

ABSTRACT

Fueling Advances in Ultra Performance Liquid Chromatography – and Ion Mobility – Mass Spectrometry as Platforms for X-omic Analyses

Sharon Michelle Munisamy, Ph.D.

Mentor: C. Kevin Chambliss, Ph.D.

Analysis of x-omics samples presents a significant challenge to analytical chemists. The complexity and diversity of x-omics samples make it difficult to determine or develop the right analytical approach. This dissertation presents novel applications of ultra performance liquid chromatography – high-resolution mass spectrometry (UPLC-HRMS) and ion mobility – high-resolution mass spectrometry (IM-HRMS) in two emerging x-omics areas: lignocellulomics and petroleomics. Compared to traditional methods of x-omics analyses, these techniques offer higher chromatographic efficiency with UPLC, which can equate to faster analysis times and greater separation, an additional degree of separation with ion mobility spectrometry (IMS), and high peak resolution and mass accuracy with HRMS, which allow for identification of known and unknown analytes.

Direct infusion electrospray ionization – ion mobility – high-resolution mass spectrometry (DIESI-IM-HRMS) has been evaluated as a rapid technique for the determination of total molecular composition in “whole-sample” biomass hydrolysates

and extracts. IM-HRMS data reveal a high molecular weight range of biomass components (up to 1100 m/z) and provide trendline isolation of feedstock components from those introduced “in process”. Carbohydrates and other lignocellulosic degradation products were identified via HRMS exact mass measurements (with typical mass errors less than 5 ppm). Analyte assignments were supported via IM-MS collision-cross-section (CCS) measurements and trendline analysis.

An automated “omics” approach utilizing UPLC-HR-TOF-MS was developed for the identification of previously unknown lignocellulosic degradation products. The approach combines traditional HRMS techniques with a “metabolomics” method for novel compound identification. Evaluation of the method for a small subset (16) of the identified peaks enabled unambiguous molecular formula assignment for 69% of the peaks.

Ion mobility – mass spectrometry and several model compounds were employed to better understand asphaltene molecules in terms of their structural types and degree of structural diversity. CCS analysis demonstrates that both monomeric and dimeric archipelago- and island-type structures might be present in asphaltenes. Comparison of arrival time peak widths for asphaltenes and model compounds indicates structural or conformational diversity for asphaltene compounds within a given nominal m/z .

Fueling Advances in Ultra Performance Liquid Chromatography – and Ion Mobility –
Mass Spectrometry as Platforms for X-omic Analyses

by

Sharon Michelle Munisamy, B.S.

A Dissertation

Approved by the Department of Chemistry and Biochemistry

Patrick J. Farmer, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

C. Kevin Chambliss, Ph.D., Chairperson

Stephen L. Gipson, Ph.D.

Touradj Solouki, Ph.D.

Charles M. Garner, Ph.D.

William C. Hockaday, Ph.D.

Accepted by the Graduate School
May 2012

J. Larry Lyon, Ph.D., Dean

Copyright © 2012 by Sharon Michelle Munisamy

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	xii
ACKNOWLEDGMENTS	xiv
CHAPTER ONE-Introduction to X-omics Areas and Analyses	1
Introduction to X-omes and X-omics: A Brief History	1
Traditional and Contemporary Techniques for X-omics Analyses	2
Established X-omics: Genomics and Proteomics	3
Emerging X-omics: Petroleomics and Lignocellulomics	6
Common Limitations in X-omics Analyses.....	10
UPLC-IM-HRMS as an Emerging Tool for X-omics Analyses.....	11
Ultra Performance Liquid Chromatography	11
Ion Mobility Spectrometry.....	12
High-Resolution Time-of-Flight Mass Spectrometry	16
Scope of the Dissertation	19
CHAPTER TWO- Direct Infusion Electrospray Ionization – Ion Mobility – High-Resolution Mass Spectrometry (DIESI-IM-HRMS) for Rapid Characterization of Potential	21
Introduction.....	21
Experimental: Materials and Methods	23
Chemicals and Reagents	23
Preparation of Standards and Generation of Aqueous Extract and Hydrolysate Samples	24

DIESI-IM-HRMS Analysis	25
Collision-Cross-Section Measurements.....	26
Results and Discussion	27
Analyte Identification via HRMS	27
Characterization of IM-MS Spectra of Biomass Pretreatment Hydrolysates and Aqueous Extracts	32
Evaluation of IM-HRMS Separation of Biomass Components.....	42
Mass-Mobility Correlations of Carbohydrates in Biomass	48
Collision-Cross-Section Measurements of Biomass Analytes	50
Conclusions.....	56
CHAPTER THREE- A “Metabolomics” Approach for Identification of Unknown Degradation Products in Lignocellulosic Samples using Ultra Performance Liquid Chromatography – High-Resolution Mass Spectrometry	58
Introduction.....	58
Experimental: Materials and Methods	60
Chemicals and Reagents	60
Generation of Extracted Hydrolysate Samples	60
Preparation of Standards	61
Ultra Performance Liquid Chromatography	61
High-Resolution Mass Spectrometry	62
MetaboLynx Methodology	63
Results and Discussion	66
UPLC Method Development	66
Evaluation of HR-TOF-MS Mass Spectral Data	68
A “Metabolomics” Approach to Unknown Component Identification	71

Development of MetaboLynx Methodology	71
Automated Detection of Unknown Peaks in Hydrolysate	73
Compositional Analysis from MetaboLynx-Generated Mass Spectral Data	75
Kendrick Mass Defect Analyses for Identification of Functional Groups.....	80
Comparison of Automated Compositional Analysis to Manual Data Inspection.....	92
Visual identification of new peaks.....	92
Manual MS/MS vs. automated MS ^e analyses	93
Conclusions.....	104
CHAPTER FOUR- Ion Mobility – Mass Spectrometry Analysis of Asphaltenes using Model Compounds to Understand Structure, Diversity, and Aggregation.....	106
Introduction.....	106
Experimental: Materials and Methods	108
Chemicals and Reagents	108
Preparation of Standards and Asphaltene Samples.....	112
Direct Infusion-ESI-IM-HRMS Analysis	112
Collision-Cross-Section Measurements.....	113
Results and Discussion	114
DIESI-IM-HRMS Spectra of Asphaltenes.....	114
Comparison of IM-MS Results for Asphaltene and Model Compounds.....	117
Collision-cross-section measurements to evaluate general structure type.....	117
Arrival time peak width measurements to assess	

structural diversity	120
Considerations in Asphaltene Aggregation	125
Conclusions.....	128
CHAPTER FIVE- Future Directions	129
APPENDICES	134
APPENDIX A: Collision Cross Section Measurements.....	135
Lignocellulomic Analytes (Chapter 2).....	135
Petroleomic Analytes (Chapter 4).....	143
APPENDIX B: Additional DIESI-IM-HRMS Data	149
APPENDIX C: Optimization and Development of the UPLC Method Presented in Chapter Three	153
APPENDIX D: Development and Optimization of MetaboLynx and MS ^e Viewer Software	158
Complete List of MetaboLynx Method Parameters.....	158
MetaboLynx Method Validation Experiments	164
REFERENCES	165

LIST OF FIGURES

Figure	Page
1.1 Illustration of the number of publications for common “omes” and “omics” areas	2
1.2 Mass spectrum of a dilute-acid-pretreated corn stover hydrolysate	9
1.3 Example of the separation of two ions during a portion of a traveling wave ion mobility separation.....	14
1.4 IM-MS spectra (m/z vs. drift time) of a mixture of compounds	16
1.5 Waters Acquity Liquid Chromatograph and Synapt HDMS Mass Spectrometer.....	19
2.1 Negative ion mode mass spectra of (a) full scan (m/z 100-1000) corn stover hydrolysate and collision-induced-dissociation spectra of (b) 6-carbon trisaccharide (827.2 m/z), (c) 6-carbon tetrasaccharide (665.1 m/z), and (d) 6-carbon trisaccharide (341.1 m/z) in a corn stover hydrolysate.....	28
2.2 Negative ion mode mass spectra of (a) 4-hydroxycoumaric acid (163.04 m/z), (b) homovanillic acid/syringaldehyde (181.05 m/z), and (c) ferulic acid/3-hydroxy-4-methoxycinnamic acid (193.05 m/z) in a corn stover hydrolysate.....	32
2.3 DIESI-IM-MS plot of m/z vs. arrival time for negative ions resulting from 1:10 dilutions of (a) 0.07% H_2SO_4 -pretreated sorghum hydrolysate with three distinct trendlines (I, II, and III) and (b) aqueous extract of sorghum with two distinct trendlines (I and II) and a region with several unresolved trendlines (III)	33
2.4 IM-MS spectra of a) MTBE-extracted raw sorghum hydrolysate and b) aqueous fraction leftover after MTBE extraction with labeled trendlines (I, II, and III)	36
2.5 Extracted high-resolution mass spectra of the middle trendline (Figure 2.3a, II) resulting from DIESI-IM-HRMS analysis of a dilute-acid pretreated sorghum hydrolysate	38
2.6 Negative ion mode mass spectra of (a) $K(HSO_4)_2^-$ (232.87 m/z), (b) $NaK_2(HSO_4)_2SO_4^-$ (390.77 m/z) , (c) $NaK_2(HSO_4)_4^-$ (488.73 m/z) , and (d) $NaK_3(HSO_4)_4^-$ (624.65 m/z) from trendline II (middle trendline).....	39

2.7	Deconvolution of a representative portion of the mass spectra of a dilute-acid pretreated sorghum hydrolysate.....	39
2.8	IM-MS spectra and inset of the mass spectrum from a control experiment, where a dilute-acid ACE pretreatment was performed without any sample (no biomass).....	40
2.9	IM-MS spectra with mass spectrum inset of H ₂ SO ₄ -pretreated sorghum hydrolysate neutralized with Ca(OH) ₂	42
2.10	Three-dimensional plot of arrival time versus <i>m/z</i> for a pentose tetramer (Pent ₄ , 545.1714 <i>m/z</i>), a tentatively identified hexuronic acid trimer (HexA ₃ , 545.1003 <i>m/z</i>), and an unknown component (unk*, 545.14 <i>m/z</i>).....	44
2.11	Representative HRMS and IM-HRMS data for corn stover and sorghum hydrolysates.....	46
2.12	Arrival time distributions for the [M – H] [–] ion of four 6-carbon trisaccharide isomer standards and the corresponding trisaccharide signal observed in corn stover and sorghum hydrolysates.....	47
2.13	Mass-mobility correlations for 5- and 6-carbon carbohydrate ions observed in corn stover hydrolysate	49
3.1	Example of the way MetaboLynx identifies expected and unexpected metabolites	64
3.2	Representative figures for primary biomass components of cellulose, hemicellulose, and lignin	65
3.3	Time-optimized UPLC chromatogram of 35 common biomass degradation products	67
3.4	UPLC chromatograms, resolution-optimized for 35 known biomass degradation products in a 5 ppm mixture(top), AFEX-pretreated corn stover hydrolysate (middle), and dilute H ₂ SO ₄ -pretreated corn stover hydrolysate (bottom).....	69
3.5	Mass spectra for a water blank and an AFEX-pretreated hydrolysate of corn stover at 21.1 minutes.....	71
3.6	<i>M/z</i> distribution of unknown peaks, expressed as a percentage of the total peaks found in the 200-3500 intensity range, for the AFEX and and dilute-acid pretreated hydrolysate samples	76

3.7	Retention time distribution of unknown peaks, expressed as a percentage of the total peaks found in the 200-3500 intensity range, for the AFEX and dilute-acid pretreated hydrolysate samples	76
3.8	Mass defect (exact mass – nominal mass) as a function of nominal mass for new m/z peaks (200-3500 area units) in the AFEX and dilute acid hydrolysate samples	78
3.9	Mass defect as a function of nominal mass for peaks in AFEX and dilute-acid hydrolysate sample at given 10-minute retention time ranges	79
3.10	Kendrick mass defect plots for four functional groups (-OH, -OCH ₃ , CO ₂ , and CH ₂) for the target analytes in a 5 ppm standard mixture.....	82
3.11	Structures for levulinic and adipic acids	85
3.12	Kendrick mass defect plot for –OH for the new m/z peaks (200-3500 area units) and target analytes in the AFEX sample	86
3.13	Kendrick mass defect plot for –OCH ₃ for the new m/z peaks (200-3500 area units) and target analytes in the AFEX sample	87
3.14	Kendrick mass defect plot for –CO ₂ for the new m/z peaks (200-3500 area units) in the AFEX sample	88
3.15	Kendrick mass defect plot for –CH ₂ for the new m/z peaks (200-3500 area units) and target analytes in the AFEX sample	89
3.16	Collision-induced-dissociation spectra of m/z 323, 484, 645, 806, and 967, which correspond to glucosamine of varying degrees (2-5) of polymerization	100
3.17	Example of how KMD analyses and fragmentation data can help narrow down molecular formula possibilities.....	102
4.1	Arrival time distribution for ions in the 600-605 m/z range of an asphaltene sample	114
4.2	Direct infusion-ESI-IM-MS spectra of a 0.1 mg/mL asphaltene sample in 60:40 toluene-methanol at an mobility wave height of 7.5 V and gas flow rate of 25 mL/min.....	115
4.3	Extracted mass spectra for (a) trendline I (asphaltene signal), (b) trendline II (non-asphaltene signal), and (c) trendline III (unassigned signal) indicated in Figure 4.2	116

4.4	Comparison of collision cross sections for asphaltenes (average for peaks in a 5 Da window) and model compounds (Table 4.1)	118
4.5	Arrival time distributions for $[M+H]^+$ ions of archipelago-type (PmPhP) and island-type (THP) structures at a wave height of 7.5 V	119
4.6	Arrival time peak width as a function of arrival time for polyaromatic hydrocarbons and other model compounds used in this study (see Table 4.1) at two mobility cell wave heights	121
4.7	Mass spectrum (top left) of asphaltenes in the m/z range 600-605. Arrival time distributions (a-e) are shown for each selected nominal m/z peak in that range and peak width measurements at half height are shown	122
4.8	Average peak widths at half height for asphaltene and model compound ions as a function of average arrival time	123
4.9	Average peak widths at half height for asphaltene and model compound ions as a function of m/z	124
4.10	Arrival time distributions for model compounds and ions at similar m/z in an asphaltene sample	125
4.11	Comparison of collision cross sections for asphaltenes (average for peaks in a 5 Da window) and observed model compound dimers	127
5.1	UPLC-HR-TOF-MS chromatogram of dilute-acid-pretreated hydrolysate of corn stover with an extended run time relative to the presented in Chapter 3	131
5.2	Chromatograms of MTBE-extracts of 0.7% H_2SO_4 -pretreated xylan, inulin, avicel, and cellulose	132
5.3	IM-MS spectra obtained with a next generation Synapt (G2) for an unextracted, dilute- acid pretreated hydrolysate of corn stover	133
A.1	Plot of $\ln(\Omega') = X \times \ln(t'_D)$ used as the first step in collision-cross-section measurement of the lignocellulosic analytes described in chapter 2	135
A.2	Calibration plot used for collision-cross-section measurement of the lignocellulosic analytes, described in chapter 2	136
A.3	Plot of $\ln(\Omega') = X \times \ln(t'_D)$ used as the first step in collision-cross-section measurement of the model compounds and	

	petroleomic analytes described in chapter 4	143
A.4	Calibration plot used for collision-cross-section measurement of the model compounds and petroleomic analytes, described in Chapter 4.....	144
B.1	DIESI-IM-MS plot of m/z vs. arrival time for negative ions resulting from a 1:10 dilution of dilute-acid-pretreated corn stover hydrolysate with four distinct trendlines (I, II, III, and IV)	149
B.2	Examples of the resolving power differences between V and W mode for several ions in a corn stover hydrolysate	150
B.3	M/z peak for the $[M+H]^+$ ion of P-2,5-PMP. The TOF detector was operating in dead time during the acquisition of most of this peak.....	151
B.4	Arrival time distributions (ATDs) for the peak segments denoted in Figure B.3	152
C.1	UPLC chromatograms of a mixture of 35 standard degradation products obtained with different “B” solvents: methanol (top) and acetonitrile (bottom), each containing 0.01% formic acid.....	154
C.2	UPLC chromatograms of 1-, 2-, 5-, and 10- μ L injection of a 5 mg/L mixture of 35 standard degradation products.....	155
C.3	UPLC chromatograms of a 5 mg/L mixture of 35 standard degradation products obtained at column temperatures of 25, 30, and 35 $^{\circ}$ C	156
C.4	UPLC chromatograms of a 5 mg/L mixture of 35 standard degradation products with varying percentages of formic acid in the mobile phase	157
D.1	Snapshot of the MetaboLynx software interface	162
D.2	Snapshot of the MS ^e software interface	163

LIST OF TABLES

Table	Page
2.1 Experimental exact mass, calculated exact mass, and error (ppm) for representative carbohydrate and organic acid analytes in corn stover hydrolysate	29
2.2 Measured average collision cross sections for negatively charged ions of select carbohydrate standards (top) and identified carbohydrates in corn stover hydrolysate (bottom)	52
2.3 Comparison of average collision cross sections for $[M - H]^-$ ions of representative aromatic acid and aldehyde standards and $[M - H]^-$ ions of the same exact mass in corn stover and sorghum hydrolysates.....	54
3.1 UPLC gradient elution profile	61
3.2 Manually-generated HR-TOF-MS results for the 35 targeted analytes in a 5 mg/L standard solution.....	70
3.3 MetaboLynx-generated HR-TOF-MS results for the 34 targeted analytes in a 5 ppm standard solution	74
3.4 Kendrick mass defect (KMD) results for the target analytes differing by an “OH”	83
3.5 Kendrick mass defect (KMD) results for the target analytes differing by an “OCH ₃ ”	83
3.6 Kendrick mass defect (KMD) results for the target analytes differing by an “CH ₂ ”	84
3.7 Kendrick mass defect (KMD) results for the target analytes differing by an “CO ₂ ”	84
3.8 New m/z peaks in the AFEX sample identified by KMD analysis to differ by “OH”	90
3.9 New m/z peaks in the AFEX sample identified by KMD analysis to differ by “OCH ₃ ”	91
3.10 New m/z peaks in the AFEX sample identified by KMD analysis to	

	differ by “CH ₂ ”	91
3.11	New <i>m/z</i> peaks in the AFEX sample identified by KMD analysis to differ by “CO ₂ ”	92
3.12	New <i>m/z</i> peaks identified both by visual inspection and by MetaboLynx software	94-97
3.13	CID fragments for MS ^e analyses and manual MS/MS experiments for the new <i>m/z</i> peaks identified both by visual inspection and by MetaboLynx software	98-99
4.1	Compound names, formulas, monoisotopic masses, and structures for the model compounds used in this study.....	109-111
A.1	Collision cross sections (CCSs) for [M – H] [–] ions at exact masses corresponding to aromatic acid and aldehyde degradation products standards at four ion mobility wave heights	137
A.2	Collision cross sections (CCSs) for [M – H] [–] ions at exact masses corresponding to aromatic acid and aldehyde degradation products in corn stover hydrolysate at four ion mobility wave heights.....	138-139
A.3	Collision cross sections (CCSs) for [M – H] [–] ions at exact masses corresponding to aromatic acid and aldehyde degradation products in sorghum hydrolysate at four ion mobility wave heights.....	139-140
A.4	Collision cross sections (CCSs) for [M – H] [–] ions of carbohydrate standards at six ion mobility wave heights	141
A.5	Collision cross sections (CCSs) for negatively charged ions at exact masses corresponding to carbohydrates in corn stover hydrolysate at six ion mobility wave heights	142
A.6	Collision cross sections (CCSs) for model compound ions at six ion mobility wave heights.....	145-146
A.7	Collision cross sections (CCSs) for asphaltene ions at four ion mobility wave heights	147-148
C.1	UPLC gradient elution profile for time-optimized separation of 35 degradation products.....	153
D.1	Compound names, mass differences (from the C ₃₃ H ₃₈ ON ₃ S ₃ “parent” mass), and expected retention times entered into the MetaboLynx method for the “expected metabolites”.....	161

ACKNOWLEDGMENTS

I gratefully acknowledge Chris Becker for the priceless guidance that he has generously provided over the past several years. I would not be where I am today without you and have come to thoroughly enjoy scientific research because of you.

I would also like to express my thanks to Dr. Kevin Chambliss for the invaluable lessons over the past 6 years, particularly in conducting scientific research and being an effective teacher. I also appreciate the freedom you have given me in both research and teaching.

Thank you to past and present group members, particularly Lekh Sharma, Pilar Perez-Hurtado and Richard Sevcik, for your friendship. A special thanks to Zach Hardie for believing in me and providing encouraging words when I really needed them.

Thank you to my friends and family for your support and encouragement. Thank you to Thiru Munisamy, who has been my inspiration and such a source of joy in my life these past few years. Finally, thank you to my parents for your constant support. Mom, thank you for always being there for me, both throughout this Ph.D. process and always. I never would have been able to accomplish this without you.

CHAPTER ONE

Introduction to X-omics Areas and Analyses

Introduction to X-omes and X-omics: A Brief History

The suffixes “ome” and “omics” in physical sciences are widely adopted as convenient and simple ways to describe the collection of components in a complex system and the *in toto* study of those components, respectively.¹ The utilization of “ome” and “omics” in this modern scientific sense originated in 1920 when the German botanist Hans Winkler combined the words “gene” and “chromosome” and coined the term “genome” to describe the collection of DNA in an organism.²⁻⁴ The first “omics” term was not introduced until 1987 when Victor McKusick and Frank Ruddle were establishing a journal focusing on genomes and titled it “Genomics”.³ After that, other related and biologically-relevant “omics” areas were established: proteomics (study of proteins, their structures, and their functions), lipidomics (study of lipids), glycomics (study of oligo- or polysaccharides), glycoproteomics (study of glyco-proteins), etc. Since then, “omics” has extended beyond traditional biology to areas such as metabolomics (study of metabolites found in a cell or organism), petroleomics (study of the components of crude oil), and lignocellulomics (study of components in lignocellulosic biomass samples). To date, hundreds of other “omic” areas have been established.⁵

Figure 1.1 shows an illustration of the number of publications in these x-omics areas, as defined by a Google Scholar search. Bubble size indicates the total number of

articles that have been published in the specific x-omic area. Research in the areas of genomics, proteomics, and metabolomics is clearly well established. Fewer reports have been published for the less established areas of glycomics, lipidomics, glycoproteomics and petroleomics. Lignocellulomics is not present in the figure because we have only recently coined the term.

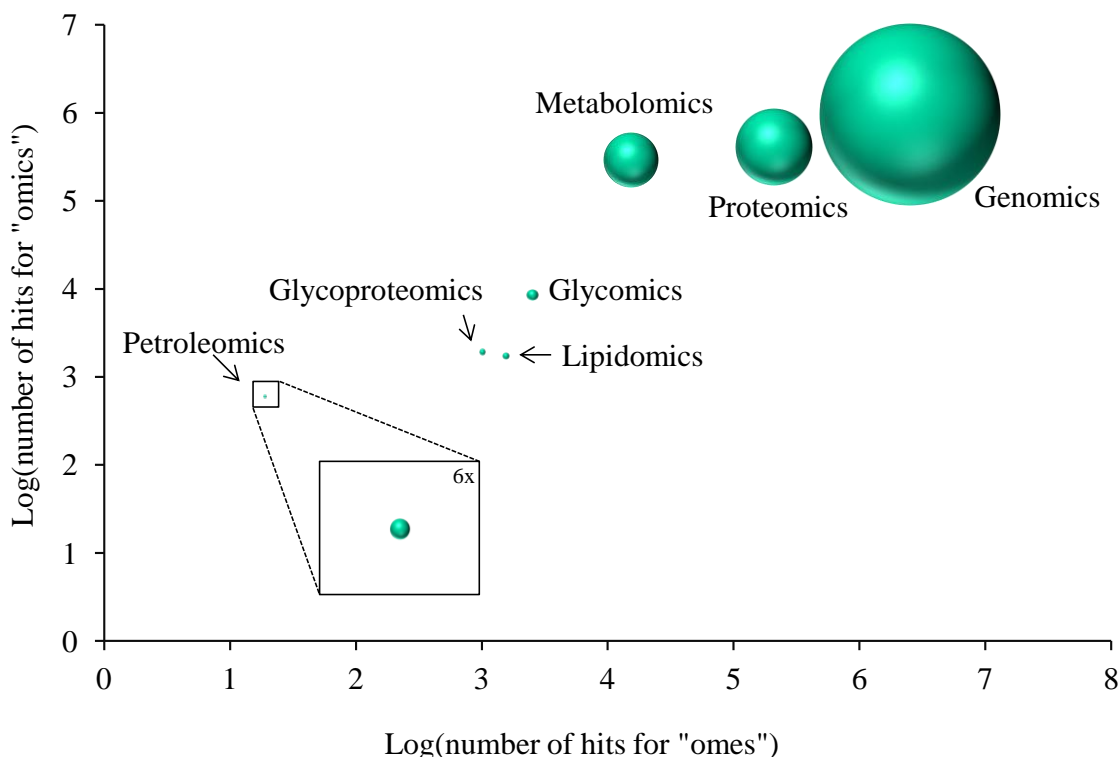


Figure 1.1. Illustration of the number of publications for common “omes” and “omics” areas. The x- and y-axes represent the number (on a log scale) of “results” or hits in a Google Scholar article search for the selected “omes” and “omics” terms, respectively. The bubble size indicates total hits for both “omes” and “omics” terms (e.g., 3,477,000 for genome/genomics). The search was performed on January 30, 2012.

Traditional and Contemporary Techniques for X-omics Analyses

The complexity and diversity of x-omics samples makes it difficult to determine or develop the best analytical approach for characterization. X-omics analyses have

made significant progress in throughput, reproducibility, and accuracy in recent years.⁶ A brief overview of traditional and more recently developed analytical approaches to x-omics analyses is presented for established x-omics areas (i.e., genomics and proteomics) and emerging x-omics areas that are relevant to the work in this dissertation (i.e., petroleomics and lignocellulomics).

Established X-omics: Genomics and Proteomics

Genomics and proteomics represent the most popular x-omics areas. From inception through January 2010, the words “genomics” and “proteomics” appeared 55,750 and 23,343 times, respectively, in scientific literature.⁷ For comparison, note that the next most common “omics” term, metabolomics, only appeared 1,686 times during the same time frame. Historically, techniques for protein and DNA analyses were time-consuming and labor-intensive.

DNA sequencing methods typically required four steps: reaction, separation, detection, and data analysis. The most popular of these methods were the Maxam-Gilbert method and the Sanger method.⁸ The Maxam-Gilbert method involves base-specific cleavage of fluorescent- or radioactive-labeled DNA fragments through four different reactions followed by separation with gel electrophoresis for each reaction and fluorescence or radioactivity detection. Sanger degradation involves combination of the single-stranded DNA fragment to be sequenced, DNA polymerase, a primer of the DNA fragment, free adenine (A), guanine (G), cytosine (C), and thymine (T) bases and fluorescent- or radioactive-labeled dideoxynucleoside triphosphates (ddNTPs) of the four bases. DNA polymerase extends the primer until a ddNTP is incorporated into the sequence and the reaction terminates. Reaction mixtures for each of the four bases are

separated via gel electrophoresis with one lane for each of the A-, G-, C- and T-terminated fragments. The gel is read from bottom to top (5' to 3'), where a band at a given position indicates the identity and position of that base in the sequence. When Sanger sequencing was first introduced, gels were read manually but now that part of the method has been automated.⁹ In addition, capillary electrophoresis (CE) with automated injection can be used instead of slab gel electrophoresis, which removes another step previously requiring laborious efforts of a researcher. The sequence output that can be obtained with CE is 100 times greater than the output of a single researcher.⁹ This high throughput capability of CE enabled sequencing of the human genome well in advance of the time frame anticipated at inception of the Human Genome Project.⁸

Protein sequencing was most commonly achieved via Edman degradation, which involves removal of the N-terminal amino group of a protein by reaction with phenylisothiocyanate and subsequent detection and identification of the removed amino acid.¹⁰ These steps are repeated for each amino acid in the protein. Each cycle takes approximately 45 minutes, so this method can become very time-consuming for larger peptides and proteins. Automation of these processes has reduced the labor-intensive aspect of this technique but the process is still time-consuming.

Beyond sequencing, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), which separates proteins based on isoelectric point and molecular weight, has been the most widely used technique for protein analysis.¹¹ This technique on its own was limited in its reproducibility and sensitivity; only the most abundant proteins could be detected unless a fractionation step was performed prior to 2D-PAGE analysis. In addition, posttranslational modifications (e.g., phosphorylation or glycosylation) can

make analysis more difficult because modified proteins can appear at multiple positions in the gel.¹¹

With the invention of soft ionization sources (i.e., electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI)), mass spectrometry has become a powerful tool in protein analyses.¹² Protein identification, including sequencing, can be achieved with mass spectrometry by dissolving individual spots resulting from 2D-PAGE, enzymatically digesting them into peptides, and analyzing the digests with mass spectrometry.¹¹ Alternatively, liquid chromatography can be used in place of 2D-PAGE by digesting a mixture of proteins and separating the digest with liquid chromatography prior to mass spectrometric analysis.¹³ Multidimensional liquid chromatography has proven effective for these types of analyses. In this approach, the protein digest is typically applied to a cation-exchange column and a fraction of peptides eluting from this column are then subjected to separation on a reversed-phase column before being introduced into the mass spectrometer. Subsequent fragmentation of peptides with tandem mass spectrometry (MS/MS), combined with software designed to match observed fragments with those of known peptides and proteins in databases, allows for protein identification through a bottom-up approach. Top-down proteomics, an alternative approach, involves direct infusion (no separation) of an intact protein into a high-resolution mass spectrometer (e.g., Orbitrap or Fourier transform ion cyclotron resonance mass spectrometer) which can provide identification based on the observed exact mass or induced fragmentation in the mass spectrometer.¹¹ Posttranslational modifications of known proteins can be easily detected with mass spectrometry by looking for expected fragment ions that have an additional mass corresponding to the

modification. For example, acetylation is indicated by a 42.0106-Da addition in mass. Alternatively, neutral loss of the modification (e.g., CH_2CO for acetylated peptides) during fragmentation of modified peptides can help identify posttranslational modifications. Despite the simplicity of these approaches, identification of some posttranslational modifications can be problematic (e.g., low ionization efficiency for phosphorylated compounds). Sequencing of an unknown protein is especially difficult when posttranslational modifications are present because a mass difference between adjacent peptides or fragments does not only correspond to a single amino acid but also to the modification. However, in most current sequencing software, a researcher can specify posttranslational modifications to include in the search.

Emerging X-omics: Petroleomics and Lignocellulomics

Applicable to the work presented in this dissertation are two emerging areas of x-omics research: petroleomics and lignocellulomics. A majority of petroleomics research has focused on low-molecular-weight samples (e.g., light or sweet crude oils) and thus, their composition is fairly well understood.¹⁴ An area that is lacking is detailed knowledge of composition for high-molecular-weight samples (e.g., heavy crude oils). Analogous to genomics, researchers are interested in determining the chemical constituents of a crude oil or its “petroleome”. Research in this area was traditionally limited to analyzing bulk properties using techniques such as UV-visible, infrared or nuclear magnetic resonance spectroscopy and vapor pressure osmometry.^{15,16} Gas chromatography (GC), often coupled with mass spectrometry (MS), was also widely utilized for characterization of petroleomic samples.¹⁷ However, because GC-MS is

limited to volatile or semi-volatile components, heavy crude oils were not extensively characterized.

The field of petroleomics has been accelerated with the development of high-resolution mass spectrometry, particularly Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). These mass spectrometers, with ultrahigh resolving power ($\geq 400,000$ up to 800 Da for some instruments) and excellent mass accuracy (sub-ppm error), allow for a greater number of components to be detected and identified.¹⁶ Ultrahigh-resolution mass spectrometers are important for separation of isobars, compounds with the same nominal mass but different exact masses. Current FTICR mass spectrometers are capable of separating compounds that only differ by 3.4 mDa (e.g., C₃ vs. SH₄). However, even with ultrahigh resolving power and mass accuracy, exact molecular formulas cannot always be assigned, which is more often a problem at higher m/z values. Data visualization plots, such as Kendrick mass defect (which allow for easy identification of homologous series), double bond equivalence (number of double bonds plus rings) vs. carbon number, and van Krevelen (H/C ratio vs. O/C ratio) plots, provide additional information about sample composition, even when exact molecular assignments cannot be made.¹⁶

Analyses in lignocellulomics have primarily centered on quantitation of carbohydrates or phenolic and organic acid degradation products resulting from pretreatment processes. Determination of carbohydrates in biomass samples has been traditionally carried out using HPLC with pulsed amperometric (PAD)¹⁸⁻²⁰ or refractive index (RI)²¹ detection. Primary techniques for the analysis of degradation products

include GC-MS²¹⁻²³ and HPLC with UV²⁴, conductivity²⁵ or mass spectrometric^{26,27} detection.

Current lignocellulomic research is focused on improving feedstocks, pretreatment processes, and fermentation conditions for greater ethanol yield. Pretreatment degradation products, which can be inhibitory to fermentation, are routinely monitored in studies that are aimed at improving pretreatment processes. If fewer inhibitory compounds are produced by a given pretreatment method, a greater volume of ethanol should be produced. Thus, to evaluate a new pretreatment method for improvement over an existing pretreatment method, the concentrations of degradation products are determined. An HPLC-UV-MS/MS method, the current state-of-the-art technique for quantitation of degradation products, only monitors 40 compounds.²⁶ Pretreated biomass, referred to as “hydrolysate”, is composed of many more than 40 compounds, as evidenced by mass spectrometric analysis of whole hydrolysates (Figure 1.2). The compounds produced during pretreatment vary depending on the feedstock and pretreatment type utilized,²⁸ so development of a routine method for identification of unknown degradation products is an important step in understanding hydrolysate composition and optimizing ethanol production.

Unknown degradation products have primarily been identified with GC-MS. Several GC-MS methods have been published for the identification of unknown components in biomass hydrolysate, including: (1) lignin-derived products and other phenolic compounds in dilute-sulfuric-acid-pretreated poplar, corn stover, and switchgrass,²³ (2) aromatic, aliphatic and furan acids in a dilute-nitric-acid-pretreated poplar hydrolysate,²¹ (3) aliphatic acids, furan derivatives, phenolic compounds, ketones

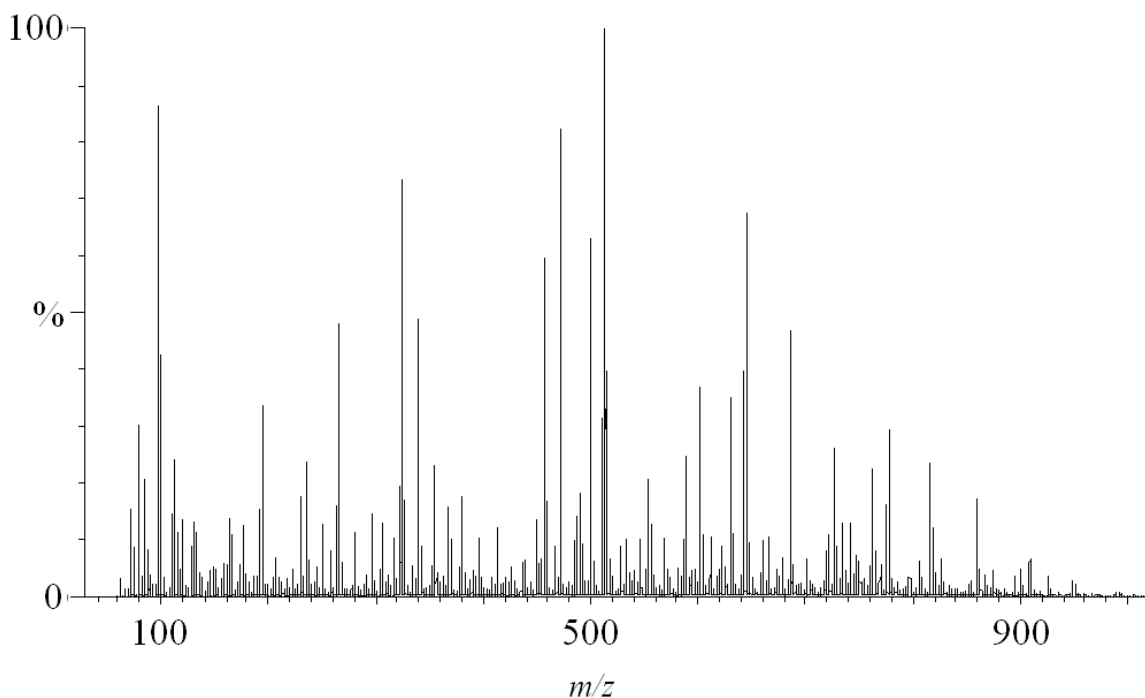


Figure 1.2. Mass spectrum of a representative dilute-acid-pretreated hydrolysate of corn stover directly infused into an ESI-TOF mass spectrometer. Note that nearly all of the m/z peaks in the figure represent singly charged components.

and a pyridine derivative in dilute-sulfuric-acid-pretreated corn stover hydrolysate,²⁹ and (4) pyrazine and imidazole derivatives in ammonia-fiber-expansion- (AFEX) pretreated corn stover.²⁷ All of these methods rely on matching observed mass spectra for unknown components with those of known compounds in the literature or a database (e.g., NIST Mass Spectral Library or National Bureau of Standards Mass Spectral Database). In addition, GC analyses of unknown degradation products are limited to the more volatile components of biomass hydrolysates or require derivatization, which adds a degree of difficulty to the analysis. Fewer LC-MS methods for identification of unknown degradation products have been reported. As a small component of a broader study, Chundawat et al.,²⁷ used LC-HRMS to identify nitrogenous cell wall decomposition products. LC-MS typically operates with electrospray ionization (ESI), so there are no

universal libraries to which one can compare mass spectra of unknown compounds, as there are for electron impact (EI)-generated spectra in GC-MS. Exact masses of unknown peaks and analysis of fragmentation patterns from MS/MS experiments must be used to identify unknown components. Although several methods have been reported,^{21,23,27,29} there is currently no widely used method for identification of unknown degradation products in biomass hydrolysates.

Common Limitations in X-omics Analyses

Analysis of x-omics samples presents a significant challenge to analytical chemists. These samples are typically very complex, comprised of hundreds to tens of thousands of components that can vary substantially in chemical nature and abundance (e.g., proteins can vary in concentration over 10 orders of magnitude).^{4,7} It is difficult to find a preparation method and analysis technique suitable for analyzing a group of highly diverse compound classes with concentration differences spanning orders of magnitude in a single analysis. In addition, small available sample volumes, which are often the case for x-omics samples, limit the types of analyses researchers can perform. Traditional techniques suffer from long analysis times, extensive sample preparation, which often introduces contaminants into the sample,³⁰ and limited resolving power, which requires a greater degree of separation or fractionation prior to detection. Contemporary techniques (e.g., ultrahigh-resolution mass spectrometry) offer many improvements in x-omics analyses, such as small sample volume requirements and higher resolving power. However, even these new and improved techniques suffer from limitations. For example, tandem mass spectrometry in most data-dependent analyses is limited to a mass selection window of 1 Da for the parent ion. The resulting fragmentation spectra can be from

multiple parent ions if there are isobars in the sample, which is common for the complex mixtures of x-omics. Continued improvement in x-omics analyses is dependent on development and application of information-rich tools that have high resolving power, require minimal sample preparation, and probe multiple characteristics (e.g., molecular weight, structure, charge, hydrophobicity, etc.) of analytes.

