each. The needs of such efforts will drive LC/MS technologies to deliver more chemical information using shorter analysis times. In this study, we describe the multiplexing of non-selective CID (muxCID) as a powerful tool for enhancing information from non-target metabolite profiling in the form of nonselective generation of ion breakdown curves. We take advantage of the rapid acquisition capabilities of time-of-flight mass spectrometry to generate energy-resolved fragmentation with accurate mass measurements by switching among five different CID voltages on the chromatographic time scale. As the following sections will describe, the muxCID approach offers additional advantages for annotation of ions, extending quantitative dynamic range, and resolving matrix effects.

2.2 Experimental Section

2.2.1 Materials. Seeds for plants were obtained from the C. M. Rick Tomato Genetics Resource Center at the University of California-Davis. Plants used for this study were grown in a greenhouse at Michigan State University, and leaves were harvested immediately before metabolite extraction.

2.2.2 Chemicals. Acetonitrile (HPLC grade), 2-propanol (HPLC grade), methanol (HPLC grade), and formic acid (88% aqueous solution) were purchased from VWR Scientific; rutin was obtained from (Sigma-Aldrich, St. Louis, MO, USA).

2.2.3 Extraction Method. To extract metabolites from *Solanum trichomes*, ~100 mg leaves were dipped in 2 ml isopropanol: CH₃CN: H₂O (3:3:2 v/v/v) for 1 min. Extracts were centrifuged, and the supernatants were analyzed without further processing. Extracts were stored at -20 °C until analyzed.

2.2.4 Liquid Chromatography. A Shimadzu (Columbia, MD) LC-20AD HPLC ternary pump with SIL-5000 autosampler was used. Separations were performed on a Thermo BetaBasic column (1 × 150 mm, 3 μm particles) held at 30 °C that was interfaced to the mass spectrometer via electrospray ionization probe. A solvent gradient was executed based on 0.15% aqueous formic acid (A) and methanol (B) as follows: initial condition 5% B; linear gradient to 50% B (5 min); 95% B (33 min); 100% B (35 min); hold at 100% B until 38 min; return to 5% B (43 min). The flow rate was 0.1 ml·min⁻¹, and the injection volume was 10 μl.

2.2.5 Mass Spectrometry. Analyses were performed by coupling the HPLC separations (above) to a Waters LCT Premier mass spectrometer (Milford, MA), operated using electrospray ionization (ESI) and under control of MassLynx version 4.1 software. The ESI conditions were as follows: capillary voltage, 3200 V; cone voltage, 10 V; source temperature 90 °C; desolvation gas flow, 300 L·h⁻¹; desolvation gas temperature, 200 °C; cone gas flow, 20 L·h⁻¹. Detection was performed in the negative ion mode over *m/z* 50-1500 using centroided peak acquisition and dynamic range enhancement. Aperture 1 voltages were 10, 25, 40, 55, 80 V for five functions (where the scan time for each mode was 0.4 s with an interscan delay of 0.1 s).

2.3 Results and Discussion

Three hard realities confront investigators designing LC/MS protocols for comprehensive and nontargeted metabolite analysis. First is the notion that no single set of ion source conditions will be optimal for all metabolites. Conditions that deposit enough energy to dissociate noncovalent clusters of one metabolite may cause excessive fragmentation in others. Second, comprehensive and non-targeted metabolite profiling using LC/MS presents challenges to annotate each ion species as molecular, fragment, adduct or noncovalent dimer ions. This can present substantial challenges when metabolites coelute, as is often the case for large-scale metabolite profiling studies that require short analysis times to enable analyses of large numbers of biological replicates. Third, plant tissues produce a diverse suite of metabolites with abundances spanning several orders of magnitude, and extended dynamic range is an important issue for metabolite identification and quantification.

Extending nontargeted metabolite analysis to low-level metabolites demands recognition of ion type and acquisition of molecular and fragment mass information for as many metabolites as possible. This can be achieved by rapid switching among multiple CID conditions during a single analysis using a predetermined protocol (Figure 2-1A). Switching between CID conditions can be achieved on the time scale of about 10 ms, and spectra for each CID condition are stored in separate data acquisition functions within a single data file. Such analyses yield quasi-simultaneous acquisition of spectra under gentle and harsh conditions in addition to several conditions of intermediate severity

(Figure 2-1B). For the purposes of this study, five different CID conditions were employed to allow for sufficient time for each stored spectrum to correspond to accumulated signal from 8000 transients. Spectra acquired at any given time present a mixture of molecular, fragment, and noncovalent oligomer ions derived from all substances eluting at the time. Switching between CID conditions yields different amounts of decomposition of each ion into fragments, with the yields of various fragment ions depending on the reaction kinetics and activation energies for each fragmentation pathway. Further data processing can be performed for each retention time-ion m/z combination to generate breakdown curves (Figure 2-1C) showing the dependence of ion abundance on collision potential. The figure depicts behavior of ions from three coeluting metabolites. The first metabolite yields molecular, noncovalent dimer, and fragment ions, even at the lowest CID potential. Its low threshold for fragmentation allows it to yield substantial amounts of fragment ions at lower CID energies than the other two metabolites. Breakdown curves can be used to recognize noncovalent oligomers, as their signals decrease sharply with CID potential, and fragment ions, which have abundances that increase with CID potential until they fragment further. Different classes of metabolites may be distinguished based on collision energy thresholds and the slopes of breakdown curves for these ions.

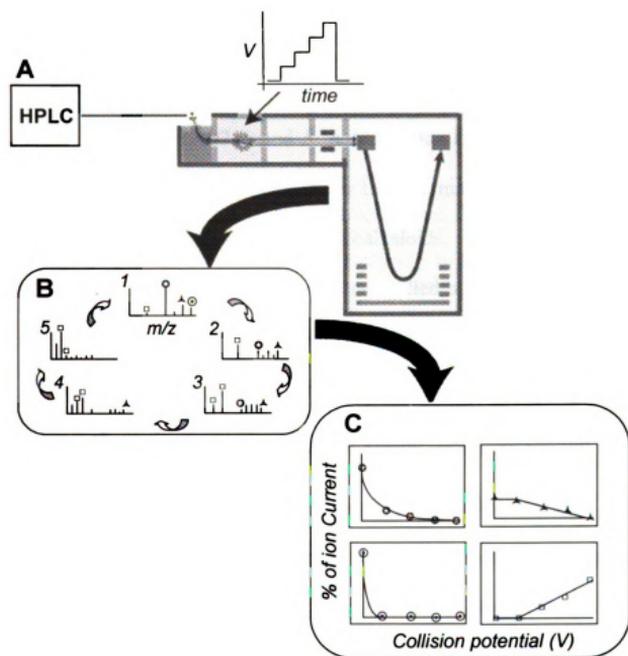


Figure 2-1. Schematic of LC/multiplexed CID method showing (A) application of multiple collision potentials to the ion transit lens to effect nonselective ion fragmentation, (B) mass spectra acquired using five different quasi-simultaneous CID conditions, showing molecular (☆), noncovalent oligomer (O), and fragment ion species (□) using color-coded symbols to distinguish ions coeluting metabolites, and (C) breakdown curves showing common behavior for molecular ion species for two coeluting metabolites, a noncovalent dimer ion, and a fragment ion.

To demonstrate proof-of-principle for the multiplexed CID approach, LC/MS metabolite analyses were performed on solvent dip extracts of leaves of two wild relatives of tomato, *S. pennellii* LA0716 and *S. habrochaites* LA1777. These extracts each contain more than 300 metabolites detectable by LC/MS, many of them acylglucose and acylsucrose esters, flavonol glycosides, glycoalkaloids, and others yet to be identified. For these analyses, pre-analyzer CID was controlled by adjusting the potential applied to Aperture 1 in the transit region between the ion source and the TOF analyzer. Pressures in this region are not controlled or measured on this instrument, but are estimated to exceed 10^{-4} mbar and are sufficient to ensure >99% conversion of some molecular ion species to fragment ions. The instrument control software allows multiple acquisition functions, each with distinct Aperture 1 settings. For these experiments, five functions were set up, with Aperture 1 voltages ranging from 10-80 V. The choice of 5 functions was made to enable switching among the functions with a duty cycle compatible with conventional HPLC separations.

The quasi-simultaneous acquisition of metabolite mass spectra is demonstrated in Figure 2-2, which presents metabolite mass spectra acquired in negative mode ESI using five different CID potentials. In each case, the bottom spectrum was generated under the gentlest conditions (Aperture 1 = 10 V), with the top spectrum acquired under harshest conditions (80 V). The first set of spectra (Figure 2-2A) demonstrates behavior for the glycoalkaloid α -tomatine. At CID settings of 10 and 25 V, only the formate adduct (m/z 1078) is observed owing to the lack of acidic groups that favor formation of the deprotonated ion $[M-H]^-$. With increasing CID potential, $[M-H]^-$ at m/z 1032 appears

as a result of loss of formic acid, but more extensive fragmentation was only observed at the highest potential (80 V), primarily involving losses of portions of the oligosaccharide group. In Figure 2-2B, behavior of a triacylglucose (two C4 and one C5 fatty acid esters) is demonstrated. At the lowest CID energy, noncovalent dimer ions corresponding to $[2M-H]^-$ and $[2M+formate]^-$ were observed at m/z 807 and 853 respectively, $[M+formate]^-$ at m/z 449 was the most abundant ion, and minor amounts of $[M-H]^-$ (m/z 403) and a fragment ion at m/z 227 were present but barely detectable. As is often the case for specialized metabolites, the acylsugar lacks groups with moderate acidity, so formate adducts were the most abundant signals at lowest CID potential. As this potential was increased, $[M-H]^-$ grew in abundance along with several fragment ions such as m/z 315 and 227, which form via successive losses of neutral C4 fatty acids. At the highest CID energy, the C4 fatty acid anion (m/z 87) dominates the spectrum, with a smaller amount of the C5 fatty acid anion observed at m/z 101. These two fragment ion masses provide evidence for the fatty acid groups. We speculate that differences in yields of the two fatty acid anions may be attributable to differences in the substitution positions of these groups, and a more detailed analysis of CID spectra of *Solanum* metabolites will be described in Chapter 3 and 4.

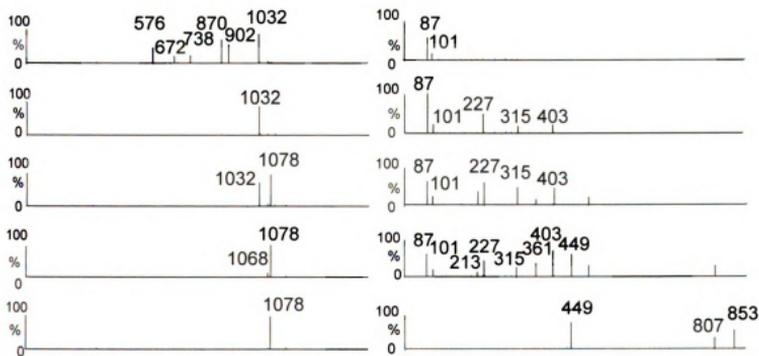
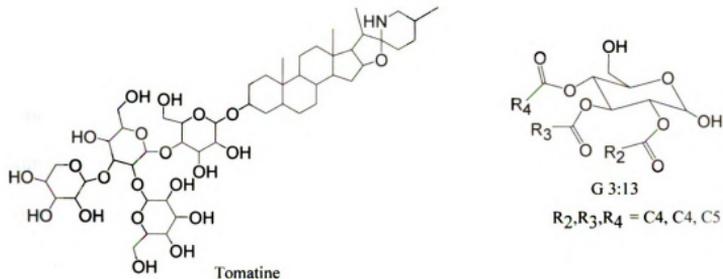


Figure 2-2. Negative mode electrospray ionization mass spectra obtained using multiplexed collision induced dissociation using five aperture 1 voltages (from bottom to top: 10, 25, 40, 55, and 80 V) for two metabolites extracted from *S. pennellii* LA0716 by leaf dip. Panel A: α -tomatine at retention time of 15.25 min; panel B: triacylglycerol G 3:13 with two *iso*-C4 and one branched C5 fatty acid esters with retention time 19.99 min.

Annotation of detected ions as molecular, noncovalent oligomer, adduct, or fragment ions serves as the foundation for interpreting nontarget metabolite mass spectra. Metabolites often coelute, and this makes such assignments more challenging. However, multiplexing of CID conditions makes possible the construction of breakdown curves for every m/z value at every retention time during a chromatographic separation. Figure 2-3 displayed TIC (total ion chromatogram) for *S. habrochaites* LA1777 and *S. pennellii* LA0716 under lowest CID voltage and the metabolites used in Figure 2-4 and 2-5 were labeled in TIC. To illustrate the practice of this feature, we constructed breakdown curves (Figure 2-4) for dimer, adduct, and fragment ions and deprotonated molecules derived from four metabolites extracted from the tomato relatives *S. pennellii* LA0716 and *S. habrochaites* LA1777. The structures for those four metabolites were labeled on the top of Figure 2-4. The absolute abundances of all dimer ions and formate adducts exhibited a sharp drop with increasing CID potential (Figure 2-4A, 4B), but most chloride adducts were more persistent owing to their higher thresholds for decomposition (data not shown). Various fragment ions appeared only at higher CID potentials (Figure 2-4D) and the abundances increased with elevated CID voltages. Such behavior allows fragment ions to be distinguished from molecular ion species, and targets these signals for use in structure assignments. Abundances of $[M-H]^-$ ions display various dependence on CID potential, often reaching maxima at intermediate voltages (Figure 2-4C) owing to their formation from decomposition of dimers and some adduct ions. The most important aspect of this finding suggests that shapes of breakdown curves can guide assignment of those ions as molecular, noncovalent oligomer, or fragment species.

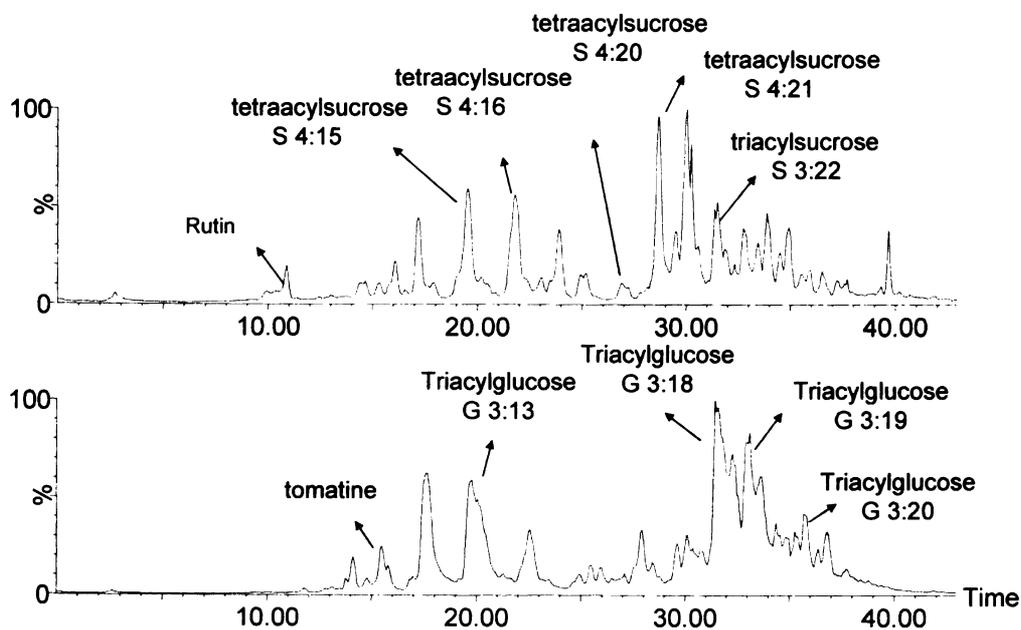
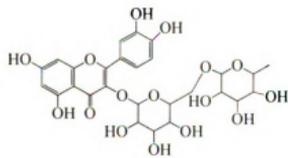
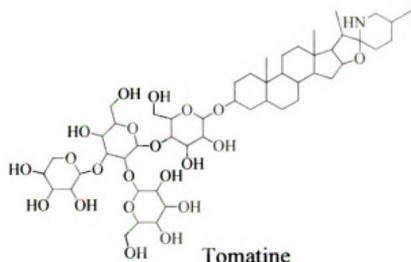


Figure 2-3. Total ion chromatogram (TIC) for *S. habrochaites* LA1777 (top) and *S. pennellii* LA0716 (bottom) using ESI negative with CID potential at 10 V. The labeled metabolites were used as examples for Figure 2-4, 2-5 and 2-6.

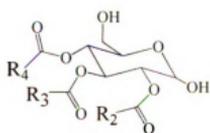
Figure 2-4. Breakdown curves showing energy dependence of various ion abundances for metabolites extracted from leaf dips of *S. pennellii* LA0716 and *S. habrochaites* LA1777. Ion abundances are normalized to the total ion current for each individual ion as summed for all five collision potentials. Panel A: Breakdown curves of non-covalent dimer ions for the flavonoid glycoside rutin with $[2M-H]^-$ (◆) and $[2M+HCOO]^-$ (◇) and two acylsugar metabolites, tetraacylsucrose with 21 carbon atoms in acyl groups detected as $[2M-H]^-$ (▲) and $[2M+HCOO]^-$ (△); dimer ions of triacylglucose with 19 carbon atoms in acyl groups detected as $[2M-H]^-$ (■) and $[2M+HCOO]^-$ (□); Panel B: breakdown curves for formate adduct ions $[M+HCOO]^-$ for the glycoalkaloid tomatine (●), rutin (□), tetraacylsucrose (○), and triacylglucose (●); Panel C: breakdown curves for monomeric $[M-H]^-$ ions for tomatine (*), rutin (×), tetraacylsucrose (□), and triacylglucose (◆); Panel D: breakdown curves for major fragment ions for tomatine at m/z 870 (■), rutin at m/z 300 (△), C5 (×) and C4 (□) fatty acid anion fragments for tetraacylsucrose, and C5 (○) and C10 (△) fatty acid anion fragments for triacylglucose. The proposed structures of four metabolites are on the top of breakdown curves.



Rutin

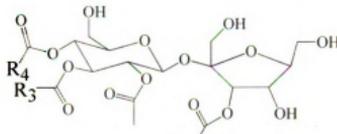


Tomatine



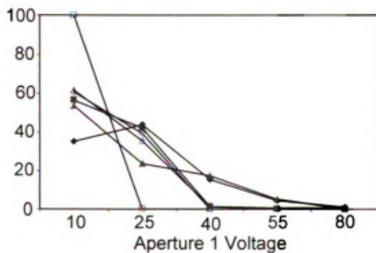
G 3:19

$R_2, R_3, R_4 = C4, C5, C10$

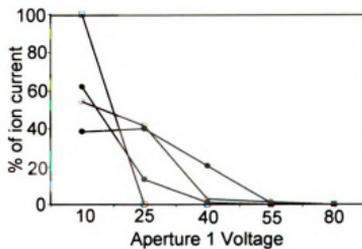


S 4:21

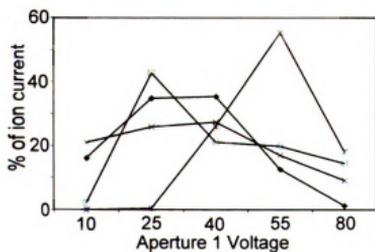
$R_3, R_4 = C4, C5, R_3' = C10$



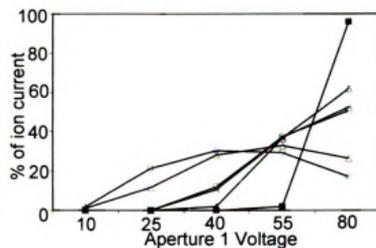
A



B



C



D

Shapes of breakdown curves are determined by the ease with which compounds form various ions. At low CID potentials, the more acidic metabolites were more likely to form $[M-H]^-$ than adducts, and this was reflected in the dominant abundances of $[M-H]^-$ and dimer ions in spectra of rutin at low CID potentials. Abundances of formate adducts of the weakly acidic metabolite rutin had the sharpest drop among adducts of these four metabolites. The relative abundances of formate adduct ions to deprotonated molecular species therefore depend on metabolite formate ion affinities, metabolite acidities (e.g. ease of conversion of $[M+formate]^-$ to $[M-H]^-$) and energy barriers toward fragment ion formation. Collision energy dependences of ions derived from six acylsugars (one triacylsucrose, two tetraacylsucroses, and three triacylglucoses) were compared respectively (Figure 2-5 and 2-6). The results highlight the consistency of shapes of breakdown curves generated by multiplexed CID within each chemical class. Application of our findings to several hundred metabolite breakdown curves generated by multiplexed CID suggested that comparisons of breakdown curve shapes have value for recognition of chemical classes of metabolites. This feature is important for plant metabolomics since it was common there were no report in database or literature for many secondary metabolites, but at least they can be assigned to chemical classes under the help of breakdown curves.

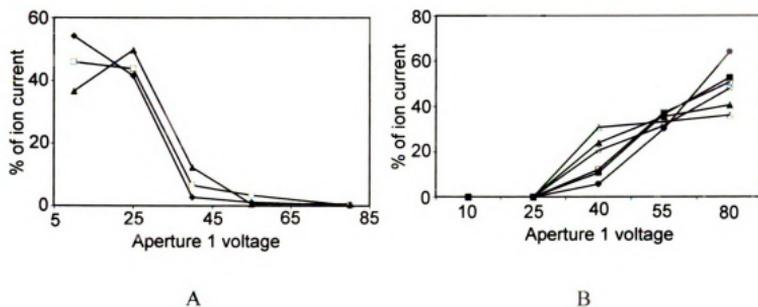
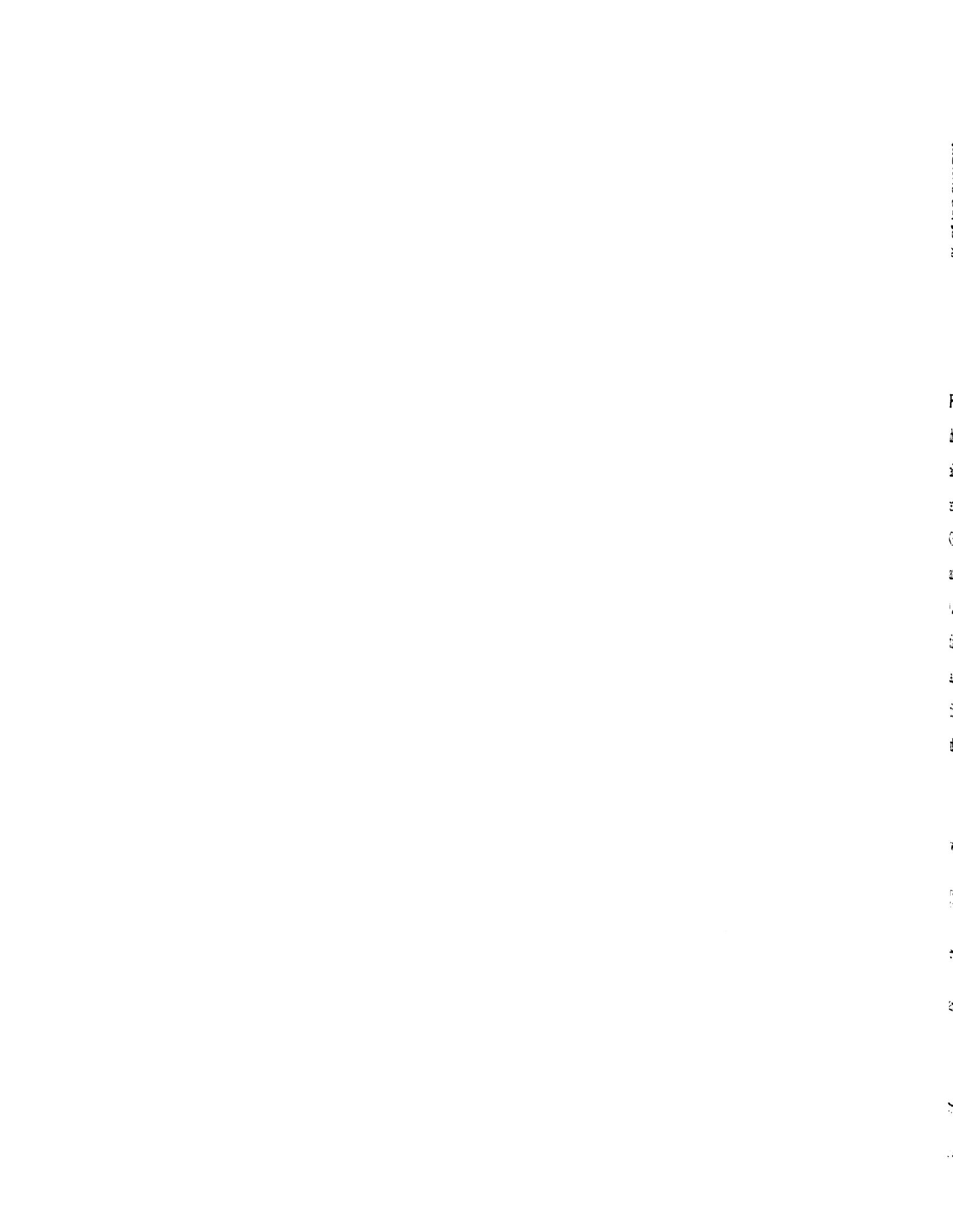


Figure 2-5 A: Breakdown curves derived from metabolites detected during LC/negative ion ESI MS analyses of a leaf dip extract from *S. habrochaites* LA1777, showing collision energy dependence of abundances of (A) formate ion adducts of three acylsucrose metabolites ($[M+HCOO]^-$ at m/z 737 S3:22 (◆); $[M+HCOO]^-$ at m/z 653 S4:15 (□); $[M+HCOO]^-$ at m/z 723 S4:20 (▲) and (B): fatty acid fragment ions of various carbon lengths (C5 fatty acid at m/z 101 (■) and C12 fatty acid at m/z 199 (□) for S3:22), (C4 fatty acid at m/z 87 (▲) and C5 fatty acid at m/z 101 (△) for S4:16), (C4 fatty acid at m/z 87 (●) and C10 fatty acid at 171 (○) for S4:20). The acylsucroses are designed as S3:22, S4:16 and S4:20, with the letter S representing the sucrose core, the first number represents the number of fatty acyl groups, and the second number corresponds to the total number of carbon atoms in the fatty acyl groups.



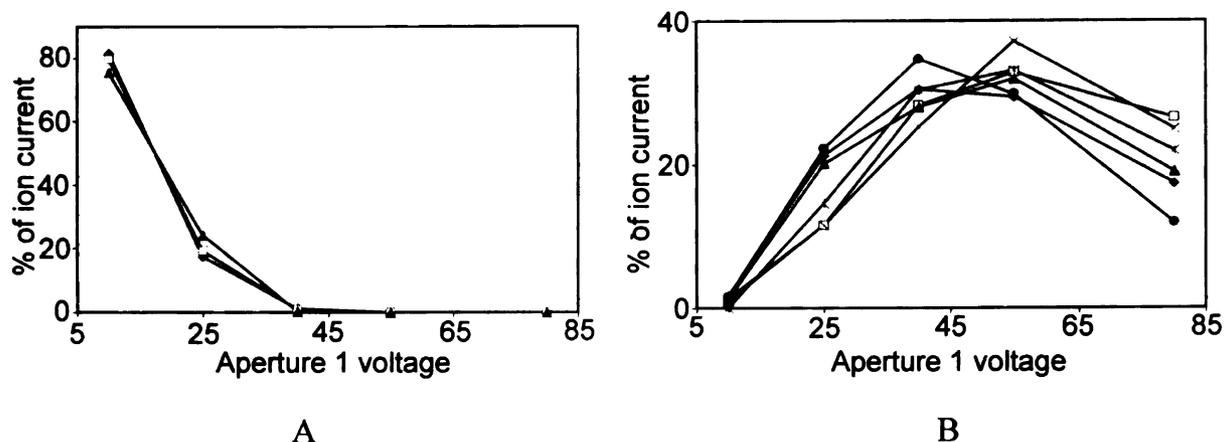


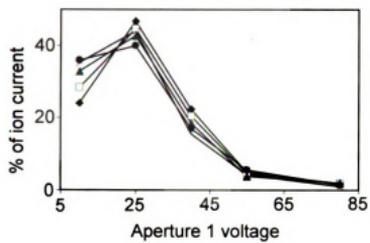
Figure 2-6 Breakdown curves derived from LC/negative ion ESI MS analyses of a leaf dip extract from *S. pennellii* LA0716 showing collision energy dependence of abundances of: (A) formate ion adducts of three triacylglucose metabolites ($[M+HCOO]^-$ at m/z 547 G3:20 (●); $[M+HCOO]^-$ at m/z 519 G3:18 (□); $[M+HCOO]^-$ at m/z 449 G3:13 (▲) and (B): fatty acid anion fragments of various carbon lengths. (C5 fatty acid at m/z 101 (◆) and C10 fatty acid at m/z 171 (□) for G 3:20); (C4 fatty acid at m/z 87 (▲) and C10 fatty acid at m/z 171 (×) for G3:18); (C4 fatty acid at m/z 87 (*) and C5 fatty acid at m/z 101 (●) for G3:13). The triacylglucoses are designated as G3:13, G3:18, and G3:20 with the letter G representing the glucose core, the number 3 refers to three fatty acyl groups, and the last number corresponds to the total number of carbon atoms in the fatty acyl groups.

The discussion above has emphasized behavior of negative ions, but similar results were obtained using positive ion mode. For acylsugar metabolites, the adduct ion $[M+NH_4]^+$ for all tested acylsugars underwent fragmentation at lower CID threshold energies than $[M+Na]^+$, and MS/MS product ion spectra showed that most fragment ions arose from $[M+NH_4]^+$ and not $[M+Na]^+$.

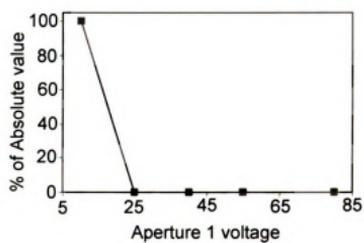
We considered the possibility that analyte concentration could influence shapes of breakdown curves owing to formation of dimer and other oligomer ions at higher analyte concentrations. Breakdown curves were generated using LC/multiplexed CID for rutin

standard ranging from 20 to 800 pmol injected (2-80 μ M solutions). Breakdown curves displayed in Figure 2-7 showed minimal concentration-dependence for fragment ions. Earlier studies have shown that influences of solvent composition and pH on breakdown curves were negligible.¹⁵⁵

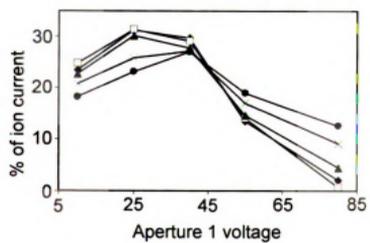
Figure 2-7 Breakdown curves for the flavonoid glycoside rutin obtained using negative mode electrospray ionization following injections of 10 μL of solutions at five different concentrations (2 (\blacklozenge), 5 (\square), 10 (\blacktriangle), 40 (\times), and 80 (\bullet) μM). A: non-covalent dimer ion $[2\text{M}-\text{H}]^-$ B: formate adduct ion $[\text{M}+\text{HCOO}]^-$ C: monomer deprotonated ion $[\text{M}-\text{H}]^-$ D: aglycone radical anion fragment with m/z 300 and E: aglycone fragment ion with m/z 301.



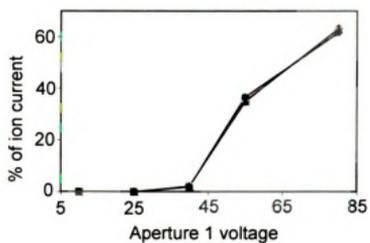
A



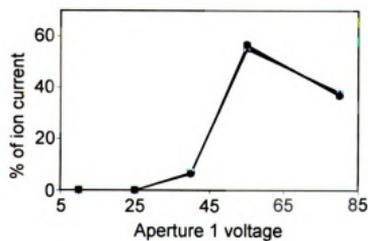
B



C



D



E

Two great advantages of the multiplexed CID approach lie in its ability to generate fragment mass information for low abundance metabolites and to discriminate coeluting metabolites based on different collision energy thresholds for fragmentation. To illustrate this point, we present an example using two coeluting metabolites derived from a leaf dip extract of *S. pennellii* LA0716 (Figure 2-8). Two metabolites of minor abundance coelute at 22.5 min, a triacylglucose (with two C4 and one C5 fatty acid ester groups) that gives a formate ion adduct at m/z 463, and trimethylquercetin that yields $[M-H]^-$ at m/z 343 that is barely evident in the low voltage spectrum. Triacylglucoses undergo facile fragmentation and lack acidic functionality. Even at this lowest collision potential, the spectrum displays multiple triacylglucose peaks including deprotonated dimer (m/z 835), formate-bound dimer (m/z 881), the deprotonated triacylglucose (m/z 417), chloride adduct (m/z 453 and 455) and a fragment ion (m/z 375, corresponding to loss of a neutral C4 fatty acid from the formate adduct) are observed at higher abundance than the deprotonated molecule of trimethylquercetin. The mass spectrum generated at 25 V CID potential shows fragment ions derived from the triacylglucose that correspond to neutral losses of the fatty acid groups (m/z 329, loss of C4 fatty acid from $[M-H]^-$ and m/z 241 and 227 arise from subsequent losses of C4 or C5 fatty acid respectively). Fragments corresponding to the anions of C4 and C5 fatty acids are also evident at m/z 87 and 101. In contrast, the spectrum at 80 V is characterized by fragment ions derived from trimethylquercetin, which requires greater collision energy than the triacylglucose to yield abundant fragment ions. At the higher collision potential, the triacylglucose is largely converted to low mass fragment ions, which no longer obscure the trimethylquercetin fragments.

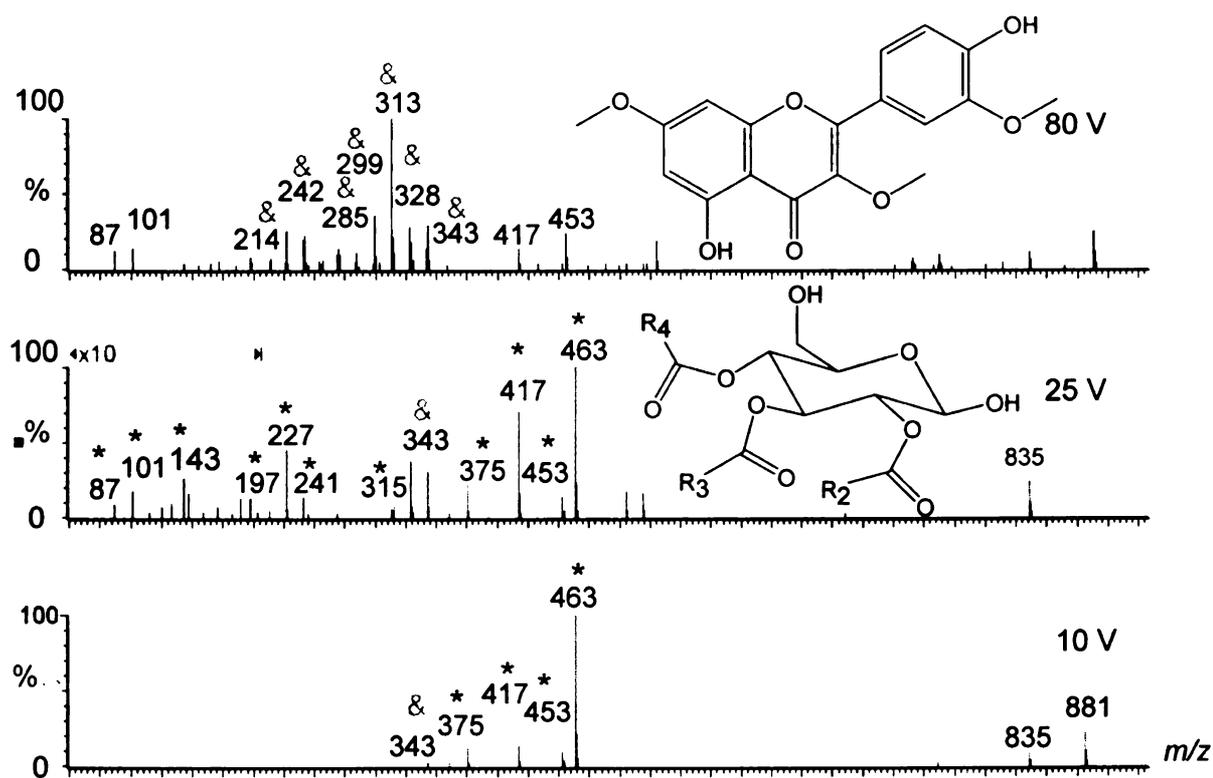


Figure 2-8. Multiplexed CID mass spectra (electrospray, negative mode) of coeluting metabolites from LC/TOF MS analysis of an extract of *S. pennelli* LA0716, showing spectra obtained from three of the five Aperture 1 potentials. The lowest collision potential (10 V) yields multiple ions from a triacylglycerol (S3:14) including formate and chloride adducts (m/z 463 and 453), deprotonated (m/z 417), fragment (m/z 375), and dimer ions (m/z 881 and 835). Of markedly lower abundance is $[M-H]^-$ of trimethylquercetin (m/z 343). Slightly higher CID potential (25 V, middle panel) shows extensive fragmentation of the triacylglycerol (ions marked with *); spreading the signal among numerous fragment ions allows more ready observation of the methylated flavonoid ion (marked with &). Fatty acyl anions are observed at m/z 87 and 101 corresponding to C4 and C5 fatty acyl groups. In the top panel, the highest collision voltage of 80 V yields fragmentation of the methylated flavonoid through losses of methyl groups.

Multiplexed CID-TOF mass spectra generate information about molecular and fragment ion species including accurate mass measurements and elemental composition based upon abundances of isotopomer ions. With proper instrument calibration, multiplexed CID-TOF spectra yielded ion masses accurate to within 0.005 Da of true values for all molecular and fragment ions that have adequate ion statistics for mass measurement. The coupling of muxCID with TOF mass measurement allows ions of overlapping nominal mass to be distinguished based upon mass defects (the fraction of ion mass following the decimal point). Among elements commonly found in metabolites, ^1H and ^{14}N make positive contributions to mass defect (+7.8 and +3.1 mDa per nucleus, respectively), heavier elements (^{16}O , ^{31}P and ^{32}S) make negative contributions to mass defect (-5.1, -26.2 and -27.9 mDa respectively), and ^{12}C contributes zero to mass defect. Owing to the high number abundance of hydrogen and low frequency of phosphorus and sulfur in most metabolites, the mass defect for molecular, dimer, and fragment ions largely reflects the number of hydrogen atoms.

High mass accuracy provides information essential for metabolite discovery, but even technologies capable of the highest accuracy mass measurements fail to yield conclusive molecular formula assignments.¹⁵⁶ The number of elemental formulas that yield agreement with a measured mass grows exponentially with mass, and even mass measurement accuracy of 1 part-per-million can result in hundreds of matching formulas for metabolites with molecular masses greater than 500 Da. Combining accurate mass measurements with quantitative assessments of abundances of ions with heavy isotopes can limit the range of matching elemental formulas. Extending this approach to include accurate fragment mass measurements and fragment isotopomer abundances adds

additional discriminating information about metabolite formulas and structures.¹⁵⁷ Linking of fragment and molecular ion masses through their common chromatographic elution profiles enjoys common practice in GC/MS analyses, and has been implemented in such context using software packages such as AMDIS.¹⁵⁸ We suggest that rapid multiplexing of CID with accurate measurements of mass and isotopomer abundances for molecular and fragment ions adds the additional dimension of measuring energy-dependence of ion abundances to accelerate discrimination of metabolites through several features: (1) suggestion of elemental formulas for molecular and fragment ions, (2) recognition of functional group types via assignments of group-specific fragment ions or neutral mass losses, and (3) discrimination of molecular structure and ion types according to ion appearance thresholds from breakdown curve analysis. Multiplexing of CID in a single analysis allows this approach to reach low-abundance metabolites providing that ion signals are abundant enough to allow for precise measurements of these signals, and does not require data-dependent ion selection for MS/MS scans. In essence, this approach provides the benefits of MS^E analysis⁹⁹ with the additional dimension of energy-resolution of ion fragmentation.

When chromatographic separations of metabolites are incomplete, multiple metabolites will arrive at the mass spectrometer at the same time. In such cases, extracted ion chromatogram profiles and mass defects are useful aids for assigning fragment ions with appropriate molecular ion species. This is illustrated in the case of coeluting glycoalkaloid and flavonoid metabolites present in extracts of leaves from an introgression line (IL) 1-1 derived from *S. lycopersicum* M82 and *S. pennellii* LA0716. Based on extracted ion chromatograms generated at low CID potential (10 V), the

chromatographic profiles of two metabolites detected at m/z 609 and 1076 overlap, with retention times of 2.46 and 2.50 min. respectively (Figure 2-9). Second pair of metabolites, detected at m/z 593.15 and 1094.54 are not resolved at all, with a retention time of 2.60 min for both.

For the first pair of metabolites which overlap, the TIC shows a skewed peak that does not resolve the two metabolites. A composite spectrum was generated for the 2.46 min peak by averaging spectra across the peak (Figure 2-9 A and B XICs). This was performed for all five CID potentials, and spectra for three of the five acquired functions are displayed in Figure 2-10. The lowest energy mass spectrum shows three prominent peaks at m/z 609.15, 655.15, and 1076.52, and additional ions are observed at higher CID potentials. Extracted ion chromatograms (XICs) were generated for all ions observed at the highest (80 V) CID potential, and these are displayed in Figure 2-11. Fragment ions commonly inherit the elution profiles from molecular ions, and two distinct sets of chromatographic profiles were observed, with one set of ions eluting with a peak apex at 2.46 min, and the other set eluting with an apex at 2.50 min. The earlier eluting peak is annotated as quercetin-3-O-rutinoside, and the second peak corresponds to an isomer of dehydrotomatine that is described in more detail in Chapter 4. In this case, resolution of the ultraperformance LC column provided distinct chromatographic elution profiles that allowed fragment ions to be associated with specific metabolites without MS/MS spectra.

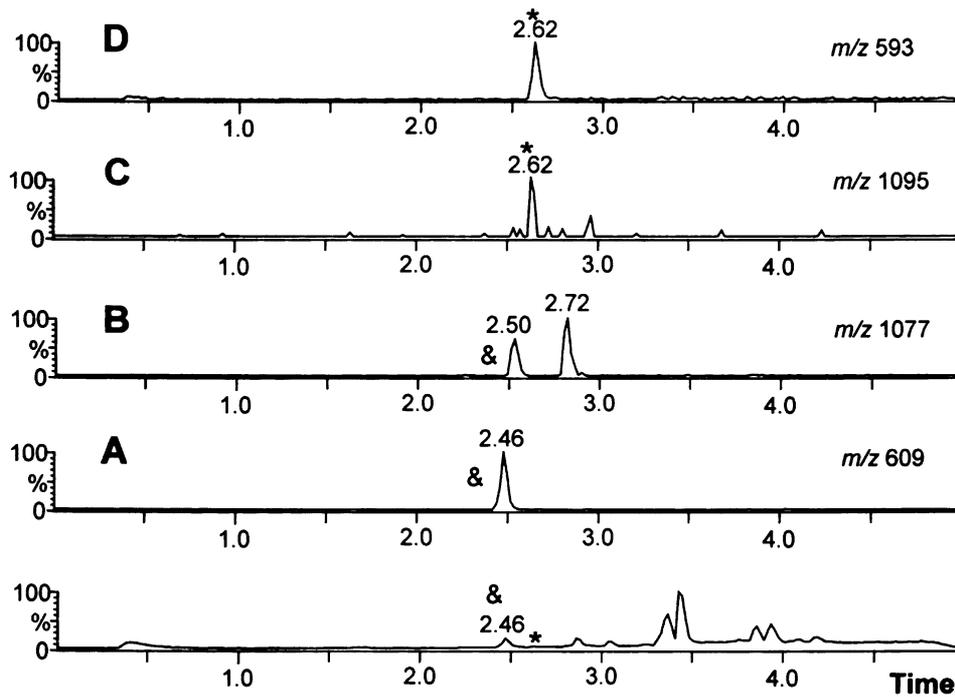


Figure 2-9 TIC for IL1-1 (Bottom) and XIC of m/z 609, 1076 (elute at 2.46 min), 1094 and 593 (elute at 2.60min) under ESI negative mode. The peaks from XICs of m/z 609 and m/z 1077 labeled with & were corresponding to coeluting peak in TIC at RT = 2.46 min; the peaks from XICs of m/z 593 and m/z 1095 labeled with * were corresponding to peak in TIC at RT = 2.62 min.

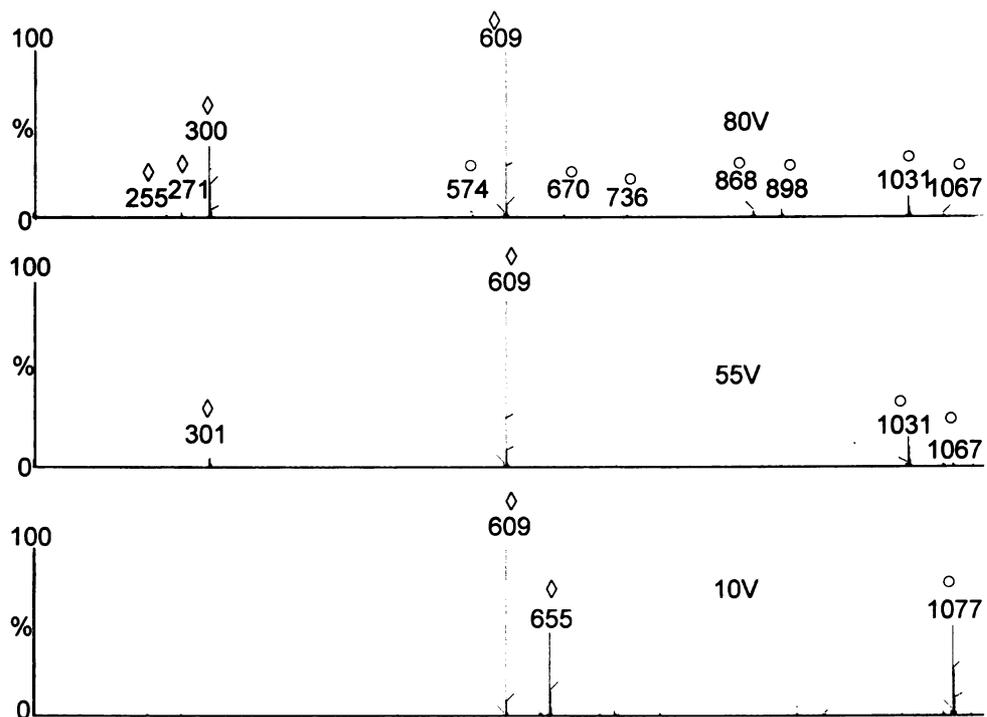


Figure 2-10 Averaged mass spectra across the LC/MS TIC peak eluting at 2.46 min using three CID voltages (10, 55 and 80 V). The metabolites are from an extract of trichomes from IL 1-1. The ions labeled with ○ belong to one dehydrotomatine isomer and the ion peaks labeled with ◇ are corresponding to quercetin-3-O-rutinoside (rutin).

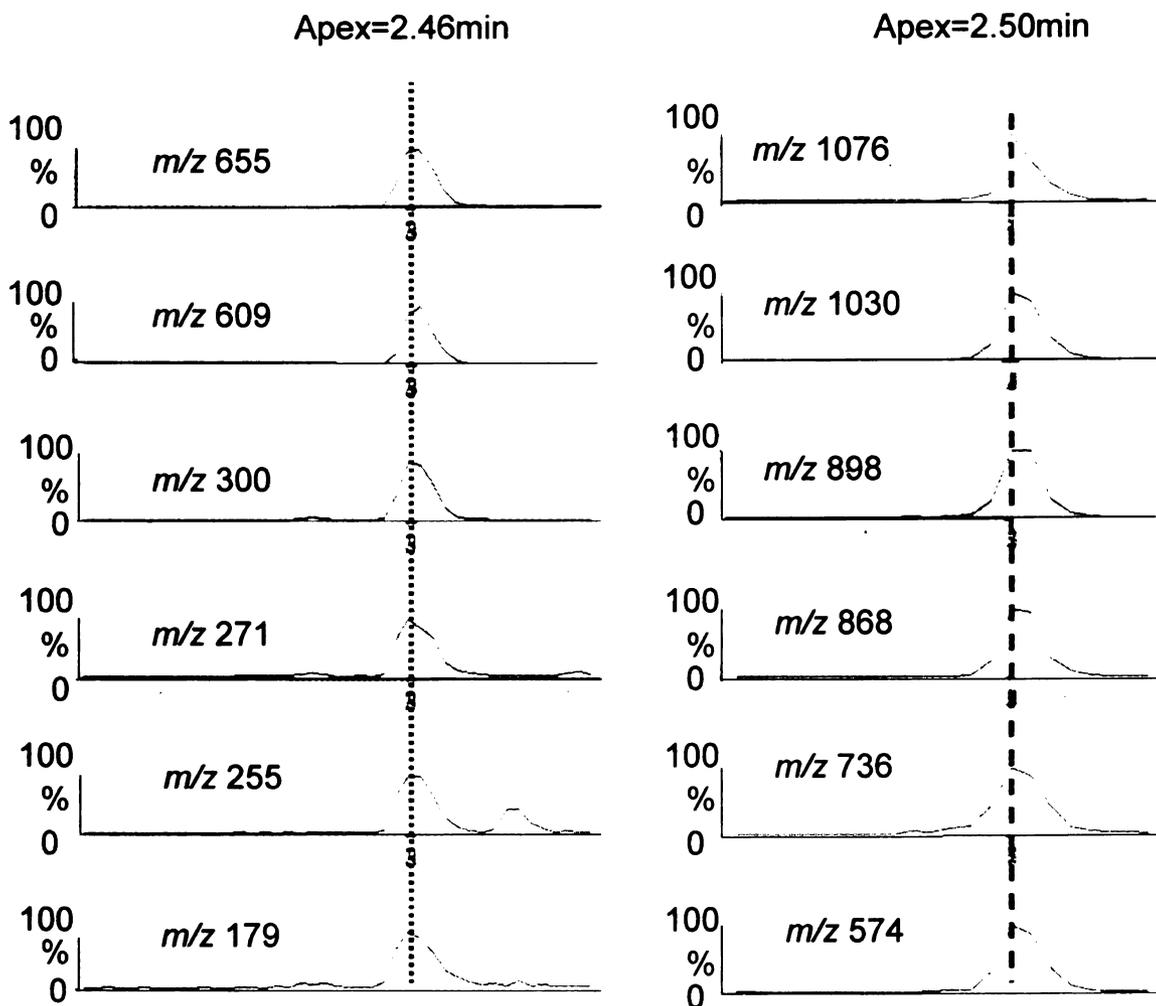
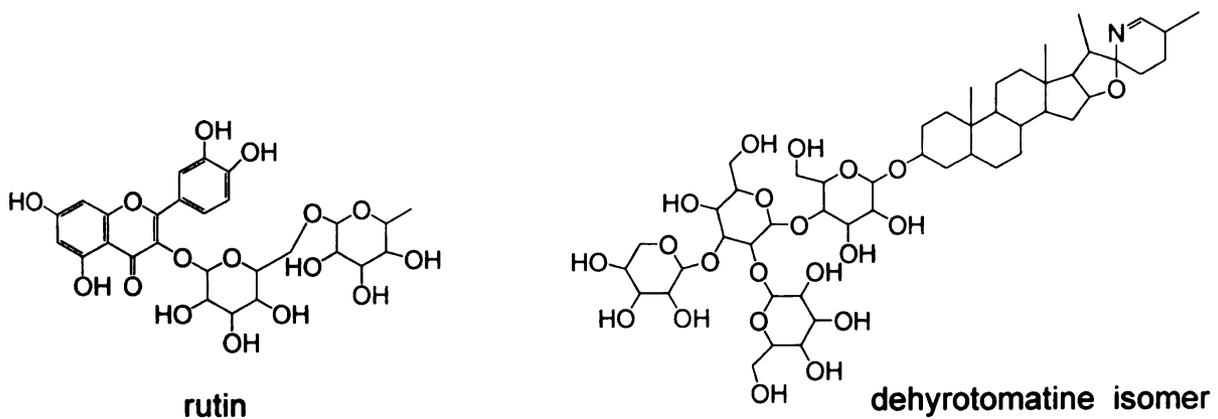


Figure 2-11 XICs for the ions labeled with \circ and \diamond from Figure 2-10 were generated for LC/MS analysis of an extract of IL 1-1. The chromatograms on the left are assigned as being derived from quercetin-3-O-rutinoside (rutin) and the chromatograms on the right are related to an isomer of dehydrotomatine.

It is almost unavoidable that some metabolites will have indistinguishable chromatographic elution profiles, and this was the case for metabolites eluting at 2.60 min. (Figure 2-9, C and D XICs). The spectrum derived from averaging across the TIC peak showed major ions at m/z 593.15, 639.15, and 1094.54 using the lowest CID potential. As before, spectra were also generated at raised CID potentials (Figure 2-12). Breakdown curves were calculated for the three ions above (Figures 2-13 and 2-14). The peaks at m/z 639.15 and 1094.54 had disappeared when the CID potential was raised to 55 V. Such behavior is consistent with both being formate adduct ions, and this assignment is supported by the appearance of ions 46 Da lower (corresponding to $[M-H]^-$) for both cases. The abundance of the ion of m/z 639.15 drops off more quickly with increasing CID potential than is the case for m/z 1094.54. This behavior suggests greater acidity or proton mobility in the former, because a proton must migrate to allow for neutral loss of formic acid. Also, at 55 V CID potential, two fragment ions appear at m/z 285.04 and 284.04 (unlabeled), and at the highest CID potential (80 V), additional fragment ions appear at m/z 255.03, 227.02, 592.39, 754.44, 886.48, 916.50. Support for annotation of these ions as fragment ions comes from slopes of their breakdown curves, which increase with increasing CID potential. It is worthy to note that fragment ions at m/z 592.39, 754.44, 886.48 and 916.50 all have even masses and positive mass defects above 390 mDa, which are traits common to the ions at m/z 1094.54 $[M+HCOO]^-$ and 1030.52 $[M-H]^-$. In addition, the shapes of breakdown curves for these molecular ion and fragment ions were similar with corresponding ions from dehydrotomatine (Figure 2-13). These similarities in mass defects and breakdown curves suggested this metabolite has similarities with other glycoalkaloids. Based on the combined information from

accurate mass measurements and breakdown curves, this metabolite was assigned as hydroxytomatine. Ions at m/z 593.15, 639.15, 284.04, and 285.04 did not fit behavior or mass defects of ions from glycoalkaloids, and were attributed to a coeluting metabolite with structural units and elemental compositions distinct from the glycoalkaloid family. Fragment ions at m/z 284.04 and m/z 285.04 gave calculated elemental formulas characteristic of fragments corresponding to the aglycone kaempferol, and breakdown curves were similar to the flavonoid glycoside rutin (Figure 2-14). Based on the sum of information from accurate mass measurements and breakdown curves, this metabolite was assigned as kaempferol-3-O-rutinoside. Thus, with the help of breakdown curves and mass defect, even coeluting metabolites can be distinguished and annotated. The annotation of these and other metabolites is described in more detail in Chapter 4.

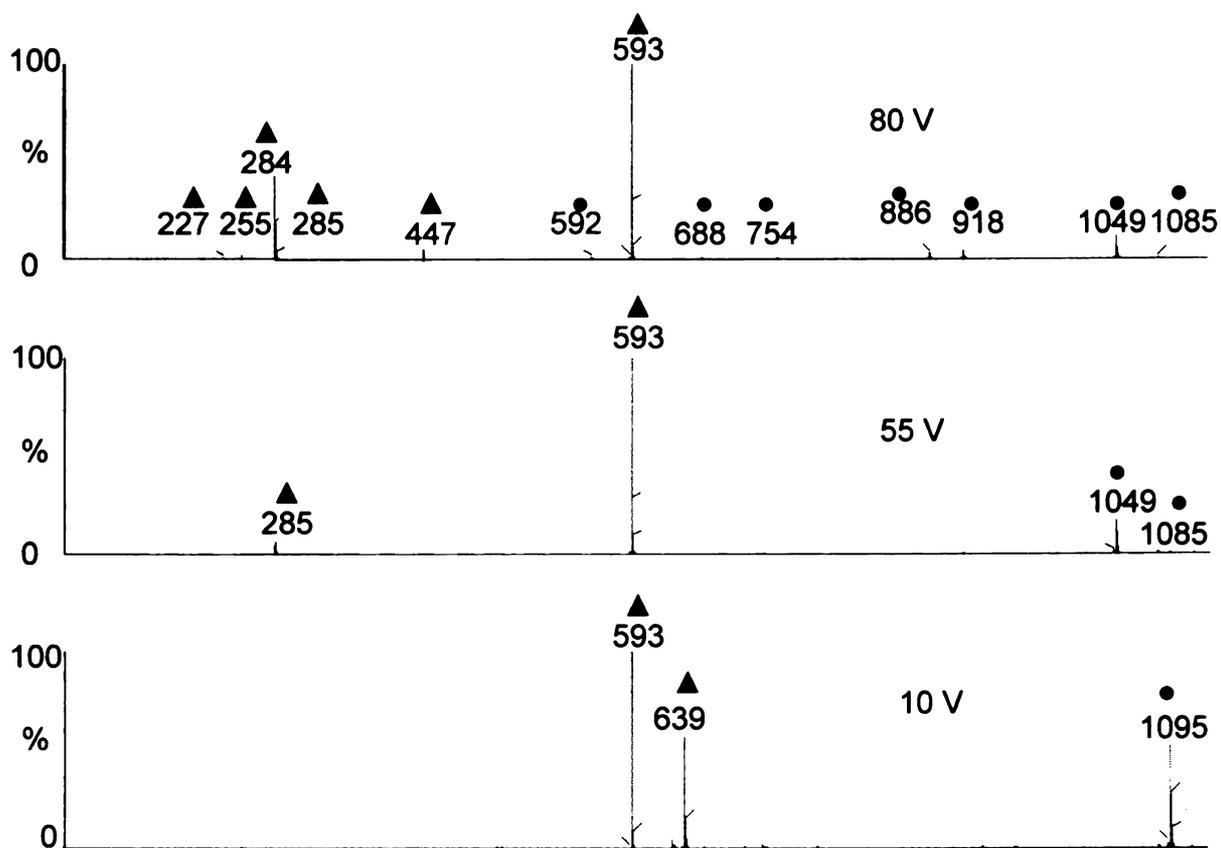
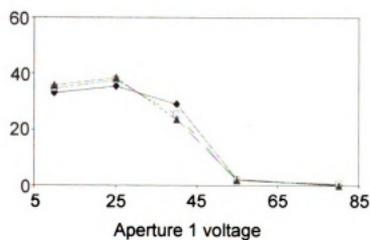
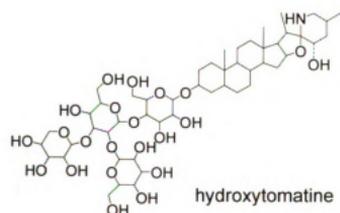
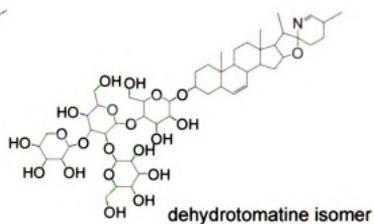
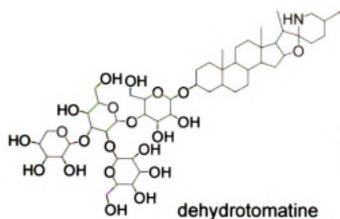
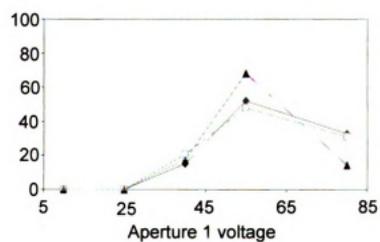


Figure 2-12. Negative mode mass spectra from an extract of IL-1-1 trichomes. Spectra were generated for three CID voltages (10, 55, and 80 V) by averaging spectra across the TIC peak eluting at 2.60 min. The peaks labeled with Δ were attributed to kaempferol-3-O-rutinoside and the peaks labeled with \bullet are attributed to hydroxytomatine.

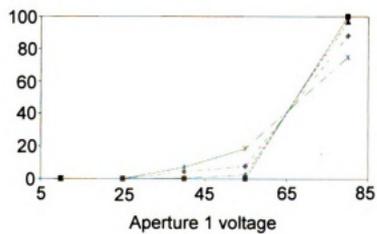
Figure 2-13 Breakdown curves showing energy dependence of various ion abundances for glycoalkaloid metabolites eluting at 2.46, 2.60 and 2.70 min extracted from trichomes of IL 1-1. Ion abundances are normalized to the total ion current for each individual ion as summed for all five collision potentials. Breakdown curves of A: formate adduct ion $[M+HCOO]^-$ for dehydrotomatine isomer (\blacklozenge), hydroxytomatine (\blacktriangle), dehydrotomatine (\square); B: monomeric $[M-H]^-$ ions for isomer of dehydrotomatine (\blacklozenge), hydroxytomatine (\blacktriangle), dehydrotomatine (\square); C: major fragment ions for dehydrotomatine isomer at m/z 898 (\blacklozenge) and at m/z 574 (\square); for dehydrotomatine at m/z 898 (\blacktriangle) and fragment ion at m/z 574 (\cdot); for hydroxytomatine at m/z 916 (\circ) and at m/z 592 (\bullet).



A

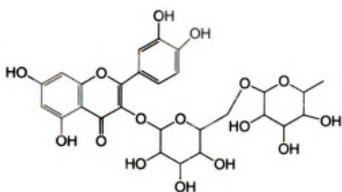


B

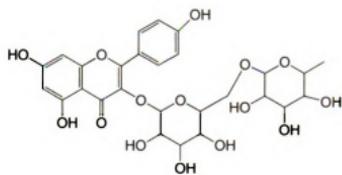


C

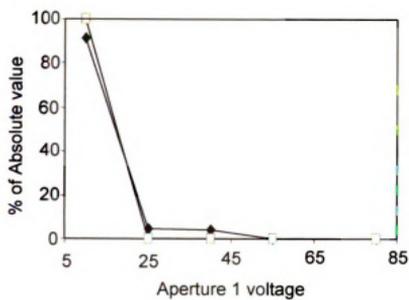
Figure 2-14 Breakdown curves showing energy dependence of various ion abundances for flavonoid metabolites eluting at 2.46 and 2.60 min extracted from trichomes of IL 1-1. Ion abundances are normalized to the total ion current for each individual ion as summed for all five collision potentials. Breakdown curves of A: formate adduct ion $[M+HCOO]^-$ for rutin (\square) and kaempferol-3-O-rutinoside (\blacklozenge), B: monomeric $[M-H]^-$ ions for rutin (\square), kaempferol-3-O-rutinoside (\blacklozenge), C: major fragment ions for rutin at m/z 301 (\square), for kaempferol-3-O-rutinoside at m/z 285 (\blacklozenge).



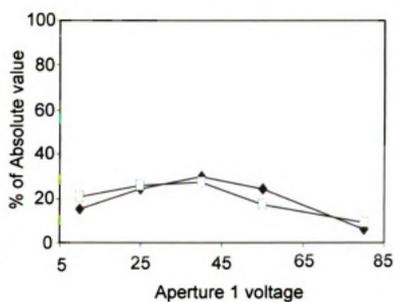
rutin



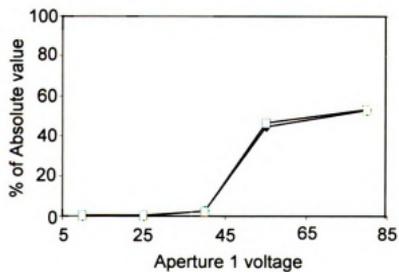
kaempferol-3-rutinoside



A



B



C

Quantitative profiling is an essential aspect of correlating metabolite levels with gene expression. The initial impetus for this work was driven by the limited dynamic range of TOF mass spectrometers, coupled with the idea that multiplexed CID could extend this range by attenuating strong molecular ion signals via inducing their fragmentation. In large-scale metabolite screening that can involve many thousands of LC/MS analyses, repeating each analysis using a more dilute sample is often not practical. In Figure 2-15, integrated ion currents for ions from the flavonoid glycoside rutin demonstrate how multiplexed CID extends the dynamic range of LC/TOF MS. At low collision potentials, the response of $[M-H]^-$ shows a steep slope but deviates from linearity at concentrations in the low μM range owing to formation of noncovalent dimer ions at higher rutin concentrations. As expected, a collision potential of 80 V yielded a less steep slope but not just from a proportional reduction in ion current. The high collision potentials induced dissociation of adducts and noncovalent complexes and yielded linear response up to the highest concentration (50 μM) tested.

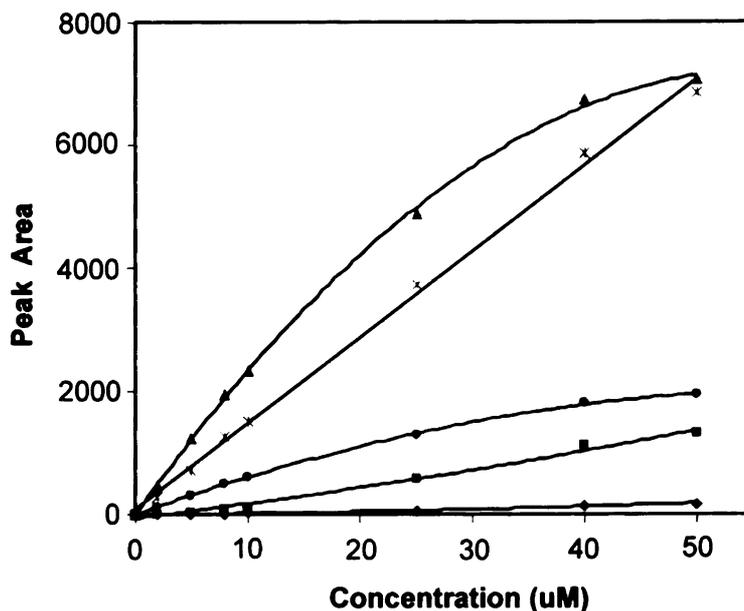


Figure 2-15. Concentration dependence of extracted ion chromatogram (XIC) peak areas for ions derived from rutin at different collision potentials using 10 μL injections of rutin standard solution and negative mode electrospray ionization. Curve A (\blacktriangle): the deprotonated molecular ion $[\text{M}-\text{H}]^-$ using the lowest CID potential (10 V); B (\blacklozenge): the deprotonated molecular ion $[\text{M}-\text{H}]^-$ using the highest CID potential (80 V); C (\bullet): fragment ion at m/z 300 from rutin using the highest CID potential (80 V); D (\blacksquare): the rutin noncovalent dimer ion $[\text{2M}-\text{H}]^-$ using the lowest CID potential (10 V); E (\blacklozenge): the rutin noncovalent dimer ion $[\text{2M}-\text{H}]^-$ using the highest CID potential (80 V). $N=3$ separate plants, error bars represent standard error of the mean.

Matrix suppression can be revealed and minimized through use of multiplexed CID. Hydrophobic components that coelute with specific analytes are notorious contributors to matrix suppression of ionization.¹⁵⁹ These components preferentially partition to surfaces of electrosprayed droplets and compete for ionization. In some cases, they form heterodimeric ions with other analytes, reducing molecular ion signals of the latter. Matrix suppression can be reduced through use of more aggressive CID conditions that drive dissociation of heterodimeric ions toward molecular ion forms. The behaviors of

three major isomers of sesquiterpene acid metabolites in extracts of the tomato relative *S. habrochaites* LA1777 are presented as examples of this effect. These metabolites elute in the midst of numerous acylsugar metabolites that are more than 10-fold more abundant, often yielding ions corresponding to noncovalent heterodimers between acylsugar and sesquiterpene acid. Figure 2-16 shows breakdown curves for three isomeric deprotonated sesquiterpene acids ($[M-H]^-$, m/z 233). The breakdown curve for the second isomer shows behavior typical of deprotonated molecules in the absence of coeluting interferences, with abundance decreasing as CID potential is increased. In contrast, abundances of corresponding ions for the other two isomers increase with CID voltage. This behavior is attributed to increased formation of $[M-H]^-$ at elevated collision energies via dissociation of heterodimeric ions. Substantial signals were observed for noncovalent heterodimer ions (not shown) formed by associations of sesquiterpene isomer 1 with a triacylsucrose, and sesquiterpene isomer 3 formed heterodimer ions with a tetraacylsucrose. These findings highlight the diagnoses of effects of coeluting interferences through examination of breakdown curves for molecular ion species.

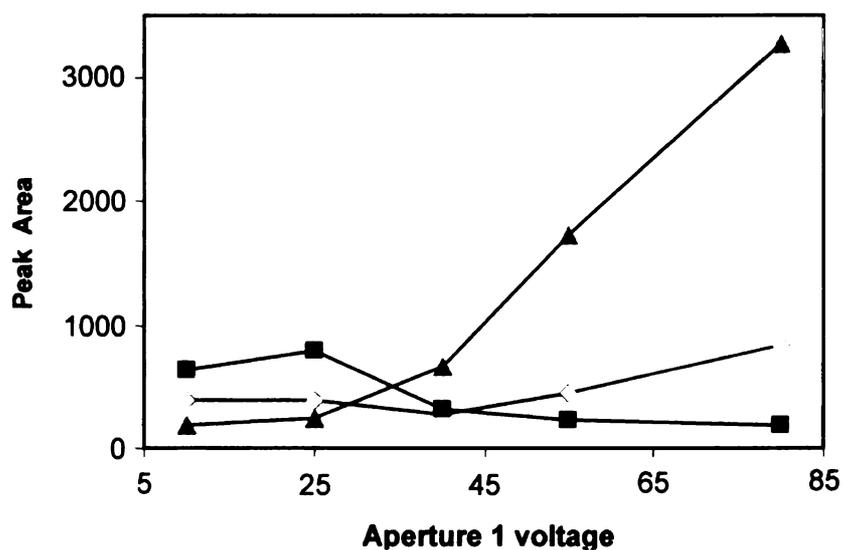
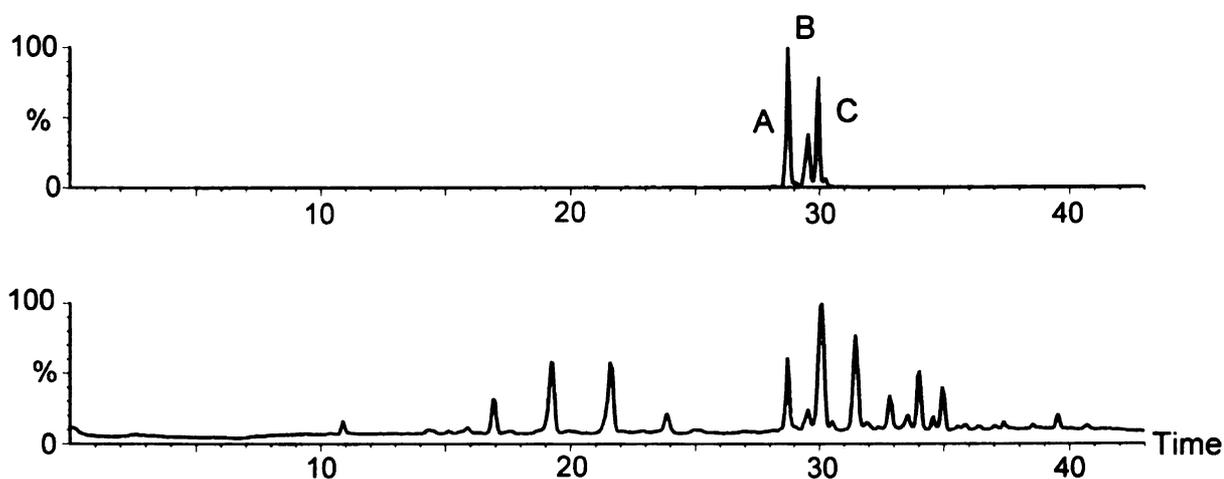


Figure 2-16 (Top) TIC of *S. habrochaites* LA1777 and XIC of m/z 233 using ESI negative mode with CID potential at 80 V. (Bottom) Breakdown curves derived from LC/negative ion ESI MS analyses of a leaf dip extract from *S. habrochaites* LA1777 corresponding to three isomers A (\diamond), B (\blacksquare) and C (\blacktriangle) of deprotonated sesquiterpene acid ($[M-H]^-$ ions at m/z 233) showing isomer-dependence of signal upon collision potential. Ion yields for isomers A and C exhibit an unusual positive slope, suggesting contributions toward these ion abundances from fragmentation of heterodimers that result from coelution with abundant acylsugar metabolite.

2.4 Conclusions

Acquisition of LC/TOF MS data using multiplexed nonselective CID with LC/TOF MS provides accurate ion masses and ion breakdown curves on the ultra-performance chromatographic time scale. The approach employs a suite of instrumental conditions that are consistent across analyses as needed for comparisons of large numbers of diverse samples in metabolomic discovery research. The approach enables deep coverage of the metabolome, yielding accurate measurements of molecular and fragment masses and ion behavior in a collision energy-resolved context. Multiplexed CID has particular value when limited analyte quantities or time constraints preclude repeat analyses.

One might ask whether more than two CID conditions are necessary, or whether there is an optimal number of CID conditions that ought to be employed. Multiplexing of CID conditions exacts minimal costs in terms of LC/MS method performance, providing the duty cycle is compatible with adequate sampling across chromatographic peaks. Our findings, using either fewer or more CID conditions, provided minimal additional information relative to five collision potentials. Use of five CID potentials provides a minimal definition of shapes of breakdown curves for all detected ions including appearance thresholds for fragment ions, and can direct assignment of those ions as molecular, noncovalent oligomer, or fragment species. The combination of this information offers the prospect for recognition of chemical classes of metabolites. These improvements in analytical throughput promise to accelerate discoveries in emerging areas including profiling of temporal and spatial dependence of metabolomes or analyses of other high complexity samples.

Multiplexing of collision conditions also provides a rich source of information for quantitative comparisons of metabolite profiles. Collision induced fragmentation can attenuate ion signals and extend dynamic range for abundant metabolites. As is the case with traditional GC/MS with electron ionization, metabolites can be quantified based on various ions including molecular and fragment species, and virtually guarantees that both kinds of ions will be detected for all metabolites. In addition, elevated collision potentials can drive dissociation of heterodimeric ions formed by coeluting metabolites, and decrease matrix suppression of molecular ion formation.

One of the most promising applications of LC/TOF MS with multiplexed CID lies in large-scale screening of metabolic phenotypes. During the past year, our laboratory has used a five-minute LC/TOF MS protocol that employs multiplexed CID to screen plant genotypes for associations with metabolic traits. This approach has generated analyses of 300 samples over a 30h period and yielded more than 1500 analytical signals after automated peak alignment and integration. The additional dimensionality of the MS data (CID energy) aids metabolite annotation even when chromatographic profiles overlap. More detailed results of this rapid method for discoveries of genes that influence metabolite levels will be described in chapter 5.

Chapter 3. Annotation of acylsugar metabolites in glandular trichomes from accessions of the genus *Solanum* based on LC/TOF MS coupled with multiplexed CID

3.1 Introduction

Secretory and glandular trichomes (SGTs) of wild tomato relatives within the genus *Solanum* exude a cocktail of secondary metabolites that play critical roles in protection against biotic and abiotic stresses including herbivory, pathogen infection, extreme temperature and excessive light.^{106,160} Many of these specialized metabolites were reported to have important potential value as commercial products such as drugs, natural pesticides, consumer products, or food additives.^{106,161-163} SGTs are specialized plant cells that are prolific chemical factories, but our understanding of the functions that confer such biochemical productivity remains limited. Identification of the genes involved in biosynthetic pathways and regulation of biosynthesis of these secondary metabolites will facilitate manipulation of metabolite production in plants, allowing for improved insect resistance in tomato and other crops. Furthermore, expression of these genes in other organisms such as bacteria and yeasts offers potential for using microorganisms as biological catalysts for specific chemical transformations. However, discoveries of genes involved in metabolite biosynthesis require profiling and identification of metabolites in an assortment of different cell types in numerous plant genotypes. Only then can gene expression be correlated with plant chemistry.

Assignment of structures to SGT metabolites has been slow to develop. Much of our knowledge has been derived for a limited range of plants, primarily *S. pennellii* LA0716^{108,132,164} and *S. habrochaites* LA1777,¹¹⁵ and has been focused on a limited number of acylsugar metabolites that were purified from plant extracts. Most knowledge of metabolite structures was obtained from NMR^{118,165-167} or GC/MS¹⁶⁸⁻¹⁷⁰ analysis. However, NMR requires previous purification of each compound and is not practical for

large numbers of different plants. Many metabolites lack volatility needed for GC/MS analysis. As a result, such compounds must be derivatized before analysis, and these procedures are often time-consuming, and structures are inferred from chemical derivatives that differ from the original metabolite. Such barriers might be overcome in part through use of LC/MS, which avoids derivatization steps, and LC/MS approaches are finding increasing use in global characterization of metabolites in mixtures.^{100,103,171-}

¹⁷⁴ Mass spectrometry is most powerful, when used for metabolite identification, when accurate molecular mass measurements allow assignment of an elemental formula to each metabolite. In recent years, coupling of LC separations to mass analyzers such as TOF^{103,175} or FTICR¹⁷⁶⁻¹⁷⁸ have facilitated assignments of formulas to previously unidentified metabolites. Yet more information can be generated by selecting metabolite molecular ion species and generating MS/MS fragmentation patterns that aid structure assignments.^{103,172,174,179}

While coupling of HPLC separation with tandem mass spectrometry provides valuable information for profiling and identification of secondary metabolites, it suffers from important limitations. The standard approach employs a data dependent strategy,¹¹⁹ in which the data system performs real-time selection of molecular ions followed by acquisition of MS/MS product ion spectra for the selected ions. The nature of these analyses dictates that the analysis is different for each analysis, and this complicates quantitative comparisons of metabolite levels in different sample extracts. Furthermore, many of the mass analyzers are used for such analyses offer limited mass measurement accuracy for fragment ions. The approach described below for metabolite identification has relied on methods that fragment all ions in a data-independent fashion, yielding

accurate measurements of fragment ion masses. This multiplexed CID approach, described in Chapter 2, provides important information about metabolite structures and quantities with high throughput needed for screening large numbers of samples.

The multiplexed CID method provides for quasisimultaneous generation of molecular and fragment ions by switching the CID potential from low to high in this approach. Accurate mass measurements on molecular and fragment ions in a single analysis simplify spectrum interpretation and support elemental formula assignment. Use of multiple collision conditions in a single analysis yields information about the collision energy-dependence of all ion abundances, providing additional evidence for assignment of ions to chemical classes of metabolites.

To test the utility of LC/multiplexed CID MS for metabolite profiling and identification, this approach was first applied for identifying metabolites extracted from two wild tomato relatives, *Solanum pennellii* LA0716 and *S. habrochaites* LA1777. A limited number of metabolites from these accessions had been characterized in earlier studies. Demonstration of the value of LC/multiplexed CID MS was then achieved through analyses of a different accession for each species, *S. pennellii* LA1522 and *S. habrochaites* LA1353. Acylsugars, alkaloids, phenolic acids, flavonoids and terpenoids are also annotated for these and additional *Solanum* accessions. Results from multiplexed CID are compared to results from product ion MS/MS spectra for many metabolites in order to support interpretation of fragment ion formation.

3.2 Experimental Section

3.2.1 Materials. Plants were grown in a greenhouse in Michigan State University and the seeds were obtained from C. M. Rick Tomato Genomics Resource Center at the University of California-Davis.

3.2.2 Chemicals. Acetonitrile (HPLC grade), 2-propanol (HPLC grade), methanol (HPLC grade), and formic acid 88% were obtained from VWR Scientific and Fisher Scientific.

3.2.3 Extraction Method. To extract metabolites from *Solanum* trichomes, ~100 mg of individual leaflets were dipped in 2 ml isopropanol: CH₃CN: H₂O (3:3:2 v/v/v) for 1 min. A 1.0 ml aliquot of each extract was transferred to a 1.5 ml polypropylene microcentrifuge tube, and the tubes were centrifuged (5000 rpm) for 2 min at 25°C, and the supernatants were analyzed without further processing. Extracts were stored at -20°C.

3.2.4 High-Performance Liquid Chromatography. A Shimadzu (Columbia, MD) LC-20AD HPLC ternary pump was used. The separation was performed on a Thermo 1 × 150 mm BetaBasic C₁₈ column (100 × 1 mm, 3 μm). Gradient elution was executed based on 0.15% aqueous formic acid (solvent A) and methanol (solvent B) as follows: initial conditions 5% B; linear gradient to 50% B (5 min), 95% B (33 min), and 100% B (35 min); hold at 100% B until 38 min, and return to 5% B (43 min). The flow rate was 0.1 ml·min⁻¹, and the injection volume was 10 μl. The column was maintained at 30 °C in a column oven.

3.2.5 Mass Spectrometry. Analyses were performed using a model LCT Premier mass spectrometer (Waters, Milford, MA). Operating parameters were set as follows: capillary voltage: 3.2 kV in both positive and negative mode; desolvation gas flow: 300 L·h⁻¹;

desolvation gas temperature: 200 °C; source temperature: 90 °C; cone gas flow: 20 L·h⁻¹.

The instrument was operated in both negative and positive ion electrospray modes using separate injections of extract for each ionization mode. Analyses were performed using rapid switching of Aperture 1 voltage in the ion transit region of the mass spectrometer, proving quasisimultaneous generation of spectra under fragmenting and nonfragmenting conditions. Mass spectra were acquired over m/z 50 to 1500 with a scan time of 0.4 s for each function. Spectra were acquired in centroid format using 'dynamic range enhancement' (DRE) enabled, which acquires separate mass spectra by switching a beam attenuator on and off, followed by post-acquisition stitching together of spectra. The instrument mass scale was calibrated daily in both ionization modes over m/z 50-1500 using 0.1% aqueous phosphoric acid as reference. Most analyses were performed using the V-mode ion pathway in the mass analyzer, which generated a resolving power of 5000 (FWHM). For some analyses requiring more exact mass measurements, 0.1% H₃PO₄ was infused through the reference sprayer of the lock-spray source as a reference lock mass, and the lockspray attachment switched between sample and reference streams.

3.2.6 Characterization of Metabolites. Measured ion masses were used to generate possible elemental formulas using the Masslynx elemental composition calculator (Waters), using a mass tolerance of 10 ppm. In addition, isotopic pattern filter or the mass defect can assist to determine the natural of elemental composition.

3.2.7 Nomenclature. To simplify the reporting of acylsugars, they are described using a single letter corresponding to the base sugar (G = glucose; S = sucrose) followed by the designation of number of fatty acyl groups and total number of acyl carbon atoms. For

example, a glucose triester acylated with three C₅ fatty acids would be designated as G 3:15.

3.3 Results and Discussion

3.3.1 Structure annotation for glucose trimesters in *S. pennellii* LA0716

In earlier studies, acylsugars were identified as major components in SGTs from several species in the plant genus *Solanum*.^{115, 164} Known acylsugars are based upon either glucose or sucrose as a core, and most of these sugars are esterified with 3-4 fatty acyl groups consisting of either straight or branched chains ranging in carbon numbers from 2 to 12.^{107,109} Acylglucoses are the major metabolites exudates from glandular trichomes of the wild tomato relative *S. pennellii*^{107-111, 164, 180}. Burke and coworkers¹⁶⁴ discovered nine 2,3,4-tri-O-acylglucose esters in *S. pennellii* using magnetic NMR, GC/MS and desorption chemical ionization (DCI) MS (Figure 3-1). Only three acylsugars were purified as needed for complete structure elucidation by NMR owing to the limited separation of metabolites by preparative chromatography. In all cases, the positions of fatty acyl groups were assigned at 2-, 3-, and 4- positions using NMR. This suite of acylglucoses consisted of variants containing *iso*-C4, *iso*- and *anteiso*-C5, and straight chain and *iso*-C10 fatty acid groups, with a total of nine different combinations of substitution.

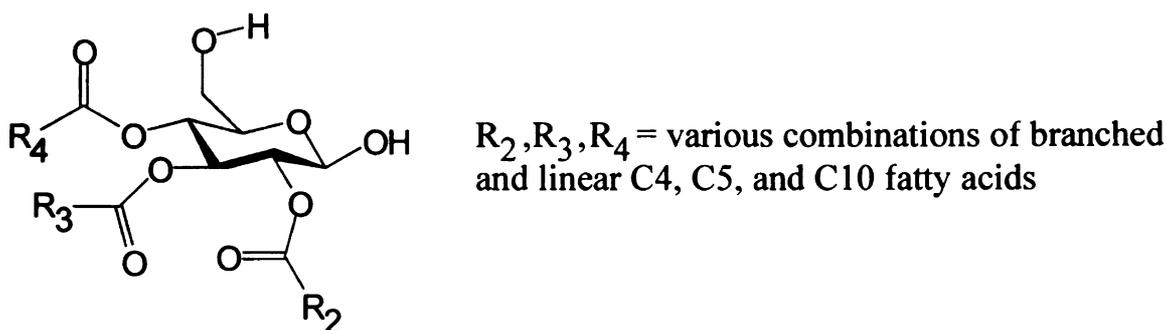


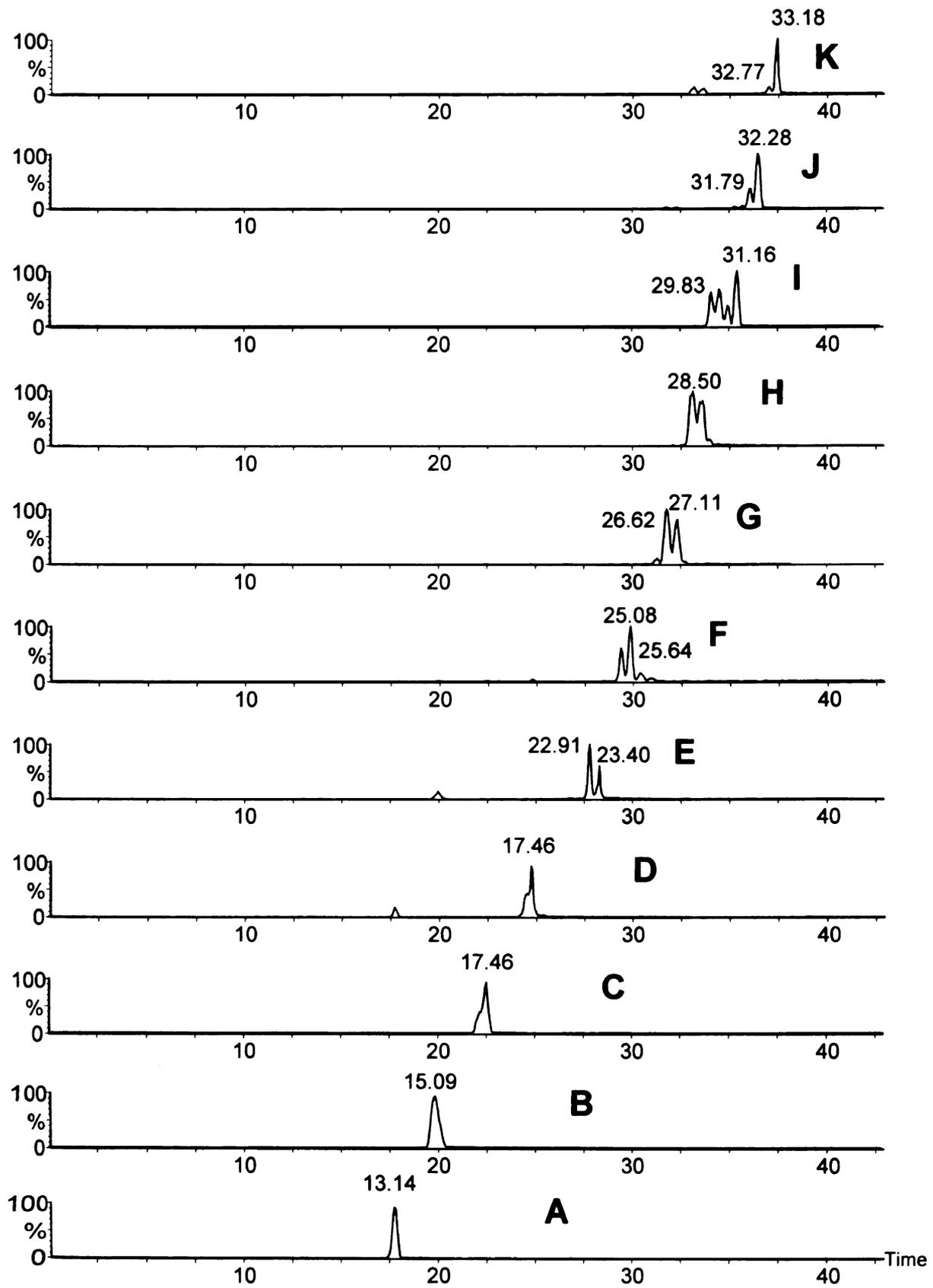
Figure 3-1. Generalized structure of triacylglucose metabolites from *S. pennellii* LA0716 as identified by Burke *et al.*¹⁶⁴

To simplify discussion of acylsugars, we propose the following nomenclature to describe them. Acylsugars are assigned a letter (G or S) corresponding to glucose or sucrose core, followed by two numbers. The first number indicates the number of fatty acyl groups attached to the core, and the second number describes the total number of carbon atoms in the fatty acyl chains. In some cases, the numbers of carbons in the individual fatty acids are included in parentheses. The report of Burke's laboratory described glucose triesters of four molecular masses that we abbreviate as follows: G 3:12 (4,4,4), G 3:13 (4,4,5), G 3:18 (4,4,10) and G 3:19 (4,5,10).

Multiplexed CID coupled with LC/TOF MS is applied for structure annotation of acylsugars in *S. pennellii* LA0716 to compare LC/MS results with Burke's original findings. Acylsugars lack readily ionized basic or acidic functional groups. As a consequence, various adducts with cations ($[M+Na]^+$, $[M+NH_4]^+$) or anions ($[M+Cl]^-$, $[M+HCOO]^-$), are formed in ESI under gentle CID conditions. The dominant peaks in the total ion chromatogram of an extract of *S. pennellii* leaf tissue displayed $[M+formate]^-$ ions for a series of glucose esters ranging from G 3:13 to G 3:23. Extracted ion chromatograms for these m/z values are presented in Figure 3-2. The acylglucoses

were resolved at different retention times using a reversed phase C₁₈ column, with retention time increasing with the number of fatty acyl carbon atoms. Findings from the current work document 23 chromatographic peaks corresponding to formate adduct of triacylglucoses, and these ranges across 11 different molecular masses. This LC/MS approach has more than doubled the known diversity of acylglucoses in *S. pennellii* LA0716.

Figure 3-2. Extracted ion chromatograms (XICs) generated from an extract of leaf tissue from *S. pennellii* LA0716 of ions of m/z values corresponding to $[M+HCOO]^-$ for glucose triesters with total fatty acid carbon atoms ranging from 12 to 22. A: G 3:12 (m/z 435). B: G 3:13 (m/z 449). C: G 3:14 (m/z 463). D: G 3:15 (m/z 477). E: G 3:16 (m/z 491). F: G 3:17 (m/z 505). G: G 3:18 (m/z 519). H: G 3:19 (m/z 533). I: G 3:20 (m/z 547). J: G 3:21 (m/z 561). K: G 3:22 (m/z 575). A-K (from bottom to top).



G

multi

Assi

oval

gene

Vu

of

and

stun

sever

zior

Stun

Ar

ins

ad C

sup

Stun

Stun

Stun

Stun

Stun

Stun

Stun

Stun

Stun

Glucose triester G 3:18(4,4,10) from *S. pennellii* LA0716 was used to illustrate how multiplexed CID spectra were used for metabolite annotation and identification. Assignment of ions as molecular ions (both adduct and deprotonated forms), non-covalent dimers, and fragment ions is aided by evaluating slopes of breakdown curves generated under multiplexed CID (Figure 2-6. G 3:18). At the lowest CID potential (10 V), $[M+HCOO]^-$ at m/z 519.29 is the most abundant ion observed, with minor amounts of $[M-H]^-$ at m/z 473.28 and non-covalent dimer ions corresponding to $[2M+HCOO]^-$ and $[2M-H]^-$ at m/z 993.59 and 947.59 respectively. As CID potential was increased, the abundance of formate adduct ions decreased, and relative abundances of $[M-H]^-$ and several fragment ions increased. At the highest CID potential of 80 V, the fatty acid anions corresponding to deprotonated C10 (m/z 171.14) and C4 (m/z 87.04) fatty acids dominated the spectrum. Between these extremes of CID potential (25 V), fragment ions at m/z 347.13, 301.13 and 213.08 reached maximum relative abundance. These fragment ions were assigned as products formed by elimination of neutral C10, C10 plus formic, and C10 plus formic plus C4 acids, respectively, from $[M+HCOO]^-$ of G 3:18 (Figure 3-4 top). It is worthy of note that the formate group remains attached after fatty acid elimination in the fragment at m/z 347. A fragment ion at m/z 143.04 was assigned as corresponding to losses of two acyl groups as fatty acids and elimination of the third as a fatty ketene. This fragment (m/z 143.04) was common to all observed triacylglucoses, and provides a characteristic marker for recognition of triacylglucoses.

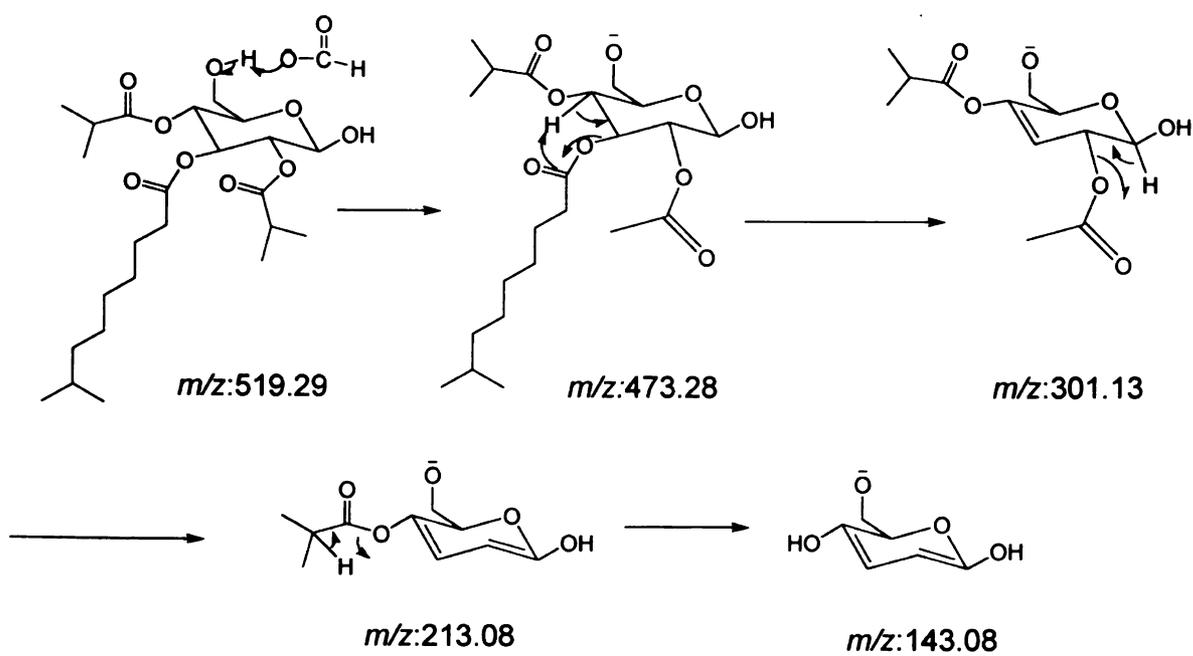
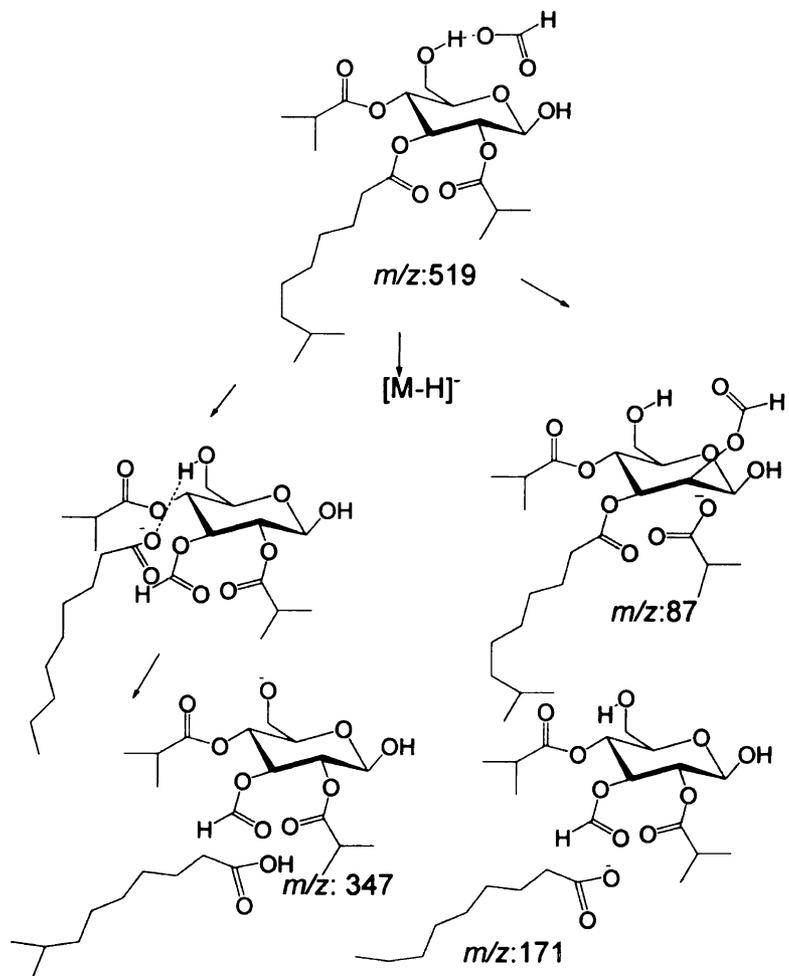
High accuracy mass measurements of all observed ions accelerate spectrum interpretation and help confirm associations of fragments with molecular ions. For example, the fragment ion at m/z 143.04 (dehydrated glucose fragment $C_6H_7O_4^-$) is

readily distinguished by the TOF mass analyzer from the C8 fatty acid anion ($C_8H_{15}O_2^-$; theoretical m/z 143.11). Therefore, the masses of fatty acid constituents of each acylsugar can be determined by multiplexed CID in negative mode based on accurate mass measurements of deprotonated fatty acid fragment ions.

In order to confirm assignments of fragments from multiplexed CID, product ion MS/MS spectra were generated for the formate adduct (m/z 519) of triacylglycerol G 3:18 using a QToF mass spectrometer (Figure 3-5). Product ions corresponding to neutral losses of C10, C10 plus formic acid, and C10, formic acid plus C4 fatty acids were observed at m/z 347.14, m/z 301.14, and m/z 213.09 respectively, remnants of the glucose core appeared at m/z 143.04 and m/z 125.03, and C10 and C4 fatty anions were observed at m/z 171.14 and m/z 87.04. All of these fragment ions were similar as observed in multiplexed CID spectra using 25 V as described above. These findings demonstrate that multiplexed CID experiments generate information in fragment ion masses similar to those generated during MS/MS experiments.

Formation of fatty acid anion fragments from formate adducts of triacylglycerols is consistent with formation of the fatty acid anion fragments ($RCOO^-$) occurring via nucleophilic displacement of the fatty acid anion by formate at the carbon where the fatty acid ester is attached on the carbohydrate backbone.^{181,182} Additional evidence in support of this mechanism comes from the appearance of $[M+formate-RCOOH]^-$ at m/z 347 in the case of loss of the neutral C10 fatty acid. Additional fragment ions corresponding to losses of neutral fatty acids from $[M-H-RCOOH]^-$ or $[M+HCOO^-RCOOH]^-$ are consistent with *cis* elimination reactions.^{181,182} A proposed fragmentation for the formate adduct of G 3:18 is presented in Figure 3-3.

Figure 3-3. Proposed fragmentation pathway of the formate adduct of triacylglycerol S3:18 extracted from *S. pennellii* LA0716 (RT = 21.11 min) using ESI negative mode with CID voltage 25 V. Specific substitution positions of individual fatty acids remain uncertain.



The structure assigned for triacylglycerol G 3:18 from negative ion mass spectra was further confirmed by generating multiplexed CID mass spectra in positive mode. The peaks observed, using a CID potential of 10 V, at m/z 492.23 and m/z 497.19 correspond to $[M+NH_4]^+$ and $[M+Na]^+$. Figure 3-4 displays ESI spectra from a triacylglycerol G 3:18 (RT = 27.11 min) under CID potential 25 V in negative (top) and 40 V in positive (bottom) mode. In both positive and negative ion modes, fragment ions are observed that correspond to neutral losses of fatty acid groups and to fragments characteristic of the fatty acid moieties. In positive mode, acylium ions corresponding to C4 and C10 fatty acids are observed at m/z 71 and 155 respectively using collision potential of 40 V, and their relative abundance increased when the potential was raised to 80 V.

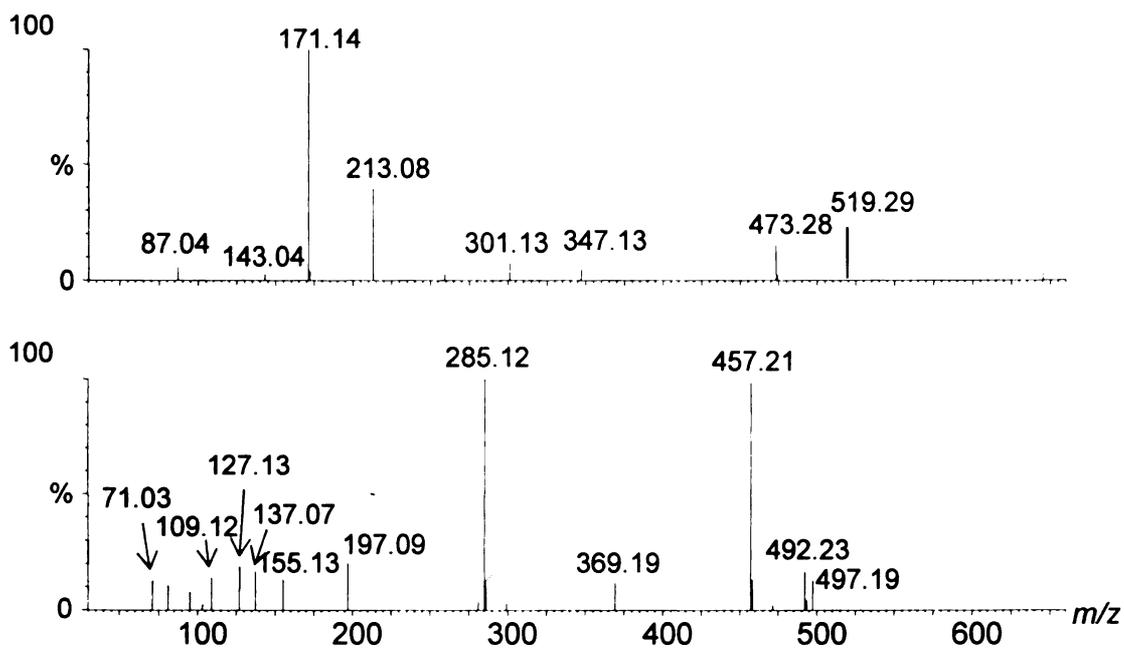


Figure 3-4. Electrospray ionization mass spectra of G 3:18 (RT = 27.11 min) extracted from *S. pennellii* LA0716 using an Aperture 1 CID potential of 25 V in negative (top) and 40 V in positive (bottom) mode.

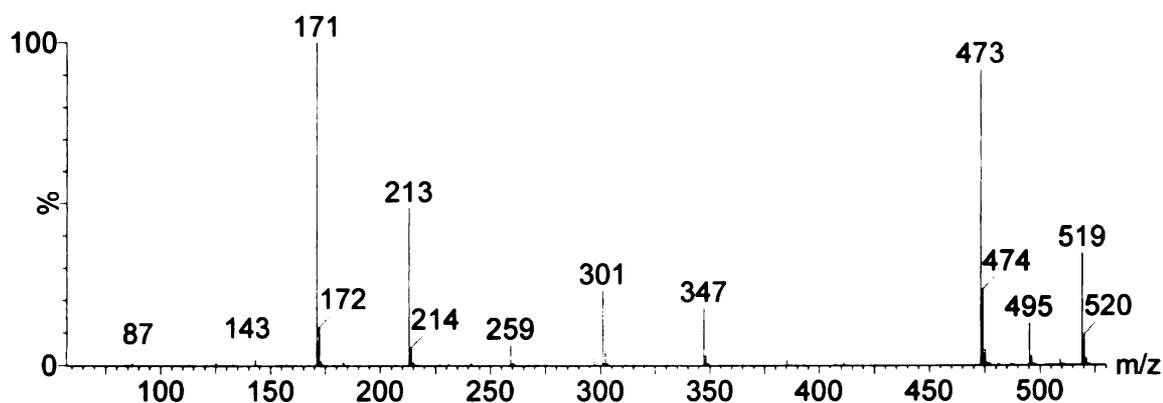


Figure 3-5 Product ion MS/MS spectrum of m/z 519 ($[M+HCOO]^-$) from triacylglycerol G 3:18 from *S. pennellii* LA0716 using collision energy 30 eV.

Most fragment ions were derived from $[M+NH_4]^+$ instead of $[M+Na]^+$, which was demonstrated by generating product ion MS/MS spectra of $[M+NH_4]^+$ generated using the QTOF instrument. The ammonium adduct of acylglycerol G 3:18 undergoes facile loss of NH_3 and H_2O to form $[M+H-H_2O]^+$ (m/z 457.21) at collision potential of 10 V. Product ion spectra for m/z 492 $[M+NH_4]^+$ and m/z 457 $[M+H-H_2O]^+$ were performed using the QTOF (Figure 3-6). The $[M+H-H_2O]^+$ fragment was generated in source without changing ion source parameters, and a product ion spectrum was generated. The major product ions using 30 V collision energy in the QTOF were the same as those generated using multiplexed CID at 40 V. All of the major product ions observed from all three precursors ($[M+NH_4]^+$, and $[M+H-H_2O]^+$) corresponded to losses of the various neutral fatty acids. None of the product ions could be attributed to cross-ring fragmentation reactions that would have been useful for establishing positions of substitution of the fatty acid groups.

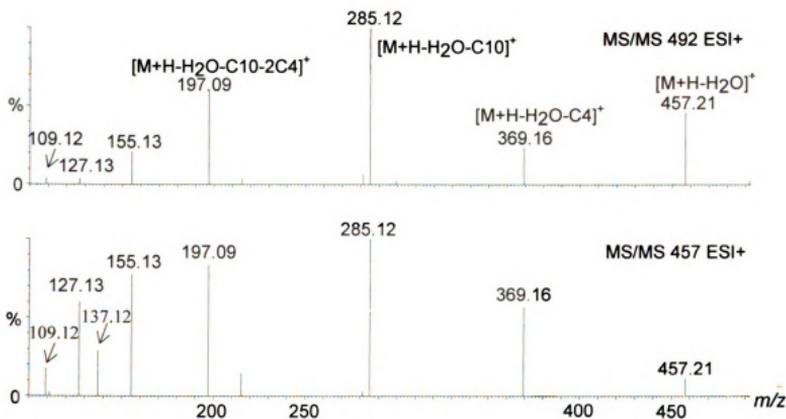
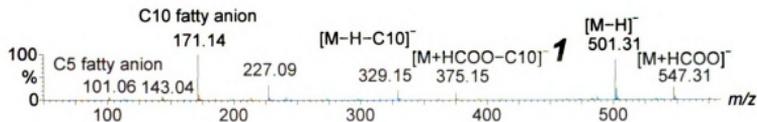
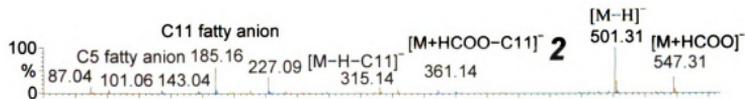
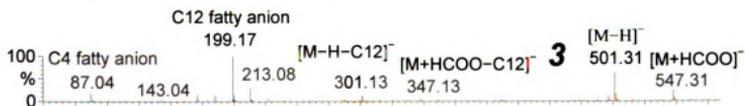
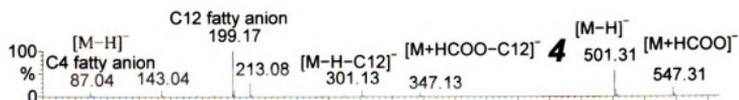
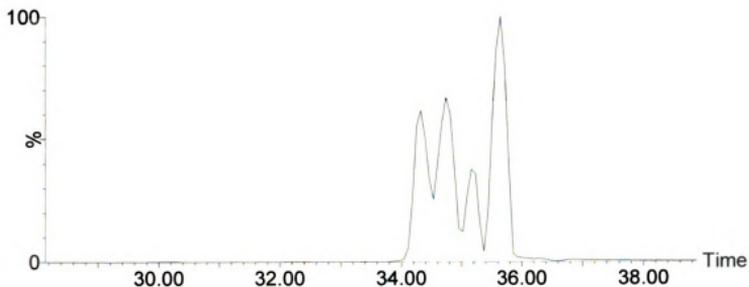


Figure 3-6 Top: Product ion MS/MS spectrum of m/z 492 ($[M+NH_4]^+$) from triacylglycerol G 3:18 from *S. pennellii* LA0716 (collision energy:30 eV). Bottom: product ion MS/MS spectrum for m/z 457 ($[M+H-H_2O]^+$) for G 3:18; m/z 457 was generated in source (Collision energy:35 eV).

Isomeric acylsugar metabolites can be distinguished based on fragment ion masses produced using more aggressive CID voltage. Triacylglycerol G 3:20 from *S. pennellii* LA0716 provides a good example, yielding at least four resolved chromatographic peaks yielding abundant formate adduct ions (Figure 3-7). Fatty acyl constituents could be determined from the anions of fatty acids cleaved from acylglucosides at CID potentials of 25 V and greater. The extracted ion chromatogram for the formate adduct of G 3:20 revealed four major isomers. The first eluting peak (peak 1) shows fragments for C5 and C10 fatty anions, peak 2 shows C4, C5, and C11 fatty anions, and peaks 3 and 4 show C4 and C12 fatty anion peaks. In order to match the observed masses, peak 1 isomer must

contain two C5 groups, and peaks 3 and 4 must contain two C4 groups. Given the limited number of observed peaks, we concluded that variation in substitution patterns for specific fatty acids was limited. For peaks 3 and 4, additional GC/MS evidence suggested the presence of both linear and branched C12 fatty acid groups. Therefore, we attribute the two isomeric forms (C4, C4, C12) to arise from different C12 fatty acid groups, and not to the presence of multiple sites of C12 substitution. All the ions were labeled in the mass spectrum of each triacylglycerol isomer in Figure 3-5.

Figure 3-7. (Top) Extracted ion chromatogram of m/z 547 (G 3:20) from LC/MS analysis of an extract of *S. pennellii* LA0716 under CID potential 10 V in ESI negative mode. Chromatographic peaks corresponding to four isomers detected as m/z 547 were resolved. (Bottom) ESI mass spectra obtained from these four isomers at 40 V in negative mode.



All detectable acylglucoses from *S. pennellii* LA0716 were annotated from multiplexed CID mass spectra and product ion MS/MS scans. Major fragment ions and fatty acyl constituents determined using multiplexed CID for each detectable acylglucose of LA0716 (*S. pennellii*) are presented in Table 3-1. Accurate masses for formate adducts for each acylglucose were determined by flow injection analysis, and are listed in Table 3-2. One original finding of this work is discovery of C8, C9, C11, and C12 fatty ester groups on triacylglucoses from this plant genotype. The C8 and C9 fatty acids were not observed in extracts from several other accessions within the genus *Solanum* (*S. habrochaites* LA1777 and LA1353), and we conclude that their occurrence in accession LA0716 may help guide future discoveries of genes involved in branched fatty acid elongation.

Table. 3-1. Fragments and fatty acid constituents for detected glucose triesters from *S. pennellii* LA0716 in negative and positive mode (CID potential, 25 V)

Acylsugar metabolite annotation	Fatty acid ester groups	[M+formate] ⁻ (m/z)	[M-H] ⁻ (m/z)	Fragment ions (m/z)
G 3:12	C4, C4, C4	435	389	347, 301, 213, 143, 125, 87
G 3:13	C4, C4, C5	449	403	361, 315, 227, 143, 125, 101, 87
G 3:14	C4, C5, C5	463	417	375, 329, 227, 143, 125, 101, 87
G 3:15	C5, C5, C5	477	431	375, 329, 227, 143, 125, 101
G 3:16	C4, C4, C8	491	445	347, 301, 213, 125, 143, 87
G 3:17	C4, C5, C8 C4, C4, C9	505	459	361, 315, 227, 125, 143, 101, 87 347, 301, 213, 125, 143, 157, 87
G 3:18	C4, C4, C10	519	473	347, 301, 213, 143, 125, 171, 87
G 3:19	C4, C5, C10	533	487	361, 315, 227, 143, 171, 101, 87
G 3:20	C5, C5, C10 C4, C4, C11 C4, C5, C12	547	501	347, 301, 213, 143, 125, 171, 101 361, 315, 213, 143, 125, 185, 101, 87 375, 329, 227, 143, 125, 199, 87
G 3:21	C4, C5, C12	561	515	361, 315, 227, 143, 199, 101, 87
G 3:22	C5, C5, C12	575	529	375, 329, 227, 143, 199, 101, 101

Acylsugar metabolite annotation	Fatty acid constituents	[M+NH ₄] ⁺ (m/z)	[M+Na] ⁺ (m/z)	Fragments (m/z)
G 3:12	C4, C4, C4	408	413	373, 285, 197, 127, 109
G 3:13	C4, C4, C5	422	427	387, 299, 197, 127, 109
G 3:14	C4, C5, C5	436	441	401, 299, 211, 127
G 3:15	C5, C5, C5	450	455	415, 313, 211, 127
G 3:16	C4, C4, C8	464	469	429, 285, 197, 127
G 3:17	C4, C5, C8	478	483	443, 299, 197, 127
G 3:18	C4, C4, C10	492	497	457, 285, 197, 127
G 3:19	C4, C5, C10	506	511	471, 299, 211, 127
G 3:20	C5, C5, C10 C4, C4, C12 C4, C5, C11	520	525	485, 313, 211, 127 485, 285, 211, 127 485, 299, 211, 109
G 3:21	C4, C5, C12	534	539	499, 299, 211, 127
G 3:22	C5, C5, C12	548	553	513, 313, 211, 127

Table 3-2. Exact mass measurements of $[M+HCOO]^-$ ions of detected glucose triesters from *S. pennellii* LA0716 detected using ESI negative mode with CID potential (10 V).

Acyl sugar metabolite annotation	Formula of formate adduct	Measured (m/z)	Calculated (m/z)	ppm (error)
G 3:12	C ₁₉ H ₃₁ O ₁₁	435.1866	435.1865	-0.2
G 3:13	C ₂₀ H ₃₃ O ₁₁	449.2022	449.2030	-0.2
G 3:14	C ₂₁ H ₃₅ O ₁₁	463.2179	463.2169	-2.2
G 3:15	C ₂₂ H ₃₇ O ₁₁	477.2336	477.2327	-1.9
G 3:16	C ₂₃ H ₃₉ O ₁₁	491.2492	491.2515	4.7
G 3:17	C ₂₄ H ₄₁ O ₁₁	505.2615	505.2649	-6.7
G 3:18	C ₂₅ H ₄₃ O ₁₁	519.2805	519.2830	4.8
G 3:19	C ₂₆ H ₄₅ O ₁₁	533.2962	533.2964	0.4
G 3:20	C ₂₇ H ₄₇ O ₁₁	547.3118	547.3122	0.7
G 3:21	C ₂₈ H ₄₉ O ₁₁	561.3275	561.3262	-2.3
G 3:22	C ₂₉ H ₅₁ O ₁₁	575.3431	575.3434	0.5

Despite the successes in annotating acylsugar metabolites based on the number of acyl chains and the number of carbons in the fatty acid groups, the CID spectra did not distinguish branched from straight chain of fatty acyl groups or their positions of attachment. This shortcoming arises because fragment ions that would distinguish fatty acid isomers were not observed. GC/MS was performed to distinguish isomeric fatty acids after transesterification from all the acylsugars in these wild type species. The GC/MS results for the composite composition of fatty acyl groups in leaf extracts of *S. pennellii* LA0716 are summarized in Figure 3-8. All fatty acid chain lengths observed in LC/MS results were confirmed by GC/MS data, and isomeric fatty acids were distinguished by their retention times and electron ionization mass spectra. The proposed structures for branched or straight chain fatty acyl groups were based on fatty acid esters identified from GC/MS results by comparing retention times and mass spectra with authentic standards or the NIST 05 mass spectrum library and support from literatures.¹¹²

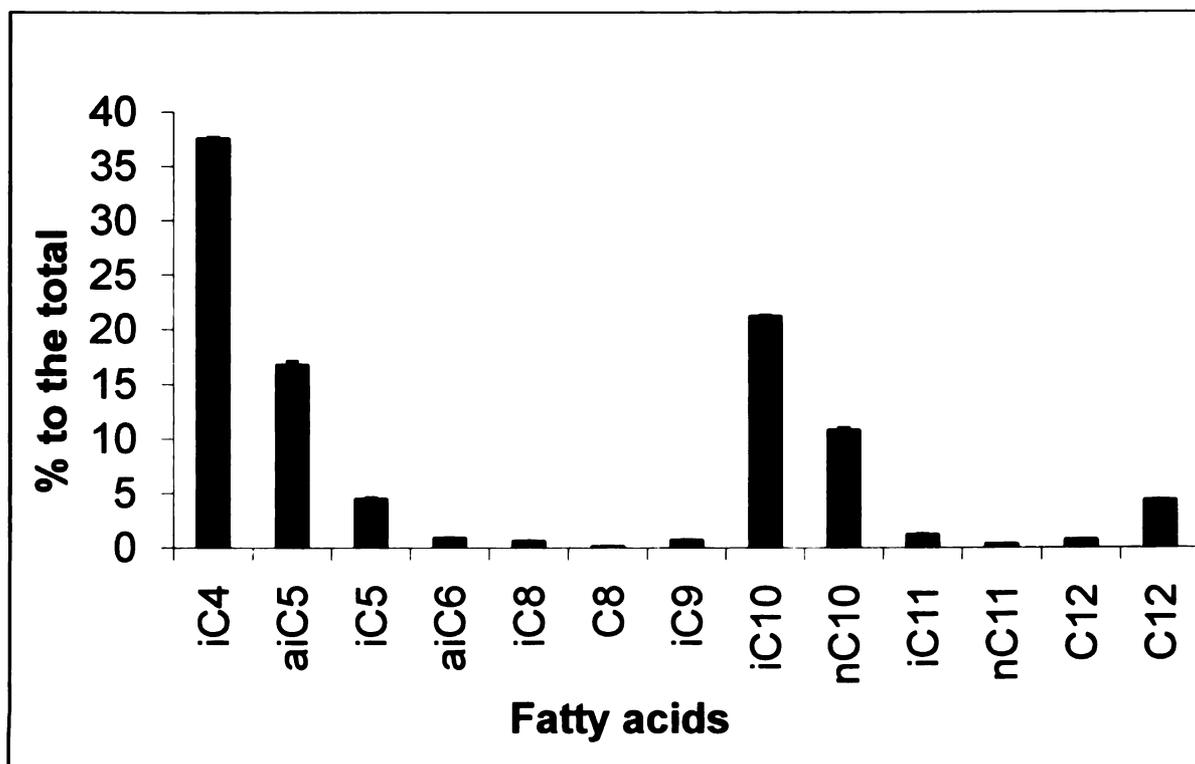


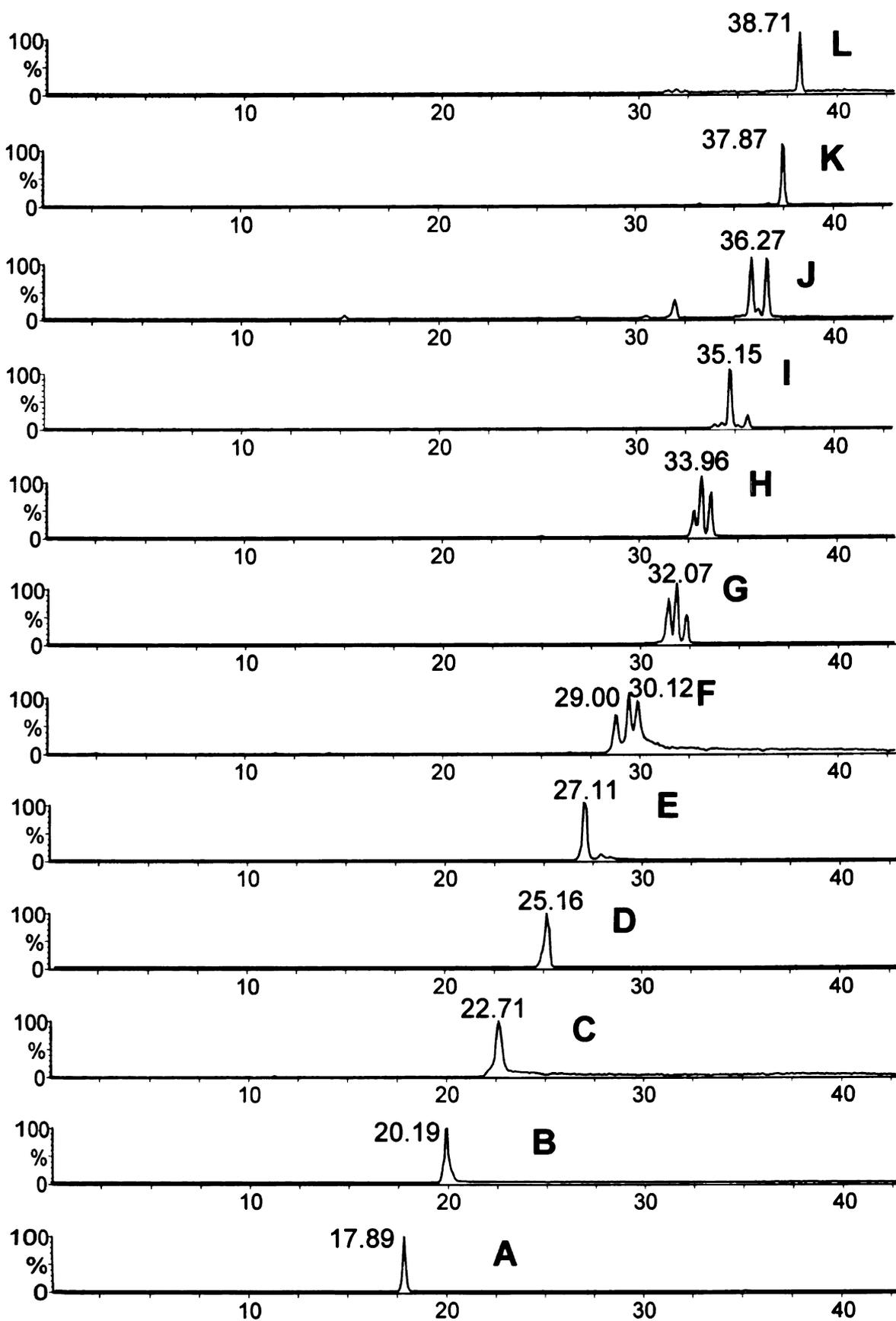
Figure 3-8 Fatty acyl groups from total acylsugars in an extract of *S. pennellii* LA0716 were transesterified to form fatty acid ethyl esters, and analyzed using GC/MS. Data for each fatty acid were calculated as a percentage of the total ion current chromatogram peak areas of fatty acids detected through GC/MS. The abbreviations “i” and “ai” refer to *iso*- and *anteiso*- branched isomers. Error bars indicate standard error with $n = 3$.

3.3.2. Structure annotation for glucose triesters and sucrose triesters in *S. pennellii* LA1522.

After the performance of multiplexed CID was verified by detecting known triacylglycerols and previously undetected acylsugars in extracts of *S. pennellii* LA0716, this approach was applied to profile acylsugar metabolites in a different accession of the same species, LA1522. No previous reports have documented acylsugars in LA1522 accession. The LC/MS results revealed triacylglycerols ranging from G 3:12 to G 3:24 in leaf trichome extracts from this accession. All the XICs for formate adducts of the

acylglucoses are displayed in Figure 3-9. Findings from the current work document 21 chromatographic peaks corresponding to formate adduct of triacylglucoses in LA1522, and these are distributed across 12 different molecular masses.

Figure 3-9. Extracted ion chromatograms (XICs) generated from an extract of leaf tissue from *S. pennellii* LA1522 for ions corresponding to $[M+HCOO]^-$ for glucose triesters with total fatty acid carbon atoms ranging from 12 to 22. A: G 3:12 (m/z 435). B: G 3:13 (m/z 449). C: G 3:14 (m/z 463). D: G 3:15 (m/z 477). E: G 3:16 (m/z 491). F: G 3:17 (m/z 505). G: G 3:18 (m/z 519). H: G 3:19 (m/z 533). I: G 3:20 (m/z 547). J: G 3:21 (m/z 561). K: G 3:22 (m/z 575). L: G 3:23 (m/z 589).



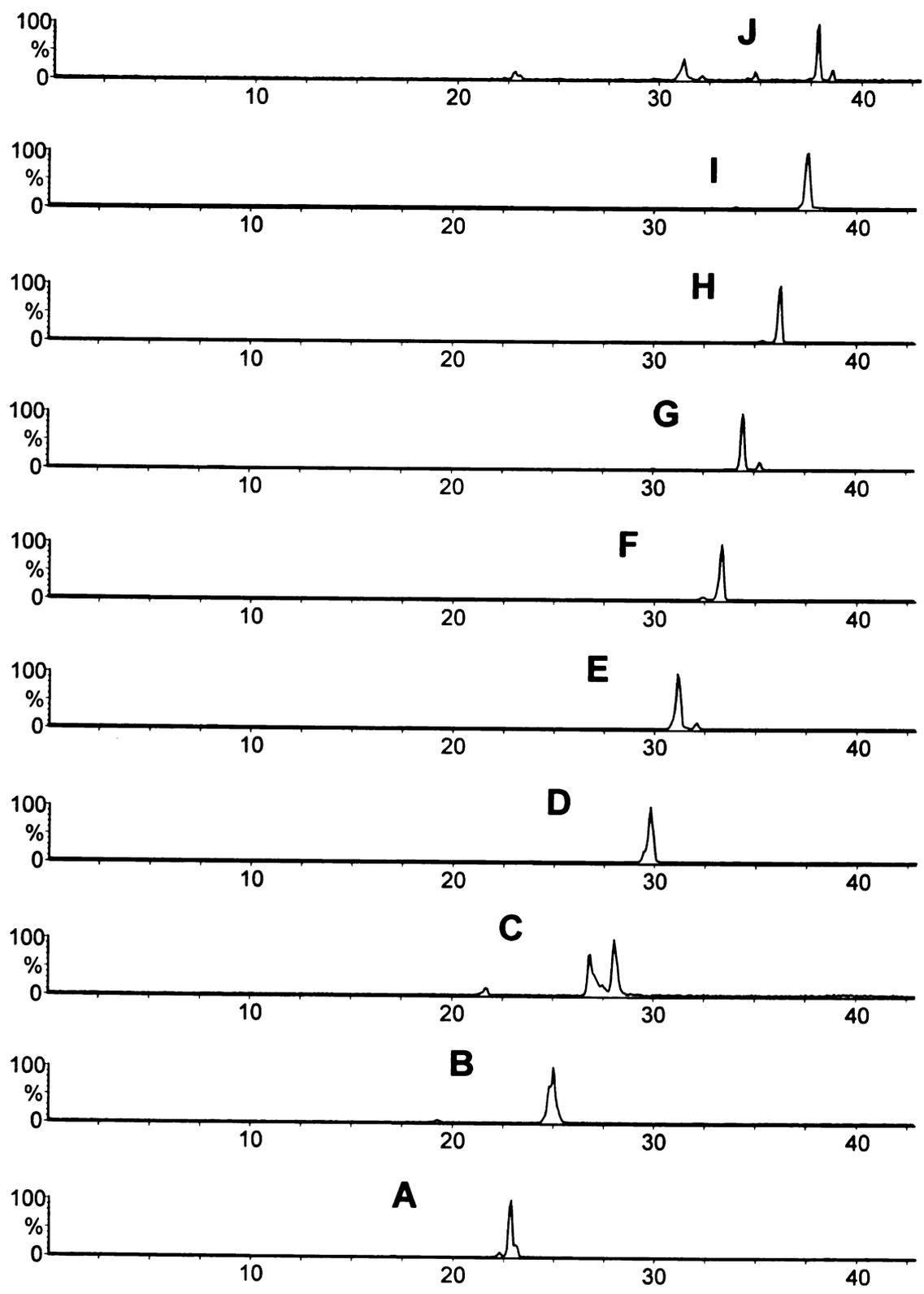
Although the patterns of acylsugars were similar with LA0716, the relative abundances of fatty acyl groups differed between accessions, as did fatty acyl substitution patterns. For example, triacylglucose G 3:19 (m/z 533) in LA0716 was substituted with C4, C5, and C10 fatty acyl groups in two isomers believed to differ only in substitution of a straight chain C10 ester by an *iso*-C10 ester. In contrast, for LA1522, the fatty acyl constituents were C5, C5, and C9 in all three observed chromatographic peaks. It is postulated that the isomers in LA1522 may arise from different isomeric C5 fatty acids (both *iso*-C5 and *anteiso*-C5 fatty acyl groups). The glucose triesters were assigned based on accurate mass measurement and fragments, as was performed before for LA0716. All the detected glucose triesters are described in table 3-3.

Table 3-3. Fragment ions and fatty acid constituents of detected glucose triesters in an extract of *S. pennellii* LA1522 leaf trichomes using negative and positive mode electrospray ionization (CID potential, 25 V)

Acylsugar metabolite annotation	Fatty acid ester groups	[M+HCOO] ⁻ (m/z)	[M-H] ⁻ (m/z)	Fragment ions (m/z)
G 3:12	C4, C4, C4	435	389	347, 301, 213, 143, 125, 87
G 3:13	C4, C4, C5	449	403	361, 315, 227, 143, 125, 101, 87
G 3:14	C4, C5, C5	463	417	375, 329, 227, 143, 125, 101, 87
G 3:15	C5, C5, C5	477	431	375, 329, 227, 143, 125, 101
G 3:16	C5, C5, C6	491	445	389, 343, 241, 227, 143, 115, 101
G 3:17	C4, C5, C8 C5, C6, C6	505	459	361, 315, 227, 125, 143, 101, 87 389, 343, 241, 227, 143, 115, 101
G 3:18	C4, C4, C10 C5, C5, C8	519	473	347, 301, 213, 143, 125, 171, 87 375, 329, 227, 143, 125, 101
G 3:19	C5, C5, C9	533	487	361, 315, 227, 143, 171, 101, 87
G 3:20	C5, C5, C10 C4, C5, C12	547	501	347, 301, 213, 143, 125, 171, 101 375, 329, 227, 143, 125, 199, 87
G 3:21	C5, C6, C10	561	515	361, 315, 227, 143, 199, 101, 87
G 3:22	C5, C5, C12	575	529	375, 329, 227, 143, 199, 101, 101
G 3:23	C5, C6, C12	589	543	389, 343, 241, 227, 199, 143, 115, 101

In addition to glucose triesters, sucrose triesters were also detected in LA1522 at levels similar to triacylglucoses. There were 15 chromatographic peaks corresponding to formate adduct of triacylsucroses in LA1522, and these range across 10 different molecular masses. All XICs of these sucrose triesters are shown in Figure 3-10. One striking difference between the triacylglucose and triacylsucrose chromatograms is evident. Specifically, fewer isomeric peaks are observed in the triacylsucroses than triacylglucoses. Two potential explanations come to mind. Either the LC separation fails to resolve some isomeric triacylsucroses, or the substitution patterns differ between acylglucoses and acylsucroses. This question remains unresolved, pending purification and NMR characterization of more acylsugar metabolites by other researchers in the Jones laboratory.