UPLC- and IM-HRMS as Emerging Tools for X-omics Analyses

Ultra performance liquid chromatography – and ion mobility – high-resolution mass spectrometry (UPLC- and IM-HRMS) present attractive techniques for x-omics analyses. Compared to traditional methods of x-omics analyses, these techniques offer higher chromatographic efficiency with UPLC, which can equate to faster analysis times and greater separation, an additional degree of separation with ion mobility spectrometry (IMS), and high resolving power and mass accuracy with HRMS, which allows for identification of known and unknown analytes.

Ultra Performance Liquid Chromatography

Because UPLC instruments have been outfitted to operate at higher pressures relative to HPLC instruments, columns are designed for increased efficiency and decreased run time, with smaller particle sizes and shorter columns. Long HPLC analyses can be scaled down to produce shorter analysis times, as well as decreased eluent consumption and smaller injection volumes (< 10 μ L). Methods utilizing UPLC have been reported in many x-omics areas, including genomics (e.g., determination of the degree of DNA methylation in extracted human blood samples);³¹ proteomics (e.g., quantitative analysis of hydrolyzed peptides);³² glycomics (e.g., detection and

quantitation of anionic polysaccharides in less than 5 minutes);³³ lipidomics (e.g. determination of various types of lipids in rat plasma);³⁴ and glycoproteomics (e.g., identification and comparison of glycosylated and nonglycosylated peptides in serum samples from a healthy individual and an individual with ovarian cancer),³⁵ to name a few. Of all of the x-omics areas, UPLC has been most widely utilized for metabolomics. More than 150 UPLC metabolomics studies have been published, which often overlap into proteomics or genomics.^{36,37} Examples of UPLC metabolomics studies include methods for quantitation of human serum metabolites,³⁸ detection of dimethoxycinnamic acid derivatives in human plasma after coffee consumption,³⁹ and metabolite fingerprinting of wounded plants.⁴⁰

Lignocellulomic UPLC studies have not been extensively reported. Ross and co-workers reported two separate UPLC methods for monitoring the metabolism of 10 organic acids⁴¹ and glucose⁴² during microbial fermentation. Although only two methods have been reported for biomass samples, analytes commonly found in biomass (e.g., organic acids) have been monitored with UPLC in other samples such as beverages,⁴³⁻⁴⁵ tobacco plants,⁴⁶ and red mustard greens.⁴⁷

Ion Mobility Spectrometry

Ion mobility spectrometry separates ions based on size, shape, and charge. In traditional ion mobility, a direct current voltage is applied across an ion mobility cell (typically, 5-25 cm) and a drift gas of nitrogen, helium or argon is introduced into the cell to give a pressure in the range of ~1 torr to atmospheric pressure, depending on the instrumental setup.⁴⁸ Smaller, compact ions will have fewer collisions with the drift gas and will be accelerated through the ion mobility cell at a higher velocity than larger, less

compact molecules which will have more collisions with the drift gas. An ion's average ion-neutral collision cross section, which is considered as the surface area of the interaction between the ion and the neutral drift gas molecule, can be directly correlated to its reduced mobility through the drift cell (K_0):

$$\Omega = \frac{(18\pi)^{1/2}}{16} \frac{ze}{(k_b T)^{1/2}} \left[\frac{1}{m_1} + \frac{1}{m_2} \right]^{1/2} \frac{1}{K_0} \frac{1}{N} \quad (1.1)$$

where Ω is the ion-neutral collision cross section (m^2 or more commonly, \AA^2), z is the ion's charge, e is elementary charge (C), k_b is Boltzmann's constant (J/K), T is temperature (K), m_1 is the mass of the ion (kg), m_2 is the mass of the drift gas (kg), K_0 is an ion's reduced mobility ($\text{m}^2/\text{V-s}$) and N is the drift gas number density at standard conditions (m^{-3}). The reduced mobility of an ion can be related to its drift time:

$$K_0 = \frac{L}{t_D E} \frac{P}{(760 \text{ torr})} \frac{(273.2 \text{ K})}{T} \quad (1.2)$$

where t_D is the drift time of the ion (s), E is electric field strength (V/m), L is length of the drift cell (m), P is pressure in the drift cell (torr), and T is temperature (K). K_0 is normalized to standard conditions of 760 torr and 273.2 K. Drift times are dependent on the instrument and parameters utilized but reduced mobilities and collision cross sections (at a given temperature) are constant for a given compound. Thus, reduced mobilities and collision cross sections can be compared for analogous ions across different samples and instruments.

A more recent approach to IMS separations uses a traveling wave voltage profile rather than static DC voltage across the IM cell. Traveling wave ion mobility drift cells

are comprised of a series of ring electrodes that have two applied voltages: a traveling wave voltage that propels ions through the mobility cell superimposed on a radiofrequency voltage that radially confines ions.⁴⁹ Just as in traditional ion mobility, the time it takes for an ion to traverse the cell (drift time) is dependent on its size and shape. A relatively small or compact ion (green ion in Figure 1.3) rides the traveling wave and is propelled through the cell at the same velocity as the wave. Conversely, a larger or less compact ion (purple ion in Figure 1.3) has more collisions with the drift gas, falls back over the top of the wave into a potential valley and must wait for the next wave to carry it through the cell, resulting in a longer drift time. Thus, two species that differ in their relative size and shape can be separated with ion mobility, even if they have the same mass. Charge also dictates how fast an ion crosses the ion mobility cell. The higher the charge, the faster the ion is accelerated by the DC traveling wave voltage. Thus, two ions of similar size and shape but different charge can also be separated in the mobility cell.

At its most basic level, ion mobility instrumentation is comprised of an ionization source, an ion mobility cell, and a detector. Commonly, mass spectrometry is incorporated after

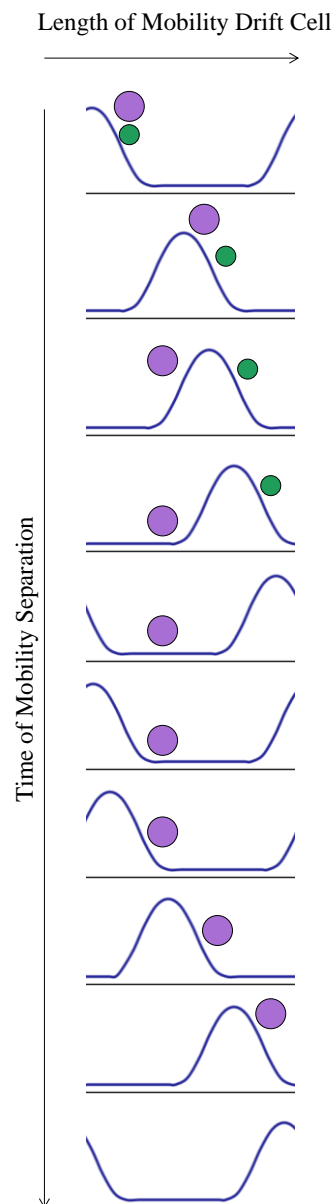


Figure 1.3. Example of the separation of two ions during a portion of a traveling wave ion mobility separation. Time of the separation proceeds down the page.

the ion mobility cell for mass identification of species exiting the mobility cell, adding additional separation. IM-MS is particularly advantageous for complex x-omics samples because of its ability to distinguish different compound classes. Figure 1.4 shows IM-MS spectra (m/z vs. ion mobility drift time in milliseconds) of a mixture of peptides, mono- and oligo-saccharides and sodium formate in a methanol–water solvent system. Mass-mobility correlations, called trendlines, can be seen for each compound class. Although the trendlines are not completely separated, this figure illustrates the capability of IM-MS in distinguishing between different compound classes. The lower trendlines represent singly charged ions, whereas the upper trendlines contain ions of higher charge states (+2, +3, etc). Higher-charge-state trendlines appear above singly charged trendlines in IM-MS spectra because for a given m/z , ions with more charge will move faster through the drift cell and thus, will have a shorter drift time. Deviation of a peak from a trendline might indicate that the compound associated with that peak is of a different compound class.

IM-MS has been applied as an effective tool for x-omics sample analyses, especially in the areas of proteomics, metabolomics, petroleomics, and glycomics. Ion mobility spectrometry has been shown to separate various biological compounds (e.g., peptides, DNA, lipids, glycoconjugates, etc.), protein conformations (e.g., α -helices, sheets, random coils) and different classes of metabolites (e.g., amino acids, carbohydrates, fatty acids, purines, etc.).⁵⁰⁻⁵² In the area of petroleomics, IM-MS has been utilized to differentiate between cyclic and planar structures in samples such as crude oil, asphaltenes, and desasphalted oils.⁵³⁻⁵⁶ Glycomic research with IM-MS has mainly focused on separation of positional and structural carbohydrate isomers.⁵⁷⁻⁶⁰ However,

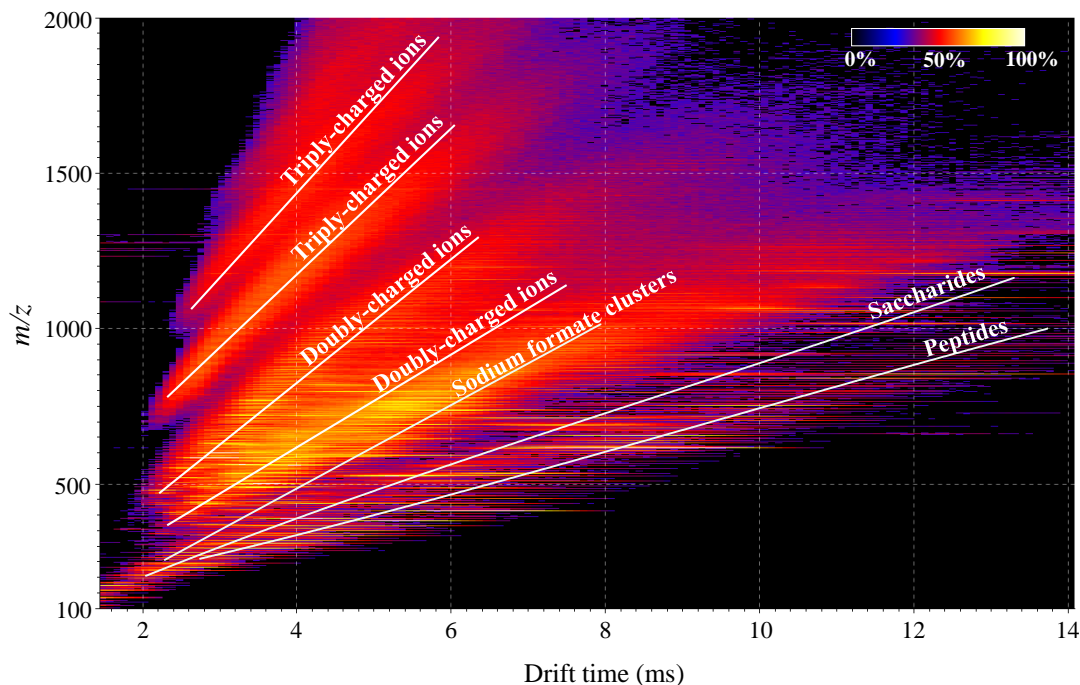


Figure 1.4. IM-MS spectra (m/z vs. drift time) of a mixture of compounds (peptides, mono- and oligo-saccharides, and sodium formate in a methanol–water solvent system). Labeled trendlines are indicated by white lines that represent linear regression of IM-MS data for each compound class. Color denotes ion signal intensity, with white representing the most intense ions followed by yellow, orange, red, purple, blue (typically noise) and black (no ion signal).

IM-MS has also been used to significantly reduce mass spectra complexity, allowing for easy detection of singly-, doubly-, and triply-charged glycoconjugates in human urine.⁶¹

High-Resolution Time-of-Flight – Mass Spectrometry

In time-of-flight – mass spectrometry (TOF-MS), ions of different mass-to-charge ratios (m/z) are separated by their flight time. Ions are accelerated into a flight tube by applying voltages on pusher and puller plates, separated in the flight tube and detected (e.g., via a multichannel plate detector) at the end of the flight tube. Ions have the same or very similar kinetic energy (E) when they are introduced into the flight tube and so

according to the basic kinetic energy equation (equation 1.3), ions with different masses (m) will have different velocities (v).

$$E = \frac{1}{2}mv^2 \quad (1.3)$$

Mass can be directly correlated to flight time (t) by substituting flight distance (d) divided by flight time for velocity and rearranging.

$$m = \left(\frac{2E}{d^2}\right)t^2 \quad (1.4)$$

This equation indicates that smaller masses will have a faster flight time and reach the detector earlier than larger masses. Directly related to flight time is the mass spectrometer's ability to resolve two ions with similar flight times. How well a mass spectrometer can separate these ions is dictated by its mass resolving power, which is defined in equation 1.5 for an ion with mass, m , and mass peak width, Δm .⁶²

$$\text{Resolving power} = \frac{m}{\Delta m} \quad (1.5)$$

Note that resolution, defined as $\Delta m/m$, is not a concept commonly utilized in mass spectrometry. Flight tube length, how fast ion detections are recorded, and kinetic energy distribution caused by varied ion position between the pusher and puller plates all have an effect on the mass resolving power. Resolving power can be maximized by ensuring fast data acquisition times, by increasing the length of the flight tube or by placing a reflectron (a type of ion mirror) at the end of the flight tube to help correct for kinetic energy distribution and also allow for longer flight paths without the need for longer flight tubes. The first commercial TOF-MS instrument (available in the 1950s) had a

resolving power up to 300.⁶² Current commercial TOF-MS instruments are capable of resolving powers up to 40,000.⁶³

These high-resolution time-of flight mass analyzers are capable of providing exact masses of ions with relatively high mass accuracy (< 1 ppm for some instruments).⁶² For known analytes, exact masses offer improved confidence in identification; for unknown ions, chemical formulas can be obtained from exact masses, significantly reducing the number of possible chemical formulas suggested by a nominal mass. For example, at nominal mass 163, there are 15 possible formulas for a molecule containing only C, H, and O atoms. For higher masses, there are even more possibilities (e.g., there are 64 possibilities for a molecule at mass 365). A high-resolution mass spectrometer can narrow down the possibilities to a few or sometimes single possibility, depending on the mass accuracy of the method. For example, 4-hydroxycoumaric acid, at m/z 163.0395 for the $[M - H]^-$ ion, has only one possible molecular formula within 50 ppm of its theoretical exact mass.

In addition, because they offer fast analysis times, time-of flight mass analyzers are well suited to couple with ion mobility. While an ion mobility separation occurs on a millisecond time frame, TOF mass spectra are acquired in microseconds so hundreds of TOF spectra can be collected for each mobility spectrum.⁶⁴ TOF mass analyzers can also be paired with a quadrupole mass analyzer, which can be used as a mass filter. MS/MS experiments can be performed by selecting a parent ion in the quadrupole, subjecting the parent ion to fragmentation (e.g., through collision-induced dissociation, CID), and then detecting the resulting daughter ions with the TOF mass analyzer.

The mass spectrometer used for research presented in this dissertation, the Waters Synapt HDMS (Figure 1.5; shown along with the UPLC instrument utilized), has been designed to incorporate all of these features: a quadrupole mass filter, a Tri-Wave traveling wave ion mobility cell with trap and transfer cells before and after the ion mobility cell, respectively, where CID can be performed, and a time-of-flight mass analyzer.

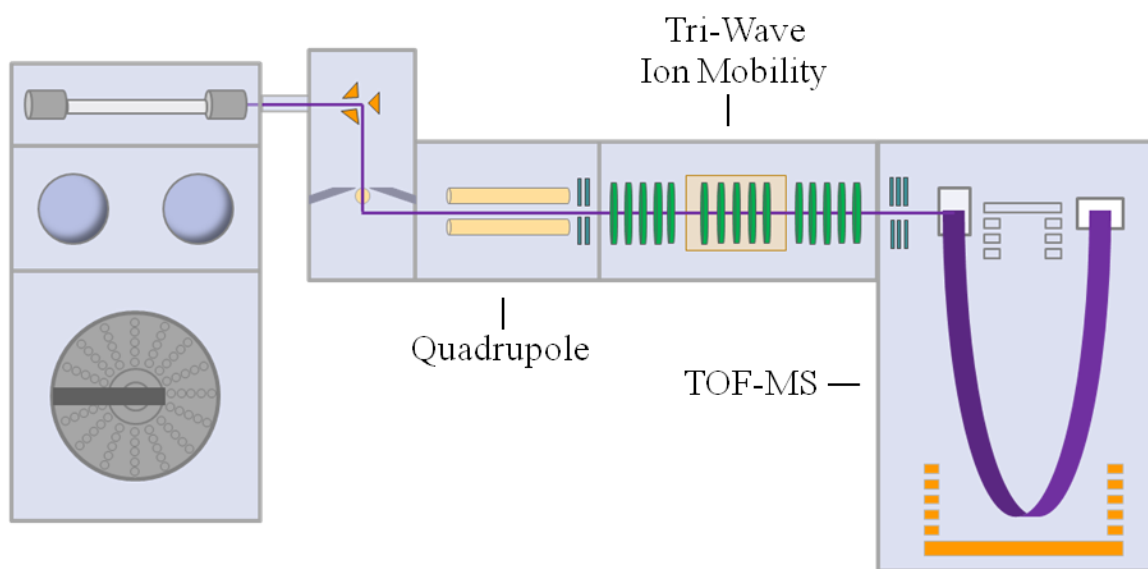


Figure 1.5. Waters Acquity Liquid Chromatograph and Synapt HDMS Mass Spectrometer.

Scope of the Dissertation

This dissertation presents novel applications of ultra performance liquid chromatography – high-resolution mass spectrometry (UPLC-HRMS) and ion mobility – high-resolution mass spectrometry (IM-HRMS) in two emerging x-omics areas: lignocellulomics and petroleomics. Previous work in our research lab focused on developing a HPLC-MS/MS method for quantitation of 37 targeted organic acid and

aldehyde degradation products in biomass hydrolysates. Chapters 2 and 3 of this dissertation detail the development of two novel methods for analysis of hydrolysates, specifically, development of (i) a IM-HRMS technique for characterization of unextracted biomass samples and (ii) a UPLC-HRMS and “metabolomics” method for identification of unknown components in biomass hydrolysates.

Previous asphaltene work demonstrated that IM-MS can be used to (i) determine the molecular weight distribution of monomeric asphaltenes and (ii) separate monomeric from dimeric asphaltenes. Preliminary studies comparing arrival times of asphaltenes to arrival times of fullerenes and linear polymers also indicated that IM-MS may be useful in revealing structural information for these sample types. Building on this previous work, Chapter 4 of this dissertation illustrates the utility of IM-MS for structural analyses in asphaltene samples and further develops understanding of asphaltene structure using model compounds. Specifically, IM-MS data of model compounds believed to be representative of “asphaltene structures” were employed to evaluate structure type and diversity in an asphaltene sample.

CHAPTER TWO

Direct Infusion Electrospray Ionization – Ion Mobility – High-Resolution Mass Spectrometry (DIESI-IM-HRMS) for Rapid Characterization of Potential Bioprocess Streams

Introduction

Understanding the correlation between sample composition and fuel production remains one of the greatest challenges facing the field of bio-based energy. Accordingly, detailed characterization of potential biofuel process streams is a critical step in optimizing biomass-to-fuel conversions, which are required for viable energy yields.⁶⁵ Most current biomass analyses target bulk components⁶⁶⁻⁶⁹ or screen only a small fraction^{19,21-23,26,27,70,71} of the total molecular composition (e.g., concentrations of carbohydrates and a few organic acid degradation products generated during bioethanol production). Moreover, these techniques can involve time-consuming preparation steps, extended analysis times, and multiple analyses for different compound classes.^{19,21-23,27,70-72} Currently, there are relatively few examples of analyses applied to “whole-sample” biomass hydrolysate. Fluorescence and near-infrared spectroscopic techniques coupled with multivariate statistics^{66,67} and ¹³C-NMR spectroscopy⁶⁸ have both been employed for rapid determination of bulk properties of whole-sample biomass hydrolysate. However, such techniques are best suited for providing scoping and/or macro-component (i.e., bulk) information on biomass samples. One approach targeting a whole-sample molecular-level understanding of biomass hydrolysate composition was recently reported by Helm and coworkers,⁶⁹ in which flow injection electrospray mass spectrometry (FIE-MS) was

employed to identify several organic acids and carbohydrates in a dilute-acid pretreated biofeedstock. Although FIE-MS permits monitoring of multiple compound classes in a single analysis, collection of only nominal mass data in the absence of a separation technique can lead to uncertainty in analyte identification and loss of relevant signal in the presence of impurities (e.g., non-native sample components such as surfactants) that are inherent in sample generation.

Ion mobility – mass spectrometry (IM-MS), a technique involving separation of gas phase ions by size, shape and m/z , offers an additional degree of separation over mass spectrometry alone. IM-MS has been applied as an effective tool for complex sample analysis in several fields, as described in Chapter 1. IM-MS has also been utilized in small molecule analyses, including separation of pharmaceutical drug formulations,^{73,74} chiral amino acid enantiomers,⁷⁵ and alkaloid stereoisomers;⁷⁶ trace analysis of drugs;⁷⁷ and screening for chemical warfare agents.⁷⁸⁻⁸⁰

Despite the wide application of IM-MS to many complex matrices, its potential for use in analysis of samples derived from biomass feedstocks has not been investigated. A few analytes common to many biomass samples have been independently investigated with IM-MS (often using commercially available standards), including acetic acid⁵² and several carbohydrates^{52,57-60,81,82} such as 5- and 6-carbon mono/oligosaccharides. However, in almost all of these studies, only positive ions have been analyzed, which is contrary to established mass spectrometry protocols for biomass analysis.^{69,72} Due to the diversity of structural components in biomass and ion mobility's potential for complex sample separation, IM-HRMS is an excellent candidate for whole-sample analysis of biomass hydrolysates and extracts.

Herein, we present direct infusion electrospray ionization – ion mobility – high-resolution mass spectrometry (DIESI-IM-HRMS) as a novel approach to whole sample biomass analysis. DIESI-IM-HRMS is applied towards the identification and structural confirmation of multiple components in 10-fold diluted biomass samples. Our results indicate that this technique has the potential to separate and identify carbohydrates and other sample components based on a combination of collision cross section (CCS) and HRMS analysis. IM-HRMS analysis also provided facile separation of analytes resulting from biomass and those introduced during pretreatment and sample preparation. We present for the first time collision cross sections for negatively charged carbohydrates and other analytes occurring natively in biomass samples. IM-MS correlations for structural carbohydrate oligomers in a bio-matrix have also been established to aid rapid identification of potential carbohydrates in a sample. While other techniques involve labor-intensive sample clean-up steps (e.g., multiple liquid and/or solid phase extractions)⁷⁰ or time-consuming chromatographic separation prior to MS analysis,^{22,70,71} DIESI-IM-HRMS is demonstrated to offer a high-throughput approach for rapid screening of whole-sample hydrolysates.

Experimental: Materials and Methods

Chemicals and Reagents

Xylobiose and xylotetraose were purchased from Megazyme International Ireland (Wicklow, Ireland). All other chemicals were reagent grade or better and purchased from standard commercial vendors (i.e., Sigma-Aldrich or Fisher Scientific). Distilled water was purified and deionized to 18.2 M Ω with a Barnstead Nanopure Diamond UV water

purification system. Dried sorghum feedstock was received from Dr. William L. Rooney, Department of Soil and Crop Sciences, Texas A&M University (College Station, TX). Corn stover hydrolysate (dilute-acid pretreated) was obtained from the National Renewable Energy Laboratory (NREL, Golden, CO).

Preparation of Standards and Generation of Aqueous Extract and Hydrolysate Samples

Individual stock solutions of carbohydrates and organic acids were prepared at 100 ppm by dissolving the neat chemical in water. From the stock solution, a 15 ppm solution of each standard was prepared by serial dilution; these solutions were used for all analyses of standard compounds. A minimal volume of methanol was used in aiding dissolution of certain organic acids in stock solutions.

An aqueous extract of sorghum was prepared via accelerated solvent extraction (ASE-200, Dionex Corp., Sunnyvale, CA) as previously reported.¹⁹ Biomass hydrolysates were generated by pretreating sorghum with 0.7% (w/w) H₂SO₄ using an ASE pretreatment protocol.⁸³ Briefly, operation conditions were as follows: 1 g of biomass per 11-mL cell, temperature: 195 °C, pressure: 1500 psi, preheat time: 9 min, heat time: 9 min, static time: 24 min, purge time: 120 s, flush volume: 150% of cell, and cycles: 1. Neutralized hydrolysate samples were prepared by adding calcium hydroxide until the pH reached 7.0. All samples were diluted 1:10 with purified water and filtered using a 25-mm syringe filter with a 0.2 µm nylon membrane (Pall Corp.) before analysis.

Extracted hydrolysates were prepared by adding 2 mL methyl-*tert*-butyl ether (MTBE) to 2 mL undiluted sorghum hydrolysate in a centrifuge tube. After agitation (< 1 min), the MTBE layer was removed from the aqueous layer and subsequently its volume was reduced under a stream of nitrogen until less than 0.5 mL remained. The

MTBE-extracted fraction and the leftover aqueous fraction were each diluted 10-fold with water and filtered.

DIESI-IM-HRMS Analysis

All analyses were performed with a Synapt HDMS ion mobility – mass spectrometer (Waters, Manchester, U.K.) equipped with an electrospray ionization source operated in negative ion mode. For all analyses except those involving collision-cross-section measurements, ion mobility spectrometry (IMS) separations were performed with a traveling-wave height of 8.0 V and applied voltages of 6.0 and 4.0 V, respectively, to the trap and transfer cells. Collision-induced-dissociation (CID) experiments were conducted by increasing the voltage across the trap cell from 6.0 V to an optimized value between 20-30 V to produce abundant fragment ions across the m/z range of interest. During IMS separations, a nitrogen bath gas was introduced into the IM cell at a flow rate of 30 mL/min for most analyses providing a typical operating IM pressure of 0.68 mbar. Where noted, a lower flow rate (i.e., 25 mL/min, 0.59 mbar) was used to compare ion mobility profiles of various analytes over a narrow mass range.

The Waters TOF is capable of operating in two reflectron modes (V and W). Unless noted, exact mass measurements were performed by operating the instrument in time-of-flight W mode to improve resulting mass accuracy (mass resolving power (MRP) ~ 15,000). Other experiments were operated in V mode (MRP ~ 9,000) to enhance instrumental sensitivity. Illustration of the resolution differences between V and W mode can be seen in Appendix B, Figure B.1. Leucine enkephalin was used for instrument calibration and as a lock mass ion as described by Wolff et al.⁸⁴ A 2 ng/ μ L lockspray solution was infused at 10 μ L/min and 1 s of lockspray data was collected for

every 10 s of sample data. The $[M - H]^-$ ion for leucine enkephalin (m/z 554.2615) was used as a lock mass in post-run data processing. All theoretical masses were calculated using masses for ^{12}C , ^1H , and ^{16}O isotopes listed in an IUPAC technical report.⁸⁵ All arrival time distributions were obtained by integrating the monoisotopic peak of a given analyte in the 2D IM-MS spectra and are presented uncorrected, as obtained from software (i.e., analyte arrival times are equal to the sum of their mobility drift time and time spent in transit between the drift cell and TOF detector).

Collision-Cross-Section Measurements

Collision cross sections of analytes were measured following the protocol reported by Ruotolo and coworkers.⁸⁶ $[M - H]^-$ ions generated from raffinose, melezitose, and α -cyclodextrin standards, as well as fragment ions of m/z 221 and 323 resulting from collision-induced dissociation of α -cyclodextrin and melezitose, respectively, were used as calibrants for collision-cross-section measurements. Collision cross sections for the calibrants were taken from literature and were 199.1, 196.4, 179.1, 134.1, and 163.2 \AA^2 for raffinose, melezitose, α -cyclodextrin, m/z 221 and m/z 323 fragment ions, respectively.⁸² Accurate calibration was verified by re-measuring the collision cross sections of the standards using the generated calibration curve; deviation from reported values was less than 1.0%. Plots used for calibration are shown in Appendix A (Figures A.1-2). IM-MS data were collected for carbohydrate standards and analytes at six different ion mobility wave heights (every 0.5 V in the range 5.5-8.0 V) at an ion mobility bath gas flow rate of 25 mL/min. IM-MS data were collected for organic acid and aldehydes standards and analytes at four wave heights (5.5, 5.8, 6.0 and 6.5 V) at an ion mobility bath gas flow rate of 25 mL/min. Collision cross sections are reported

as an average over all wave heights, with the exception of (i) ions having a molecular weight $< 200\ m/z$ at larger wave heights, where field-strength-dependent effects⁸⁶ were present and (ii) ions having a molecular weight $\sim > 540\ m/z$ at smaller wave heights, where analyte distributions were outside of the ion mobility window (see Tables A.1-5 in Appendix A for all collision cross sections). At larger wave heights, smaller ions will travel with the front of the wave and thus, experience a stronger average electric field over the course of the mobility separation. This phenomenon translates into shorter observed arrival times than those that are expected and thus, data for small ions at large wave heights results in inaccurate collision cross section values. Arrival times used in collision-cross-section measurements were obtained from the monoisotopic peak of each analyte.

Results and Discussion

Analyte Identification via HRMS

Accurate mass data was initially collected for a corn stover hydrolysate. Representative data for the full mass range (m/z 100-1000) can be observed in Figure 2.1a. Preliminary analysis of these data reveal a number of masses that can be assigned to components expected to occur in biomass pretreatment hydrolysates (e.g., carbohydrates and organic acids; representative data shown in Table 2.1). In general, experimentally measured masses were found to agree with the theoretical masses of the assigned compounds with errors of 5 ppm or less (average error 3.0 ppm). For those compounds with errors less than 5 ppm, only one possible $C_xH_yO_z$ formula had a theoretical mass within 5 ppm of the experimental mass; the only exception was the 6-

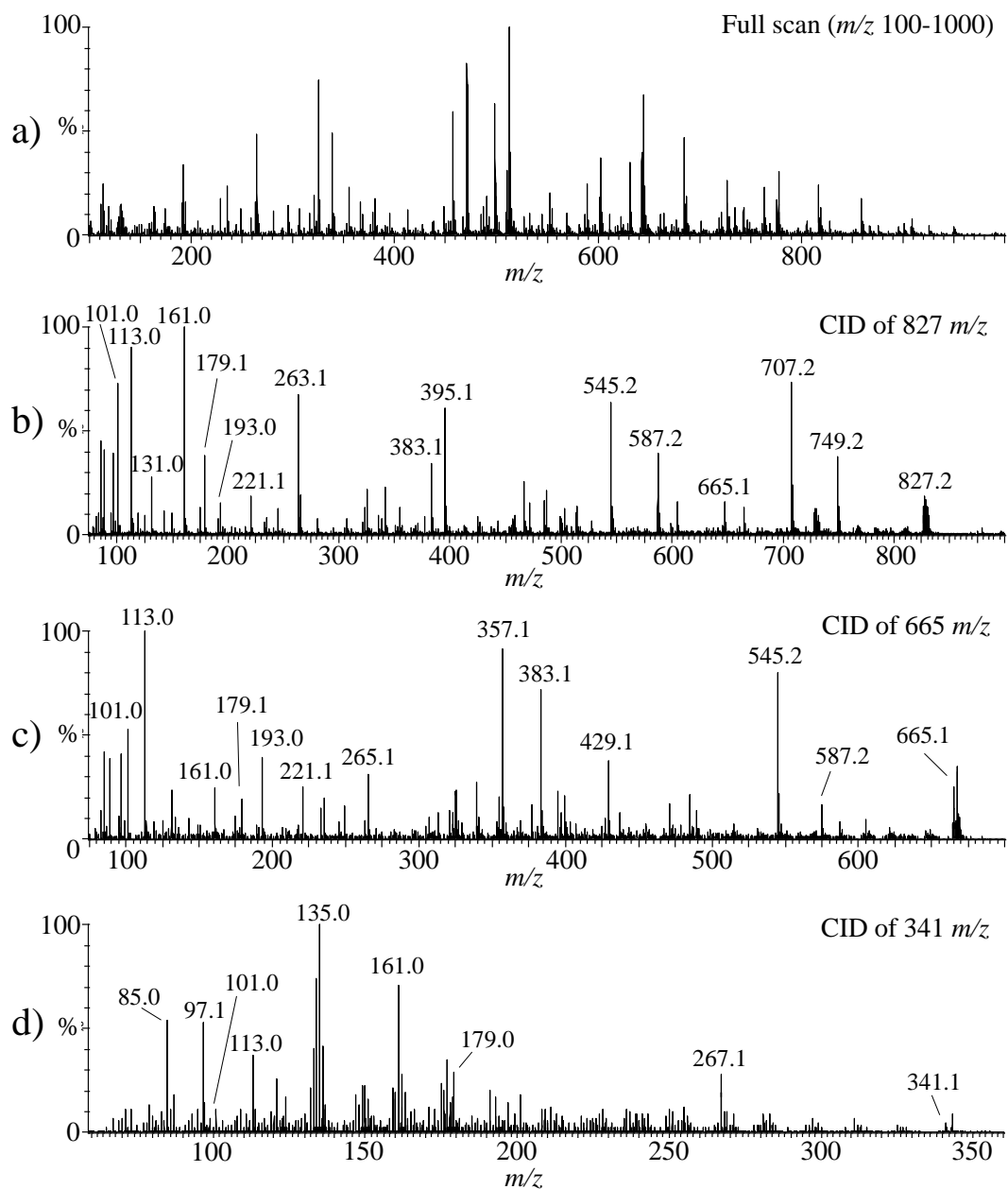


Figure 2.1. Negative ion mode mass spectra of (a) full scan (m/z 100-1000) corn stover hydrolysate and collision-induced-dissociation spectra of (b) 6-carbon trisaccharide (827.2 m/z), (c) 6-carbon tetrasaccharide (665.1 m/z), and (d) 6-carbon trisaccharide (341.1 m/z) in a corn stover hydrolysate.

Table 2.1. Experimental exact mass, calculated exact mass, and error (ppm) for representative carbohydrate and organic acid analytes in corn stover hydrolysate. Data is shown for the monoisotopic $[M - H]^-$ ion of each analyte. DP number represents degree of polymerization.

Analyte Assignment	Exp ^a mass	Calc mass	Error (ppm)
5-carbon carbohydrates			
DP1	149.0450	149.0450	0.0
DP2	281.0869	281.0873	-1.4
DP3	413.1283	413.1295	-2.9
DP4	545.1714	545.1718	-0.7
6-carbon carbohydrates			
DP1	179.0560	179.0556	2.2
DP2	341.1024	341.1084	-17.6
DP3	503.1586	503.1612	-5.2
DP4	665.2073	665.2140	-10.1
DP5	827.2668	827.2669	-0.1
organic acids			
ferulic acid	193.0500	193.0501	-0.5
caffeic acid	179.0340	179.0344	-2.2
itaconic acid	129.0182	129.0188	-4.7
glutaric acid	131.0339	131.0344	-3.8
p-coumaric acid	163.0396	163.0395	0.6
vanillic acid	167.0337	167.0344	-4.2
adipic acid	145.0499	145.0501	-1.4
homovanillic acid	181.0502	181.0501	0.6
syringic acid	197.0452	197.0450	1.0

^a All experimental masses are reported for data collected in TOF W mode, with the exception of data for the 6-carbon DP5 carbohydrate, which was collected in V mode for increased sensitivity.

carbon DP5 carbohydrate at m/z 827, for which there were three possible formulas. Careful inspection of HRMS data reveals multiple peaks at a given nominal mass (see Figure B.1 in Appendix B for examples), suggesting that nominal-mass analysis in the absence of a front-end separation would be inadequate to fully characterize the ionized species present in the mass spectrometer. Small organic species were identified with less than 5.0 ppm error. Experimental exact masses corresponding to 6-carbon and 5-carbon monosaccharides (e.g., glucose and xylose), disaccharides, and trisaccharides can be assigned with less than 6 ppm error (Table 2.1), with the exception of the 6-carbon disaccharide (m/z 341) and tetrasaccharide (m/z 665). Mass error for these peaks were greater (-17.6 and -10.1 ppm, respectively) indicating that either another species is present at a similar m/z , disrupting the exact mass measurements, or incorrect assignment of the ion signal was made. Asymmetric MS peaks for both of these analytes (data not shown) suggest that another ion is present which differs in mass by less than 0.03 and 0.04 Da for the disaccharide and tetrasaccharide, respectively.