Figure 3-10. Extracted ion chromatograms (XICs) generated from an extract of leaf tissue from *S. pennellii* LA1522 of ions of m/z values corresponding to $[M+HCOO]^-$ for sucrose triesters with total fatty acid carbon atoms ranging from 12 to 23. A: S 3:15 (m/z 639). B: S 3:16 (m/z 653). C: S 3:17 (m/z 667). D: S 3:18 (m/z 681). E: S 3:19 (m/z 695). F: S 3:20 (m/z 709). G: S 3:21 (m/z 723). H: S 3:22 (m/z 737). I: S 3:23 (m/z 751). J: S 3:24 (m/z 765).



The process of annotating acylsucrose metabolites is illustrated using S 3:22 from LA1522 as an example. This metabolite was chosen because only one isomer was apparent from LC/MS results. Based on the slope of breakdown curves generated using multiplexed CID, the ions were assigned as adducts and fragments (Figure 3-11). At the lowest CID potential (10 V), $[M+HCOO]^-$ was the most abundant ion observed, accompanied by minor amounts of $[M+Cl]^-$. At a setting of 25 V, $[M-H]^-$ at m/z 691.39 appeared, but its abundance decreased at higher CID voltages. Yet higher CID potential yielded more extensive fragmentation, most notably the fatty acid anions at m/z 101.06 (C5) and 199.17 (C12). Other major fragments at m/z 607.33 (loss of C5), m/z 509.22 (loss of C12), m/z 425.17 (losses of C5 and C12) and m/z 341.11 (losses of two C5 and C12) are attributed to neutral losses of each of the three fatty acid groups as fatty ketenes. It is worthy of emphasis that eliminations of fatty ketenes from triacylsucroses stand as distinctly different from eliminations of neutral fatty acids from triacylglucoses. This point will be discussed in more detail below.

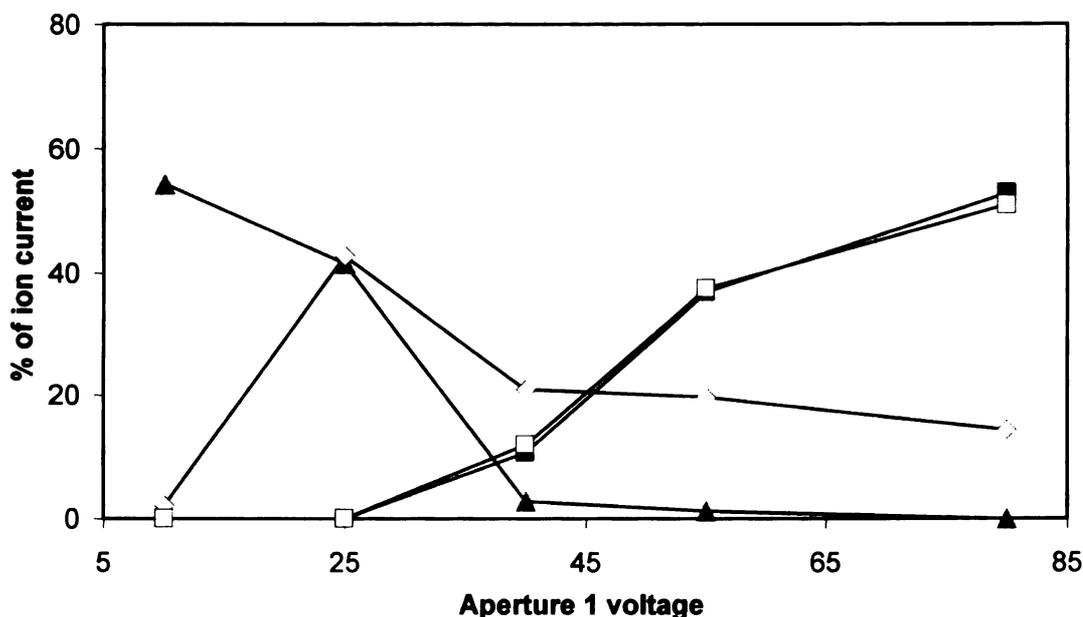


Figure 3-11. Breakdown curves derived from LC/negative ion ESI MS analyses of a leaf dip extract from *S. pennellii* LA1522 showing collision energy dependence of abundances of: (A) formate ion adducts of three triacylsucrose metabolites ($[M+HCOO]^-$ at m/z 737 S3:22 (▲); $[M-H]^-$ at m/z 691 (◇); C12 fatty acid at m/z 199 (■); C5 fatty acid anion (□).

To establish relationships between precursor and fragment ions, a product ion MS/MS spectrum was performed for the formate adduct of triacylsucrose S 3:22 from accession LA1522. Figure 3-12 compares ESI spectrum of S 3:22 obtained using LC/MS and multiplexed CID (top) to the product ion MS/MS spectrum (bottom) for $[M+HCOO]^-$. The two spectra share remarkable similarity, and the results suggest that all major fragments could be explained by decomposition of formate adducts. These findings allow

Losses of fatty ketene are characteristic of negative mode CID spectra of triacylsucroses. Their formation could be attributed to either a “charge-directed” process, which involves attack of anion charge site on the hydrogen of the fatty acid constituents, or a “charge-remote” mechanism occurring through *cis*-elimination reaction, as observed in CID experiments performed on acyl glycerophospholipids.^{181, 182} Loss of neutral fatty acid might arise from “charge-remote” mechanisms through *cis*- elimination reactions, but those fragment ions only were observed to have relatively low abundance in this work. The proposed fragmentation pathway for the formate adduct of S 3:22 is presented in Figure 3-15.

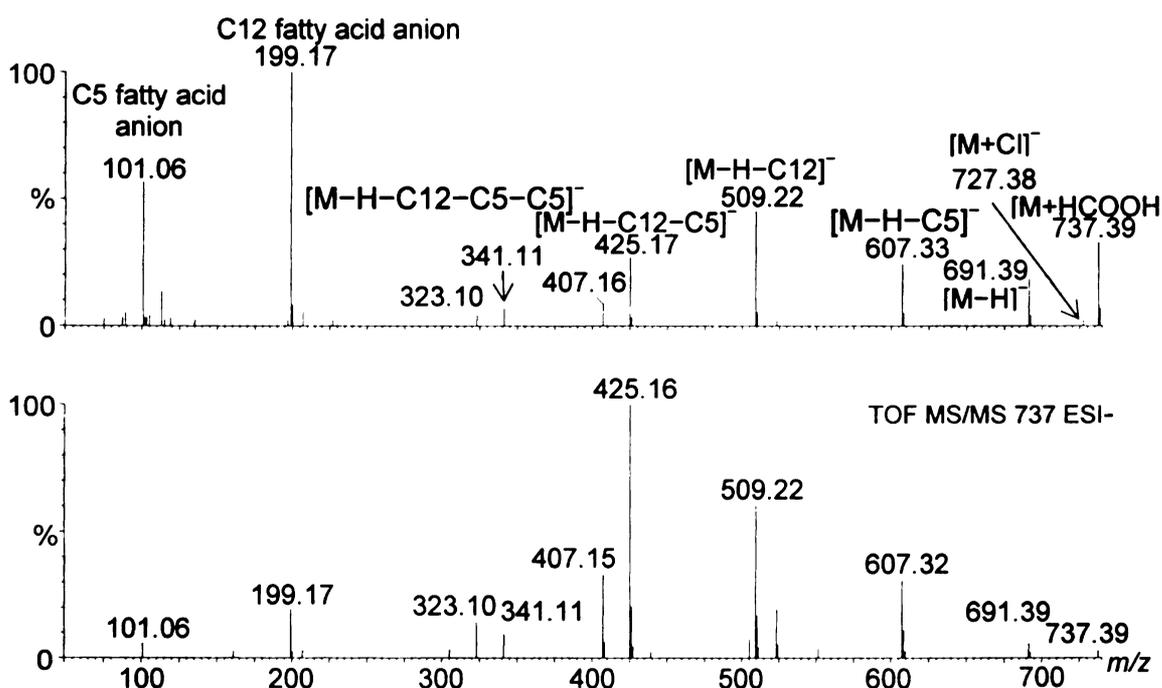


Figure 3-12 (TOP) ESI spectrum of triacylsucrose S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using CID potential of 55 V. (Bottom) Product ion MS/MS spectrum for formate adduct of S 3:22 (products of *m/z* 737.39 using collision potential of 35 eV).

Positions of fatty acid substitution on acylsugars were not evident from negative mode CID spectra, and structural information was limited to information about the number of carbons in fatty acyl moieties. CID spectra generated using positive mode provide complementary information about positions of fatty acyl groups. No $[M+H]^+$ ions were observed in positive mode ESI spectra of acylsugars, and the low CID potential spectra were dominated by adducts $[M+NH_4]^+$ and $[M+Na]^+$. Figure 3-13 displays a positive mode ESI spectrum of triacylsucrose S 3:22 from LA1522 leaf extracts using CID potential of 40 V.

Product ion MS/MS spectra were generated for ammonium and sodium adducts of S 3:22 (Figure 3-14) in order to establish the source(s) of fragment ions observed using multiplexed CID. The results suggested that most fragments under multiplexed CID in positive mode were derived from $[M+NH_4]^+$ because the predominant products of $[M+Na]^+$ were sodium-containing fragment ions not observed in the multiplexed CID spectra. Collision induced dissociation of $[M+Na]^+$ gave results similar to a report on acylsucroses from oriental tobacco reported by Xu and coworkers,¹⁸³ and a proposed fragmentation pathway of S 3:22 $[M+Na]^+$ is displayed in Figure 3-16. These findings are consistent with greater activation barriers to fragmentation of $[M+Na]^+$ owing to limited proton mobility.

Information from positive mode CID spectra of triacylsucroses complements assignments of fatty acyl composition derived from negative ion CID spectra. The most obvious example of this comes from the dominant loss of NH_3+180 (fructofuranose unit) through “charge-directed” cleavage of the glycosidic bond to produce a fragment ion at m/z 513.34. This fragment consists of the anhydroglucose core ion with all three fatty

acyl groups attached. The importance of this fragmentation reaction is that the fragment mass serves to indicate which sugar groups are substituted by specific fatty ester groups. This information is not evident in negative mode CID spectra. For all of the triacylsucroses from *S. pennellii* LA1522, all fatty acid groups are attached on the glucopyranose ring. Nearly all remaining products from $[M+NH_4]^+$ can be attributed to neutral losses of fatty acids or fatty ketenes. Proposed fragment pattern of S 3:22 $[M+NH_4]^+$ is shown in Figure 3-17.

A summary of the information from positive and negative mode CID spectra of triacylsucrose S 3:22 from LA1522 suggest C5, C5 and C12 fatty acyl groups, all of which are attached on the glucopyranose ring. As described in subsequent chapters, this substitution pattern is less common among *Solanum* accessions than substitution divided between the two rings. A compilation of fragment ions and fatty acyl constituents for acylsucroses in accession LA1522 was prepared from multiplexed CID spectra, and is assembled in Table 3-4. Accurate masses for sucrose esters are displayed in Table 3-5. In contrast to its close relative LA0716, the accession LA1522 contained both triacylglucoses and triacylsucroses, with the latter having all three fatty acyl groups on the glucopyranose ring. This discovery highlights substantial biochemical differences between two accessions of the same plant species, and points to remarkable chemical diversity within the genus *Solanum*. Of equal importance, the observed acylsugars exhibit some specificity in substitution patterns, and are not random combinations of attachments of fatty acids to sugars. The selectivity of substitution reflects enzyme selectivity in biosynthesis. The chemical diversity of trichome metabolites is explored in more detail across a broader range of genetic variants in subsequent chapters.

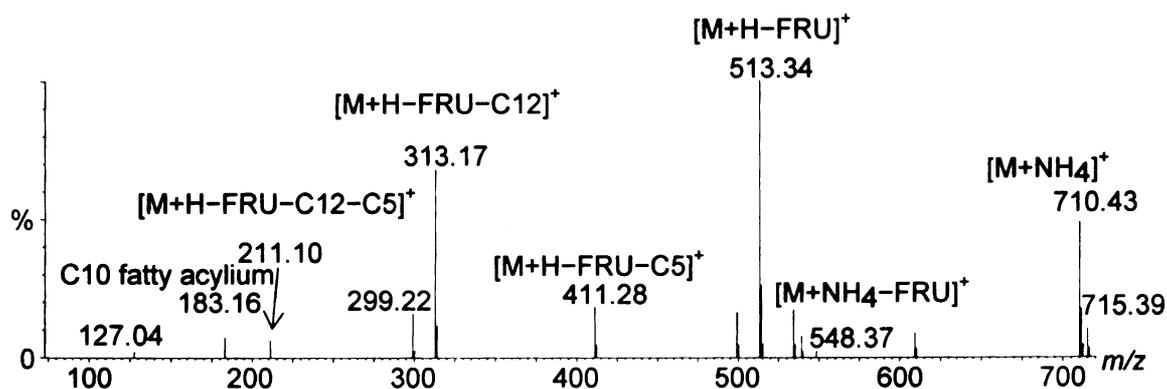


Figure 3-13 Positive mode ESI spectrum of triacylsucrose S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using CID potential of 55 V. The abbreviation ‘FRU’ refers to a fragment derived from the fructofuranose ring.

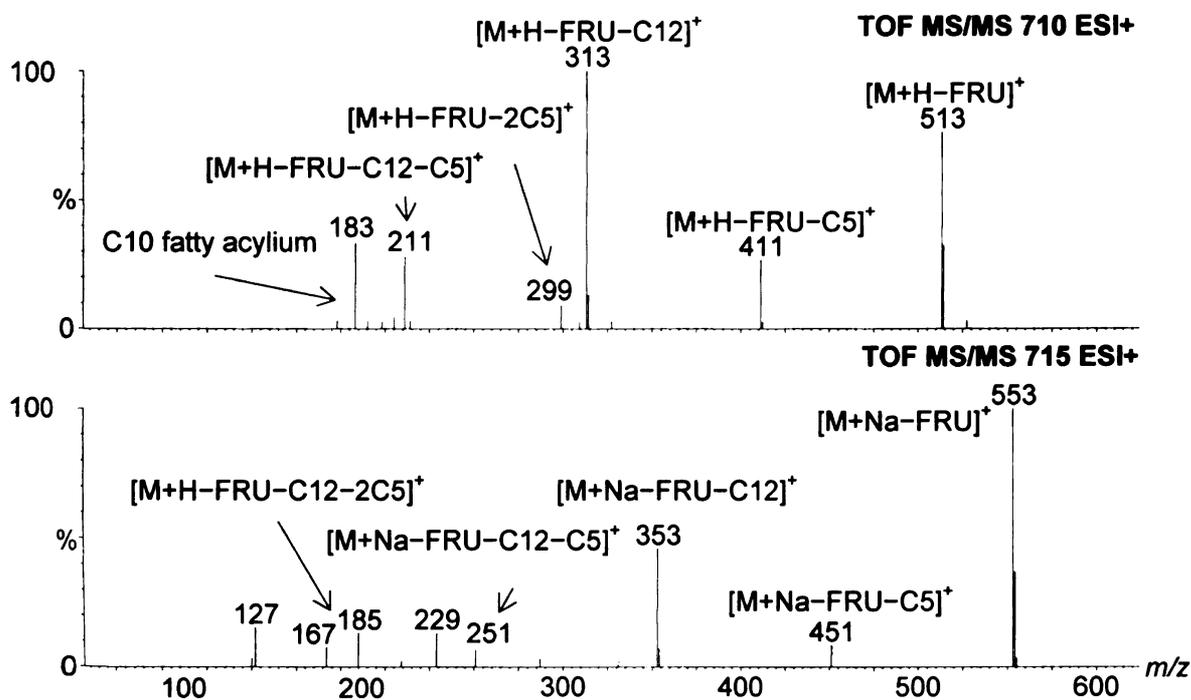
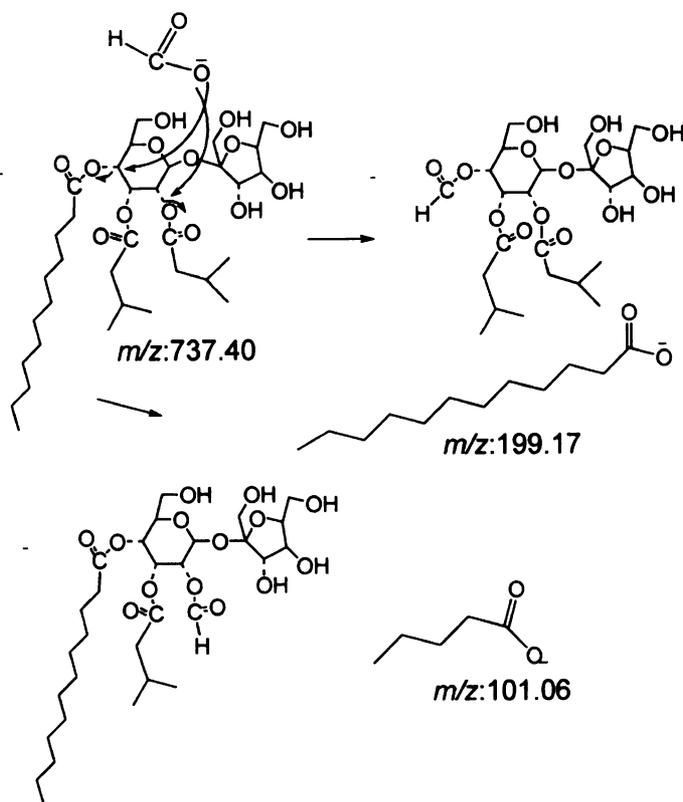
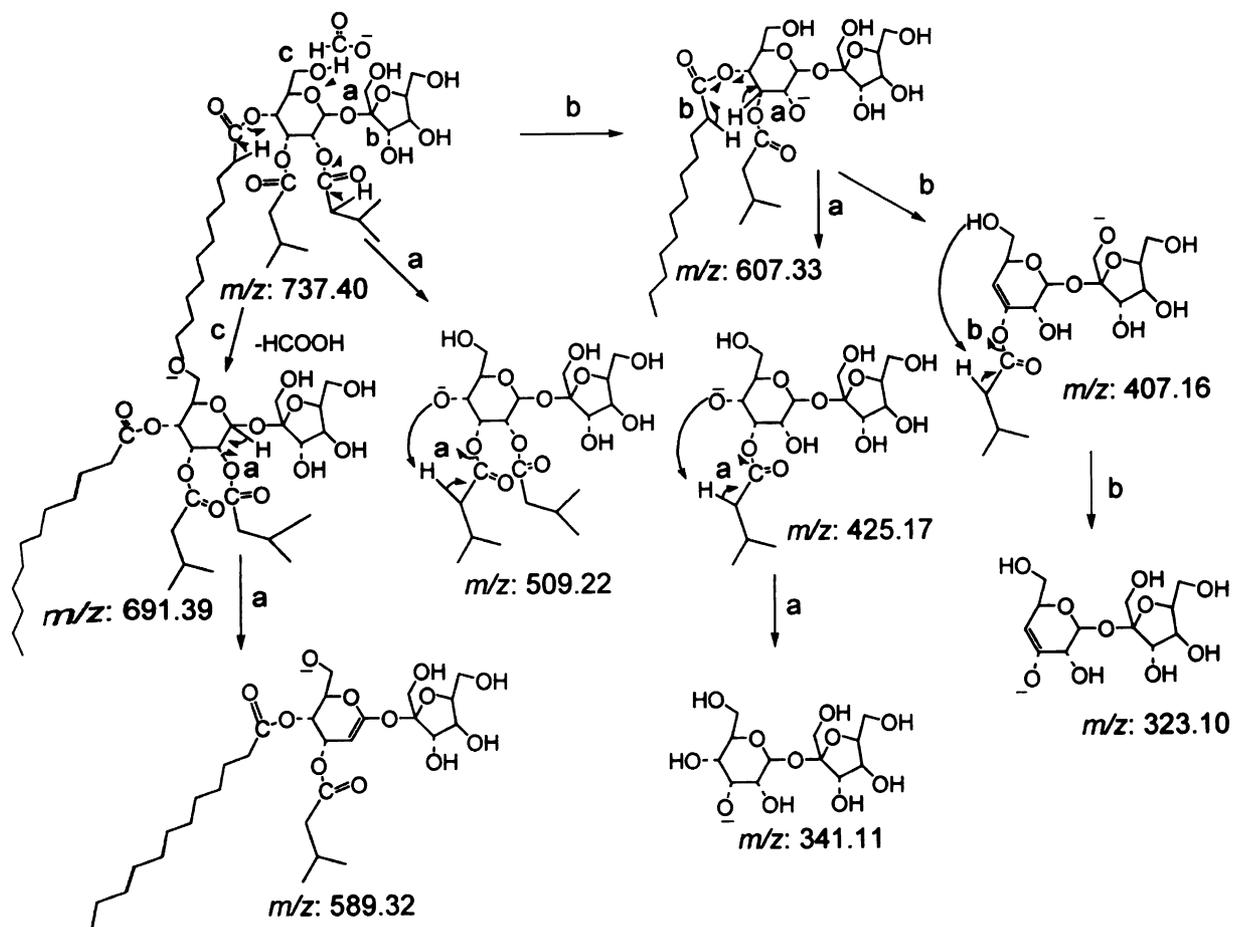


Figure 3-14 (Top) Product ion MS/MS spectra of (Top) $[M+NH_4]^+$ at m/z 710.43 (collision energy: 45 eV) and (Bottom) $[M+Na]^+$ at m/z 715.39 (collision energy: 55 eV) for S 3:22 extracted from *S. pennellii* LA1522. The abbreviation ‘FRU’ refers to a fragment derived from the fructofuranose ring.

Figure 3-15 Proposed fragment pathway of formate adduct of S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using ESI negative mode with CID voltage 55 V. Specific substitution positions of individual fatty acids remain uncertain.



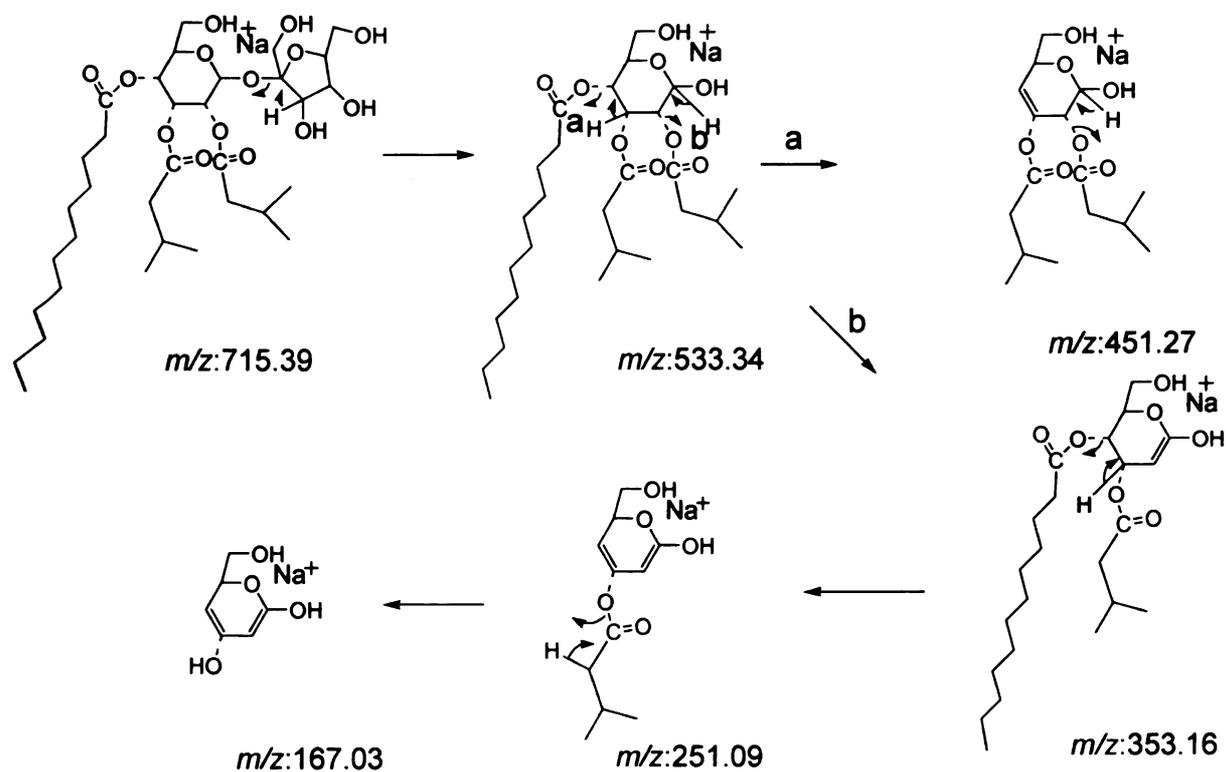


Figure 3-16 Proposed fragment pathway for sodium adduct of S 3:21 extracted from *S. pennellii* LA1522 based on MS/MS product ion spectrum. Specific substitution positions of individual fatty acids remain uncertain

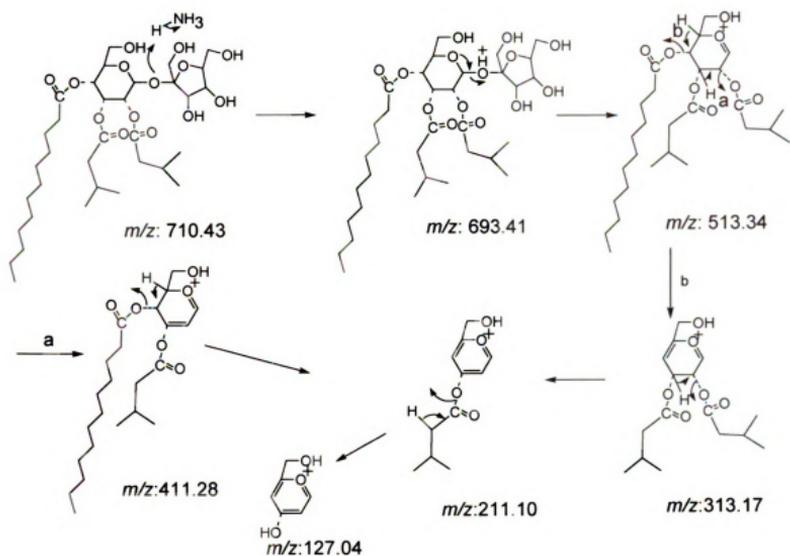


Figure 3-17 Proposed fragment pathway of ammonium adduct for S 3:22 (RT = 36.50 min) extracted from *S. pennellii* LA1522 using ESI positive mode with CID voltage 40 V. Specific substitution positions of individual fatty acids remain uncertain.

Table.3-4. Fragments and fatty acid constituents for detected sucrose triesters in *S. pennellii* LA1522 using ESI negative (CID potential, 55 V) and positive mode (CID potential, 40 V)

Acylsugar metabolite and Fatty acid constituents	[M+HCOO] ⁻ (m/z)	[M+ ³⁵ Cl] ⁻ (m/z)	Fragment ions (m/z)
S 3:15 C5,C5,C5	639	629	593, 425, 407, 341, 323, 179, 101
S 3:16 C5,C5,C6	653	643	607, 523, 425, 341, 323, 115, 101
S 3:18 C5,C5,C8	681	671	635, 551, 425, 407, 341, 179, 143, 101
S 3:19 C5,C5,C9	695	685	649, 565, 509, 425, 341, 323, 157, 101
S 3:20 C5,C5,C10	709	699	663, 579, 509, 425, 407, 341, 323, 171, 101
S 3:21 C5,C6,C10	723	713	677, 593, 579, 509, 425, 407, 341, 323, 171, 115, 101
S 3:22 C5,C5,C12	737	727	691, 607, 509, 425, 407, 341, 323, 199, 179, 101
S 3:23 C5,C6,C12	751	741	705, 621, 607, 523, 439, 421, 341, 323, 199, 115, 101
S 3:24 C6,C6,C12	765	755	719, 537, 439, 341, 323, 199, 115

Acylsugar metabolite and Fatty acid constituents	[M+NH ₄] ⁺ (m/z)	[M+Na] ⁺ (m/z)	Fragment ions (m/z)
S 3:15 C5,C5,C5	612	617	455, 415, 313, 211, 127
S 3:16 C5,C5,C6	626	631	469, 429, 313, 211, 127
S 3:18 C5,C5, C8	654	659	497, 457, 355, 313, 211, 127
S 3:19 C5,C5,C9	668	673	471, 313, 211, 127
S 3:20 C5,C5,C10	682	687	485, 313, 211, 127
S 3:21 C5,C6,C10	696	701	499, 327, 225, 127
S 3:22 C5,C5,C12	710	715	513, 411, 313, 309, 211, 127
S 3:23 C5,C6,C12	724	729	527, 327, 211, 127
S 3:24 C6,C6,C12	738	743	541, 341, 225, 127

Table 3-5. Exact mass measurement of $[M+HCOO]^-$ Ions for detectable sucrose triesters in accession LA1522 detected using ESI negative mode under CID potential 10 V.

Acylsugar metabolite annotation	Formula of formate adduct	Measured m/z $[M+HCOO]^-$	Calculated m/z	Error (ppm)
S 3:15	C ₂₈ H ₄₇ O ₁₆	639.2885	639.2864	3.3
S 3:16	C ₂₉ H ₄₉ O ₁₆	653.3036	653.3021	2.3
S 3:17	C ₃₀ H ₅₁ O ₁₆	667.3150	667.3177	-4.0
S 3:18	C ₃₁ H ₅₃ O ₁₆	681.3334	681.3336	0.3
S 3:19	C ₃₂ H ₅₅ O ₁₆	695.3505	695.3490	2.2
S 3:20	C ₃₃ H ₅₇ O ₁₆	709.3619	709.3647	-3.9
S 3:21	C ₃₄ H ₅₉ O ₁₆	723.3803	723.3822	2.6
S 3:22	C ₃₅ H ₆₁ O ₁₆	737.3960	737.3972	1.6
S 3:23	C ₃₆ H ₆₃ O ₁₆	751.4116	751.4102	-1.9

As was the case for *S. pennellii* LA0716, the branched or straight chain fatty acyl groups for acylsugars from accession LA1522 could not be distinguished from CID spectra. Therefore, GC/MS was performed to identify fatty acids after transesterification from all the acylsugars in these wild type species. The GC/MS results for the composite composition of fatty acyl groups in leaf extracts of *S. pennellii* LA1522 are summarized in Figure 3-18. All fatty acid chain lengths observed in LC/MS results were supported by GC/MS data. The identities of fatty acyl groups were based comparisons of EI mass spectra from GC/MS analyses with retention times and mass spectra of authentic standards. When standards were not available, the fatty acid ethyl esters were identified using comparisons to the NIST 05 mass spectrum library. GC/MS results revealed that in LA1522, the iC5 was most abundant (~40% of peak area), with about 15% each of aiC5, nC10, and nC12. In contrast, iC5 made up less than 5% of total TIC area for LA0716, for which iC4 accounted for nearly 40% of the peak area. The aiC5 and iC10 esters made up about 15% and 20% respectively, and nC10 was the next most abundant chain (~

10%). It is also worthy to note the inverse ratio of *anteiso* to *iso*-C5 isomers in LA0716 relative to LA1522. The *iso*- and *anteiso*-C5 are believed to be derived from different amino acid precursors, leucine and isoleucine respectively. Whether the shift in C5 isomer abundances results from different relative sizes of the leucine and isoleucine pools or from different pathway efficiencies cannot be determined from available evidence. Likewise, the *iso*-C4 fatty acid is believed to be derived from valine, but reasons for the high abundance of *i*C4 esters in LA0716 cannot yet be established with certainty. It is recommended that future work address these questions, perhaps through use of metabolic flux measurements using stable isotope labeled tracers.

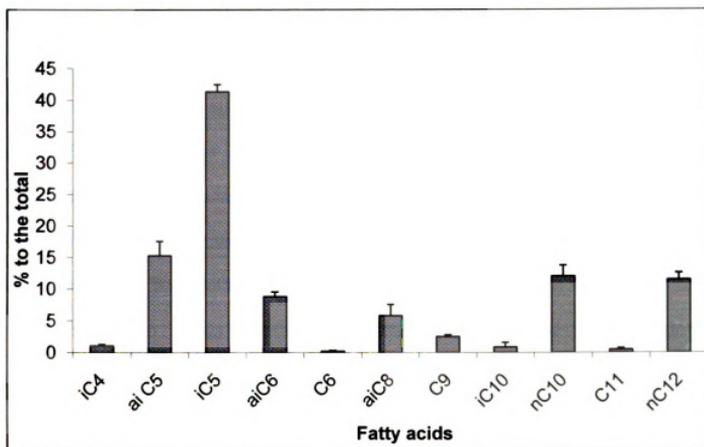


Figure 3-18 Fatty acyl groups from total acylsugars in an extract of *S. pennellii* LA1522 were transesterified to form fatty acid ethyl esters, and analyzed using GC/MS. Data for each fatty acid were calculated as a percentage of the total ion current chromatogram peak areas of fatty acids detected through GC/MS. The abbreviations “i” and “ai” refer to *iso*- and *anteiso*- branched isomers. Error bars indicate standard error with $n = 3$.

Generation of fragment ions from triacylsucroses required higher collision potentials than for analogous triacylglycerols, as was evident from inspection of plots of various ion abundances as a function of collision potential (breakdown curves) described in Chapter 2 (Figures 2-5 and 2-6). For acylglucosides, $[M+\text{formate}]^-$ abundances drop sharply as collision potential is increased above 10 V, while for acylsucroses, the drop occurs at collision potentials greater than 25 V. While this behavior can be attributed in part to

lower center-of-mass collision energies for the more massive triacylsucroses, a comparison of the behavior of the formate adduct of the lowest mass triacylsucrose formate (m/z 625) shows striking differences relative to the most massive triacylglucose (m/z 589), with the latter forming greater yields of CID products at lower collision potentials. Based on these observations, we judge the activation energy for fragmentation of triacylglucoses to be lower than triacylsucroses based on the threshold CID potential needed to observe fragment ions.

Differences in energy thresholds for fragmentation can be exploited using multiplexed CID to distinguish acylglucoses and acylsucroses when they coelute, as is often the case when analyzing plant extracts with diverse composition of acylsugar metabolites. Resolution of fragments from coeluting acylsugars is illustrated in Figure 3-19, using coeluting triacylglucose G 3:14 (4,5,5) and triacylsucrose S 3:15 (5,5,5). At the lowest collision potential (10 V) the spectrum displays multiple ions from a triacylsucrose (S3:15) including formate (m/z 639) and chloride adducts (m/z 629 and 627), and a formate-bound dimer ion (m/z 1223). Of lower abundance is $[M-H]^-$ of triacylglucose G 3:14 (m/z 463). Raising the collision potential to 25 V (middle panel) leads to extensive fragmentation of the triacylglucose (ions marked with *), but no fragments (e.g. $[M-H]^-$) derived from CID of the triacylsucrose are observed at this potential. Fatty acyl anions are observed at m/z 87 and 101 corresponding to C4 and C5 fatty acyl groups from G 3:14. In the bottom panel, the highest collision voltage of 80 V yields fragmentation of the triacylsucrose as evident from appearance of $[M-H]^-$ from S3:15 (m/z 593) and neutral losses of fatty ketenes or fatty acids. This feature aided profiling of metabolites of accession LA1522 which produces complex mixtures of

triesters of both glucose and sucrose. This multiplexed CID approach offers great potential to resolve coeluting compounds based on different fragmentation energy thresholds.

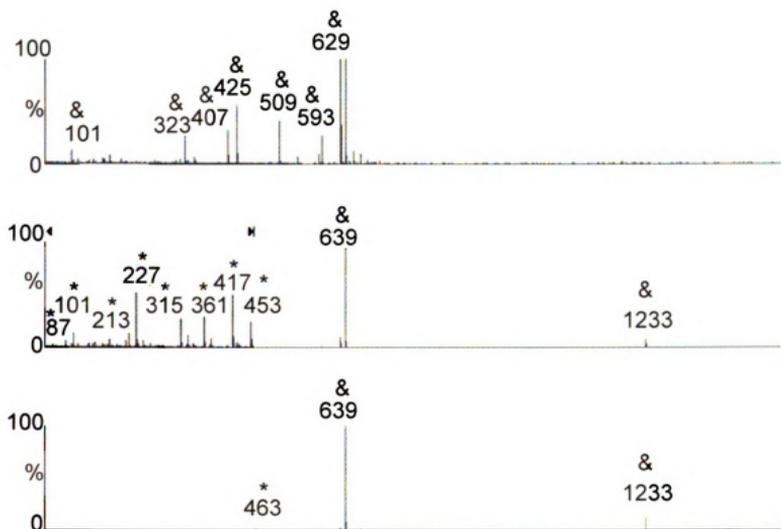


Figure 3-19. Multiplexed CID mass spectra (electrospray, negative mode) of coeluting metabolites from LC/TOF MS analysis of an extract of *S. pennellii* LA1522 showing spectra obtained from three of the five Aperture 1 potentials. Top panel: 10 V; middle panel: 25 V; bottom panel: 80V. m/z of formate adduct ions at m/z 639 is sucrose triester S 3:15 (5,5,5) and m/z 463 is glucose triester G 3:14 (4,5,5) at RT = 22.8 min.

3.3.3. Structure annotation of sucrose tetraesters in *S. habrochaites* LA1777.

Although the acylsucroses described in *S. pennellii* LA1522 were all triesters, other *Solanum* accessions exhibited metabolites of the same nominal mass but earlier

chromatographic elution times. Closer examination of the m/z values of these metabolites suggested elemental formulas more consistent with sucrose tetraesters. Sucrose tetraesters have the same nominal mass as sucrose triesters but with one less carbon atom in the fatty acyl chains. Accurate mass measurements allowed us to distinguish them by mass difference, as the tetraesters are lower in mass than the isobaric triester by 0.036 Da. In addition, sucrose tetraesters can be distinguished from triesters based on different fragmentation behavior with ESI positive mode, which is described in detail in following example for structure annotation for S 4:21 from LA1777.

Based on differences in exact masses and fragmentation behavior, most acylsugars in extracts of *S. habrochaites* LA1777 were assigned as sucrose tetraesters. Figure 3-20 displays extracted ion chromatograms for m/z values corresponding to formate adduct of acylsucroses for LA1522 and LA1777 accessions. Observed peaks in these chromatograms separate into two distinct retention time groups. Peaks corresponding to tetraesters of sucrose are highlighted with a blue dashed line, and those corresponding to triacylsucroses are highlighted with a red dashed line. The tetraesters were easily distinguished by chromatography, as all eluted about 5 minutes earlier than the triesters of the same nominal mass. Five sucrose tetraesters were reported by King¹¹³ as constituents of *S. habrochaites* that we abbreviate as follows: S 4:14 (2,4,4,4), S 4:17 (2,5,5,5), S 4:20 (2,4,4,10) and S 4:21 (2,4,5,10). All these acylsucroses were tetraesters with one acetate group on glucopyranose. The LC/MS profiling recognized 18 chromatographic peaks corresponding to formate adducts of tetraacylsucroses from accession LA1777, and these span 11 different molecular masses ranging from 14 to 24 total fatty acyl carbons.

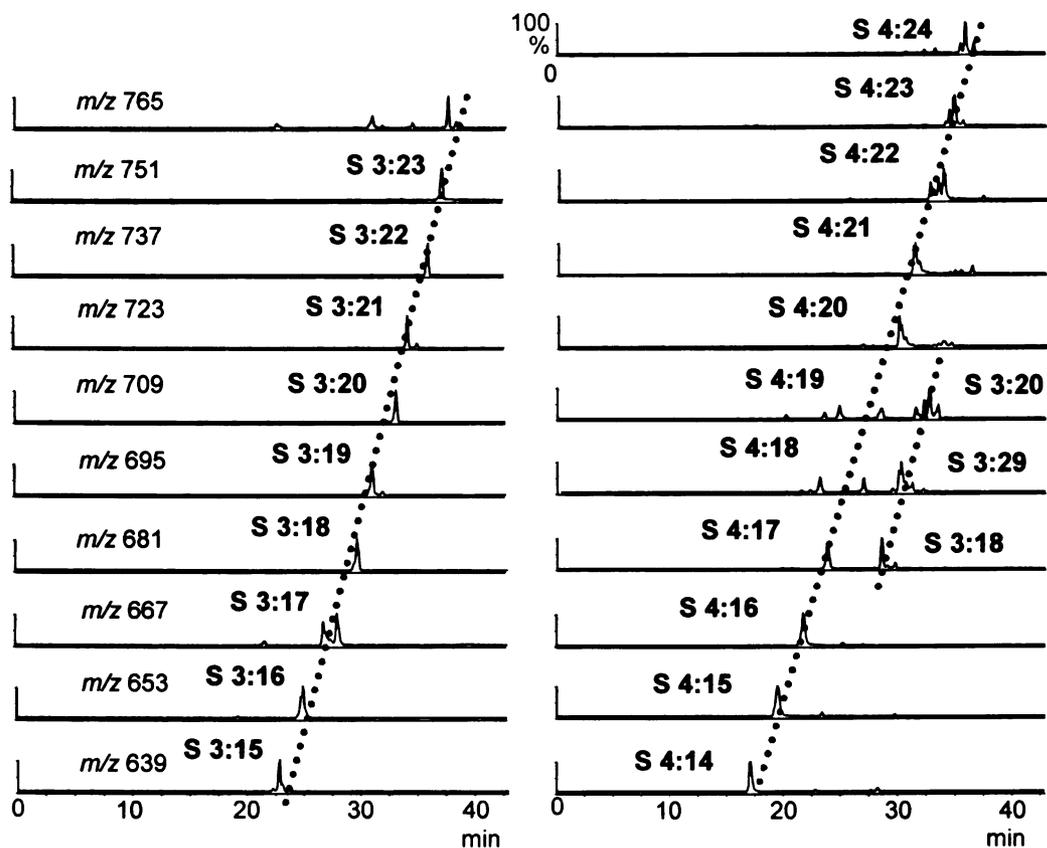


Figure 3-20. (A) Extracted ion chromatograms (XICs) from negative ion mode LC/MS analyses of (A) *S. pennellii* LA1522 and (B) *S. habrochaites* LA1777. Selected m/z values correspond to $[M+HCOO]^-$ of sucrose triesters (highlighted with dashed lines) and sucrose tetraesters (solid line).

The tetraacylsucrose S 4:21 from *S. habrochaites* LA1777 is used to illustrate how sucrose tetraesters were annotated. Figure 3-21 (top) shows the mass spectrum obtained using CID potential of 55 V for the most abundant isomer of S 4:21. Fragments of m/z 87.04, 101.06 and 171.12 were assigned as C4, C5 and C10 fatty acid anions, correspondingly. This, and all other tetraacylsucroses from this accession showed

87.04, 101.06 and 171.12 were assigned as C4, C5 and C10 fatty acid anions, correspondingly. This, and all other tetraacylsucroses from this accession showed formation of an ion corresponding to loss of formic acid and ketene from the formate adduct ($[M-H-C_2H_2O]^-$; m/z 649.34 for S 4:21) that was not present for any of the sucrose triesters. This fragment serves as evidence of an acetyl ester. Additional fragments at m/z 537.22, 495.21, 425.17, and 341.11 correspond to further losses of the longer chain neutral fatty ketenes. Product ion MS/MS spectrum were generated for formate adducts of S 4:21 in order to establish the source(s) of fragment ions observed using multiplexed CID, as we did for S 3:22 from LA1522 (Figure 3-21 bottom). The two spectra share remarkable similarity, and the results suggest that all major fragments could be explained by decomposition of formate adducts. These findings allow annotation of S 4:21 from LA1777 as a sucrose tetraester with one C2, one C4, one C5 and one C10 fatty acyl groups. For sucrose tetraesters in *S. habrochaites* LA1777, the fragmentation pathway in negative mode was similar with sucrose trimesters in *S. penellii* LA1522. Losses of fatty ketene are characteristic of negative mode CID spectra of triacylsucroses. Their formation could be attributed to either a “charge-directed” process, which involves attack of anion charge site on the hydrogen of the fatty acid constituents, or a “charge-remote” mechanism occurring through *cis*-elimination reaction.

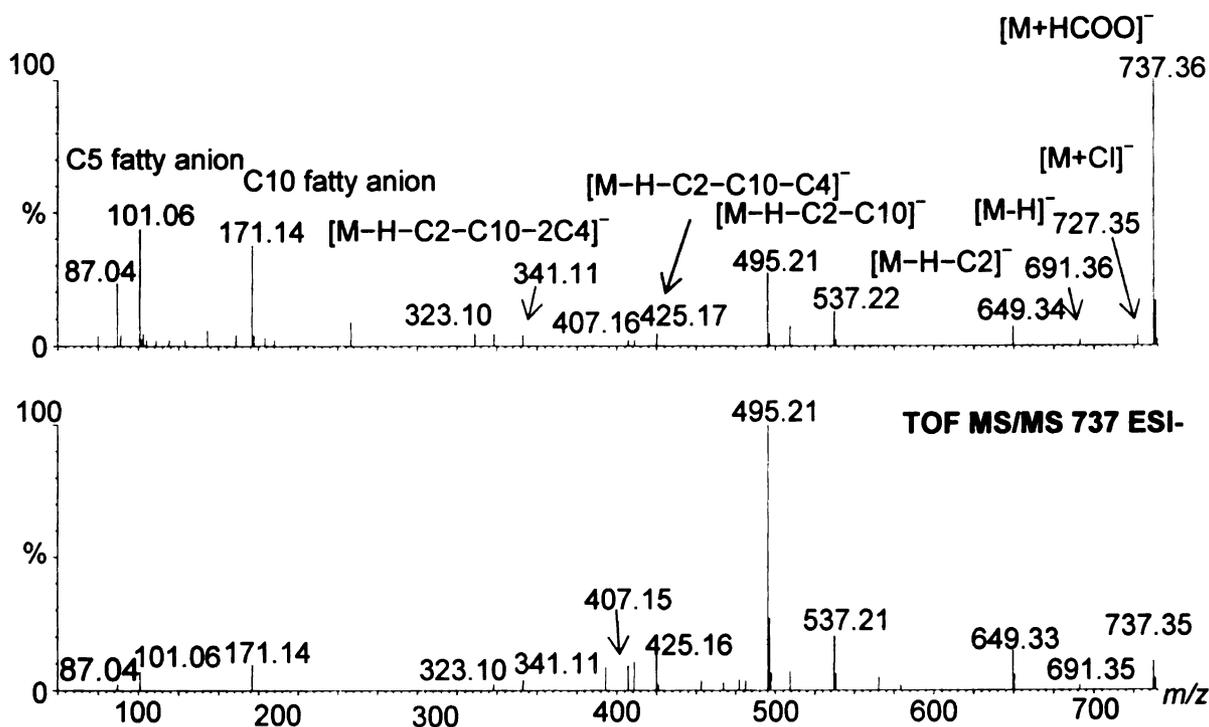


Figure 3-21. (Top) Negative mode ESI spectrum of tetraacylsucrose S 4:21 (RT = 31.50 min) under CID potential 55 V and (Bottom) product ion MS/MS spectrum of formate adduct (m/z 737.35) for S 4:21 from *S. habrochaites* LA1777.

More information about which rings bear individual fatty acids position was yielded from positive mode CID spectra. Figure 3-22 displays a positive mode ESI spectrum of triacylsucrose S 4:21 from *S. habrochaites* LA1777 leaf extracts using a CID potential of 40 V. Product ion MS/MS spectra were also generated for ammonium and sodium adducts of S 4:21 in order to establish the source(s) of fragment ions observed using multiplexed CID (Figure 3-23). The results suggested that most fragments observed in positive mode using multiplexed CID were derived from $[M+NH_4]^+$, in similar fashion as the triesters from LA1522 discussed above. $[M+NH_4]^+$ for sucrose tetraesters from LA1777 was cleaved at the glycosidic bond that connects the two rings, forming two fragments corresponding to acyl substituted glucopyranose and fructofuranose ions.

Similar ion behavior was observed for acylsugars in *Nicotiana benthamiana*.¹¹⁹ Dissociation of $[M+NH_4]^+$ preferentially cleaved the glycosidic bond and formed cations stabilized by the adjacent oxygen. This selectivity in yields of fragment ions is attributed to formation of a more stable tertiary carbocation from the fructofuranose ring relative to a secondary carbocation from the glucopyranose ring (Figure 3-24). For S 4:21, CID of $[M+NH_4]^+$ generated fragment ions at m/z 317.20 and m/z 359.18, with the former being more abundant under a wide range of collision conditions. In the case of S 4:21, the mass of the fructofuranose fragment ion (m/z 371.20) is consistent with substitution by one C10 fatty acyl group. A pseudo MS³ spectrum of product ions from m/z 371 was generated after forming this ion in the source by raising the source cone voltage (Figure 3-23, bottom). The prominent product ion at m/z 155.12 was assigned as the C10 acylium ion. The remaining C2, C4, and C5 acyl groups were on the glucopyranose unit for S 4:21 as supported by the fragment ion at m/z 359.17. Fragmentation reactions that generate structurally useful fragment ions for acylsugars from CID of $[M+formate]^-$ and $[M+NH_4]^+$ are summarized in Figures 3-25 and 3-26.

Results from CID in positive ion mode allow the fatty acid groups to be assigned to either the glucopyranose or fructofuranose rings. Sucrose tetraesters in *S. habrochaites* LA1777 were substituted with combinations of C2, C4, C5, C10, and C12 fatty ester groups, but these groups were not randomly distributed. Fragments derived from cleavage of the glycosidic bonds showed that the glucopyranose ring was substituted with three variations of a single fatty ester. All C2 esters were on the glucopyranose ring, with the variations consisting of either two C4, one each of C4 and C5, or two C5 groups. These are reflected in the minor abundance fragments at m/z 345, 359, and 373 (Table 3-

6). Substitution on the fructofuranose ring was limited to a single acyl group, varying among C4, C5, C10, and C12 esters. Note that the longer C10 and C12 esters are never observed on the glucopyranose ring. This behavior is the opposite of tomato (*S. lycopersicum* M82), where the longer fatty acyl groups were always on the glucopyranose ring. Accurate mass measurements for major acylsucroses from accession LA1777 are presented in Table 3-7, and these measurements are consistent with the assignments described above. Findings from this work indicate that tomato and its relatives exert tight control over which fatty acyl groups are attached at each position in acylsugars, yet there is dramatic diversity in acylgroup substitution among related species within the genus *Solanum*. Mass spectra that exploit CID behavior offer a powerful window into the diversity of acylsugar metabolite structures that should accelerate discoveries of the genetic basis for this selectivity.

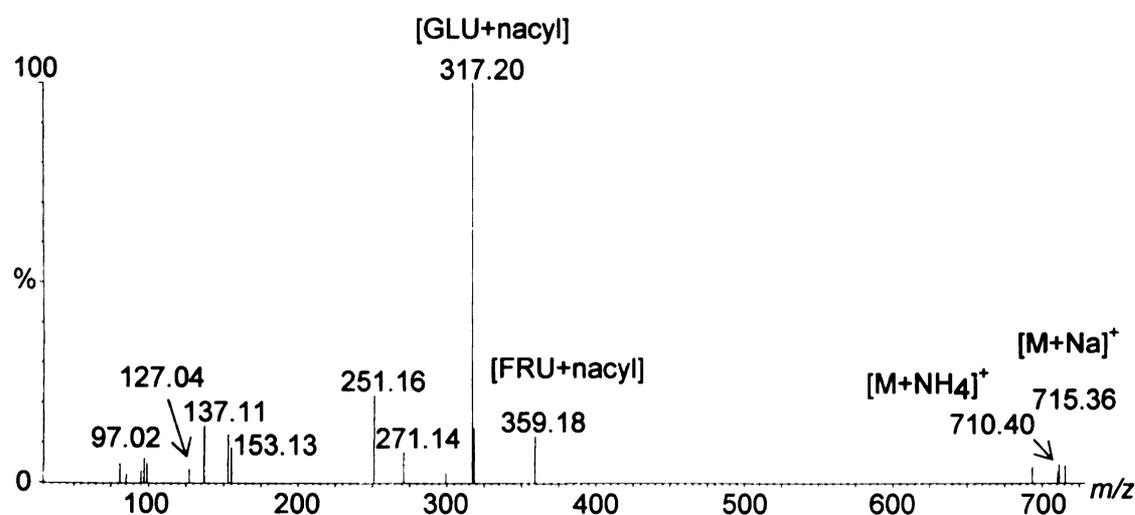


Figure 3-22. Positive mode ESI spectrum of tetraacylsucrose S 4:21 (RT = 31.50 min) from *S. habrochaites* LA1777 using CID potential 40 V. The abbreviation ‘FRU’ refers to a fragment derived from the fructofuranose ring and ‘GLU’ refers to a fragment derived from glucopyranose ring.

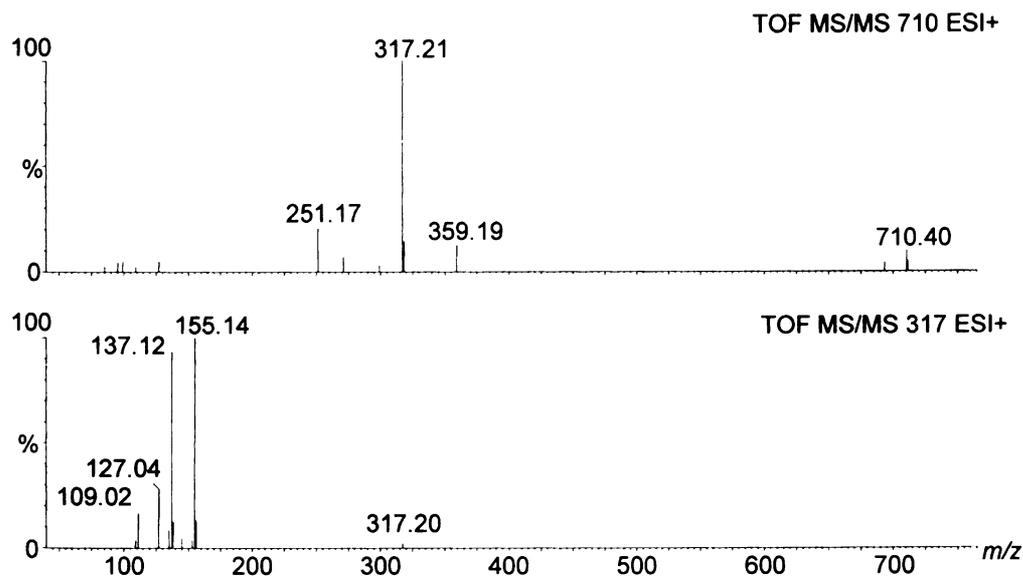


Figure 3-23. (Top) Product ion MS/MS spectrum of $[M+NH_4]^+$ at m/z 710.40 and (Bottom) product ion MS/MS spectrum for m/z 317 ($[FRU+acyl]^+$) of S 4:21 from *S. habrochaites* LA1777; m/z 317 was generated in source from tetraesters S4:21 from *S. pennellii* LA11777. The abbreviation 'FRU' refers to a fragment derived from the fructofuranose ring.

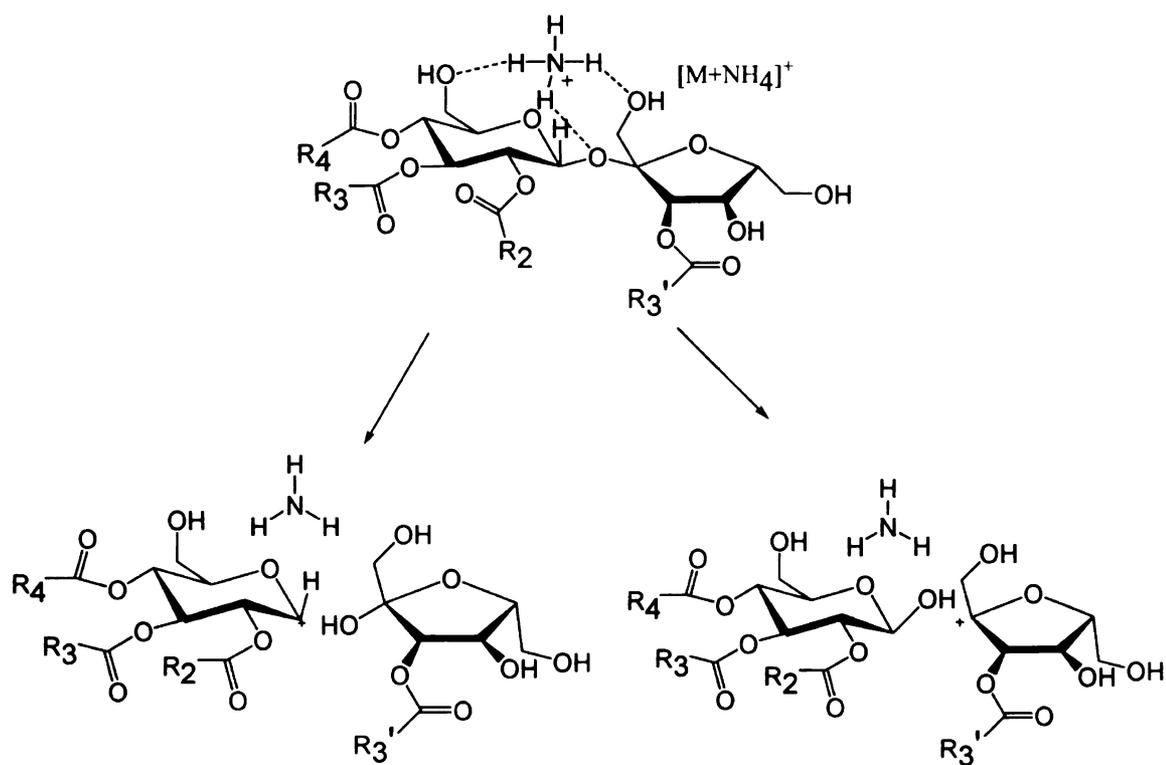


Figure 3-24. Collision induced dissociation of $[M+NH_4]^+$ acylsugar ions yield a dominant acylfructofuranose fragment ion, and a less abundant acylglucopyranose fragment ion.

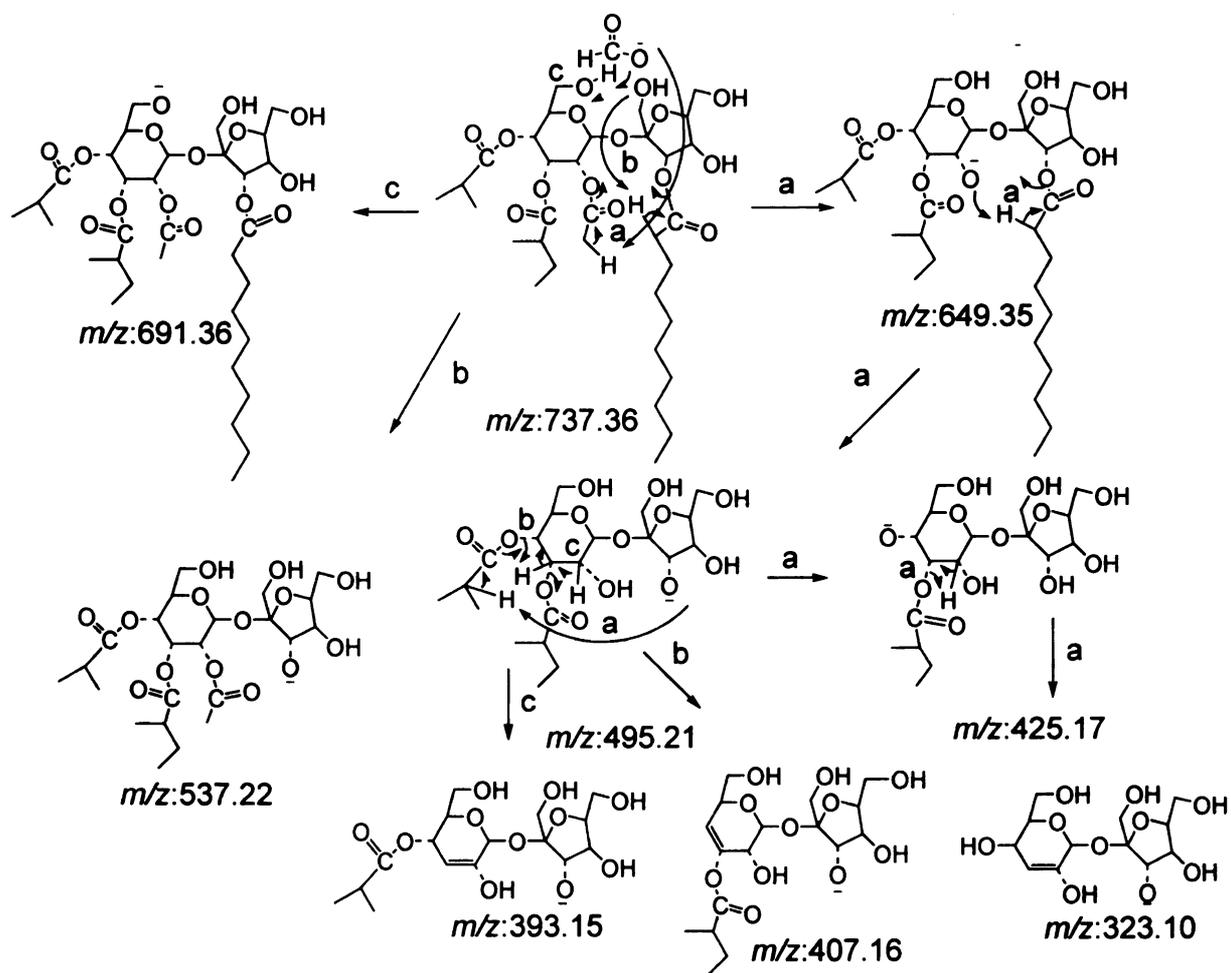


Figure 3-25 Proposed fragment pathway upon CID of the formate adduct for S 4:21 (RT = 31.5 min) extracted from *S. habrochaites* LA1777 using ESI negative mode with CID voltage 55 V.

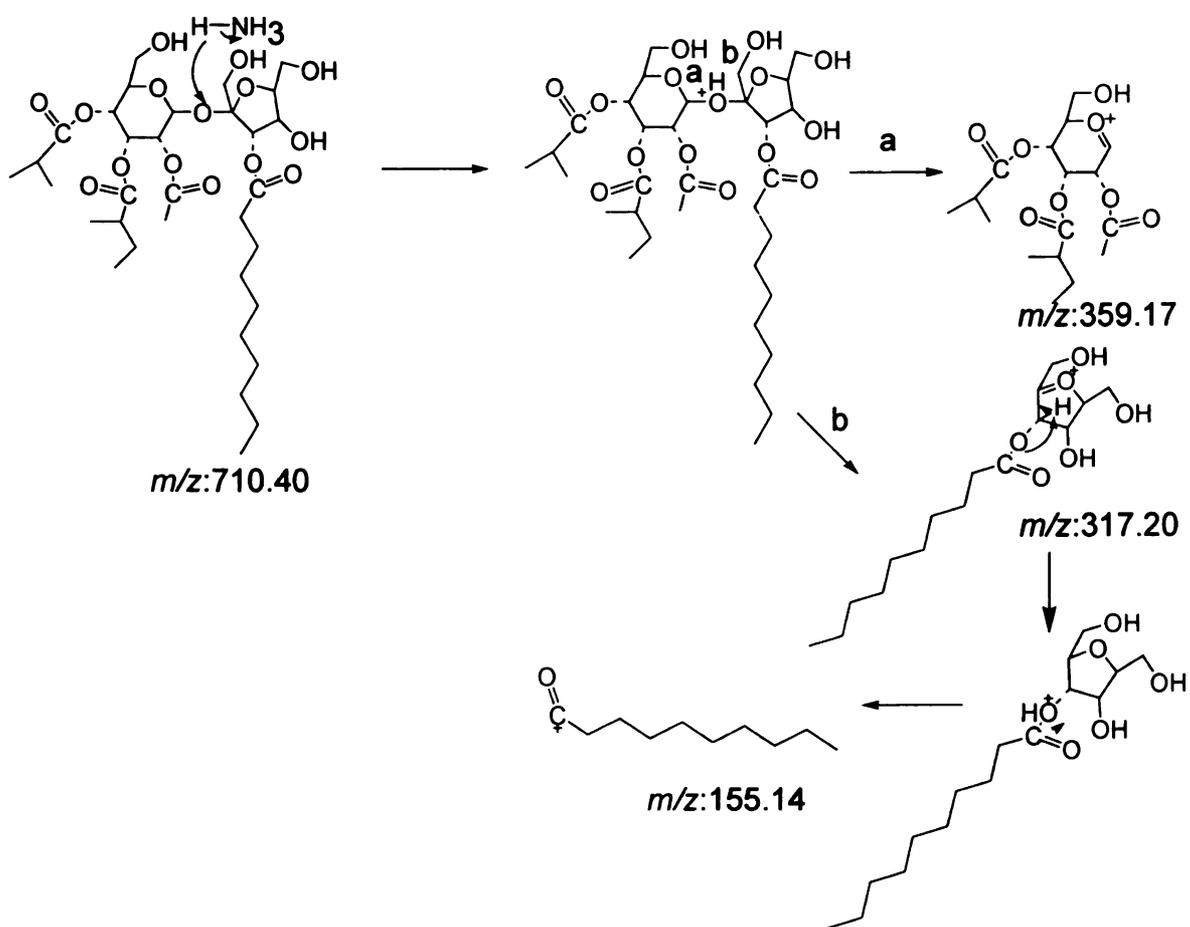


Figure 3-26 Proposed fragment pathway of ammonium adduct for S 4:21 (RT = 31.5 min) extracted from *S. habrochaites* LA1777 using ESI positive mode with CID voltage 55 V

Table 3-6. Fragments and fatty acid constituents of detected sucrose tetraesters from *S. habrochaites* LA1777 using ESI positive mode (CID potential, 40 V) and negative (CID potential, 55 V).

Acylsugar metabolite and Fatty acid constituents	[M+NH ₄] ⁺ (<i>m/z</i>)	[M+Na] ⁺ (<i>m/z</i>)	Major Fragments in order of decreasing abundance
S 4:14 C2,C4,C4,C4	612	617	233, 345
S 4:15 C2,C4,C4,C5	626	631	247, 345
S 4:16 C2,C4,C5,C5	640	645	247, 359
S 4:17 C2,C5,C5,C5	654	659	247, 373
S 4:20 C2,C4,C4,C10	696	701	317, 345
S 4:21 C2,C4,C5,C10	710	715	317, 359
S 4:22 C2,C4,C4,C12	724	729	345, 345
S 4:23 C2,C4,C5,C12	738	743	345, 359
S 4:24 C2,C5,C5,C12	752	757	345, 373

Acylsugar metabolite and Fatty acid constituents	[M+HCOO] ⁻ (<i>m/z</i>)	[M+ ³⁵ Cl] ⁻ (<i>m/z</i>)	Fragment ions (<i>m/z</i>)
S 4:14 C2,C4,C4,C4	639	629	593, 551, 87
S 4:15 C2,C4,C4,C5	653	643	607, 565, 101, 87
S 4:16 C2,C4,C5,C5	667	657	621, 579, 537, 495, 393, 341, 323, 101, 87
S 4:17 C2,C5,C5,C5	681	671	635, 551, 509, 425, 407, 341, 323, 101
S 4:20 C2,C4,C4,C10	723	713	677, 635, 481, 411, 393, 341, 323, 101, 87
S 4:21 C2,C4,C5,C10	737	727	691, 649, 495, 393, 341, 323, 171, 101, 87
S 4:22 C2,C4,C4,C12	751	741	709, 663, 481, 393, 341, 323, 305, 199, 87
S 4:23 C2,C4,C5,C12	765	755	719, 677, 495, 411, 341, 323, 199, 101, 87
S 4:24 C2,C5,C5,C12	779	769	733, 691, 509, 425, 407, 341, 323, 199, 101

Table 3-7. Exact mass measurements of $[M+HCOO]^-$ ions for detected sucrose tetraesters in *S. habrochaites* LA1777 detected using ESI negative mode with CID potential 10 V.

Acylsugar metabolite annotation	Formula of formate adduct	Measured (m/z)	Calculated m/z	ppm (error)
S 4:14	C ₂₇ H ₄₃ O ₁₇	639.2500	639.2520	3.1
S 4:15	C ₂₈ H ₄₅ O ₁₇	653.2657	653.2686	4.4
S 4:16	C ₂₉ H ₄₇ O ₁₇	667.2813	667.2815	0.3
S 4:17	C ₃₀ H ₄₉ O ₁₇	681.2970	681.2992	3.2
S 4:19	C ₃₂ H ₅₃ O ₁₇	709.3283	709.3290	2.5
S 4:20	C ₃₃ H ₅₅ O ₁₇	723.3439	723.3432	-1.0
S 4:21	C ₃₄ H ₅₇ O ₁₇	737.3596	737.3577	-2.6
S 4:22	C ₃₅ H ₅₉ O ₁₇	751.3752	751.3737	-2.0
S 4:23	C ₃₆ H ₆₁ O ₁₇	765.3909	765.3875	-4.4
S 4:24	C ₃₇ H ₆₃ O ₁₇	779.4065	779.4094	3.7

In order to provide additional information about structures of fatty acyl groups in acylsugars from *S. habrochaites* LA1777 detected using LC/MS and multiplexed CID, GC/MS analyses were performed for the composite mixture of acylsugar fatty acids in this accession after transesterification to ethyl esters (Figure 3-27). The result confirmed the LC/MS findings that C4, C5, C10, and C12 were the dominant fatty acid groups. For this accession, the ratio of *iso*-C5 to *anteiso*-C5 was about 2:1, a value similar to the *S. pennellii* accession (LA1522), but the inverse of the ratio found in *S. pennellii* LA0716. The *iso*-C4 fatty acyl group was dominant in LA1777 and LA0716 (about 35% of total chains), but C4 was much less abundant (about 2% of total) in LA1522. For C10 and C12 fatty acyl groups, ratios of branched (*iso*) to linear fatty acyl chains showed intriguing variation. For *S. pennellii*, formation of the branched isomer was favored in LA0716 by about 2:1, whereas accession LA1522 favored the linear isomer (1:8 branched/linear). In

contrast, *S. habrochaites* LA1777 favored the branched C10 chain by about 11:1. Corresponding ratios for the C12 isomers followed a different pattern. For *S. pennellii*, accession LA0716 favored the linear isomer (1:4 branched/linear ratio), and only the linear C12 isomer was observed in LA1522. For *S. habrochaites* LA1777, the branched C12 isomer was favored over the linear chain by 2:1. These findings suggest that all three accessions synthesize and attach the same fatty acyl groups onto sugars, with the exception that LA1522 does not make acylsugars with branched C12 fatty acyl attachments. However, the relative amounts of the different fatty acid groups show marked diversity among these three accessions. It is tempting to attribute the variability in fatty acyl substitution to differences in substrate selectivity during either fatty acid biosynthesis, attachment of fatty acyl groups to sugars, or a combination of both. The major conclusion from this exploration of acylsugar structure points to substantial chemical diversity in acylsugar composition. These findings will be explored in further detail in Chapter 6.

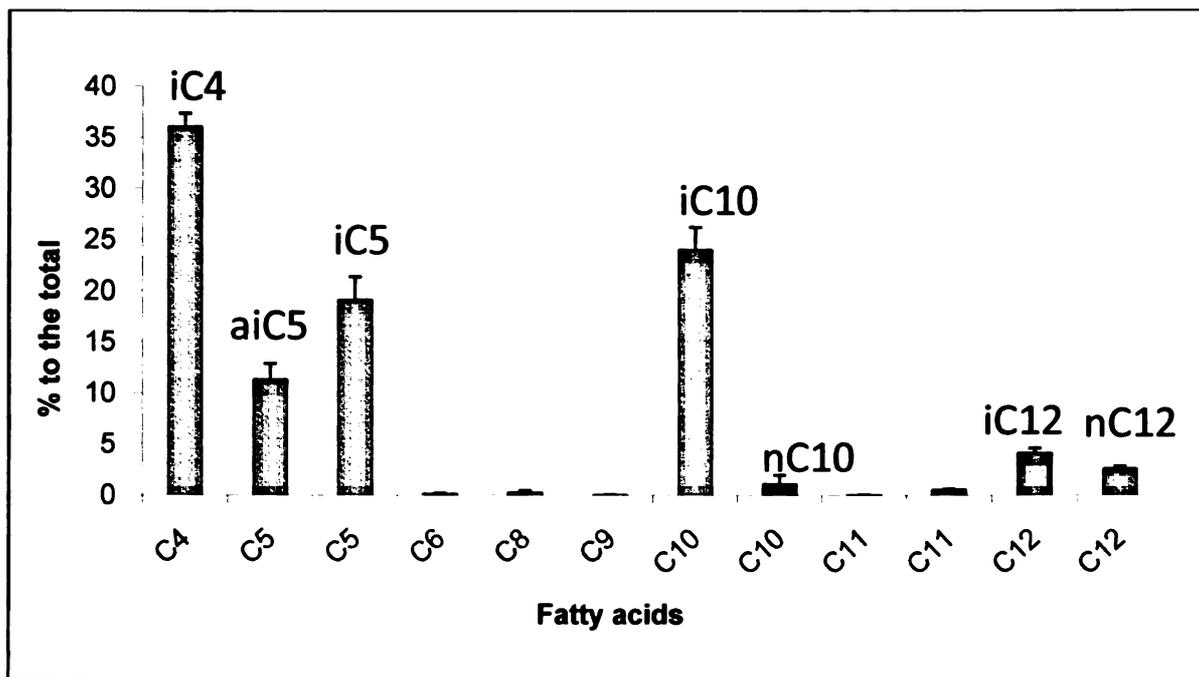


Figure 3-27 Fatty acyl groups from total acylsugars in an extract of *S. habrochaites* LA1777 were transesterified to form fatty acid ethyl esters, and analyzed using GC/MS. Levels for each fatty acid were calculated as a percentage of the total ion current chromatogram peak areas of fatty acids detected through GC/MS. The abbreviations “i” and “ai” refer to *iso*- and *anteiso*- branched isomers, respectively. Fatty acids with the same number of carbon atoms are listed in order of increasing retention time from left to right. Error bars indicate standard error with $n = 3$.

3.3.4 Structure elucidation of sucrose triesters and tetraesters in *S. habrochaites* LA1353.

Another *S. habrochaites* accession, LA1353, was extracted, and metabolite profiles were generated to broaden our knowledge about acylsugar diversity, especially since there were no published reports about acylsugars for this accession. Analyses using LC/MS suggested that comparable amounts of sucrose tri- and tetraesters are accumulated by LA1353, but acylglucoses were not detected. Table 3-8 displays all the detected acylsugars from *S. habrochaites* LA1353. Fatty acid ester constituents were C4,

C5, C10, C11, and C12, but no acetate ester were present and only one chain of C10 or longer was present in any of the tri- or tetraesters. Accurate mass measurements and a relatively short chromatographic retention time identified one acylsugar as a pentaester, with five C5 chains. No other pentaesters were observed in substantial abundance. For this accession, GC/MS analyses were not performed, so the composition of linear and branched isomeric fatty acid groups remains undetermined.

Positive ion CID spectra indicated that all triester fatty acid groups were substituted on the glucopyranose ring. All tetraesters contained a single C5 fatty acid on the fructofuranose ring (evident as m/z 247 fragment), with all three other ester groups, including all C10, C11, and C12 groups, were attached on the glucopyranose moiety. The tetraesters correspond to observed sucrose triesters modified by addition of one C5 fatty acyl group on the fructofuranose ring. For the pentaester (S 5:25), two fatty acyl groups reside on the fructofuranose ring (evidence as m/z 313 fragment), while three substitutions are on the glucopyranose ring, based on the fragment at m/z 415. Specific positions of substitutions of individual ester groups remain unknown, pending metabolite purification and NMR characterization.

Fatty acyl substitution patterns in *S. habrochaites* LA1353 are distinct from all other accessions described above. For sucrose triesters, all fatty acyl groups are on glucopyranose. This pattern differs from LA1777 where one group (including the longer chains) is on the fructofuranose ring, but is similar to *S. pennellii* LA1522. For sucrose tetraesters, accession LA1353 always has the longer chain esters on the glucopyranose ring, a pattern differing from LA1777. Accession LA1522 did not produce tetraesters.

Substitution of two fatty ester groups, both C5, on the fructofuranose ring is a trait unique to LA1353.

Table 3-8. Fragments and fatty acid constituents of detected acylsugars from *S. habrochaites* LA1353 using ESI positive mode (CID potential, 40 V) and negative (CID potential, 55 V).

Acylsugar metabolite annotation	Fatty acid constituents	[M+NH ₄] ⁺ (m/z)	[M+Na] ⁺ (m/z)	Fragment ions (m/z)
S 3:19	C4,C5,C10	668	673	511,339,251
S 3:20	C5,C5,C10	682	687	525,485,353,313,229
S 3:21	C5,C5,C11	696	701	539,499,397,313,211
S 3:22	C5,C5,C12	710	715	553,513,353,313,211
S 4:19	C4,C5,C5,C5	682	687	247,441
S 4:20	C5,C5,C5,C5	696	701	247,455
S 4:24	C4,C5,C5,C10	752	757	247,511
S 4:25	C5,C5,C5,C10	766	771	247,525
S 4:26	C5,C5,C5,C11	780	785	247,539
S 4:27	C5,C5,C5,C12	794	799	247,553
S 5:25	C5,C5,C5,C5,C5	780	785	331,415

Acylsugar metabolite annotation	Fatty acid constituents	[M+HCOO] ⁻ (m/z)	[M+ ³⁵ Cl] ⁻ (m/z)	Fragment ions (m/z)
S 3:19	C4,C5,C10	695	685	565,495,425,407,341,171,101
S 3:20	C5,C5,C10	709	699	579,495,425,341,171,101
S 3:21	C5,C5,C11	723	713	593,509,425,407,341,323,185,101
S 3:22	C5,C5,C12	737	727	607,509,425,407,341,323,199,101
S 4:19	C4,C5,C5,C5	709	699	663,579,495,425,341
S 4:20	C5,C5,C5,C5	723	713	677,593,509,425,341,101
S 4:24	C4,C5,C5,C10	779	769	677,593,509,425,407,341,171,101
S 4:25	C5,C5,C5,C10	793	783	663,579,425,341,171,101
S 4:26	C5,C5,C5,C11	807	797	761,677,593,509,425,185,341,101
S 4:27	C5,C5,C5,C12	821	811	775,691,607,509,425,341,323,199,101
S 5:25	C5,C5,C5,C5,C5	807	797	761,677,593,509,425,407,341,323,101

Accurate mass measurements played a key role in the identification of the sucrose tetraester S5:25 from accession LA1353. Two sucrose esters with the same formate adduct nominal mass (m/z 807) were observed, but subtle mass differences allowed them to be assigned as S 4:26 (5,5,5,11) and S 5:25 (5,5,5,5,5) (Figure 3-28). Additional support for these assignments came from differences in fragment masses (36 mDa theoretical difference between isobaric fragments from tetra- and penta-esters). These small mass differences facilitate data mining for acylsugars with different numbers of ester groups via generation of extracted ion chromatograms using a narrow mass window, such as 10 mDa.

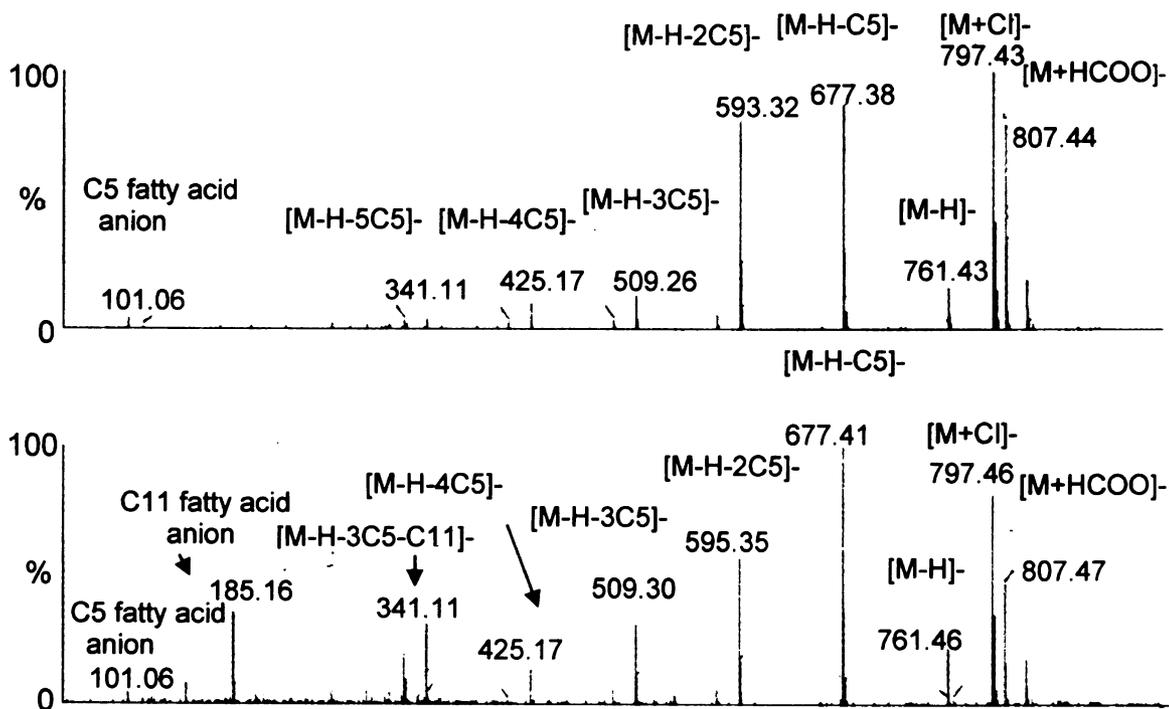
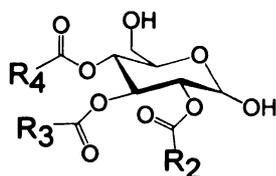


Figure 3-28 (Top) Negative mode ESI spectra of S 5:25 (RT = 31.91 min) and (Bottom) S 4:26 (RT = 36.04 min) from *S. habrochaites* LA1353 using CID potential 55 V.

This survey of two accessions from each of two plant species (*S. pennellii* and *S. habrochaites*) showed dramatic differences in acylsugar fatty acyl chains and positions of fatty ester substitutions. To facilitate comparisons of these genotypes, the structures of acylsugars are summarized in Figure 3-29.

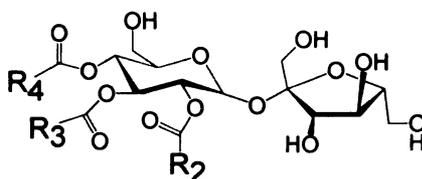
Figure 3-29 Putative structures for detected acylsugars among *S. pennellii* LA0716 and LA1522 and *S. habrochaites* LA1777 and LA1353. Specific substitution positions of individual fatty acids remain uncertain for some metabolites.

S. pennellii
LA0716



Glucose triesters

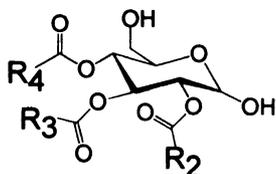
$R_2, R_3, R_4 = C_4, C_5, C_8, C_9, C_{10}, C_{11}, C_{12}$
Only one chain C8 or longer



Sucrose triesters

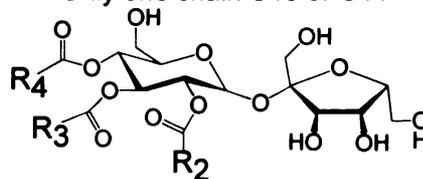
$R_2, R_3, R_4 = C_4, C_5, C_{10}, C_{11}$
Only one chain C10 or C11

S. pennellii
LA1522



Glucose triesters

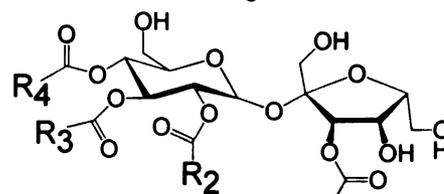
$R_2, R_3, R_4 = C_4, C_5, C_6, C_8, C_9, C_{10}, C_{12}$
Only one chain C8 or longer



Sucrose triesters

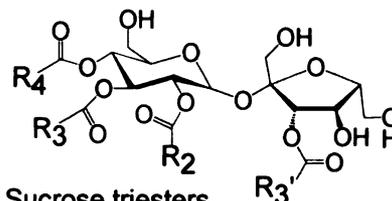
$R_2, R_3, R_4 = C_5, C_6, C_8, C_9, C_{10}, C_{12}$
Only one chain C8 or longer, no R3' substitution

S. habrochaites
LA1777



Sucrose tetraesters

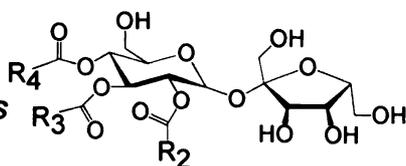
$R_2 = CH_3; R_3, R_4, R_3' = C_4, C_5, C_{10}, C_{12}$
Only one chain C10 or longer



Sucrose triesters

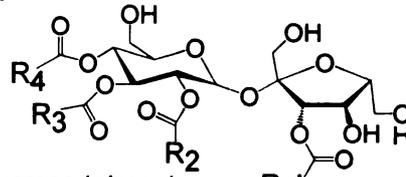
$R_3, R_4, R_3' = C_4, C_5, C_{10}, C_{12}$
Only one chain C10 or C12

S. habrochaites
LA1353



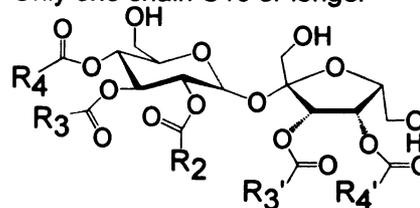
Sucrose triesters

$R_2, R_3, R_4 = C_4, C_5, C_{10}, C_{11}, C_{12}$
Only one chain C10 or longer



Sucrose tetraesters

$R_3 = C_5; R_2, R_3, R_4 = C_4, C_5, C_{10}, C_{11}, C_{12}$
Only one chain C10 or longer



Sucrose pentaesters

$R_2 = R_3 = R_4 = R_3' = R_4' = C_5$

3.4 Conclusions

This research has presented application of LC/TOF with multiplexed CID for comprehensive profiling of acylsugar metabolites extracted from trichomes of four wild tomato relatives. This approach has generated accurate masses for both adduct and fragment ions, yielding information content similar to MS/MS spectra but without the loss of information that can occur during LC/MS/MS analyses. These analyses led to detection of about 50 different kinds of acylsugars, not including isomers within specific accessions. The analyses were performed using a relatively inexpensive model of mass spectrometer, and the successes of this work suggest that laboratories that do not enjoy access to more expensive instruments can make substantial contributions to metabolite discovery.

However, the exact positions of fatty acyl chains on acylsugars and assignment of fatty acyl groups as branched or straight chains remain to be determined. Efforts to produce position-specific fragment ions through metal cationization yielded unsatisfactory results owing to facile losses of fatty acid groups during collision induced dissociation. Purification and characterization using NMR are underway in the Jones laboratory. Even though complete structural details for acylsugar metabolites have not yet been achieved, these findings raise awareness of diversity of secondary metabolites in the genus *Solanum*. More extensive profiling of other accessions, not described in this dissertation, has shown that every accession displays a profile of secondary metabolites that distinguish it from all others. Experimental determination of metabolic diversity provides researchers with an opening to discover how genes and environment combine to produce different biochemical outcomes, and rapid screening for differences in plant

chemistry can guide efforts toward gene discovery.¹⁸⁴ Future research should exploit information about the secondary metabolome to establish a scientific foundation for understanding of biological roles of genes and metabolites.

Chapter 4. Annotation of polyphenol, glycoalkaloid, and terpenoid secondary metabolites in glandular trichomes from tomato, its wild relatives, and near isogenic lines, based on LC/TOF MS coupled with multiplexed CID

4.1 Introduction

Glandular trichomes of wild relatives of tomato produce secondary metabolites in abundance. Chapter 3 of this dissertation described annotations of a diverse set of acylsugar metabolites, and in this chapter, additional metabolites from other compound classes are the focus.

Three classes of secondary metabolites, the polyphenols, glycoalkaloids and terpenoids, are the most recognized chemical defenses used by plants as protection against insects and pathogens.^{123,185,186} These compounds also have been touted as providing benefits, in the case for of polyphenolic flavonoid antioxidants, and harm, as observed with teratogenic behavior of glycoalkaloids, to human diet and health.¹²² The biological potency of these compounds lends impetus to efforts aimed at discovering genes responsible for their biosynthesis in plants. All three classes of metabolites present potential scaffolds for development of novel pharmaceutical and nutraceutical agents. Complete genetic maps of the enzymes involved in their biosynthesis are not available for any of these classes of metabolites. Glandular trichomes are prolific chemical factories known to synthesize chemical defense compounds in plants, and are therefore attractive targets for investigating relationships between gene expression and chemical composition.

Flavonoids are a prominent group of polyphenols that consist of two aromatic rings with six carbon atoms (A and B ring) and linked through a heterocyclic six-membered ring (C ring)¹⁸⁷ containing an ether oxygen (Figure 4-1). In tomato and its relatives, structural diversity among flavonoids is due to differences in positions of hydroxyl substitution, and the presence or absence of methylated, acylated and glycosylated

hydroxyl groups.¹⁸⁸ Owing to their importance in tomato flavor (they impart bitterness), flavonoid content has primarily been investigated in fruits of tomato plants, where flavonoid glycosides are abundant and have been catalogued in a tomato metabolite database.¹⁰³ This database reported that observed glycosylation primarily occurred at the 3-position. Flavonoids and their structurally related derivatives^{126,189} are important secondary metabolites in trichomes of *Solanum* species that contribute color to plant organs, protect against biotic and abiotic stress, and regulate normal plant development.¹⁹⁰ However, closely related plants often differ in flavonoid profiles, and more advances are needed to identify the genetic basis for differences in flavonoid composition in specific plants and cell types.

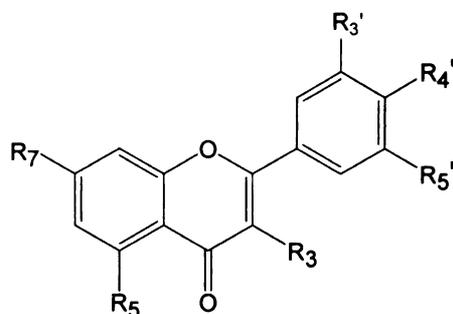


Figure 4-1. Generalized structure of flavonoid metabolites known in tomato and its relatives, where R groups can be hydrogen, hydroxyl, or other substituted oxygen-containing groups.

Glycoalkaloids are secondary metabolites common in the family Solanaceae that contain a basic nitrogen in a heterocyclic ring and are conjugated to an assortment of carbohydrates. These compounds are produced by a variety of Solanaceous plants, including potato, tomato and eggplant,¹⁹¹ and largely differ in the structure of the

oligosaccharide and in the number of double bonds (Figure 4-2). Glycoalkaloids are bioactive secondary metabolites with toxicity to bacteria, fungi, viruses, and insects, which play an important role in protection of disease in plants.¹⁹² In tomato, glycoalkaloids include α -tomatine and dehydrotomatine, which contain a spiro-ring geometry somewhat unusual in plant metabolites, and exhibit an assortment of subtle structural differences relative to chaconine or solanine and leptine in potato or solamargine and solasonine in eggplant.¹⁹² Alkaloids are reported to have a variety of pharmacological and nutritional properties in animal and humans, thus, we need a better understanding of the role of these compounds both in the plant and in the diet.^{127,128} Although much effort has been put into study of biosynthesis of glycoalkaloids,¹⁹³ many biosynthetic steps remain unknown, especially the formation of the nitrogen heterocyclic F ring of tomatine.

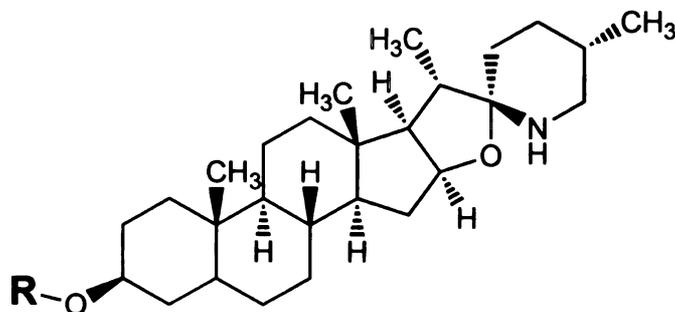


Figure 4-2. Generalized structure of glycoalkaloid metabolites known in tomato and its relatives, where R groups are glycosides.

Terpenes, the most abundant and structurally diverse group of plant secondary metabolites,¹²⁰ consist in tomato largely of the C10 mono-, and C15 sesquiterpenes.^{8,194}

Mono- and sesqui-terpenes are abundant and volatile metabolites that accumulate in tomato glandular trichomes.¹⁹⁵ Aside from these volatiles, relatively few oxidized terpenoid metabolites have been observed in tomato and its relatives, with sesquiterpene carboxylic acids from *S. habrochaites* LA1777 being an exception.^{196,197} The potential diversity of terpenoids arises from different precursors, multiple structures that can arise during cyclization, and variations in substitution patterns of metabolic oxidation. Metabolite identifications offer potential for accelerating discoveries of gene functions, and findings of this kind help integrate metabolite information into system-level studies of how plants synthesize defense chemicals.¹⁹⁸

One long standing approach used for discovery of gene functions is called forward genetics, which involves searching for genes associated with specific traits. For the purpose of this study, those traits are levels of individual secondary metabolites. In order for this approach to be efficient, it is necessary to perform deep and rapid screening of genetic variants for differences in phenotype, which in this case is plant chemistry. Once variations in chemistry are recognized in specific plants, the metabolite profiles serve to guide discovery of genes responsible for plant chemistry. These investigations often explore differences between naturally occurring plant species and between offspring produced by genetic crosses.

One of the most successful examples of the use of metabolite profiling for discovery of plant gene functions was reported by the group of Kazuki Saito of the RIKEN Plant Science Center.¹⁹⁹ In this study, metabolite profiles generated using combinations of HPLC and mass spectrometry were employed to establish functions for two *Arabidopsis* glycosyltransferases in the biosynthesis of flavonoid and anthocyanin glycosides.

In the studies described below, LC/TOF MS with nonselective and multiplexed collision induced dissociation (multiplexed CID) was developed and applied to compare metabolite profiles for tomato and several genetic variants. These include several wild tomato relatives and lines derived from crossing tomato (*S. lycopersicon* M82) with the wild tomato *S. pennellii* (LA0716), and generating genetic variants of tomato that contain small segments of genomic DNA from the wild tomato line. These new genotypes, termed introgression lines, were generated in the laboratory of Dani Zamir (REF), and they are useful for establishing roles of specific DNA regions on the chemical composition in specific plant tissues. Comparisons of metabolite profiles of conventional tomato with these introgression lines and wild tomato relatives help target subsequent gene mapping efforts to narrow regions of specific tomato chromosomes.

4.2 METHODS

4.2.1 Materials. Plants were grown in a greenhouse in Michigan State University from seeds obtained from C. M. Rick Tomato Genomics Resource Center at the University of California-Davis.

4.2.2 Chemicals. Acetonitrile (HPLC grade), 2-propanol (HPLC grade), methanol (HPLC grade), and formic acid 88% were obtained from VWR Scientific and Fisher Scientific.

4.2.3 Extraction Method. To extract metabolites from *Solanum* trichomes, ~100 mg of individual leaflets were dipped in 2 mL isopropanol: CH₃CN: H₂O (3:3:2 v/v/v) for 1 min. A 1.0 mL aliquot of each extract was transferred to a 1.5 mL polypropylene

microcentrifuge tube, and the tubes were centrifuged (5000 rpm) for 2 min at 25°C, and the supernatants were analyzed without further processing. Extracts were stored at -20°C.

4.2.4 High-Performance Liquid Chromatography. A Shimadzu (Columbia, MD) LC-20AD HPLC ternary pump was used. The separation was performed on a Thermo 1 × 150 mm BetaBasic C₁₈ column (100 × 1 mm, 3µm). Gradient elution was executed based on 0.15% aqueous formic acid (solvent A) and methanol (solvent B) as follows: initial conditions 5% B; linear gradient to 50% B (5 min), 95% B (33 min), and 100% B (35 min); hold at 100% B until 38 min, and return to 5% B (43 min). The flow rate was 0.1 ml·min⁻¹, and the injection volume was 10 µl. The column was maintained at 30 °C in a column oven.

4.2.5 Mass Spectrometry. Analyses were performed using a model LCT Premier mass spectrometer (Waters, Milford, MA). Operating parameters were set as follows: capillary voltage: 3.2 kV in both positive and negative mode; desolvation gas flow: 300 L·h⁻¹; desolvation gas temperature: 200 °C; source temperature: 90 °C; cone gas flow: 20 L·h⁻¹. The instrument was operated in both negative and positive ion electrospray modes using separate injections of extract for each ionization mode. Analyses were performed using rapid switching of Aperture 1 voltage in the ion transit region of the mass spectrometer, proving quasisimultaneous generation of spectra under fragmenting and nonfragmenting conditions. Mass spectra were acquired over *m/z* 50 to 1500 with a scan time of 0.4 s for each function. Spectra were acquired in centroid format using 'dynamic range enhancement' (DRE) enabled, which acquires separate mass spectra by switching a beam attenuator on and off, followed by post-acquisition stitching together of spectra. The instrument mass scale was calibrated daily in both ionization modes over *m/z* 50-1500

using 0.1% aqueous phosphoric acid as reference. Most analyses were performed using the V-mode ion pathway in the mass analyzer, which generated a resolving power of 5000 (FWHM). For some analyses requiring more exact mass measurements, 0.1% H₃PO₄ was infused through the reference sprayer of the lock-spray source as a reference lock mass, and the lockspray attachment switched between sample and reference streams.

4.3 RESULTS

4.3.1 Structural annotation for polyphenols

Flavonoid metabolites were detected in extracts of all of the plants described above, and their annotations were based on chromatographic retention times, accurate measurements of molecular and fragment ion masses, proposed elemental formulas derived from these measurements, and knowledge of previously-identified metabolites from the tomato fruit metabolite database (MoToDB).²⁰⁰ Flavonoid metabolites from three *Solanum* accessions consist of several flavonoid glycosides and methylated flavonoid aglycones as described below.

Putative structures for detected flavonoid glycosides are displayed in Figure 4-3. Rutin (quercetin-3-rutinoside) is used to illustrate how LC/TOF MS data were used to drive metabolite annotation. First, candidate polyphenol metabolite ions were distinguished in negative ion mode based on having lower mass defects (the value following the decimal place) than other metabolite ions of major abundance. Positive mass defects in the range of 100-200 mDa were considered likely candidates. Second, candidate flavonoid glycosides eluted earlier than acylsugar metabolites, making them easy to distinguish. Third, shapes of breakdown curves generated using multiplexed CID

(Figure 2-4 for rutin), were used to assign ions of major abundance as fragments, adducts or molecular ions for each candidate flavonoid glycoside. In one specific case, an ion at m/z 609.15 was most abundant at the lowest CID potential of 10 V, and gradually decreased in abundance with increasing CID potential, suggesting a pseudomolecular ion ([M-H]). Calculation of possible elemental formulas was performed using Masslynx software, using a mass window of 10 ppm, and the assumption that no nitrogen, sulfur, or phosphorus atoms were present. Only one elemental formula, $C_{27}H_{29}O_{16}^-$ was consistent with the measured m/z . The calculated value for the double bond equivalents (DBE, corresponding to the number of C atoms minus half the number of H atoms, plus 1) for this formula was 13.5, which indicated it had numerous rings plus double bonds, as would be consistent with an aromatic compound. Additional evidence about glycoside structures came from masses of fragments in both negative and positive modes. In the latter, fragments at m/z 465 corresponding to loss anhydrorhamnose (146 Da) and m/z 303 (loss of anhydrorhamnosylglucose) were suggestive of glycosylation pattern consisting of a rhamnosylglucose attachment (Figure 4-4). In negative mode, both sugar groups were eliminated together, showing fragments for this metabolite at m/z 301 and 300 (Figure 4-5). These ions are annotated as quercetin aglycone [M-H-anhydrorhamnosylglucose] $^-$ and a corresponding aglycone radical anion, respectively. Aglycone refers non-sugar compounds after replacement of glycosyl group from a glycoside by a hydrogen atom. Such fragment ions are consistent with earlier work from the laboratory of Magda Claeys.²⁰¹ The facile formation of the radical anion (m/z 300) by homolytic cleavage between sugar and aglycone is common for flavonoids glycosylated in the 3-position, as many of these flavonoids form stable radical anion fragments. Based

on these results, this metabolite was annotated as quercetin-3-rutinoside (rutin). Figure 4-6 displays proposed fragmentation pathway for rutin (quercetin-3-rutinoside) using negative mode.

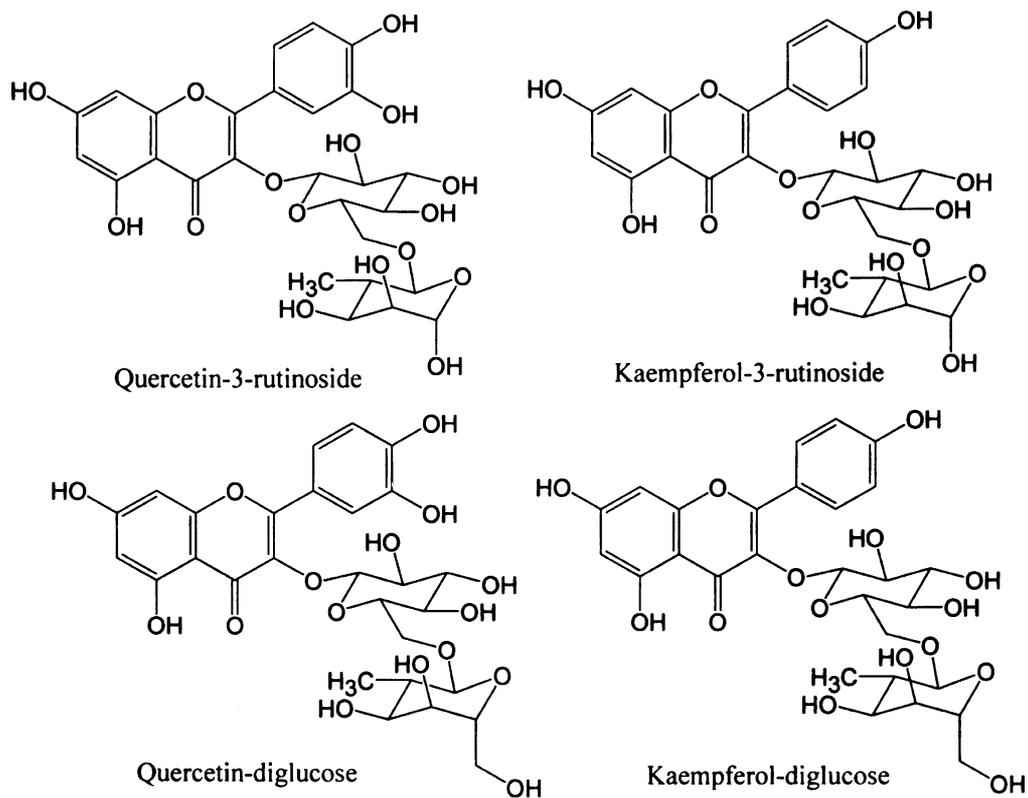


Figure 4-3 Putative structures for several flavonoid derivatives from *Solanum* wild species.

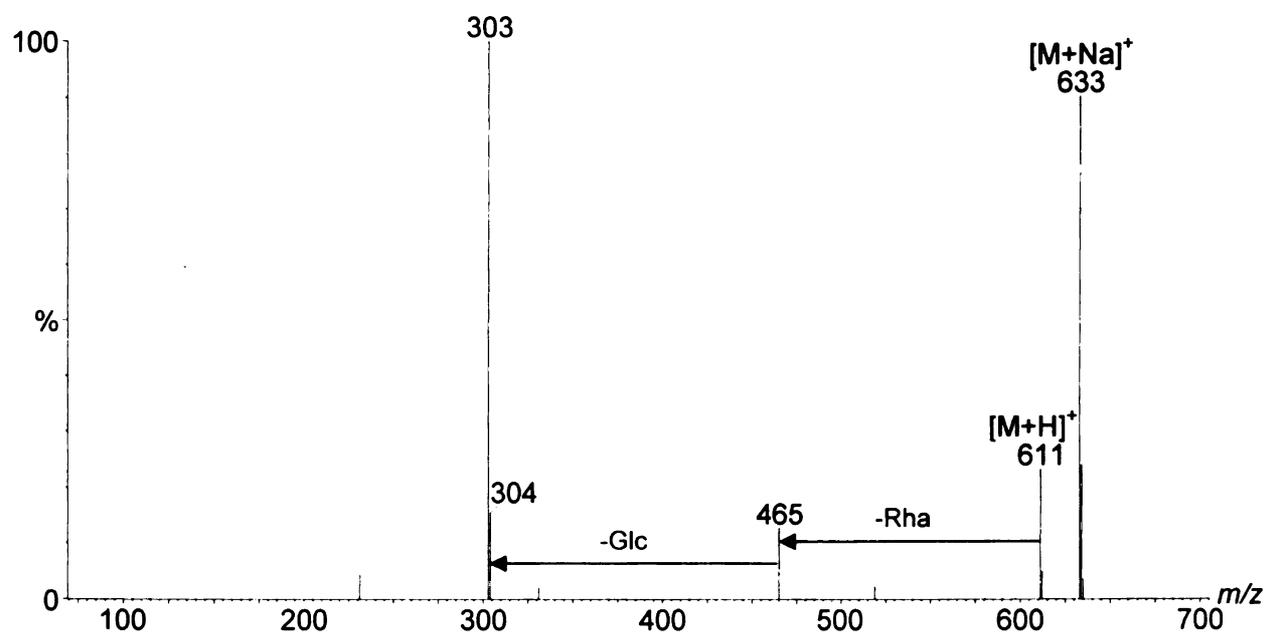


Figure 4-4 ESI mass spectra for quercetin-3-rutinoside from *S. pennellii* LA0716 using positive mode with CID 40 V.

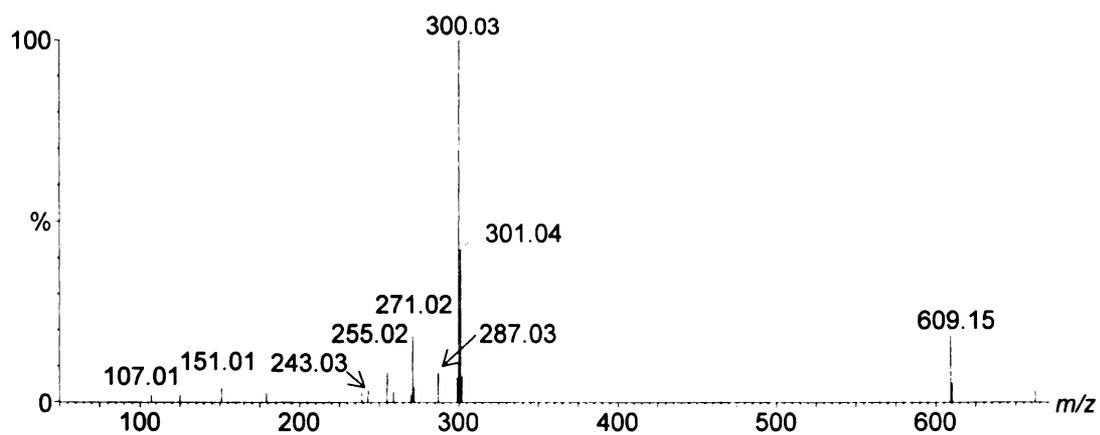
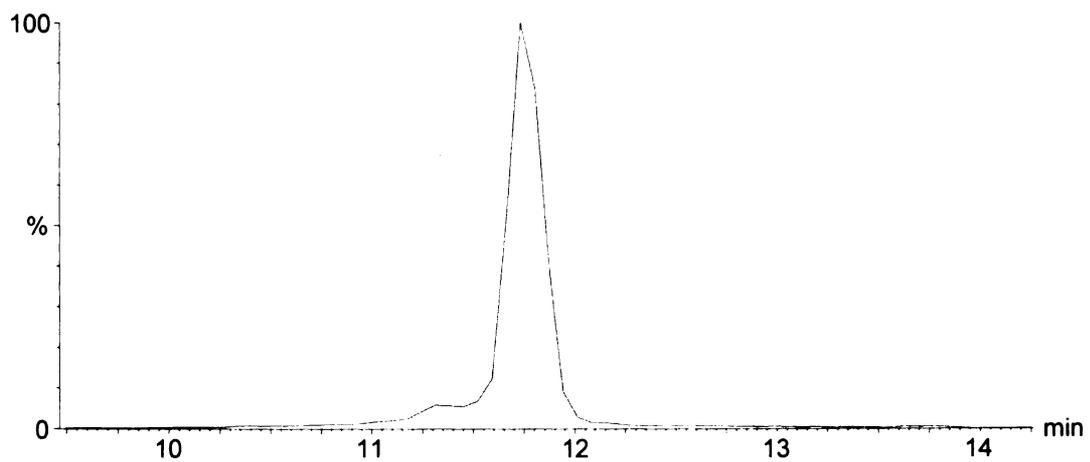


Figure 4-5 (Top) XIC for m/z 609 from *S. pennellii* LA0716 using ESI negative mode with CID voltage 10 V. (Bottom) ESI spectrum of quercetin-3-rutinoside (RT = 11.52 min) using ESI negative mode with CID voltage 55 V.

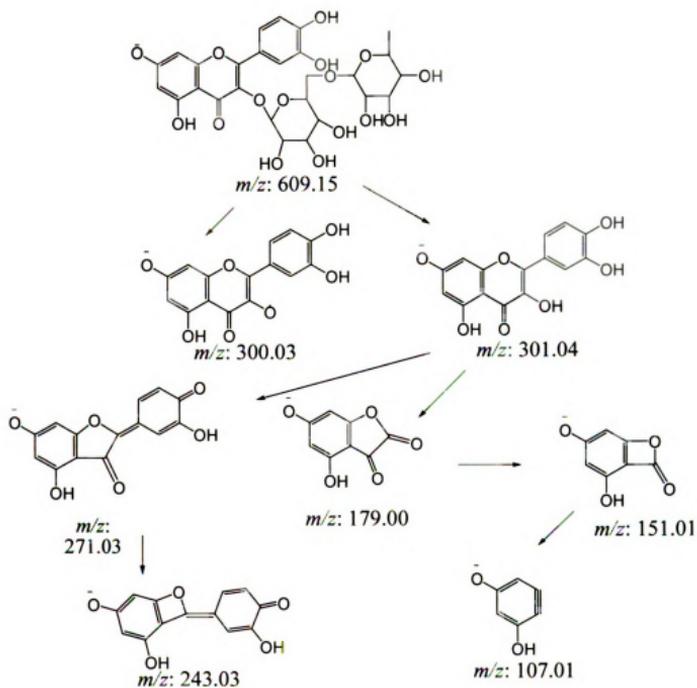


Figure 4-6 Proposed fragmentation pathway for quercetin-3-rutinoside from *S. pennellii* LA0716 using ESI negative mode with CID potential 55 V.

Two chromatographic peaks exhibiting the same exact mass at m/z 609.15 in negative ion mode (Figure 4-7) were observed in an extract of *S. habrochaites* LA1777. Based on exact mass measurement and fragment ion masses, the second peak was assigned as quercetin-3-rutinoside due to characteristic fragment ions at m/z 300.03 and 301.04 for this metabolite. In contrast, the later eluting metabolite yielded fragment ions at m/z 284.03 and 285.04 consistent with the aglycone of kaempferol, 16 Da less than aglycone of quercetin. These findings together suggested the major and earlier eluting metabolite was kaempferol-3-glucosylglucoside, although the mass spectra cannot distinguish the various hexose isomers. Additional support for this annotation is based on positive mode CID results including fragments at m/z 449.03 (-162, loss of one hexose) and 287.04 (-162-162, loss of two hexoses).

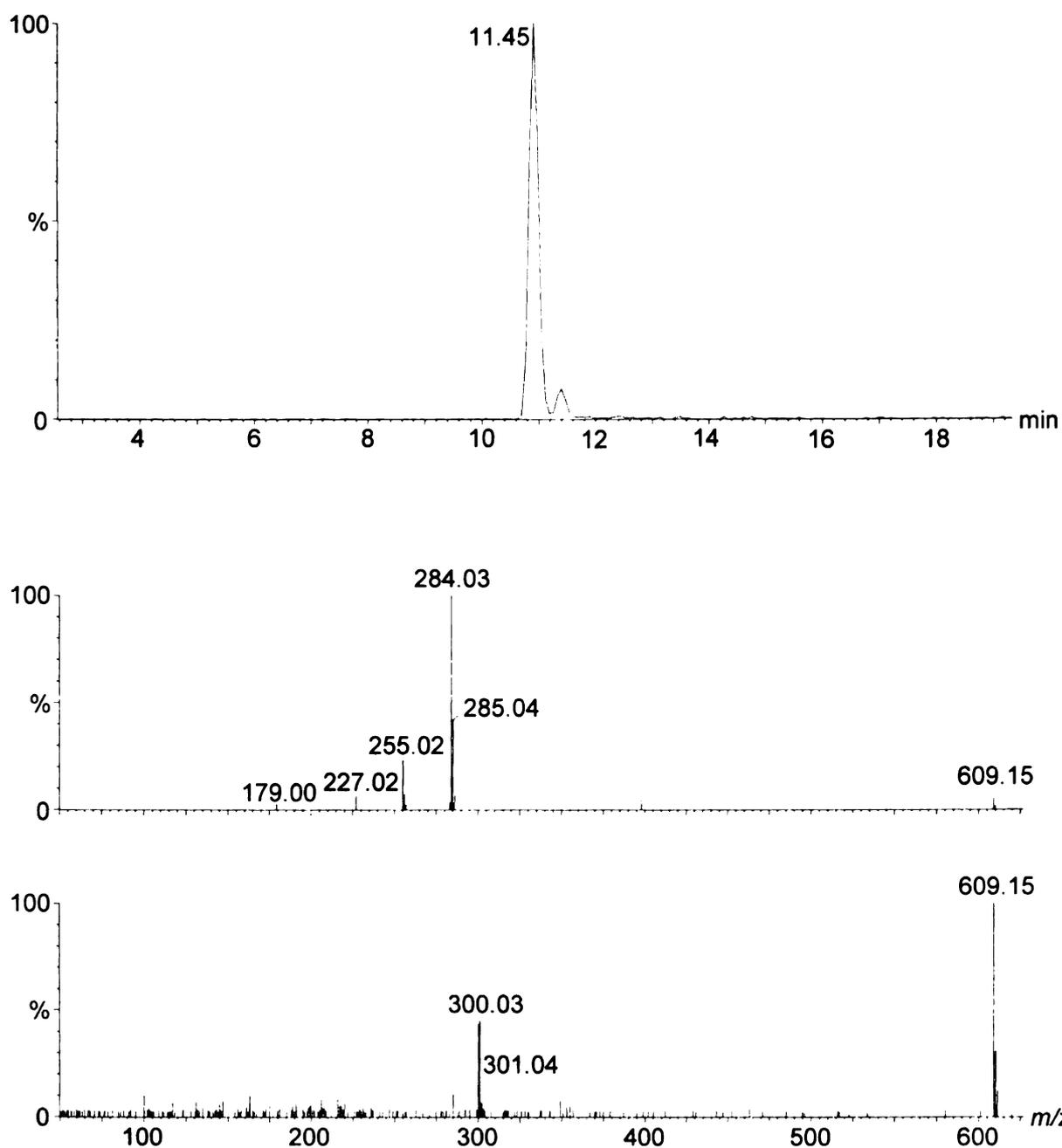
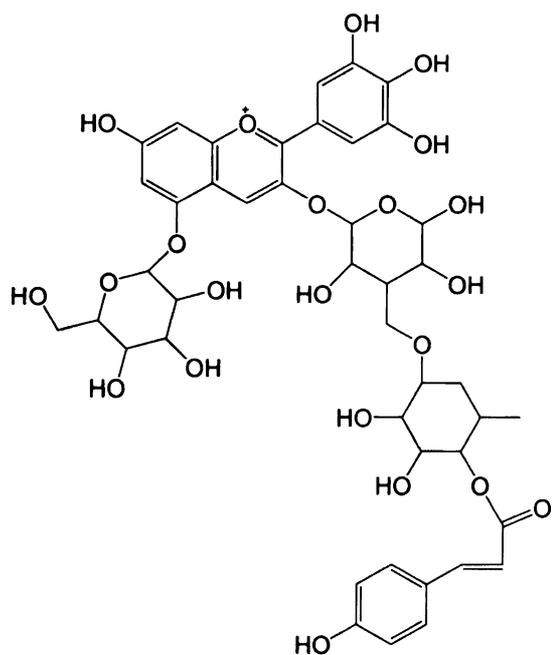


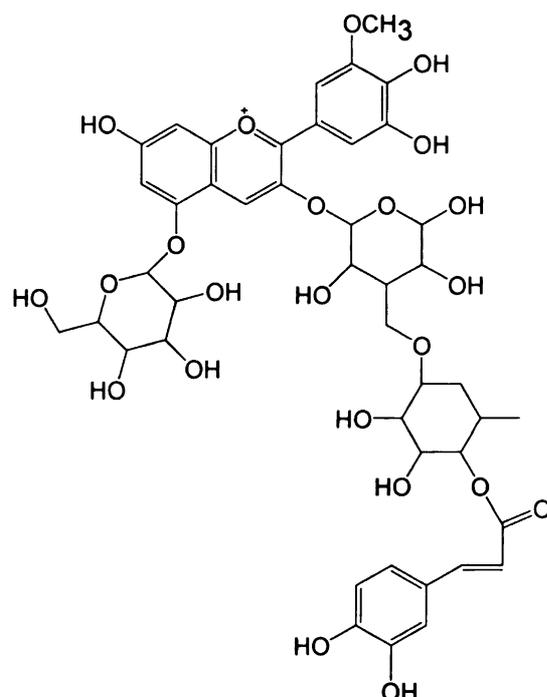
Figure 4-7 (Top) Two isomeric metabolites appear in the extracted ion chromatogram m/z 609 for an extract of *S. habrochaites* LA1777 using CID voltage 10 V. (Middle) ESI spectrum of first peak at 11.04 min for kaempferol-3-glucosylglucoside at CID voltage of 55 V. (Bottom) ESI spectrum of second peak (RT = 11.45 min) annotated as quercetin-3-rutinoside using ESI negative mode with CID voltage 55 V.

In addition to flavonoid glycosides, anthocyanins are also observed in leaf tissue from *Solanum* species. Anthocyanins possess a permanent positive charge, and are best characterized using positive mode ESI. Structures of anthocyanins from extracts of glandular trichomes of tomato and its relatives were annotated based on comparisons of exact masses and fragmentation patterns with metabolites in the MoTo tomato fruit metabolite database.²⁰² Putative structures of four anthocyanins were detected and displayed in Figure 4-8. They were assigned as coumaroyl esters of glycosides of petunidin, malvidin, and delphinidin and a caffeoyl ester of petunidin glycosides based on ions at m/z 933.27, 947.28, 919.25 and 946.26 respectively. In all cases, the masses and CID spectra were consistent with glucose conjugation at the 5-position and conjugation by the disaccharide rhamnosylglucose at the 3-position. The phenolic acid esters, *p*-coumaroyl and caffeoyl, are attached on the disaccharide. Anthocyanin aglycones were annotated based on fragment ions at m/z 303, 317 and 331 corresponding to delphinidin, petunidin and malvidin respectively. For example, for petunidin-3-O-(*p*-coumaroyl)- and Malvidin-3-O-(*p*-coumaroyl)-rutinoside-5-O-glucoside, fragments ions at m/z 317.06 and 331.07 (Figure 4-9) were due to loss of glucoside (-162 Da), rutinoside (-162) and *p*-coumaroyl (-292) from m/z 933.27 and 947.28, so the aglycones for those two anthocyanins are determined as petunidin and malvidin, respectively.¹⁰³

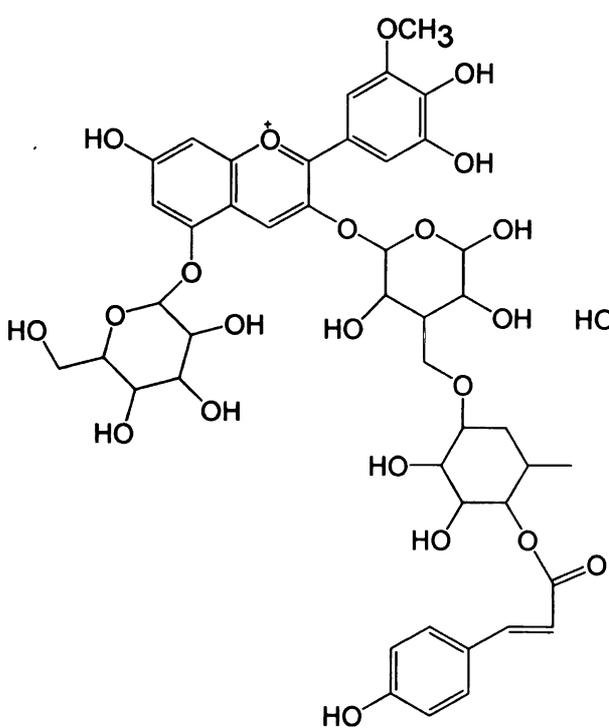
Figure 4-8 Putative structures of four anthocyanins petunidin-3-O-(p-coumaroyl)-rutinoside-5-O-glucoside (m/z 933), malvidin-3-O-(p-coumaroyl)-rutinoside-5-O-glucoside(m/z 947), delphinidin-3-O-(p-coumaroyl)-rutinoside-5-O-glucoside (m/z 919) and petunidin-3-(caffeoyl)-rutinoside-5-O-glucoside (m/z 949).



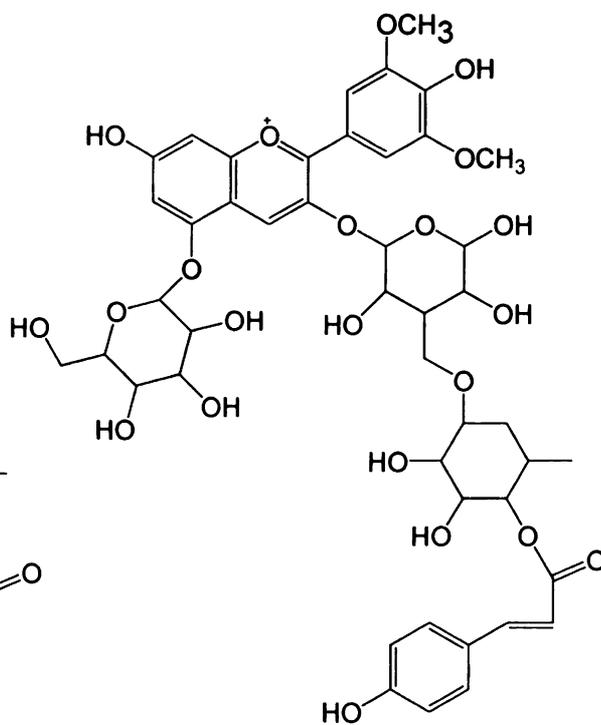
m/z: 919.25



m/z: 949.26



m/z: 933.27



m/z: 947.28

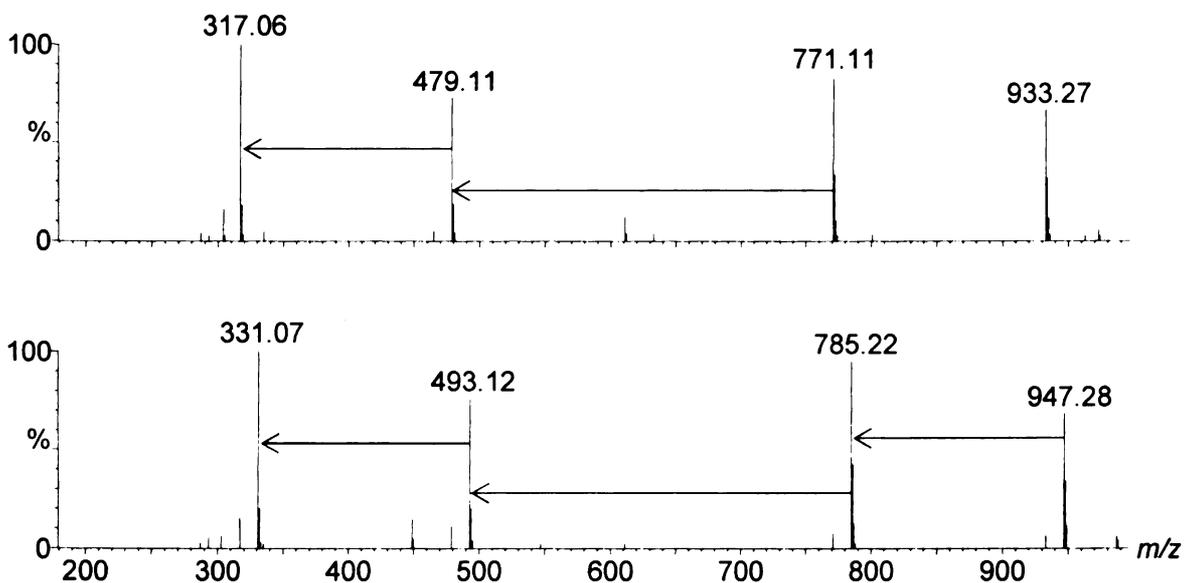


Figure 4-9 ESI mass spectra from (top) anthocyanins at m/z 933 (RT = 10.97 min) and (bottom) m/z 947 (RT = 11.25 min) from *S. lycopersicum* M82 using positive mode with CID 40 V.

Nontargeted analysis and annotation of polyphenolic metabolites is critical to establish whether they represent end products such as flavonoid glycosides, products of competing side reactions such as anthocyanins, or intermediates that accumulate owing to inefficient catalysis of biosynthetic steps. Although flavonoid biosynthesis has been widely studied,²⁰³⁻²⁰⁵ identification of all genes involved of the flavonoid pathway, including metabolic modification of flavonoids, has yet to be achieved in specific plants and cell types.⁴³ A generalized depiction of the biosynthetic steps in synthesis of flavonoids and anthocyanins is presented in Figure 4-10.

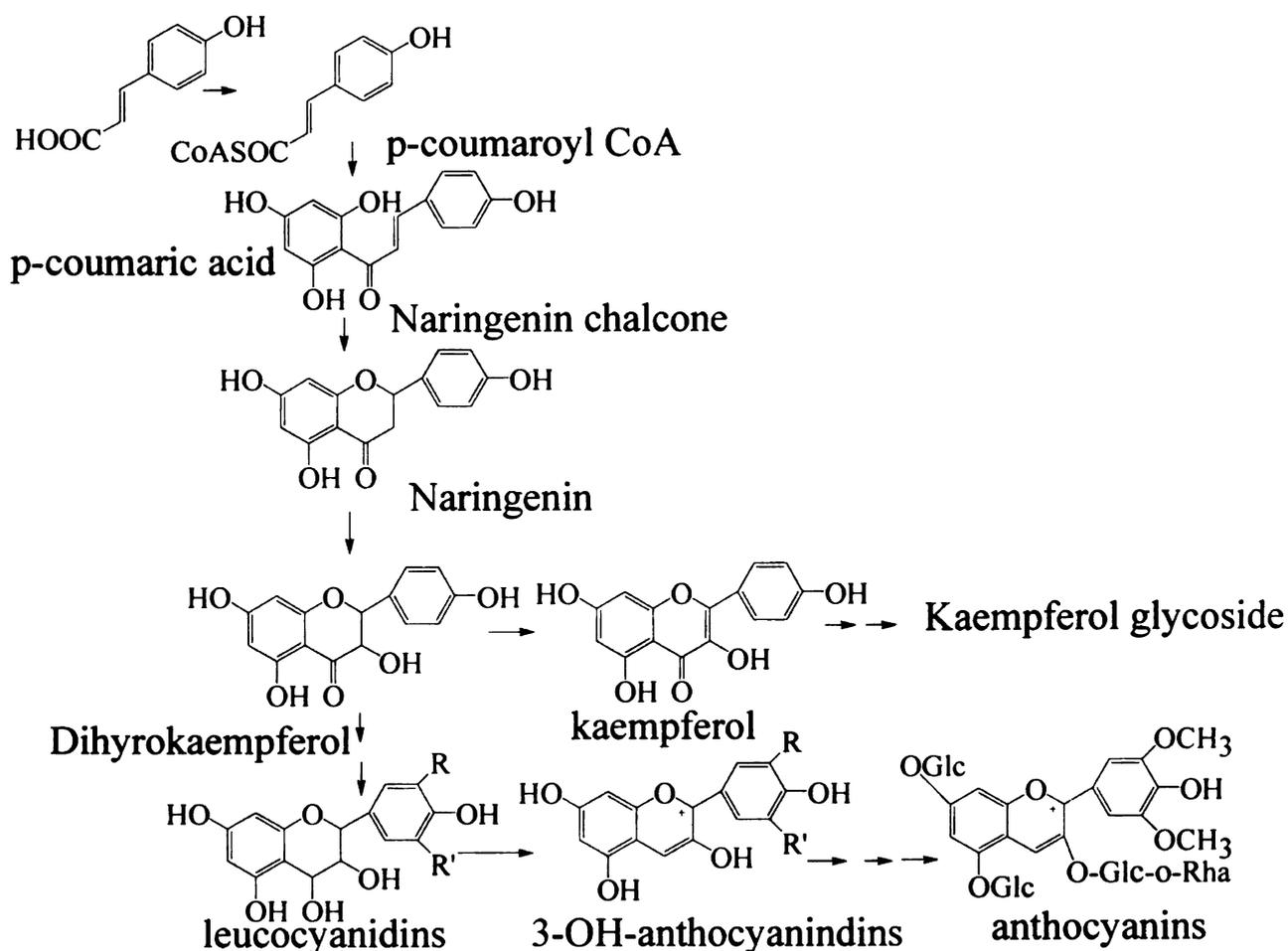


Figure 4-10 Important steps in the biosynthesis of anthocyanins and flavonoids (adapted from B. Winkel-Shirley and K. Saito^{43,206}).

Parallel investigations were conducted in collaboration with the laboratory of Professor Gregg Howe of Michigan State University to explore the role of trichome chemistry in susceptibility of plants to herbivory of insects using wild type tomato and mutant *af* (anthocyanin free). The mutant line, which was recognized by its lack of red pigmentation, was more susceptible to herbivory by the insect *Manduca sexta*, and the biological findings are described in more detail elsewhere (Kang et al., in preparation). Metabolites extracted from leaf trichomes by solvent dip were analyzed using LC/TOF

MS and multiplexed CID. An assortment of glycosides of the flavonols quercetin and kaempferol were detected in both wild type and mutant plants, with quercetin-3-rhamnoside the most abundant. Levels of all flavonoid glycosides, were about 6-fold lower in the *af* mutant, and anthocyanins were below detectable levels in the mutant. Metabolites with masses consistent with aglycones and glycoconjugates of the flavonoid precursors naringenin and/or naringenin chalcone were detected in wild type but not in the *af* mutant. These results suggest some steps in flavonoid pathway are inefficient in the *af* mutant, but this mutation does not completely eliminate flavonoid biosynthesis. Further investigations of differences in secondary metabolite profiles between wild type and *af* mutant are continuing, with one goal being the discovery of tomato genes that might contribute greater insect resistance in tomato plants.