To confirm accurate mass assignments made via HRMS data, collision-induced dissociation (CID) of the assigned analytes (Table 2.1) was performed. These data revealed fragmentation spectra of the selected carbohydrate and organic acid ions that were consistent with fragmentation spectra observed during previous carbohydrate and organic acid CID analyses,⁸⁷⁻⁸⁹ supporting the assignments of 5- and 6-carbon oligosaccharides and organic acids shown in Table 2.1. Representative CID data from these analyses can be seen in Figure 2.1b and 2.1c for 6-carbon oligosaccharides having nominal masses of m/z 827 and 665, respectively. CID spectra for m/z 827 (Figure 2.1b) shows dominant fragment ions of m/z 749.2 $[M - H - 2CH_2O - H_2O]^-$, 707.2 $[M - H -$

$\text{C}_4\text{H}_8\text{O}_4]^-$, 665.1 $[\text{M} - \text{H} - \text{C}_6\text{H}_{10}\text{O}_5]^-$, 587.2 $[\text{M} - \text{H} - \text{C}_6\text{H}_{10}\text{O}_5 - 2\text{CH}_2\text{O} - \text{H}_2\text{O}]^-$, 545.2 $[\text{M} - \text{H} - \text{C}_6\text{H}_{10}\text{O}_5 - \text{C}_4\text{H}_8\text{O}_4]^-$, 395.1 $[\text{M} - \text{H} - 2\text{C}_6\text{H}_{10}\text{O}_5 - 3\text{CH}_2\text{O} - \text{H}_2\text{O}]^-$, 383.1 $[\text{M} - \text{H} - 2\text{C}_6\text{H}_{10}\text{O}_5 - \text{C}_4\text{H}_8\text{O}_4]^-$, 263.1 $[\text{M} - \text{H} - 3\text{C}_6\text{H}_{10}\text{O}_5 - 2\text{CH}_2\text{O} - \text{H}_2\text{O}]^-$, 221.1 $[\text{M} - \text{H} - 3\text{C}_6\text{H}_{10}\text{O}_5 - \text{C}_4\text{H}_8\text{O}_4]^-$, 193.0 $[\text{M} - \text{H} - 3\text{C}_6\text{H}_{10}\text{O}_5 - \text{C}_5\text{H}_8\text{O}_5]^-$, 179.1 $[\text{M} - \text{H} - 4\text{C}_6\text{H}_{10}\text{O}_5]^-$, 161 $[\text{M} - \text{H} - 4\text{C}_6\text{H}_{10}\text{O}_5 - \text{H}_2\text{O}]^-$, 131.0 $[\text{M} - \text{H} - 4\text{C}_6\text{H}_{10}\text{O}_5 - \text{CH}_2\text{O} - \text{H}_2\text{O}]^-$, 113.0 $[\text{M} - \text{H} - 4\text{C}_6\text{H}_{10}\text{O}_5 - \text{CH}_2\text{O} - 2\text{H}_2\text{O}]^-$, and 101.0 $[\text{M} - \text{H} - 4\text{C}_6\text{H}_{10}\text{O}_5 - 2\text{CH}_2\text{O} - \text{H}_2\text{O}]^-$. Similar fragment ions and neutral losses were observed for CID of m/z 665 (Figure 2.1c). Although the 665 m/z ion was assigned as a 6-carbon tetrasaccharide with high mass error (-10.1 ppm), its CID data suggests that it was identified correctly as its fragmentation pattern is consistent with that observed for other carbohydrates. One interesting observation made during analysis of CID data was that in addition to the expected carbohydrate fragment ions for the 6-carbon disaccharide (m/z 341), a fragment ion at m/z 135 was also observed (Figure 2.1d). This fragment is not native to known carbohydrate fragmentation pathways. However, it has previously been observed as a fragment of a caffeic acid-hexose conjugate.⁹⁰ The $[\text{M} - \text{H}]^-$ ion of the conjugate has a lower exact mass (341.0873) than the 6-carbon disaccharide ion (341.1084). The presence of a species at lower m/z offers a possible explanation for the high negatively biased error (-17.6 ppm) at m/z 341, assuming both of these species are present. Representative CID mass spectra of three organic acids are shown in Figure 2.2. CID spectra for m/z 163 (Figure 2.2a) show a prominent fragment ion of 119.05, which corresponds to loss of CO_2 from the parent ion. CID spectra for m/z 181 (Figure 2.2b) show dominant fragment ions of m/z 166.03 $[\text{M} - \text{H} - \text{CH}_3]^-$, 163.04 $[\text{M} - \text{H} - \text{H}_2\text{O}]^-$, 151.05 $[\text{M} - \text{H} - 2\text{CH}_3]^-$, 137.06 $[\text{M} - \text{H} - \text{CO}_2]^-$, 135.05 $[\text{M} - \text{H} - \text{CO} - \text{H}_2\text{O}]^-$, and

123.01 $[M - H - COH - 2CH_3]^-$. CID spectra of m/z 193.05 (Figure 2.2c) show fragment ions of 178.03 $[M - H - CH_3]^-$ and 134.04 $[M - H - CO_2 - CH_3]^-$.

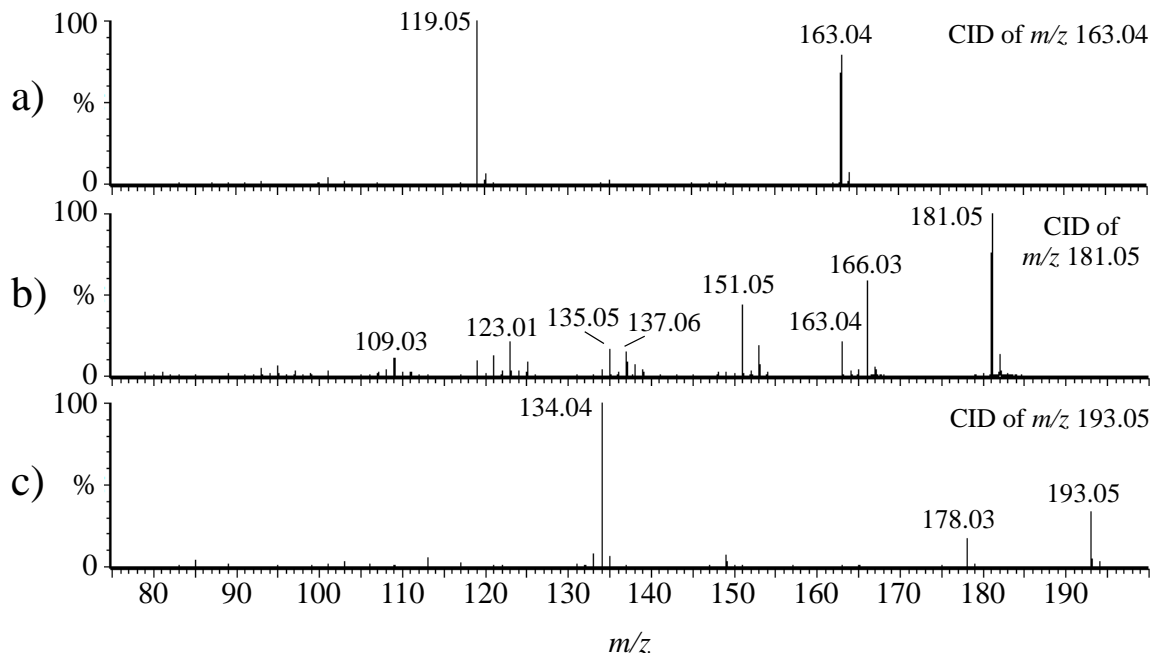


Figure 2.2. Negative ion mode CID mass spectra of (a) 4-hydroxycoumaric acid (163.04 m/z), (b) homovanillic acid/syringaldehyde (181.05 m/z), and (c) ferulic acid/3-hydroxy-4-methoxycinnamic acid (193.05 m/z) in a corn stover hydrolysate.

Characterization of IM-MS Spectra of Biomass Pretreatment Hydrolysates and Aqueous Extracts

Feasibility of DIESI-IM-HRMS for analysis of biomass samples was evaluated using two sample types derived from a sorghum feedstock: (i) a dilute-acid pretreated hydrolysate and (ii) an aqueous extract. These samples were selected as representative of the relatively complex matrices typically encountered in bioprocess workflows. Although pretreatment hydrolysates have long been a subject of investigation, relatively little is known about their total molecular composition. On the other hand, the relevance

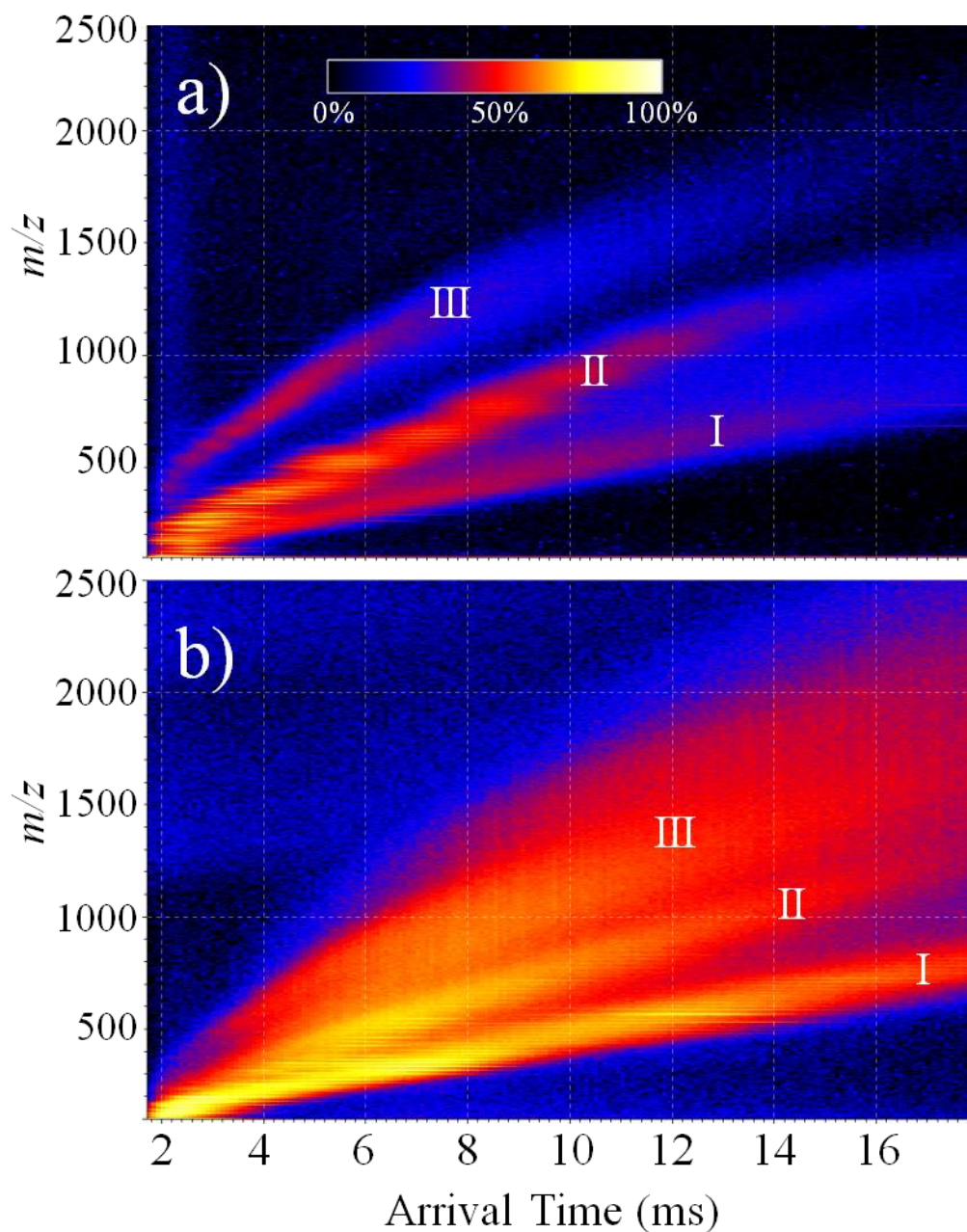


Figure 2.3. DIESI-IM-MS plot of m/z vs. arrival time for negative ions resulting from 1:10 dilutions of (a) 0.7% H_2SO_4 -pretreated sorghum hydrolysate with three distinct trendlines (I, II, and III) and (b) aqueous extract of sorghum with two distinct trendlines (I and II) and a region with several unresolved trendlines (III).

of aqueous extracts to valuations of potential energy-yield from a feedstock has only recently been realized.⁹¹

A dilute-acid pretreated sorghum hydrolysate (diluted 10-fold) was analyzed and resulting ion mobility – mass spectra from this analysis are contained in Figure 2.3a.

Note that data in Figure 2.3 were collected over 2 min. Compared to traditional chromatographic separations of biomass hydrolysates,^{21,70,71,92} a 2-min analysis time represents a 10- to 20-fold increase in throughput. In addition, these data exhibit an increase in peak capacity from ~79,000 with HRMS-only analysis to ~350,000 with IM-HRMS. The total ion signal observed via mass analysis is separated into three distinct trendlines as indicated by the signal groupings in Figure 2.3a, I-III. The charge state of each trendline was determined by isotopic analysis. The ¹²C and ¹³C isotope peaks differ by 1 *m/z* for peaks within the lower two trendlines, indicating that ions in these trendlines are singly charged. Isotopes in the upper trendline (Figure 2.3a, III) differ by 0.5 *m/z*, revealing it to consist of doubly charged ions. Separation of singly from doubly charged species assists in deconvolution of mass spectra and allows for easier identification of individual components. In Figure 2.3a, peaks are clearly observed between arrival times of 2 and 18 ms and up to *m/z* 2000. The most intense ion signal is observed along the middle trendline (between 2-12 ms and 100-1100 *m/z*). Note that the components typically monitored in biomass hydrolysates (i.e., organic acid degradation products and monosaccharides) via LC-UV⁷⁰ and LC-MS^{26,71} fall below *m/z* 250. However, data in Figure 2.3a reveal that these analytes would only represent a small fraction of total hydrolysate composition, as a majority of the ion signal falls outside of that mass range.

The identified carbohydrates and lignin degradation products (Table 2.1) were all observed to occur along the bottom trendline (Figure 2.3a, I).

An aqueous extract of the same sorghum feedstock was also analyzed with DIESI-IM-HRMS. Ion mobility – mass spectra of the extract show two clear trendlines (Figure 2.3b, I and II) and other poorly resolved trendlines at higher m/z and arrival time (Figure 2.3b, III). A majority of the same carbohydrate oligomers that were observed in the hydrolysate were also observed in the lower trendline of this data (Figure 2.3b, I). We note that the two lower trendlines in Figure 2.3a roughly overlay with those in Figure 2.3b, indicating that the same two compound classes are present in both samples. The bottom two trendlines (Figure 2.3b, I and II) are designated as singly charged by isotope analysis and the upper, poorly resolved trendlines (Figure 2.3b, III) contain doubly charged ions. Of particular note in the aqueous extract, intense ion signal is seen up to m/z 2500 and over a wide arrival time range (2-18 ms). One explanation for the presence of higher molecular weight species in the aqueous extract compared to the hydrolysate is that some of the larger components of biomass might be broken down to smaller degradation products during dilute-acid pretreatment.⁶⁸ DIESI-IM-MS spectra for corn stover hydrolysate can be seen in Appendix B (Figure B.2).

Although exact mass and CID data confirm that the lower trendlines shown in Figure 2.3a and 2.3b (labeled “I”) contain organic species native to plant material, no organic species identified by HRMS or CID were observed in the other trendlines (as detailed below, these trendlines were ultimately determined to be composed of inorganic ion clusters through exact mass and CID analysis). In initial experiments, organic extractions of the biomass samples were performed by placing them in contact with an

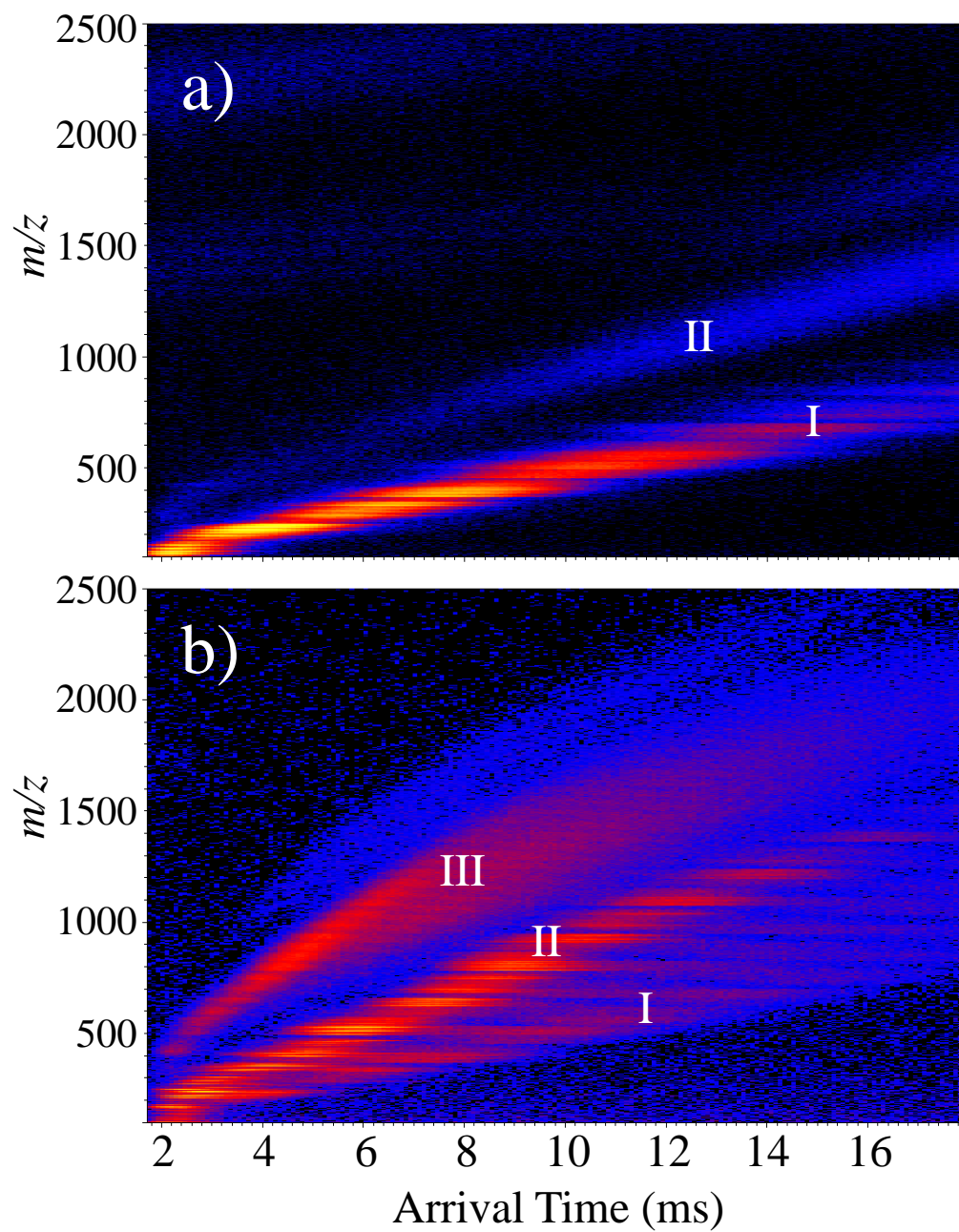


Figure 2.4. DIESI-IM-MS spectra of a) MTBE-extracted raw sorghum hydrolysate and b) aqueous fraction leftover after MTBE extraction with labeled trendlines (I, II, and III).

equal volume of MTBE, a polar, aprotic solvent which has previously been demonstrated as suitable for extracting organic degradation products from pretreatment hydrolysates.⁷⁰ Subsequent analysis of the two fractions revealed that only components of the bottom trendline (Figure 2.4a, I) partition to the MTBE solvent, while the components from the middle and upper trendlines (Figure 2.4b, II and III) remain in the aqueous phase. This result suggests ions in the middle and upper trendlines may be highly polar or ionic (e.g., inorganic salts), either natively present in the sample, or potentially introduced during the pretreatment process. Although species in the lower trendline were only partially extracted into the MTBE solvent under the specified conditions, all analytes in the lower trendline exhibited similar distribution behavior as evidenced by uniform intensity across the trendline in both the extracted and aqueous fractions. The partitioning behavior of analytes in the lower trendline is thus consistent with the assumption that it is composed predominantly of species resulting from the components native to lignocellulosic biomass (i.e., polar organic species derived from cellulose, hemicellulose, and lignin).

The mobility-extracted mass spectra of trendline II (Figure 2.3a) is shown in Figure 2.5. Most intense ions in these data exhibit a strong M+2 isotope, consistent with the presence of a Cl or S atom in these analytes (e.g., in the form of Cl^- or HSO_4^- adducts). For example, a relatively intense M+2 peak for m/z 216.91 can be observed in the mobility-extracted mass spectra expansion of sorghum hydrolysate (Figure 2.5, inset). HRMS exact masses of these peaks are consistent with inorganic adducts of HSO_4^- , H_2SO_4 , SO_4^{2-} , K^+ , and Na^+ , combined in various ways to give a single negative charge. Presence of a peak at m/z 96.96 in CID analysis (see Figure 2.6 a-d for examples) for nearly all of these peaks confirmed that they contain bisulfate, HSO_4^- . Deconvolution of

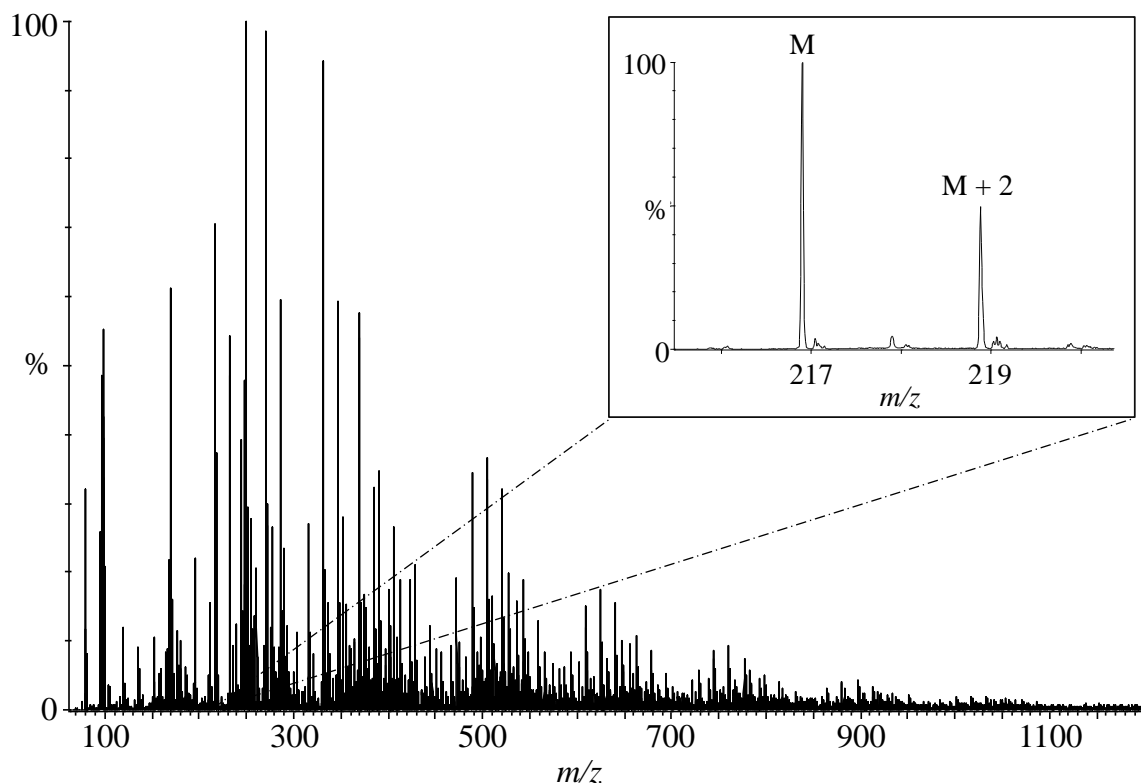


Figure 2.5. Extracted high-resolution mass spectra of the middle trendline (Figure 2.3a, II) resulting from DIESI-IM-HRMS analysis of a dilute-acid pretreated sorghum hydrolysate. The region between 216-220 m/z is expanded (inset) to demonstrate the prominent M and M + 2 isotopes typical of ions in this trendline.

the extracted mass spectra of the middle trendline (Figure 2.7) reveals several series of peaks that differ by 16 Da, which corresponds to an exchange of Na^+ for K^+ . In fact, a majority of the adducts are of the form $[\text{K}_{2n-1}(\text{SO}_4)_n]^-$, $[\text{K}_n(\text{HSO}_4)_{n+1}]^-$, $[\text{Na}_{2n-1}(\text{SO}_4)_n]^-$, $[\text{Na}_n(\text{HSO}_4)_{n+1}]^-$, $[\text{Na}_n\text{K}_m(\text{HSO}_4)_{n+m+1}]^-$, or $[(\text{H}_2\text{SO}_4)_n(\text{HSO}_4)]^-$ according to exact mass data. These inorganic ions are likely the result of the dilute-sulfuric-acid pretreatment process, as they were also present in a control experiment (see Figure 2.8 for IM-MS and MS spectra), where dilute-acid accelerated solvent extractor pretreatment was performed in an empty cell (i.e., in the absence of a feedstock).

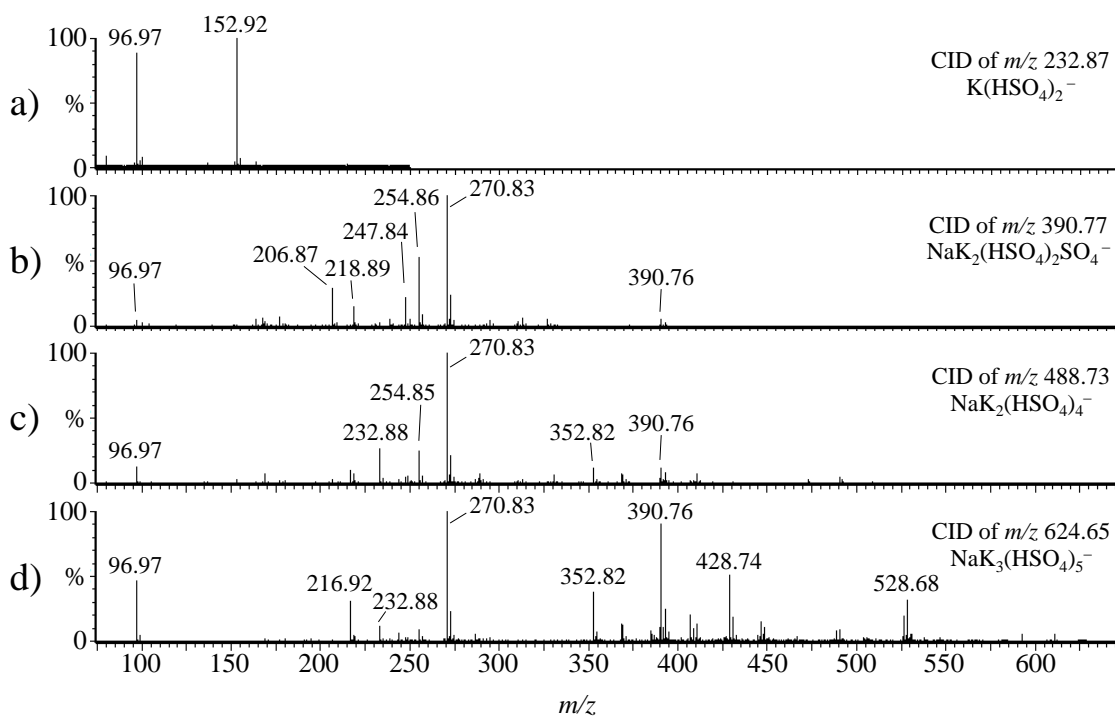


Figure 2.6. Negative ion mode CID mass spectra of (a) $\text{K}(\text{HSO}_4)_2^-$ (232.87 m/z), (b) $\text{NaK}_2(\text{HSO}_4)_2\text{SO}_4^-$ (390.77 m/z), (c) $\text{NaK}_2(\text{HSO}_4)_4^-$ (488.73 m/z), and (d) $\text{NaK}_3(\text{HSO}_4)_5^-$ (624.65 m/z) from trendline II (Figure 2.3a).

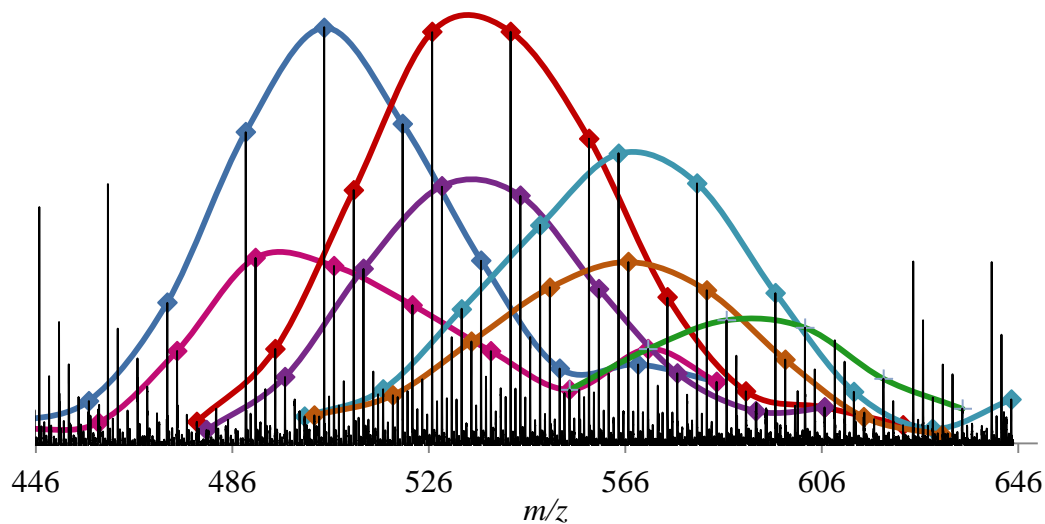


Figure 2.7: Deconvolution of a representative portion of the mass spectra of a dilute-acid-pretreated sorghum hydrolysate (i.e., ion signal in trendline II, Figure 2.3a). Each distribution (represented by different colored lines) corresponds to m/z peaks that are separated by 16 Da.

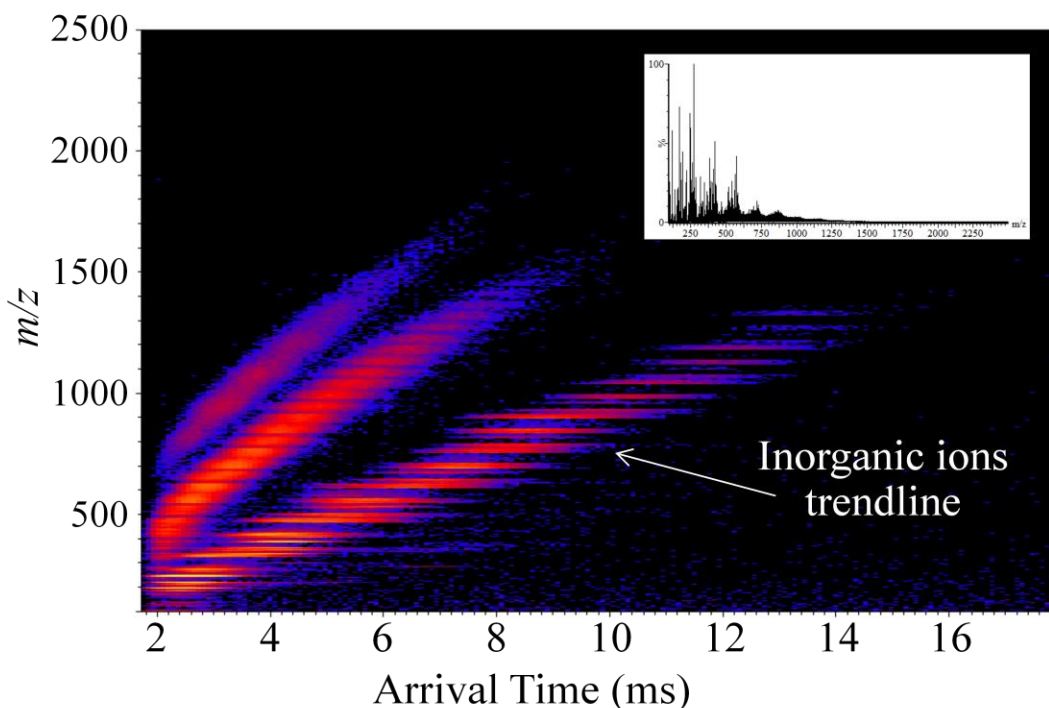


Figure 2.8. IM-MS spectra and inset of the mass spectrum from a control experiment, where a dilute-acid accelerated solvent extractor pretreatment was performed without any sample (no biomass). The trendline corresponding to inorganic ions is denoted. The inset represents the total mass spectra of this sample and was added for comparison to the extracted mass spectra (Figure 2.5) of trendline II in sorghum hydrolysate (Figure 2.3a).

Although bisulfate adducts were not as prevalent in the middle trendline of aqueous extracts (Figure 2.3b, II), a few were observed, likely resulting from residual acid in the ASE system that was used to perform the extraction. The remaining ion signal in the middle trendline of Figure 2.3b can be partially explained by adducts of K^+ and Na^+ with Cl^- , OH^- , or other inorganic anions. However, exhaustive assessment of the ion signal in this region has not been performed. It is important to note that HSO_4^- adducts of carbohydrates and lignin degradation products were found to fall on the bottom trendline along with the non-adducted biomass analytes. However, no peaks were detected in the bottom trendline that were purely ionic/inorganic. Conversely, no non-ionic/organic components were observed in the middle or upper trendlines. Although ion

signal is too weak in the upper trendlines to achieve adequate accurate mass data, this signal's correlation between m/z and mobility arrival time is consistent with doubly charged salt adducts (in the case of Figure 2.3a, III) and both doubly charged salt adducts and doubly charged organic species in the broad upper regions of Figure 2.3b (III).

In order to reduce the intensity of the inorganic ions, as well as minimize the formation of HSO_4^- adducts with analyte ions, Ca(OH)_2 was added to the hydrolysate to remove H_2SO_4 . The resulting CaSO_4 precipitate was removed by filtration. The mass spectra of the treated hydrolysate did exhibit a reduction in intensity of the inorganic ions. However, the ion mobility spectra (Figure 2.9) of the hydrolysate were complicated with the addition of multiple new trendlines, corresponding to greater abundance of multiply-charged ions resulting from the increased pH of the sample. Due to the increased complexity of the spectra upon addition of Ca(OH)_2 , overliming samples before ion mobility analysis, which involves treatment of a pretreatment hydrolysate with Ca(OH)_2 , would not be preferred without carefully readjusting the pH with a suitable volatile acid (e.g., formic acid).

The persistence of inorganic salts and surfactants that are introduced during sample processing and cleanup is a significant hindrance for traditional mass spectrometry techniques, requiring exhaustive labor to remove this interfering signal from the spectra and often significantly limiting observation of relevant ion signal. The presence of such contaminants is especially problematic for MS analysis of routine biomass hydrolysates where pretreatment with dilute sulfuric acid is among the most popular approaches.⁹³ However, as shown above, IM-MS provides a two-dimensional separation that enables signal resulting from inorganic salt adducts to be resolved from

target organic species. In fact, m/z data can be extracted for a particular trendline (a fixed mobility and m/z correlation range), which excludes interfering ions from other trendlines and results in mass spectra with increased sensitivity for analytes composing that trendline. Such abilities are particularly useful when the chemical noise potentially swamps analyte signal, as would be the case for MS-only analysis of the sorghum hydrolysate.

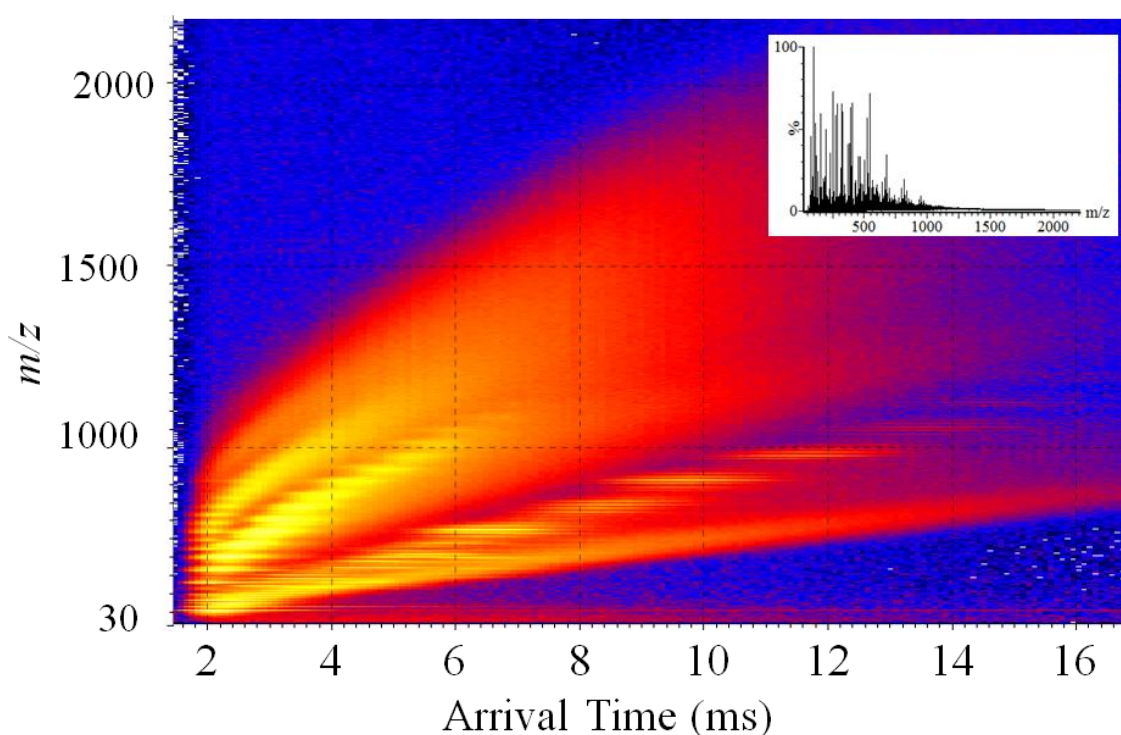


Figure 2.9. IM-MS spectra with mass spectrum inset of H_2SO_4 -pretreated sorghum hydrolysate neutralized with $\text{Ca}(\text{OH})_2$. Additional trendlines of multiply-charged ions are present.

Evaluation of IM-HRMS Separation of Biomass Components

The utility of IM-HRMS over low-resolution mass spectrometry or even HRMS only for whole-sample biomass analysis can be further evaluated through careful inspection of sample data. For example, a majority of the carbohydrates identified in

these samples (Table 2.1) were observed to have another ion occurring at the same nominal m/z (separated by as little as 16 mDa) and often only partially resolved. Figure 2.10 shows the IM-HRMS data for a dilute-acid-pretreated corn stover hydrolysate in the range 544.94-545.36 m/z , which is the molecular weight range of a carbohydrate oligomer believed to be present in the sample. These data (Figure 2.10, left projection) reveal two peaks in the mass spectrum with the same nominal mass but differing in exact mass by 0.0711 m/z (i.e., 545.1003 and 545.1714). Although nominal-mass analyzers would not be able to separate these components, HRMS is able to distinguish between the two species without the need for front-end prefractionation. More significantly, despite only achieving partial baseline resolution for these two components in the mass domain, addition of IMS separation was able to provide baseline resolution of the two main components (Figure 2.10, top projection) and also reveals the presence of a third species that was not observable by HRMS analysis alone (i.e., Figure 2.10, “unk*”). The peak at m/z 545.1714 corresponds to a 5-carbon tetrasaccharide (e.g., xylotetraose) as established by exact mass and CID (described above). The peak at m/z 545.1003 is believed to be a hexuronic acid trimer (HexA₃) based on exact mass (2.4 ppm error) and CID spectra which show losses of 60 ($-2\text{CH}_2\text{O}$), 176 ($-\text{C}_6\text{H}_8\text{O}_6$ [one polymer unit]), 194 ($-\text{C}_6\text{H}_{10}\text{O}_7$ [one polymer unit and H_2O]), and 254 ($-\text{C}_8\text{H}_{14}\text{O}_9$ [one polymer unit, H_2O and $2\text{CH}_2\text{O}$]). Arrival times of the two components identified in Figure 2.10 are centered at 8.874 and 10.005 ms for the 545.1003 and 545.1714 m/z ions, respectively. As expected, the hexuronic acid oligomer with only three polymer units is observed at a shorter average arrival time than the 5-carbon tetrasaccharide with four polymer units.