Flavonoids are often present as flavonoid glycosides in tomato plant flowers, leaves, stems or roots.²⁰¹ No flavonoid aglycones have been reported in tomato tissues, however, based on our results, various methylated aglycones of flavonoids were found in extracts of trichomes from *S. habrochaites* LA1777 and *S. lycopersicum* M82 (commercial tomato). Figures 4-11 displays extracted ion chromatograms for pseudomolecular ions of methylated myricetins in an extract of *S. habrochaites* LA1777, with masses corresponding to 2-5 methyl groups.

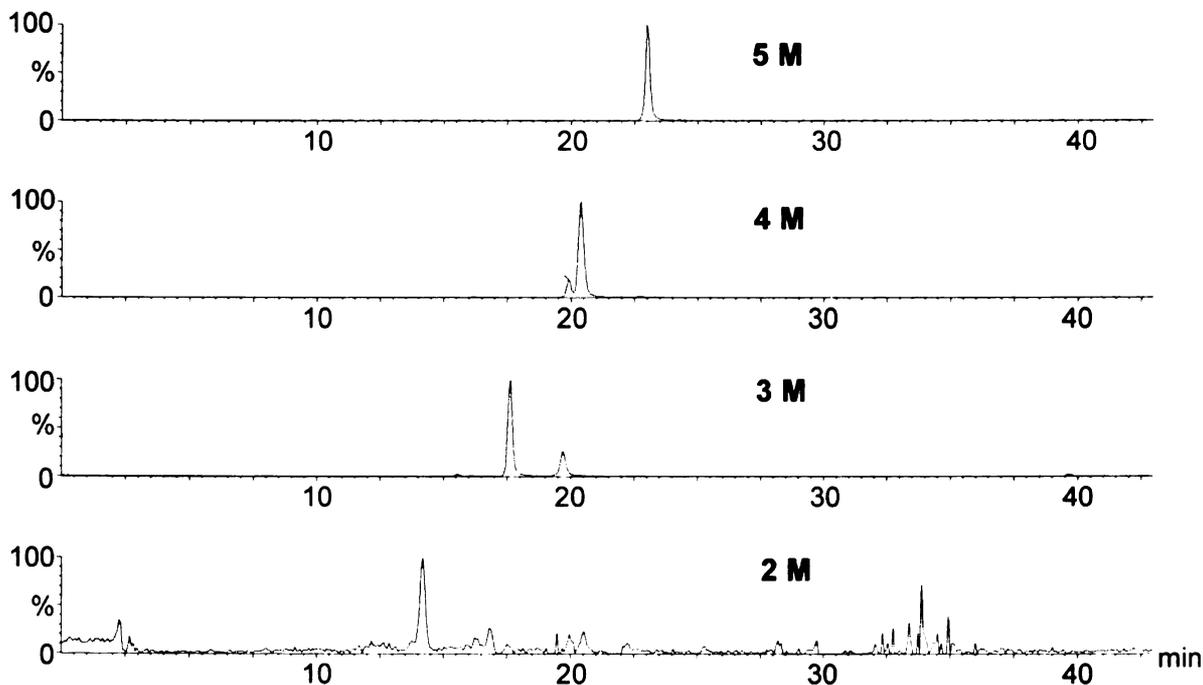


Figure 4-11 LC/MS XICs for a trichome extract from *S. habrochaites* LA1777 using ESI positive mode with CID voltage 10 V showing: m/z 347 (dimethylmyricetin, 2M), m/z 361 (trimethylmyricetin, 3M), m/z 375 (tetramethylmyricetin, 4M) and m/z 389 (pentamethylmyricetin, 5M).

Annotation of methylated flavonoids is based on masses of pseudomolecular and fragment ions generated using ESI in both positive mode. Proposed structures from mono- to pentamethylmyricetins are shown in Figure 4-12. Methylated flavonoid candidates have positive mass defects in the range of 40-100 mDa. These metabolites elute later than flavonoid glycosides owing to the lack of the polar carbohydrate groups, but earlier than acylsugars. Fragment ions generated using elevated CID energies were consistent with multiple losses of $-CH_3$ ($\Delta 15$ Da) groups in both positive and negative modes. Additional fragmentation involves cross-ring cleavages, and common positions of ring cleavage of flavonoids are displayed in Figure 4-13.²⁰⁷ Fragmentation pathways of

tri- and tetramethylmyricetins are used as examples to show how these structures were annotated. Trimethylmyricetin from *S. habrochaites* LA1777 has its three methyl groups distributed among the A, B, and C rings based on fragment ion masses produced using ESI positive mode. Determination of the positions of methylation proved challenging owing to the facile losses of methyl radicals as discussed in more detail below. Although evidence for one methyl group on the A ring for this metabolite was suggested by the fragment ion at m/z 167.03, attributed to 1,3A⁺ cleavage using the nomenclature developed by Mabry and Markham³³ However, the same fragment mass also suggests localization of a methyl group on the B ring as a 1,2B⁺ fragment ion. Extensive homolytic cleavage of the bond between the oxygen and methyl group was observed in multiplexed CID and in MS/MS product ion spectra. The dominant fragments for trimethylmyricetin included those arising from losses of one and two methyl radicals from [M+H]⁺. Figure 4-14 displays ESI spectrum of trimethylmyricetin using ESI positive mode and product ion MS/MS spectrum for trimethylmyricetin. More detailed characterization of methylated flavonoid aglycones is being conducted by graduate student Chao Li who is using MS/MS and MS³, and these results will be presented separately. Proposed fragmentation pathways for trimethyl myricetin using positive mode are shown in Figure 4-15.

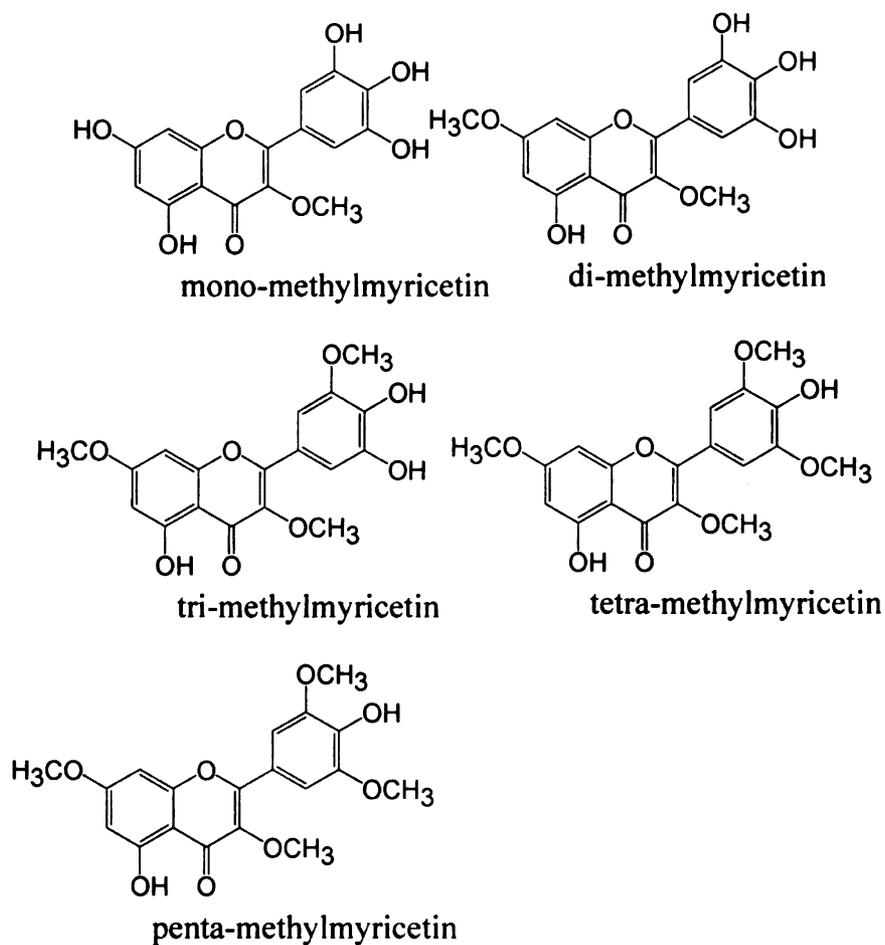


Figure 4-12 Putative structures of major methylated flavonoids in *Solanum* species LA1777. The exact positions of methyl groups on these flavonoids remain unknown.

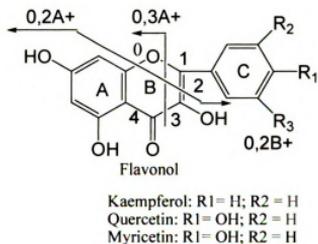


Figure 4-13 Nomenclature of fragment ions generated from flavonoids.²⁰⁷

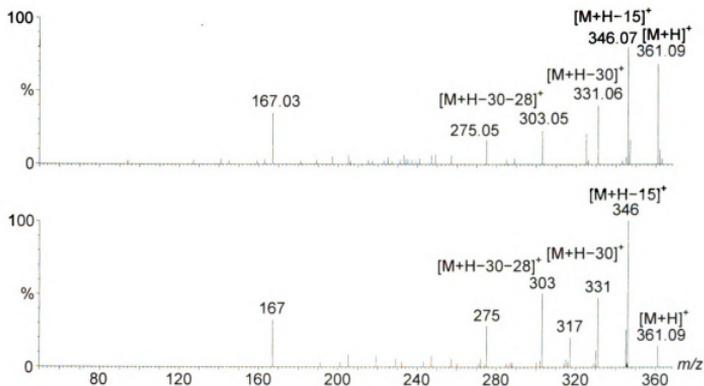


Figure 4-14 (Top) Positive mode mass spectrum of trimethylmyricetin from *S. habrochaites* LA1777 using CID potential 80 V. (Bottom) Product ion MS/MS spectrum of $[M+H]^+$, m/z 361 trimethylmyricetin (positive mode) using collision energy of 60 eV.

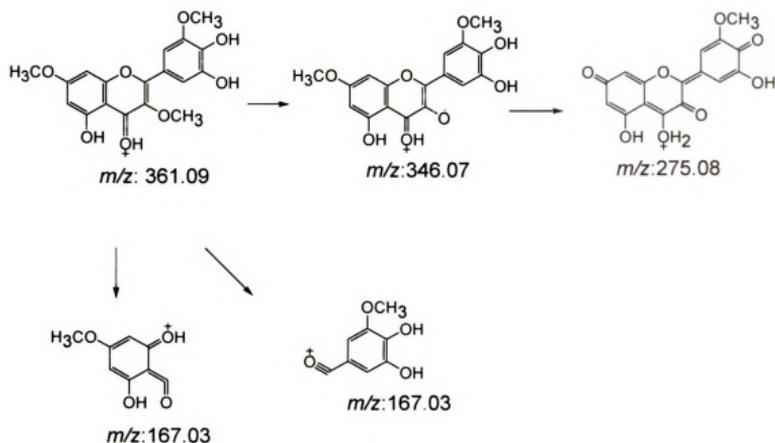


Figure 4-15 Proposed formation of fragment ions of trimethylmyricetin in ESI positive mode, obtained from extract of *S. habrochaites* LA1777

Positive mode mass spectra and MS/MS product ion spectra were generated for tetramethylmyricetin from *S. habrochaites* LA1777 (Figure 4-16). Fragment ions suggested two methyl groups on B ring and another two on A and C ring respectively. Two methyl groups attached on B ring were suggested by fragment ion at m/z 181.05, but this fragment might also be explained by methyl substitutions on the 5 and 7 positions on the A ring. Therefore, positions of methyl group substitution were not established from these experiments.

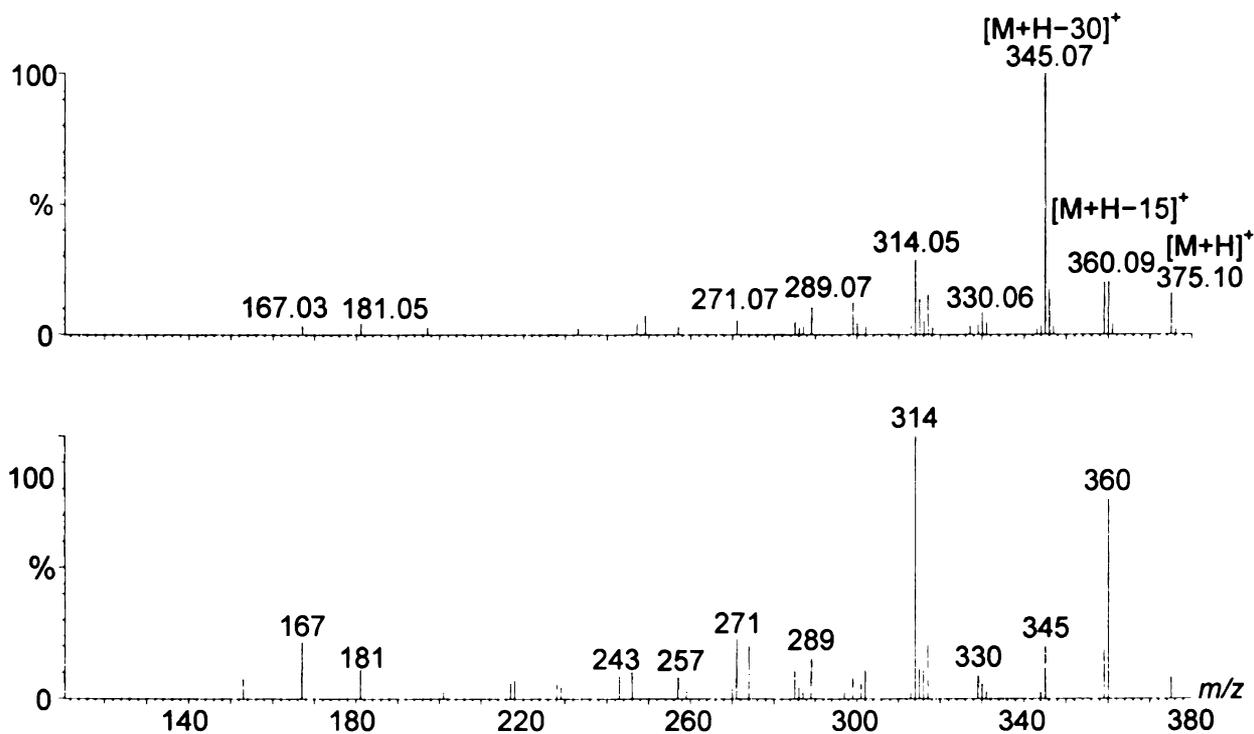


Figure 4-16. (Top) Positive mode mass spectrum of tetramethylmyricetin extracted from *S. habrochaites* LA1777 with CID potential 80 V. (Bottom) Positive mode product ion MS/MS spectrum of tetramethylmyricetin at m/z 361 using collision energy 40 V.

Although the exact positions of methyl groups on flavonoids could not be definitively assigned without standards, annotation of these metabolites is important for metabolic phenotype screening and association of methylation patterns with investigations of specific methyltransferase enzymes. Methylation patterns differed in various tomato lines. For example, *S. lycopersicum* M82 contain methylmyricetins ranging from 1-3 methyl groups, whereas in *S. pennellii* LA0716, the core aglycones differed, spanning methylated kaempferol, and quercetin, but barely detectable amounts of methylmyricetins. Kaempferol aglycones contained 1-2 methyl groups and quercetin had 1-3 methyl groups in this genotype. When metabolic profiles of flavonoids were surveyed among introgressions of LA0716 \times M82, introgression lines (ILs) IL6-3 and IL6-4 displayed

higher ratios of tri- to dimethylmyricetin. This ratio was inverted relative to other ILs and recurrent parent M82 (detailed in Chapter 5), implying that methyltransferases originating from chromosome number 6 of *S. pennellii* somehow alter the amount of trimethylmyricetins among these two ILs. These findings serve as the basis for continuing efforts to characterize O-methyltransferases involved in myricetin methylation in *S. habrochaites* LA1777 in the laboratories of Jones and Eran Pichersky (University of Michigan).

4.3.2 Structural annotations for glycoalkaloids

Glycoalkaloids make up another important class of specialized metabolites in tomato plants.²⁰⁰ These compounds act as toxic defense compounds and exist in numerous plant tissues, including the fruit and root.¹²⁷ The name refers to two characteristics – the term “glyco” implies that these metabolites are conjugated to carbohydrates. Alkaloids are metabolites containing basic nitrogen functional groups. In tomato, the best-known glycoalkaloids are variations of tomatine (Figure 4-17).

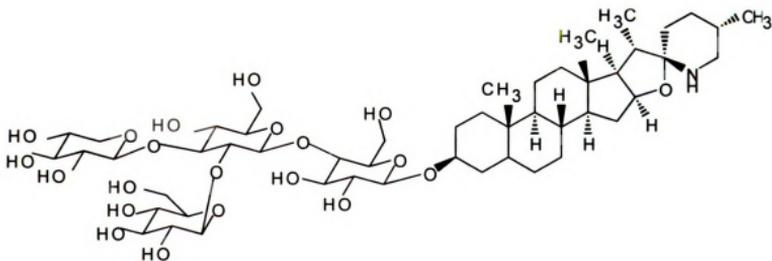


Figure 4-17 Chemical structure of tomatine

Glycoalkaloids are easily protonated in positive mode owing to the basic secondary amine group, and they are easily detected using electrospray ionization in positive mode. They lack acidic functional groups, and are more likely to be observed as adduct ions (e.g. formate adducts) in negative mode. The polar carbohydrate groups and the positively-charged protonated amine confer sufficient interaction with polar solvent for these compounds to elute earlier than acylsugars on reversed phase separations used in this work. Glycoalkaloids are recognized based on their large positive mass defects in the range of 400-600 mDa, owing to the substantial hydrogen content in the steroid aglycones.

Mass spectrometry has been a useful tool for characterization of glycoalkaloids. Positive mode MS/MS product ion scans generated using an ion trap mass spectrometer²⁰⁸ were generated for steroid glycoalkaloids extracted from tomato leaves to elucidate fragmentation mechanisms. The findings can be summarized as follows: protonated glycoalkaloids undergo ready loss of water accompanied by losses of carbohydrate moieties, with relatively low levels of fragments useful for investigating the structure of the aglycone portion (minus carbohydrates).

Based on characteristic fragments summarized in a 2005 report,²⁰⁸ the glycoalkaloid metabolites were identified in trichome extracts using LC/TOF MS with multiplexed CID. Manual inspection of negative ion LC/MS results for an extract from tomato *S. lycopersicum* M82 trichomes led to recognition of two abundant peaks with even nominal masses, indicative of an odd number of nitrogen atoms present. Extracted ion chromatograms (XICs) for masses of formate adducts of glycoalkaloids α -tomatine and dehydrotomatine, which were previously known tomato metabolites, are shown in Figure

4-18. Identifications of these metabolites were supported by chromatographic retention time, accurate mass measurements in both positive and negative ion modes, fragment ion masses, proposed elemental formulas derived from these measurements, and knowledge of previously-identified metabolites from the tomato fruit metabolite database (MoToDB) and Kazusa OMICS (KOMICS) database.¹⁹⁸

The known tomato glycoalkaloid α -tomatine is used to illustrate how mass spectrometry supports glycoalkaloid structure elucidation. LC/MS and CID spectra are particularly useful because these are large molecules with complex NMR spectra in the aliphatic region, and NMR-based structure elucidation requires prior purification of milligram quantities and extensive spectrum interpretation. First, the accurate mass measurement of $[M+H]^+$ for the strongest glycoalkaloid signal was an even nominal mass at m/z 1034.55, suggesting this molecule contained an odd number of nitrogen atoms. Fragment ions generated at higher CID potentials appeared at m/z 902.50, 740.45, and 578.41, are consistent with consecutive neutral losses of anhydro forms of xylose (132 Da), xylose-glucose unit (132+162 Da) and xylose-glucose-glucose (132+162+162 Da), in agreement with the documented sequence of the oligosaccharide conjugate. Neutral loss of another glucose (162 Da) would account for the fragment ion at m/z 416.36, which corresponds to the characteristic fragment of the protonated aglycone ($[\text{tomatidine}+H]^+$). In addition to fragments mentioned above, a doubly-charged ion at m/z 528.78 is assigned as $[M+H+Na]^{2+}$ and other fragment ions observed at m/z 884.52, 722.43, 560.38 and 398.33 using the highest CID potential (80 V) are in agreement with consecutive loss glycosides plus H₂O through a rearrangement process in the E ring,²⁰⁸ as described in Figure 4-19. Lower mass fragment ions at m/z 273.22 and 255.21 are consistent with

cleavage of E-ring from the protonated aglycone [tomatidine+H]⁺ at m/z 416.36 with CID voltage 55 V, which was consistent with earlier studies by Cataldi and colleagues.²⁰⁸ The ion at m/z 161.13 can be assigned as the result of water loss and multiple cleavages in the C-ring that can be explained by a charge driven mechanism, involving the generation of resonance stabilized allylic carbocations²⁰⁹ (Figure 4-20).

Comparisons of mass spectra of other putative glycoalkaloids assist their structure annotation. Figure 4-21 displays ESI mass spectra of α -tomatine and dehydrotomatine using ESI positive mode with CID voltage 80 V. Another metabolite, assigned as dehydrotomatine, gives a protonated molecule 2 Da lower (m/z 1032 [M+H]⁺) than tomatine. All major fragment ions are also 2 Da lower in mass than the corresponding fragment ions from α -tomatine, indicative that the unsaturation lies in the aglycone portion and not in the carbohydrate portion. These results are consistent with the difference of one unsaturation relative to tomatidine,²⁰⁸ and with a double bond at position 5, 6 in the steroidal aglycone, as reported for tomato.¹⁹⁸

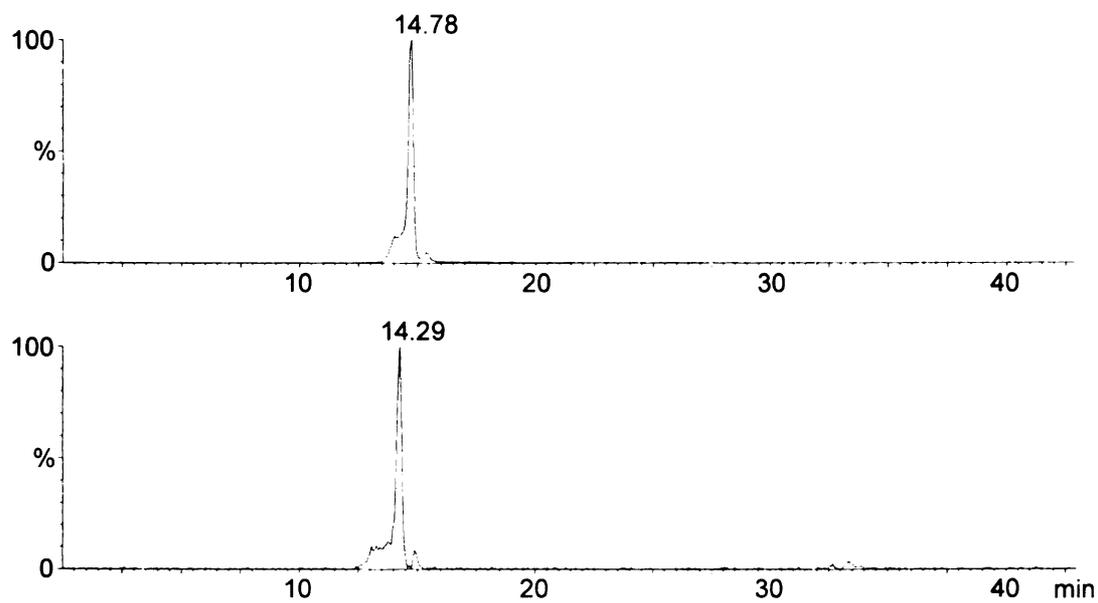


Figure 4-18 (Top) LC/MS XICs of $[M+H]^+$ for α -tomatine m/z 1034.5 (RT = 14.78 min) and (Bottom) dehydrotomatine m/z 1032.5 (RT = 14.29 min) from a leaf dip extract of tomato (*S. lycopersicum* M82).

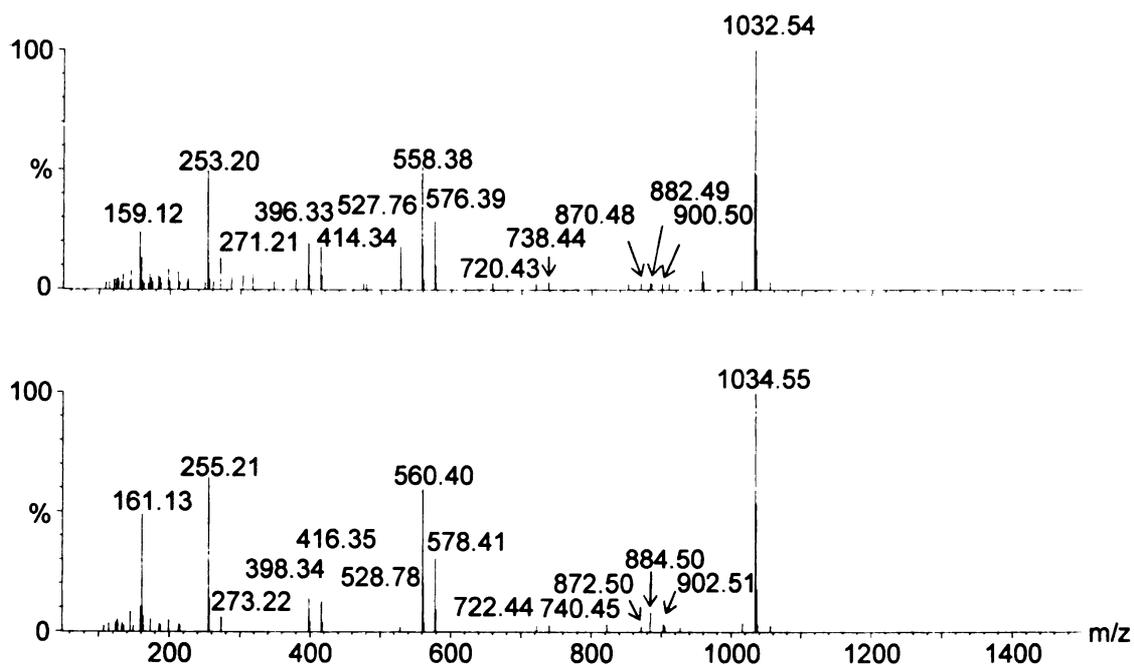
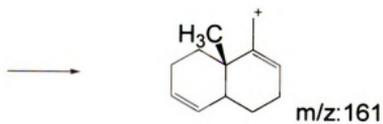
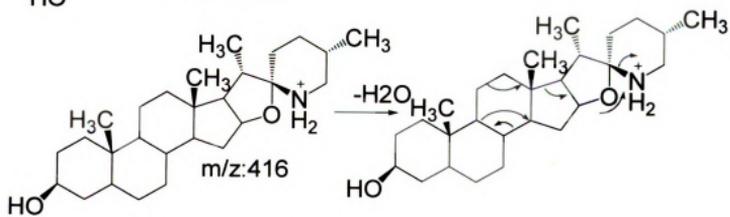
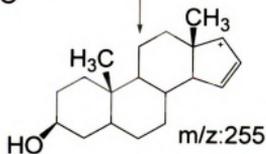
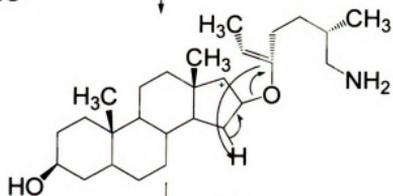
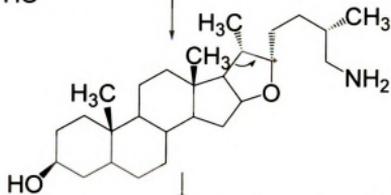
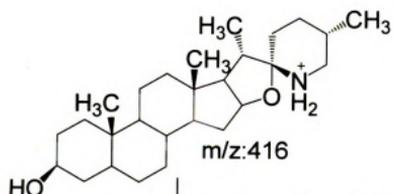


Figure 4-19 Mass spectra of α -tomatine (Top) and dehydrotomatine (Bottom) from LC/MS analysis of an extract from *S. lycopersicum* M82 using ESI positive mode with CID potential 80 V.

Figure 4-20. A proposed fragmentation pathway for α -tomatine using ESI positive mode based on literature reports.^{208, 209}



In contrast to *S. lycopersicum* M82, two chromatographic peaks suggestive of additional isomers for both α -tomatine and dehydrotomatine are present in extracts of *S. pennellii* LA0716 (Fig.4-21). These two isomeric peaks produce same fragment ions using ESI positive mode, suggesting they are either stereoisomers or regioisomers. Metabolic phenotypes of glycoalkaloids were also surveyed among more than 50 introgression lines of LA0716 \times M82. Chromatographic peaks corresponding to novel glycoalkaloids were discovered in extracts of IL1-1 and 1-1-3 that were not present in other ILs or M82, suggesting that a region at the top of chromosome 1 is responsible for the metabolic difference. Two peaks with same exact mass as protonated dehydrotomatine, at m/z 1032.56 were detected for IL1-1 and IL1-1-3 (Figure 4-22 top). Based on fragment ions generated using a CID voltage 55 V, both yield losses of neutral carbohydrate fragments at m/z 900.50, 738.44 and 576.39, and give the same mass for the steroid aglycone at m/z 414.34. However, lower mass fragment ions attributed to fragments of the aglycone are different after E-ring is cleaved. The early eluting isomer (RT = 2.73 min) generates fragment ions (Figure 4-22 middle) at m/z 271.21 and 253.20, but the later eluting isomer (RT = 3.00 min) is consistent with the known dehydrotomatine metabolite, yielding fragment ions (Figure 4-22 bottom) 2 Da mass greater at m/z 273.22, 255.21. The fragments from the early eluting isomer are identical with the corresponding peaks from α -tomatine, which does not have any double bonds in rings A-D. These observations suggest the early eluting isomer corresponds to tomatine, but with an additional unsaturation, probably a double bond, on the F ring which contains the nitrogen atom. A proposed pathway for formation of the fragments at m/z 273 and 255 is displayed in Figure 4-23. This is a novel metabolite with no previous literature

reported. Purification of this metabolite is underway by another student in Jones' lab to confirm the structure by NMR. If the metabolite is indeed the imine, it may reveal a view toward the key biosynthetic steps in formation of the heterocyclic F ring, and the association of this trait with chromosome 1 should steer research toward the genes in this region.

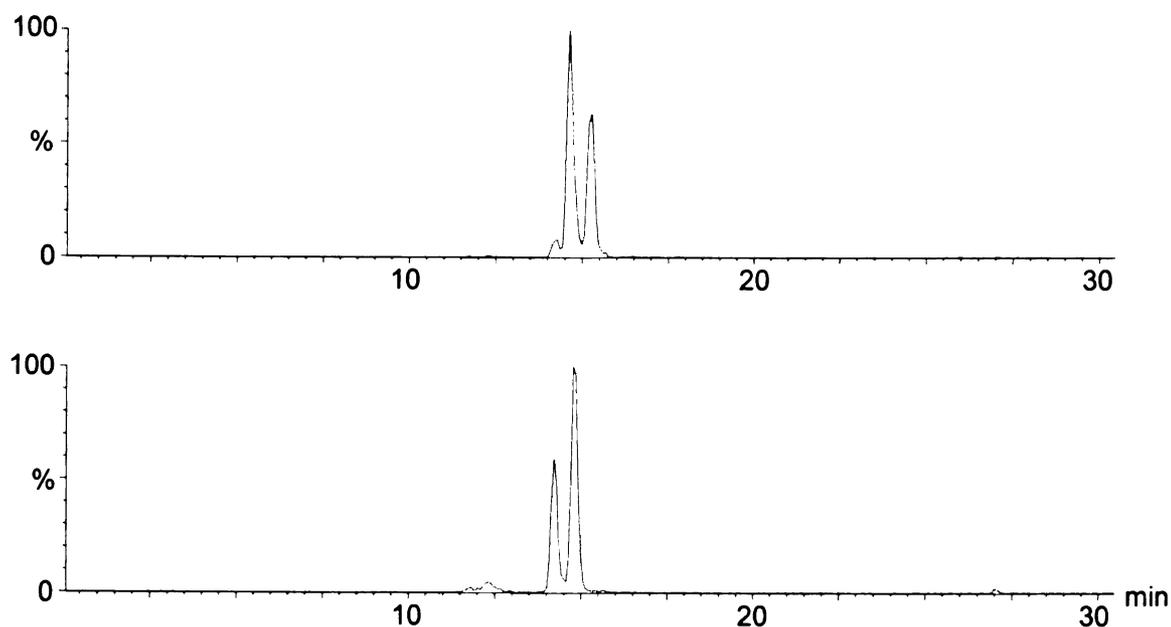


Figure 4-21 LC/MS XICs of an extract of *S. pennellii* LA0716 using ESI in positive ion mode corresponding to protonated molecules of (Top) α -tomatine at m/z 1034.5 and (Bottom) dehydrotomatine at m/z 1032.5.

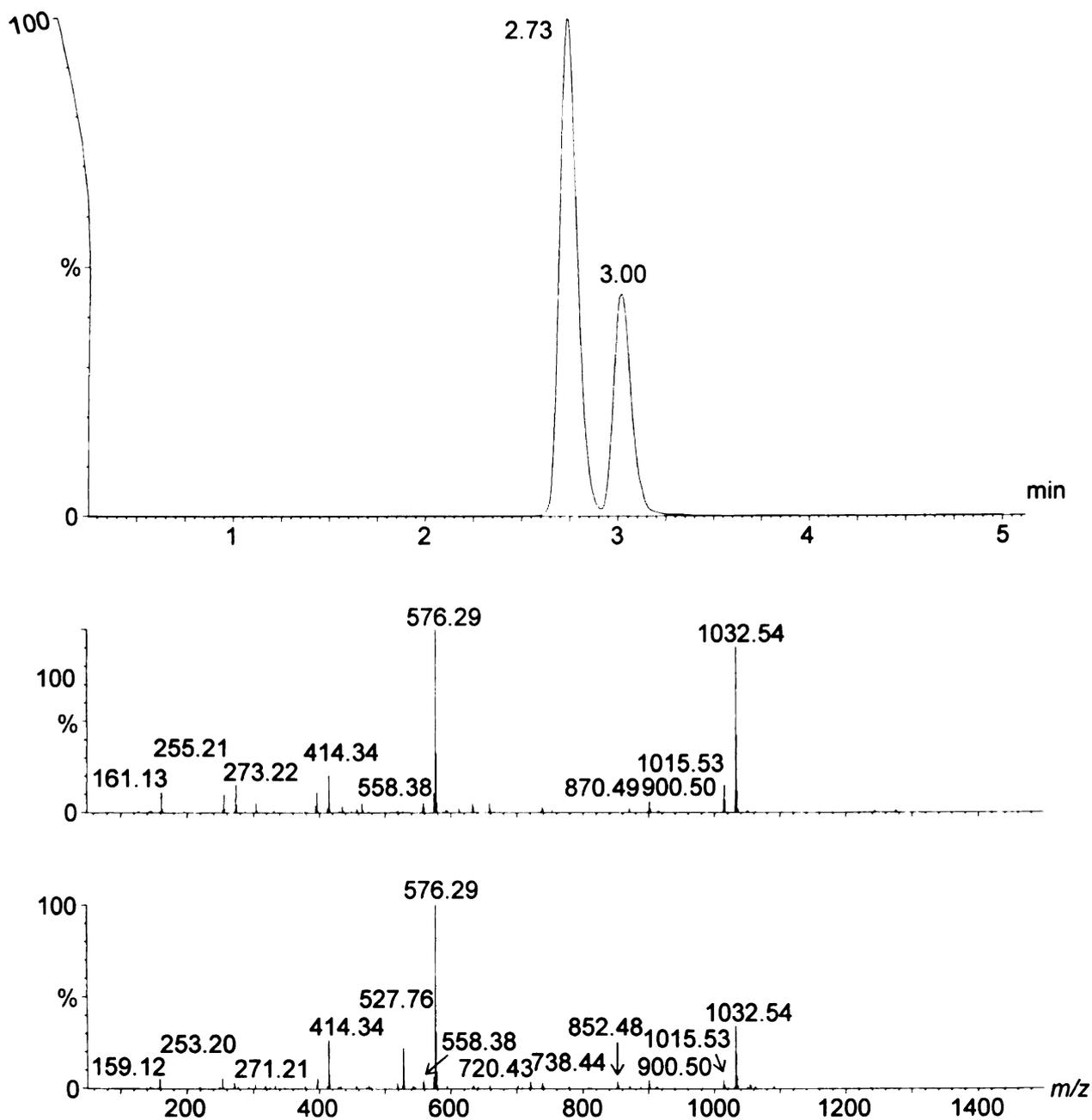


Figure 4-22 (Top) LC/MS XIC of m/z 1032.5 from IL1-1 using ESI positive mode and CID voltage 10 V. ESI spectrum (Middle) of first peak (RT = 2.73 min) corresponding to dehydrotomatine isomer and ESI spectrum (Bottom) of second peak (RT = 3.00 min) corresponding to dehydrotomatine (double bond in 5,6-position) using CID 80 V.

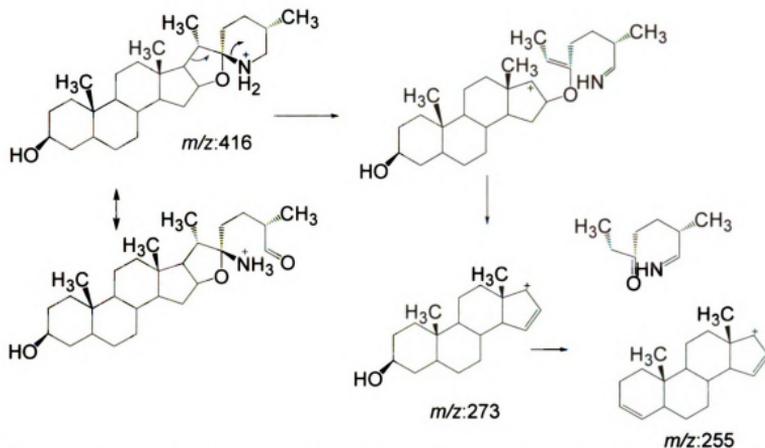


Figure 4-23 A proposed fragmentation pathway for first eluting dehydrotomatine isomer peak (RT = 2.73 min) from IL 1-1 using ESI positive mode with CID voltage 80 V. Fragment ions at m/z 273 and 255 indicate that the position of the double bond is not on rings A-D.

In addition, another putative glycoalkaloid at m/z 1050.55 ($[M+H]^+$) was detected in extracts of IL1-1 and 1-1-3. For this metabolite, the aglycone fragment is observed at m/z 432.35, suggesting one hydroxyl group based on the 16 Da increase in mass relative to the corresponding fragment from α -tomatine. Spectrum generated with higher CID potentials revealed associated fragment ions at m/z 273.23 and 255.22, as are found also for α -tomatine, attributed to cleavage of E-ring. These fragment ions suggest that the hydroxyl group is on the heterocyclic F ring of the metabolite, though an alternative explanation would involve a tautomeric form of a steroidal amino aldehyde. Figure 4-24 displays XIC of m/z 1050.56 ($[M+H]^+$) and ESI spectrum for this metabolite using ESI positive mode with CID potential 80 V. A fragmentation pathway that accounts for

formation of the fragments at m/z 273 and 255 from the protonated aglycone fragment is proposed in Figure 4-25.

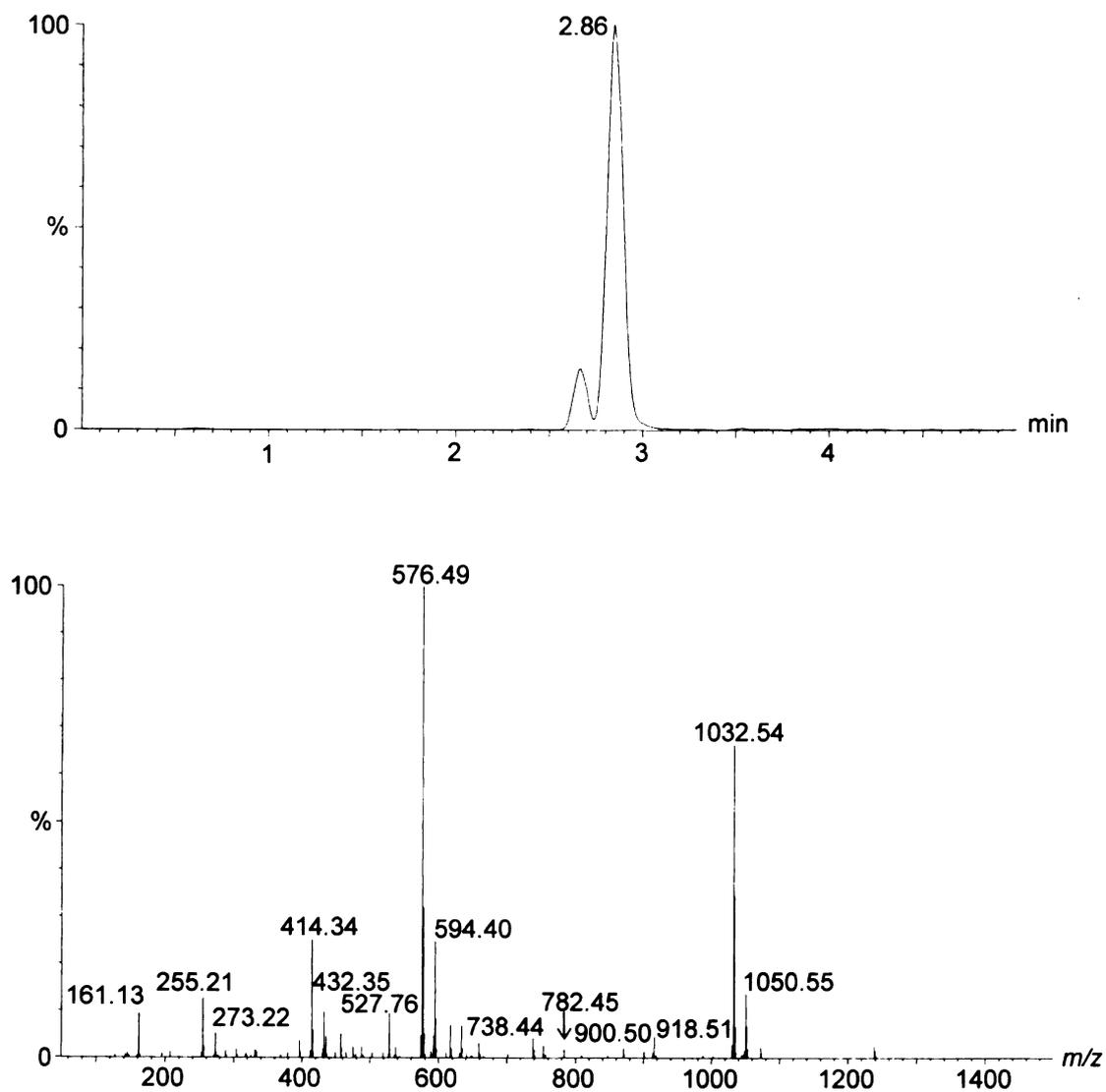


Figure 4-24 LC/MS XIC of m/z 1050 corresponding to a putative hydroxytomatine metabolite from IL1-1 using ESI positive mode. (Bottom) ESI spectrum for hydroxytomatine from IL1-1 using ESI positive mode with CID potential of 80 V.

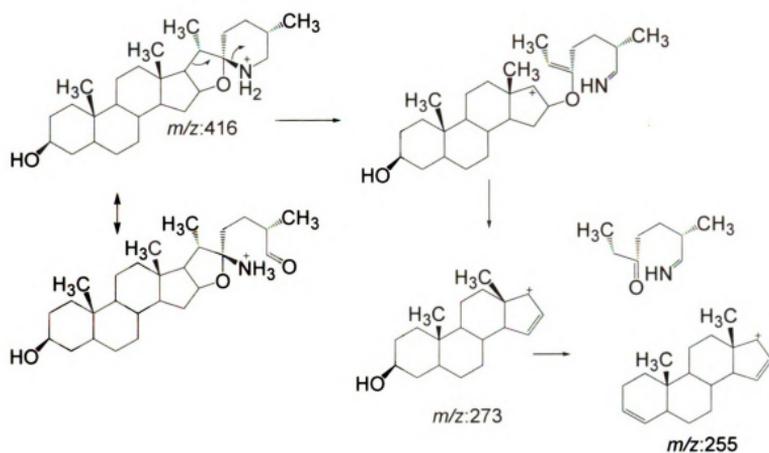


Figure 4-25 A proposed fragmentation pathway for hydroxytomatine from IL 1-1 using ESI positive mode.

To summarize the results from above, characteristics of fragmentation for glycoalkaloids²⁰⁸ were used to assign this class of metabolites using multiplexed CID. The carbohydrate side chain is easily cleaved to leave fragments corresponding to the protonated aglycone. Lower mass fragment ions caused by cleavage of E-ring appear and are useful for diagnosing structural modifications that occur on the F-ring.

4.3.3 Structural annotation of terpenoid metabolites

Although volatile monoterpenes (10 carbon atoms) and sesquiterpenes (15 carbon atoms) impart much of the characteristic “tomato” smell to *S lycopersicum* leaves, characterization of less volatile terpenoids has been less explored in the genus *Solanum*¹⁹⁷.

Oxidized and more polar terpenoids are less common, and are usually not observed in GC/MS analyses of tomato or its relatives.

One exception to this rule is the wild tomato relative, known now as *S. habrochaites* LA1777. This plant produces three isomeric sesquiterpene acids: α -santalenoic acid and two isomers of bergamoten-12-oic acid, all of which have the formula $C_{15}H_{22}O_2$ and molecular mass of 234 Da. Analyses of this accession using LC/MS revealed three major isomers of sesquiterpene acid at m/z 233.15 (Figure 4-26) in extracts of *S. habrochaites* LA 1777 and their assignment was based on exact mass, fragments and prior literature report.¹⁹⁷ One unusual feature was observed in the product ion MS/MS spectrum for deprotonated ion of α -santalenoic acid and is displayed in Figure 4-27. Products derived from $[M-H]^-$ included an abundant fragment ion at m/z 98 that is a radical anion. Such odd electron fragments are relatively unusual products from CID of aliphatic carboxylates, but stable radical due to the near double bond. A proposed fragmentation pathway is shown in Figure 4-28.

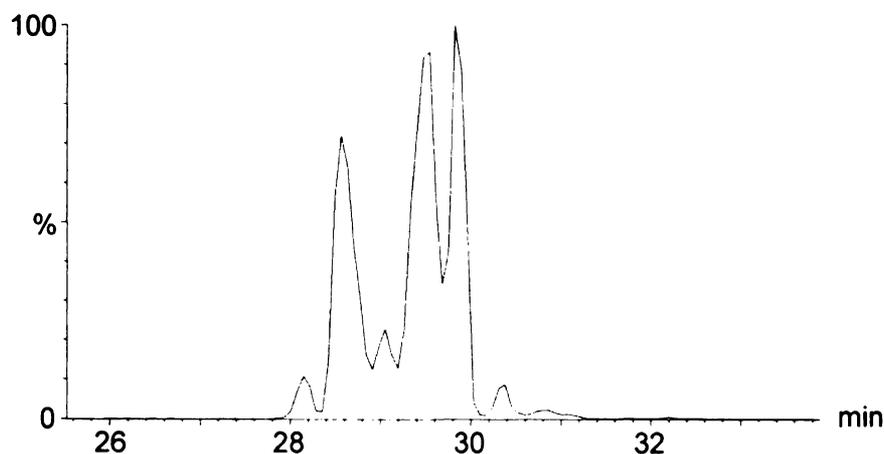


Figure 4-26. LC/MS XIC of sesquiterpene acid at m/z 233 from *S. habrochaites* using ESI negative mode.

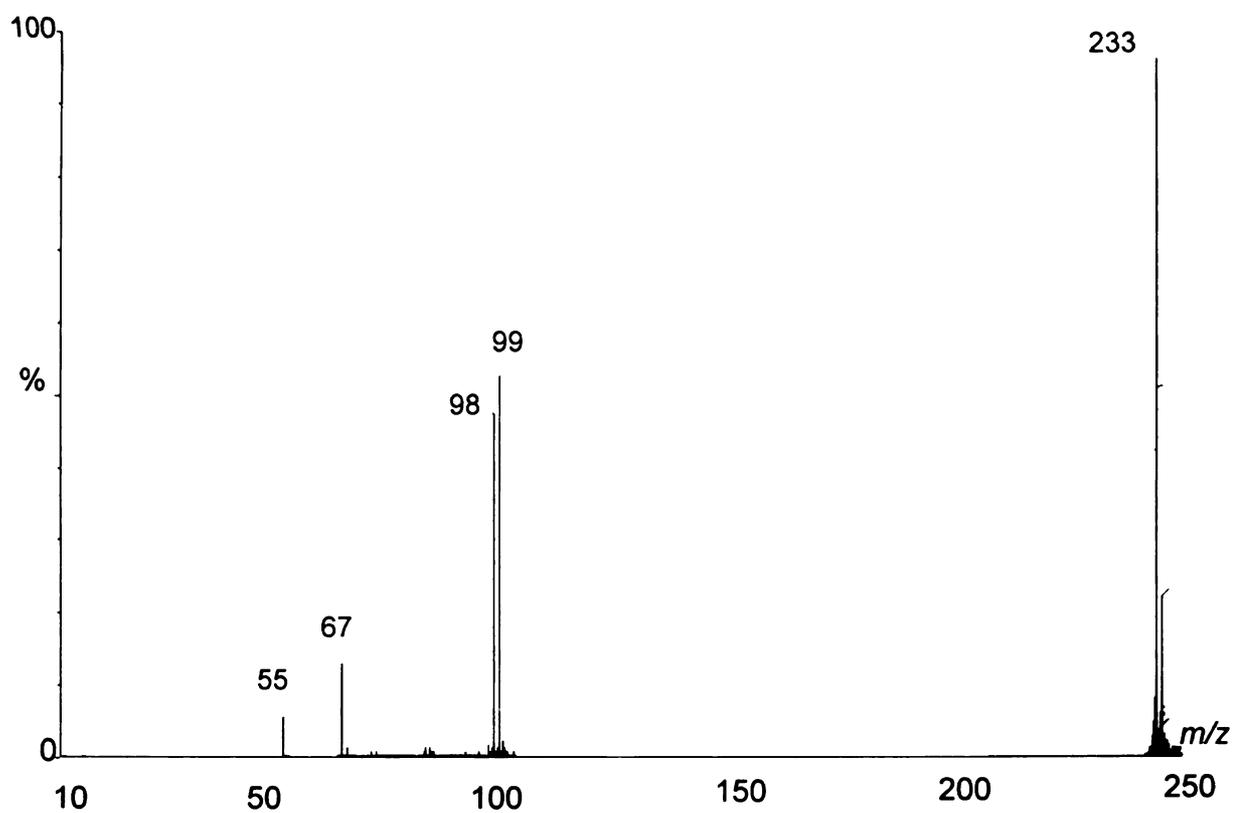


Figure 4-27 Product ion MS/MS spectrum of sesquiterpene acid at m/z 233 from *S. habrochaites* LA1777 using ESI negative mode with collision energy 45eV using flow injection analysis.

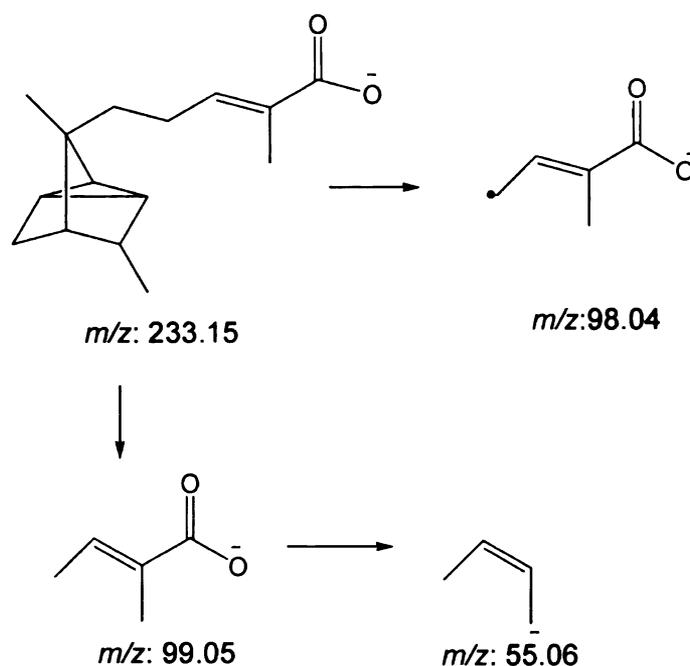


Figure 4-28 A proposed fragmentation pathway for α -santalenoic acid from *S. habrochaites* using ESI negative mode.

Several additional terpenoid acids from *S. habrochaites* LA1777 were also observed and annotated based on accurate mass measurements. These include three peaks annotated as oxidized sesquiterpene acids and a diterpene acid. (Figure 4-29). For the former, the mass difference between these and the sesquiterpene acids (m/z 249.15 and m/z 233.15) was Δ 16.00 Da, suggestive of one oxygen atom, so m/z 249.15 was tentatively assigned as hydroxylated variants of the sesquiterpene acids. The three chromatographic peaks may correspond to the three isomeric sesquiterpene acids, with each modified by addition of an oxygen atom. For the diterpene acid at m/z 319.22, the mass difference relative to m/z 233.15 was Δ 86.02 Da, corresponding to $C_5H_{10}O$. Though the negative mode CID spectra did not yield sufficient information to determine

more structural details, the difference in formula would be consistent with addition of an isoprene unit to sesquiterpene acid plus H₂O.

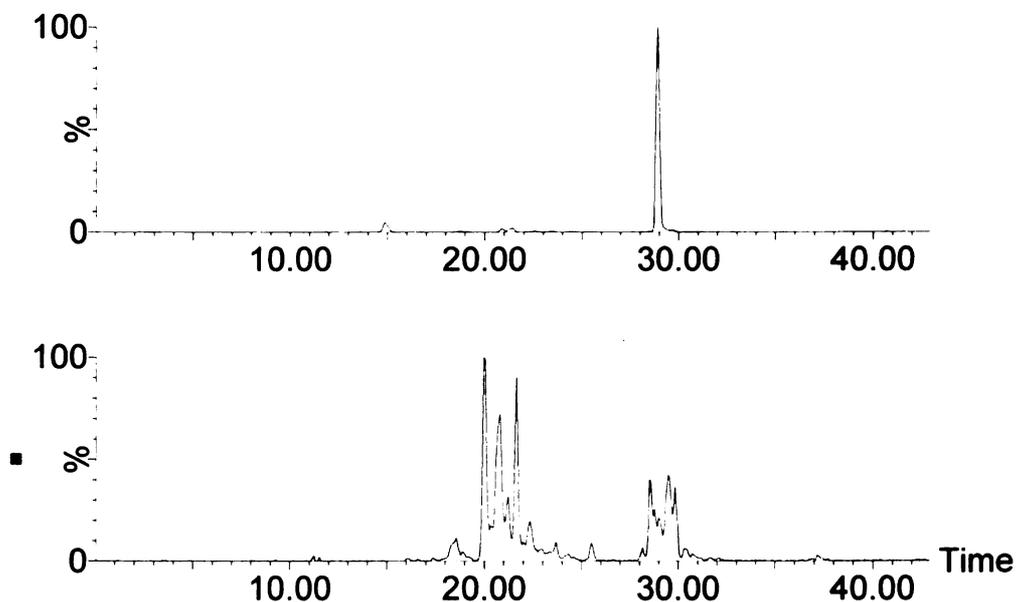


Figure 4-29 (Top) LC/MS XICs of diterpene acid at m/z 319 and (Bottom) hydroxysesquiterpene acid at m/z 249 from *S. habrochaites* LA1777 using ESI negative mode with CID potential of 10 V.

In addition to sesquiterpene acids and diterpene acids, several metabolite ions were discovered in extracts of *S. habrochaites* LA1777, with the analyses relying heavily on accurate mass measurements. Sesterterpenoids are a group of metabolites containing 25 carbon atoms (5 isoprene units) in core structures that are rarely found in plants, though several have been isolated from marine sponges. The few known examples exhibit a number of interesting pharmaceutical properties.^{210,211} One prominent signal detected in extracts of accession LA1777 showed a putative [M-H]⁻ peak at m/z 649.32

was used as an example to show characterization of these metabolites. First, three possible elemental formula were obtained based on accurate mass measurement. Then, fragment ions are generated for one isomer of m/z 649.32 (RT = 16.83 min) with CID voltage 55 V (Figure 4-30). A product ion MS/MS spectrum was generated for this metabolite to confirm that the major fragments observed using multiplexed CID were also observed in products of m/z 649. For both kinds of spectra, major fragments were observed at m/z 605.32, corresponding to neutral loss of CO₂. This neutral loss occurred at low CID potentials as would be consistent of a carboxylic acid group that undergoes decarboxylation easily such as a β -keto acid. Mass difference between m/z 605.32 and m/z 563.31 is $\Delta 42.02$, consistent with neutral loss of ketene (CH₂=C=O). These findings are consistent with a malonic ester group that undergoes successive losses of CO₂ and ketene. The next prominent fragment is observed at m/z 401.26, corresponding to subsequent neutral loss of anhydrohexose ($\Delta 162.05$). Exact mass measurements suggested an elemental formula for m/z 401.26 to be C₂₅H₃₇O₄⁻. No further fragmentation of this ion was observed, therefore this is assigned as the core of this terpenoid metabolite. Several mass peaks at m/z 811, 729, 691, 567, 581, 499 and 487 were observed to consecutively loss $\Delta 44.01$ (COO⁻) and then $\Delta 42.02$ (CH₂CO) corresponding to malonic ester in addition to m/z 649. Neutral loss scan of $\Delta 86$ was generated for *S. habrochaites* LA1777 to support this result (Figure 4-31). Based on exact mass measurements and their fragmentation pattern, they were all assigned as sesterterpene acid derivatives. For example, m/z 811.38, mass difference with m/z 649.32 is $\Delta 162.0650$, corresponding to hexose. The retention time of mass peak at m/z 811.38

also elute earlier (1.05 min) than mass peak at m/z 649.32 since the polarity increased due to one more hexose. Elucidation of those metabolite structures is important for exploring the metabolite pathway of terpene acids. Therefore, much effort was put into the purification of these metabolites by preparative HPLC in our lab now. Although exact structures can not determined so far, the classes of metabolites at least can be assigned based on characteristic fragments and exact mass with the aid of multiplexed CID.

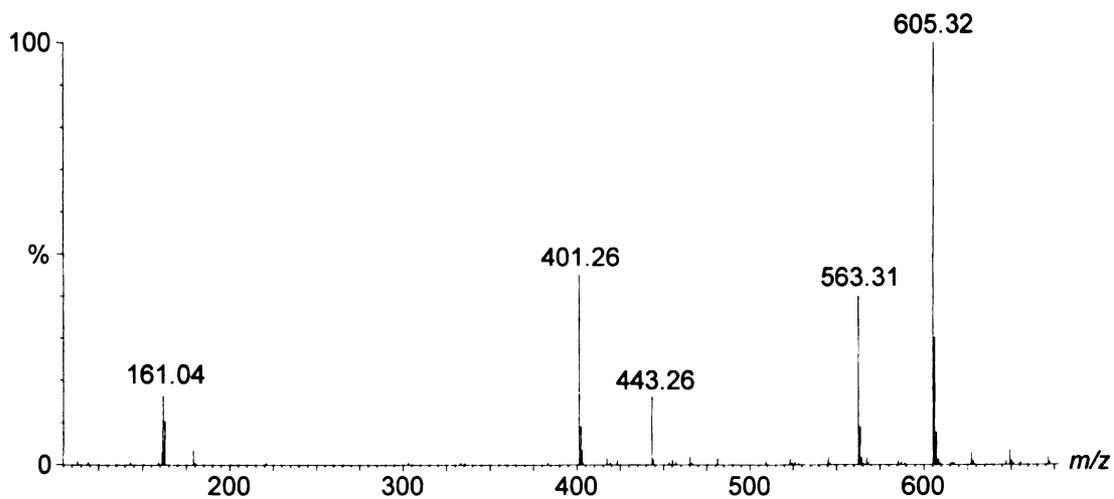
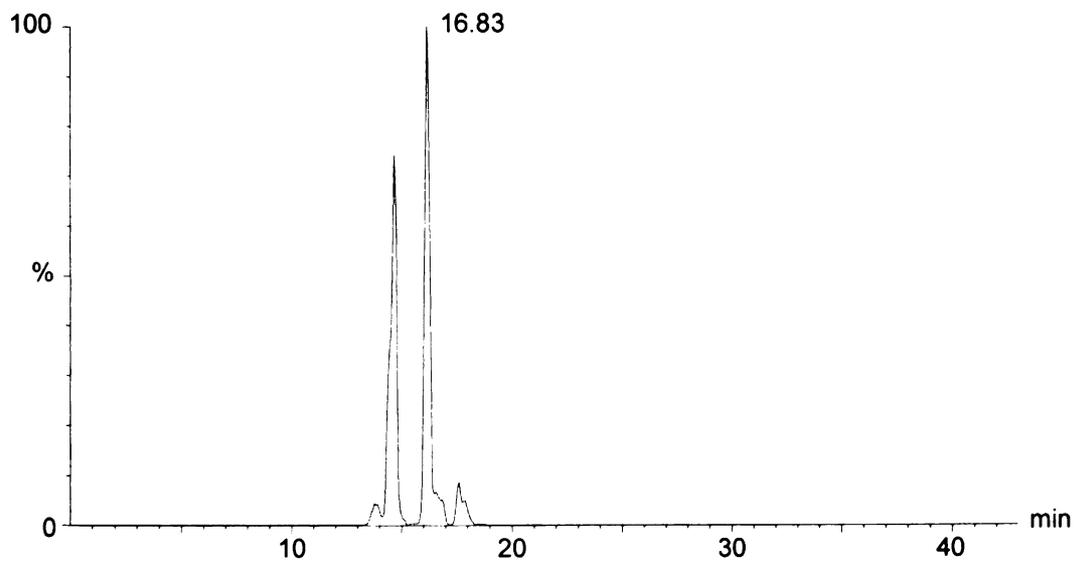


Figure 4-30 (Top) LC/MS XICs of m/z 649 [M-H]⁻ of sesterterpene metabolite using ESI negative mode. (Bottom) ESI spectrum of m/z 649 (RT = 16.83 min) from *S. habrochaites* LA1777 using ESI negative mode with CID potential 55 V.

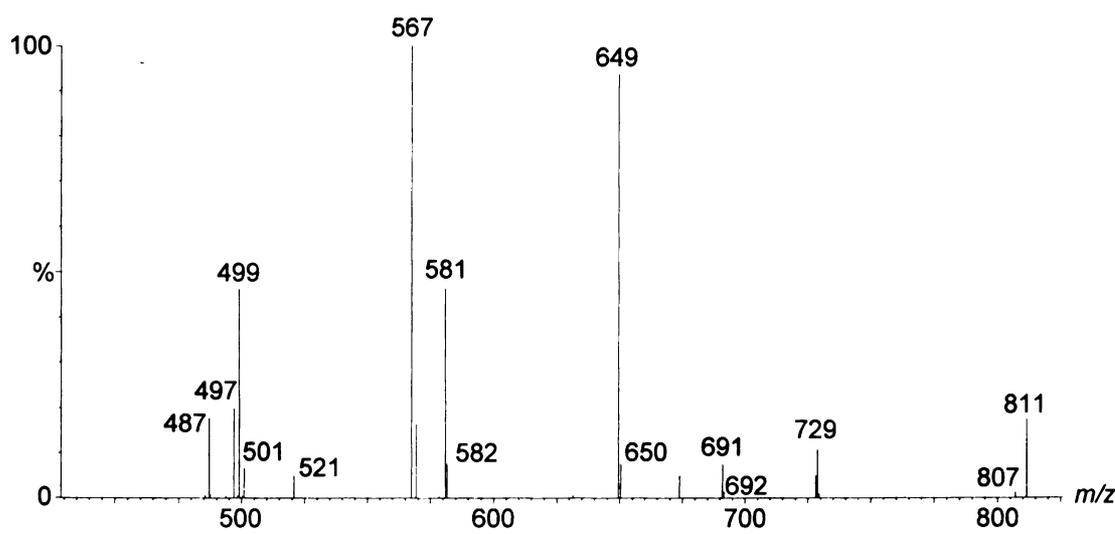


Figure 4-31 Constant neutral loss (86 Da) spectrum from an extract *S. habrochaites* LA1777 using flow injection analysis and ESI negative mode.

4.4 Conclusions

These efforts have annotated several novel secondary metabolites from trichomes within the genus *Solanum* using LC/TOF with multiplexed CID and LC/MS/MS performed on a QToF instrument. These discoveries have included findings of methylated variants of the flavonoid myricetin, novel glycoalkaloids with F-ring modifications that point to potential biosynthetic intermediates, and new terpenoid acids including malonate esters of sesterterpene glycosides, which constitute a family of specialized metabolites rarely described within the plant kingdom. The observation of new metabolites in several genotypes of tomato and its relatives suggest that these plants may represent a rich resource for mining of genes useful as biological catalysts.

Chapter 5. Fast LC/time-of flight mass spectrometry for screening metabolic phenotypes for tomato chromosomal substitution lines

5.1 Introduction

Glandular secreting trichomes (GSTs) are epidermal protuberances that are found on a wide variety of plant species.^{212,213} These structures typically manufacture relatively large amounts of specialized (secondary) metabolites, and either store these molecules in an extracellular space or deposit them on the epidermal surface, sometimes in amounts sufficient to cause stickiness to the touch. Trichome compounds such as those in basil,²¹⁴ mint²¹⁵ and hops²¹⁶ are responsible for their smell and taste. Other plants manufacture compounds of medicinal importance such as the anti-nausea agent tetrahydrocannabinol from Cannabis and antimalarial artemisinin from wormwood. The combination of active biosynthesis, species-specificity of metabolite accumulation and ease of purification of trichomes make them an excellent system for specialized metabolite pathway elucidation and identification of genes encoding the pathway enzymes.^{212,213} The combination of cDNA sequencing and metabolite analysis of SGTs to discover candidates for *in vitro* enzyme analysis has been very effective in elucidation of specialized metabolic pathways.^{217-56,218} However, this molecular biochemistry approach can be very challenging or impossible when seeking enzymes that have structures unrelated to known activities or are members of multigene families. Forward genetics (screening for an altered phenotype in a collection of genetically diverse individuals) is a complementary approach to discovery of enzymes, transporters or regulators of poorly defined biosynthetic pathways or primary sequences not recognizable by homology. One of the advantages to this approach is that, by directly screening genetic variants for changes in metabolite levels, it is possible to identify genes encoding completely novel products.²¹⁹ Therefore, we applied forward genetics into our research. A series of nearly isogenic

lines (ILs) (Figure 5-1) are chosen as targets for study widespread natural diversity. Among those lines, the individual chromosomal segments from the wild species *S. pennellii* (accession LA0716) were introduced into cultivated tomato (*S. lycopersicum*) M82 by recurrent backcrossing and genotyping.²²⁰ These lines were used for identification or verification of genes affecting a variety of phenotypes, including differences in metabolites.²²¹⁻²²³ The phenotypic and genetic study of ILs will help with the identification of diverse genetic variants and understanding the gene function. Tomato trichome research can take advantage of the available genetic resources to characterize specialized metabolites and to identify loci/genes resulting in the production of the compounds. Exploiting the genetic basis of the chemical phenotype in trichomes will enable further understanding of what chemicals are produced, what genes are involved in the production of the compounds and how they are regulated. In addition, it will also provide important knowledge for plant metabolic engineering to improve plant defense and human nutrition by regulating the production of specific metabolites. Therefore, successful analysis of the complex network of plant metabolism requires analytical method able to record information on as many metabolites as possible.

IL bin mapping

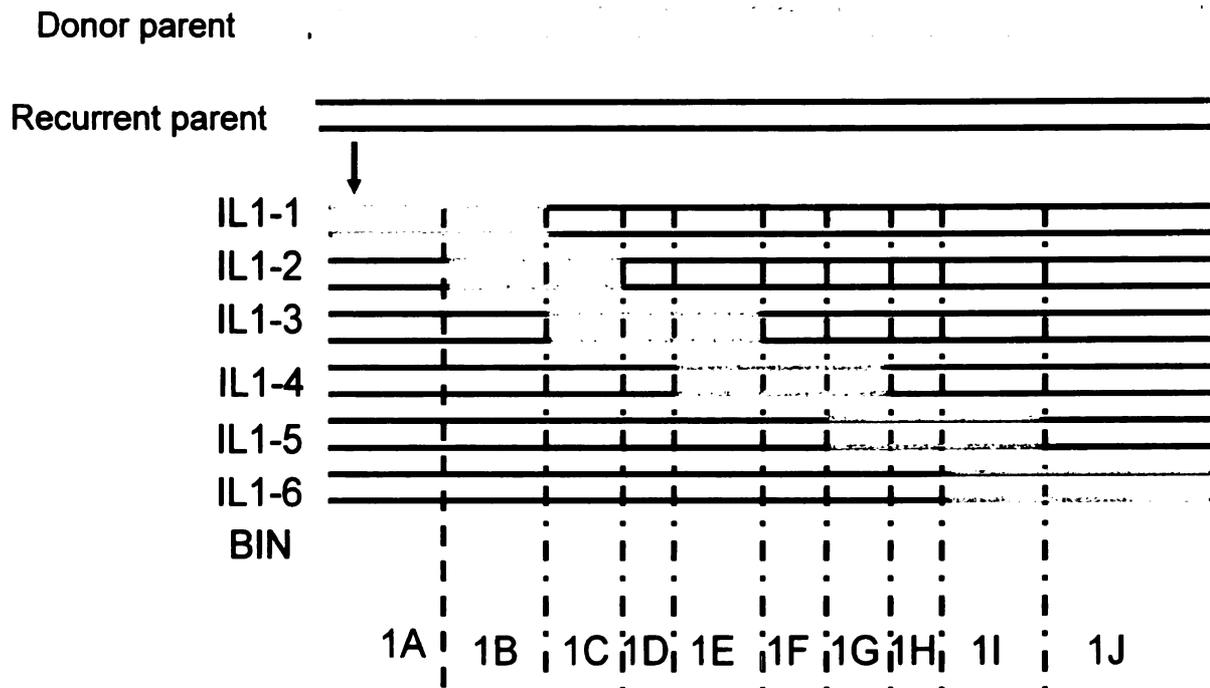


Figure 5-1. Incorporation of genomic DNA from *S. pennellii* donor parent (red) and tomato recurrent parent (black) into DNA for introgression lines. (adapted from http://zamir.sgn.cornell.edu/Qtl/il_story.htm)

Currently, however, no single technique can provide detection of all metabolites due to the complexity of chemical compounds.²²⁴ Over the past decade, several methods suitable for large-scale analysis and comparison of metabolites in plant extracts have been established,²²⁵ including gas chromatography coupled to mass spectrometry (GC-MS),^{5,24,56,226} direct flow injection–mass spectrometry (DFI-MS),^{224,227,228} liquid chromatography–mass spectrometry (LC-MS),^{103,229-233} capillary electrophoresis–mass spectrometry (CE-MS)^{234,235} and NMR technologies.^{48,235} LC-MS-based approaches are expected to be of particular importance in plants, owing to the highly rich biochemistry

of plants, which covers many semi-polar compounds, which can be separated and detected by LC-MS approaches.²³⁶ Here, in our study, a fast LC/time of flight mass spectrometry with multiplexed collision CID method is developed for discovery of plant phenotypes. The multiplexed CID system switches voltages from low to high and generates molecular and fragments in a single analysis. Exact mass measurements for all of ions using this approach assist to limit elemental formula and associate fragments with molecular ions. Another advantage of the method is that it is fast and reliable. For instance, ~300 plants can be screened within 30 h with reproducible and accurate results. Also, pre-existing software– Quanlynx and Markerlynx- in Waters Masslynx 4.1 is helpful for data processing including peak integration and statistical tests such as Principal Component Analysis (PCA). Five specific metabolic phenotypes are discovered among 65 ILs using this approach. In summary, this high-throughput approach offers a great opportunity for large-scale chemical screening.

5.2 Experimental Method

5.2.1 Plant Growth Conditions. Tomato seed, *S. lycopersicum* cv. M82 and *S. pennellii* LA0716, was obtained from the Tomato Genetic Resource Center (<http://tgrc.ucdavis.edu>). *S. pennellii* ILs²²⁰ were obtained from Dr. Dani Zamir (Hebrew University Faculty of Agriculture, Rehovot, Israel). Plant seedlings were grown in Jiffy peat pots (Hummert International) for 3 weeks in a growth chamber maintained for 16 h at 28°C in the light (300 $\mu\text{E}/\text{m}^2/\text{s}$, mixed cool white and incandescent light bulbs) and 8 h at 20°C in the dark.

5.2.2 Plant Extractions For analysis of total trichome metabolite, leaflets from the second leaf after the newly emerging leaf of 3-week-old plants were dipped with gentle rocking for 1 min in 750 μl of isopropanol : CH_3CN : H_2O (3:3:2 v/v/v) containing 10 μM of internal standard propyl-4-hydroxybenzoate.

5.2.3 Chromatography Reverse-phase analysis was performed on a Shimadzu (Columbia, MD) LC-20AD HPLC system using a fused core Ascentis Express C18 column, 2.1 \times 50 mm with 2.7 μm particles. The column was maintained at 30 $^\circ\text{C}$. A steep gradient was executed for rapid metabolite screening based on solvent 0.15% formic acid in MilliQ water (A) and Methanol (B) with analysis time of 5 min/sample. Gradient profile consisted of: initial conditions 10%B; linear gradient to 60% B (2 min); 100% B (3 min); hold at 100% B until 4 min.; return to 10% B (5 min). Flow rate was 0.4 $\text{ml}\cdot\text{min}^{-1}$.

5.2.4 Mass Spectrometry A Waters LCT Premier mass spectrometer (Milford, MA) coupled to a Shimadzu LC-20AD HPLC system and SIL-5000 auto sampler were used. Ionization mode: electrospray negative mode; capillary voltage: 3.00 kV; Source temperature: 90 $^\circ\text{C}$; desolvation temperature: 300 $^\circ\text{C}$; nebulizer nitrogen flow rate: 20 $\text{L}\cdot\text{h}^{-1}$; desolvation nitrogen gas flow rate: 300 $\text{L}\cdot\text{h}^{-1}$; mass range: 50-1500 Da. Analyses were performed using rapid switching of the Aperture 1 voltage in the ion transit region of the mass spectrometer, providing quasisimultaneous generation of spectra under fragmenting and nonfragmenting conditions. Aperture 1 voltage were 10, 25, 40, 55, 80 V for five functions. Data were acquired in centroid format with 'dynamic range enhancement' (DRE) enabled for a wide dynamic range for quantification. Peak areas of extracted ion chromatograms for characteristic masses of each metabolite and the internal

standard were integrated using Waters QuanLynx software. Measured levels of secondary metabolites were normalized to the dried weight of the tissue used for each extraction.

5.2.5 Chemometric Data Analysis The ESI⁻ raw data were analyzed by the Markerlynx applications manager version 4.1 (Waters, Manchester, UK). The parameters were RT range 0-5 min, mass range 100-1500 Da, mass tolerance 20 ppm, internal standard detection parameters were deselected for peak retention time alignment, isotopic peaks were excluded for analysis, noise elimination level was set at 50, minimal intensity was set to 1% of base peak intensity, maximum mass per RT was set at 50 and, finally, RT tolerance was set at 0.2 min. No specific mass or adduct ions were excluded. Ions of different samples were considered to be the same ion when they demonstrated the same RT and *m/z* value within tolerance above. If the peak was not detected in the sample, the ion intensity was documented as zero in the final data table. A list of the intensity of the peaks detected was generated for all the samples, using RT and *m/z* data pairs as the identifier of each peak. The resulting three-data comprising of peak number (RT-*m/z* pair), sample name, and ion intensity were analyzed by unsupervised method principal component analysis (PCA) within the Markerlynx software. Following that, the supervised method orthogonal projection to latent structures-discriminant analysis (OPLS) analysis is carried out using SIMCA-P.

5.3 Results and Discussion

5.3.1. Analysis of Chromosome Substitution Lines Using LC/TOF MS

Initial evaluation of rapid methods for screening metabolic phenotypes focused on determining whether physical separation of metabolites upstream of the mass spectrometer was desirable. Direct injection MS^{37, 228, 237-240} offers the most rapid option for mass spectrometric analyses, but suffers from suppression of ionization,³⁷ limited ability to distinguish of chemical isomeric metabolites, and complications that arise from in-source formation of adduct, oligomer, and fragment ions. Chromatographic separation offers a mechanism to resolve these features, but usually involves long analysis times (30 min or more) per sample. Recent improvements in LC column technologies provide enhanced chromatographic resolution through use of smaller particle sizes and superficially porous particles.²⁴¹⁻²⁴³ Thus, a fast LC separation employing a 5 min/sample analysis time coupled with exact mass measurements from time-of-flight mass spectrometry was developed for nontargeted metabolite profiling.

Fast 5-minute LC/TOF MS method was compared with a slower 43-min method on a longer column using a leaf dip extract of metabolites from of IL 8-1 (Figure 5-2). The major observed metabolites, including malic acid, quinic acid, chlorogenic acid, alkaloidglycosides and several acylsugars were resolved using both methods. The slower separation yielded chromatographic peak widths about 4-fold broader (20-30 s vs. 6 s) than the fast separation. Extracted ion chromatograms (XICs) for some specific metabolites were also compared between these two techniques to assess separation of isomeric metabolites. We generated 11 different XICs corresponding to formate adduct ions of acylsugars differing in molecular mass owing to different fatty acyl groups. (Figure 5-3). Fast HPLC separation resolved as many isobaric acylsugar metabolite peaks as the slower method. For example, three peaks at m/z 695 were resolved by fast HPLC

and the slower HPLC separations. One peak is assigned as a tetraacylsucrose with 18 fatty acyl carbon atoms, and the other two are triacylsucroses with 19 fatty acyl carbons. These assignments could be made owing to the high mass accuracy measurements which distinguish the tetraester from the triester from the 36.4 milliDalton difference in molecular masses.

Figure

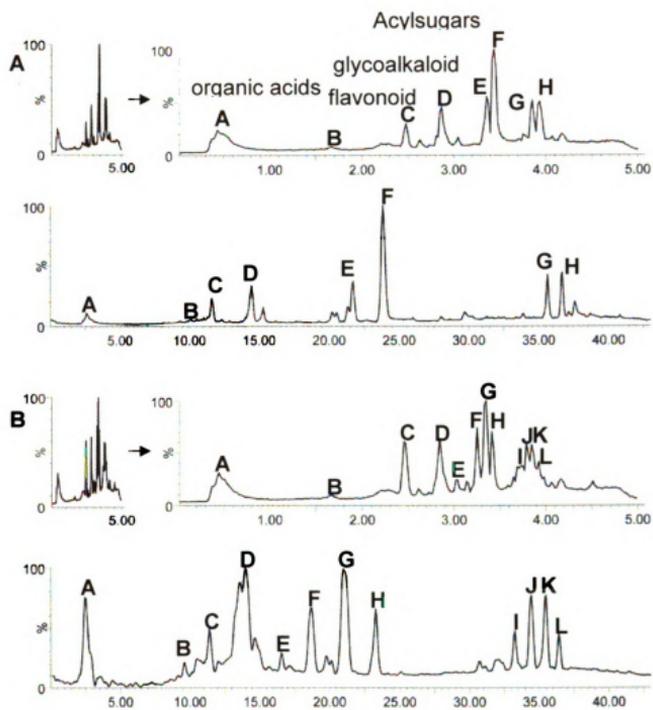
13

14

15

16

Figure 5-2. Total ion chromatograms obtained from leaf dip extraction of M82 (A) and IL8-1-1 (B) with a 5 min gradient using an Ascentis Express C18 fused core column, 2.1 x 50 mm; 2.7 μ m (Top of A and B) compared to a 43 min gradient with a Thermo BetaBasic C18 column, 1 x 150 mm, 3 μ m (Bottom of A and B) in ESI negative mode. A: Malic acid and quinic acid; B: Chlorogenic acid; C: Rutin; D: Tomatine; E-L: Acylsugars.



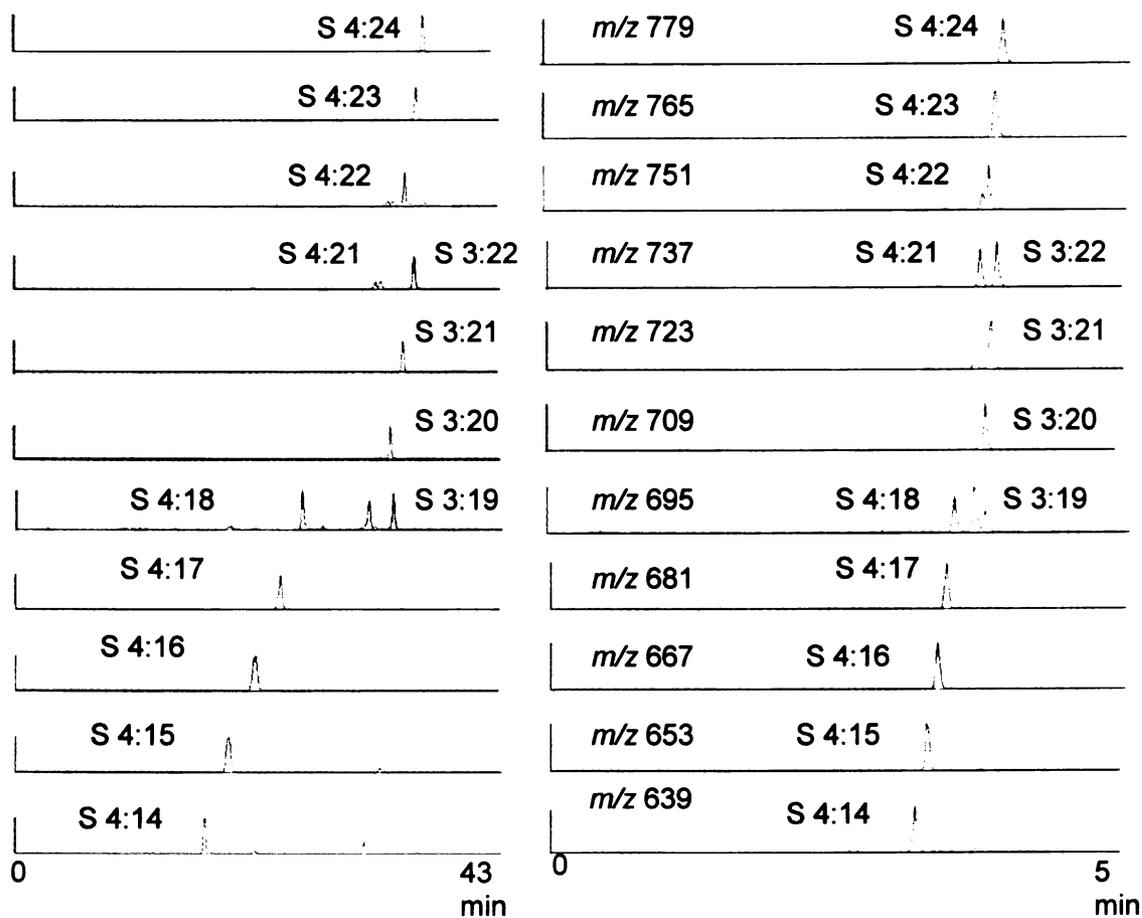


Figure 5-3 Extracted ion chromatograms (XICs) of $[M+HCOO]^-$ for leaf dip extract of IL 8-1-1 showing peaks corresponding to sucrose triesters or tetraesters from m/z 639 to m/z 779 (homologs differing by 14 Da in molecular mass owing to additional methylene group in the acid moieties). Left: 43 min gradient in ESI negative mode. Right: 5 min gradient in ESI negative mode.

We have frequently used the fast LC/MS method for screening several hundred metabolite extracts from a single sample set. For one set of experiments, 296 samples including blanks were screened over one weekend using this fast HPLC/MS method. Automated peak detection, integration, and alignment was performed, generating an extensive data matrix of analytical signals (annotated as retention time-mass tags), with each represented as the value for the integrated peak area for the specific mass and chromatographic retention time (Figure 5-4). In a typical analysis, this data matrix compiles information from more than 1500 analytical signals after automated peak alignment and integration with Markerlynx software. This was not equivalent to 1500 individual analytes as it was confirmed some of the RT- m/z pairs were adducts or minor fragment ions. Based on our analysis of mass spectra of a single peak and considering the number of adducts and fragments ions associated with each compound (approximately 5-7 ions), it was predicted that the total number of metabolites close to 200 or more, which means one metabolite/1.5 second by this approach.

Data analysis and visualization was performed on normalized, mean-centered and Pareto-scaled peak areas, using unsupervised principal component analysis (PCA). The results were displayed as loadings plots for the first two principal components, in which the loadings of each signal on the latent variables (principal components) are displayed. The sum of the loadings for each sample extract is then calculated, and these values are displayed as a scores plot, where each symbol represents a distinct sample. Clustering of samples based on metabolite profiles may be evident from the scores plot. PCA scores plot generated from LC/MS metabolite data of recurrent parent M82 and 65 ILs is displayed in Figure 5-5. Two distinct sample groups emerge, with ILs 1-3, 1-4, 8-1, 8-1-

1, 11-3, and 5-3 separated from all other ILs and M82. Further examination suggested the outlying samples could be divided into four clusters: ILs 8-1 and 8-1-1; IL 11-3, 5-3; 1-3, 1-4 and M82 (Figure 5-6, Left). The corresponding loadings plot (Figure 5-6, Right) aided identification of the analytical signals contributing most to the measured variance, and steered efforts toward metabolite identification of these metabolites. Once metabolites are recognized that distinguish specific genotypes, these compounds must be identified to aid attribution of functions to the responsible genes.

Figure 5-4. Metabolite data extracted using MarkerLynx software from LC/MS analyses showing a subset of 5 introgression line samples. The software first generates extracted ion chromatograms (XICs) for each mass, then detects and integrates the peak intensity, and aligns the peaks according to retention time. The metabolite identifications are assigned in the first column. Values in the various columns correspond to extracted ion peak areas for the retention times and masses listed in the third and fourth columns.

Metabolite annotation	Retention time (min)	Sample #		LC0720	LC0778	LC0657	LC0700	LC0821
		IL #:	m/z					
		4-4	4-3				3-5	3-1
S4:17 (5,5,5,2) [M+formate]-	3.4331	681.3257	2539.585	3026.702	2539.585	3041.419	3000.957	2573.2961
Tomatine [M+formate]-	2.8532	1078.547	920.7722	1107.242	920.7722	1252.612	1159.081	1076.7679
Rutin [M-H]-	2.4666	609.1765	850.0944	578.5131	850.0944	541.8746	911.7477	620.9218
S4:24 (5,5,1,2,2) [M+formate]-	3.9226	779.4285	892.3435	1392.364	892.3435	1262.97	1469.61	1001.6091
S4:17 (5,5,5,2) [2M+formate]-	3.4237	1317.604	783.0968	935.386	783.0968	993.8664	1402.32	854.6617
Internal standard 1 [M+formate]-	3.3644	379.2263	1713.306	1761.733	1713.306	1681.03	1561.184	1753.2177
S3:22 (5,5,1,2) [M+formate]-	3.839	737.4204	564.9627	1092.471	564.9627	858.3376	934.7832	801.7933
Rutin [M+formate]-	2.4658	655.1806	403.9035	241.0746	403.9035	251.4741	493.7012	304.1516
S3:15 (5,5,5) [M+formate]-	3.3419	639.3144	226.096	324.0673	226.096	314.6945	282.7648	298.0453
S4:16 (4,5,5,2) [M+formate]-	3.3413	667.3105	372.3629	503.1761	372.3629	431.5696	668.6205	448.3788
MW 356 [M+formate]-	4.1631	401.321	902.882	837.2664	902.882	896.6526	503.3472	840.4512
Dehydrotomatine isomer 2 [M+formate]-	2.8016	1076.528	108.4641	124.743	108.4641	196.6897	139.3052	106.3352
S4:15 (4,4,5,2) [M+formate]-	3.2495	653.297	0	9.1742	0	11.6098	22.2506	9.9217
S3:21 (5,5,1,1 or 4,5,1,2) [M+formate]-	3.7675	723.4037	20.8174	25.3627	20.8174	22.1522	38.4598	22.6197
Rutin [2M-H]-	2.4627	1219.3	63.4763	22.2724	63.4763	29.8779	98.5548	34.0199
Rhamnosyl glucosyl kaempferol [M-H]-	2.6172	593.1831	85.6215	92.8603	85.6215	185.4221	100.5453	215.7483
S4:23(4,5,1,2,2 or 5,5,1,1,2) [M+formate]-	3.8115	765.4122	25.422	30.7183	25.422	33.5228	47.2314	31.5883

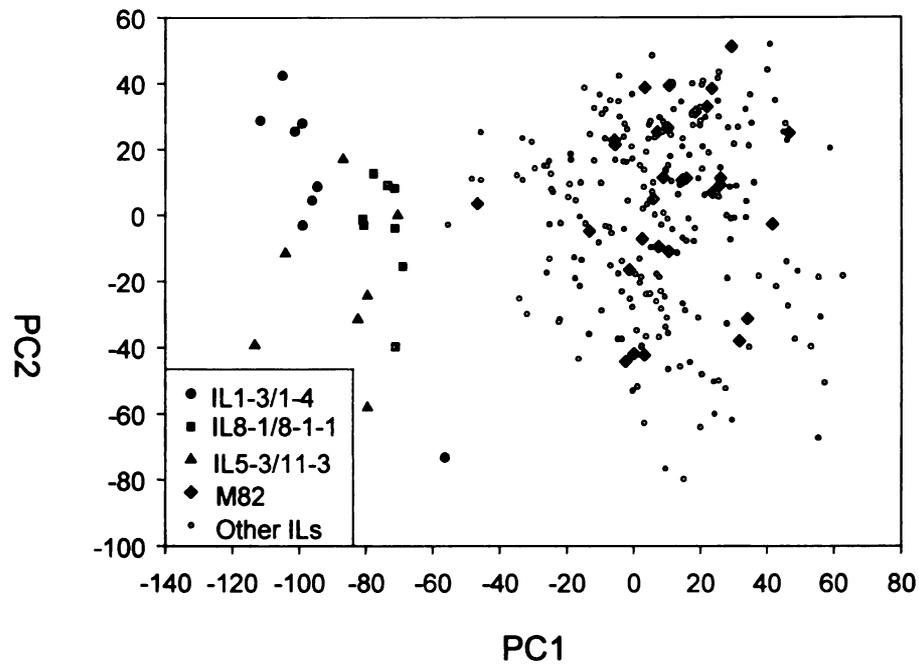
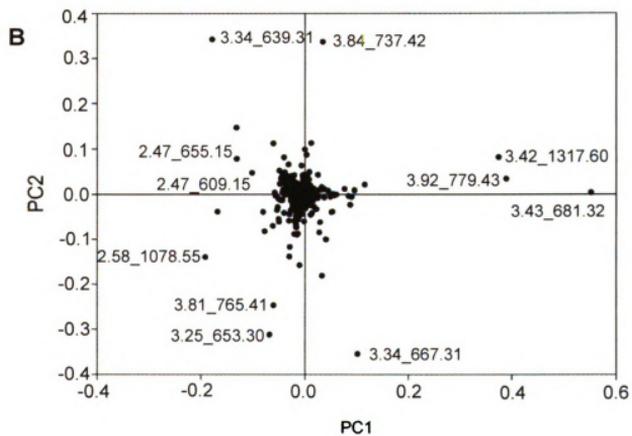
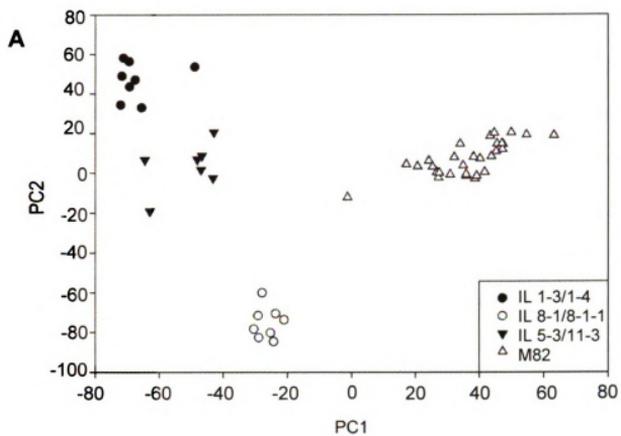


Figure 5-5 Principal component analysis (PCA) on 66 ILS plus recurrent parent M82 (total 276 samples) obtained using Pareto scaling with mean centering. PCA scores plot; Symbols: ILS 1-3, 1-4 (●); 5-3, 11-3 (▲); 8-1, 8-1-1 (■); Squares (◆) represent recurrent parent M82 (n=31); Cross (⊗) represent other 60 ILS (n = 3 or 4 for each IL).

Figure 5-6 (Left) PCA scores plot from LC/TOF MS metabolite data represent three clusters of ILs separated from recurrent parent M82 using Pareto scaling with mean centering. The clusters are 1-3, 1-4 (●); 5-3, 11-3 (▼); 8-1, 8-1-1 (○) and M82 (Δ). (Right) PCA loadings plot showed corresponding analytical signals (retention time-mass pairs) contributing to the difference among clusters in PCA scores plot.



5.3.1.1 ILs 1-3 and 1-4 are missing an acetyl group on major acyl sucrose metabolites

Two metabolites, of m/z 639 and 737, contributed most to distinguishing ILs 1-3 and 1-4 as judged by their positions near the top of the loadings plot. These contributions are clearly observed in the LC/MS chromatograms for M82 and IL1-3 shown in Figure 5-7. Metabolites at m/z 639.28 (RT 3.35 min) and 737.40 (RT 3.84 min) are more abundant in IL1-3, but were largely less from M82, which displayed abundant metabolites at m/z 681.30 (RT 3.42 min) and m/z 779.41 (RT 3.92 min) that were largely absent from IL1-3. ILs 1-3 and 1-4 stands out as having substantially greater abundance of this metabolite than the parent line M82 and all other ILs. Multiplexed CID approach was performed and yielded ion masses and mass accuracy within 10 ppm of true values for all molecular and fragment ions that have adequate ion statistics for mass measurement. Multiplexed CID provided spectra with minimal fragmentation at the lowest CID voltage, and progressively more extensive fragmentation at higher potentials. Processing of data generated under different quasi-simultaneous instrument settings aided recognition of substructures based on structure-specific fragment ions and assignments of putative metabolite structures. The great aspect of using multiplexed CID lies in it minimized need for doing MS/MS, which makes this fast LC/MS even more compelling.

S 3:15 from IL1-3 and S 4:17 from *S. lycopersicum* M82 were used to illustrate how to annotate metabolite structures. The annotation of metabolite structures were described in detail in chapter 3. Mass spectra of those two metabolites with CID voltage 55 V are described in Figure 5-8. For mass spectra of S 4:17 from *S. lycopersicum* M82, raising CID potential to 55 V generated numerous ions observed at coeluting retention time with

593.28 (losses of C2), 509.22 (loss of C5 plus C2), 425.17 (loss of two C5 plus C2), 341.11 (loss of three C5 plus C2) due to loss of fatty acyl groups as neutral fatty ketene. Fatty anion at m/z 101.06 corresponding to fatty anion of C5 fatty acid supported S 4:17 had C5 fatty acyl groups. Thus, acylsugar with formate adduct ion at m/z 681.29 was determined as a sucrose tetraester (S 4:17) with one C2 and three C5 fatty acyl groups. The structure of this metabolite is supported by NMR data (see detail, Table 6-1). Compared to formate adduct ion at m/z 681.30 for S 4:17, ion at m/z 639.28 for S 3:15 has fragments at 509.22 (loss of C5), 425.17 (loss of two C5), 341.11 (loss of three C5) corresponding to consecutive loss of neutral fatty ketenes and ion at m/z 101.06 (fatty anion of C5). Therefore, acylsugar with formate adduct ion at m/z 639.28 is sucrose tetraester (S 5:15), with one less acetate group compared to ion at m/z 681.30. Similar with these, acylsugars with ions at m/z 737.40 and m/z 779.41 were assigned as S 3:22 (5,5,12) and S 4:24 (2,5,5,12). The result indicated IL1-3 and 1-4 were two specific traits with much higher abundance of sucrose triesters, with one less acyl group compared to tetraesters. Therefore, this is consistent with the hypothesis that this introgression region is missing a gene necessary for normal acetyltransferase activity.

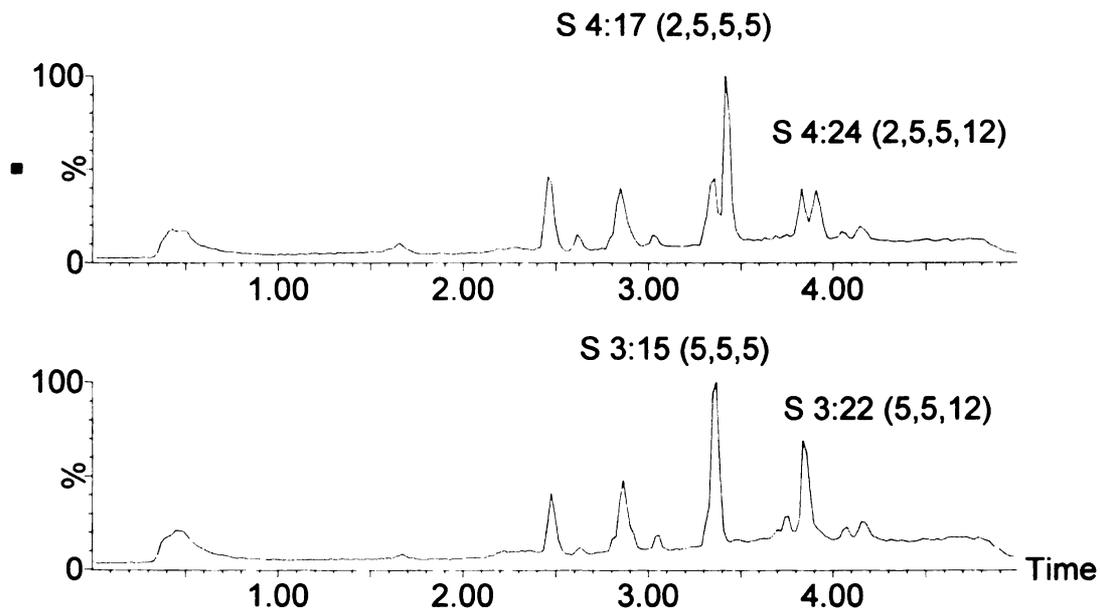


Figure 5-7 Total ion LC/MS chromatograms obtained from leaf dip of M82 (Top) and IL 1-3 (Bottom) using ESI negative mode. The peaks for most variance between two lines are labeled.

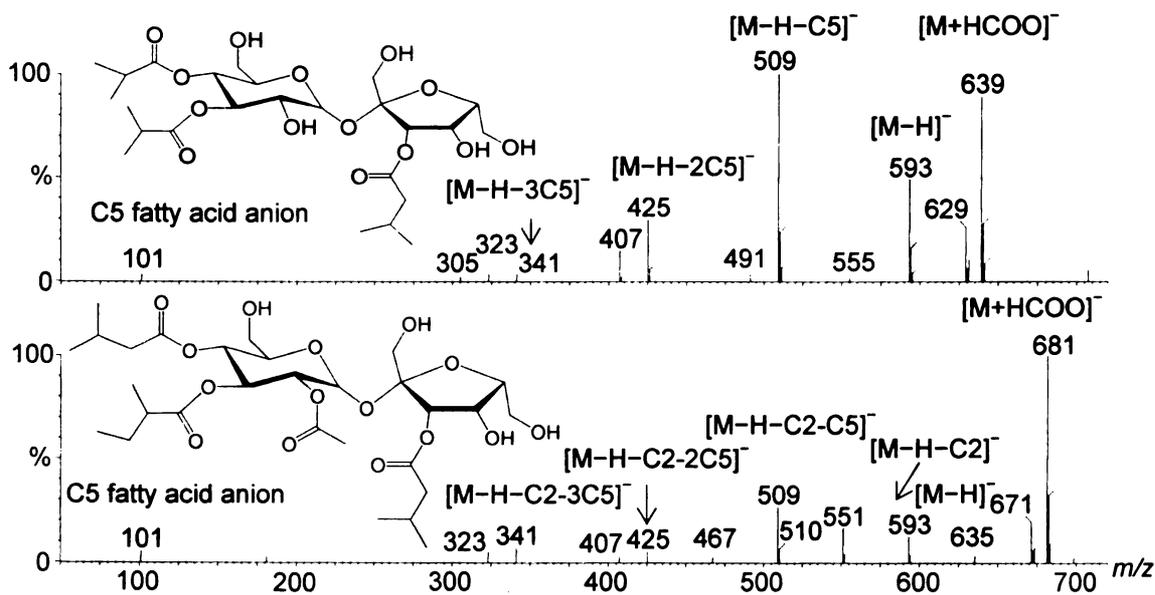


Figure 5-8 (Top) Mass spectrum of peak at RT = 3.35 min in IL 1-3 and (Bottom) Mass spectrum of peak at RT = 3.42 min from *S. lycopersicum* M82 with CID potential 55 V.

5.3.1.2 Two introgression lines with lower total acylsugars

Results from screening the ILs illustrate that the rapid LC-TOF MS method can reveal quantitative metabolite changes. In contrast to the effect of the *S. pennellii* chromosomal regions 1-3 and 1-4 and 8-1-1 on acylsugar substitution, two other ILs were found to cause an unexpected change in total acylsugar levels. ILs 5-3 and 11-3 were found to have changes in the proportions of acylsugars relative to tomatine (m/z 1078) and rutin (m/z 609) from PCA data. Quanlynx software in Masslynx 4.1 was performed to quantify the total amounts of detectable acylsugars among all ILs and M82. The nominal mass, retention time of corresponding acylsugars were edited into the quantification method (Table 5-1) and use chemical compound propyl-4-hydroxybenzoate (synthesized) as internal standard. Total peak areas of observable acylsugars were normalized to internal standard and dry leaf weight. ILs 5-3 and 11-3 showed the lowest amounts of acylsugars (Figure 5-9) compared to other ILs and M82, which supported the result from PCA. Decreased acylsugar levels caused by introgression of chromosomal segments from *S. pennellii* LA0716 into M82 tomato is counterintuitive because this wild tomato species accumulates much higher amounts of acylsugars than the cultivated tomato.²⁴⁴ Two general hypotheses would account for this unexpected behavior: either there are alleles of M82 important for normal acylsugar levels or alleles from LA0716, which when removed from the LA0716 genomic context, negatively influence acylsugar accumulation compared to the genes from M82 tomato. This high-throughput with PCA statistic tool greatly shortens the time to discover specific traits with phenotype among hundreds of samples.

Table 5-1. The detectable acylsugars for ILs were listed with retention time (min) and molecular mass for $[M+HCOO]^-$.

Acylsugars	Retention time	Molecular mass [M+HCOO]⁻
S 4:14	3.14	639.25
S 4:15	3.25	653.27
S 4:16	3.35	667.28
S 4:17	3.43	681.30
S 4:18	3.49	695.33
S 4:20	3.65	723.34
S 4:21	3.69	737.36
S 4:22(1)	3.74	751.37
S 4:22(2)	3.79	751.37
S 4:23	3.80	765.39
S 4:24	3.92	779.41
S 3:14	3.25	625.27
S 3:15	3.35	639.29
S 3:16	3.43	653.31
S 3:18	3.58	681.33
S 3:19(1)	3.65	695.35
S 3:19(2)	3.75	695.35
S 3:20	3.70	709.36
S 3:21	3.76	723.38
S 3:22	3.84	737.40

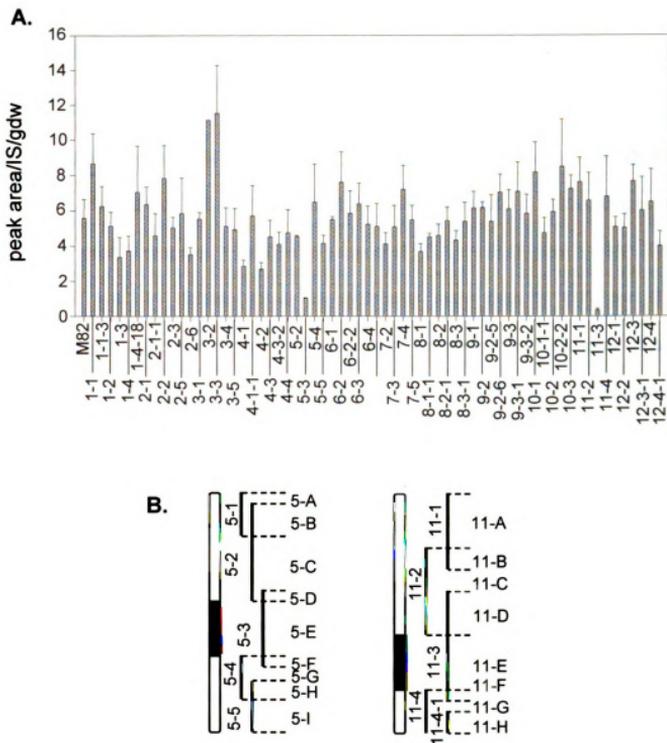


Figure 5-9 (A) Total amount of acylsugars for 66 IL_s and recurrent parent M82 based on Quanlynx software in Masslynx 4.1 (waters). Sum of all the peak areas of acylsugars were normalized to internal standard (Propyl-4-hydroxybenzoate) and dry leaf dip. Acylsugars are detected under ESI negative mode. The retention time and m/z of $[M+HCOO]^-$ corresponding to acylsugars are edit into method in Quanlynx. N=3, or 4 were tested for each IL that was grown except for IL3-2 (n = 1) and IL2-6 (n = 2). (B) Map of chromosomal introgressions for chromosomes 5 and 11.

5.3.1.3 ILs 8-1 and 8-1-1 causes a shift in acyl chain lengths without altering numbers of substitutions

ILs 8-1 and 8-1-1 also stand out from other ILs and M82 based on PCA scores plot and the signals responsible for this cluster in PCA loadings plot are m/z 667 (S 4:16), m/z 653 (S 4:15) and m/z 765 (S 4:23). TICs for IL8-1 and IL8-1-1 with changes in major acylsugar peaks compared to M82 also supported PCA result (Figure 5-10A). The major compound was 14 daltons lower in molecular mass than S 4:17 for IL 8-1-1. Fragmentation and accurate mass measurements indicated that this peak is a S 4:16 acylsucrose with a C4 fatty acid substituted for one of the C5 moieties (Figure 5-10 A and B). Note that this is a quantitative change and M82 still accumulates S 4:16 at ~ 50% of the level seen in IL8-1 and 8-1-1. Further analysis of the acylsugars in IL8-1-1 showed an increase in the abundance of acylsugars having one or more C4 acyl chains compared to M82 (Figure 5-10 A and B). Acylsugars lacking a C4 acyl chain however were decreased in IL8-1-1 compared to M82 (Figure 5-10 B).

This feature was also supported through comparing specific functional groups using multiplexed CID with aggressive CID condition. XICs of functional groups of C4 and C5 fatty acid from *S. lycopersicum* M82, *S. pennellii* LA0716 and IL 8-1 using CID potential 55 V are shown in Figure 5-11. Through comparison of the sum of integrated peak areas of fatty anions for C5 to C4 among those species, C5 shifts to C4 in IL 8-1 compared to M82, but similar with LA0716. Therefore, a great aspect for multiplexed CID is, observation of characteristic functional groups for specific class of metabolites produced under higher CID potential using multiplexed CID can assist profiling metabolic phenotypes.

The acyl chains present on the acylsugars are proposed to come from precursors of branched chain amino acid biosynthesis.^{245, 246} Because the C5 acyl chains can come from precursors of either leucine or isoleucine, two types of C5 acyl chains can be present. 3-methyl-butyrate (iso-C5 or iC5) comes from the leucine biosynthetic pathway and 2-methyl-butyrate (anteiso-C5 or aiC5) comes from the isoleucine pathway. The higher proportion of C4 acyl chains and the corresponding decrease in acylsugars with C5 acyl chains prompted us to further characterize the types of acyl chains present on acylsugars in M82 and IL8-1-1. Acyl chains on the acylsugars in leaf dips were trans-esterified to the corresponding ethyl esters for analysis by GC-MS. This type of analysis allows for determining whether acyl chains are branched or straight chain, and more specifically, the position of the branch for the C5 branched chain. Following ethyl esterification, the leaf dips from IL8-1-1 showed an increase in isobutyrate ethyl ester compared to M82 (Figure 5-2.) This increase in the C4 chain was associated with a decrease in 3-methylbutyrate ethyl ester (iC5) and not in the 2-methylbutyrate ethyl ester (aiC5). One possible reason for this phenotype could be that since the iC5 precursors are derived from the C4 precursor, a specific decrease in only the iC5 side chains in IL8-1-1 suggests that elongation of the C4 precursor to iC5 is perturbed in this line.

Figure 5-10 A. Total ion chromatograms from LC-TOF MS analysis of acylsugars in leaf dips of M82 and IL8-1-1. Labeling nomenclature for acylsugars - S 3:22 (5,5,12) is an acylsucrose with 3 acyl chains having a total of 22 carbons and numbers in parentheses indicate the lengths of the individual acyl chains. B. Amounts of acylsugars showing differences in abundance between M82 and IL8-1-1 are shown as integrated peak areas normalized to the internal standard and the dry weight of the extracted leaflet. Error bars indicate standard deviation with n = 4. C. Schematic representation of chromosome 8 introgressions showing locus controlling the acylsugar phenotype located on bin 8-B. Structure of S 4:17 (2,5,5,5) is based on LC-MS and NMR data. Structure of S 4:14 (2,4,4,4), an acylsugar detected only in IL 8-1-1 and not in M82, is inferred based on negative and positive mode LC-MS.

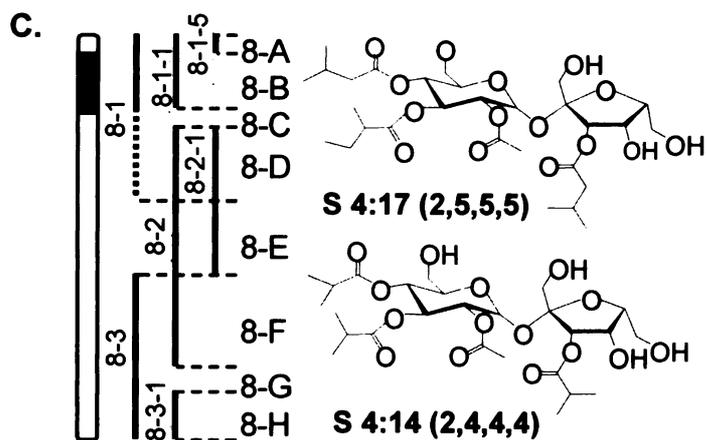
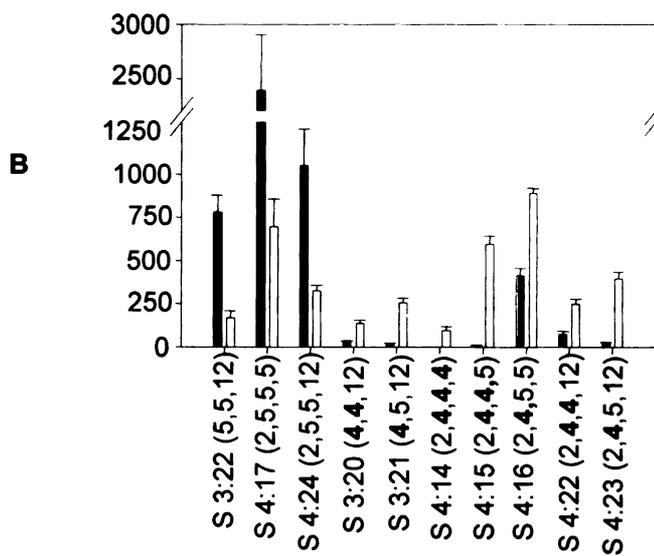
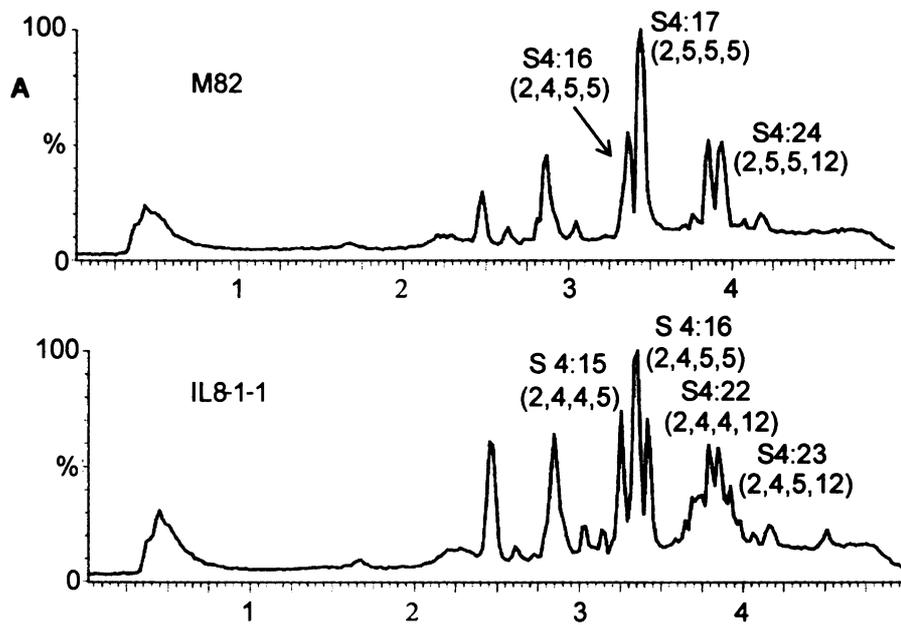
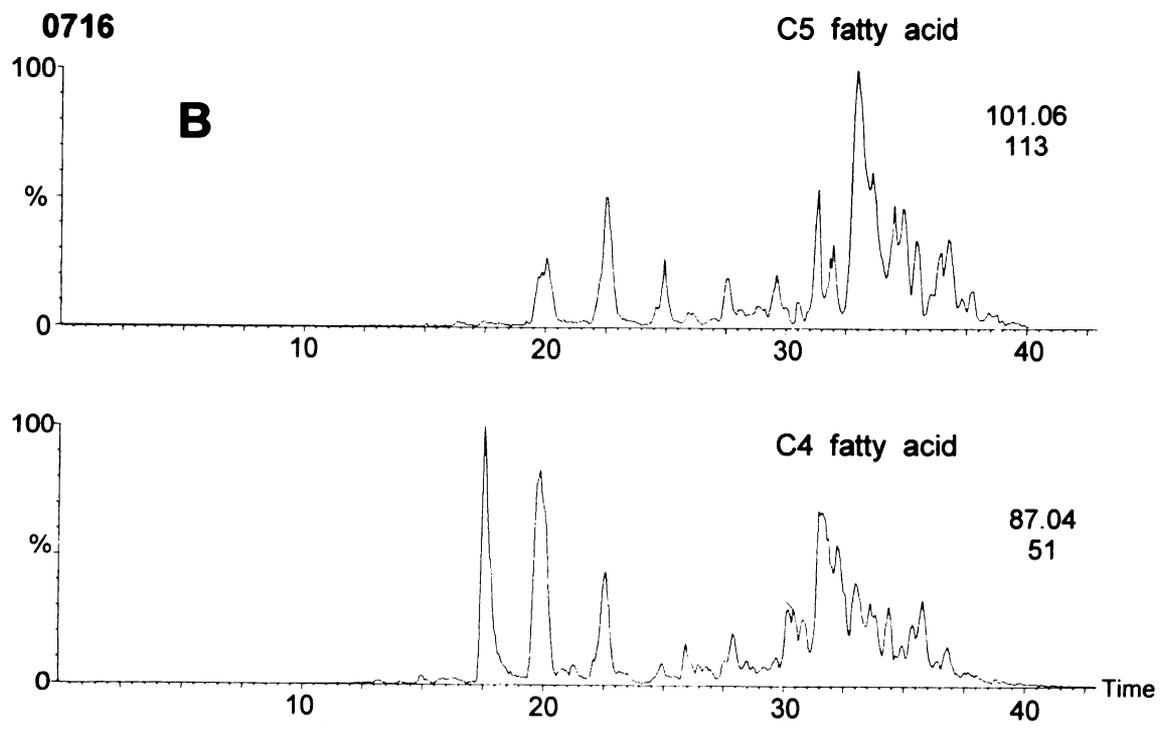
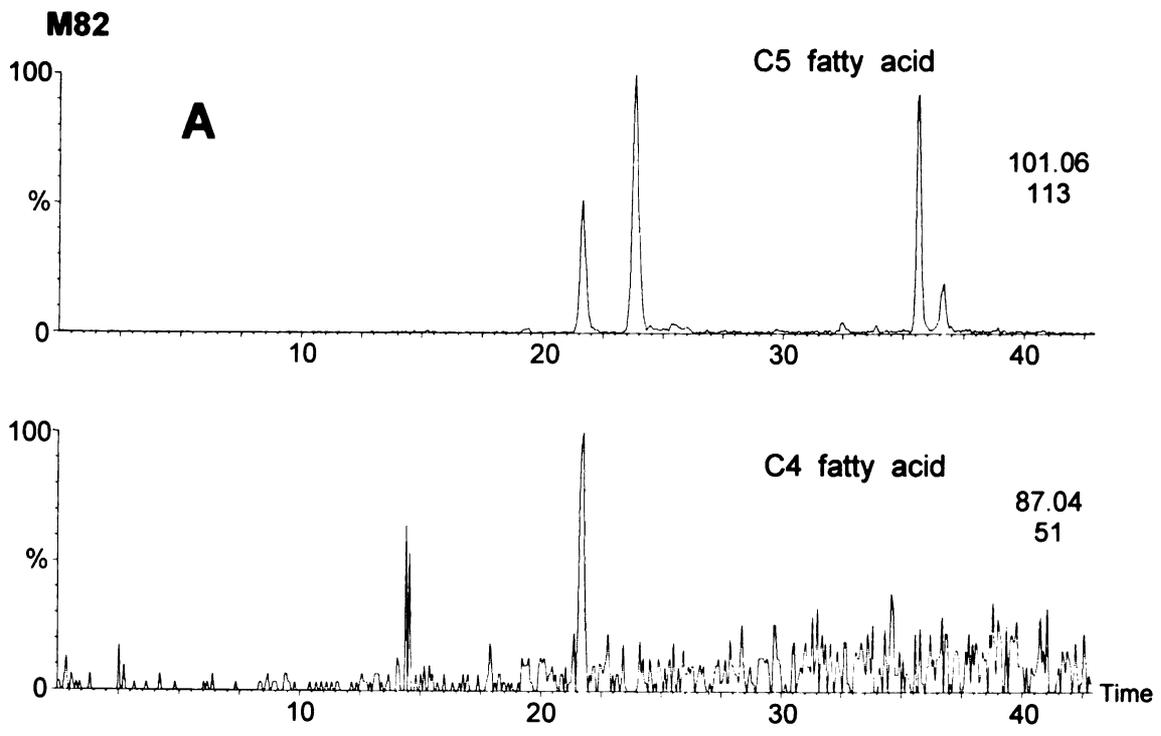




Figure 5-11 ESI negative LC/MS XICs of C4 m/z 87.04 and C5 m/z 101.06 fatty acyl groups cleaved from acylsugars with CID potential 55 V for A: *S. lycopersicum* M82 and B: *S. pennellii* LA0716. C: ILs 8-1. C5 fatty acyl group was predominated in M82, while comparable amount of C4 and C5 fatty acyl groups were in LA0716 and C5 fatty acyl groups shifted to C4 from IL8-1 based on the integration of peak areas for each fatty acyl group. D: Mass spectra of peak A for S 3:20 from IL8-1 using ESI negative mode with CID potential 55 V.



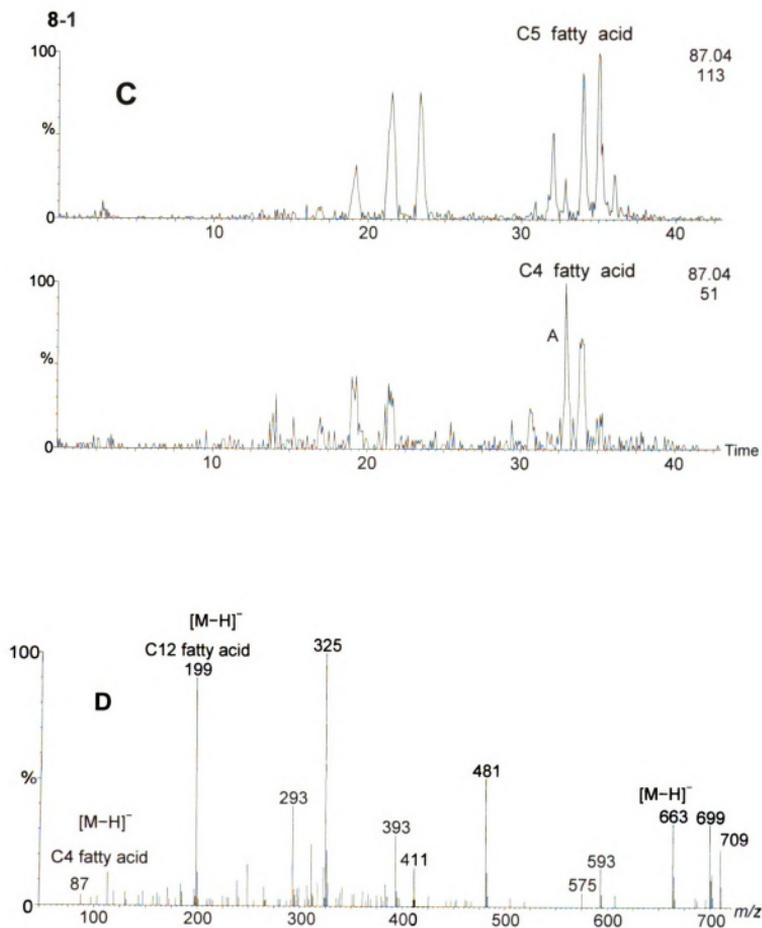


Figure 5-11 (Con't)

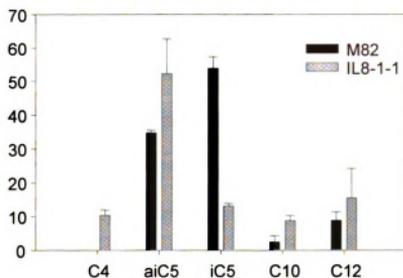


Figure 5-12 Side chains on acylsugars from leaf dip samples were transesterified to the corresponding ethyl esters and analyzed by GC/MS as described in chapter 6. The peak area % values for the base peak XIC for each fatty acid in M82 and IL8-1-1 are shown along with the standard error with $n = 3$ for IL 8-1-1 and $n=5$ for M82. For these data C4 ethylester levels were below limits of quantification for M82, which is presumed to arise from the high volatility of this product.

5.3.1.4 Discovery of differences in accumulation of metabolites of lower abundance

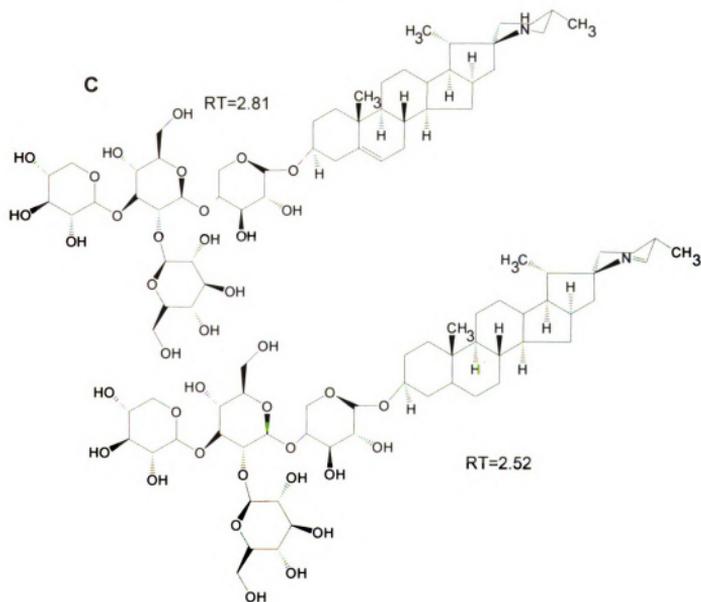
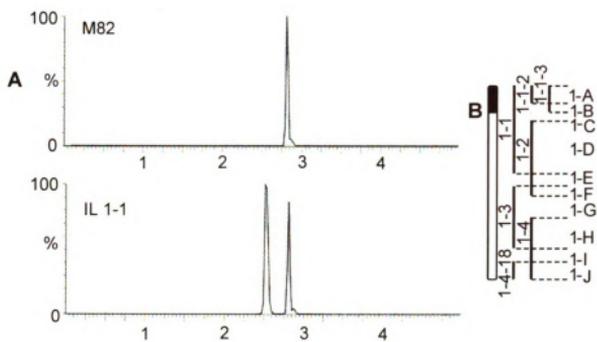
ILs 1-1 and 1-1-3 with novel glycoalkaloids

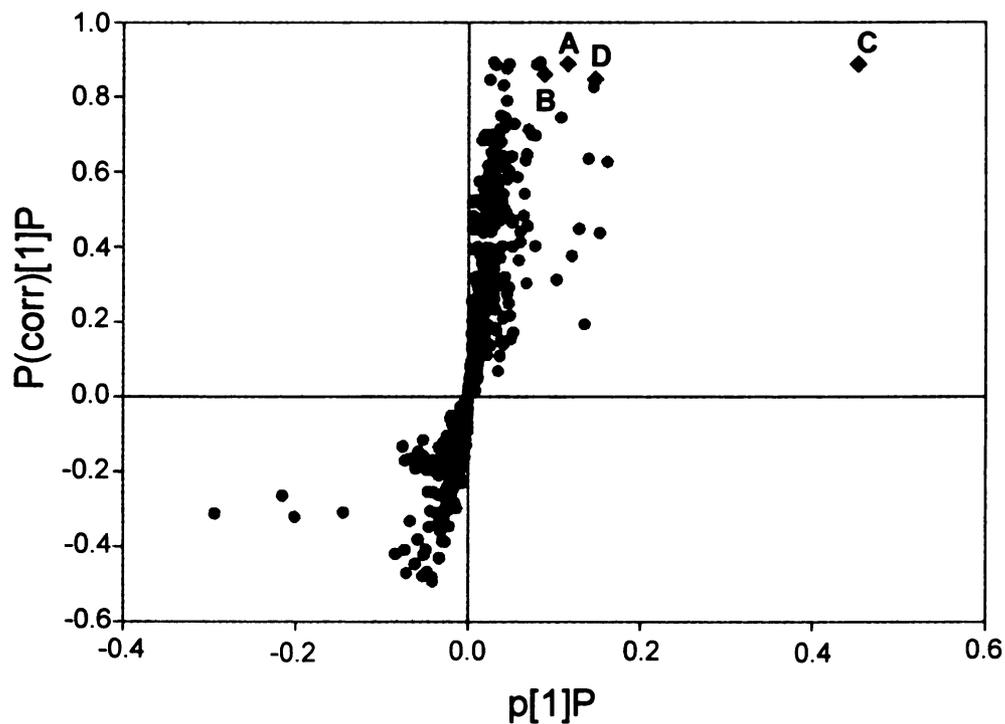
Abundant metabolites that distinguish NILs from a reference genotype can be recognized by visual inspection of LC/MS chromatograms, but discovery of accumulation of less abundant metabolites requires alternative approaches. Low abundance metabolites often are obscured in PCA plots, and rapid separations increase

the likelihood of metabolite coelution. These challenges can be overcome through automated generation of extracted ion chromatograms (XICs) for specific molecular or fragment masses of anticipated metabolites or metabolite fragment ions. In the current study, XICs for the glycoalkaloid dehydrotomatine (m/z 1076.5 for the formate adduct ion, negative ion mode) were generated and integrated, as this dehydrogenated form of tomatine had been reported in earlier studies²⁴⁷ (Figure 5-13). While the quantitative measurements showed modest differences in dehydrotomatine levels across the series of all NILs, the chromatograms revealed an earlier-eluting isomer of dehydrotomatine that was present only in the overlapping IL 1-1 and 1-1-3, and not observable in recurrent parent M82 and about 50 folds in another parent LA0716. Fragmentation patterns generated using positive ion mode suggested the position of the double bond in the early eluting isomer to be in the F-ring,¹⁹⁸ probably in the form of an amine group. Once NILs exhibiting a metabolic difference are identified, deeper probing of differences in metabolite accumulation is achieved using a supervised multivariate statistical analysis termed Orthogonal Projection to Latent Structures-Discriminant Analysis (OPLS-DA).⁹² This algorithm employs multiple linear regression and orthogonal filtering to aid recognition of metabolites that discriminate sample classes, in this case the two classes were M82 parent (class 1) and IL1-1-3 (class 2) (Figure 5-14) and the LC/MS data were automatically extracted using MarkerLynx software to yield an array of aligned and integrated peak areas. These processing steps generate lists of peak areas annotated with specific LC retention times and masses, and the statistical analysis assigns a $p(\text{corr})$ score suggestive of how well they discriminate classes, and this score is independent of the metabolite abundance. Analysis by OPLS-DA suggested three metabolites abundant in

ILs 1-1 and 1-1-3, and these were manually annotated as didehydrotomatine, hydroxytomatine, and the early eluting dehydrotomatine isomer based on fragmentation patterns and accurate mass measurements. All three were undetected in every one of the other 63 NILs. LC/MS signals for didehydrotomatine, hydroxytomatine metabolites ranged from 0.8-6 % of the signal for the abundant S4:17 tetraacylsucrose.

Figure 5-13. A. Extracted ion chromatograms of M82 and IL1-1 for m/z 1076.5 show an earlier eluting peak found only in the IL and not in M82. B. Schematic representation of chromosome 1 introgressions showing locus controlling the glycoalkaloid phenotype is located on bin 1-A or 1-B (IL1-1-2 was not analyzed in this study). C. Structures for dehydrotomatine isomers. Compound with RT = 2.81 min having the double bond in the B-ring has been described before.²⁴⁷ Structure of compound with RT = 2.52 min showing the double bond in the F-ring is inferred based on fragmentation from positive ion mode MS.





	Retention time_ <i>m/z</i>	p[1]P	p(corr)[1]P
A	2.51_1066.5019	0.114	0.889
B	2.47_1074.4997	0.088	0.860
C	2.49_1076.5256	0.453	0.887
D	2.60_1094.5345	0.146	0.847

Figure 5-14. OPLS-DA loading plot showed interclass difference between IL 1-1 (Right) and recurrent parent M82 (Left) using fused core column with 5 min gradient. The ions were labeled with retention time and *m/z*. The upper right corner and the lower left corner are corresponding to IL 1-3 (*n* = 3) and M82 (*n* = 4), respectively. A: $[M+Cl]^-$ of isomer of dehydrotomatine at *m/z* 1066; B: $[M+HCOO]^-$ of didehydrotomatine at *m/z* 1074; C: $[M+HCOO]^-$ of isomer of dehydrotomatine at *m/z* 1076; D: $[M+HCOO]^-$ of hydroxytomatine at *m/z* 1094.

5.3.1.5 ILs 6-3 and 6-4 with higher ratio of tri- to dimethylmyricetin

Mono-, di- and trimethylmyricetins were observed in *S. lycopersicum* M82 and structure annotation was described in chapter 4. Quanlynx software was used to quantify these three methylmyricetins among all ILs and *S. lycopersicum* M82 (Fig. 5-15) based on extracted ion chromatograms. IL6-3 and IL6-4 were discovered to have more percentage of trimethylmyricetin compared to mon- and di-methylmyricetin, but dimethylmyricetin was most abundance one among other ILs and recurrent parent M82, implying that methyltransferases originating from *S. pennellii* somehow alter the amount of trimethylmyricetin among these two ILs. These findings serve as the basis for continuing efforts to characterize O-methyltransferases involved in myricetin methylation in *S. habrochaites* LA1777 in the laboratories of Jones and Eran Pichersky (University of Michigan).