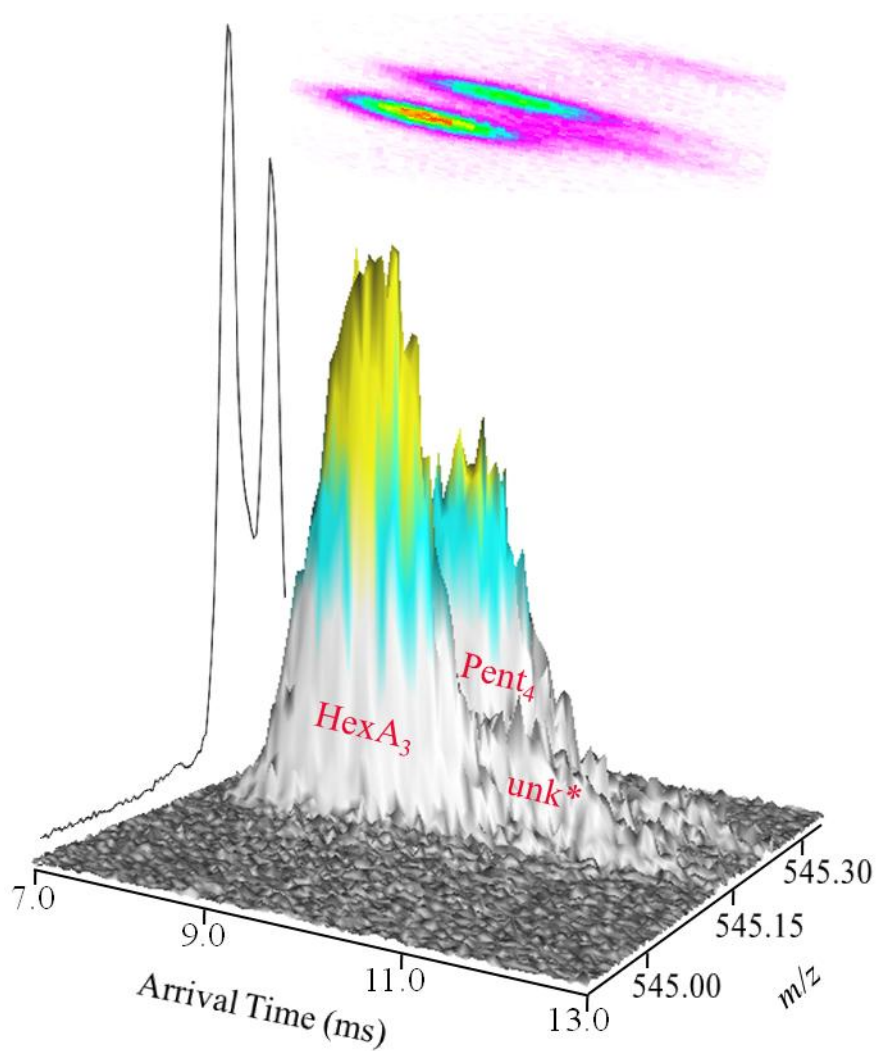


Figure 2.10. Three-dimensional plot of arrival time versus m/z for a pentose tetramer (Pent_4 , 545.1714 m/z), a tentatively identified hexuronic acid trimer (HexA_3 , 545.1003 m/z), and an unknown component (unk^* , 545.14 m/z). These data demonstrate the potential for increasing component resolution when incorporating ion mobility as an alternate post-ionization separation. See text for details.

In addition to enhancing resolving power, DIESI-IM-HRMS can also be employed towards rapid fingerprinting of potential biofuel process streams. Chemical fingerprinting of lignocellulosomic samples by MS alone can be particularly difficult owing to the high degree of structural diversity, which results in (potentially multiple) ion signal(s) at every nominal mass. With the incorporation of ion mobility, unique arrival times of ions are obtained allowing rapid differentiation between samples and/or sample components. For example, Figure 2.11 shows three-dimensional IM-MS plots for an expansion of IM-MS spectra for sorghum and corn stover hydrolysates. Note that assessment of mass spectral data alone (Figure 2.11, top) would reveal similar data for both sample types, that is, a peak at every nominal mass with modest variation in intensity. However, the combination of MS and IMS (Figure 2.11, bottom) enables more facile differentiation between the two samples. In particular, differences in arrival time distributions between the two samples at a given nominal mass are indicative of different components present in each sample (as might be expected for two different feedstocks). Rapid fingerprinting of biofuel process streams may be particularly useful when monitoring stream “health”, assessing sample viability, and during chemometric feedstock valuations.⁷²

Both peak width and arrival times may be used to evaluate the presence of specific species. For example, a 5-carbon trisaccharide was identified at m/z 503 in corn stover and sorghum hydrolysates. To determine the potential of ion mobility to differentiate between negatively charged ions for structural isomers of carbohydrates at this mass, four common glucose trisaccharide isomers (varying in linkage position) were analyzed with DIESI-IM-HRMS (Figure 2.12). Ion mobility traces for the monoisotopic

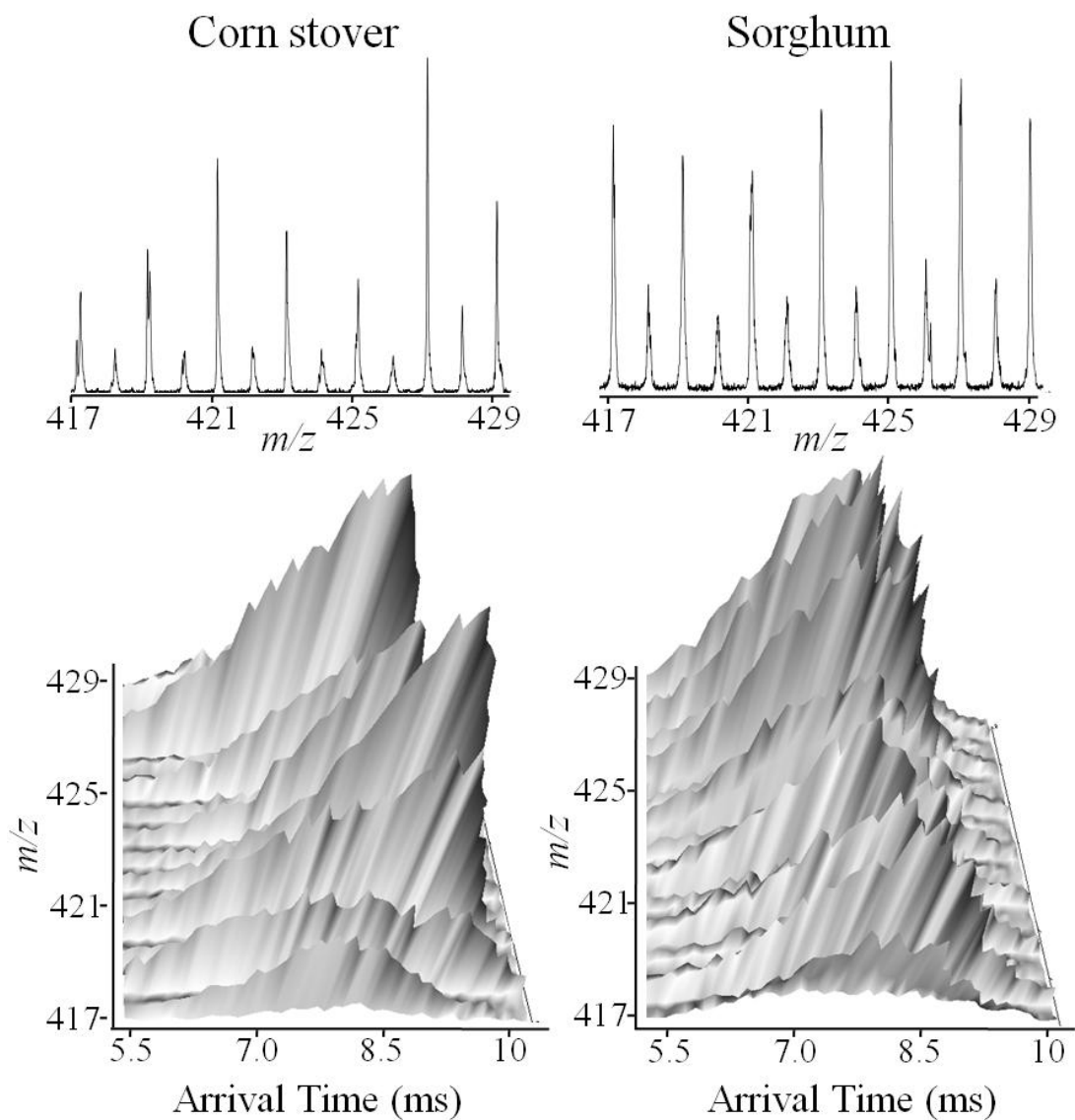


Figure 2.11. Representative HRMS (top) and IM-HRMS (bottom) data for corn stover (left) and sorghum (right) hydrolysates. Note the differences in ion arrival times for ions in the range 417-429 m/z , which is indicative of the different sample components in this range.

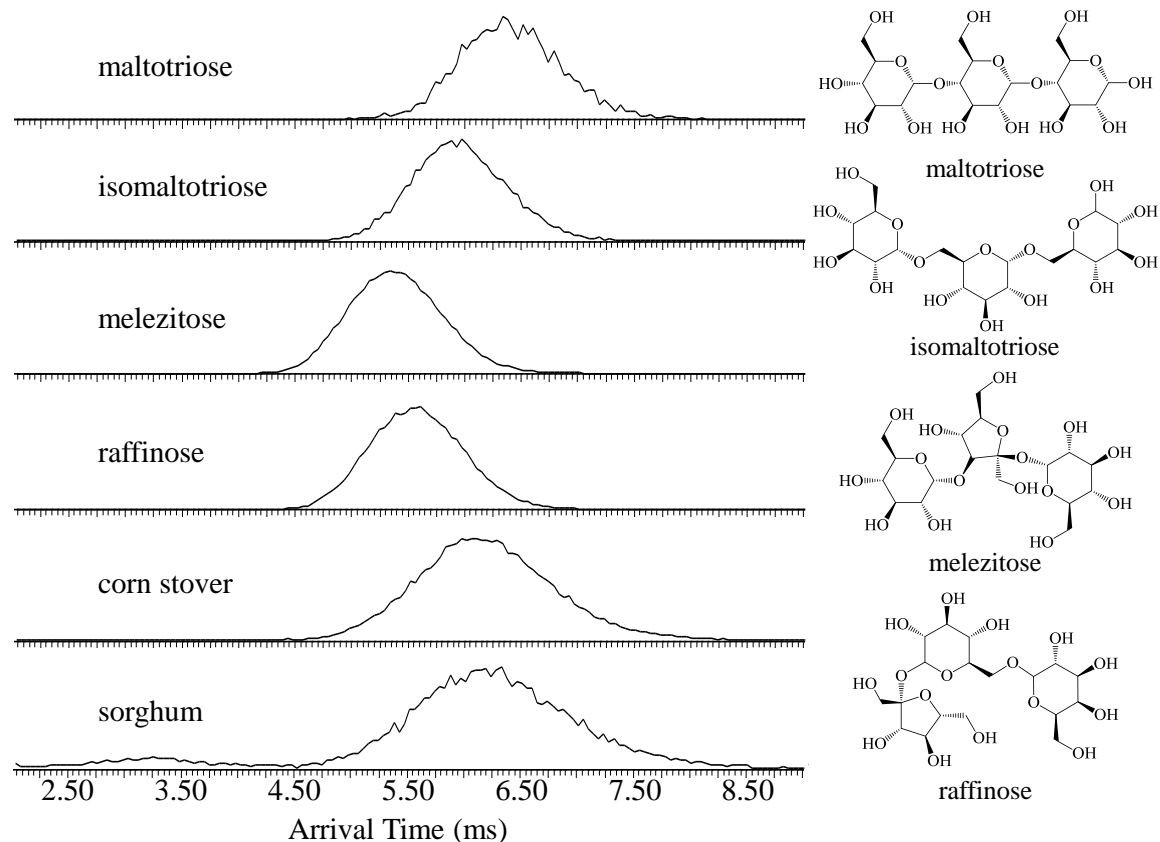


Figure 2.12. Arrival time distributions for the $[M - H]^-$ ion of four 6-carbon trisaccharide isomer standards and the corresponding trisaccharide signal observed in corn stover and sorghum hydrolysates. Structures for the carbohydrate standards reported in this figure are shown on the right. Note that an IM gas flow rate of 25 mL/min was used in the collection of these data.

mass of each isomer ($[M - H]^-$) are shown in Figure 2.12 (top 4 traces). Each ion was observed to have an arrival time distribution occurring between 4.4 and 7.5 ms with a base peak width of about 2.2 ms. Maltotriose had the longest arrival time with a peak apex of 6.4 ms. Isomaltotriose, which has greater structural flexibility between monomer units (thus may fold more compactly in the gas phase), had an average arrival time of 5.9 ms. The more compact trisaccharides, melezitose and raffinose, each with one 5-membered ring, had the shortest average arrival times of 5.3 ms and 5.5 ms, respectively. To compare the arrival time distributions of these four trisaccharides to the peak at the

same m/z observed in real samples, ion mobility traces observed for m/z 503 in two biomass hydrolysates are provided in the bottom 2 traces of Figure 2.12. Average arrival times are similar (6.2 ms) and fall between that of maltotriose and the other three isomers. These peaks are also broader (3.5-4.0 ms at the base) than the width of a single-component peak, indicating that multiple isomers may be present in the biomass samples (these data are described in more detail below).

Mass-Mobility Correlations of Carbohydrates in Biomass

The plot of m/z vs. arrival time in Figure 2.13 shows IM-MS correlations for 5- and 6-carbon carbohydrates and their various adducts (i.e., $[M - H]^-$, $[M + Cl]^-$, and $[M + HSO_4]^-$) in a corn stover hydrolysate. No notable differences in the IM-MS correlations for 5- and 6-carbon carbohydrates or the different adduct types are observed. A few ions appear to fall slightly above or below the trendline; however, Ruotolo and coworkers⁹⁴ have previously demonstrated that analytes in a given compound class can deviate from a trendline by up to 11%. In Figure 2.13, all carbohydrate oligomers are within 6% of the average trendline suggesting abnormal structural variation does not occur for carbohydrates under the instrumental conditions utilized. Most importantly, these data suggest that any ion signal deviating by more than 6% from the average carbohydrate trendline is likely not a carbohydrate, or at least should be carefully evaluated before making a positive carbohydrate assignment.

Utilizing mass-mobility correlations can be important for analyte determinations in complex samples. For example, when investigating a biomass hydrolysate using direct injection on a nominal-mass mass spectrometer, Helm and researchers⁶⁹ observed a peak at m/z 246.8 that was identified as a pentose- HSO_4^- adduct. In the mass spectrum of

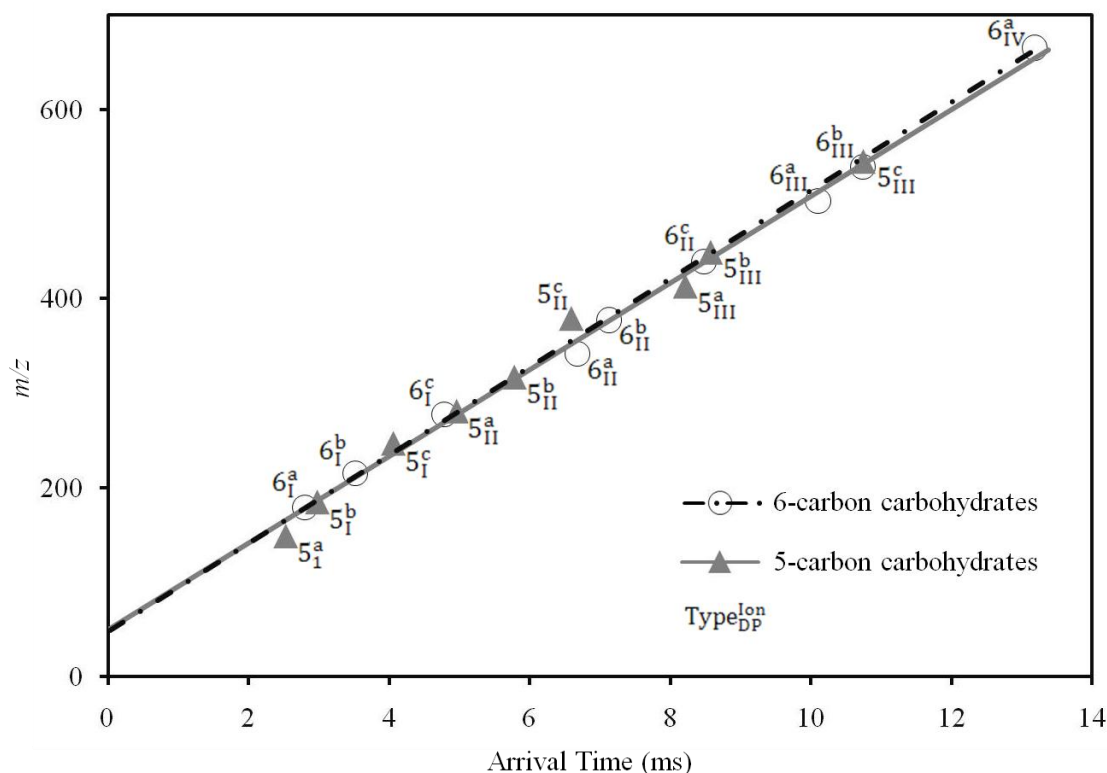


Figure 2.13. Mass-mobility correlations for 5- and 6-carbon carbohydrate ions observed in corn stover hydrolysate. The numbers “5” and “6” are added to the plot for easy identification of 5- and 6-carbon data points with superscripts and subscripts representing, respectively, the ion type and degree of polymerization. Superscript letters designate ion type as follows: (a) $[M - H]^-$, (b) $[M + Cl]^-$ and (c) $[M + HSO_4]^-$. Degree of polymerization (i.e., DP units) for each ion is indicated by the subscripts I, II, III and IV, respectively, for the monomer through increasing polymer units.

biomass hydrolysate used in this work, an intense peak at m/z 246.85 was observed with a corresponding average arrival time of 2.62 ms. Based on previous work,⁶⁹ the analyst would be tempted to assume this analyte is the same pentose- HSO_4^- ion as reported previously. However, the mass and arrival time of this ion was observed to correspond to an analyte positioned on the middle trendline (i.e., Figure 2.3a, II), indicating that it is an inorganic ion. A less intense peak at m/z 247.01 and 4.24 ms was identified as a potential candidate for the pentose- HSO_4^- adduct, which was confirmed by exact mass. With unique mass-mobility correlations for different compound classes (e.g., carbohydrates

and inorganic ions), identification of peaks with close or even the same molecular weight can be made with improved confidence without requiring additional analyses (e.g., LC-MS, CID, etc).

Collision-Cross-Section Measurements of Biomass Analytes

Comparison of ion mobility arrival times for standard and sample peaks can be useful in analyte identification. However, arrival times can vary between instruments and are dependent on various ion mobility parameters such as field strength, cell length, gas pressure or flow rate, wave height, etc. Collision cross sections (CCS), on the other hand, are (i) independent of both the instrument used and instrumental parameters (assuming low electric-field conditions are maintained during IMS separations), and (ii) provide an effective measure of an ion's gas-phase size, which is constant for a given analyte. Thus, measurement of CCS can provide increased confidence in analyte assignments, particularly in the case of isomers. To generate representative data, collision cross sections of carbohydrate standards and select carbohydrate ions in a corn stover hydrolysate were measured. Table 2.2 shows CCS values for the $[M - H]^-$ ions of xylose, glucose, and seven oligosaccharide standards. Generally, CCS values ranged from 110.4 to 258.2 Å² and varied linearly with molecular weight as expected from evaluation of arrival time distributions. Collision cross sections for the negative carbohydrate ions were found to be larger than those reported for positive ions (i.e., $[M + Na]^+$).^{57-59,81} Differences are smallest for the monosaccharides (about 2% larger than literature values for xylose and glucose) and largest for maltose (16% difference) with the four 6-carbon trisaccharides in between at 7-13%. It is not known whether these differences in cross section result from differences in ion structure or from differences

between the two approaches for measurement. However, the collision cross sections found in this work for negative trisaccharide ions exhibit the same trend (maltotriose > isomaltotriose > raffinose > melezitose) as those reported for positive ions in literature.

57,58,81

CCS data for carbohydrates identified in the corn stover hydrolysate were determined for $[M - H]^-$, $[M + Cl]^-$, and $[M + HSO_4]^-$ ions and are reported in Table 2.2 (Cl^- and HSO_4^- adducts of the 6-carbon tetra- and pentasaccharides were not observed in the hydrolysate). These CCS values (113.8 to 256.6 Å²) have a similar range as the analyzed standards and agree with the values for the $[M - H]^-$ ions of the standard carbohydrates ($\leq 3\%$ difference), suggesting that the molecular structures of the standard analytes are probably the same as or similar to those analytes observed in corn stover. We note that direct comparison of Cl^- and HSO_4^- adducts between the hydrolysate and the standards is not possible as they were not observed in the standard solutions. One exception to the agreement between the standards and analytes observed in corn stover is the CCSs of the 6-carbon trisaccharides. Four different 6-carbon trisaccharide isomer standards (see Figure 2.12 for structures and ion arrival times) were investigated in this work and their collision cross sections range from 197.1 (melezitose) to 210.6 (maltotriose) Å² (Table 2.2). As described above, the broad arrival time distribution for the 6-carbon trisaccharide in corn stover suggests the presence of multiple isomers. However, the average collision cross section of this peak (210.1 Å²) is very similar to that of maltotriose (0.24% difference), indicating that maltotriose might be the most abundant isomer in the sample.

A particularly interesting feature of the collision-cross-section data for carbohydrates (Table 2.2) is that a significant difference in collision cross sections for the

Table 2.2. Measured average collision cross sections for negatively charged ions of select carbohydrate standards (top) and identified carbohydrates in corn stover hydrolysate (bottom).

Analyte	DP unit	m/z^a	Collision cross section ^b (Å ²)		
Standards			[M-H] ⁻		
xylose	1	149	110.4		
xylobiose	2	281	150.3		
xylotetraose	4	545	214.7		
glucose	1	179	118.3		
maltose	2	341	166.0		
maltotriose	3	503	210.6		
isomaltotriose	3	503	205.3		
melezitose	3	503	197.1		
raffinose	3	503	200.2		
maltotetraose	4	665	227.2		
stachyose	4	665	227.7		
maltopentaose	5	827	258.2		
Corn stover			[M-H] ⁻	[M+Cl] ⁻	[M+HSO ₄] ⁻
5-carbon	1	149	113.8	127.1	145.5
	2	281	150.8	158.3	169.5
	3	413	191.2	194.3	202.3
	4	545	212.3	219.8	220.6
6-carbon	1	179	120.6	136.6	144.7
	2	341	167.9	185.4	185.4
	3	503	210.1	215.2	216.9
	4	665	235.7	—	—
	5	827	256.6	—	—

^a m/z represents nominal mass for [M - H]⁻ ions. m/z of chloride and bisulfate adducts in corn stover (not shown) are equivalent to the m/z of [M - H]⁻ + 36 and + 98 Da, respectively.

^b Relative standard deviations for all CCSs are less than 2%.

deprotonated and adducted (Cl^- and HSO_4^-) ions is observed. In the absence of conformational change, ion adduction generally does not significantly alter the collision cross section of ions in large molecules (e.g., peptides and proteins), especially in the case of sodium or potassium adducts.⁹⁵ However, adducting ions have been observed to change CCSs for smaller molecules $< 200\ m/z$ (up to 14% and 24% change in CCS with Na^+ and $\text{Ca}(\text{CH}_3\text{COO})^+$, respectively).^{59,78,80} It is currently not clear whether the observed change in cross section is due to conformational change or to the relative size of the anion. We note that the difference is larger for lower molecular weight carbohydrates (up to 32%) than for higher molecular weight carbohydrates (as low as 2.4%), which may indicate that the relative size of the adducting ion to the analyte is alone responsible for influencing the observed CCS. In addition, bisulfate adducts were found to have larger collision cross sections than chloride adducts, with the exception of the 6-carbon disaccharide and trisaccharide where the values are equal or nearly equal.

In addition to carbohydrates, collision cross sections were also measured for select small molecule ($< 200\ m/z$) biomass degradation products (i.e., aromatic acids and aldehydes). Collision cross sections of standards and corresponding ions found in corn stover and sorghum hydrolysates are listed in Tables A.1-3 in Appendix A. A representative subset of these data is contained in Table 2.3. A comparison of cross sections for the standard and hydrolysate ions can provide information about sample composition. For example, caffeic acid appears to be present in both hydrolysate samples based on a less than 1% difference in cross section between the standard ($123.6\ \text{\AA}^2$) and hydrolysate samples ($123.0\ \text{\AA}^2$ for corn stover and $124.9\ \text{\AA}^2$ for sorghum). Similarly, the CCSs of ferulic acid in the standard ($134.3\ \text{\AA}^2$) and in the sorghum sample ($132.5\ \text{\AA}^2$) are

Table 2.3. Comparison of average collision cross sections for $[M - H]^-$ ions of representative aromatic acid and aldehyde standards and $[M - H]^-$ ions of the same exact mass in corn stover and sorghum hydrolysates.

Compound	m/z	Collision cross section ^a (\AA^2)		
		Standard	Corn stover	Sorghum
ferulic acid	193	134.3	130.7	132.5
caffeic acid	179	123.6	123.0	124.9
homovanillic acid	181	130.6	127.3	127.2
syringaldehyde	181	126.0		
2,5-dihydroxybenzoic acid	153	111.1	113.3	116.0
3,4-dihydroxybenzoic acid	153	116.8		
salicylic acid	137	105.7	109.8	113.2
4-hydroxybenzoic acid	137	107.0		
3,4-dihydroxybenzaldehyde	137	116.7		

^a Relative standard deviations for all CCSs are less than ~2%.

closely matched, indicating that the ion signal in sorghum hydrolysate at m/z 193 is consistent with the structure of ferulic acid. A slightly smaller CCS is observed in the corn stover hydrolysate (130.7 \AA^2), which might indicate that the peak at m/z 193 (in corn stover) is not only generated from ferulic acid but might also contain an ion with a smaller cross section such as a hexuronic acid, which has a molecular weight that would only be partially resolved from ferulic acid under the conditions used for cross-section determinations.

The remainder of the data in Table 2.3 shows collision cross sections for three different sets of isomeric organic acid standards and the corresponding ions in the hydrolysate samples. These three comparisons yield three different scenarios that can occur in IM-MS analysis. First, consider isomers homovanillic acid and syringaldehyde,

which have cross sections of 130.6 and 126.0 Å², respectively. The cross section of the ion in corn stover (127.3 Å²) matches the cross section of syringaldehyde, indicating that the m/z peak in the hydrolysate might belong to syringaldehyde. These data are in good agreement with quantitative determination of syringaldehyde and homovanillic acid in this sample (by an LC-MS/MS method²⁶), which indicates syringaldehyde is the most abundant isomer at 19.8 ppm while homovanillic acid is only present at a much lower concentration (1.08 ppm). A similar concentration comparison in sorghum is not possible as quantitative analysis on that sample has not been performed. However, the cross section for the 181 m/z ion in sorghum is very similar to the cross section of the ion in corn stover, which suggests that the two hydrolysates might have similar compositions in the case of these two isomers. Second, the collision cross section for m/z 153 in corn stover (113.3 Å²) was in between the cross sections for 2,5-hydroxybenzoic acid (111.1 Å²) and 3,4-dihydroxybenzoic (116.8 Å²) standards, both of which were confirmed by previous analysis to be in the sample at low concentrations. A greater width of the arrival time peak compared to the widths of the standard peaks also supports the presence of both of these isomers. Based on a larger observed cross section (116.0 Å²) for the 153 m/z ion in sorghum, 3,4-dihydroxybenzoic acid (116.8 Å²) is presumed to be the most abundant isomer. Third, the average collision cross sections for m/z 137 in corn stover (109.8 Å²) and sorghum (113.2 Å²) have values in between those of the standard compounds: salicylic acid (105.7 Å²), 4-hydroxybenzoic acid (107.0 Å²), and 3,4-dihydroxybenzaldehyde (116.7 Å²). However, the widths of the arrival time peaks for m/z 137 are not larger than the peak width of a single component, suggesting that the

peaks are probably not a combination of these isomers but rather that another isomer/analyte with the same molecular weight is present in the samples.

Evaluation of CCS for analyte determination does bring about a few interesting perspectives and/or concerns regarding the use of DIESI-IM-HRMS analysis for these sample types. First, while syringaldehyde and homovanillic acid were both present in the sample described above, only syringaldehyde was detected. As would be expected, only the most abundant analytes (or those with extremely high response factors) would typically be observed via this technique owing to the inherent qualities of DIESI. Second, although a number of analytes with identical chemical formulas have cross sections that are significantly different to the extent that analytical assignments are possible, the resolving power of the instrument employed in this work prohibits baseline (or even partial) resolution if these analytes were to occur simultaneously. However, even at current resolution, analyte arrival-time peak width and shape (i.e., Gaussian vs. non-Gaussian) can be relied on when assessing peak purity.

Conclusions

Owing to its ability to isolate relevant biomass signal from interfering sample components, DIESI-IM-HRMS is an attractive technique for assessing total molecular composition, which may ultimately provide a more thorough understanding of the correlation between hydrolysate composition and biofuel production. Although samples in biofuel process streams are typically screened for a handful of low-molecular weight components (i.e., $< 250\ m/z$), this work reveals highly complex samples with molecular weights of ions ranging up to $1500\ m/z$. With a 4-fold increase in peak capacity, IM-HRMS offers improved separation over HRMS alone. Increased sensitivity and

decreased spectral complexity can be achieved via trendline-refined IM-HRMS, enabling useful data collection without requiring extensive sample clean-up steps prior to MS analysis. In addition, ion-neutral collision cross sections provide a degree of analyte confirmation that supersedes analyte chromatography retention time (i.e., cross section is a constant for a given analyte independent of instrumental conditions). Although the resolving power of the mobility separation was not sufficient to provide baseline resolution for analytes of very similar size and shape, baseline resolution was often attainable with both IMS and HRMS separations combined. We note that current commercially available instrumentation offers on average a 10-fold increase in mobility resolving power which would substantially improve separation and identification of various isomers (e.g., carbohydrate and lignin degradation products) when they are present in the same sample.

CHAPTER THREE

A “Metabolomics” Approach for Identification of Unknown Degradation Products in Lignocellulomic Samples using Ultra Performance Liquid Chromatography – High-Resolution Mass Spectrometry

Introduction

Ultra performance liquid chromatography – high-resolution time-of-flight mass spectrometry (UPLC-HR-TOF-MS) combined with automated data processing presents an attractive technique for identification of unknown compounds in “omics” samples. UPLC-HR-TOF-MS offers many advantages over HPLC – low-resolution mass spectrometry.⁶² Compared to HPLC, UPLC utilizes smaller stationary-phase particles and higher operating pressures that provide increased chromatographic efficiency (i.e., higher-resolution and/or decreased analysis times). HR-TOF-MS provides high resolving power compared to “nominal mass” low-resolution MS (e.g., quadrupole mass analyzers), significantly increasing peak capacity of MS analyses. Additionally, fast data acquisition afforded by TOF-MS enables data-dependent analyses (e.g., MSⁿ analysis), which expedite detailed characterization of complex samples.⁶² Such information-rich analyses rely heavily on automation in both data collection and processing.

Lignocellulomics represents an emerging “omics” area that could greatly benefit from the methodology described above. The majority of lignocellulomic research centers around optimizing conversion of lignocellulosic biomass to biofuels and bio-based products. While numerous conversion schemes exist, most rely on three independent processing steps: (i) pretreatment – designed to improve accessibility of cellulose to

enzymes, (ii) enzymatic hydrolysis – which converts cellulose to monomeric glucose units, and (iii) microbial fermentation – which converts monomeric sugars to product. Degradation products resulting from pretreatment of biomass are known to have inhibitory effects on downstream enzymatic and/or microbial steps. However, current understanding of the relationship between pretreatment and inhibition is largely empirical. Thus, the ability to more accurately define the composition of pretreatment liquids has become increasingly important in fundamental biofuels research.

Analytical methods for lignocellulosic samples typically target a limited number of compounds compared to the number of potentially inhibitory degradation products present in a pretreatment sample. The most comprehensive LC-MS method currently employed utilizes tandem mass spectrometry to target 37 known organic acid, aldehyde and ketone degradation products.²⁶ While targeted MS is adequate for identifying known analytes of interest, a set of 37 compounds is only a small fraction of the total number of components in pretreated samples, as demonstrated in Chapter 2. A more comprehensive LC-MS approach that includes identification of non-targeted analytes has not previously been reported.

The current work focuses on development of an automated “omics” approach utilizing UPLC-HR-TOF-MS and tailored to identify previously unknown lignocellulosic degradation products. The objective of this work was not to exhaustively determine all unknown components in a biomass hydrolysate but rather to develop a method to do so and to evaluate its potential for future analyses. The approach combines traditional HRMS techniques with a “metabolomics” approach for novel compound identification. The validity of the approach was first evaluated with 35 standard compounds, known to

be present in biomass samples. The approach was subsequently used to identify degradation products in corn stover hydrolysate resulting from two different pretreatment methods: ammonia fiber expansion (AFEX) and dilute acid (DA). Results of the developed “metabolomics” identification method (i.e., MetaboLynx software and MS^e analyses) are compared to results obtained through manual identification (i.e., visual inspection of peaks and MS/MS experiments) for the AFEX-pretreated sample.

Experimental: Materials and Methods

Chemicals and Reagents

All chemicals were reagent grade or better and purchased from standard commercial vendors. Distilled water was purified to 18.2 MΩ with a Barnstead Nanopure Diamond UV water purification system. Ammonia-fiber-expansion (AFEX) pretreated corn stover samples were received from Dr. Bruce Dale (Michigan State University, Ann Arbor, MI) and were generated according to a previously reported method.⁹⁶

Generation of Extracted Hydrolysate Samples

AFEX-pretreated samples were extracted via accelerated solvent extraction (ASE) using a Dionex ASE-200 (Dionex Corp., Sunnyvale, CA) as previously reported.²⁷ Briefly, operation conditions were: 0.6 g of biomass per 11-mL cell, temperature: 70 °C, pressure: 1500 psi, preheat time: 3 min, heat time: 5 min, static time: 10 min, purge time: 60 s, flush volume: 150% of cell, and cycles: 2.

The dilute-acid hydrolysate was generated by pretreating corn stover with 0.7% (w/w) H₂SO₄ using the ASE pretreatment protocol described in Chapter 2. All

hydrolysates were extracted with methyl-*tert*-butyl ether (MTBE) according to the method developed by Chen et al.²⁴ All samples were filtered using a 25-mm syringe filter with a 0.2 µm nylon membrane (Pall Corp.) before analysis.

Preparation of Standards

A stock solution of 35 organic acid and aldehyde standards (see Table 3.2 for compounds) was prepared at 100 mg/L by dissolving neat chemicals in methanol. Standard solutions of 1.0, 5.0, 10.0, and 15.0 mg/L were prepared by dilution of the 100 mg/L stock solution with water. A 0.5 mg/L standard was prepared by diluting the 10.0 mg/L solution with water.

Ultra Performance Liquid Chromatography

UPLC analyses were performed with a Waters Acquity UPLC BEH C18 column (1.0 mm x 100 mm; particle size = 1.7 µm) at a flow rate of 0.09 mL/min, column temperature of 35 °C and sample injection volume of 5 µL. Gradient elution was utilized with the A and B solvents specified in Table 3.1 below.

Table 3.1. UPLC gradient elution profile

Time (min)	Mobile phase composition	
	0.01% formic acid in water (%A)	0.01% formic acid in methanol (%B)
0.00	99	1.0
7.00	99	1.0
7.10	95	5.0
12.00	95	5.0
26.00	82	18
35.00	50	50
35.10	2.0	98
41.00	2.0	98
41.10	99	1.0
51.00	99	1.0

An HPLC-MS method previously developed in our research group by Sharma et al.²⁶ utilized a C30 column with 0.025% formic acid in water (solvent A) and 90:10 acetonitrile–water (solvent B). To obtain better MS signal, solvent B was switched to 100% methanol for the UPLC-MS method. Additionally, formic acid concentration in the mobile phase was lowered to 0.01% to achieve separation of (i) fumaric and maleic acid isomers and (ii) methylmalonic and succinic acid isomers. In addition to run time and gradient elution profile, injection volume and column temperature were optimized for separation of 35 standard degradation products. Evaluation of these conditions can be seen in the chromatograms in Appendix C (Figures C.1-4).

High-Resolution Mass Spectrometry

HRMS analysis was performed with a Synapt time-of-flight mass spectrometer (Waters, Manchester, U.K.) operated in negative electrospray ionization mode. MS conditions were as follows: m/z range, 30 to 500; ESI capillary, 2.8 kV; sampling cone, 40.0 V; extraction cone, 4.0 V; source temperature, 120 °C; desolvation gas temperature, 350 °C; cone gas flow rate, 50 L/hr; and desolvation gas flow rate, 600 L/hr. The trap and transfer collision energies were maintained at 6.0 and 4.0 V, respectively, except during MS/MS and MS^e experiments. For MS/MS experiments, the parent ion was selected in the quadrupole and fragmented in the trap cell (at 25.0 V). During CID segments of MS^e analyses, the trap and transfer voltages were ramped from 15.0 to 40.0 V and from 4.0 to 8.0 volts, respectively. Sodium formate was used for instrument calibration and as a lock mass ion as described by Wolff et al.⁸⁴ A 10 mg/L lockspray solution was infused at 3 µL/min, and 1 s of lockspray data was collected for every 10 s of sample data. The Na(HCOO)₂[−] cluster ion (m/z 112.9851) was used as a lock mass in

post-run data processing. All theoretical masses were calculated using masses for ^{12}C , ^1H , ^{16}O , ^{14}N , and ^{32}S isotopes listed in an IUPAC technical report.⁸⁵

MetaboLynx Methodology

MetaboLynx software (see Appendix D, Figure D.1 for software features) is designed to detect metabolites from a specific parent drug or drugs in an “analyte” sample. The molecular weight or formula of the drug(s) is imported into the software, along with mass (e.g., +79.9568 for sulfate conjugation) or formula (e.g., $+\text{SO}_3$) differences for expected metabolites. The software searches for m/z ’s corresponding to the expected metabolites and generates a list of the metabolites it finds. Additionally, the software can provide a list of unexpected metabolites, any m/z over a preset intensity threshold that is not an expected metabolite and not present in a specified “control” sample (e.g., solvent or standard mixture). An illustration of these MetaboLynx processes can be seen in Figure 3.1, a plot of m/z as a function of retention time for mock control and analyte samples. The control sample (left) represents a solution of expected metabolite standards. The two m/z peaks in the control sample, X and Y, represent an expected metabolite and a non-metabolite background peak, respectively. The analyte sample (right) corresponds to a sample containing both expected and unexpected metabolites. Peaks X and Y are also seen in the analyte sample, with an additional m/z peak, Z, which is an unexpected metabolite. For the control sample, MetaboLynx will recognize X as an expected metabolite and Y as a background peak that it should exclude from further samples. For the analyte sample, MetaboLynx will (i) detect X as an expected metabolite, (ii) exclude Y because it was a non-metabolite peak that was in the control and (iii) identify Z as an unexpected metabolite.

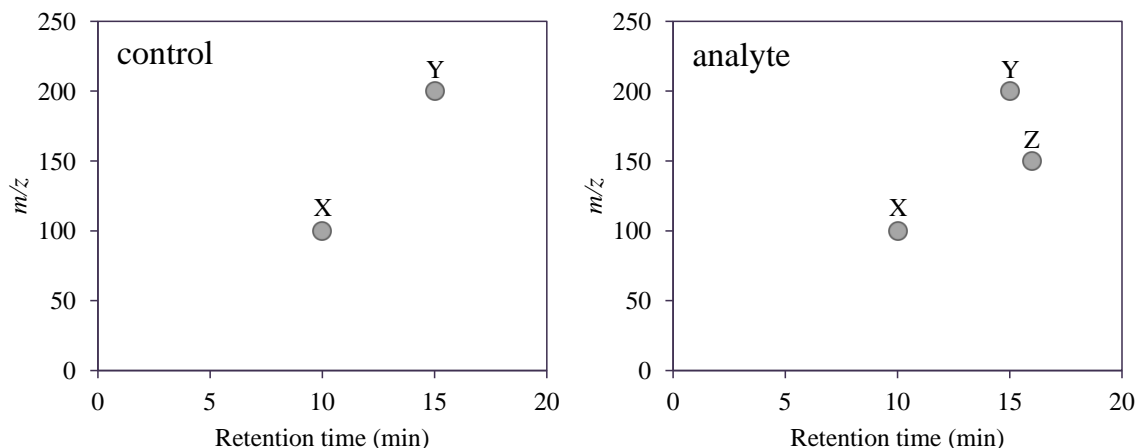


Figure 3.1. Example of the way MetaboLynx identifies expected and unexpected metabolites. The control sample on the left is a standard mixture and the analyte sample on the right is a real sample containing unexpected and expected metabolites. X, Y, and Z, represent m/z peaks of an expected metabolite, background ion, and unexpected metabolite, respectively.

In order to use MetaboLynx software to identify unknown degradation products rather than true metabolites, a parent “drug” candidate had to be formulated for lignocellulosic degradation products. I chose a parent formula of $C_{33}H_{38}O_{17}N_3S_3$, which is not a specific compound in biomass but represents the maximum number of carbon, hydrogen, and oxygen atoms (restricted as a cellulose, hemicellulose or lignin derivative) possible for m/z 500. The rationale for selection of this formula follows. First, lignocellulosic biomass is composed of cellulose (a 6-carbon carbohydrate oligomer), hemicellulose (an oligomer of 5-carbon and 6-carbon carbohydrate units) and lignin (an oligomer of aromatic sinapyl, coniferyl and *p*-coumaryl alcohol units). The chemical composition of these classes of molecules is known to consist of carbon, hydrogen and oxygen. The scope of this study was limited to m/z 500 because the most abundant ion signal in the samples of interest was below this m/z and unique chemical formulas at a given exact mass become quickly unreasonable to consider above this value given the

current resolving power/mass accuracy limitations of the instrumentation utilized. An illustration of the cellulose, hemicellulose and lignin units from which the parent formula, $C_{33}H_{38}O_{17}N_3S_3$, was specifically chosen can be seen in Figure 3.2. An oligomer for each structure type with mass greater than or equal to 500 (which allows for a conservative estimate) was considered. The maximum numbers of carbon, hydrogen, and oxygen atoms in any of the subunits in Figure 3.2 are 33, 38, and 17, respectively. The values for carbon and hydrogen were derived from a sinapyl alcohol lignin trimer and the number of oxygen atoms from the hemicellulose tetramer.

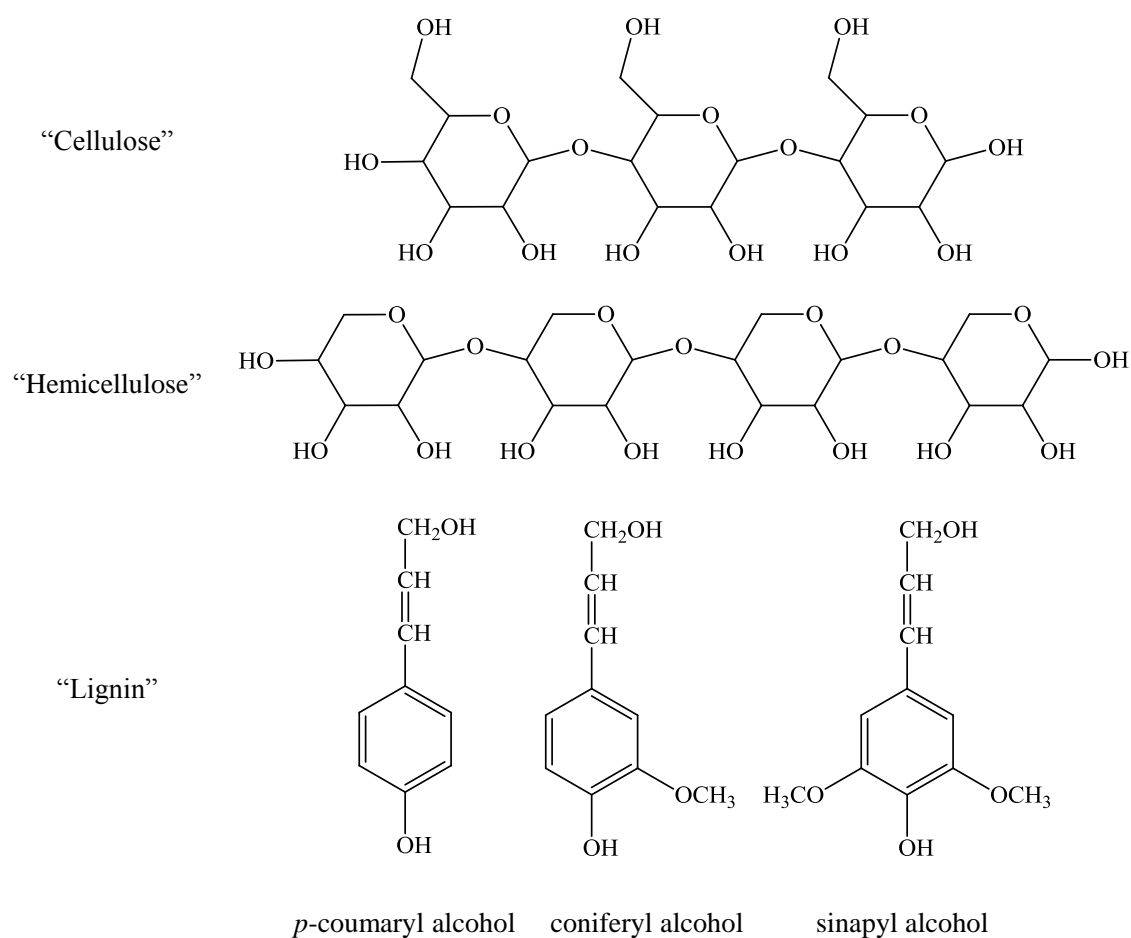


Figure 3.2. Representative figures for primary biomass components of cellulose, hemicellulose, and lignin from which a “parent” MetaboLynx formula was determined.

In addition to carbon, hydrogen, and oxygen atoms, three nitrogen atoms were allowed to account for possible reactivity during AFEX pretreatment, which has been shown to produce nitrogen-containing degradation products (e.g., amides and pyrazine derivatives).²⁷ Three sulfur atoms were also included to allow for any sulfur incorporated into dilute-sulfuric-acid-pretreated hydrolysates. In addition, due to the acidification of the extracted hydrolysates with sulfuric acid before dilution and analysis, bisulfate (HSO_4^-) adducts could be present, particularly early in the chromatographic run. Inclusion of three nitrogen and sulfur atoms allows us to look for incorporation of multiple hetero-atoms in a given structure but still limits the number of possible molecular formulas, which would become unreasonable with additional sulfur and nitrogen atoms.

The 35 standard analytes, known to be common pretreatment degradation products, were included in the MetaboLynx method as “expected metabolites”. A solution of the 35 standard degradation products at 5 mg/L was treated as the “control” sample and extracted hydrolysate samples were treated as “analyte” samples. Only compounds corresponding to peaks that were present in the analyte (hydrolysate) sample and not in the control were treated as new, unknown degradation products.

Results and Discussion

UPLC Method Development

An advantage of UPLC over HPLC is faster analysis without compromising analyte resolution. Several UPLC methods have been reported that demonstrate relatively short run times and target a high number of analytes including detection of 47

phenolic acids in 18 minutes⁹⁷, 105 veterinary drugs in a 30-minute separation⁹⁸, and 55 anabolic and androgenic steroids in equine plasma in under 5 minutes⁹⁹. A similar improvement in chromatographic run time can be achieved for the analytes typically targeted in pretreatment hydrolysates. For example, Figure 3.3 demonstrates separation and detection of 35 target analytes in under 7 minutes (for chromatographic conditions, see Appendix C, Table C.1). Note that although a number of these analytes are co-eluting, all co-eluting analytes can be resolved with mass analysis.

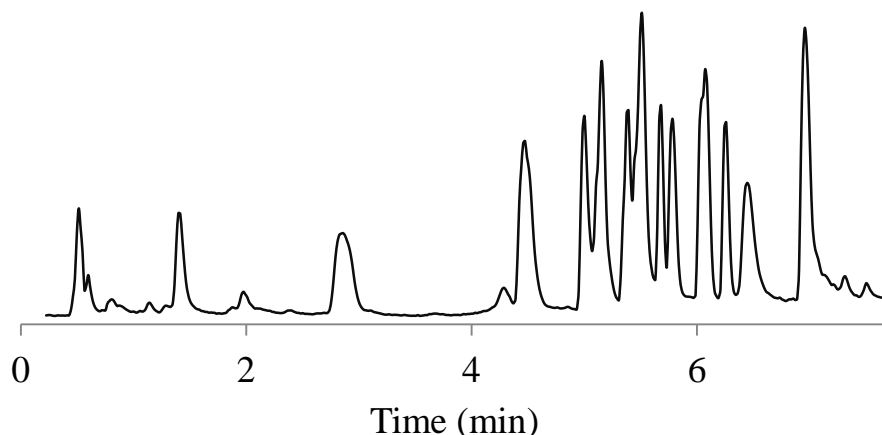


Figure 3.3. Time-optimized UPLC chromatogram of 35 common biomass degradation products.

Although ~10-fold improvement in throughput may be considered substantial for routine targeted analysis, such compressed chromatograms are not ideal for detection of unknowns. Several reports^{100,101} have demonstrated co-eluting components can unpredictably effect ionization efficiencies to the extent of completely suppressing ion signal. In addition, co-elution will result in increased spectral complexity at a given run-time as well as a lower probability of resolving unknown compounds that are isomers or structurally similar. Accordingly, conditions that provide improved chromatographic separation of sample components would be most desirable for detection of unknowns

existing in a complex matrix such as a biomass pretreatment hydrolysate. The UPLC method we utilized gave comparable, if not better, separation than what was achieved previously with HPLC. For example, 13 of the 35 analytes were chromatographically resolved in 34 min with UPLC compared to 11 analytes in 65 min with the HPLC method.²⁶ A ~5-fold reduction in plate height was also observed (e.g., 0.327 mm and 1.79 mm for caffeic acid with UPLC and HPLC, respectively). Although higher resolution could have been obtained with longer run times, optimal conditions were sought that could achieve separation in less than an hour.

The method was also applied to standard AFEX- and dilute-acid-pretreatment samples (Figure 3.4, middle and bottom chromatograms). These chromatograms reveal several abundant peaks (e.g., peak A) not observed in the standard as well as less intense peaks (e.g., peak B) spread across a majority of the chromatographic retention time. A higher abundance of ion signal is observed in the 30-35 min region of the hydrolysate samples, evidence of the tradeoff between run time and chromatographic resolution.

Evaluation of HR-TOF-MS Mass Spectral Data

To evaluate the level of mass accuracy achieved with the HR-TOF-MS instrumentation utilized in this study, a 5 mg/L standard solution containing 35 known degradation products was analyzed. Manual processing, which involves combining mass spectra over the chromatographic peak of each analyte and then smoothing, centering, and applying a lockmass correction to the mass spectral data, was performed for each of the 35 analytes to obtain exact masses. Table 3.2 contains retention times, exact masses and mass errors (ppm) for all 35 analytes. Errors in the measured experimental masses were generally less than 10 ppm. However, five compounds had large errors (>10 ppm).

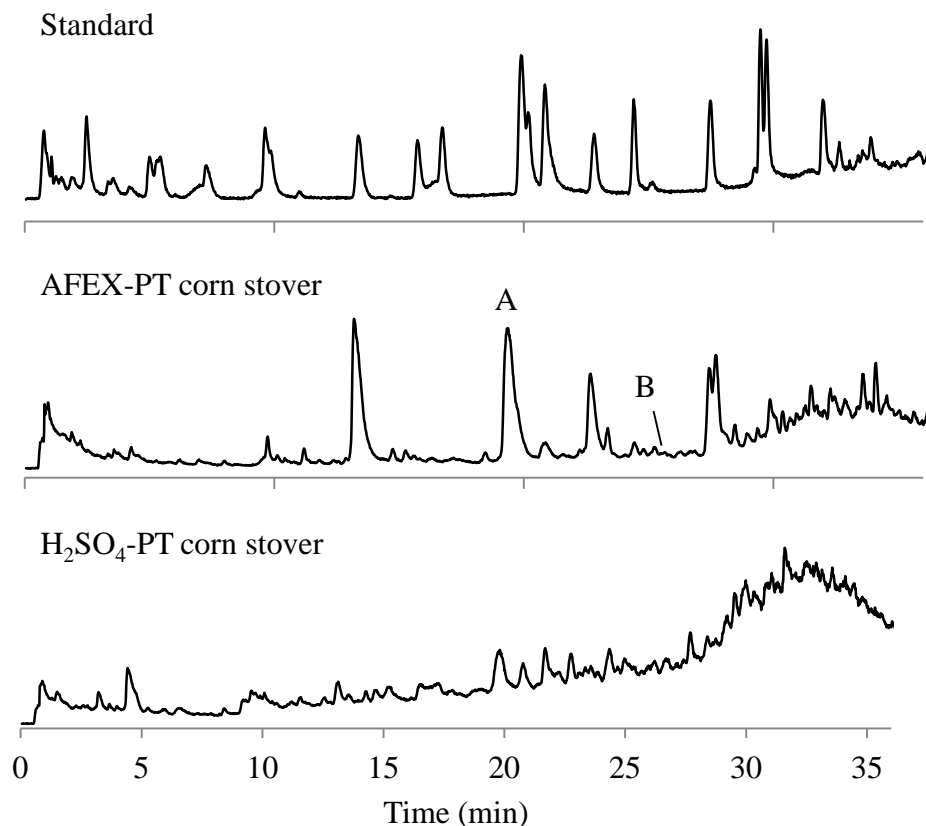


Figure 3.4. UPLC chromatograms, resolution-optimized for 35 known biomass degradation products in a 5 ppm mixture (top), AFEX-pretreated corn stover hydrolysate (middle), and dilute H₂SO₄-pretreated corn stover hydrolysate (bottom).

The signal observed for these compounds was generally very strong and these peaks were determined to be a result of the detector operating in dead time,¹⁰² as evidenced by asymmetric m/z peaks and exact masses shifted to values much lower than expected. All errors in exact mass for these compounds improved after decreasing sample concentration (rightmost column in Table 3.2). Note that although malonic acid had the largest error at -62.1 ppm, the error was large at all concentrations analyzed (0.5, 1.0, 5.0, 10, and 15 ppm). After a closer look at the mass spectrum around m/z 103, it was determined that poor m/z peak purity due to co-eluting species at a very similar m/z was

Table 3.2. Manually-generated HR-TOF-MS results for the 35 targeted analytes in a 5 mg/L standard solution and in a 1 or 10 mg/L for those ions with poor error at 5 mg/L.

Analyte	Time (min)	<i>m/z</i>	Error (ppm) 5 mg/L std	Error (ppm) 10 ^a or 1 ^b mg/L std
malonic acid ^a	0.78	102.9978	-51.5	-62.1
lactic acid	1.26	89.0234	-5.6	
succinic acid	1.21	117.0188	-0.9	
fumaric acid	1.47	115.0031	0.0	
maleic acid	1.78	115.0031	0.0	
methylmalonic acid	1.88	117.0186	-1.7	
trans/cis-aconitic acid	2.13	173.0041	-1.2	
gallic acid	2.49	169.0137	-3.6	
levulinic acid	3.36	115.0395	-1.7	
glutaric acid	3.55	131.0304	-3.1	
itaconic acid	4.20	129.0186	-1.6	
2-hydroxy-2-methylbutyric	5.00	117.0544	-6.8	
2-furoic acid	5.43	111.0082	0.0	
3,4-dihydroxybenzoic acid	5.45	153.0182	-3.9	
3,4-dihydroxybenzaldehyde ^b	9.43	137.0238	-14.6	11.7
4-hydroxybenzoic acid	9.63	137.0227	-8.8	
2,5-hydroxybenzoic acid	9.63	153.0191	-2.6	
adipic acid	9.89	145.0494	-4.8	
4-hydroxybenzaldehyde ^b	13.39	121.0252	-33.9	-5.8
vanillic acid	15.72	167.0340	-2.4	
caffeic acid	16.47	179.0332	-6.7	
vanillin	19.86	151.0401	4.0	
4-hydroxyacetophenone ^b	19.92	135.0393	-39.6	-2.2
syringic acid	20.84	197.0448	-1.0	
homovanillic acid	20.84	181.0497	-2.2	
salicylic acid	20.86	137.0240	0.7	
4-hydroxycoumaric acid	22.82	163.0395	-5.5	
syringaldehyde	24.50	181.0495	-3.3	
benzoic acid	25.15	121.0291	0.8	
sinapic acid	27.47	223.0594	-5.4	
ferulic acid	27.49	193.0493	-4.1	
3-hydroxy-4-	29.72	193.0487	-3.6	
4-hydroxycoumarin ^b	32.00	161.0178	-37.9	10.6
<i>o</i> -toluic acid	32.64	135.0445	-0.7	
<i>p</i> -toluic acid	33.92	135.0443	-2.2	

^a Data in right column is from a 10 mg/L standard solution.

^b Data in right column is from a 1 mg/L standard solution.

responsible for the large error. For this reason, malonic acid was removed as one of the target analytes from further analyses.

A “Metabolomics” Approach to Unknown Component Identification

Development of MetaboLynx methodology. HR-TOF-MS provides such a high degree of information that it can be difficult to differentiate m/z peaks of interest from background or erroneous peaks when you are performing an untargeted analysis. For example, mass spectra corresponding to a retention time of 21.1 minutes for a water blank (i.e., 5- μ L injection of purified water) and a representative hydrolysate can be seen in Figure 3.5. The water blank mass spectrum exhibits such complexity that trying to

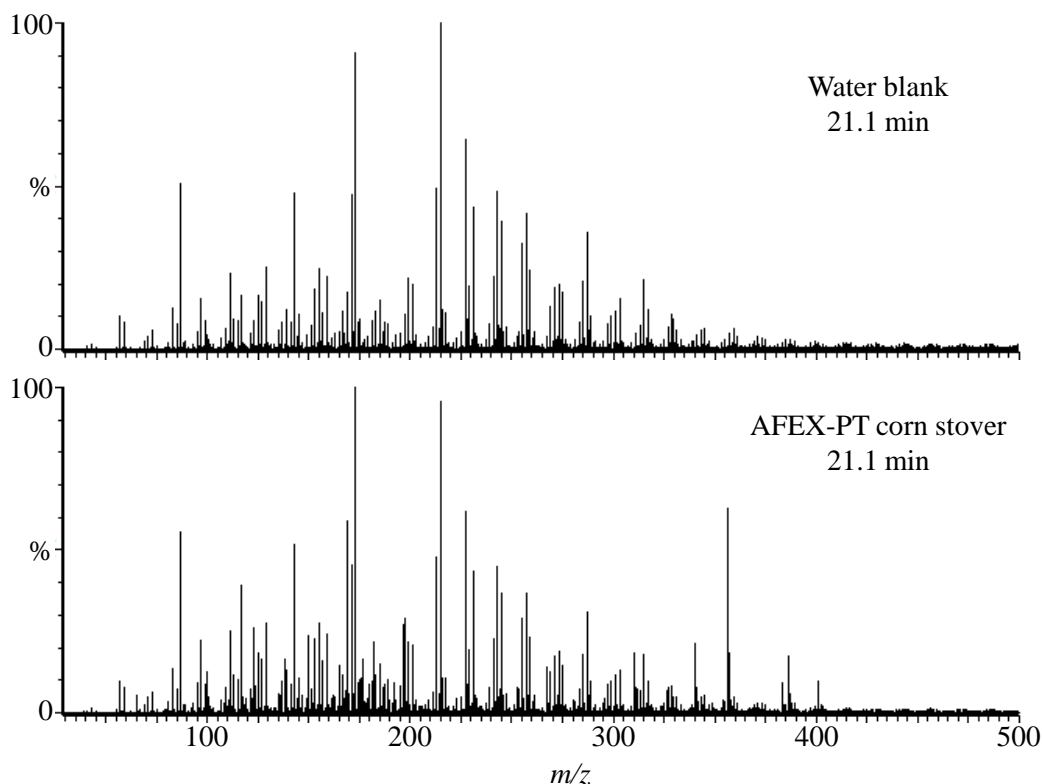


Figure 3.5. Mass spectra for a water blank and an AFEX-pretreated hydrolysate of corn stover at 21.1 minutes.

identify all peaks unique to the hydrolysate is quite difficult. Beyond the obviously new m/z peaks in the 300-400 m/z , visual identification of less obvious differences between the two spectra requires much more effort. Thus, automation for peak identification (e.g., via MetaboLynx) is essential in processing the high level of information obtained with HR-TOF-MS.

The initial MetaboLynx method outlined in the experimental section was first evaluated to ensure that the software could successfully (i) detect expected “metabolites” (aka, standard degradation products) in a sample with reasonable error in exact mass, (ii) identify background peaks in a control and exclude them from an analyte sample, and (iii) identify unexpected “metabolites”. A 5 mg/L solution containing 35 known degradation products was utilized to optimize MetaboLynx parameters to meet these three expectations. A complete list of MetaboLynx parameters and experiments for optimization can be found in Appendix D.

The final step in method development was to evaluate the accuracy of automated processing and identify intensity thresholds for which accurate mass data can be obtained. To achieve this, spectral data was smoothed and centered, and a lockmass correction applied as part of the MetaboLynx software processing, enabling exact masses to be reported automatically for the expected analytes. Errors in the measured experimental masses were generally less than 5 ppm. However, seven compounds had large errors (>15 ppm). The signal observed for these compounds were either very weak (<50 area units; S/N = 9-13) or very strong. Below 50 area units, mass measurements were less reliable, which may be related to greater noise in this intensity range. As described before, intense peaks were determined to be a result of the detector operating in

dead time.¹⁰² All errors in exact mass improved after increasing or decreasing sample concentration, for the weak and intense peaks, respectively. Table 3.3 contains exact masses, mass errors (ppm), and absolute chromatographic peak areas for all 35 analytes at the appropriate concentration. Although this method and the manual method described above for HR-TOF-MS evaluation utilized the same data file, the differences in experimental exact masses may arise due to a deficiency in the proprietary algorithm used to obtain centered m/z data. Based on the large or variable error for low and high area units, a middle area unit range of 200-3500 was used to restrict further data sets. Although exclusion of abundant ion signal may not be desirable from a compositional point of view, this work focuses on the approach rather than sample composition.

Automated detection of unknown peaks in hydrolysate. The developed MetaboLynx method was then utilized to identify unknown degradation products in dilute-acid- and AFEX- pretreated hydrolysates of corn stover. Corn stover is one of the most widely studied lignocellulosic feedstocks. Dilute-acid and AFEX pretreatments are leading candidates for commercial production and represent two pH extremes. Variation in pretreatment pH has been shown to produce differing hydrolysate compositions.²⁸ Thus, AFEX- and dilute-acid- pretreated samples should provide diversity in sample composition, allowing us to look at molecular level differences between the two samples.

The MetaboLynx software initially found 3870 and 3912 unexpected m/z peaks (which will be called “new peaks” in the remainder of the chapter) in the AFEX and dilute-acid samples, respectively, indicating similar sample complexities. First, new peaks with a negative mass defect larger than 0.01 Da were rejected because they were considered unlikely to come from biomass. Ions resulting from sulfuric acid (used to

Table 3.3. MetaboLynx-generated HR-TOF-MS results for the 34 targeted analytes in a 5 ppm standard solution, unless otherwise noted.

Analyte	<i>m/z</i>	Error (ppm)	Peak Area
lactic acid	89.0233	-6.0	9
succinic	117.0146	5.6	82
fumaric acid	115.0030	-0.3	39
maleic acid	115.0030	-0.3	39
methylmalonic acid ^a	117.0184	-2.9	115
trans/cis-aconitic acid ^a	173.0083	-1.4	78
gallic acid	169.0131	-3.2	895
levulinic acid	115.0393	-1.2	98
glutaric acid	131.0304	-2.6	324
itaconic acid	129.0184	-2.6	22
2-hydroxy-2-methylbutyric acid	117.0544	-6.2	1551
2-furoic acid	111.0081	-0.3	121
3,4-dihydroxybenzoic acid	153.0178	-6.1	2463
3,4-dihydroxybenzaldehyde	137.0238	-0.3	483
4-hydroxybenzoic acid ^b	137.0233	-3.9	1043
2,5-hydroxybenzoic acid	153.0191	2.4	158
adipic acid	145.0496	-3.0	902
4-hydroxybenzaldehyde ^b	121.0252	-6.0	2298
vanillic acid	167.0340	-2.0	855
caffeic acid	179.0342	-0.8	391
vanillin	151.0400	3.7	584
4-hydroxyacetophenone ^b	135.0391	-11.3	3036
syringic acid	197.0441	-4.2	831
homovanillic acid	181.0497	-1.9	89
salicylic acid	137.0229	-6.8	3661
4-hydroxycoumaric acid	163.0387	-4.5	1769
syringaldehyde	181.0501	-3.9	1101
benzoic acid	121.0291	1.3	288
sinapic acid	223.0610	2.1	111
ferulic acid	193.0493	-3.8	1728
3-hydroxy-4-methoxycinnamic acid	193.0487	-6.9	1849
4-hydroxycoumarin ^b	161.0178	-10.1	2494
<i>o</i> -toluic acid	135.0439	-4.7	879
<i>p</i> -toluic acid	135.0443	-1.7	833

^a Data is from a 10 ppm standard solution due to weak signal at 5 ppm.

^b Data is from a 1 ppm standard solution due to ion being in TOF dead time at 5 ppm.

acidify the extracted hydrolysate), formic acid (in the mobile phase), and sodium accounted for several of these peaks with negative defects. Biomass pretreatment products extractable into MTBE should be composed of mostly C, H, and O atoms, which do not contribute to a significant negative mass defect (less than -0.01 Da). Additional application of the intensity restriction promoting mass accuracy specified above (200-3500 area units) resulted in 209 and 746 new peaks for AFEX and dilute-acid samples, respectively, in which we were confident in assigning possible molecular formulas. All molecular formula possibilities of the form $C_vH_wO_xN_yS_z$ with theoretical masses within 15 ppm of the experimental masses were carried forward at this stage.

Compositional Analysis from MetaboLynx-Generated Mass Spectral Data

Several visual representations of the MetaboLynx data were generated to elucidate bulk differences or trends in composition between the two pretreatment samples. First, there was a noticeable difference between the m/z distributions of the samples. Figure 3.6 shows the percentage of new peaks in specified m/z ranges for each sample. As can be seen in the figure, the dilute-acid hydrolysate sample has a lower m/z distribution, that is, a higher percentage of new compounds in the lower middle m/z ranges (150-250 m/z) than the AFEX sample. The new compounds of the AFEX sample are more evenly distributed across the middle m/z ranges (150-400 m/z) than the dilute-acid sample. One possible explanation for this result is that the dilute-acid pretreatment breaks down biomass into smaller components more readily than AFEX pretreatment.

New m/z peaks, as a percentage of the total of new peaks, were also plotted as a function of retention time range (Figure 3.7). For both samples, the highest percentage of

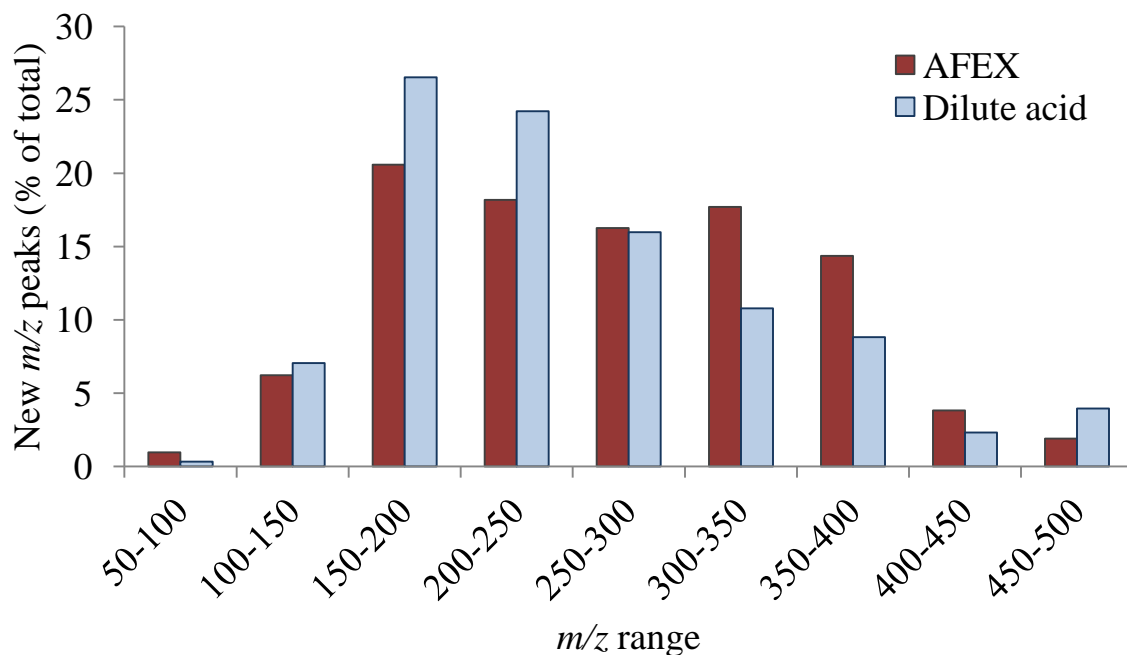


Figure 3.6. M/z distribution of unknown peaks, expressed as a percentage of the total peaks found in the 200-3500 intensity range, for the AFEX and dilute-acid-pretreated hydrolysate samples.

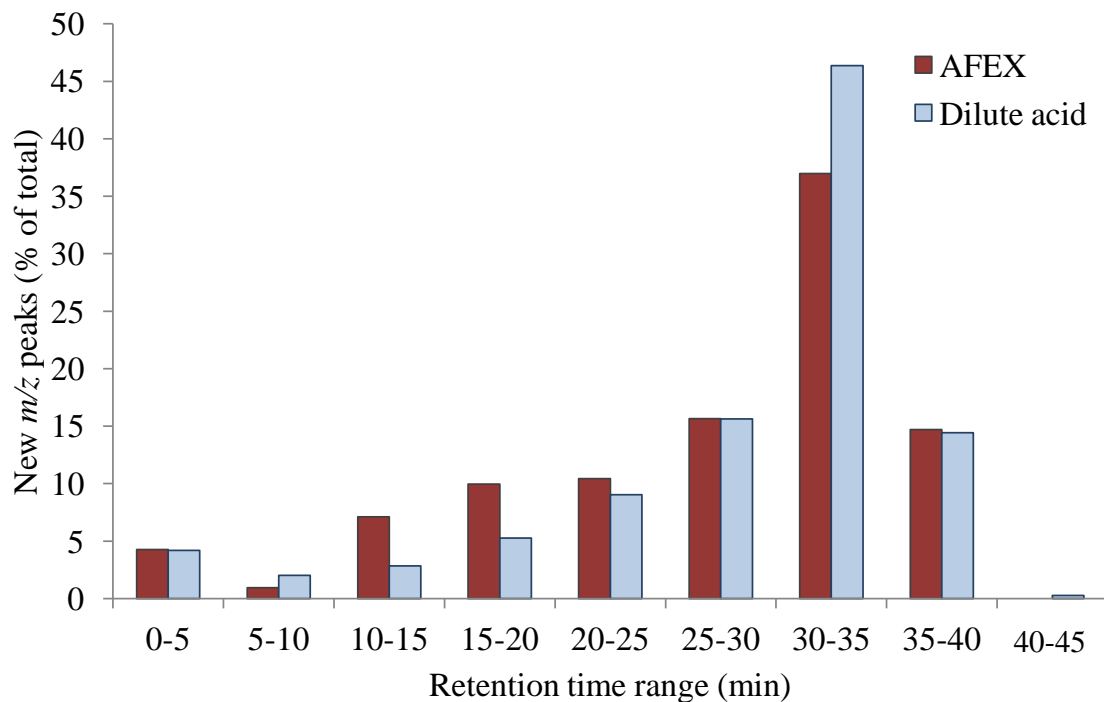


Figure 3.7. Retention time distribution of unknown peaks, expressed as a percentage of the total peaks found in the 200-3500 intensity range, for the AFEX and dilute-acid-pretreated hydrolysate samples.

new peaks eluted in the 30-35-minute region, which is consistent with the chromatograms shown in Figure 3.4. The AFEX sample had a higher percentage of peaks eluting in the middle of the chromatographic run (10-25 minutes) than the dilute-acid sample, suggesting the AFEX sample contains a slightly higher percentage of more polar compounds than the dilute-acid sample. In addition, more ion signal at longer retention times for both sample types suggests that a greater number of degradation products may be derived from lignin than from carbohydrates. Although higher abundance does not necessarily indicate potency in the biomass conversion process, higher ion signal at longer retention times does present a reason to look more closely at lignin-derived products.

Mass defects (exact mass – nominal mass) for m/z peaks can also reveal differences in compound type between the two samples. A majority of biomass degradation products contain only carbon (12.0000 Da), hydrogen (1.00783 Da) and oxygen (15.99491 Da) so for a given m/z , a lower mass defect implies higher O/C and/or lower H/C content and vice versa. Nitrogen (14.0031 Da) and sulfur (31.9721 Da) atoms can also be present, contributing to more positive and negative defects, respectively. Due to the presence of more oxygen atoms, carbohydrate-derived degradation products from cellulose or hemicellulose (e.g., levulinic acid and 2-furic acid) are generally expected to have a lower mass defect than lignin-derived degradation products (e.g., 4-hydroxybenzoic acid, ferulic acid, etc.). Figure 3.8 shows the mass defect for new m/z peaks as a function of nominal m/z for both hydrolysate samples. Overall, the observed trend is as expected, as m/z increases, mass defect increases. Some data points fall above and below the general trend, indicating that their molecular compositions diverge from

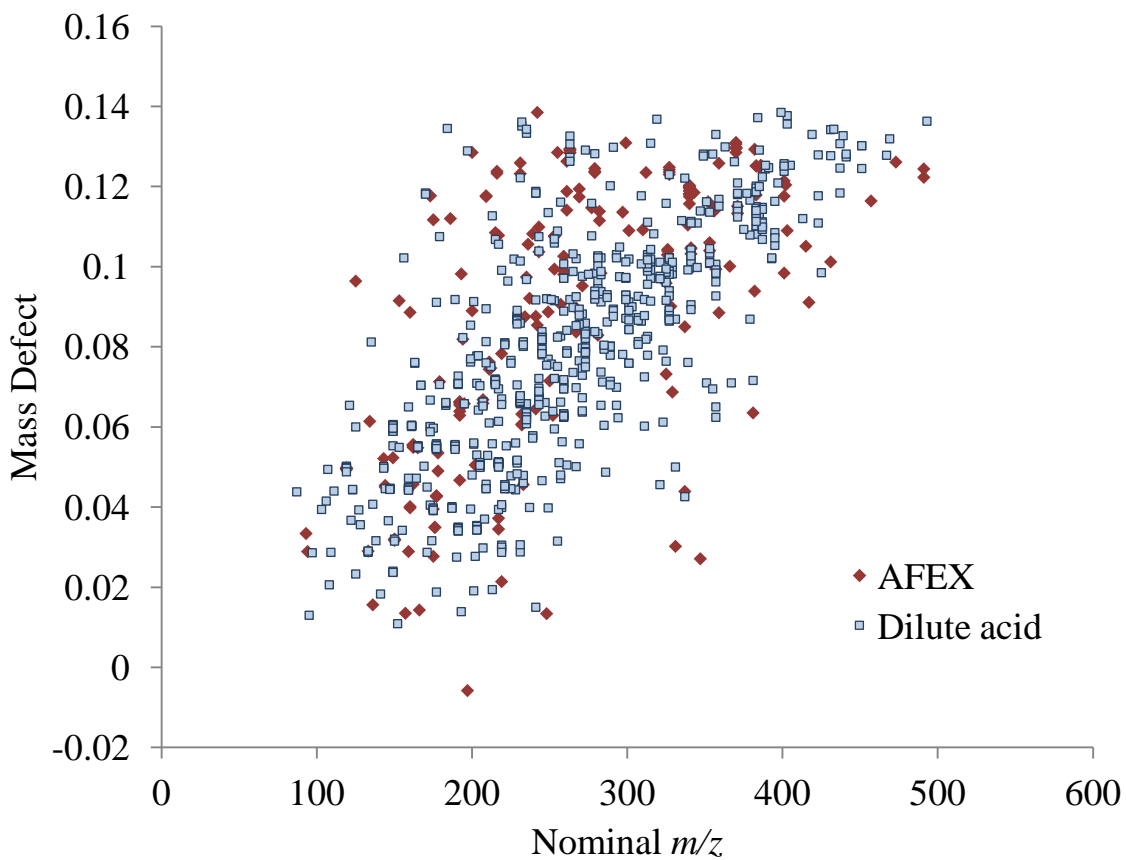


Figure 3.8. Mass defect (exact mass – nominal mass) as a function of nominal mass for new m/z peaks (200-3500 area units) in the AFEX and dilute-acid hydrolysates.

those of most components. The most significant observation is that there appears to be no obvious difference in correlation between the AFEX and dilute-acid samples, suggesting that bulk compound types are probably similar between the two samples.