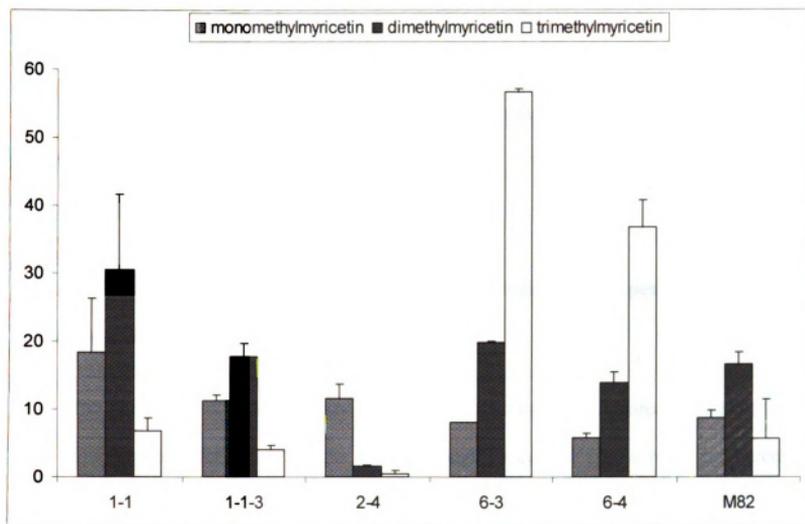


Figure 5-15. Total amount of mono-, di-, trimethylmyricetin for 66 ILs and recurrent parent M82 were obtained based on Quanlynx software in Masslynx 4.1 (waters). The peak area of mono-, di-, and trimethylmyricetin is normalized to internal standard (propyl-4-hydroxybenzoate) and dry leaf dip individually for all the samples. Myricetin is detected under ESI negative mode. The retention time and m/z of $[M+HCOO]^-$ corresponding to mono-, di- and trimethylmyricetin are edited into method in Quanlynx. $N = 3$ for ILs and $n = 30$ for M82. ILs 6-3 and 6-4 showed the different phenotypes compared to other ILs and samples. ILs 6-3 and 6-4 has more abundance trimethylmyricetin, while for M82 and other ILs, dimethylmyricetin is predominant.

Combination of PCA, OPLS-DA and fast LC/MS, five specific phenotypes were discovered from 65 ILs within minimal analysis time. Then genotypes mapping for F2 plants were performed to determine the phenotypes dominant or recessive compared to their recurrent parent *S.lycopersicum* M82. The result of fast screening of F2 plants using fast LC/MS was also used to guide genotype mapping. F2 plants were obtained by self-fertilization of F1 plants.

5.3.2 F2 plants are screened whether the metabolic phenotypes dominant or recessive for recurrent parent M82

The phenotype of IL1-4 was sucrose trimesters predominant compared to tetraesters. PCA scores plot for 22 F2 plants of IL1-4 and 10 control plants recurrent parent *S.lycopersicum* M82. 17 out of 22 (~80%) displayed the same phenotype with *S.lycopersicum* M82, which suggested phenotype of IL1-4 was *S.lycopersicum* M82 predominant. Quanlynx software was also performed for acylsugars and compared the ratio of sucrose trimesters to tetraesters. Figure 5-18 displayed the ratio of sucrose trimesters to tetraesters for all those F2 plants of 1-4 and control plants M82. The result showed 17 F2 IL1-4 out of 22 had predominant sucrose tetraesters instead of trimesters, same as M82, which supported PCA result.

Figure 5-18. Total amount of sucrose triesters and tetraesters were obtained based on Quanlynx software in Masslynx 4.1 (Waters) for 22 F2 plants of 1-4 and 12 control plants M82. 17 out of 22 showed the similar ratio of sucrose triesters to tetraesters as *S.lycopersicum* M82, which supported PCA scores plot for F2 plants of 1-4.

Around 2000 F2 plants were screened using this fast LC/MS approach during the last two years in our lab to guide genotype mapping, demonstrating the suitability of this approach for large-scale phenotype analysis. Exploiting these capabilities continues to assist our efforts to locate genes to specific chromosomal regions and accelerate the discovery of gene functions relevant to plant chemistry.

5.4. Conclusions

Fast LC/TOF MS with multiplexed CID has been demonstrated as a powerful tool to profile metabolic phenotypes in plant metabolomics within minimal analysis time. Three advances provided by this fast LC/TOF MS approach for metabolite profiling of complex *Solanum* extracts are fast detection of many metabolites (roughly one metabolite per second of instrument time), high chromatographic resolution, and accurate mass measurements for molecular and fragment ions for structure elucidation. This approach has generated analyses of 300 samples over a 30 h period and yielded more than 1500 analytical signals after automated peak alignment and integration. Five specific phenotypes including 10 ILs were discovered among 65 ILs within one week under assistance by PCA and OPLS-DA. The phenotypes including low abundance acylsugars in IL5-3, 11-3; fatty acyl groups shift from C5 to C4 in IL8-1, 8-1-1; acyl group missing in IL1-3, 1-4; methylflavonoid ratio shift in IL6-3, 6-4 and isomers of dihydrotomatine in IL1-1, 1-1-3. Significant novel markers responsible for those phenotypes were identified using LC/MS with multiplexed CID.

Around 4000 plant extracts were screened during last two years in our lab for gene mapping using this fast LC/MS approach. The genes regulating metabolic phenotype for

low abundance acylsugars were limited to a small chromosome region. Once gene sequence of tomato become public, we can rapidly limit the gene location for those phenotypes with the help of fast LC/MS.

Chapter 6. Qualitative and quantitative profiling of acylsugar metabolic phenotypes for tomato breeding lines using liquid chromatography/time-of-flight mass spectrometry

6.1 Introduction

Glandular trichomes of *Solanum* species secrete acylsugars enhance resistance of the plants to several species of insects.^{128,248-251} These highly viscous lipids constitute a significant proportion of leaf biomass in the *Solanaceae* and are produced in particularly large amounts in the wild tomato species *Solanum pennellii* accession LA0716 (~ 20% leaf dry weight).²⁵² Certain wild tobacco (*Nicotiana*) species produce acylsugars at up to 15% leaf dry weight, and these metabolites are also produced, but at lower levels, in the tobacco species *Nicotiana benthamiana*.²⁵³⁻¹⁰⁸ Increasing acyl sugar production has long been a target of tomato and potato (*Solanum tuberosum*) breeding programs^{248, 254-256} because of the potent antifeedant, antioviposition, and in some cases toxic properties of these compounds toward insect pests. Many acyl sugars exhibit insecticidal activity exceeding the efficiency of a comparable standard, insecticidal soap.¹⁶¹ Trichome-derived acyl sugars have stimulated great interest owing to their functions as alternatives to chemically synthesized sucrose ester insecticides, as antibiotics, and for consumer products such as cosmetics.^{163,257} Therefore, development of cultivars with increased levels of acylsugar metabolites would provide an important alternative that could to reduce applications of chemical insecticides, allowing for diminished environmental impact. Knowledge of the genes involved in acylsugar biosynthesis will provide information needed to manipulate production of acylsugars and other metabolites. While production of genetically modified transgenic plants would stimulate controversy owing to the potential that a modified genome will cause changes in the patterns or concentrations of metabolites,⁴⁰ traditional breeding approaches that involve genetic crosses of two naturally occurring plants is considered more acceptable by many.

Therefore, understanding the difference of metabolic phenotypes between “normal” plants and “modified” plants through classical breeding, mutation or manipulation can facilitate generation of “natural” variants that incorporate a metabolic trait from a wild species into a commercially viable plant such as tomato.

The research described in this chapter has involved collaboration with an expert in tomato breeding for acylsugar content, Professor Martha Mutschler of Cornell University, to refine a series of unique acylsugar-producing tomato breeding lines. The long-term goal is to improve acylsugar production in tomato and develop insect-resistant tomato lines. The lines investigated in this study were derived from crosses between the conventional tomato *S. lycopersicum* and its wild relative *S. pennellii* LA0716. The latter accumulates large amounts of glucose-based acylsugars in trichomes, but before this study, conventional tomato was considered to not produce any acylsugars. These efforts involved both identification and quantitative comparisons of acylsugar profiles in various plant genotypes. Profiling of acylsugars was accomplished using LC/TOF MS, and multiplexed CID was applied to enhance metabolite identification and for quantitative comparisons of metabolite levels across breeding lines. The basics underlying annotation of acylsugar metabolite structures using LC/TOF mass spectrometry with multiplexed CID was discussed earlier in chapter 3. This chapter will focus on quantitative comparisons of glucose and sucrose esters and branched- or straight-chain fatty acid esters among different breeding lines. One of the aims of this study has been to demonstrate the performance of metabolite profiling using LC/TOF MS coupled with multiplexed CID for accelerating development of crops with desired metabolic traits.

6.2 Experimental Section

6.2.1 Materials. Plant breeding lines were grown at Cornell University, and extracts of plant leaves were prepared from freshly harvested leaves. These extracts were shipped to Michigan State University on dry ice. Leaf extracts from wild type species LA0716, LA1522, LA1777 were grown in green house of Michigan state university.

6.2.2 Chemicals. Acetonitrile (HPLC grade), 2-propanol (HPLC grade), methanol (HPLC grade), formic acid (88%, CHCl_3 (HPLC grade) were from Fisher Scientific of VWR Scientific. Thionyl chloride (analytical grade), 1,2,5,6-di-O-isopropylidene- α -D-glucofuranose (98%), and decanoic acid were purchased from Sigma-Aldrich.

6.2.3 Preparation of internal standard (3-decanoyl-glucofuranose). 570 mg of decanoic acid was dissolved in 1 ml thionyl chloride and refluxed in a 25 ml round bottom flask for 4 h. After cooling to room temperature, volatiles were evaporated under aspirator vacuum to dryness, and 5 ml CHCl_3 was added to the flask, followed by 100 mg 1,2,5,6-di-O-isopropylidene- α -D-glucofuranose. The mixture was refluxed for another 5 h, then cooled to room temperature. The reaction mixture was evaporated to dryness under vacuum using a rotary evaporator. Raw product was dissolved in 2 ml methanol and hydrolyzed under 1 ml 0.1 M NaOH, and then purified by C18 solid phase extraction column, eluting with methanol/water. Yield was 31.3 %, and consisted of a mixture of four isomers as determined by LC/MS. It is expected that the isomers represent products of epimerization at the 1-position and contributions of two epimers of 3-decanoylglucopyranose formed by ring opening and closing to the pyranose form. The

structure of one isomer of the 3-decanoylglucofuranose product used as an internal standard for acylsugar analyses is shown in Figure 6-1.

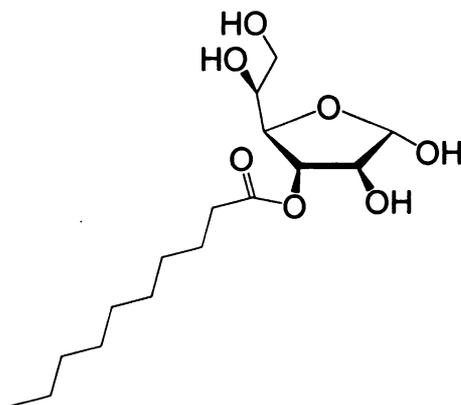


Figure 6-1 Structure of synthesized internal standard 3-decanoylglucofuranose.

6.2.4 Extraction Method. To extract metabolites from *Solanum* trichomes, ~10 mg leaves were dipped in 2 ml isopropanol : CH₃CN : H₂O (3:3:2 v/v/v) for 1 min. Extracts were centrifuged, and the supernatants were analyzed without further processing. Extracts were stored at -20 °C until they were thawed for LC/MS analysis.

6.2.5 High-Performance Liquid Chromatography. A Shimadzu (Columbia, MD) LC-20AD HPLC ternary pump was used. The separation was performed on a Thermo 1 × 150 mm BetaBasic C₁₈ column with 3 μm particles. A gradient was executed based on 0.15% aqueous formic acid (A) and methanol (B). Gradient profile consisted of: initial conditions 5% B; linear gradient to 50% B (5 min); 95% B (33 min); 100% B (35 min); hold at 100% B until 38 min; return to 5% B (43 min). The flow rate was 0.1 ml·min⁻¹, and the injection volume was 10 μl. The column was maintained at 30 °C.

6.2.6 Mass Spectrometry. Analyses were performed using a Waters LCT Premier mass spectrometry (Milford, MA). Capillary voltage of 3.2 kV was used in both positive and negative modes, using the following experimental parameters: desolvation gas flow, 300 L·h⁻¹; desolvation gas temperature, 200 °C; source temperature, 90 °C; cone gas flow, 20 L·h⁻¹. Mass spectra were acquired into five different acquisition functions using rapid switching of Aperture 1 voltage in the ion transit region of the mass spectrometer, providing quasisimultaneous generation of spectra under fragmenting and nonfragmenting conditions. Mass spectra were acquired over *m/z* 50 to 1500 with a scan time of 0.4s per function. Spectra were acquired in centroid format with ‘dynamic range enhancement’ (DRE) enabled. Prior to performing all experiments, the instrument was calibrated in both ionization modes over *m/z* 50-1500 range. This instrument was equipped with a 4.0-GHz time-to-digital converter and a reflectron, and typically yielded a resolving power of 5000 (fwhm). 0.1% H₃PO₄ was infused through the reference sprayer of the lock-spray source as a lock mass.

6.2.7 Gas Chromatography/Mass Spectrometry

The mixed solvent (acetonitrile: isopropanol: water (3:3:2, v/v/v) acidified with 0.1% formic acid) were evaporated using speed-vac to dryness. Then, add 20 µl 21% sodium ethoxide (in denatured alcohol) to 1.5 ml microcentrifuge tubes with dry trichome extraction residue and the transesterification allowed reaction to precede for 30 min at room temperature. The reaction mixture was washed with a saturated sodium chloride solution (50 µl) to remove 50 µl of the top, hexane layer and transfer to autosampler vials with low volume inserts. The sample was analyzed by capillary gas chromatography–mass spectrometry (GC-MS on a 10 m, 0.1 mm (i.d.) fused-silica column with a 0.34 µm

thick stationary phase (DB-5; Agilent, Santa Clara, CA, USA). The GC program was as follows: injector temperature 280°C; initial column temperature 36°C held for 4 min; ramped at 10°C/min to 150°C, 20°C/min to 220°C; and finally at 40°C/min to 300°C, held for 2 min. The helium carrier gas flow was set to 0.4 ml/min. All compounds were analyzed using an Agilent 5975B quadrupole mass spectrometer (Santa Clara, CA, USA) operated in 70 eV electron ionization mode using an Agilent 6890N GC system. This approach is modified from Walters and coworkers' paper.¹¹²

6.2.8 Chemometric data analysis

All data obtained from ESI-MS of the trichome samples were extracted using MarkerLynx 4.1 (Waters, Manchester, U.K.). Mass spectral data were processed over the entire retention time and m/z ranges using automated peak detection, alignment, and integration. The parameters used included mass tolerance of 20 ppm, heavy isotopic peaks were excluded for analysis, noise elimination level was set at 10.00, minimum intensity was set to 10 % of base peak intensity, maximum masses per retention time was set at 20 and finally, chromatographic retention time tolerance was set at 0.5 min. A list of the integrated areas of the peaks detected was generated for all the samples, using RT and m/z data pairs as the identifier of each analytical signal. These integrated areas were normalized to the sum of all integrated signals for each sample. In separate analysis, Quanlynx software was used to integrate peak areas of XICs for formate adducts of targeted acylsugar metabolites.

6.2.9 Extraction and isolation of S 4:17

Sixty grams of freshly harvested leaves from 3-week old *S. lycopersicum* cv. M82 were extracted by dipping leaf material into 1.0 liter of isopropanol: acetonitrile: water (3:3:2, v/v/v) containing 0.1% formic acid for 5 min. The resulting extract was partitioned into two phases by addition of 500 ml CHCl₃. The lower chloroform layer was collected and evaporated to dryness on a rotary evaporator. The residue was dissolved in 10 ml Methanol: H₂O (1:1, v/v) for further purification using a Supelclean™ LC-C18 solid phase extraction (SPE) (10 ml column). A 0.2 ml aliquot was loaded onto the SPE column for each step of the process, and the process was repeated 25 times to accumulate sufficient quantity for NMR analysis since we used two columns each run. Material was

eluted using mixtures of methanol/water, using 0.1% formic acid in the aqueous component for the following methanol/aqueous compositions: Fraction 1 = 20 ml of Methanol:H₂O (9:11, v/v); Fraction 2 = 30 ml of Methanol:H₂O (11:9, v/v); Fraction 3 = 30 ml of Methanol:H₂O (13:7, v/v), and Fraction 4 = 50 ml of Methanol:H₂O (7:3, v/v). A wash with 20 ml 100% methanol was used to clean the column after each elution. After the presence of the desired component was confirmed by flow injection analysis-electrospray ionization MS in negative ion mode, the solvent was evaporated and all of the material collected for Fraction 4 was pooled to obtain S 4:17 for NMR analysis.

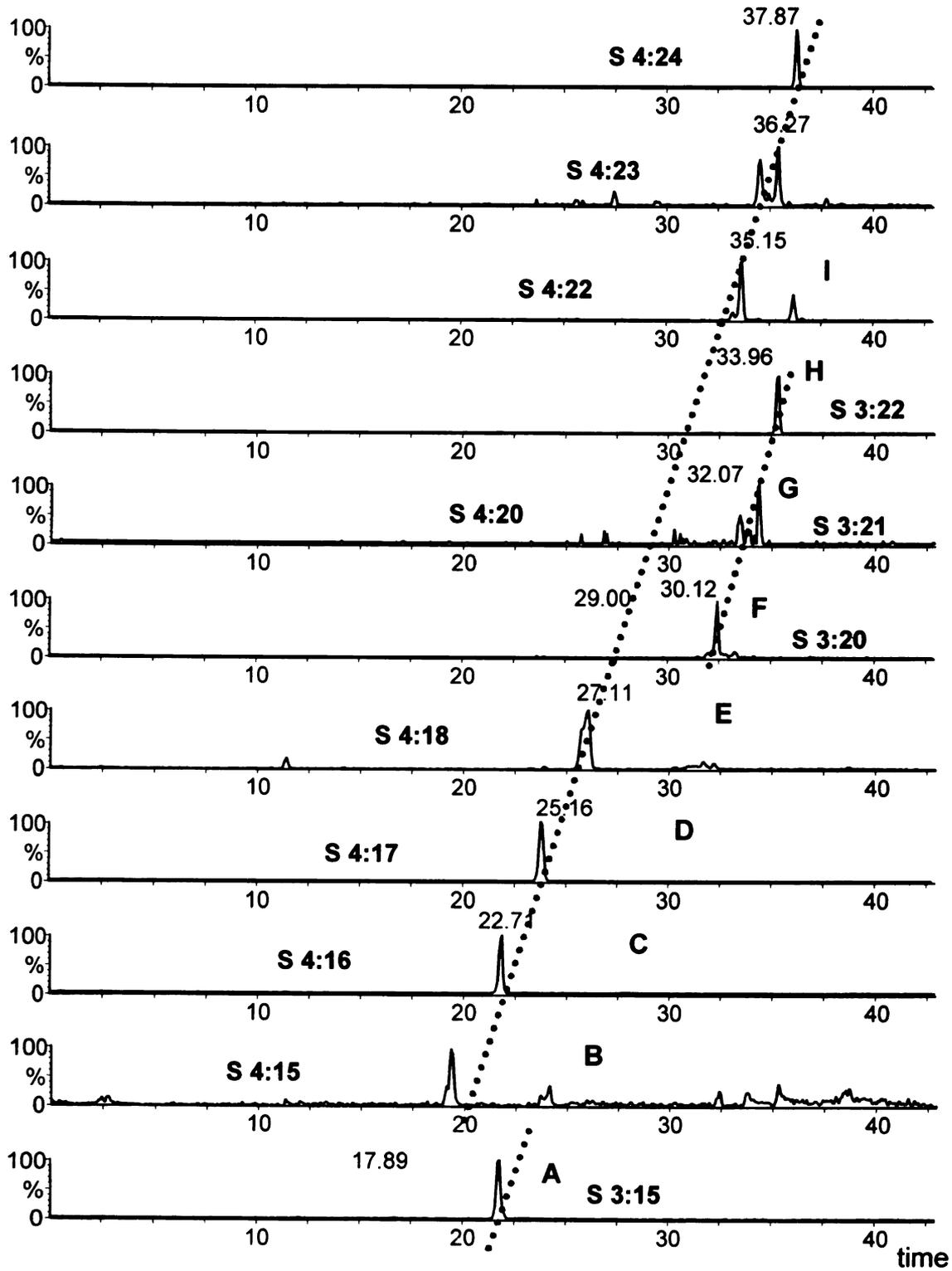
6.3 Results and Discussion

6.3.1 Annotation of acylsugar from conventional tomato *Solanum lycopersicum* M82.

One parent for some of Professor Mutschler's breeding lines is the conventional tomato, line M82 of *S. lycopersicum*. Therefore, understanding the metabolic pattern of

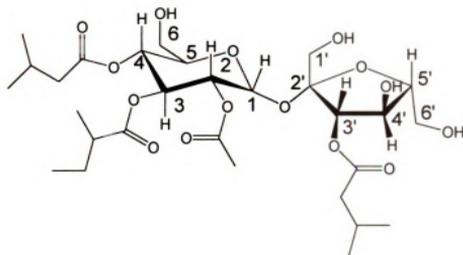
acylsugars from this species is important for breeders to track the genetic basis of phenotypes for progeny of this line. However, acylsugars had not been detected from *S. lycopersicum* M82 prior to this study, presumably due to the limited sensitivity of those original analytical procedures.^{132,169} This report describes, for the first time, the accumulation of acylsugars in cultivated M82, with twelve distinct peaks in the LC/MS results suggestive of tri- and tetraacylsucroses. Figure 6-2 shows the extracted ion chromatograms for formate adducts of acylsugars for an extract of *S. lycopersicum* M82.

Figure 6-2 XICs of formate adducts from S 4:15 to S 4:24 (highlighted with red dash line) and S 3:15, S 3:20 to S 3:22 (highlighted with blue dash line) from *S. lycopersicum* M82.



Four major acylsugars annotated as S 3:15, S 4:17, S 3:22 and S 4:24 were detected in accession M82. First, the most abundant component S 4:17 was purified and the structure was characterized using NMR. Isolation of other three is underway by undergraduate student Thomas Westbrook. Structural information for S 4:17 was generated from ^1H , ^{13}C , DEPT, gHMQC, gHMBC, gCOSY, and TOCSY NMR experiments. NMR spectra are in Appendix.

Table 6-1. Structure of S 4:17 in *Solanum lycopersicum* M82 characterized by NMR (^1H , ^{13}C , DEPT, gHMQC, gHMBC, gCOSY, and TOCSY). Chemical shifts of proton and ^{13}C resonances are listed with proton-proton coupling constants.



Carbon # (multiplicity)	¹ H mult. (J, Hz)	¹³ C
1(CH)	5.61 <i>d</i> (3.7)	89.3
2(CH)	4.85 <i>dd</i> (10.6, 3.7)	70.5
2-O-		
-1 (CO)	----	170.2
-2 (CH ₃)	1.98 <i>s</i>	20.6
3 (CH)	5.44 <i>dd</i> (10.6, 10.0)	68.9
3-O-		
-1 (CO)	----	175.4
-2 (CH)	2.29 <i>m</i>	40.9
-2 (-CH ₃)	1.03 <i>d</i> (7.0)	16.4
-3 (CH ₂)	1.56, 1.37 <i>sext</i> (7.0)	26.5
-4 (CH ₃)	0.81 <i>t</i> (7.4)	11.4
4 (CH)	4.92 <i>dd</i> (10.2, 10.0)	68.5
4-O-		
-1 (CO)	----	172.2
-2 (CH ₂)	2.1 - 2.4 <i>m</i> ^b	42.9
-3 (CH)	2.15 <i>m</i>	25.8
-4 (CH ₃)	0.9 <i>d</i> (7) ^a	22.3
-5 (CH ₃)	1.0 <i>d</i> (7) ^a	22.3
5 (CH)	4.09 <i>m</i>	71.5
6 (CH ₂)	3.56, 3.60 <i>m</i>	61.4
1' (CH ₂)	3.47, 3.58 <i>d</i> (12.3)	64.2
2' (C)	----	103.9
3' (CH)	5.21 <i>d</i> (7.9)	78.7
3'-O-		
-1 (CO)	----	174.0
-2 (CH ₂)	2.3 - 2.4 <i>m</i> ^b	43.0
-3 (CH)	2.03 <i>m</i>	25.3
-4 (CH ₃)	0.91 <i>d</i> (7) ^a	22.3
-5 (CH ₃)	1.01 <i>d</i> (7) ^a	22.3
4' (CH)	4.45 <i>dd</i> (8.0, 8.0)	71.5
5' (CH)	3.91 <i>m</i>	82.5
6' (CH ₂)	3.70, 3.85 <i>m</i>	60.3

Twelve signals between δ 59 and 103 in the ^{13}C NMR spectrum, which were assigned (by DEPT) to seven oxymethyngs, three oxymethylenes, one anomeric methyne (δ 89.6), and one nonprotonated anomeric carbon (δ 102.2), were indicative of a disaccharide. The ^1H NMR signals and the connectivities observed in the gCOSY spectrum revealed both glucose and furanose rings in the disaccharide. The HMBC correlation from H-1 of glucose (δ 5.61) to C-2' of fructose (δ 102.2) completes the identification of the disaccharide as sucrose²⁵⁸. The low-field shifts of the H-3, H-3', H-4, H-2 signals and the presence of carbonyl signals at δ 170.2, 175.4, 172.2, 174.0 were assigned as 2, 3, 4, and 3' carbons of sucrose with ester attachments. ^1H , ^{13}C , COSY, HMQC and HMBC spectra allowed identification of the acylgroups as one 2-methylbutanoyl, two 3-methylbutanoyls and one acetyl. The positions of these groups were established by the HMBC correlations from H3 to C-1 (δ 175.4) for 2-methylbutanoyl, H-4 to C-1 (δ 172.2) for 3-methylbutanoyl, H-2 to C-1 (δ 170.2) for acetyl, H-3' to C-1 (δ 174.0) for another 3-methylbutanoyl. Therefore, acylsugar S 4:17 is substituted with an acetate ester at the 2-position, iso-C5 esters at the 4- and 3'- positions, and an anteiso-C5 ester at the 3-position of sucrose.

Knowledge of the structure of S 4:17 made it possible to better interpret multiplexed CID spectra generated using ESI and both negative and positive ion modes (Figure 6-3). The anion at m/z 101.06 corresponds to deprotonated C5 fatty acid, and several fragment ions correspond to losses of fatty ketenes from $[\text{M}-\text{H}]^-$ as follows: m/z 593.27 (loss of C2), 551.24 (loss of C5), 509.23 (loss of C2 plus C5), 425.16 (loss of C2 plus two C5), 341.11 (loss of C2 plus three C5, corresponds to $[\text{M}-\text{H}]^-$ of sucrose). In positive mode, $[\text{M}+\text{NH}_4]^+$ ions undergo cleavage of the glycosidic bond to form two prominent fragment

ions with the more abundant at m/z 247.12 corresponding to the residue from the fructofuranose ring with one C5 fatty acid ester, and the other at m/z 373.22 corresponding to the glucopyranose residue with two C5 and one C2 ester groups.

The acylsucrose metabolites observed in trichome extracts of leaflet of *S. lycopersicum* M82 are described in Table 6-2, which includes major fragment ions generated for both positive and negative mode ESI mass spectra. It is worthy to note that all the detected sucrose esters produce the fragment ion at 247.12 in positive ion mode, which suggested all have one C5 fatty acyl group on the fructofuranose ring. This result, combined with the NMR data, suggest remarkable selectivity of substitution positions of fatty ester groups in tomato acylsucroses.

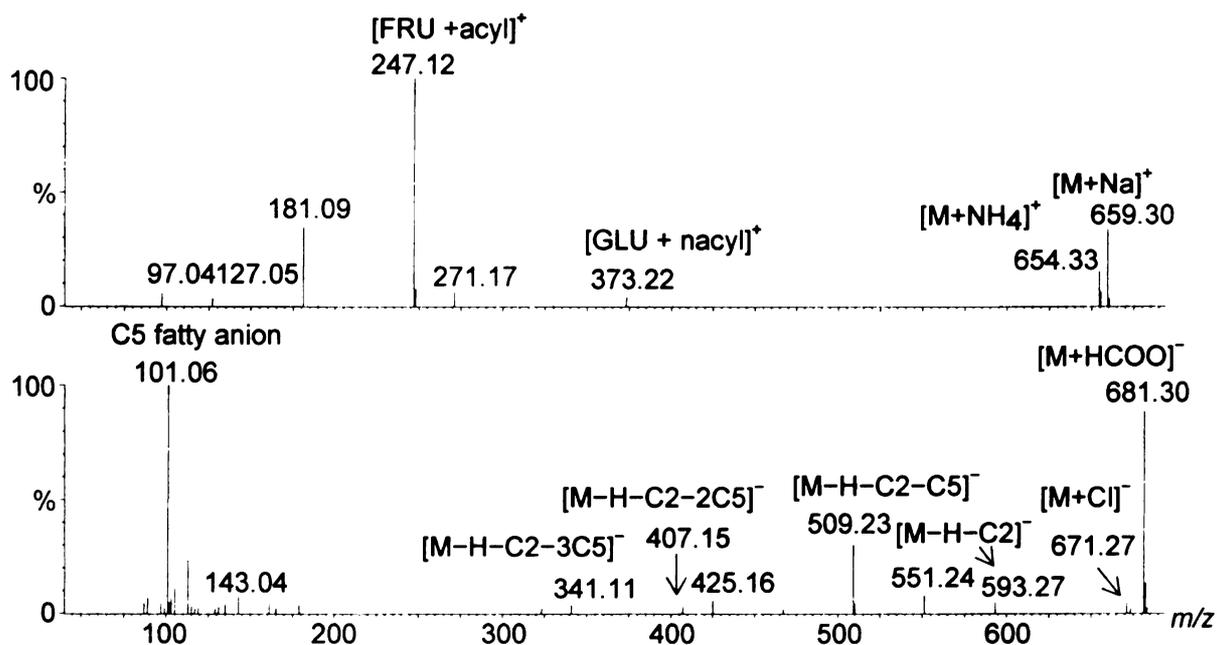


Figure 6-3 Electrospray ionization mass spectra of S 4:17 (RT = 3.11 min) extracted from *S. lycopersicum* M82 using CID potential of 55 V (top) in negative and 40 V (bottom) in positive mode.

Table 6-2. Acyl sucrose metabolites detected in trichome extracts from *S. lycopersicum* M82 and their adduct ions and fragment ions observed using ESI negative (CID potential 55 V) and positive mode (40 V).

Acylsugar metabolite annotation	Fatty acid constituents	[M+NH ₄] ⁺ (m/z)	[M+Na] ⁺ (m/z)	Fragments (m/z)
S 3:15	C5, C5, C5	612	617	247,331
S 4:17	C2, C5, C5, C5	654	659	247,373
S 3:20	C5, C5, C10	682	687	247,401
S 4:22	C2,C5, C5, C10	724	729	247,443
S 3:22	C5, C5, C12	710	715	247,429
S 4:24	C2, C5, C5,C12	752	757	247,471

Acylsugar metabolite annotation	Fatty acid ester groups	[M+formate] ⁻ (m/z)	[M+Cl] ⁻ (m/z)	Fragment ions (m/z)
S 3:15	C5, C5, C5	639	629	593,509,425,407,341,323,101
S 4:17	C2, C5, C5, C5	681	671	635,593,509,425,407,341,305,101
S 3:20	C5, C5, C10	709	699	663,579,425,407,341,323,171,101
S 4:22	C2,C5, C5, C10	751	741	705, 621,579,425,407,341,323,171,101
S 3:22	C5, C5, C12	737	727	691,607,509,425,407,341,323,199,101
S 4:24	C2,C5,C5,C12	779	769	733,691,607,425,407,341,323,199,101

6.3.2 Compare total amounts of acylsugars and distributed amounts of acylsugars from *S. pennellii* LA0716, LA1522, *S. habrochaites* LA1777 and *S. lycopersicum* M82 using LC/MS

Compared to cultivated *S. lycopersicum* M82, many wild type *Solanum* species accumulate large amounts of acylsugars, which is the reason why plant breeders cross the cultivated tomato with wild species to produce lines of cultivated tomato that have inherited the genetic capacity to accumulate substantial amounts of acylsugars. This study

has aimed to make quantitative comparisons of acylsugar levels in tomato and several lines of wild tomato previously known to accumulate acylsugars. To this end, analytical responses for glucose triesters, sucrose triesters and sucrose tetraesters were quantified for *S. pennellii* LA0716, L1522, *S. habrochaites* LA1777 and *S. lycopersicum* M82 by adding all the LC/MS extracted ion chromatogram peak areas of formate adducts for detected acylsugars normalized to the internal standard and dry leaf weight (Figure 6-4). Unfortunately, no authentic standards are available for determination of instrument response factors for the various acylsugar metabolites.

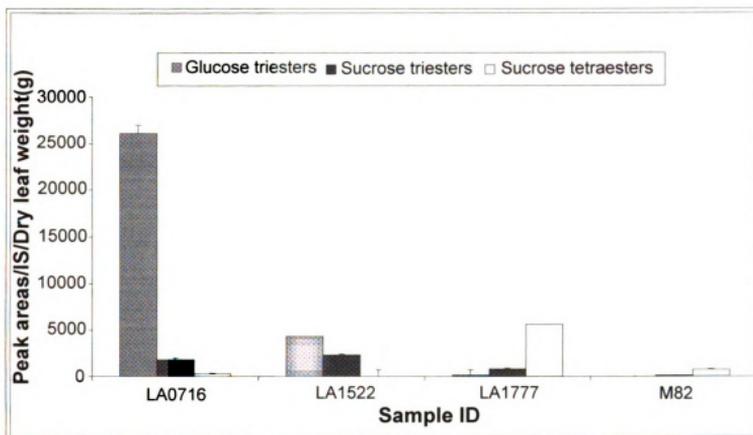


Figure 6-4 Distributed amounts of glucose triesters, sucrose triesters, and sucrose tetraesters of *S. pennellii* LA0716, LA1522, *S. habrochaites* LA1777 and *S. lycopersicum* M82 based on sum of total peak areas of formate adducts for detected acylsugars normalized to internal standard and dry leaf weight for each acylsugar. N=3, SE. Plants were grow in MSU green house.

Based on the LC/MS results, the *S. pennellii* accession LA0716 produced the greatest total amount of acylsugars, accumulating about 5-fold more than LA1522 and LA1777, and about 36-fold more than tomato (M82). Glucose triesters predominated in both accessions of *S. pennellii*, whereas sucrose esters were the dominant form in *S. habrochaites* and *S. lycopersicum*. Glucose triester levels from *S. pennellii* LA0716 were about 6-fold greater than from LA1522 (same species). Levels of glucose triesters from *S. habrochaites* LA1777 were much lower, around 0.6% of the LA0716 levels. Tomato

(M82) did not yield detectable glucose triesters. The two *S. pennellii* lines (LA0716 and 1522) produced substantial amounts of sucrose triesters, but virtually no sucrose tetraesters. The reverse was observed for *S. habrochaites* LA1777 and *S. lycopersicum* M82, where sucrose tetraesters were the dominant forms.

6.3.3 Fatty acyl substituents of acylsugars among wild type *S. pennellii* LA0716, LA 1522, *S. habrochaites* LA1777 and *S. lycopersicum* M82 using GC/MS

In addition to differences among the four lines in amounts and ratios of acylsugars, the distribution of fatty acyl substituents for acylsugars was also different, as established by performing GC/MS analyses of fatty acid ethyl esters after transesterification. Acylsugars from *S. pennellii* LA0716 contained the following branched chain fatty acids: 2-methylpropanoic acid (iC4), 2-methylbutanoic acid and 3-methylbutanoic acid (aiC5 and iC5), and 8-methylnonanoic acid (iC10), as well as some straight-chain fatty acids (nC10 and nC12). These findings were consistent with a literature report that employed GC/MS.^{110,164,169} For some species LA1522, the ratio of aiC5 to iC5 was about 2:1, and was similar to the ratios observed for *S. habrochaites* LA1777 and *S. lycopersicum* M82. However, in the close relative LA0716 gave the inverse ratio (about 1:2 aiC5:iC5).

Quantitative differences in the amounts of the individual fatty acid ester groups were also determined using GC/MS (Figure 6-5). The iso-C4 fatty acyl group was dominant in LA1777 and LA0716 (about 35% of total chains), but C4 was much less abundant (about 2% of total) in LA1522 and not detectable in extracts from *S. lycopersicum* M82 used in this study. However, in separate analyses of M82 extracts from different sources obtained at different times, low levels of iC4 fatty acid groups were detected.

Accumulation of large amounts of metabolites containing short or medium-length branched chain fatty acid esters is not common in plants.¹⁶⁹ Walters and Steffens¹¹² proposed the short branched chain fatty acids, such as 2-methylpropanoic acid (iC4), 2-methylbutanoic acid and 3-methylbutanoic acid (aiC5 and iC5), arise from branched chain amino acids such as valine, leucine and isoleucine. However, the mechanisms of fatty acid elongation and the genetic control of carbon flux from amino acid to fatty acid pathways in these systems are still uncertain.²⁵⁹ Combining information about identities and levels of fatty acyl esters across numerous species with information about gene expression may provide critical knowledge that will guide metabolic engineering of plants and microbes capable of biosynthesis of these and other metabolites.

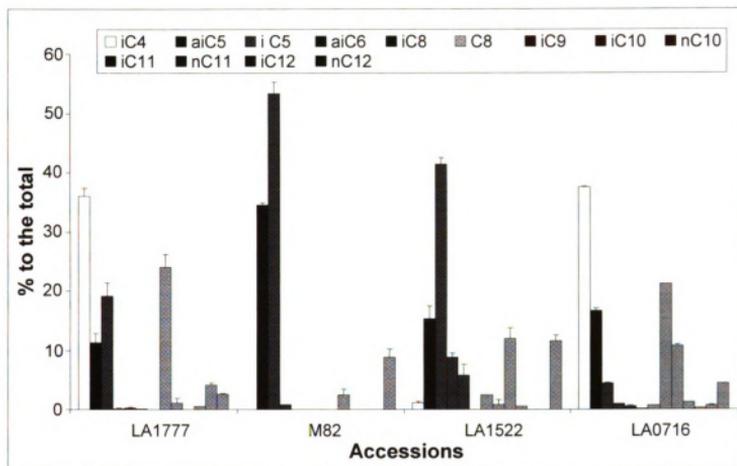


Figure 6-5 Distribution of fatty acyl groups for total acylsugars among *S. pennellii* 0716, 1522, *S. habrochaites* 1777 and *S. lycopersicum* M82 obtained from GC/MS. N=3 separate plants, error bars represent standard error of the mean.

6.3.4 Compare total amounts of acylsugars and distributed amounts of glucose triesters, sucrose triesters and sucrose tetraesters, and fatty acyl substituents of acylsugars among Cornell breeding lines using LC/MS.

To determine how well a large scale metabolite profiling effort could guide breeding of tomato lines for acylsugar production, extracts from leaflets from 90 different plants samples were detected including the parent lines *S. lycopersicum* M82, *S. pennellii* LA0716, introgression lines derived from crosses of LA0716 x M82, Second- to fifth generation progeny from crosses of introgression lines with *S. pennellii* LA0716, and

crosses between a separate high acylsugar producing line (termed “97 fixed-lines”, usually derived from several introgressions of LA0716 genes into the tomato parent) with introgression lines, and several new lines developed at Cornell. The description of these breeding lines is shown in Appendix Table A-1. Crosslines were detected for acylsugar accumulation and subjected to QTL (Quantitative Trait Locus) analysis to determine the genomic regions associated with the accumulation of acylglucoses, acylsucroses, and total acylsugars. Total amounts of acylsugars and distribution of acylglucoses, acylsucroses were surveyed to determine variance among those breeding lines. This information is then used to identify breeding lines that carry desirable metabolic traits. Several quantitative measures were used in this study: (1) total amounts of acylsugars determined by LC/MS, (2) the ratio of acylglucoses to acylsucroses (LC/MS), (3) relative amounts of aiC5, iC5, and iC4 fatty acid groups (determined by GC/MS). These measurements help plant breeders identify desirable traits in these breeding lines, and select lines for field testing of insect resistance.

Figure 6-6 displays the total amounts of acylsugars for crosslines and fixed acylsugar lines. Distributions of glucose triesters, sucrose triesters and sucrose tetraesters are shown in Figure 6-7. The parent line *S. pennellii* LA0716 accumulated ~3-folds more total acylsugars than the Cornell fixed line 071402, which ranked second most abundance acylsugars. Fixed lines and crosslines contained at least 10-fold more total acylsugars than the parent tomato M82 and assorted introgression lines. All fixed lines produced similar relative amounts of individual of sucrose triesters and tetraesters, but differed in total amounts of acylsucroses. Therefore any difference in the susceptibility of the lines to insect damage was attributed to variance in the level of total acylsugars but not in

relative levels of different acylsugars. These findings led to recommendations that breeding efforts should focus on enhancing total acylsugar yields rather than engineering a specific acylsugar composition.

Distribution of acylsugars was different among introgression lines, which are tomato lines that contain small amounts of genomic DNA from an acylsugar-producing wild parent. In comparing parent tomato M82 with introgression lines, sucrose tetraesters were ~10-folds of sucrose triesters except for IL1-3 and 1-4, for which the triesters were dominant as described in chapter 5 (Figure 5-6). Several introgression lines were selected based upon changes in either acylsugar amounts or profiles, and were crossed with the high acylsugar producer *S. pennellii* LA0716 and with Cornell fixed lines. The purpose of this experiment was to determine whether introducing genes that shift fatty acid composition would also influence total acylsugar yields. Results from LC/MS analyses were generated for crosses between introgression lines and *S. pennellii* (8-1 × 0716, 8-2 × 0716, 3-2 × 0716 and 7-5 × 0716). Three of these genotypes (8-2 × 0716, 3-2 × 0716 and 7-5 × 0716) produced similar levels of acylglucoses and acylsucroses, but for one genotype (8-1 × 0716) acylglucoses were only about 30% of acylsucrose levels.,

In a related set of experiments, the Cornell acylsugar fixed line (97FL) was crossed with tomato M82 and several introgression lines. Three genotypes (M82 × 97FL, 3-2 × 97FL, 7-4 × 97FL), did not accumulate detectable levels of acylglucoses, as expected because none of the parent lines accumulated acylglucoses.

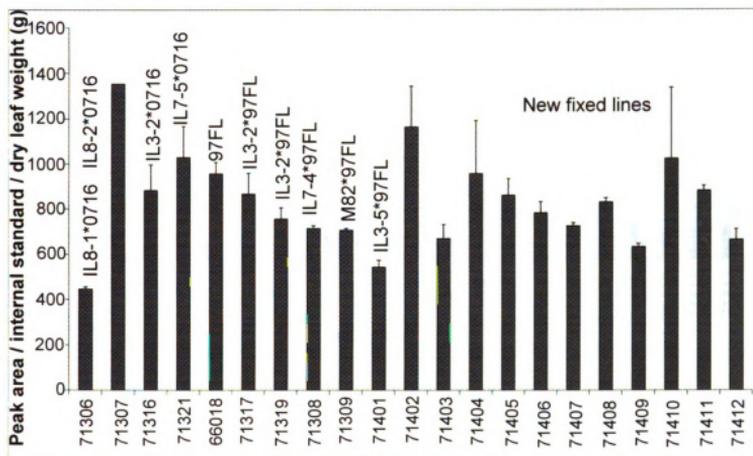


Figure 6-6 Total amounts of acylsugars in Cornell tomato breeding lines. Reported levels represent the sum of XIC peak areas of acylsugar formate adducts normalized to internal standard and dry leaf weight. N=2 or 3 for those lines except N=1 for IL8-2 × 0716. Error bars represent standard error of the mean

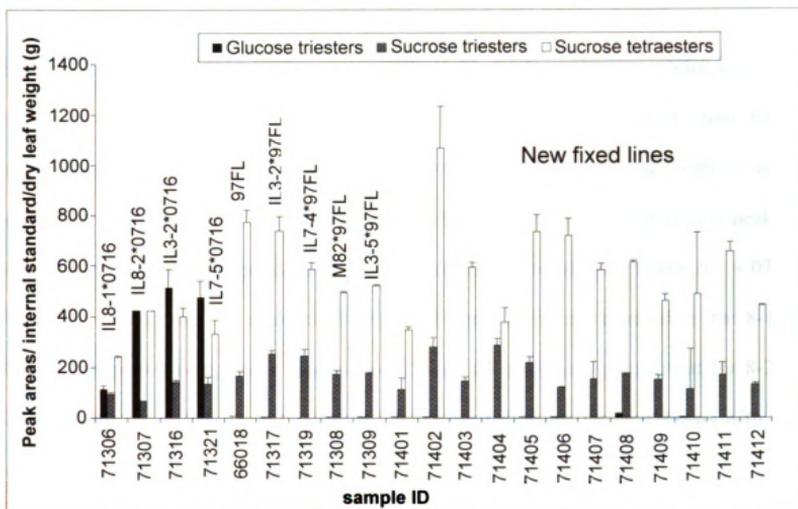


Figure 6-7. Distribution of glucose triesters, sucrose triesters and sucrose tetraesters among all the breeding lines. N=2 or 3 for those lines except N=1 for IL8-2 × 0716. Error bars represent standard error of the mean.

In addition to variations in total amounts of acylsugars and in distributions of acylglucoses versus acylsucroses among those breeding lines, another feature of great interest was a shift in fatty acyl substitutions from C5 to C4 fatty acids observed for 8-1 × 0716 compared to other crosslines and fixed lines. Such a shift in metabolite abundances is not readily apparent from LC/MS data generated using gentle conditions, but the presence of specific fatty acid groups is apparent in high CID energy spectra where fatty acid anions are abundant. Screening for this shift in fatty acyl composition can be assessed by generating XICs for specific fatty acid anions (e.g. *m/z* 87 and 101 for C4 and C5 fatty acids respectively) using high CID potentials.

Figure 6-8 shows the XICs of C4 and C5 fatty acid anion fragments in 8-1 × 0716 and 8-2 × 0716 using 55 V CID potential. Both lines accumulated acylglucoses and acylsucroses. Several differences in acylsugar composition are evident from these chromatograms. First, the acylsugars of line 8-1×0716 display more peaks with C4 fatty acid fragments than 8-2×0716 (approximately 18 notable peaks compared to 4 peaks), suggesting a more diverse population of C4-containing acylsugars in crossline 8-1 × 0716. In addition, integration of chromatographic peak areas led to the conclusion that 8-1 × 0716 accumulated comparable amounts of C5 and C4 fatty acyl groups, but for 8-2 × 0716, acylsugars with C5 groups were ~ 5-folds more abundant than acylsugars with C4 chains. The latter distribution was similar to results from other crosses between introgression lines and wild type *S. pennellii* LA0716.

Acylsugar metabolites were identified with the aid of multiplexed CID for crosses between introgression lines and wild type *S. pennellii* LA0716 (Table 6-3 and 6-4). For in 8-1 × 0716, most sucrose triesters and tetraesters have at least one C4 fatty acyl group, but almost no C4 on acylsucroses was observed for other cross lines. This result indicated that genes related to C4 fatty acid substitution are located at the top of chromosome 8, which is in agreement with GC data from Martha Mutschler's group.¹⁶⁹ The remarkable diversity of acylsugar metabolites from this line presents substantial challenges for metabolite isolation for more detailed structure analysis.

The LC/MS results were also employed to generate quantitative information about levels of individual acylsugars in the lines described above. These profiles were derived by manually constructing a table of retention times and *m/z* values for formate adduct ions, and using QuanLynx software to integrate all observed peaks corresponding to

acylsugar formate adducts. The results demonstrated that line 8-1 × 0716 had lower accumulation of glucose triesters and lower total acylsugars compared to lines 8-2 × 0716, 3-2 × 0716 and 7-5 × 0716. Compared to all other lines of tomato, wild relatives, and genetic crosses, the suite of acylsugars accumulated by the cross of IL 8-1 with *S. pennellii* LA0716 accumulates about twice as many distinguishable acylsugars than any other plants investigated in this research. This investigation therefore documents two important findings. First, the LC/multiplexed CID TOF MS approach generates sufficient information to describe acylsugar type (tri- vs. tetra-esters, glucose vs. sucrose esters) to be useful for phenotypic screens. Second, this tool has demonstrated its utility in development of “gain-of-function” lines that produce metabolites not synthesized by either parent line. The combination of these approaches lends support to metabolic engineering of plants for desired biochemical traits.

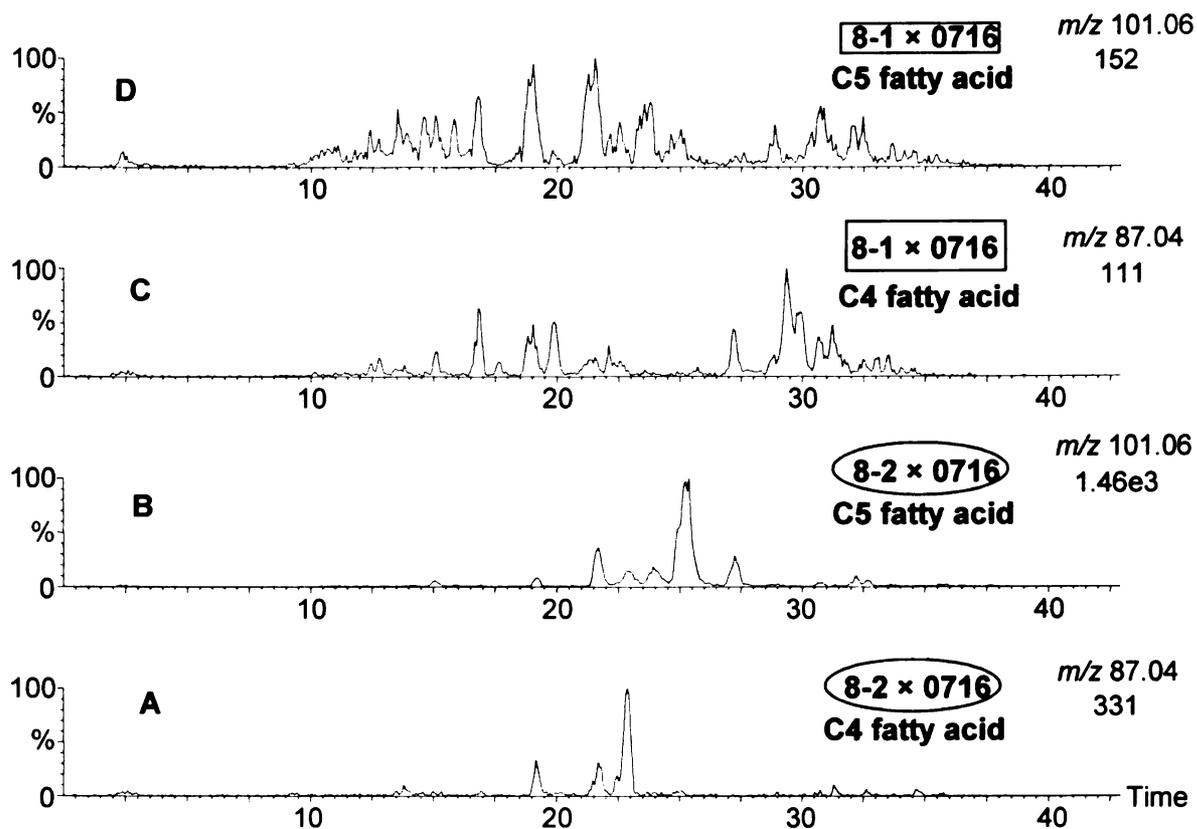


Figure 6-8. ESI negative XICs of C4 m/z 87.04 and C5 m/z 101.06 fatty acyl groups using CID potential 55 V for crosses 8-2 \times 0716 (A and B) and 8-1 \times 0716 (C and D). The number labeled below mass to charge ratio on the right of chromatogram indicate the base peak intensity,

Table 6-3. Fatty acyl constituents for detected acylsucroses in 8-1 × 0716 using ESI negative mode with CID potential at 55 V.

Acylsugar metabolite annotation	Formula of formate adduct	Measured m/z [M+HCOO] ⁻	Fatty acids constituents
S 3:12	C ₂₆ H ₄₃ O ₁₆	597.20	C4,C4,C4
S 3:13	C ₂₆ H ₄₃ O ₁₆	611.23	C4,C4,C5
S 3:14	C ₂₇ H ₄₅ O ₁₆	625.25	C4,C5,C5
S 3:15	C ₂₈ H ₄₇ O ₁₆	639.28	C5,C5,C5
S 3:17	C ₃₀ H ₅₁ O ₁₆	667.31	C4,C4,C9
S 3:18	C ₃₁ H ₅₃ O ₁₆	681.33	C4,C4,C10 C4,C5,C9
S 3:19	C ₃₂ H ₅₅ O ₁₆	695.35	C4,C5,C10 C4,C4,C11
S 3:20	C ₃₃ H ₅₅ O ₁₆	709.36	C4,C4,C12 C4,C5,C11 C5,C5,C10
S 3:21	C ₃₄ H ₅₇ O ₁₆	723.38	C4,C5,C12
S 3:22	C ₃₅ H ₅₉ O ₁₆	737.39	C5,C5,C12

Acylsugar metabolite annotation	Formula of formate adduct	Measured m/z [M+HCOO] ⁻	Fatty acids constituents
S 4:14	C ₂₇ H ₄₃ O ₁₇	639.25	C2,C4,C4,C4
S 4:15	C ₂₈ H ₄₅ O ₁₇	653.26	C2,C4,C4,C5
S 4:16	C ₂₉ H ₄₇ O ₁₇	667.28	C2,C4,C5,C5
S 4:17	C ₃₀ H ₄₉ O ₁₇	681.29	C2,C5,C5,C5
S 4:19	C ₃₂ H ₅₃ O ₁₇	709.32	C2,C4,C4,C9
S 4:20	C ₃₃ H ₅₅ O ₁₇	723.34	C2,C4,C4,C10
S 4:21	C ₃₄ H ₅₇ O ₁₇	737.36	C2,C4,C5,C10 C2,C4,C4,C11
S 4:22	C ₃₅ H ₅₉ O ₁₇	751.38	C2,C4,C4,C12 C2,C4,C5,C11 C2,C5,C5,C10
S 4:23	C ₃₆ H ₆₁ O ₁₇	765.39	C2,C4,C5,C12 C2,C5,C5,C11
S 4:24	C ₃₇ H ₆₃ O ₁₇	779.41	C2,C5,C5,C12

Table 6-4. Fatty acyl constituents for detected acylsucroses in 8-2 × 0716, 3-2 × 0716 and 7-5 × 0716 using ESI negative mode with CID potential at 55 V.

Acylsugar metabolite annotation.	Formula of formate adduct	Measures m/z [M+HCOO] ⁻	Fatty acids constituents
S 3:15	C ₂₈ H ₄₇ O ₁₆	639.28	C5,C5,C5
S 3:19	C ₃₂ H ₅₅ O ₁₆	695.35	C5,C5,C9
S 3:20	C ₃₃ H ₅₅ O ₁₆	709.36	C5,C5,C10 C4,C5,C11
S 3:21	C ₃₄ H ₅₇ O ₁₆	723.38	C5,C5,C11
S 3:22	C ₃₅ H ₅₉ O ₁₆	737.39	C5,C5,C12
S 3:23	C ₃₆ H ₆₁ O ₁₆	751.41	C5,C6,C12

Acylsugar metabolite annotation	Formula of formate adduct	Measures m/z [M+HCOO] ⁻	Fatty acids constituents
S 4:16	C ₂₉ H ₄₇ O ₁₇	667.28	C2,C4,C5,C5
S 4:17	C ₃₀ H ₄₉ O ₁₇	681.29	C2,C5,C5,C5
S 4:21	C ₃₄ H ₅₇ O ₁₇	737.36	C2,C5,C5,C9
S 4:22	C ₃₅ H ₅₉ O ₁₇	751.38	C2,C5,C5,C10
S 4:23	C ₃₆ H ₆₁ O ₁₇	765.39	C2,C5,C5,C11
S 4:24	C ₃₇ H ₆₃ O ₁₇	779.41	C2,C5,C5,C12
S 4:25	C ₃₈ H ₆₅ O ₁₇	793.43	C2,C5,C6,C12

Although the metabolite profiling results for 90 breeding lines (described above) identified lines that accumulate high levels of acylsugars and display significant reduction of insect infestation in field trials, these lines exhibit undesirable traits from the parent *S. pennellii* LA0716 including small fruit, and poor germination and growth. In order to eliminate those negative traits, the Cornell group has combined data from these acylsugar profiles with DNA marker information to select plants with high acylsugar production, but reduced contributions of *S. pennellii* DNA in regions not associated with acylsugar production. These plants have undergone further field trials, and new

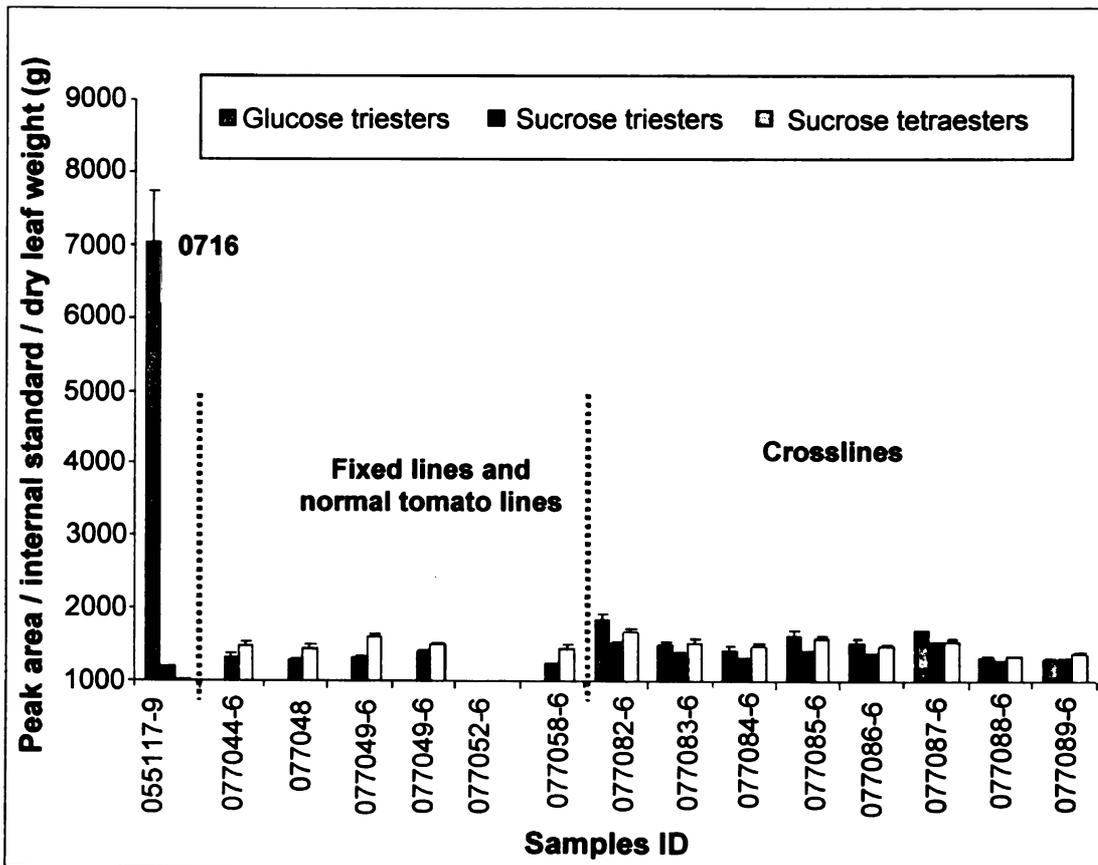
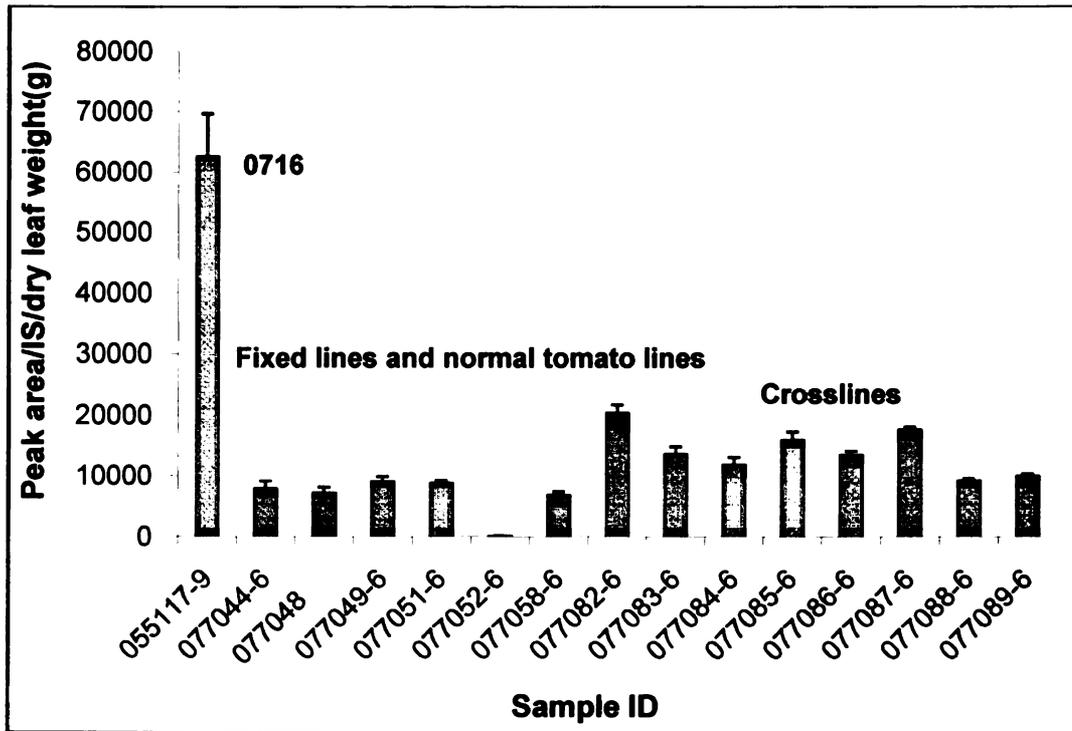
commercial tomato lines with high trichome acylsugar accumulation may be introduced to the commercial market within the next few years.

Continued efforts to improve acylsugar accumulation in tomato trichomes has continued with investigation of 13 additional lines not described above. For this study, LC/MS metabolite profiles were generated for a total of 60 plants including *S. pennellii*, normal tomato lines, five fixed-lines and eight crosses between fixed lines and *S. pennellii* LA0716. The description of these breeding lines was displayed in Appendix Table A-2.

Annotation of metabolites as acylsugars was performed for the set of 60 plants mentioned above, using LC/TOF MS and multiple CID conditions. From this information, a list of target acylsugar metabolites was constructed, and XIC for the formate adduct of each was integrated using QuanLynx software. For each extract, the sum of all integrated peak areas was calculated (Figure 6-9 Top). *S. pennellii* LA0716 accumulates 3 to 7-fold more total acylsugars compared to all fixed lines, and the conventional tomato line accumulated minimal amounts relative to all others. Acylsugar totals in crosses between fixed lines and *S. pennellii* LA0716 roughly doubled compared to fixed-lines. These findings were expected, since inter-specific crosses between a species with high levels of any of these compounds and a species with low levels usually results in plants with intermediate levels of the compounds^{259,260}. Substantial diversity in acylsugar accumulation was observed across this set of plants. For *S. pennellii* LA0716, ~90% of acylsugar signal was attributed to triacylglucoses. There were no detectable acylglucoses in fixed lines and normal tomato lines. Glucose triesters account for 30-50 % of the total

acylsugars (Figure 6-9 bottom) among crosses, which suggests a biochemical inheritance from their parent *S. pennellii*.

Figure 6-9 Amounts of acylsugars in *S. pennellii*, normal tomato lines, fixed-lines and crosslines based on sum of peak areas, then normalized to internal standard and dry leaf weight (Top). Composition is broken down into glucose triesters, sucrose triesters and sucrose tetraesters were among all the breeding lines (Bottom). N=4. Error bars represent standard error of the mean.



One of the more interesting aspects about these breeding lines was the increased diversity of acylsugars in plants derived by crossing fixed lines with wild type *S. pennellii* LA0716. Figure 6-10 compares LA0716, fixed line 071026 and crosses 071026 × LA0716. Many acylsugars that were not present in either parent appeared after crossing. Fatty acyl substituents C11 and C9 were prominent in the cross, but had low abundance in both parents. (Figure 6-11). This new variation in metabolite composition may derive from two sources: (1) changed abundances of specific fatty acid precursors, and (2) additional variation in positions of fatty acid substitution. One may ask how new substitution patterns arise after crossing. One hypothesis is that the cross leads to expression of enzymes in the cross that was not expressed in either parent. Future investigations will be needed to test this hypothesis.