Correlations between mass defect and nominal m/z for different retention time ranges can be seen in Figure 3.9 for the AFEX and dilute-acid samples. In the AFEX sample, a majority of the early-eluting ions (0-10 minutes, circled in the figure) have mass defects lower than the trend, signifying that they have less hydrogen and/or more oxygen relative to carbon. These compounds may be more polar and thus, elute earlier on a reversed-

phase column. Similar observations cannot be made for the dilute-acid sample; no significant deviations from the general trend can be discerned.

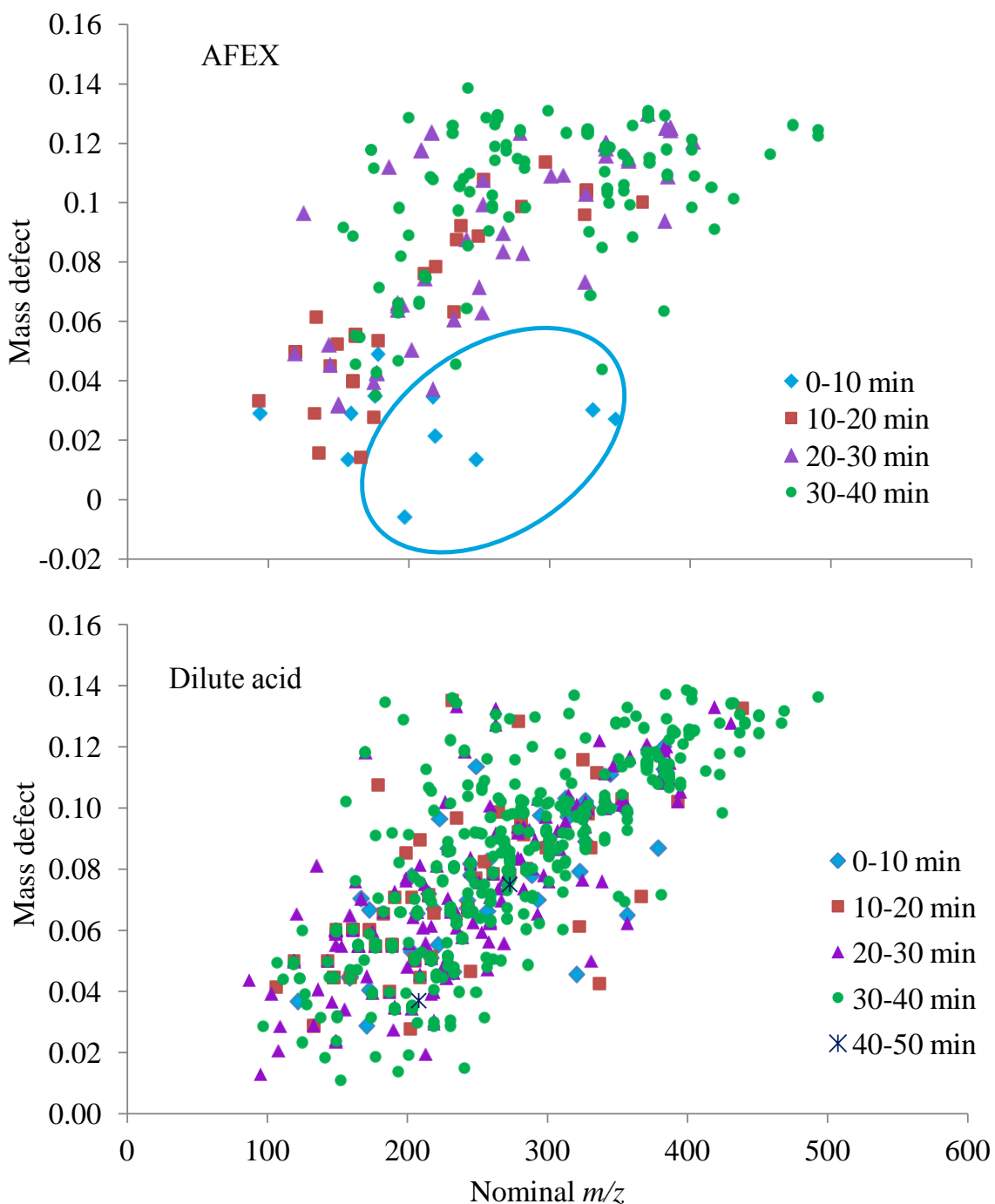


Figure 3.9. Mass defect as a function of nominal mass for peaks in AFEX and dilute-acid hydrolysate sample at given 10-minute retention time ranges. Early-eluting peaks with a lower mass defect than the general trend are circled for the AFEX sample.

Kendrick Mass Defect Analyses for Identification of Functional Groups

In order to easily classify the new compounds by their functional groups, Kendrick mass defect (KMD) plots were prepared for OH, OCH₃, CO₂, and CH₂ groups, which were chosen because they are present in known cellulose, hemicellulose and lignin degradation products. Members of a homologous series (molecules that differ only in their number of a certain functional group) can be identified in a KMD plot. In addition, molecular formula identification for a lower MW member of a KMD series allows for unambiguous assignment of molecular formula to higher MW members of that same series.¹⁰³ These plots were prepared by first converting m/z 's to Kendrick masses for each peak. Experimental m/z 's were multiplied by the nominal mass of the functional group and divided by the exact mass of the functional group. For example, Kendrick mass for the functional group OH (which corresponds to an exchange of "OH" for "H" in a molecule and thus, molecules differing in their number of OH groups will differ by their number of "O" atoms) can be calculated as follows for the syringaldehyde ion (m/z 181.0501):

$$\text{Kendrick mass} = 181.0501 \times \frac{16}{15.99491} = 181.1077$$

This calculation converts the mass of O from its IUPAC mass of 15.99491 to its nominal mass of exactly 16. Subtracting the exact Kendrick mass from the nominal mass of the Kendrick mass gives the Kendrick mass defect.

$$\text{Kendrick mass defect} = 181 - 181.1077 = -0.1077$$

In a plot of Kendrick mass defect versus Kendrick mass, ions of a homologous series appear on the same horizontal line. KMD analyses are normally performed with ultrahigh-resolving-power Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, which allows for high mass accuracy and KMDs with very low relative range (i.e., horizontal lines (slope = 0) connecting members of a homologous series in a KMD plot). Although KMD analyses in the present study were achieved with HR-TOF-MS, which does not have the same level of resolving power and mass accuracy, many series in the KMD plots described below exhibit fairly horizontal lines. However, some do have a non-zero slope, which is to be expected based on the mass accuracy of the TOF mass spectrometer. To evaluate the accuracy of KMDs using HR-TOF-MS, KMD plots for OH, OCH₃, CO₂, and CH₂ were prepared for the 5 mg/L standard mixture (Figure 3.10). All analytes of a homologous series are denoted by connecting dotted lines in Figure 3.10. Tables 3.4-7 show the results of these plots for each functional group. In general, the Kendrick mass defects were very similar for related analytes, with most relative ranges less than 2% (calculated as the difference between the highest and lowest KMD values for each set). In cases where the direction (positive/negative) or magnitude of the exact mass error varied across the analytes, the KMDs also varied noticeably, with relative ranges of 3-7%.

It should be noted that the same mass defect for a set of analytes does not always indicate that the analytes are related by their number of the specific functional group (OH's, OCH₃'s, etc.). The set of atoms in the functional group can be incorporated into the molecule in a different manner. An example can be seen in Figure 3.11, adipic acid (C₆H₁₀O₄) has one additional oxygen, one additional carbon, and two additional

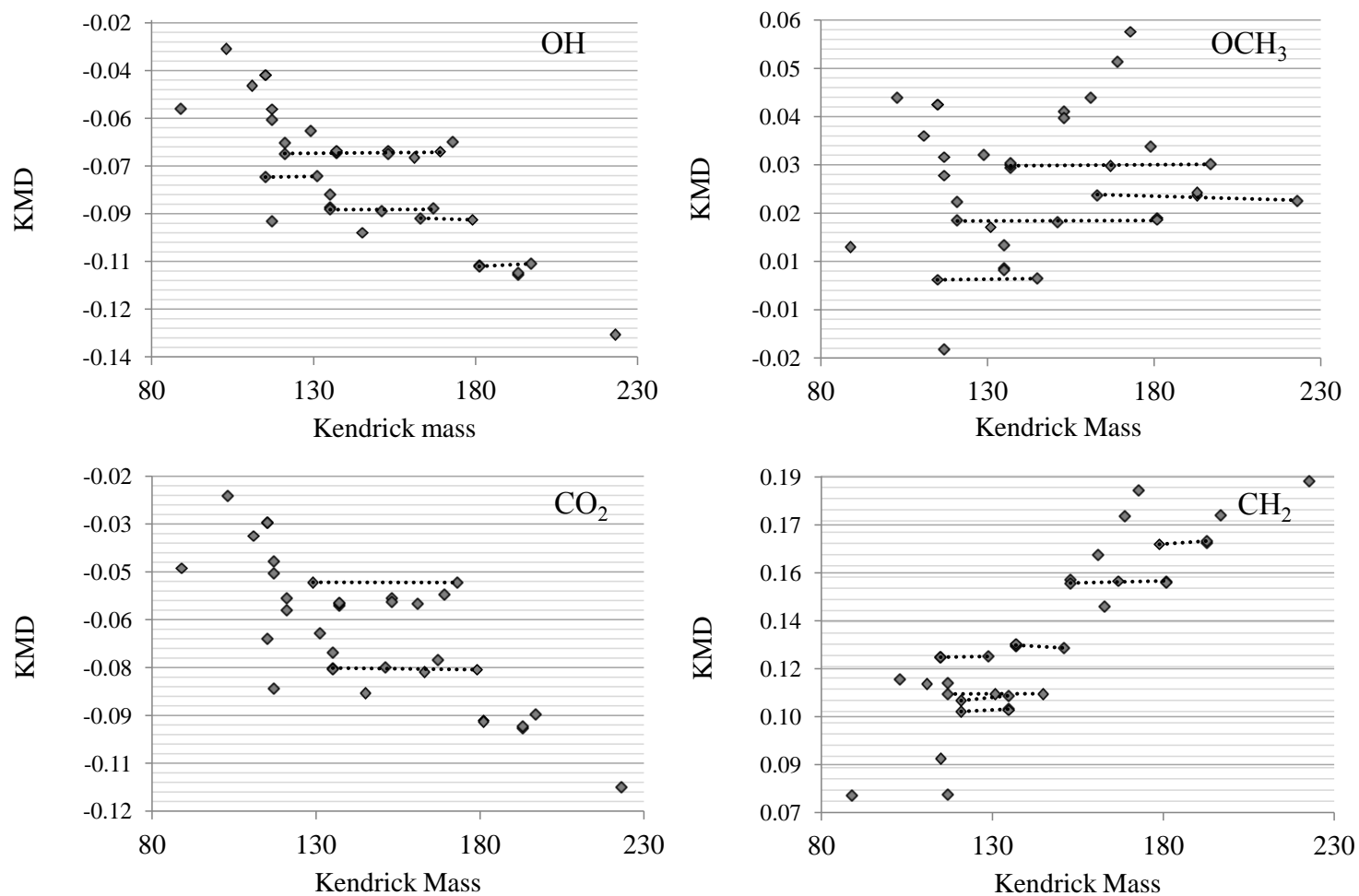


Figure 3.10. Kendrick mass defect plots for four functional groups (-OH, -OCH₃, CO₂, and CH₂) for the target analytes in a 5 ppm standard mixture. Analytes that differ by the number of functional groups are denoted by dotted lines.

Table 3.4. Kendrick mass defect (KMD) results for the target analytes differing by an “O”. Relative ranges (%) are given for the KMD values.

# of addt'l O's	analyte	exact mass	<i>m/z</i>	error (ppm)	KMD	relative range, %
-	benzoic acid	121.0290	121.0291	0.8	-0.0676	1.35
1	OHbenzoic acid	137.0239	137.0233	-4.4	-0.0669	
2	diOHbenzoic acid	153.0188	153.0191	2.0	-0.0678	
3	gallic acid	169.0137	169.0131	-3.6	-0.0669	
-	syringaldehyde	181.0501	181.0501	0.0	-0.1077	0.85
1	syringic acid	197.0450	197.0441	-4.6	-0.1068	
-	toluic acid	135.0446	135.0443	-2.2	-0.0873	1.04
1	vanillin	151.0395	151.0400	3.3	-0.0881	
2	vanillic acid	167.0344	167.0340	-2.4	-0.0872	
-	levulinic acid	115.0395	115.0393	-1.7	-0.0759	0.28
1	glutaric acid	131.0344	131.0340	-3.1	-0.0757	
-	4-hydroxycoumaric acid	163.0395	163.0387	-4.9	-0.0906	0.65
1	caffeic acid	179.0344	179.0342	-1.1	-0.0912	

Table 3.5. Kendrick mass defect (KMD) results for the target analytes differing by an “OCH₂”. Relative ranges (%) are given for the KMD values.

# of addt'l OCH ₂ 's	analyte	exact mass	<i>m/z</i>	error (ppm)	KMD	relative range, %
-	4-hydroxybenzoic acid	137.0239	137.0233	-4.4	0.0249	1.84
1	vanillic acid	167.0344	167.034	-2.4	0.0248	
2	syringic acid	197.0450	197.0441	-4.6	0.0252	
-	benzoic acid	121.0290	121.0291	0.8	0.0135	3.43
1	vanillin	151.0395	151.0400	3.3	0.0132	
2	syringaldehyde	181.0501	181.0501	0.0	0.0136	
-	4-hydroxycoumaric acid	163.0395	163.0387	-4.9	0.0187	6.46
1	ferulic acid	193.0501	193.0493	-4.1	0.0186	
2	sinapic acid	223.0606	223.0610	1.8	0.0175	
-	levulinic acid	115.0395	115.0393	-1.7	0.0012	19.85
1	adipic acid	145.0501	145.0496	-3.4	0.0014	

Table 3.6. Kendrick mass defect (KMD) results for the target analytes differing by a “CH₂”. Relative ranges (%) are given for the KMD values.

# of addt'l CH ₂ 's	analyte	exact mass	<i>m/z</i>	error (ppm)	KMD	relative range, %
-	malonic acid	103.0031	102.9974	-55.3	0.1176	4.70
1	succinic	117.0188	117.0184	-3.4	0.1123	
2	glutaric acid	131.0344	131.0340	-3.1	0.1123	
3	adipic acid	145.0501	145.0496	-3.4	0.1124	
-	dihydroxybenzoic	153.0188	153.0191	2.0	0.1517	0.49
1	vanillic acid	167.0344	167.0340	-2.4	0.1525	
2	homovanillic acid	181.0501	181.0497	-2.2	0.1525	
-	caffeic acid	179.0344	179.0342	-1.1	0.1657	0.69
1	ferulic acid	193.0501	193.0487	-7.3	0.1669	
-	4-hydroxybenzoic acid	137.0239	137.0233	-4.4	0.1297	0.81
1	vanillin	151.0395	151.0400	3.3	0.1287	
-	fumaric acid	115.0031	115.0030	-0.9	0.1254	0.20
1	glutaric acid	129.0188	129.0184	-3.1	0.1257	
-	benzoic acid	121.029	121.0291	0.8	0.1060	0.80
1	o-toluic acid	135.0446	135.0439	-5.2	0.1069	
-	4-hydroxybenzaldehyde	121.0290	121.0252	-31.4	0.1099	1.58
1	4-hydroxyacetophenone	135.0446	135.0391	-40.7	0.1117	

Table 3.7. Kendrick mass defect (KMD) results for the target analytes differing by a “CO₂”. Relative ranges (%) are given for the KMD values.

# of addt'l CO ₂ 's	analyte	exact mass	<i>m/z</i>	error (ppm)	KMD	relative range, %
-	itaconic acid	129.0188	129.0184	-3.1	-0.0483	0.17
1	aconitic acid	173.0086	173.0083	-1.7	-0.0483	
-	o/p-toluic acid	135.0446	135.0439	-5.2	-0.0752	0.64
1	caffeic acid	179.0344	179.0342	-1.1	-0.0756	

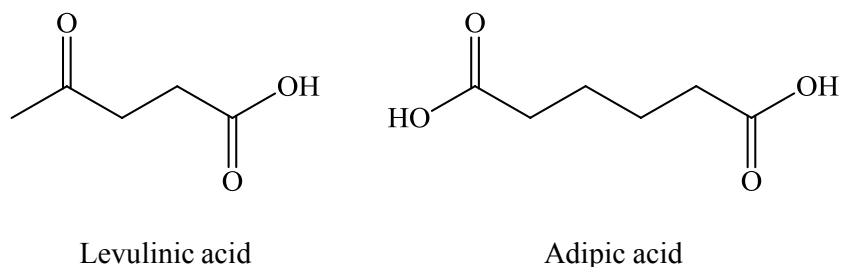


Figure 3.11. Structures for levulinic and adipic acids, which differ by “OCH₂” but do not contain a single methoxy group.

hydrogen atoms when compared to levulinic acid (C₅H₈O₃), which gives them the same KMD value for the methoxy functional group “OCH₃”, even though neither of them has a methoxy functional group. This type of situation is one caveat in using KMD to identify functional groups.

KMD analyses were then carried out for the AFEX sample (Figures 3.12-15).

Criteria for identification of homologous series were (i) KMD differences less than 7% (based on what was observed in the standard solution) and (ii) m/z differences corresponding to the specified functional groups. Series for target analytes and new peaks are designated by blue, solid lines and green, dashed lines, respectively, in Figures 3.12-15 and KMD results are presented in Tables 3.8-11. Overall, 17, 8, 13, and 14 sets of peaks were identified as homologous series of OH, OCH₃, CO₂, and CH₂, respectively. It is interesting that all of the series found by KMD in the AFEX sample contain only two or three members, whereas KMD plots of petroleomics samples, for example, contain series with greater than 50 members.^{103,104} Although more than a few members in a homologous series would not be expected in this work due to the 500 m/z limit, it is unclear if the level of mass accuracy is preventing further identification of members (within the imposed 7% relative range restriction).

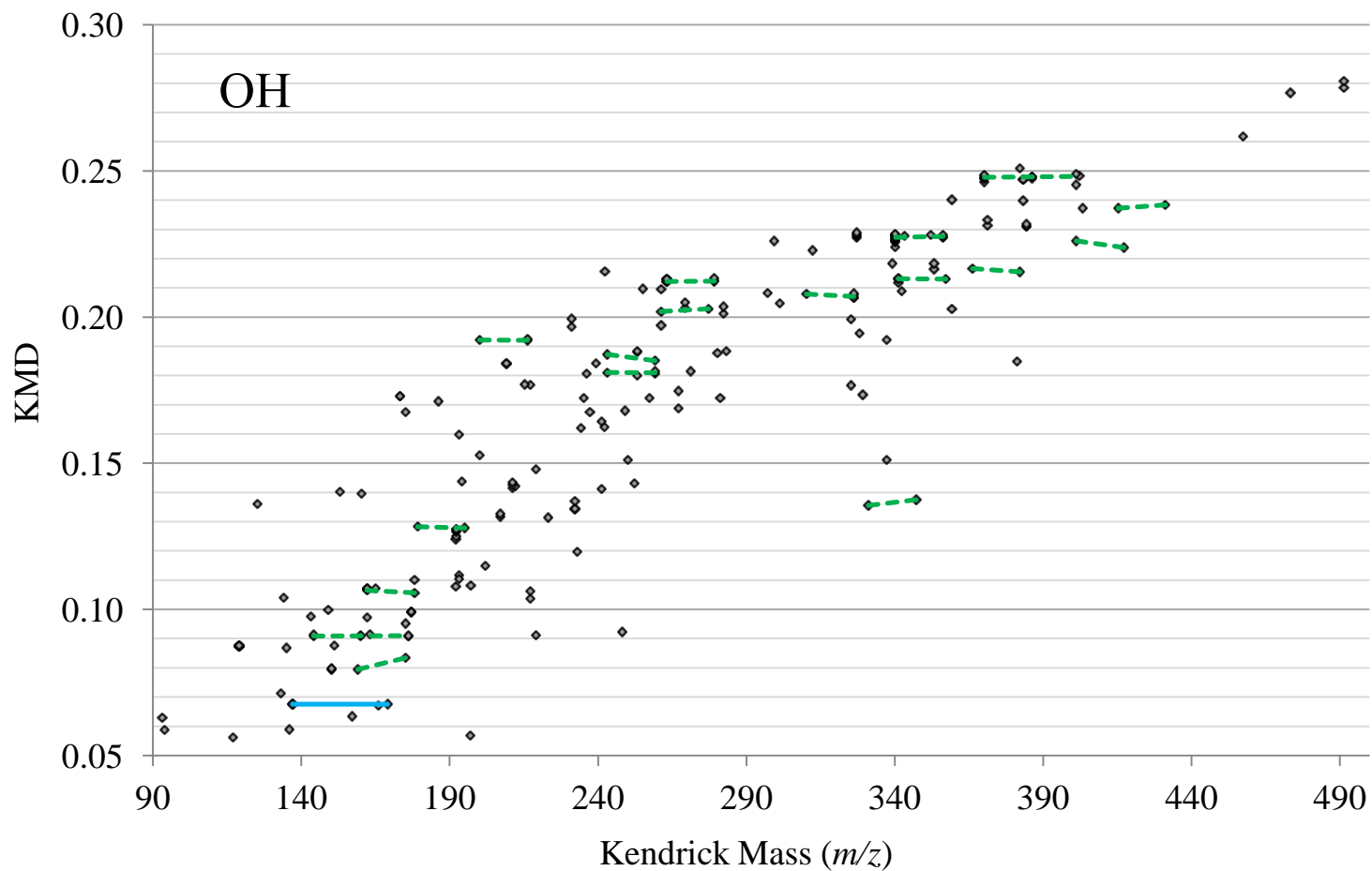


Figure 3.12. Kendrick mass defect plot for $-OH$ for the new m/z peaks (200-3500 area units) and target analytes in the AFEX sample. Ions that differ by the number of functional groups are denoted by blue solid lines for the target analytes and green dotted lines for the new m/z peaks.

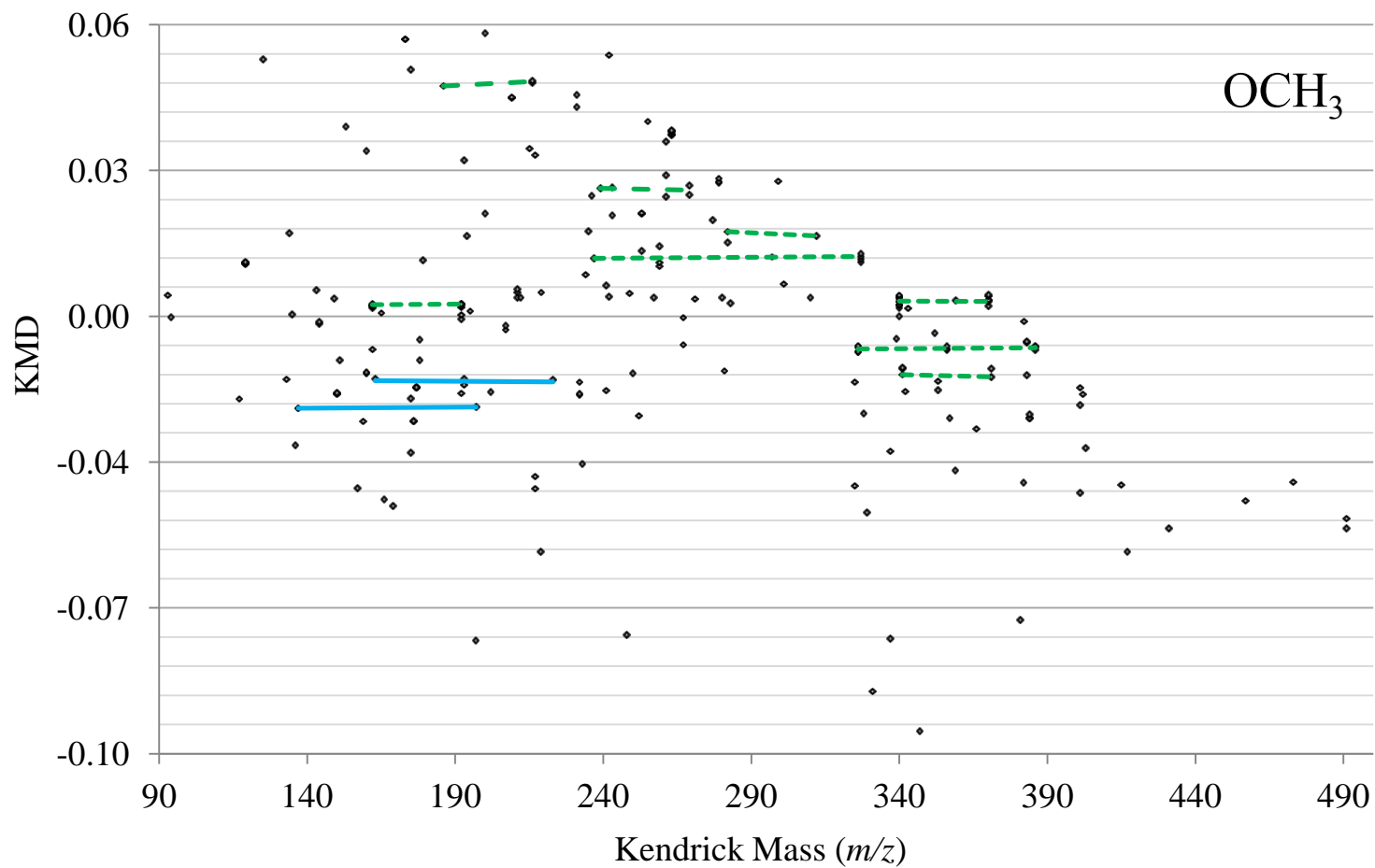


Figure 3.13. Kendrick mass defect plot for $-\text{OCH}_3$ for the new m/z peaks (200-3500 area units) and target analytes in the AFEX sample. Ions that differ by the number of functional groups are denoted by blue solid lines for the target analytes and green dotted lines for the new m/z peaks.

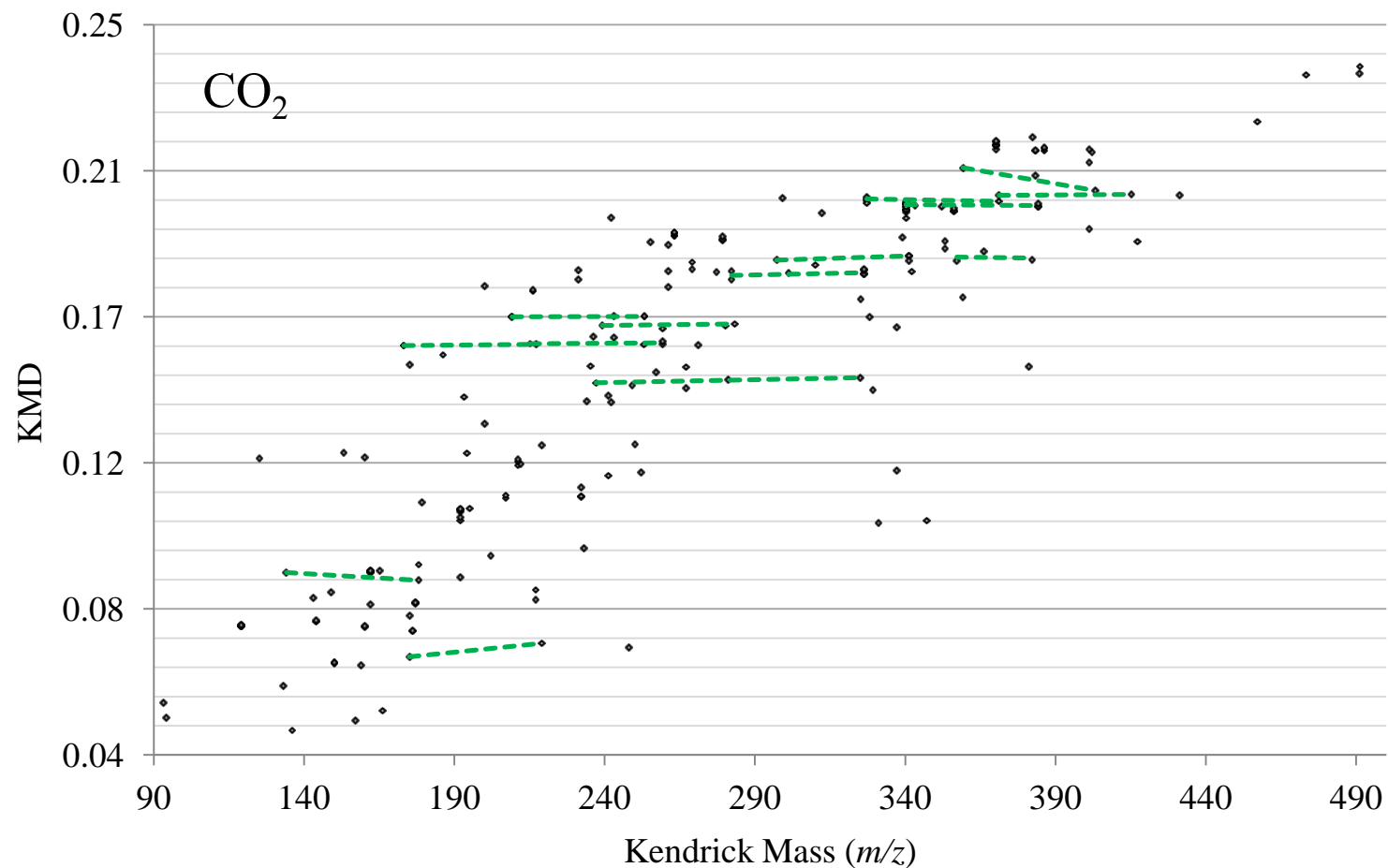


Figure 3.14. Kendrick mass defect plot for $-\text{CO}_2$ for the new m/z peaks (200-3500 area units) in the AFEX sample. Ions that differ by the number of functional groups are denoted by green dotted lines. No target analytes within the given intensity range (200-3500 area units) and differing by CO_2 were found in the AFEX sample.

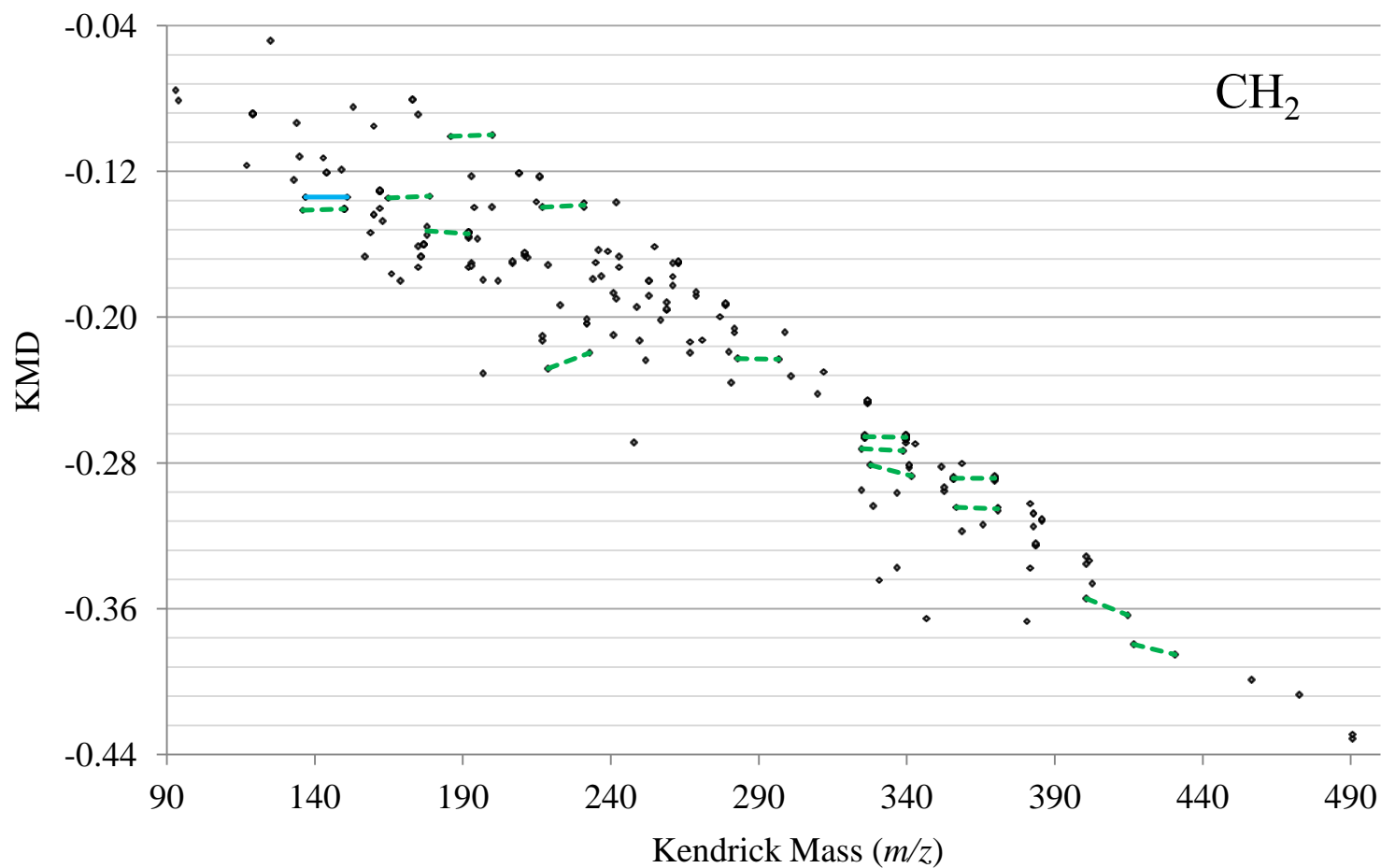


Figure 3.15. Kendrick mass defect plot for $-\text{CH}_2$ for the new m/z peaks (200-3500 area units) and target analytes in the AFEX sample. Ions that differ by the number of functional groups are denoted by blue solid lines for the target analytes and green dotted lines for the new m/z peaks.

Table 3.8. New m/z peaks in the AFEX sample identified by KMD analysis to differ by “OH”. Relative ranges (%) are given for the KMD values.

# of addt'l O's	m/z	KMD	relative range, %	# of addt'l O's	m/z	KMD	relative range, %
-	144.0450	0.0908		-	310.1092	0.2079	
1	160.0398	0.0907	0.21	1	326.1043	0.2081	0.42
2	176.0349	0.0909		2	342.0999	0.2088	
-	162.0549	0.1065		-	415.1051	0.2372	
1	178.0490	0.1057	0.76	1	431.1012	0.2384	0.50
-	179.0713	0.1283		-	370.1298	0.2476	
1	195.0658	0.1279	0.32	1	386.1251	0.2480	0.16
-	243.1037	0.1811		-	457.1164	0.2619	
1	259.0990	0.1815	0.22	1	473.1261	0.2767	5.49
-	243.1099	0.1873		-	159.0289	0.0795	
1	259.1026	0.1851	1.19	1	175.0277	0.0834	4.78
-	200.1285	0.1922		-	331.0302	0.1355	
1	216.1238	0.1926	0.20	1	347.0271	0.1375	1.46
-	261.1188	0.2019		-	366.1001	0.2166	
1	277.1147	0.2029	0.49	1	382.0939	0.2155	0.51
-	340.1190	0.2272		-	401.0984	0.2260	
1	356.1141	0.2274	0.08	1	417.0911	0.2238	0.98
-	327.1236	0.2277					
1	343.1185	0.2277	0.00				

Table 3.9. New m/z peaks in the AFEX sample identified by KMD analysis to differ by “OCH₃”. Relative ranges (%) are given for the KMD values.

# of addt'l OCH ₂ 's	m/z	KMD	relative range, %	# of addt'l OCH ₂ 's	m/z	KMD	relative range, %
-	326.1033	-0.0118		-	237.0924	0.0087	
1	356.1144	-0.0114	3.80	1	297.1140	0.0091	4.70
2	386.1249	-0.0115		2	327.1246	0.0091	
-	341.1052	-0.0153		-	282.1142	0.0145	
1	371.1156	-0.0155	1.04	1	312.1239	0.0137	6.10
-	162.0558	-0.0014		-	239.1082	0.0241	
1	192.0665	-0.0014	2.85	1	269.1194	0.0247	1.58
-	340.1194	-0.0008		-	186.1120	0.0465	
1	370.1300	-0.0007	5.26	1	216.1233	0.0473	1.71

Table 3.10. New m/z peaks in the AFEX sample identified by KMD analysis to differ by “CH₂”. Relative ranges (%) are given for the KMD values.

# of addt'l CH ₂ 's	m/z	KMD	relative range, %	# of addt'l CH ₂ 's	m/z	KMD	relative range, %
-	417.0911	-0.3746		-	219.0214	-0.2232	
1	431.1012	-0.3802	1.47	1	233.0456	-0.2146	3.90
-	401.0984	-0.3495		-	283.0984	-0.2177	
1	415.1051	-0.3584	2.53	1	297.1136	-0.2182	0.21
-	357.0994	-0.2993		-	178.0490	-0.1498	
1	371.1151	-0.2993	0.02	1	192.0655	-0.1490	0.57
-	356.1148	-0.2828		-	136.0156	-0.1363	
1	370.1285	-0.2848	0.69	1	150.0317	-0.1358	0.33
-	328.0901	-0.2762		-	217.1078	-0.1346	
1	342.0999	-0.2821	2.09	1	231.1232	-0.1349	0.19
-	325.0959	-0.2671		-	165.0548	-0.1295	
1	339.1104	-0.2683	0.43	1	179.0713	-0.1287	0.66
-	326.1039	-0.2602		-	186.1120	-0.0958	
1	340.1189	-0.2609	0.25	1	200.1285	-0.0950	0.89

Table 3.11. New m/z peaks in the AFEX sample identified by KMD analysis to differ by “CO₂”. Relative ranges (%) are given for the KMD values.

# of addt'l CO ₂ 's	m/z	KMD	relative range, %	# of addt'l CO ₂ 's	m/z	KMD	relative range, %
-	237.0921	0.1470		-	327.1248	0.2005	
1	281.0829	0.1479	0.99	1	371.1151	0.2010	0.33
2	325.0732	0.1484		2	415.1051	0.2012	
-	175.0277	0.0682		-	209.1174	0.1659	
1	219.0214	0.0721	5.53	1	253.1074	0.1661	0.12
-	134.0614	0.0924		-	282.1115	0.1768	
1	178.0490	0.0902	2.43	1	326.1030	0.1785	0.96
-	173.1177	0.1578		-	297.1136	0.1824	
1	217.1078	0.1580	0.18	1	341.1045	0.1834	0.59
-	215.1085	0.1583		-	312.1235	0.1957	
1	259.0990	0.1581	0.14	1	356.1139	0.1963	0.30
-	239.1082	0.1635		-	340.1190	0.1969	0.40
1	283.0984	0.1639	0.23	1	384.1088	0.1977	
-	236.1056	0.1602		-	359.1258	0.2089	3.22
1	280.0986	0.1634	3.22	1	403.1090	0.2023	

Comparison of Automated Compositional Analysis to Manual Data Inspection

Visual identification of new peaks. To assess the capabilities (in terms of thoroughness and utility) of the automated compositional analysis approach, unknown peaks in the AFEX sample were also identified based on visual inspection of the UPLC chromatogram and corresponding mass spectra. Combined mass spectra for each peak in the AFEX chromatogram were compared to mass spectra of the same retention time range in a water blank run under the same conditions. Any m/z peak that was (i) obviously present (i.e., at an intensity much greater in the mass spectrum of the AFEX

sample relative to the mass spectrum of the water blank) and (ii) not a target analyte was identified as a new, unknown peak. Overall, of the 30 and 209 peaks that were identified by visual inspection and MetaboLynx, respectively, 14 peaks were identified by visual inspection only, 193 peaks by MetaboLynx only and 16 peaks by both methods, (as indicated by a similar m/z peak at a similar retention time).

Manual MS/MS vs. automated MS^e analyses. Results for the 16 peaks identified by both visual inspection and MetaboLynx are contained in Table 3.12. Molecular formulas of the form $C_vH_wN_xO_y$ with exact masses within 15 ppm of experimental masses are listed for each method (with mass error and i-FIT provided for comparison). Although both methods utilized the same data file, there are small differences (generally less than 2 ppm) in reported experimental exact masses. (One possible explanation for these differences was described previously in this chapter.) For 10 out of the 16 peaks, the molecular formula with the lowest error in exact mass was the same for both methods. For five of the six remaining peaks, the MetaboLynx data set typically included the top (i.e., lowest error) visually identified chemical formula in its list of possibilities. To narrow down the possible number of chemical formulas, CID fragmentation experiments were performed for the AFEX sample. Visually-identified m/z peaks were subjected to MS/MS experiments set up manually, where the parent ion was selected in the quadrupole, fragmented in the trap cell and its daughter ions observed in the TOF. An MS^e analysis was also performed for the sample, in which low and high collision energy was alternatively applied in the trap cell to induce fragmentation over the course of the UPLC run. Fragments were identified with MS^e data viewer software (see Appendix

Table 3.12. New m/z peaks identified both by visual inspection and by MetaboLynx software. Molecular formulas with theoretical masses within 15 ppm of the experiment mass, mass errors (ppm) and isotope fit are listed for each method.

Retention Time (min)	MetaboLynx-Identified m/z			Visually Identified m/z		
	Formula Possibilities	Error (ppm)	i-FIT	Formula Possibilities	Error (ppm)	i-FIT
20.11	326.1030			326.1031		
	C₁₈H₁₆NO₅	0.6	5.6	C₁₈H₁₆NO₅	0.9	0.03
	C ₂₁ H ₁₄ N ₂ O ₂	-7.7	3.0	C ₁₃ H ₁₆ N ₃ O ₇	13.2	3.5
	C ₁₅ H ₁₈ O ₈	8.6	7.1			
	C ₁₃ H ₁₆ N ₃ O ₇	12.9	8.4			
20.80	386.1253			386.1255		
	C₂₀H₂₀NO₇	3.4	5.4	C₂₀H₂₀NO₇	3.9	0.04
	C ₂₃ H ₁₈ N ₂ O ₄	-3.6	0.4	C ₂₆ H ₁₆ N ₃ O	-9.8	3.3
	C ₁₇ H ₂₂ O ₁₀	10.4	8.5	C ₁₃ H ₂₄ NO ₁₂	-11.4	5.8
	C ₂₆ H ₁₆ N ₃ O	-10.4	6.8	C ₁₅ H ₂₀ N ₃ O ₉	14.2	5.8
	C ₁₃ H ₂₄ NO ₁₂	-11.9	10.2			
21.27	356.1148			356.1141		
	C ₂₂ H ₁₆ N ₂ O ₃	-3.7	1.2	C₁₉H₁₈NO₆	2.0	0.07
	C₁₉H₁₈NO₆	3.9	5	C ₁₄ H ₁₈ N ₃ O ₈	13.2	5.2
	C ₂₅ H ₁₄ N ₃	-11.2	3.4	C ₂₅ H ₁₄ N ₃	-13.2	2.9
	C ₁₆ H ₂₀ O ₉	11.5	7.8	C ₁₂ H ₂₂ NO ₁₁	-14.6	5.8
	C ₁₂ H ₂₂ NO ₁₁	-12.6	9.3			
22.52	383.1252			383.1254		
	C ₂₀ H ₁₉ N ₂ O ₆	2.3	3.3	C ₂₀ H ₁₉ N ₂ O ₆	2.9	1.3
	C ₂₃ H ₁₇ N ₃ O ₃	-4.7	2.5	C ₂₅ H ₁₉ O ₄	-7.6	0.4
	C ₂₅ H ₁₉ O ₄	-8.1	2.3	C ₁₃ H ₂₃ N ₂ O ₁₁	-12.5	2.6
	C ₁₇ H ₂₁ NO ₉	9.4	4			
	C ₁₃ H ₂₃ N ₂ O ₁₁	-13.1	4.7			

Table 3.12. New m/z peaks identified both by visual inspection and by MetaboLynx software. Molecular formulas with theoretical masses within 15 ppm of the experiment mass, mass errors (ppm) and isotope fit are listed for each method. (cont'd.)

Retention Time (min)	MetaboLynx-Identified m/z			Visually Identified m/z		
	Formula Possibilities	Error (ppm)	i-FIT	Formula Possibilities	Error (ppm)	i-FIT
22.52	386.1251			386.1252		
	C₂₀H₂₀NO₇	2.8	5.8	C₂₀H₂₀NO₇	3.9	0.04
	C ₂₃ H ₁₈ N ₂ O ₄	-4.1	3.4	C ₂₆ H ₁₆ N ₃ O	-9.8	3.3
	C ₁₇ H ₂₂ O ₁₀	9.8	8.1	C ₁₃ H ₂₄ NO ₁₂	-11.4	5.8
	C ₂₆ H ₁₆ N ₃ O	-10.9	0.2	C ₁₅ H ₂₀ N ₃ O ₉	14.2	5.8
	C ₁₃ H ₂₄ NO ₁₂	-12.4	9.6			
22.77	C ₁₅ H ₂₀ N ₃ O ₉	13.2	9.3			
	325.0732			325.0726		
	C ₆ H ₁₇ N ₂ O ₁₃	0.3	4	C ₆ H ₁₇ N ₂ O ₁₃	-1.5	10.0
	C ₂₁ H ₁₁ NO ₃	-2.2	2.5	C₁₈H₁₃O₆	4.3	6.2
	C₁₈H₁₃O₆	6.2	2.8	C ₂₄ H ₉ N ₂	-12.3	0.0
	C ₉ H ₁₅ N ₃ O ₁₀	-7.7	3.8	C ₁₁ H ₁₇ O ₁₁	-13.8	8.7
27.50	C ₁₆ H ₁₁ N ₃ O ₅	10.2	3			
	C ₂₄ H ₉ N ₂	-10.5	2			
	C ₁₁ H ₁₇ O ₁₁	-12	3.5			
	402.1204			402.1198		
	C ₂₃ H ₁₈ N ₂ O ₅	-3	4.6	C ₂₀ H ₂₀ NO ₈	2.2	1.9
	C ₂₀ H ₂₀ NO ₈	3.7	5.7	C ₂₆ H ₁₆ N ₃ O ₂	-11.2	0.3
27.57	C ₂₆ H ₁₆ N ₃ O ₂	-9.7	3.5	C ₁₅ H ₂₀ N ₃ O ₁₀	11.2	3.0
	C ₁₇ H ₂₂ O ₁₁	10.4	6.8	C ₁₃ H ₂₄ NO ₁₃	-12.4	4.1
	C ₁₃ H ₂₄ NO ₁₃	-10.9	7.6			
	C ₃₀ H ₁₄ N ₂	11.7	0.3			
28.35	186.1120			186.1118		
	C₉H₁₆NO₃	-5.4	-	C₉H₁₆NO₃	-6.4	-
28.35	192.0655			192.0665		
	C₁₀H₁₀NO₃	-3.1	2.2	C₁₀H₁₀NO₃	2.1	-
	C ₇ H ₁₂ O ₆	10.9	3.4			

Table 3.12. New m/z peaks identified both by visual inspection and by MetaboLynx software. Molecular formulas with theoretical masses within 15 ppm of the experiment mass, mass errors (ppm) and isotope fit are listed for each method. (cont'd.)

Retention Time (min)	MetaboLynx-Identified m/z			Visually Identified m/z		
	Formula Possibilities	Error (ppm)	i-FIT	Formula Possibilities	Error (ppm)	i-FIT
29.47	211.0744			211.0741		
	$C_{12}H_9N_3O$	-0.9	1.7	$C_{14}H_{11}O_2$	-8.5	0.07
	$C_{14}H_{11}O_2$	-7.1	0.3	$C_9H_{11}N_2O_4$	10.4	2.8
	$C_9H_{11}N_2O_4$	11.8	3.5			
32.23	217.1078			217.1078		
	$C_{10}H_{17}O_5$	0.9	4.4	$C_{10}H_{17}O_5$	0.9	-
	$C_8H_{15}N_3O_4$	6.9	4.1			
	$C_{13}H_{15}NO_2$	-11.5	2.1			
32.34	340.1175			340.1178		
	$C_{19}H_{18}NO_5$	-2.9	3	$C_{19}H_{18}NO_5$	-2.1	0.5
	$C_{16}H_{20}O_8$	5	3.4	$C_{14}H_{18}N_3O_7$	9.7	0.9
	$C_{14}H_{18}N_3O_7$	8.8	3.3			
	$C_{22}H_{16}N_2O_2$	-10.9	2.6			
	$C_{26}H_{14}N$	14.4	2.1			
32.34	352.1162			352.1168		
	$C_{17}H_{20}O_8$	1.1	4.5	$C_{20}H_{18}NO_5$	-4.8	4.9
	$C_{15}H_{18}N_3O_7$	4.8	5.1	$C_{15}H_{18}N_3O_7$	6.5	6.6
	$C_{20}H_{18}NO_5$	-6.5	3.7	$C_8H_{22}N_3O_{12}$	-9.9	8.1
	$C_{27}H_{14}N$	10.2	0.5	$C_{27}H_{14}N$	11.9	0.01
	$C_8H_{22}N_3O_{12}$	-11.6	6.5			
	$C_{12}H_{20}N_2O_{10}$	12.5	5.7			
	$C_{23}H_{16}N_2O_2$	-14.2	2.7			
33.70	282.1115			282.1118		
	$C_{14}H_{18}O_6$	4.3	3.4	$C_{17}H_{16}NO_3$	-4.3	0.2
	$C_{17}H_{16}NO_3$	-5.3	2.4	$C_{12}H_{16}N_3O_5$	9.9	1.9
	$C_{12}H_{16}N_3O_5$	8.9	4.1			
	$C_{20}H_{14}N_2$	-14.9	0.8			

Table 3.12. New m/z peaks identified both by visual inspection and by MetaboLynx software. Molecular formulas with theoretical masses within 15 ppm of the experiment mass, mass errors (ppm) and isotope fit are listed for each method. (cont'd.)

Retention Time (min)	MetaboLynx-Identified m/z			Visually Identified m/z		
	Formula Possibilities	Error (ppm)	i-FIT	Formula Possibilities	Error (ppm)	i-FIT
33.85	401.1214			401.0891		
	$C_{19}H_{19}N_3O_7$	-2.2	6.8	$C_8H_{21}N_2O_{16}$	0.0	0.6
	$C_{31}H_{15}N$	2.5	4.1	$C_{20}H_{17}O_9$	4.5	1.6
	$C_{16}H_{21}N_2O_{10}$	4.5	8	$C_{26}H_{13}N_2O_3$	-8.7	2.5
	$C_{21}H_{21}O_8$	-5.5	6	$C_{13}H_{21}O_{14}$	-10	2.2
	$C_{28}H_{17}O_3$	9	0.5	$C_{15}H_{17}N_2O_{11}$	14.7	3.1
34.14	$C_{13}H_{23}NO_{13}$	11.2	9.1			
	312.1235			312.1240		
	$C_{18}H_{18}NO_4$	-0.3	1.7	$C_{18}H_{18}NO_4$	1.3	0.1
	$C_{15}H_{20}O_7$	8.3	2.3	$C_{13}H_{18}N_3O_6$	14.1	2.6
	$C_{21}H_{16}N_2O$	-9	1.4			
	$C_{13}H_{18}N_3O_6$	12.5	3.8			

D, Figure D.2 for features), which time-aligns fragment ions by identifying m/z peaks in similar retention time windows. Results of these experiments can be seen in Table 3.13. Fragmentation data between the two methods vary, which is expected based on differences in how the experiments are designed. In MS/MS experiments, m/z peaks are selected in the quadrupole and thus, only fragments from parent peaks are detected. In MS^e experiments, a single m/z is not selected but rather everything eluting at the same retention time is subjected to fragmentation. That is, fragments resulting from two co-eluting species will appear together in the MS^e data, making it difficult to assign fragment peaks to a particular parent peak. Note that for 3 of the 16 peaks, the MS/MS spectra were unclear due to multiple parent peaks (at very similar m/z) being selected in the quadrupole. Despite the differences in approach, a majority of the peaks do have one or

Table 3.13. CID fragments for MS^c analyses and manual MS/MS experiments for the new *m/z* peaks identified both by visual inspection and by MetaboLynx software. KMD data includes functional groups identified by KMD analyses for the new peaks and the other members of the same homologous series in parentheses.

<i>m/z</i>	RT (min)	MS ^c Fragments MS/MS Fragments	KMD Data
326	20.11	309 , 280, 265 , 237 , 211 , 162 , 143 , 119 309 , 265 , 237 , 211 , 162 , 143 , 119 , 117, 91	OH (310, 342), OCH ₃ (356, 386), CO ₂ (282), CH ₂ (340)
386	20.80	372, 341, 340, 325, 299, 297 , 287, 283, 282 , 267 , 192, 149, 135 327, 312, 297 , 282 , 267 , 203, 188, 176, 136, 109, 92	OH (370), OCH ₃ (326, 356)
356	21.27	323, 312 , 296, 285, 253 , 251, 243, 162, 144, 120 312 , 253 , 229, 176, 134, 119	OH (340), OCH ₃ (326, 386), CH ₂ (370)
383	22.52	368 , 352, 340 , 338, 325 , 324, 310 , 295, 281 , 267 , 238, 225, 216, 159 368 , 340 , 325 , 310 , 297, 281 , 267 , 201, 177, 159, 134	
386	22.52	342 , 326, 297, 283, 279, 237, 190 , 146, 134 368, 353, 342 , 327, 310, 284, 190 , 177, 176, 147, 134 , 121, 109, 87	OH (370), OCH ₃ (326, 356)
325	22.77	297, 293, 281 , 237 , 163, 119 , 93 307, 281 , 263, 237 , 211, 219, 135, 119 , 93	CO ₂ (237, 281)
402	27.50	186, 175, 168, 125, 97 Unclear spectra	
186	27.57	175, 168, 125, 97 145, 123, 97 , 80	OCH ₃ (216), CH ₂ (200)
192	28.35	177, 165, 150, 121 , 93 176, 149, 121 , 93	OCH ₃ (162), CH ₂ (178)
211	29.47	203, 177, 173, 163, 139, 119 Unclear spectra	
217	32.23	199, 183, 169, 166, 164, 155, 107, 83 171, 153, 111, 95, 89	CO ₂ (173), CH ₂ (231)

Table 3.13. CID fragments for MS^e analyses and manual MS/MS experiments for the new *m/z* peaks identified both by visual inspection and by MetaboLynx software. KMD data includes functional groups identified by KMD analyses for the new peaks and the other members of the same homologous series in parentheses.

<i>m/z</i>	RT (min)	MS ^e Fragments MS/MS Fragments	KMD Data
340	32.34	337, 325 , 322, 310 , 294, 282 , 267 , 239 , 215 325 , 310 , 282 , 267 , 239 , 200, 119	OH (356), OCH ₃ (370) CO ₂ (384)
352	32.34	340, 337 , 325, 322 , 310, 294 , 282, 267, 239, 215 337 , 322 , 306, 294 , 266, 251	
282	33.70	251, 242, 174 , 162 , 145 , 125, 119 , 93 174 , 162 , 145 , 119 , 93	CO ₂ (326), OCH ₃ (312)
401	33.85	282, 251, 242, 174, 162, 145, 125, 119, 93 341, 311, 283, 269, 117, 87	
312	34.14	297, 177, 173, 145, 135, 117 Unclear spectra	OCH ₃ (282) CO ₂ (356)

more fragments in common between the two methods. (Fragment ions resulting from both methods are bolded.)

Before using fragmentation data to assign molecular formulas, it is important to point out a feature unique to MS^e experiments. A limitation in LC-MS analyses of unknowns is that if multiple *m/z* peaks are present for a given chromatographic peak, it is unclear if each *m/z* peak is its own species in the sample or if lower *m/z* peaks are fragmenting (in the ESI source or in the trap cell) from higher *m/z* peaks. Retention time alignment (with a narrow time window) of MS^e software allows for identification of these situations when a lower *m/z* ion is likely the result of a higher *m/z* ion. For example, a prominent peak of *m/z* 162 at 13.9 minutes was initially designated as p-coumaryl amide (C₉H₉NO₂), a known AFEX degradation product.¹⁰⁵ However, several higher *m/z* peaks

were observed in the MS^e retention-time-aligned low collision energy mass spectrum of m/z 162, including m/z 237, 280, 323, and 484. Manual CID fragmentation of these ions revealed them to originate from the same parent ion (Figure 3.16). Similar fragmentation losses (18, 43, and 86 Da) for m/z 162, 323 and 484 suggest that they compose a polymer of 161 Da polymer units. We increased our m/z window (beyond 500) to determine if higher MW polymer units were present. Subsequent fragmentation of m/z 645, 806, and 945 peaks yielded similar fragments and losses, including m/z 162, 323, and 484. It is not

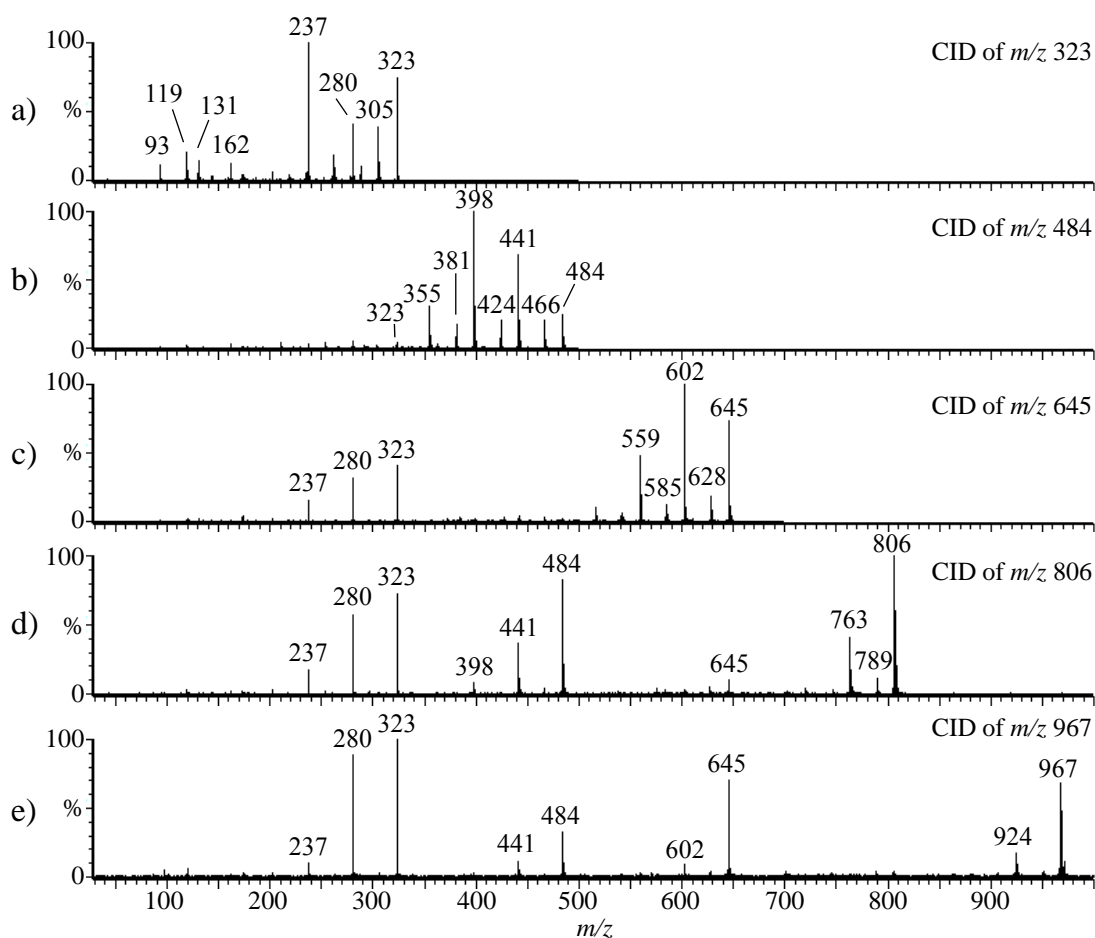


Figure 3.16. Collision-induced-dissociation spectra of m/z 323, 484, 645, 806, and 967, which correspond to varying degrees (2-6) of polymerization.

known whether this compound is naturally occurring in corn stover or produced during AFEX pretreatment. However, similar m/z peaks were not observed in the dilute-acid sample so this compound is either produced during AFEX pretreatment or degraded during dilute-acid pretreatment.

KMD analyses described earlier in this chapter provide an additional tool in assignment of molecular formulas. Identification of a lower MW member of a homologous series can allow for identification of higher MW members of that same series. Since there are fewer molecular formula possibilities at lower MW, it is easier to identify lower MW peaks. For example, m/z 192 in the AFEX sample is related to m/z 162 by a methoxy group (OCH_3). These peaks correspond to compounds that have been previously identified as AFEX pretreatment degradation products: feruloyl amide and coumaryl amide, respectively, which in fact differ in structure only by a single methoxy group. For the m/z 162 peak, there was only one molecular formula (i.e., $\text{C}_9\text{H}_8\text{O}_2\text{N}$) with an exact mass within 15 ppm of the experimental mass. For the MetaboLynx-identified m/z 192 peak, only one (i.e., $\text{C}_{10}\text{H}_{10}\text{NO}_3$) of the two possible molecular formulas within 15 ppm differs by “ OCH_2 ” from $\text{C}_9\text{H}_8\text{O}_2\text{N}$. Thus, KMD analyses can help lead to assignment of molecular formulas or at least narrow down the number of possibilities.

Functional groups identified by KMD, along with nominal masses of their fellow members in the homologous series, are listed in Table 3.13 for the peaks identified by both MetaboLynx and visual inspection. For example, m/z 326 is related to m/z 310 and 342 by “OH”, to m/z 356 and 386 by “ OCH_3 ” and to m/z 282 by “ CO_2 ”. The data in Table 3.13 can be used to narrow down the molecular formula possibilities given in Table 3.12. An illustration of this process is shown in Figure 3.17 for MetaboLynx-identified

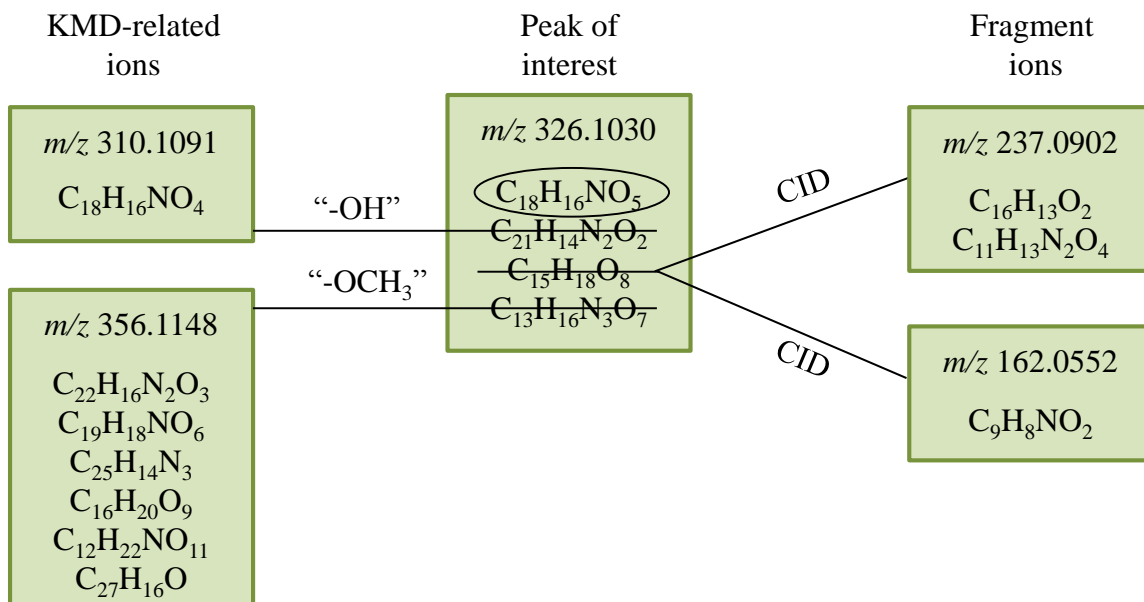


Figure 3.17. Example of how KMD analyses and fragmentation data can help narrow down molecular formula possibilities. The molecular formula possibilities for exact masses within 15 ppm of the experimental masses are listed for KMD-related ions, fragment ions, and the peak of interest. Three molecular formulas can be removed from the list, leaving one to be assigned.

m/z 326.1030, for which there are initially four possible molecular formulas with exact masses within 15 ppm of m/z 326.1030. Molecular formula possibilities (with exact masses within 15 ppm) are also listed for two fragments of m/z 326 (i.e., m/z 162.0552 and 237.0902) and two members of separate homologous series with m/z 326 (i.e., m/z 310.1091 and m/z 356.1148). Considering that fragments must have a fewer number of individual C, H, N, and O atoms than their parent ion, one of the possible molecular formulas for m/z 326 can be rejected: $C_{15}H_{18}O_8$. Formula possibilities for fragment ion 237 m/z , $C_{16}H_{13}O_2$ and $C_{11}H_{13}N_2O_4$, have a greater number of C and N atoms, respectively, than the m/z 326 possibility ($C_{15}H_{18}O_8$) so it can be ruled out. Likewise, the formula for m/z 162.0552, $C_9H_8NO_2$, has also a greater number of N atoms than the $C_{15}H_{18}O_8$ possibility, confirming its rejection. Two of the remaining three possibilities

can be ruled out based on KMD data. The molecular formula for m/z 326 must contain one “OCH₂” (due to exchange of “OCH₃” for “H”, as described previously) less than m/z 356.1148. The formula C₁₃H₁₆N₃O₇ of m/z 326 can be eliminated because it is not composed of one “OCH₂” less than any of the possible formulas for m/z 356. Likewise, m/z 310 (related to m/z 326 by an “O” atom”) rules out the formula C₂₁H₁₄N₂O₂. The only formula possibility left (i.e., C₁₈H₁₆NO₅) can be assigned to the m/z 326 ion. Note that the m/z 310 ion actually rules out all three other formulas, not just one. However, a larger weight of evidence approach was employed to achieve higher confidence in formula assignment. It should also be noted that although the formula possibilities for m/z 282 (related to m/z 326 by “CO₂”) could not rule out any of the possibilities for m/z 326, assignment of C₁₈H₁₆NO₅ to m/z 326 allowed assignment of C₁₇H₁₆NO₃ to m/z 282.

Application of KMD and fragmentation data in this way allowed for a single molecular formula to be assigned for 11 of the 16 new peaks in Table 3.12 (shown in bold). Of the remaining 5 peaks, molecular formulas were narrowed down to two or three possibilities for 4 peaks (shown in italics in Table 3.12). No molecular formulas could be ruled out for the highest MW ion, m/z 402, which had no KMD pair and did not have clear fragmentation data (i.e., other ions around m/z 402 were also being selected in the quadrupole). Additional KMD analyses for other functional groups or better parent ion selection in CID experiments might help in assigning a molecular formula for this peak.

Based on the peaks with a single molecular formula assignment, low mass error appears to be a better indicator of accurate molecular formula than isotope fit for the approach presented in this chapter. Molecular formulas corresponding to exact masses

with the lowest error accounted for 8 out of the 11 MetaboLynx-identified peaks (that were assigned a single formula) compared to only 1 out of the 11 with the best i-FIT value. Although 8 out of 11 seems to be a large percentage, using mass error alone would have incorrectly identified molecular formulas for approximately 30% of ions. Neither mass error nor i-FIT should be considered exclusively but rather they should be used in conjunction with KMD and fragmentation data to improve the accuracy of formula assignment.

Conclusions

The UPLC separations presented in this chapter clearly offer improvements in chromatographic efficiency (e.g., increased analyte separation and reduced analysis time) compared to traditional HPLC separations. HRMS data proved useful in elucidating bulk differences in molecular composition between the two pretreated samples. Although assignment of unique molecular formulas could not be achieved for every new peak and was especially problematic for larger m/z compounds with our current instrumentation, the combined UPLC-HRMS and MetaboLynx methodology described in this chapter presents a promising approach to identification of new degradation products in biomass. The molecular understanding of composition this approach provides is expected to significantly improve opportunities to find inhibitory components that may significantly limit fuel yields, and also may be utilized to identify potential value-added products. The primary hurdle in successfully employing this methodology is the resolving power and mass accuracy of the mass spectrometer. I note that commercial solutions currently exist that address these limitations. Application of this same method with a higher resolving power instrument such as an Orbitrap, Fourier transform ion cyclotron resonance, or the

next generation time-of-flight (Waters Synapt G2) mass spectrometer would likely improve mass accuracy and utility of this method. Other ideas for extensions and applications of the method presented in this chapter will be discussed in Chapter 5.

CHAPTER FOUR

Ion Mobility – Mass Spectrometry Analysis of Asphaltenes using Model Compounds to Understand Structure, Diversity, and Aggregation

Introduction

Current trends in the petroleum industry are towards heavier crude oils with high sulfur, metal, and asphalt content owing to the depletion of light crude-oil reserves.^{15,106-112} Asphaltenes represent the largest molecular weight (MW) fraction of crude oil and remain detrimental in the petroleum industry in both upstream and downstream refining processes.^{17,113-119} The presence of asphaltenes is problematic because they create drastic variations in viscosity and have a tendency to flocculate, ultimately resulting in costly system fouling.¹¹⁷⁻¹²⁰ Accordingly, the economic value of a crude oil diminishes as a function of increasing asphaltene content.^{121,122} To date, problems caused by asphaltenes present in crude oils remain unresolved. Ultimate industrial goals are to (i) eliminate the detrimental effect asphaltenes cause during transportation and refining of crude oils and (ii) develop new technologies to convert these materials into useful fuels, thus obtaining the greatest value from heavy crude oil.^{121,123} Efficient processing of these materials and development of new methods for their conversion to useful fuels requires quantitative characterization of asphaltene chemical properties, including molecular weight distribution, functional groups, composition, molecular structures, etc.

Operationally, petroleum asphaltenes are defined by their solubility characteristics (toluene soluble, n-heptane insoluble) and exist as a complex mixture of compounds which vary in molecular class and structure. A number of issues have been resolved

concerning asphaltene structures and several excellent texts review the current knowledge base.^{15,119,124-126} By definition, asphaltenes are the most aromatic and most polar (i.e., containing the highest percentage of heteroatoms) portion of crude oil. It is generally thought that asphaltenes have a single aromatic core surrounded by alkyl substituents,^{15,119,127} and several studies have suggested that asphaltenes have an average molecular weight of approximately 800 g/mol with a distribution between 300 and 2500 g/mol.^{15,128-130} However, the exact molecular weight distributions of heavy petroleum samples, especially that of asphaltenes, remains a subject of some debate, despite extensive research in this area.^{120,127,131-134}

A variety of techniques have been employed to determine the average molecular features of asphaltenes. Approximately 50% of asphaltene carbon is aromatic, as determined by ¹³C NMR, which suggests a peri-condensed structure similar to coronene.¹³⁵ A peri-condensed structure has also been supported by X-ray Raman spectroscopy.¹³⁶ Remaining carbon exists primarily as alkyl substituents on the aromatic rings,¹⁵ although the average length and degree of branching for these substituents is not known. A majority of heteroatoms present in asphaltenes are embedded in the aromatic core, and data has been interpreted to indicate that nitrogen and sulfur are typically present in 5-membered aromatic rings^{15,119,123,126} with a smaller percent of 6-membered aromatic rings containing nitrogen.¹³⁷ On the other hand, although furan-like structures have also been reported,¹³⁸ oxygen is primarily present in carbonyl, alcohol, or carboxylic acid substituents.^{139,140} Nonetheless, asphaltenes in crude oils remain largely uncharacterized owing to inherent difficulties in their isolation and subsequent analysis.^{120,141}

In the work presented in this chapter, ion mobility – mass spectrometry (IM-MS) and several model compounds were employed to better understand asphaltenes in terms of their structural types and degree of structural diversity. Electrospray ionization (ESI)-generated asphaltene ions were compared with model compound monomers and dimers generated from a dual ESI/ESCI (electrospray chemical ionization) source, in terms of their collision-cross-section (CCS) values and arrival-time-peak widths. CCS analysis demonstrates that both monomeric and dimeric archipelago- and island-type structures may be present in asphaltenes. Comparison of arrival time peak widths for asphaltenes and model compounds indicates structural or conformational diversity for asphaltene compounds within a given nominal m/z .

Experimental: Materials and Methods

Chemicals and Reagents

Asphaltenes were precipitated (with n-heptane) from a San Andro region crude oil received from Dr. Daulat D. Mamora in the Department of Petroleum Engineering at Texas A&M University (College Station, TX). Cholestanon-thiophene (chol-thio), tetrahexylpyrene (THP), 1,3-bis(2-pyren-1-yl-ethyl)-benzene (PmPhP), 2,5-bis(2-pyren-1-yl-ethyl)-thiophene (PThP), and 2,5-bis(2-pyren-1-yl-ethyl)-pyridine (P-2,5-PMP) were received from Dr. Murray Gray, Director of The Centre for Oil Sands Innovation at The University of Alberta (Edmonton, Canada). All other chemicals (>97% purity, except for rubrene and pentacene, whose purities are not specified) were purchased from standard commercial vendors (i.e., Sigma-Aldrich and Fisher-Scientific). All model compounds used in this study can be found in Table 4.1.

Table 4.1. Compound names, formulas, monoisotopic masses, and structures for the model compounds used in this study.

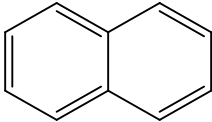
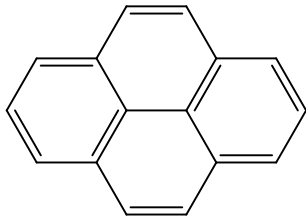
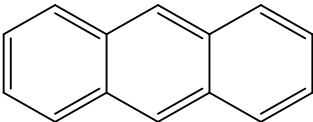
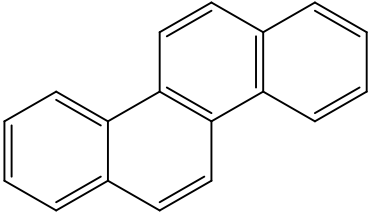
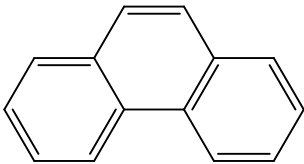
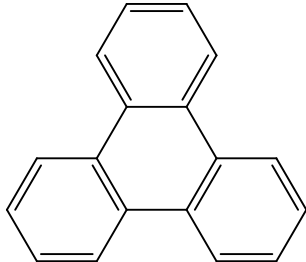
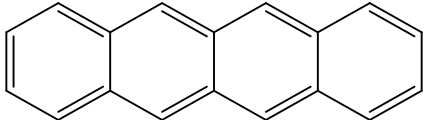
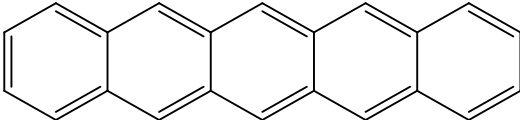
Compound Formula Monoisotopic Mass Structure Type	Structure	Compound Formula Monoisotopic Mass Structure Type	Structure
Naphthalene $C_{10}H_8$ 128.06259 PAH		Pyrene $C_{16}H_{10}$ 202.07824 PAH	
Anthracene $C_{14}H_{10}$ 178.07824 PAH		Chrysene $C_{18}H_{12}$ 228.09389 PAH	
Phenanthrene $C_{14}H_{10}$ 178.07824 PAH		Triphenylene $C_{18}H_{12}$ 228.09389 Fused Ring	
Benz[b]anthracene $C_{18}H_{12}$ 228.09389 PAH		Pentacene $C_{22}H_{14}$ 278.10954 PAH	

Table 4.1. Compound names, formulas, monoisotopic masses, and structures for the model compounds used in this study (cont'd.)