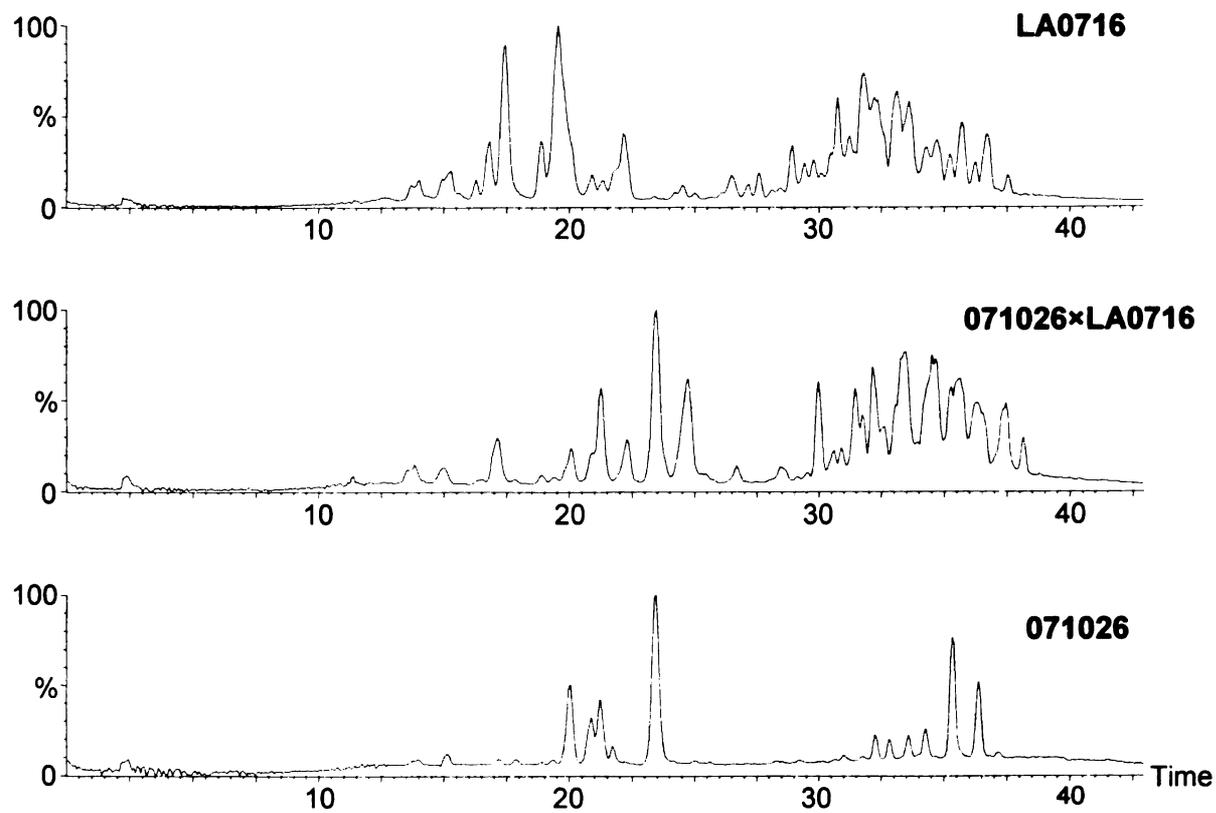


Figure 6-10. TIC from wild type *S. pennellii* LA0716 (top), cross of 071026 × LA0716 (middle) and 071026 (bottom) using ESI negative mode with CID potential at 10 V.

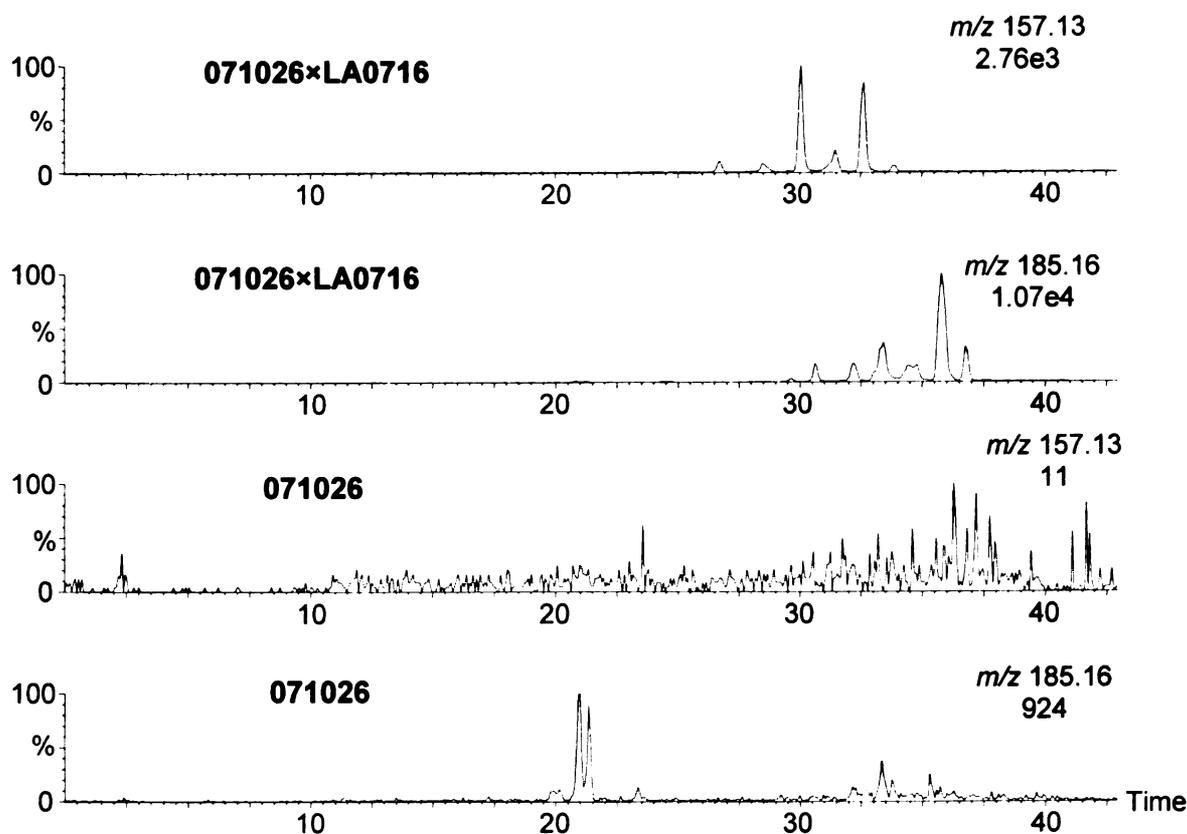


Figure 6-11 ESI negative XICs of C9 m/z 157.13 and C10 m/z 185.16 fatty acyl groups using CID potential 55 V for fixed line 071026 and crosses between 071026 and *S. pennellii* LA0716. The numbers below m/z on the right top of chromatograms indicate the base ion peak intensity.

The results from LC/MS were supported by using GC/MS for fatty acyl group analysis. Fatty acyl groups among fixed lines, crosses between fixed lines and *S. pennellii* LA0716 were analyzed using GC/MS after transesterification in Figure 6-12. For fixed lines, they contain predominantly: 2-methylbutanoic acid and 3-methylbutanoic acid (aiC5 and iC5), and straight-chain fatty acid nC12, the same as their parent normal tomato line. But in plants generated by crossing fixed lines with *S. pennellii* LA0716, the distribution of fatty acids changed. iC4, aiC6, iC9 and iC11 fatty acids appear that were

barely detectable in the fixed lines. For example, the branched chain iC11 fatty acyl group became the second most abundant fatty acyl group (about 40% to the total fatty acyl) after iC5 among those crosslines, but iC11 was not only 1% of fatty acyl groups in LA0716 and was not detected in the fixed lines.

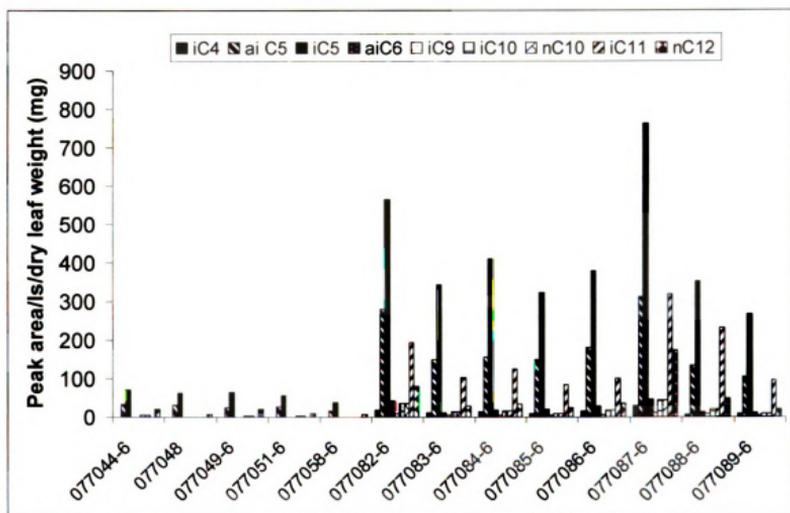


Figure 6-12 Distribution of amounts of individual fatty acyl groups among fixed-lines, crosslines and *S. pennellii* LA0716 based on GC/MS profiles. 077044-6, 077048, 077049-6, 077051-6, 077058-6 are fixed lines; 077082-6 to 077089-6 are crosses from fixed lines and *S. pennellii* LA0716; N=4, Error bars represent standard error of the mean.

Annotation of acylsugars were based on LC/TOF MS data for retention time and accurate masses of formate adducts and fragment ions, following the same approach described in Chapter 3. Fatty acyl chains for acylsugars were determined based on fatty acid anions in negative mode. Fatty acid compositions for observed glucose triesters for

fixed lines and crosslines between fixed lines and *S. pennellii* LA0716 are displayed in the Tables.6-5 and 6-6. The suite of acylsugars accumulated by the cross of fixed lines with *S. pennellii* LA0716 accumulates about three folds as many distinguishable acylsugars (about 22) than fixed lines themselves (about 7) investigated in this research.

Table 6-5. Fatty acid constituents of detected acylsucroses from fixed lines using ESI negative (CID potential, 55 V).

Acylsugar metabolite annotation	Formula of formate adduct	Measures m/z $[M+HCOO]^-$	Fatty acids constituents
S 3:15	C ₂₈ H ₄₇ O ₁₆	639.28	C5,C5,C5
S 3:20	C ₃₃ H ₅₅ O ₁₆	709.36	C5,C5,C10
S 3:21	C ₃₄ H ₅₇ O ₁₆	723.37	C4,C5,C12
S 3:22	C ₃₅ H ₅₉ O ₁₆	737.39	C5,C5,C12

Acylsugar metabolite annotation	Formula of formate adduct	Measures m/z $[M+HCOO]^-$	Fatty acids constituents
S 4:17	C ₃₀ H ₄₉ O ₁₇	681.29	C2,C5,C5,C5
S 4:22	C ₃₅ H ₅₉ O ₁₇	751.38	C2,C5,C5,C10
S 4:24	C ₃₇ H ₆₃ O ₁₇	779.41	C2,C5,C5,C12

Table 6-6. Fatty acid constituents of detected acylsugars from crosses between fixed lines and *S. pennellii* LA0716 using ESI negative (CID potential, 55 V).

Acylsugar metabolite annotation	Formula of formate adduct	Measures m/z [M+HCOO] ⁻	Fatty acids constituents
G 3:14	C ₂₀ H ₃₃ O ₁₁	463.21	C4,C5,C5
G 3:15	C ₂₁ H ₃₅ O ₁₁	477.23	C5,C5,C5
G 3:16	C ₂₂ H ₃₇ O ₁₁	491.24	C5,C5,C6
G 3:18	C ₂₃ H ₃₉ O ₁₁	519.28	C4,C5,C9
G 3:19	C ₂₅ H ₄₃ O ₁₁	533.29	C5,C5,C9
G 3:20	C ₂₆ H ₄₅ O ₁₁	547.31	C5,C5,C10 C4,C5,C11
G 3:21	C ₂₇ H ₄₇ O ₁₁	561.32	C5,C5,C11
G 3:22	C ₂₈ H ₄₉ O ₁₁	575.34	C5,C5,C12
G 3:23	C ₂₉ H ₅₁ O ₁₁	589.36	C5,C6,C12

Acylsugar metabolite annotation	Formula of formate adduct	Measures m/z [M+HCOO] ⁻	Fatty acids constituents
S 3:15	C ₂₈ H ₄₇ O ₁₆	639.28	C5,C5,C5
S 3:19	C ₃₂ H ₅₅ O ₁₆	695.35	C5,C5,C9
S 3:20	C ₃₃ H ₅₅ O ₁₆	709.36	C5,C5,C10 C4,C5,C11
S 3:21	C ₃₄ H ₅₇ O ₁₆	723.38	C5,C5,C11
S 3:22	C ₃₅ H ₅₉ O ₁₆	737.39	C5,C5,C12
S 3:23	C ₃₆ H ₆₁ O ₁₆	751.41	C5,C6,C12

Acylsugar metabolite annotation	Formula of formate adduct	Measures m/z [M+HCOO] ⁻	Fatty acids constituents
S 4:16	C ₂₉ H ₄₇ O ₁₇	667.28	C2,C4,C5,C5
S 4:17	C ₃₀ H ₄₉ O ₁₇	681.29	C2,C5,C5,C5
S 4:21	C ₃₄ H ₅₇ O ₁₇	737.36	C2,C5,C5,C9
S 4:22	C ₃₅ H ₅₉ O ₁₇	751.38	C2,C5,C5,C10
S 4:23	C ₃₆ H ₆₁ O ₁₇	765.39	C2,C5,C5,C11
S 4:24	C ₃₇ H ₆₃ O ₁₇	779.41	C2,C5,C5,C12
S 4:25	C ₃₈ H ₆₅ O ₁₇	793.43	C2,C5,C6,C12

6.4 Conclusion

This chapter has presented results from profiling of acylsugar metabolites in (about 100) different genotypes of tomato breeding lines. Several aspects of this research represent original scientific contributions that demonstrate how nontargeted metabolomic profiling can guide plant breeding efforts to develop desirable traits.

1. Contrary to prior belief, conventional tomato (line M82) does accumulate acylsugars, but at low levels compared to other lines. The LC/MS approach developed for this study proved capable of detecting acylsugar production in **all** lines studied. This advance allows investigation of factors that regulate acylsugar production, as distinct from measuring the presence or absence of acylsugars. Recognition that all tomato lines produce measureable quantities of acylsugars offers the prospect for a paradigm shift in understanding factors that regulate acylsugar accumulation.
2. Before this work, acylsugar analyses were almost entirely performed by performing base hydrolysis of organic plant extracts, followed by colorimetric measurement of glucose. Measurements of sucrose esters required enzymatic conversion to glucose (using invertase), that could also be detected by colorimetry. The LC/MS method developed in this work has led to more than 100-fold lowering of detection limits, and has allowed for comprehensive profiling of 50 or more different acylsugars in a single plant extract.
3. Tomato breeding lines derived by crossing *S. pennellii* LA0716 with other fixed lines of accumulated acylsugar mixtures with greater diversity than combinations

of the two parents might suggest. These new crosslines demonstrate that a bottleneck to metabolic engineering has been overcome through traditional breeding, without need for recombinant DNA techniques. The LC/MS methodology developed here offers a powerful tool for screening plants for biochemical composition, and should aid future plant breeding efforts.

Chapter 7. Concluding remarks

For most of the modern era, discoveries of genes have been driven by a reductionist approach in which organisms were reduced to individual genes that were studied one-at-a-time and in isolation. Although these efforts generated a great deal of knowledge, complex networks control most biological processes, and these are not amenable to yield a comprehensive understanding when they are studied in isolated model systems. For several reasons, engineering of desirable secondary metabolite levels in plants has found limited success. Fortunately, new high-throughput technologies that can link gene expression levels with comprehensive chemical profiles offer the prospect for overcoming many technical obstacles to metabolic engineering of plants and other organisms. These technologies have brought about an era of systems biology, which is a more holistic way to view biological functions. To move such efforts forward, this thesis has presented development and application of a new metabolomic methodology using LC/MS for discovery and profiling of specialized metabolites in glandular trichomes of numerous plants in the genus *Solanum*. One central goal of my efforts has been to push the limits of a relatively low-cost mass spectrometer to provide rich information regarding metabolite identities and quantities that could be accessible to laboratories without the resources to purchase expensive and more powerful mass spectrometers.

Technologies for metabolite profiling have lagged behind other genomic and proteomic methods owing to the chemical diversity of metabolites. Improved capabilities in analytical throughput for both metabolite identification and quantification are needed to generate comprehensive metabolite databases that will be useful for gene discoveries. To this end, Chapter 2 describes development of a new LC/TOF MS approach for metabolite profiling that uses nonselective and multiplexed CID. This analytical

methodology enables deep coverage of the metabolome using a consistent protocol for identification and quantification, yielding accurate measurements of molecular and fragment masses and ion behavior in a collision energy-resolved context. The combination of this information offers the prospect for annotating ions as molecular, oligomer, or fragment ions and for improved recognition of chemical classes of metabolites based on mass and ease of fragmentation, even when multiple metabolites coelute. These improvements in analytical throughput allow faster metabolic profiling analyses to be performed to identify major and minor metabolites in high complexity biological extracts. In Chapters 3 and 4, the LC/multiplexed CID MS approach was applied for nontargeted profiling of metabolites extracted from trichomes of genus *Solanum*. These analyses led to detection and annotation of many secondary metabolites, covering acylsugars, flavonoids, alkaloids and terpenoids, many of which were previously unknown in specific species. Experimental determination of metabolic diversity provides researchers with an opening to discover how genes and environment combine to produce different biochemical outcomes, and rapid screening for differences in plant chemistry can guide efforts toward gene discovery.

Chapter 5 takes multiplexed CID into the realm of high-throughput screening, and describes a new fast LC/MS approach used to screen trichome metabolic phenotypes for thousands of transgenic plants. The additional dimensionality of the MS data (CID energy) aids metabolite annotation even when chromatographic profiles overlap, and enables measurements of approximately 100 metabolites per minute of instrument time, and analyses of several thousand samples per year on a single mass spectrometer. Results from screening of tomato introgression lines have identified several chromosome

regions that influence trichome chemical composition and amounts of trichome metabolites and these findings are guiding gene mapping efforts. Finally, Chapter 6 applies LC/TOF MS with GC/MS for qualitative and quantitative profiling of anti-insect acylsugar metabolites for Cornell breeding lines of tomato. These results demonstrated nontargeted metabolite profiling using LC/MS reveals remarkable chemical diversity among plant genotypes, and can guide plant breeding efforts to develop desirable traits in economically important crops.

The fast LC/TOF MS-multiplexed CID method offers great value for deep profiling of metabolomes but requiring minimal instrument time per sample. It allows metabolic phenotype screening for thousands of samples to become practical, and generates qualitative and quantitative analysis results for unidentified metabolites to assist discovery of their biosynthetic pathways. Additional applications of the LC/TOF and multiplexed CID approach, not included in this dissertation, have generated metabolite profiles as parts of two collaborative studies, one involving plant resistance to insect herbivory with the laboratory of Professor Gregg Howe, and the other a survey of other plants within the family Solanaceae with Professor Cornelius Barry's group. In addition, this approach is widely performed by other members of the Jones laboratory for profiling of plant metabolic responses to pathogen stress, responses of cancer cells to chemotherapeutic agents, discovery of novel bioactive oxylipins in plants and algae, and profiling of soluble byproducts of biomass processing for generation of cellulosic ethanol.

Perhaps the most striking conclusion that can be drawn from the thousands of LC/MS analyses performed during these studies is the remarkable chemical diversity observed

within a single plant genus. Even closely related lines within the same species showed distinct metabolite profiles. These are largely attributed to small changes in DNA sequence resulting in substantial changes to plant chemistry. Improved understanding of the biochemical basis for these differences will come from ongoing and future research, and offers hope that the current century will see improved use of biological sources of chemicals.

Appendix

Table A-1 Description of breeding lines from Cornell including introgression lines, fixed lines, crosses between introgression lines and *S. pennellii* LA0716, and new fixed lines.

MI sample #	Samples for GH dips	pedigree
1	66018	97FL
2	71302	M82
3	71304	M82
4	71305	IL 10-1
5	71306	IL8-1 × LA716
6	71307	IL8-2 × LA716
7	71308	M82 × 97FL
8	71309	IL3-5 × 97FL
9	71311	IL 3-5
10	71312	IL 8-1
11	71313	IL 8-1-1
12	71314	IL 8-2
13	71315	IL3-2
14	71316	IL3-2 × LA716
15	71317	IL3-2 × 97FL
16	71318	IL7-4
17	71319	IL7-4 × 97FL
18	71320	IL 7-5
19	71321	IL7-5 × LA716
20	71322	IL-10-1 / control
21	71401	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F2
22	71402	(short 3 × 99FL F2 selection 01-1014-B03)F6
23	71403	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F3
24	71404	(short 3 × 99FL F2 selection 01-1014-B03) F5
25	71405	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F3
26	71406	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F2
27	71407	(short 3 × 99FL F2 selection 01-1014-B03)F5
28	71408	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F3
29	71409	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F3
30	71410	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F3
31	71411	(short 3 × 99FL F2 selection 01-1014-D07)F4
32	71412	[LP cyto × 011014-DO7] × (prog sel from 011014-B03)F3

Table A-2 Description of breeding lines from Cornell including *S. pennellii* LA0716, normal tomato lines, fixed lines and crosses between fixed lines and *S. pennellii* 0716.

MI sample #	samples for GH dips	ID and Types of crosses	pedigree
1	055117-9	LA716	<i>S. pennellii</i> LA716 long internode parent
2	055117-9	LA716	<i>S. pennellii</i> LA716 long internode parent
3	055117-9	LA716	<i>S. pennellii</i> LA716 long internode parent
4	055117-9	LA716	<i>S. pennellii</i> LA716 long internode parent
5	077044-6	066076-7	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071025 (066076-4)
6	077044-6	066076-7	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071025 (066076-4)
7	077044-6	066076-7	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071025 (066076-4)
8	077044-6	066076-7	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071025 (066076-4)
9	077048	071026	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
10	077048	071026	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
11	077048	071026	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
12	077048	071026	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
13	077049-6	66060-5	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
14	077049-6	66060-5	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
15	077049-6	66060-5	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
16	077049-6	66060-5	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4
17	077051-6	066254-7 (also 071024)	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071024
18	077051-6	066254-7 (also 071024)	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071024
19	077051-6	066254-7 (also 071024)	((LPC X 011014-DO7] X (prog sel from 011014-B03)F2)F4 '071024

Table A-2 (continued)

20	077051-6	066254-7 (also 071024	((LPC X 011014-DO7] X (prog sel from 011014- B03)F2)F4 '071024
21	077052-6	066083	NC33EB-1 (PH-2+PH-3)
22	077052-6	066083	NC33EB-1 (PH-2+PH-3)
23	077052-6	066083	NC33EB-1 (PH-2+PH-3)
24	077052-6	066083	NC33EB-1 (PH-2+PH-3)
25	077058-6	066112-8 also 071021	((LPC X 011014-DO7] X (prog sel from 011014- B03)F2)F3 '071021
26	077058-6	066112-8 also 071021	((LPC X 011014-DO7] X (prog sel from 011014- B03)F2)F3 '071021
27	077058-6	066112-8 also 071021	((LPC X 011014-DO7] X (prog sel from 011014- B03)F2)F3 '071021
28	077058-6	066112-8 also 071021	((LPC X 011014-DO7] X (prog sel from 011014- B03)F2)F3 '071021
29	077082-6	071021-3 x 055117-9	071021-3 x LA716 long internode
30	077082-6	071021-3 x 055117-9	071021-3 x LA716 long internode
31	077082-6	071021-3 x 055117-9	071021-3 x LA716 long internode
32	077082-6	071021-3 x 055117-9	071021-3 x LA716 long internode
33	077083-6	071026-3 x 055117-9	071026-3 x LA716 long internode
34	077083-6	071026-3 x 055117-9	071026-3 x LA716 long internode
35	077083-6	071026-3 x 055117-9	071026-3 x LA716 long internode
36	077083-6	071026-3 x 055117-9	071026-3 x LA716 long internode

Table A-2 (continued)

37	077084-6	066060-5 x 055117-9	066060-5 x LA716 long internode
38	077084-6	066060-5 x 055117-9	066060-5 x LA716 long internode
39	077084-6	066060-5 x 055117-9	066060-5 x LA716 long internode
40	077084-6	066060-5 x 055117-9	066060-5 x LA716 long internode
41	077085-6	071024-1 x 055117-9	071024-1 x LA716 long internode
42	077085-6	071024-1 x 055117-9	071024-1 x LA716 long internode
43	077085-6	071024-1 x 055117-9	071024-1 x LA716 long internode
44	077085-6	071024-1 x 055117-9	071024-1 x LA716 long internode
45	077086-6	066076-4 x 055117-9	066076-4 x LA716 long internode
46	077086-6	066076-4 x 055117-9	066076-4 x LA716 long internode
47	077086-6	066076-4 x 055117-9	066076-4 x LA716 long internode
48	077086-6	066076-4 x 055117-9	066076-4 x LA716 long internode
49	077087-6	055113-1 x 055117-9	055113-1 x LA716 long internode
50	077087-6	055113-1 x 055117-9	055113-1 x LA716 long internode
51	077087-6	055113-1 x 055117-9	055113-1 x LA716 long internode

Table A-2 (continued)

52	077087-6	055113-1 x 055117-9	055113-1 x LA716 long internode
53	077088-6	061121-23 x 055117-9	NC33EB-1 x long internode LA716
54	077088-6	061121-23 x 055117-9	NC33EB-1 x long internode LA716
55	077088-6	061121-23 x 055117-9	NC33EB-1 x long internode LA716
56	077088-6	061121-23 x 055117-9	NC33EB-1 x long internode LA716
57	077089-6	061120-23 x 055177-9	LIL x LA716 long
58	077089-6	061120-23 x 055177-9	LIL x LA716 long
59	077089-6	061120-23 x 055177-9	LIL x LA716 long
60	077089-6	061120-23 x 055177-9	LIL x LA716 long

Bibliography

Bibliography

- (1) Pimentel, D.; Pimentel, M. *American Journal of Clinical Nutrition* **2003**, *78*, 660s-663s.
- (2) Wittstock, U.; Gershenzon, J. *Current Opinion in Plant Biology* **2002**, *5*, 300-307.
- (3) Theis, N.; Lerdau, M. *International Journal of Plant Sciences* **2003**, *164*, S93-S102.
- (4) Duke, S. O.; Canel, C.; Rimando, A. M.; Tellez, M. R.; Duke, M. V.; Paul, R. N. *Advances in Botanical Research Incorporating Advances in Plant Pathology, Vol 31 2000* **2000**, *31*, 121-151.
- (5) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nature Biotechnology* **2000**, *18*, 1157-1161.
- (6) Schauer, N.; Fernie, A. R. *Trends in Plant Science* **2006**, *11*, 508-516.
- (7) Xie, Z.; Kapteyn, J.; Gang, D. R. *Plant Journal* **2008**, *54*, 349-361.
- (8) Schilmiller, A. L.; Last, R. L.; Pichersky, E. *Plant Journal* **2008**, *54*, 702-711.
- (9) Fridman, E.; Wang, J. H.; Iijima, Y.; Froehlich, J. E.; Gang, D. R.; Ohlrogge, J.; Pichersky, E. *Plant Cell* **2005**, *17*, 1252-1267.
- (10) Gang, D. R.; Wang, J. H.; Dudareva, N.; Nam, K. H.; Simon, J. E.; Lewinsohn, E.; Pichersky, E. *Plant Physiology* **2001**, *125*, 539-555.
- (11) Iijima, Y.; Gang, D. R.; Fridman, E.; Lewinsohn, E.; Pichersky, E. *Plant Physiology* **2004**, *134*, 370-379.
- (12) Iijima, Y.; Davidovich-Rikanati, R.; Fridman, E.; Gang, D. R.; Bar, E.; Lewinsohn, E.; Pichersky, E. *Plant Physiology* **2004**, *136*, 3724-3736.
- (13) Suzuki, H.; Reddy, M. S. S.; Naoumkina, M.; Aziz, N.; May, G. D.; Huhman, D. V.; Sumner, L. W.; Blount, J. W.; Mendes, P.; Dixon, R. A. *Planta* **2005**, *220*, 696-707.
- (14) Goossens, A.; Hakkinen, S. T.; Laakso, I.; Seppanen-Laakso, T.; Biondi, S.; De Sutter, V.; Lammertyn, F.; Nuutila, A. M.; Soderlund, H.; Zabeau, M.; Inze, D.; Oksman-Caldentey, K. M. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 8595-8600.
- (15) Morikawa, T.; Mizutani, M.; Aoki, N.; Watanabe, B.; Saga, H.; Saito, S.; Oikawa, A.; Suzuki, H.; Sakurai, N.; Shibata, D.; Wadano, A.; Sakata, K.; Ohta, D. *Plant Cell* **2006**, *18*, 1008-1022.

- (16) Martzen, M. R.; McCraith, S. M.; Spinelli, S. L.; Torres, F. M.; Fields, S.; Grayhack, E. J.; Phizicky, E. M. *Science* **1999**, *286*, 1153-1155.
- (17) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nature Biotechnology* **2001**, *19*, 173-173.
- (18) Koch, K. *Current Opinion in Plant Biology* **2004**, *7*, 235-246.
- (19) Mitchell-Olds, T.; Pedersen, D. *Genetics* **1998**, *149*, 739-747.
- (20) Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Current Opinion in Plant Biology* **1999**, *2*, 83-85.
- (21) Trethewey, R. N. *Current Opinion in Biotechnology* **2001**, *12*, 135-138.
- (22) Oliver, D. J.; Nikolau, B.; Wurtele, E. S. *Metabolic Engineering* **2002**, *4*, 98-106.
- (23) Sumner, L. W.; Mendes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817-836.
- (24) Roessner, U.; Luedemann, A.; Brust, D.; Fiehn, O.; Linke, T.; Willmitzer, L.; Fernie, A. R. *Plant Cell* **2001**, *13*, 11-29.
- (25) Harrigan, G. G.; Stork, L. G.; Riordan, S. G.; Reynolds, T. L.; Ridley, W. P.; Masucci, J. D.; MacIsaac, S.; Halls, S. C.; Orth, R.; Smith, R. G.; Wen, L.; Brown, W. E.; Welsch, M.; Riley, R.; Mcfarland, D.; Pandravada, A.; Glenn, K. C. *Journal of Agricultural and Food Chemistry* **2007**, *55*, 6177-6185.
- (26) Trethewey, R. N. *Current Opinion in Plant Biology* **2004**, *7*, 196-201.
- (27) Oksman-Caldentey, K. M.; Saito, K. *Current Opinion in Biotechnology* **2005**, *16*, 174-179.
- (28) Wink, M. *Theoretical and Applied Genetics* **1988**, *75*, 225-233.
- (29) Last, R. L.; Jones, A. D.; Shachar-Hill, Y. *Nature Reviews Molecular Cell Biology* **2007**, *8*, 167-174.
- (30) Fiehn, O. *Plant Molecular Biology* **2002**, *48*, 155-171.
- (31) Devaux, P. G.; Horning, M. G.; Horning, E. C. *Analytical Letters* **1971**, *4*, 151-&.
- (32) Horning, E. C.; Horning, M. G. *Journal of Chromatographic Science* **1971**, *9*, 129-&.
- (33) Sweeley, C. C.; Horning, E. C. *Nature* **1960**, *187*, 144-145.
- (34) Oldiges, M.; Lutz, S.; Pflug, S.; Schroer, K.; Stein, N.; Wiendahl, C. *Applied Microbiology and Biotechnology* **2007**, *76*, 495-511.

- (35) Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W. *Trends in Plant Science* **2004**, *9*, 418-425.
- (36) Wagner, C.; Sefkow, M.; Kopka, J. *Phytochemistry* **2003**, *62*, 887-900.
- (37) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrometry Reviews* **2007**, *26*, 51-78.
- (38) Villas-Boas, S. G.; Bruheim, P. *Omics-a Journal of Integrative Biology* **2007**, *11*, 305-313.
- (39) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrometry Reviews* **2005**, *24*, 613-646.
- (40) Schwab, W. *Phytochemistry* **2003**, *62*, 837-849.
- (41) Wyss, R. *Journal of Chromatography B-Biomedical Applications* **1995**, *671*, 381-425.
- (42) Besseau, S.; Hoffmann, L.; Geoffroy, P.; Lapierre, C.; Pollet, B.; Legrand, M. *Plant Cell* **2007**, *19*, 148-162.
- (43) Yonekura-Sakakibara, K.; Tohge, T.; Matsuda, F.; Nakabayashi, R.; Takayama, H.; Niida, R.; Watanabe-Takahashi, A.; Inoue, E.; Saito, K. *Plant Cell* **2008**, *20*, 2160-2176.
- (44) Cuesta-Rubio, O.; Piccinelli, A. L.; Fernandez, M. C.; Hernandez, I. M.; Rosado, A.; Rastrelli, L. *Journal of Agricultural and Food Chemistry* **2007**, *55*, 7502-7509.
- (45) Rieger, G.; Muller, M.; Guttenberger, H.; Bucar, F. *Journal of Agricultural and Food Chemistry* **2008**, *56*, 9080-9086.
- (46) Yeh, T. F.; Morris, C. R.; Goldfarb, B.; Chang, H. M.; Kadla, J. F. *Tree Physiology* **2006**, *26*, 1497-1503.
- (47) Wang, G.; Zhang, R.; Sun, Y.; Xie, K.; Ma, C. *Chromatographia* **2007**, *65*, 363-366.
- (48) Ward, J. L.; Harris, C.; Lewis, J.; Beale, M. H. *Phytochemistry* **2003**, *62*, 949-957.
- (49) Bailey, N. J. C.; Stanley, P. D.; Hadfield, S. T.; Lindon, J. C.; Nicholson, J. K. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 679-684.
- (50) Lenz, E. M.; Wilson, I. D. *Journal of Proteome Research* **2007**, *6*, 443-458.
- (51) Want, E. J.; Nordstrom, A.; Morita, H.; Siuzdak, G. *Journal of Proteome Research* **2007**, *6*, 459-468.

- (52) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrometry Reviews* **1998**, *17*, 1-35.
- (53) Brown, S. C.; Kruppa, G.; Dasseux, J. L. *Mass Spectrometry Reviews* **2005**, *24*, 223-231.
- (54) Birkemeyer, C.; Kolasa, A.; Kopka, J. *Journal of Chromatography A* **2003**, *993*, 89-102.
- (55) Broeckling, C. D.; Huhman, D. V.; Farag, M. A.; Smith, J. T.; May, G. D.; Mendes, P.; Dixon, R. A.; Sumner, L. W. *Journal of Experimental Botany* **2005**, *56*, 323-336.
- (56) Schauer, N.; Steinhauser, D.; Strelkov, S.; Schomburg, D.; Allison, G.; Moritz, T.; Lundgren, K.; Roessner-Tunali, U.; Forbes, M. G.; Willmitzer, L.; Fernie, A. R.; Kopka, J. *Febs Letters* **2005**, *579*, 1332-1337.
- (57) Cook, D.; Fowler, S.; Fiehn, O.; Thomashow, M. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 15243-15248.
- (58) Tolstikov, V. V.; Fiehn, O. *Analytical Biochemistry* **2002**, *301*, 298-307.
- (59) Huhman, D. V.; Sumner, L. W. *Phytochemistry* **2002**, *59*, 347-360.
- (60) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 763-769.
- (61) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306*, 471-473.
- (62) Bedair, M.; Sumner, L. W. *Trac-Trends in Analytical Chemistry* **2008**, *27*, 238-250.
- (63) Pongsuwan, W.; Bamba, T.; Harada, K.; Yonetani, T.; Kobayashi, A.; Fukusaki, E. *Journal of Agricultural and Food Chemistry* **2008**, *56*, 10705-10708.
- (64) Shen, Y. F.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 3090-3100.
- (65) Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S. *Journal of Proteome Research* **2005**, *4*, 591-598.
- (66) Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Analytical Chemistry* **2006**, *78*, 3289-3295.
- (67) Hanhineva, K.; Rogachev, I.; Kokko, H.; Mintz-Oron, S.; Venger, I.; Karenlampi, S.; Aharoni, A. *Phytochemistry* **2008**, *69*, 2463-2481.

- (68) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends in Biotechnology* **2004**, *22*, 245-252.
- (69) Chan, E. C. Y.; Yap, S. L.; Lau, A. J.; Leow, P. C.; Toh, D. F.; Koh, H. L. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 519-528.
- (70) Xie, G. X.; Plumb, R.; Su, M. M.; Xu, Z. H.; Zhao, A. H.; Qiu, M. F.; Long, X. B.; Liu, Z.; Jia, W. *Journal of Separation Science* **2008**, *31*, 1015-1026.
- (71) Zaikin, V. G.; Halket, J. M. *European Journal of Mass Spectrometry* **2006**, *12*, 79-115.
- (72) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
- (73) Wilm, M.; Mann, M. *Analytical Chemistry* **1996**, *68*, 1-8.
- (74) Fernandez-Alba, A. R.; Garcia-Reyes, J. F. *Trac-Trends in Analytical Chemistry* **2008**, *27*, 973-990.
- (75) Ferrer, I.; Thurman, E. M. *Trac-Trends in Analytical Chemistry* **2003**, *22*, 750-756.
- (76) Mihaleva, V. V.; Vorst, O.; Maliepaard, C.; Verhoeven, H. A.; de Vos, R. C. H.; Hall, R. D.; van Ham, R. C. H. J. *Metabolomics* **2008**, *4*, 171-182.
- (77) Xie, G. X.; Ni, Y.; Su, M. M.; Zhang, Y. Y.; Zhao, A. H.; Gao, X. F.; Liu, Z.; Xiao, P. G.; Jia, W. *Metabolomics* **2008**, *4*, 248-260.
- (78) Senyuva, H. Z.; Gilbert, J.; Ozturkoglu, S. *Analytica Chimica Acta* **2008**, *617*, 97-106.
- (79) Trygg, J. *Journal of Chemometrics* **2002**, *16*, 283-293.
- (80) Trygg, J.; Wold, S. *Journal of Chemometrics* **2002**, *16*, 119-128.
- (81) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2005**, *828*, 14-20.
- (82) Lenz, E. M.; Bright, J.; Knight, R.; Westwood, F. R.; Davies, D.; Major, H.; Wilson, I. D. *Biomarkers* **2005**, *10*, 173-187.
- (83) Whitfield, P. D.; Noble, P. J. M.; Major, H.; Beynon, R. J.; Burrow, R.; Freeman, A. I.; German, A. J. *Metabolomics* **2005**, *1*, 215-225.
- (84) Griffiths, W. J.; Wang, Y. Q. *Chemical Society Reviews* **2009**, *38*, 1882-1896.

- (85) Trygg, J.; Holmes, E.; Lundstedt, T. *Journal of Proteome Research* **2007**, *6*, 469-479.
- (86) Wang, Y. L.; Bollard, M. E.; Keun, H.; Antti, H.; Beckonert, O.; Ebbels, T. M.; Lindon, J. C.; Holmes, E.; Tang, H. R.; Nicholson, J. K. *Analytical Biochemistry* **2003**, *323*, 26-32.
- (87) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Communications in Mass Spectrometry* **2002**, *16*, 1991-1996.
- (88) Kim, S. W.; Ban, S. H.; Ahn, C. Y.; Oh, H. M.; Chung, H.; Cho, S. H.; Park, Y. M.; Liu, J. R. *Journal of Plant Biology* **2006**, *49*, 271-275.
- (89) Xiao, C. N.; Dai, H.; Liu, H. B.; Wang, Y. L.; Tang, H. R. *Journal of Agricultural and Food Chemistry* **2008**, *56*, 10142-10153.
- (90) Consonni, R.; Cagliani, L. R.; Stocchero, M.; Porretta, S. *Journal of Agricultural and Food Chemistry* **2009**, *57*, 4506-4513.
- (91) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. *Siam Journal on Scientific and Statistical Computing* **1984**, *5*, 735-743.
- (92) Rantalainen, M.; Cloarec, O.; Beckonert, O.; Wilson, I. D.; Jackson, D.; Tonge, R.; Rowlinson, R.; Rayner, S.; Nickson, J.; Wilkinson, R. W.; Mills, J. D.; Trygg, J.; Nicholson, J. K.; Holmes, E. *Journal of Proteome Research* **2006**, *5*, 2642-2655.
- (93) Stella, C.; Beckwith-Hall, B.; Cloarec, O.; Holmes, E.; Lindon, J. C.; Powell, J.; van der Ouderaa, F.; Bingham, S.; Cross, A. J.; Nicholson, J. K. *Journal of Proteome Research* **2006**, *5*, 2780-2788.
- (94) Moco, S.; Bino, R. J.; De Vos, R. C. H.; Vervoort, J. *Trac-Trends in Analytical Chemistry* **2007**, *26*, 855-866.
- (95) Grange, A. H.; Winnik, W.; Ferguson, P. L.; Sovocool, G. W. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 2699-2715.
- (96) Kind, T.; Fiehn, O. *Bmc Bioinformatics* **2006**, *7*, 234.
- (97) Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. C. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 792-803.
- (98) Kosaka, T.; Yoneyama-Takazawa, T.; Kubota, K.; Matsuoka, T.; Sato, I.; Sakaki, T.; Tanaka, Y. *Journal of Mass Spectrometry* **2003**, *38*, 1281-1287.

- (99) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 1989-1994.
- (100) Bottcher, C.; von Roepenack-Lahaye, E.; Schmidt, J.; Schmotz, C.; Neumann, S.; Scheel, D.; Clemens, S. *Plant Physiology* **2008**, *147*, 2107-2120.
- (101) Rochfort, S. J.; Trenerry, V. C.; Imsic, M.; Panozzo, J.; Jones, R. *Phytochemistry* **2008**, *69*, 1671-1679.
- (102) Schauer, N.; Semel, Y.; Roessner, U.; Gur, A.; Balbo, I.; Carrari, F.; Pleban, T.; Perez-Melis, A.; Bruedigam, C.; Kopka, J.; Willmitzer, L.; Zamir, D.; Fernie, A. R. *Nature Biotechnology* **2006**, *24*, 447-454.
- (103) Moco, S.; Bino, R. J.; Vorst, O.; Verhoeven, H. A.; de Groot, J.; van Beek, T. A.; Vervoort, J.; de Vos, C. H. R. *Plant Physiology* **2006**, *141*, 1205-1218.
- (104) Dayan, F. E.; Watson, S. B.; Nanayakkara, N. P. D. *Journal of Experimental Botany* **2007**, *58*, 3263-3272.
- (105) Werker, E. *Advances in Botanical Research Incorporating Advances in Plant Pathology, Vol 31 2000* **2000**, *31*, 1-35.
- (106) Wagner, G. J. *Plant Physiology* **1991**, *96*, 675-679.
- (107) Kandra, L.; Severson, R.; Wagner, G. J. *European Journal of Biochemistry* **1990**, *188*, 385-391.
- (108) Kroumova, A. B.; Wagner, G. J. *Planta* **2003**, *216*, 1013-1021.
- (109) Kroumova, A. B.; Xie, Z. Y.; Wagner, G. J. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 11437-11441.
- (110) Li, A. X.; Eannetta, N.; Ghangas, G. S.; Steffens, J. C. *Plant Physiology* **1999**, *121*, 453-460.
- (111) van der Hoeven, R. S.; Steffens, J. C. *Plant Physiology* **2000**, *122*, 275-282.
- (112) Walters, D. S.; Steffens, J. C. *Plant Physiology* **1990**, *93*, 1544-1551.
- (113) King, R. R.; Calhoun, L. A.; Singh, R. P.; Boucher, A. *Journal of Agricultural and Food Chemistry* **1993**, *41*, 469-473.
- (114) Severson, R. F.; Jackson, D. M.; Johnson, A. W.; Sisson, V. A.; Stephenson, M. G. *Acs Symposium Series* **1991**, *449*, 264-277.
- (115) King, R. R.; Calhoun, L. A.; Singh, R. P.; Boucher, A. *Phytochemistry* **1990**, *29*, 2115-2118.

- (116) King, R. R.; Calhoun, L. A. *Phytochemistry* **1988**, *27*, 3761-3763.
- (117) King, R. R.; Calhoun, L. A.; Singh, R. P. *Phytochemistry* **1988**, *27*, 3765-3768.
- (118) King, R. R.; Singh, R. P.; Calhoun, L. A. *Carbohydrate Research* **1987**, *166*, 113-121.
- (119) Slocombe, S. P.; Schauvinhold, I.; McQuinn, R. P.; Besser, K.; Welsby, N. A.; Harper, A.; Aziz, N.; Li, Y.; Larson, T. R.; Giovannoni, J.; Dixon, R. A.; Broun, P. *Plant Physiology* **2008**, *148*, 1830-1846.
- (120) Yu, F. N. A.; Utsumi, R. *Cellular and Molecular Life Sciences* **2009**, *66*, 3043-3052.
- (121) Dudareva, N.; Andersson, S.; Orlova, I.; Gatto, N.; Reichelt, M.; Rhodes, D.; Boland, W.; Gershenzon, J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 933-938.
- (122) Chappell, J. *Annual Review of Plant Physiology and Plant Molecular Biology* **1995**, *46*, 521-547.
- (123) Dicke, M.; van Loon, J. J. A. *Entomologia Experimentalis Et Applicata* **2000**, *97*, 237-249.
- (124) Valkama, E.; Salminen, J. P.; Koricheva, J.; Pihlaja, K. *Annals of Botany* **2004**, *94*, 233-242.
- (125) Wollenweber, E.; Mann, K. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences* **1984**, *39*, 303-306.
- (126) Voirin, B.; Bayet, C.; Colson, M. *Phytochemistry* **1993**, *34*, 85-87.
- (127) Friedman, M. *Journal of Agricultural and Food Chemistry* **2002**, *50*, 5751-5780.
- (128) Kennedy, G. G. *Annual Review of Entomology* **2003**, *48*, 51-72.
- (129) Carter, C. D.; Gianfagna, T. J.; Sacalis, J. N. *Journal of Agricultural and Food Chemistry* **1989**, *37*, 1425-1428.
- (130) Li, L.; Zhao, Y. F.; McCaig, B. C.; Wingerd, B. A.; Wang, J. H.; Whalon, M. E.; Pichersky, E.; Howe, G. A. *Plant Cell* **2004**, *16*, 783-783.
- (131) Eshed, Y.; Zamir, D. *Genetics* **1995**, *141*, 1147-1162.
- (132) Goffreda, J. C.; Steffens, J. C.; Mutschler, M. A. *Journal of the American Society for Horticultural Science* **1990**, *115*, 161-165.
- (133) DellaPenna, D.; Last, R. L. *Science* **2008**, *320*, 479-481.

- (134) Want, E. J.; Cravatt, B. F.; Siuzdak, G. *Chembiochem* **2005**, *6*, 1941-1951.
- (135) Tiller, P. R.; Land, A. P.; Jardine, I.; Murphy, D. M.; Sozio, R.; Ayrton, A.; Schaefer, W. H. *Journal of Chromatography A* **1998**, *794*, 15-25.
- (136) Blackler, A. R.; Klammer, A. A.; MacCoss, M. J.; Wu, C. C. *Analytical Chemistry* **2006**, *78*, 1337-1344.
- (137) Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chen, K.; Wahl, R. C.; Yu, W.; Luethy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 93-107.
- (138) Williams, J. D.; Flanagan, M.; Lopez, L.; Fischer, S.; Miller, L. A. D. *Journal of Chromatography A* **2003**, *1020*, 11-26.
- (139) Chakraborty, A. B.; Berger, S. J.; Gebler, J. C. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 730-744.
- (140) Wang, Y. S.; Vivekananda, S.; Men, L. J.; Zhang, Q. B. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 697-702.
- (141) Seymour, J. L.; Costello, C. E.; Zaia, J. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 844-854.
- (142) Zaia, J.; Miller, M. J. C.; Seymour, J. L.; Costello, C. E. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 952-960.
- (143) Lips, A. G. A. M.; Lameijer, W.; Fokkens, R. H.; Nibbering, N. M. M. *Journal of Chromatography B* **2001**, *759*, 191-207.
- (144) Tian, Q. G.; Duncan, C. J. G.; Schwartz, S. J. *Journal of Mass Spectrometry* **2003**, *38*, 990-995.
- (145) Bateman, K. P.; Castro-Perez, J.; Wrona, M.; Shockcor, J. P.; Yu, K.; Oballa, R.; Nicoll-Griffith, D. A. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 1485-1496.
- (146) Rainville, P. D.; Stumpf, C. L.; Shockcor, J. P.; Plumb, R. S.; Nicholson, J. K. *Journal of Proteome Research* **2007**, *6*, 552-558.
- (147) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. *Analytical Chemistry* **2005**, *77*, 2187-2200.
- (148) Ramos, A. A.; Yang, H.; Rosen, L. E.; Yao, X. D. *Analytical Chemistry* **2006**, *78*, 6391-6397.

- (149) Wrona, M.; Mauriala, T.; Bateman, K. P.; Mortishire-Smith, R. J.; O'Connor, D. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 2597-2602.
- (150) Cheng, F. Y.; Blackburn, K.; Lin, Y. M.; Goshe, M. B.; Williamson, J. D. *Journal of Proteome Research* **2009**, *8*, 82-93.
- (151) Zhang, H. Y.; Grubb, M.; Wu, W.; Josephs, J.; Humphreys, W. G. *Analytical Chemistry* **2009**, *81*, 2695-2700.
- (152) Grata, E.; Boccard, J.; Guillarme, D.; Glauser, G.; Carrupt, P. A.; Farmer, E. E.; Wolfender, J. L.; Rudaz, S. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2008**, *871*, 261-270.
- (153) Kaufmann, A.; Butcher, P.; Maden, K.; Widmer, M. *Analytica Chimica Acta* **2007**, *586*, 13-21.
- (154) Glauser, G.; Grata, E.; Rudaz, S.; Wolfender, J. L. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 3154-3160.
- (155) Weinmann, W.; Wiedemann, A.; Eppinger, B.; Renz, M.; Svoboda, M. *Journal of American Society for Mass Spectrometry* **1999**, *10*, 1028-1037.
- (156) Kind, T.; Fiehn, O. *Bmc Bioinformatics* **2006**, *7*.
- (157) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2007**, *8*, 105.
- (158) Stein, S. E. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 770-781.
- (159) Mei, H.; Hsieh, Y. S.; Nardo, C.; Xu, X. Y.; Wang, S. Y.; Ng, K.; Korfmacher, W. A. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 97-103.
- (160) Wagner, G. J.; Wang, E.; Shepherd, R. W. *Annals of Botany* **2004**, *93*, 3-11.
- (161) Puterka, G. J.; Farone, W.; Palmer, T.; Barrington, A. *Journal of Economic Entomology* **2003**, *96*, 636-644.
- (162) Xia, Y. L.; Johnson, A. W.; Chortyk, O. T. *Journal of Economic Entomology* **1997**, *90*, 1015-1021.
- (163) Hill, K.; Rhode, O. *Fett-Lipid* **1999**, *101*, 25-33.
- (164) Burke, B. A.; Goldsby, G.; Mudd, J. B. *Phytochemistry* **1987**, *26*, 2567-2571.
- (165) King, R. R.; Singh, R. P.; Boucher, A. *American Potato Journal* **1987**, *64*, 529-534.
- (166) Spring, O.; Heil, N.; Vogler, B. *Phytochemistry* **1997**, *46*, 1369-1373.

- (167) Wollenweber, E.; Stevens, J. F.; Ivanic, M.; Deinzer, M. L. *Phytochemistry* **1998**, *48*, 931-939.
- (168) Asai, T.; Hara, N.; Kobayashi, S.; Kohshima, S.; Fujimoto, Y. *Helvetica Chimica Acta* **2009**, *92*, 1473-1494.
- (169) Shapiro, J. A.; Steffens, J. C.; Mutschler, M. A. *Biochemical Systematics and Ecology* **1994**, *22*, 545-561.
- (170) SerratoValenti, G.; Bisio, A.; Cornara, L.; Ciarallo, G. *Annals of Botany* **1997**, *79*, 329-336.
- (171) Juo, C. G.; Chiu, D. T. Y.; Shiao, M. S. *Biofactors* **2008**, *34*, 159-169.
- (172) Yang, S.; Sadilek, M.; Synovec, R. E.; Lidstrom, M. E. *Journal of Chromatography A* **2009**, *1216*, 3280-3289.
- (173) Griffiths, W. J.; Karua, K.; Hornshaw, M.; Woffendin, G.; Wanga, Y. *European Journal of Mass Spectrometry* **2007**, *13*, 45-50.
- (174) Stobiecki, M.; Skiryecz, A.; Kerhoas, L.; Kachlicki, P.; Muth, D.; Einhorn, J.; Mueller-Roeber, B. *Metabolomics* **2006**, *2*, 197-219.
- (175) Farag, M. A.; Huhman, D. V.; Lei, Z. T.; Sumner, L. W. *Phytochemistry* **2007**, *68*, 342-354.
- (176) Kong, H.; Wang, M.; Venema, K.; Maathuis, A.; van der Heijden, R.; van der Greef, J.; Xu, G.; Hankemeier, T. *Journal of Chromatography A* **2009**, *1216*, 2195-2203.
- (177) Nakamura, Y.; Kanaya, S.; Sakurai, N.; Iijima, Y.; Aoki, K.; Okazaki, K.; Suzuki, H.; Kitayama, M.; Shibata, D. *Plant Biotechnology* **2008**, *25*, 377-380.
- (178) Holman, S. W.; Wright, P.; Langley, G. J. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 2355-2365.
- (179) Kachlicki, P.; Marczak, L.; Kerhoas, L.; Einhorn, J.; Stobiecki, M. *Journal of Mass Spectrometry* **2005**, *40*, 1088-1103.
- (180) Li, A. X.; Steffens, J. C. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 6902-6907.
- (181) Hsu, F. F.; Turk, J. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 892-899.
- (182) Hsu, F. F.; Turk, J. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 797-803.

- (183) Ding, L.; Xie, F. W.; Zhao, M. Y.; Xie, J. P.; Xu, G. W. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 2816-2822.
- (184) Matsuda, F.; Yonekura-Sakakibara, K.; Niida, R.; Kuromori, T.; Shinozaki, K.; Saito, K. *Plant Journal* **2009**, *57*, 555-577.
- (185) Friedman, M. *Journal of Agricultural and Food Chemistry* **1997**, *45*, 1523-1540.
- (186) Barbour, J. D.; Kennedy, G. G. *Journal of Chemical Ecology* **1991**, *17*, 989-1005.
- (187) Schijlen, E. G. W.; de Vos, C. H. R.; van Tunen, A. J.; Bovy, A. G. *Phytochemistry* **2004**, *65*, 2631-2648.
- (188) Bovy, A.; Schijlen, E.; Hall, R. D. *Metabolomics* **2007**, *3*, 399-412.
- (189) Horper, W.; Marner, F. J. *Phytochemistry* **1996**, *41*, 451-456.
- (190) Lazar, G.; Goodman, H. M. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 472-476.
- (191) McCue, K. F.; Shepherd, L. V. T.; Rockhold, D. R.; Allen, P. V.; Davies, H. V.; Belknap, W. R. *Abstracts of Papers of the American Chemical Society* **2005**, *229*, U83-U83.
- (192) Friedman, M. *Journal of Agricultural and Food Chemistry* **2006**, *54*, 8655-8681.
- (193) McCue, K. F.; Shepherd, L. V. T.; Allen, P. V.; Maccree, M. M.; Rockhold, D. R.; Corsini, D. L.; Davies, H. V.; Belknap, W. R. *Plant Science* **2005**, *168*, 267-273.
- (194) Colby, S. M.; Crock, J.; Dowdle-Rizzo, B.; Lemaux, P. G.; Croteau, R. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 2216-2221.
- (195) van Schie, C. C. N.; Haring, M. A.; Schuurink, R. C. *Plant Molecular Biology* **2007**, *64*, 251-263.
- (196) Coates, R. M.; Ho, J. Z.; Klobus, M.; Zhu, L. J. *Journal of Organic Chemistry* **1998**, *63*, 9166-9176.
- (197) Besser, K.; Harper, A.; Welsby, N.; Schauvinhold, I.; Slocombe, S.; Li, Y.; Dixon, R. A.; Broun, P. *Plant Physiology* **2009**, *149*, 499-514.
- (198) Iijima, Y.; Nakamura, Y.; Ogata, Y.; Tanaka, K.; Sakurai, N.; Suda, K.; Suzuki, T.; Suzuki, H.; Okazaki, K.; Kitayama, M.; Kanaya, S.; Aoki, K.; Shibata, D. *Plant Journal* **2008**, *54*, 949-962.
- (199) Tohge, T.; Nishiyama, Y.; Hirai, M. Y.; Yano, M.; Nakajima, J.; Awazuhara, M.; Inoue, E.; Takahashi, H.; Goodenowe, D. B.; Kitayama, M.; Noji, M.; Yamazaki, M.; Saito, K. *Plant Journal* **2005**, *42*, 218-235.

- (200) Moco, S.; Forshed, J.; De Vos, R. C. H.; Bino, R. J.; Vervoort, J. *Metabolomics* **2008**, *4*, 202-215.
- (201) Cuyckens, F.; Claeys, M. *Journal of Mass Spectrometry* **2004**, *39*, 1-15.
- (202) Moco, S.; Capanoglu, E.; Tikunov, Y.; Bino, R. J.; Boyacioglu, D.; Hall, R. D.; Vervoort, J.; De Vos, R. C. H. *Journal of Experimental Botany* **2007**, *58*, 4131-4146.
- (203) Tohge, T.; Matsui, K.; Ohme-Takagi, M.; Yamazaki, M.; Saito, K. *Biotechnology Letters* **2005**, *27*, 297-303.
- (204) Tohge, T.; Yonekura-Sakakibara, K.; Niida, R.; Watanabe-Takahashi, A.; Saito, K. *Pure and Applied Chemistry* **2007**, *79*, 811-823.
- (205) Lepiniec, L.; Debeaujon, I.; Routaboul, J. M.; Baudry, A.; Pourcel, L.; Nesi, N.; Caboche, M. *Annual Review of Plant Biology* **2006**, *57*, 405-430.
- (206) Winkel-Shirley, B. *Plant Physiology* **2001**, *126*, 485-493.
- (207) Ma, Y. L.; Li, Q. M.; VandenHeuvel, H.; Claeys, M. *Rapid Communications in Mass Spectrometry* **1997**, *11*, 1357-1364.
- (208) Cataldi, T. R. I.; Lelario, F.; Bufo, S. A. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 3103-3110.
- (209) Claeys, M.; VandenHeuvel, H.; Chen, S.; Derrick, P. J.; Mellon, F. A.; Price, K. R. *Journal of the American Society for Mass Spectrometry* **1996**, *7*, 173-181.
- (210) Dal Piaz, F.; Imperato, S.; Lepore, L.; Bader, A.; De Tommasi, N. *Journal of Pharmaceutical and Biomedical Analysis* **2010**, *51*, 70-77.
- (211) Ungur, N.; Kulcitki, V. *Tetrahedron* **2009**, *65*, 3815-3828.
- (212) Schillmiller, A. L.; Last, R. L.; Pichersky, E. *Plant Journal* **2008**, *54*, 702-711.
- (213) Wagner, G. J. *Plant Physiol* **1991**, *96*, 675-679.
- (214) Iijima, Y.; Davidovich-Rikanati, R.; Fridman, E.; Gang, D. R.; Bar, E.; Lewinsohn, E.; Pichersky, E. *Plant Physiol* **2004**, *136*, 3724-3736.
- (215) Croteau, R. B.; Davis, E. M.; Ringer, K. L.; Wildung, M. R. *Naturwissenschaften* **2005**, *92*, 562-577.
- (216) Nagel, J.; Culley, L. K.; Lu, Y.; Liu, E.; Matthews, P. D.; Stevens, J. F.; Page, J. E. *Plant Cell* **2008**, *20*, 186-200.
- (217) Gang, D. R.; Wang, J.; Dudareva, N.; Nam, K. H.; Simon, J. E.; Lewinsohn, E.; Pichersky, E. *Plant Physiol* **2001**, *125*, 539-555.

- (218) Schillmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 10865-10870.
- (219) Benning, C. *Analytical Biochemistry* **2004**, *332*, 1.
- (220) Eshed, Y.; Zamir, D. *Genetics* **1995**, *141*, 1147-1162.
- (221) Liu, Y. S.; Gur, A.; Ronen, G.; Causse, M.; Damidaux, R.; Buret, M.; Hirschberg, J.; Zamir, D. *Plant Biotechnol J* **2003**, *1*, 195-207.
- (222) Ronen, G.; Carmel-Goren, L.; Zamir, D.; Hirschberg, J. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 11102-11107.
- (223) Schillmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 10865-10870.
- (224) Overy, S. A.; Walker, H. J.; Malone, S.; Howard, T. P.; Baxter, C. J.; Sweetlove, L. J.; Hill, S. A.; Quick, W. P. *Journal of Experimental Botany* **2005**, *56*, 287-296.
- (225) Hall, R. D. *New Phytologist* **2006**, *169*, 453-468.
- (226) Roessner, U.; Willmitzer, L.; Fernie, A. R. *Plant Cell Reports* **2002**, *21*, 189-196.
- (227) Hirai, M. Y.; Klein, M.; Fujikawa, Y.; Yano, M.; Goodenowe, D. B.; Yamazaki, Y.; Kanaya, S.; Nakamura, Y.; Kitayama, M.; Suzuki, H.; Sakurai, N.; Shibata, D.; Tokuhsa, J.; Reichelt, M.; Gershenzon, J.; Papenbrock, J.; Saito, K. *Journal of Biological Chemistry* **2005**, *280*, 25590-25595.
- (228) Goodacre, R.; York, E. V.; Heald, J. K.; Scott, I. M. *Phytochemistry* **2003**, *62*, 859-863.
- (229) Jander, G.; Norris, S. R.; Joshi, V.; Fraga, M.; Rugg, A.; Yu, S. X.; Li, L. L.; Last, R. L. *Plant Journal* **2004**, *39*, 465-475.
- (230) Tolstikov, V. V.; Lommen, A.; Nakanishi, K.; Tanaka, N.; Fiehn, O. *Analytical Chemistry* **2003**, *75*, 6737-6740.
- (231) von Roepenack-Lahaye, E.; Degenkolb, T.; Zerjeski, M.; Franz, M.; Roth, U.; Wessjohann, L.; Schmidt, J.; Scheel, D.; Clemens, S. *Plant Physiology* **2004**, *134*, 548-559.
- (232) Vorst, O.; de Vos, C. H. R.; Lommen, A.; Staps, R. V.; Visser, R. G. F.; Bino, R. J.; Hall, R. D. *Metabolomics* **2005**, *1*, 169-180.

- (233) Rischer, H.; Oresic, M.; Seppanen-Laakso, T.; Katajamaa, M.; Lammertyn, F.; Ardiles-Diaz, W.; Van Montagu, M. C. E.; Inze, D.; Oksman-Caldentey, K. M.; Goossens, A. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 5614-5619.
- (234) Sato, S.; Soga, T.; Nishioka, T.; Tomita, M. *Plant Journal* **2004**, *40*, 151-163.
- (235) Le Gall, G.; Colquhoun, I. J.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E. *Journal of Agricultural and Food Chemistry* **2003**, *51*, 2447-2456.
- (236) De Vos, R. C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall, R. D. *Nature Protocols* **2007**, *2*, 778-791.
- (237) Smedsgaard, J.; Frisvad, J. C. *Journal of Microbiological Methods* **1996**, *25*, 5-17.
- (238) Smedsgaard, J. *Journal of Chromatography A* **1997**, *760*, 264-270.
- (239) Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. *Nature Biotechnology* **2003**, *21*, 692-696.
- (240) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J.; Oliver, S. G. *Phytochemistry* **2003**, *62*, 929-937.
- (241) Kirkland, J. J.; Langlois, T. J.; DeStefano, J. J. *American Laboratory* **2007**, *39*, 18-21.
- (242) Cunliffe, J. M.; Adams-Hall, S. B.; Maloney, T. D. *Journal of Separation Science* **2007**, *30*, 1214-1223.
- (243) Salisbury, J. J. *Journal of Chromatographic Science* **2008**, *46*, 883-886.
- (244) Fobes, J. F.; Mudd, J. B.; Marsden, M. P. *Plant Physiol* **1985**, *77*, 567-570.
- (245) Kandra, G.; Severson, R.; Wagner, G. J. *European Journal of Biochemistry* **1990**, *188*, 385-391.
- (246) van der Hoeven, R. S.; Steffens, J. C. *Plant Physiol* **2000**, *122*, 275-282.
- (247) Friedman, M.; Kozukue, N.; Harden, L. A. *Journal of Agricultural and Food Chemistry* **1997**, *45*, 1541-1547.
- (248) Lawson, D. M.; Lunde, C. F.; Mutschler, M. A. *Molecular Breeding* **1997**, *3*, 307-317.
- (249) Liedl, B. E.; Lawson, D. M.; White, K. K.; Shapiro, J. A.; Cohen, D. E.; Carson, W. G.; Trumble, J. T.; Mutschler, M. A. *Journal of Economic Entomology* **1995**, *88*, 742-748.

- (250) Hawthorne, D. J.; Shapiro, J. A.; Tingey, W. M.; Mutschler, M. A. *Entomologia Experimentalis et Applicata* **1992**, *65*, 65-73.
- (251) Nonomura, T.; Xu, L.; Wada, M.; Kawamura, S.; Miyajima, T.; Nishitomi, A.; Kakutani, K.; Takikawa, Y.; Matsuda, Y.; Toyoda, H. *Plant Science* **2009**, *176*, 31-37.
- (252) Fobes, J. F.; Mudd, J. B.; Marsden, M. P. F. *Plant Physiology* **1985**, *77*, 567-570.
- (253) Shinozaki, Y.; Matsuzaki, T.; Suhara, S.; Tobita, T.; Shigematsu, H.; Koiwai, A. *Agricultural and Biological Chemistry* **1991**, *55*, 751-756.
- (254) Bonierbale, M. W.; Plaisted, R. L.; Pineda, O.; Tanksley, S. D. *Theoretical and Applied Genetics* **1994**, *87*, 973-987.
- (255) Chortyk, O. T.; Kays, S. J.; Teng, Q. *Journal of Agricultural and Food Chemistry* **1997**, *45*, 270-275.
- (256) McKenzie, C. L.; Weathersbee, A. A.; Puterka, G. J. *Journal of Economic Entomology* **2005**, *98*, 1242-1247.
- (257) Chortyk, O. T.; Severson, R. F.; Cutler, H. C.; Sisson, V. A. *Bioscience Biotechnology and Biochemistry* **1993**, *57*, 1355-1356.
- (258) Maldonado, E.; Torres, F. R.; Martinez, M.; Perez-Castorena, A. L. *Journal of Natural Products* **2006**, *69*, 1511-1513.
- (259) Blauth, S. L.; Steffens, J. C.; Churchill, G. A.; Mutschler, M. A. *Theoretical and Applied Genetics* **1999**, *99*, 373-381.
- (260) Blauth, S. L.; Churchill, G. A.; Mutschler, M. A. *Theoretical and Applied Genetics* **1998**, *96*, 458-467.