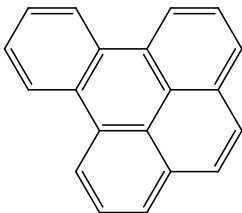
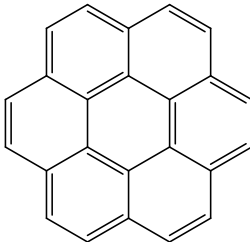
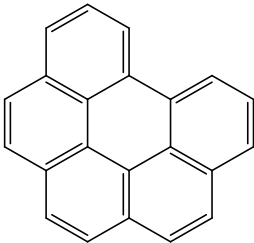
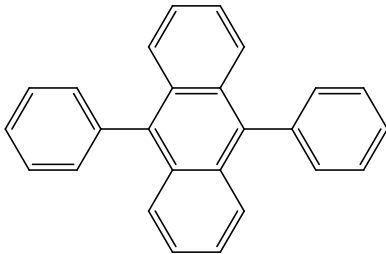
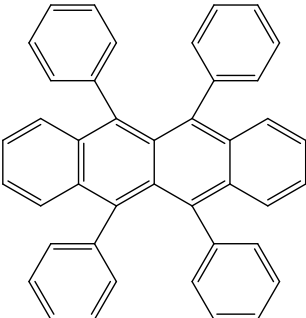
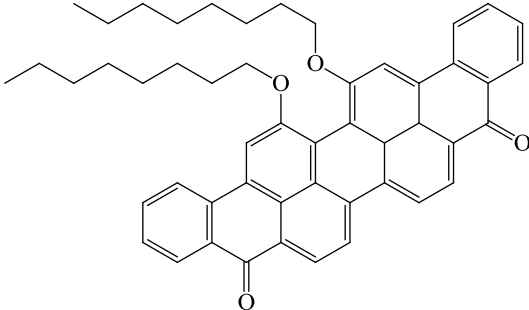
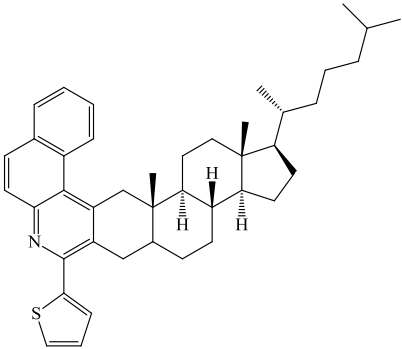
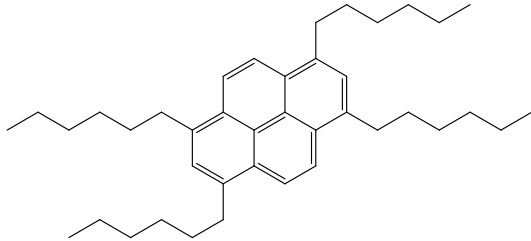
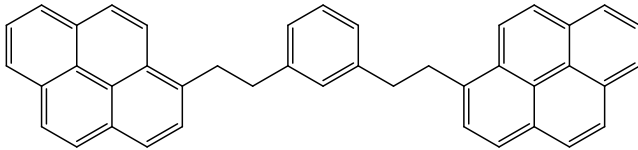
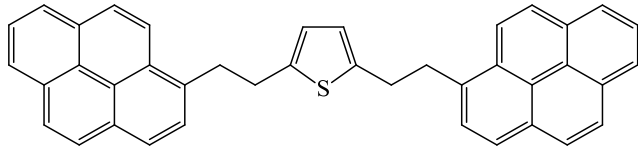
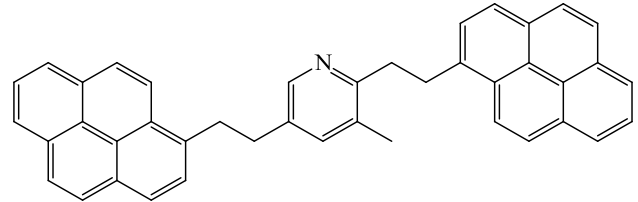
	Compound Formula Monoisotopic Mass Structure Type	Structure	Compound Formula Monoisotopic Mass Structure Type	Structure
110	Benzo[e]pyrene $C_{20}H_{12}$ 252.09389 PAH		Coronene $C_{24}H_{12}$ 300.09389 PAH	
	Benzo[ghi]perylene $C_{22}H_{12}$ 276.09389 PAH		9,10-Diphenylanthracene $C_{26}H_{18}$ 330.42110 Archipelago	
	Rubrene $C_{42}H_{28}$ 532.21908 Archipelago		Violanthrone-79 $C_{50}H_{48}O_4$ 712.35523 Island	

Table 4.1. Compound names, formulas, monoisotopic masses, and structures for the model compounds used in this study (cont'd.)

Compound (Abbrev.) Formula Monoisotopic Mass Structure Type	Structure	Compound (Abbrev.) Formula Monoisotopic Mass Structure Type	Structure
Cholestanon-Thiophene (Chol-Thio) $C_{42}H_{53}NS$ 603.38987 Island		Tetrahexylpyrene (THP) $C_{40}H_{58}$ 538.45385 Island	
1,3-bis(2-pyren-1-yl-ethyl)-benzene (PmPhP) $C_{42}H_{30}$ 534.23475 Archipelago			
2,5-bis(2-pyren-1-yl-ethyl)-thiophene (PThP) $C_{40}H_{28}S$ 540.1912 Archipelago			
2,5-bis(2-pyren-1-yl-ethyl)-3-methyl-pyridine (P-2,5-PMP) $C_{42}H_{31}N$ 549.2457 Archipelago			

Preparation of Standards and Asphaltene Samples

To prepare for spray ionization, discrete amounts of asphaltene or model compound were weighed, dissolved in toluene, and diluted to volume with methanol to obtain a stock solution of ~0.1 mg/mL in ~60:40 toluene–methanol. As necessary, dilutions of these stocks were prepared in 60-40 toluene–methanol.

DIESI-IM-HRMS Analysis

Asphaltenes and model compounds were analyzed with a Synapt HDMS ion mobility – mass spectrometer (Waters, Manchester, U.K.) that is equipped with a dual electrospray- / electrospray-chemical ionization (ESI/ESCI) source that has been described elsewhere.¹⁴² ESI settings utilized were as follows: source temperature, 110 °C; desolvation gas flow rate, 300 L/hr; desolvation temperature, 300 °C; and cone gas flow rate, 50 L/hr. IMS separations were achieved in a traveling wave (T-wave) mobility cell with wave heights between 6.0-8.5 V and a wave velocity of 300 m/s. A nitrogen bath gas was passed through the IM cell at a flow rate of 25 mL/min providing a typical operating IM pressure of 0.59 mbar. In most experiments, instrumental defaults were utilized for T-wave values in the trap and transfer cells (6 and 4 V, respectively). All arrival time distributions were obtained by integrating the monoisotopic peak of a given analyte in the 2D IM-MS spectra and are presented uncorrected, as obtained from software (i.e., analyte arrival times = mobility drift time + time spent in transit to TOF detector).

Collision-Cross-Section Measurements

Collision cross sections of analytes were measured using a calibration protocol reported by Ruotolo and coworkers.⁸⁶ $[M + H]^+$ ions generated from polypropylene glycol (PPG) standards were used as calibrants and collision cross sections for the calibrants were taken from literature¹⁴³ (118.9, 132.2, 143.9, 153.5, 165.4, 176.5, 189.3, 201.7, 214.3, 228.0, and 236.5 Å²) for DP (degree of polymerization) units 5-15, respectively. Plots used for calibration can be found in Appendix A (Figures A.3-4). Accurate calibration was verified by re-measuring the collision cross sections of the standards using the generated calibration curve; deviation from reported values was less than 2%. IM-MS data was collected at six different ion mobility wave heights (every 0.5 V in the range 6.0-8.5 V). Most collision-cross-section values for model compounds and asphaltenes are reported as an average over at least three wave heights. However, for asphaltene analytes of high m/z (i.e., ~1100-1300 m/z) and for dimers of model compounds, collision cross sections are reported at one or an average of two wave heights. Arrival times used in collision cross section measurements of model compounds were obtained by integrating ion signal across the monoisotopic peak of each standard ion. Arrival times used for average CCS measurements of asphaltenes were obtained by integrating all peaks in a 5- m/z window every 100 m/z units (e.g., 400-405 m/z , 500-505 m/z , etc.). To obtain CCSs representing threshold values for asphaltene ions, collision cross sections were also calculated from the arrival times at 13.5% maximum peak height which represents 95.5% (2 standard deviations) of the average arrival time peak area. For example, Figure 4.1 shows these measurements for the arrival time distribution of the 600-605 m/z window. An average arrival time peak height of 7.21 ms was used to

determine average CCS. Arrival times at 2σ , 5.17 ms and 9.26 ms, were used for CCSs representing threshold values for asphaltenes. All collision cross section measurements can be found in Appendix A (Tables A.6-7).

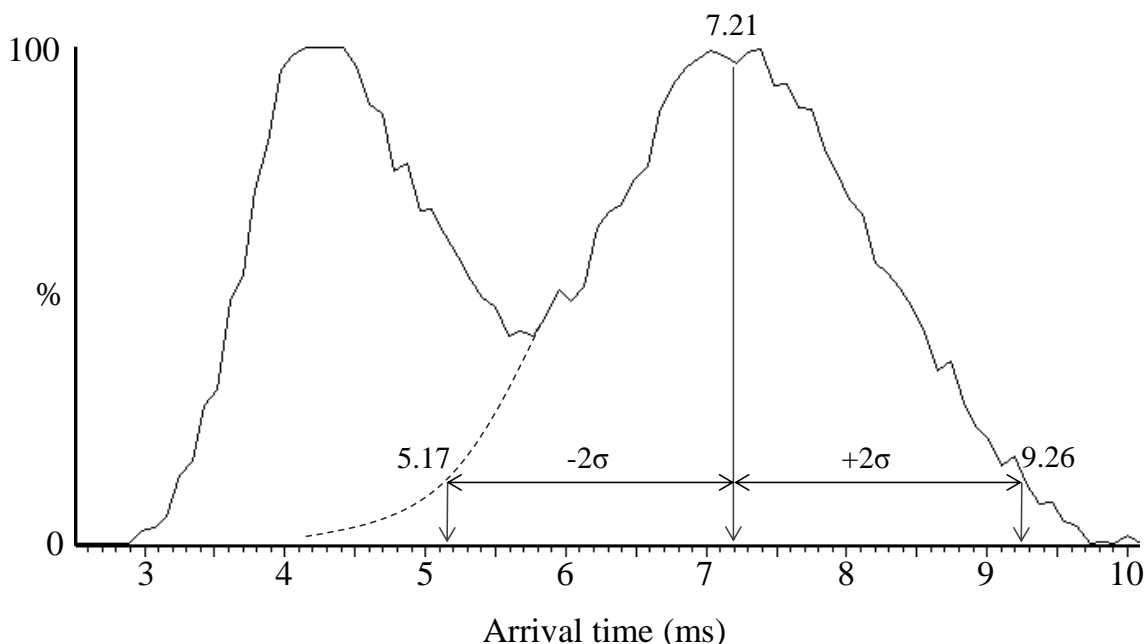


Figure 4.1. Arrival time distribution for ions in the 600-605 m/z range of an asphaltene sample. The peaks at shorter (~ 4.3 ms) and longer arrival times correspond to non-asphaltene and asphaltene signal, respectively. Average (peak apex) arrival time and arrival times for two standard deviations from the mean are indicated, assuming a Gaussian arrival time distribution for asphaltene ions.

Results and Discussion

DIESI-IM-HRMS Spectra of Asphaltenes

A sample of asphaltenes was analyzed with direct infusion electrospray ionization – ion mobility – high-resolution – mass spectrometry. The resulting IM-MS spectra (Figure 4.2) clearly show three distinct trendlines with ion signal across most of the arrival time (2-16 ms) and mass ranges (300-2000 m/z). The most intense ion signal is observed below m/z 800 and between 2 and 12 ms. Extracted mass spectra for each

trendline can be seen in Figure 4.3. As seen in Figure 4.3a, singly charged ion signal is distributed across ~ 400 - 1400 m/z with a maximum around m/z 750. This ESI-generated distribution and its maxima is in general agreement with MW distributions of asphaltene

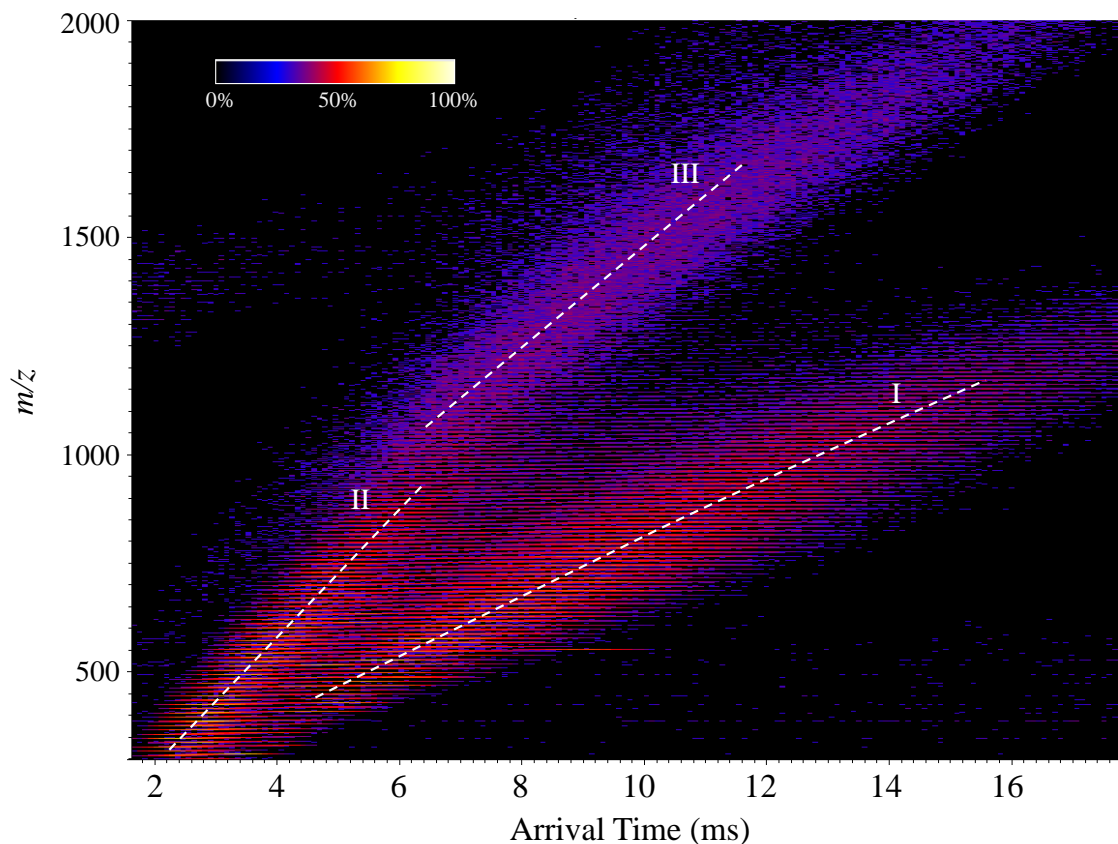


Figure 4.2. DIESI-IM-MS spectra of a 0.1 mg/mL asphaltene sample in 60:40 toluene-methanol at a mobility wave height of 7.5 V and gas flow rate of 25 mL/min. Three trendlines (I, II, and III) are observed.

molecules determined by LDI-IM-MS^{53,128} and other techniques^{130,144} which have shown that most asphaltene molecules are observed between ~ 350 and 2500 m/z and have a distribution maxima typically between 600 - 850 m/z . Thus, trendline I was presumed to be composed of singly charged asphaltene ions. Signal for trendline II (Figure 4.3b) has a more even intensity distribution across a lower m/z range (300 - 1000) and was determined to correspond to singly charged, background (non-asphaltene) ions because this trendline

was also observed in a toluene-methanol blank. Unassigned signal in trendline III (Figure 4.3c) is composed of singly charged ions with a much higher MW distribution (900-2400 m/z).

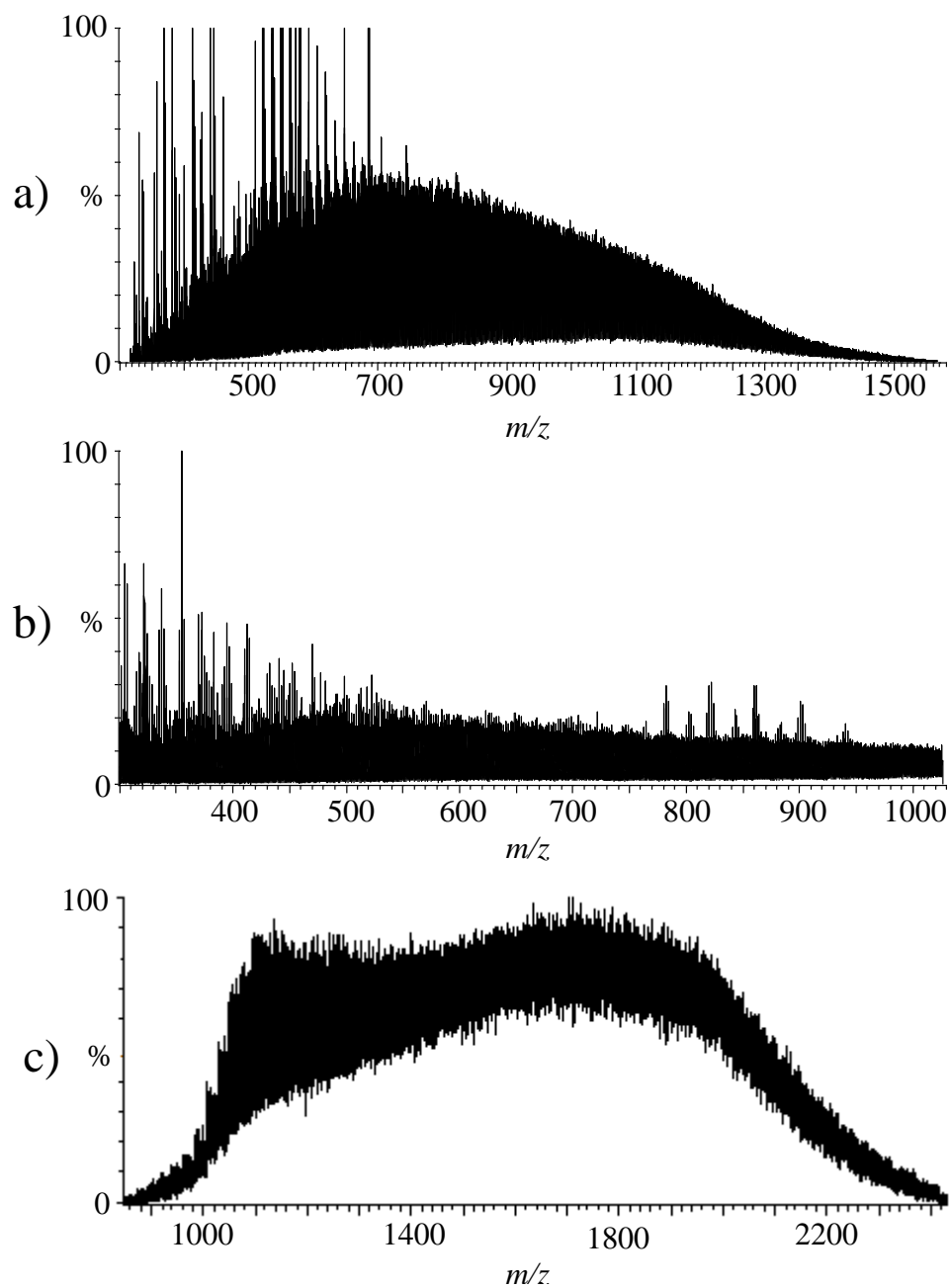


Figure 4.3. Extracted mass spectra for (a) trendline I (asphaltene signal), (b) trendline II (non-asphaltene signal), and (c) trendline III (unassigned signal) indicated in Figure 4.2.

Previous analyses of asphaltenes by IM-MS revealed the presence of gas-phase asphaltene aggregates above ~ 2000 m/z by IM-MS trendline analysis.¹²⁸ Specifically, asphaltene multimers exhibited higher mobility than monomers at similar m/z . The differences in mobility resulted in a noticeable change in trendline slope as ion signal transitioned from predominantly monomer to predominantly multimer species. A similar change in slope of the asphaltene trendline was not observed for singly charged asphaltene ions in the current IM-MS experiments. It is possible that aggregate species are not present in the current work due to differences in the ionization technique (ESI vs. LDI) or are generated but not as extensively and without sufficiently different CCSs to differentiate observed aggregates from monomer ions (i.e., trendline I may contain contributions from both monomer and multimer species). It is also possible that trendline III, which contains higher mobility ions with a higher molecular weight distribution (~ 900 - 2400 m/z) is composed of multimer asphaltene species.

Comparison of IM-MS Results for Asphaltenes and Model Compounds

Collision-cross-section measurements to evaluate general structure type. In order to determine average structure types of asphaltene molecules, collision cross sections were measured for asphaltene and model compound ions (see Table 4.1 for structures and other details). Figure 4.4 shows collision cross sections for model compounds and asphaltenes as a function of m/z . As can be observed in the plot, average asphaltene collision-cross-section values scale linearly with m/z . Average CCSs determined in this ESI-IM-MS study are slightly higher than those observed previously for asphaltenes with LDI-IM-MS.¹⁴⁵ Specifically, across the m/z range (400-1300) in Figure 4.3, the

difference in CCS for the average asphaltene trendline is 6-8% between the two methods.

The potential significance of this observation is presently unknown.

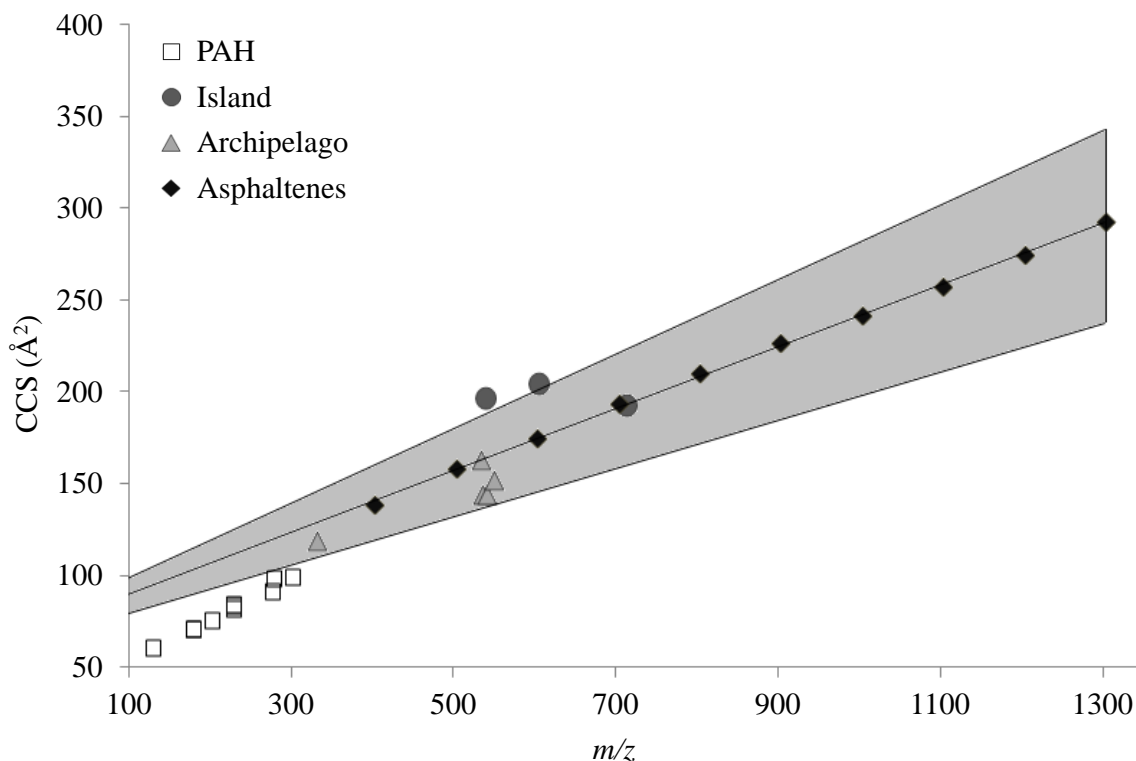


Figure 4.4. Comparison of collision cross sections for asphaltenes (average for peaks in a 5 Da window) and model compounds (Table 4.1), averaged over multiple mobility wave heights. Note that error bars for variation in CCS over the wave heights are not shown because they are smaller than the data points. The gray area around the asphaltene trendline accounts for 95.5% (2σ) of the collision cross section distribution of asphaltene ions. CCSs for outer asphaltene lines were calculated based on arrival times at 13.5% of max height. See Figure 4.1 for an illustration.

Model compounds are plotted for the average arrival time of the monoisotopic m/z peak and are classified by structure type (see Table 4.1 for classifications). Archipelago-type structures are generally observed to have a lower CCS for a given m/z than island-type structures, indicating that archipelago structures may fold over on themselves and present a smaller surface area to the buffer gas. The distinct difference in arrival time

(and CCS) can be clearly seen in Figure 4.5 for archipelago- (top) and island- (bottom) type structures. Average arrival time distributions for these two compounds of similar m/z (535 and 539 Da for $[M+H]^+$ ions) are separated by ~ 4 ms, with the island-type structure (THP) at a $\sim 67\%$ longer arrival time than the archipelago-type structure (PmPhP).

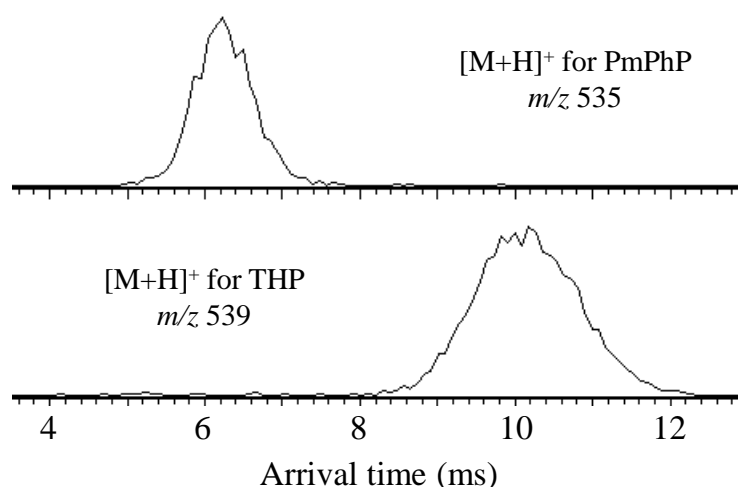


Figure 4.5. Arrival time distributions for $[M+H]^+$ ions of archipelago-type (PmPhP) and island-type (THP) structures at a wave height of 7.5 V.

Polyaromatic hydrocarbons (PAHs) fall off or just outside the lower asphaltene boundary trendline, assuming a linear extrapolation. These highly condensed structures with no alkyl chains may be expected to represent structures of low abundance for asphaltene molecules. Average H/C ratios for asphaltene molecules (1.0-1.2) are much higher than the H/C ratios of PAHs (0.4-0.8),¹⁴⁶ supporting CCS results that fully condensed PAHs are not very prevalent in asphaltenes. Archipelago-type structures fall on the average asphaltene trendline or within the outer boundaries, whereas one island-type structure (i.e., violanthrone-79) falls directly on the average asphaltene trendline and

both other island-type structures fall on the edge or just beyond the upper trendline. These results may indicate that the archipelago-type or certain island-type structures (i.e., a combination of extended alkyl chains and condensed aromatic structures) are more representative of asphaltene structures in this sample than the other evaluated island structures. The island structures that fall on the edge and beyond the upper trendline have H/C ratios of 1.26 and 1.45, respectively. These ratios are higher than the average of those for asphaltenes (1.0-1.2), supporting CCS analysis that these may be prevalent in low abundance in asphaltene samples. Note that archipelago structures with a larger central core connected to smaller outer aromatic cores via flexible alkyl linkages were not available for investigation and may also have CCSs that are consistent with asphaltene ion CCSs.

Arrival-time-peak-width measurements to assess structural diversity. Arrival-time-peak widths were measured at half maximum height for the monoisotopic peak of different model compound ions (i.e., all $M^{\bullet+}$, $[M+H]^+$, and $[M+Na]^+$ species observed). For model compounds, peak width scales linearly with average arrival time (Figure 4.6). A significant dependence of peak width on other traveling-wave IMS parameters (i.e., wave height and speed) was not observed. Thus, diffusion of ions in a traveling-wave drift cell appears directly correlated with ion residence time in the drift cell.¹⁴⁷ Peak widths for rigid PAHs (ions below 5 and 4 ms, respectively for 7.0 and 7.5 V in Figure 4.6) represent widths for single structures undergoing diffusion in the mobility cell. That is, these PAHs are not able to exist as multiple conformational isomers in the gas phase owing to their rigid structure. Model compounds plotted in Figure 4.6 having greater

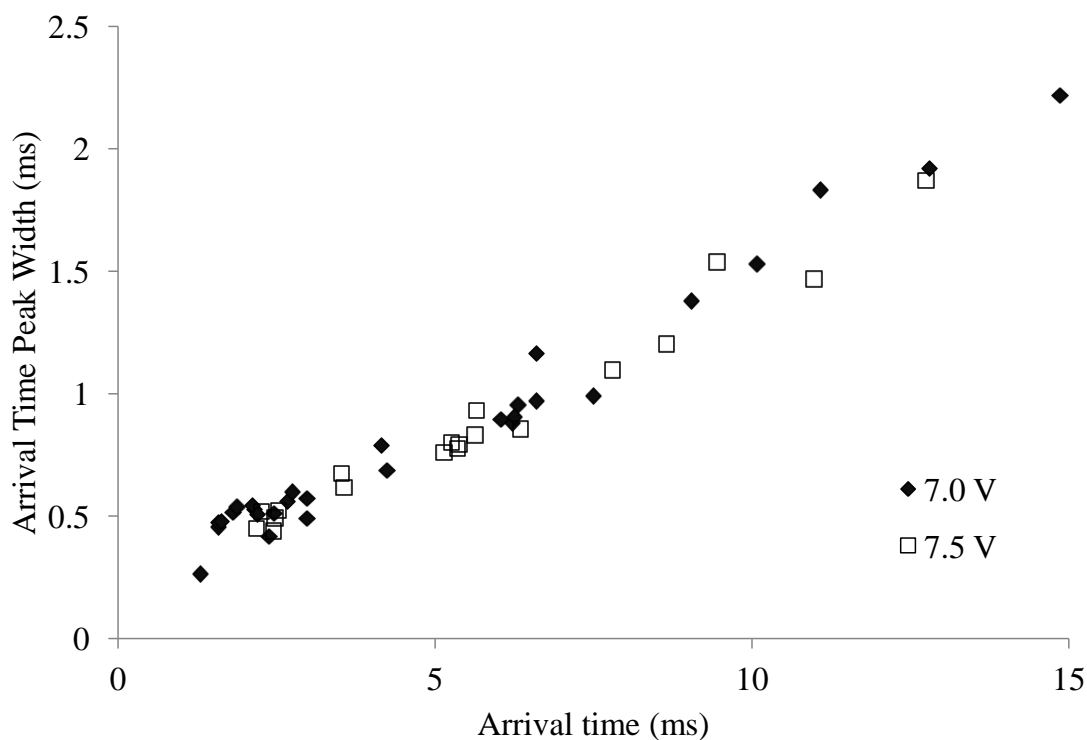


Figure 4.6. Arrival-time-peak-width as a function of arrival time for polyaromatic hydrocarbons and other model compounds used in this study (see Table 4.1) at two mobility cell wave heights.

flexibility from rotation about sp^3 -hybridized C-C bonds also follow the same linear trend as the rigid PAHs. We interpret these data to indicate the flexible model compounds exist predominately in a single gas-phase conformation that is stable under the investigated experimental conditions. Accordingly, other molecules which exist as a single conformation or structure would be expected to fall on the same line. However, ion signal which exhibits peak widths that deviate positively from the data shown in Figure 4.7 are expected to be from conformational or structural isomers that are not fully resolved at the timescale and conditions of the IMS experiment.¹⁴⁸ That is, the extent of peak broadening at a given arrival time relative to a rigid structure may be used to assess structural diversity.

For asphaltene ions, peak widths were measured for each nominal m/z asphaltene peak in the same 5-Dalton window used for collision-cross-section measurements. An example of these measurements for the 600-605 m/z window can be seen in Figure 4.7. Arrival time distributions (Figure 4.7, a-e) are shown for integration of each asphaltene peak (top left in the figure). The average of the five peak widths with error bars for standard deviation of those peak widths are plotted as a function of average arrival time in Figure 4.8, along with analogous measurements at intervals of 100 m/z across a 400-1100 m/z region. Peak widths for asphaltene ions are clearly much different than the peaks widths of model compound ions, with asphaltene ions having 66-95% larger peak widths than those on the model compound trendline (calculated based on the smallest

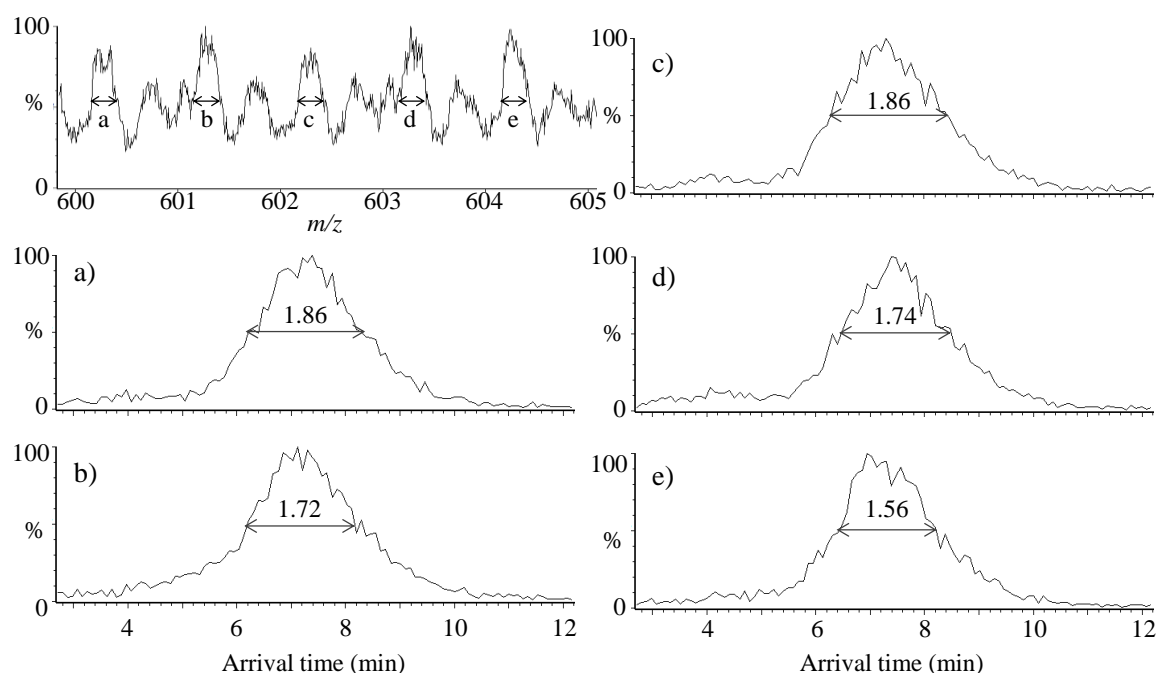


Figure 4.7. Mass spectrum (top left) of asphaltenes in the m/z range 600-605. Arrival time distributions (a-e) are shown for each selected nominal m/z peak in that range and peak-width measurements at half height are shown. M/z peaks in between the selected peaks are not derived from singly-charged asphaltene ions, as determined from extracted mass spectra of the trendlines in Figure 4.2.

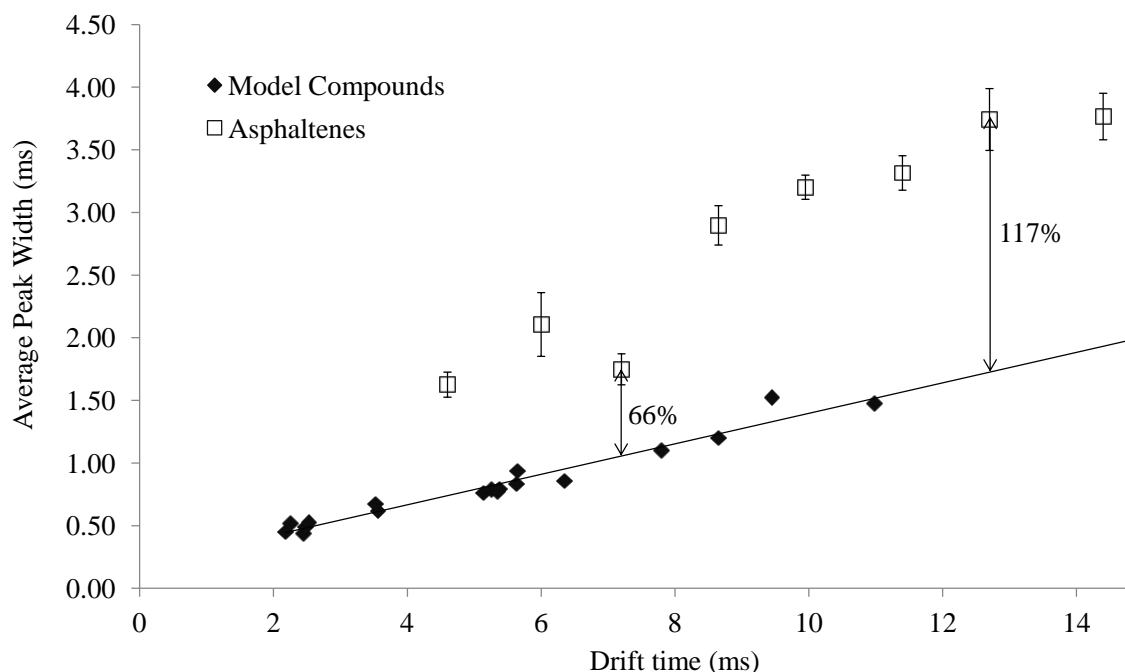


Figure 4.8. Average peak widths at half height for asphaltene and model compound ions as a function of average arrival time. Asphaltene peak widths are an average of the peak widths for each asphaltene peak in 5 Da windows. For example, the mobility peak widths for m/z 600, 601, 602, 603, and 604 were averaged for the 600-605 region (see Figure 4.7). Error bars represent one standard deviation for the 5 asphaltene peaks. Note that error bars for model compounds are smaller than the data points. A linear regression for the model compounds shows 66-117% larger peak widths for asphaltenes than the model compounds.

difference at 600-605 m/z and largest projected difference at 1000 m/z). This result suggests that asphaltene ions are adopting more than one conformation or that multiple structure types are present within the specified m/z window. While the former cannot be ruled out, there is ample evidence in literature supporting the latter.^{15,149} Additionally, no significant peak width deviations (from the rigid PAH trend) were observed for the flexible model compounds, also supporting multiple structure types for asphaltene ions in 5-Da windows.

Figure 4.9 represents the same peak width data but as a function of m/z for model compound and asphaltene ions. A general difference in peak width for model compounds

and asphaltenes can be observed. Although a linear correlation between peak width and m/z is observed for the PAHs (data below 400 m/z in Figure 4.9), other compound types (specifically those in the 530-605 m/z region) show a higher degree of diversity in peak width. Model compounds in this mass range include both island-type (e.g., chol-thio, THP and rubrene) and archipelago-type (e.g., PmPhP, PThP, and P-2,5-PMP) structures. The presence of these structural types at the same nominal m/z in an asphaltene sample would correspond to a wide arrival time distribution. For example, if equal contributions of PmPhP and rubrene (see Figure 4.10, top and middle, respectively, for their arrival time distributions) were present in a sample, the resulting arrival time peak would be much wider than a single component peak. Note that the arrival time distribution for asphaltene ions in a similar m/z range (Figure 4.10, bottom) is even wider than what

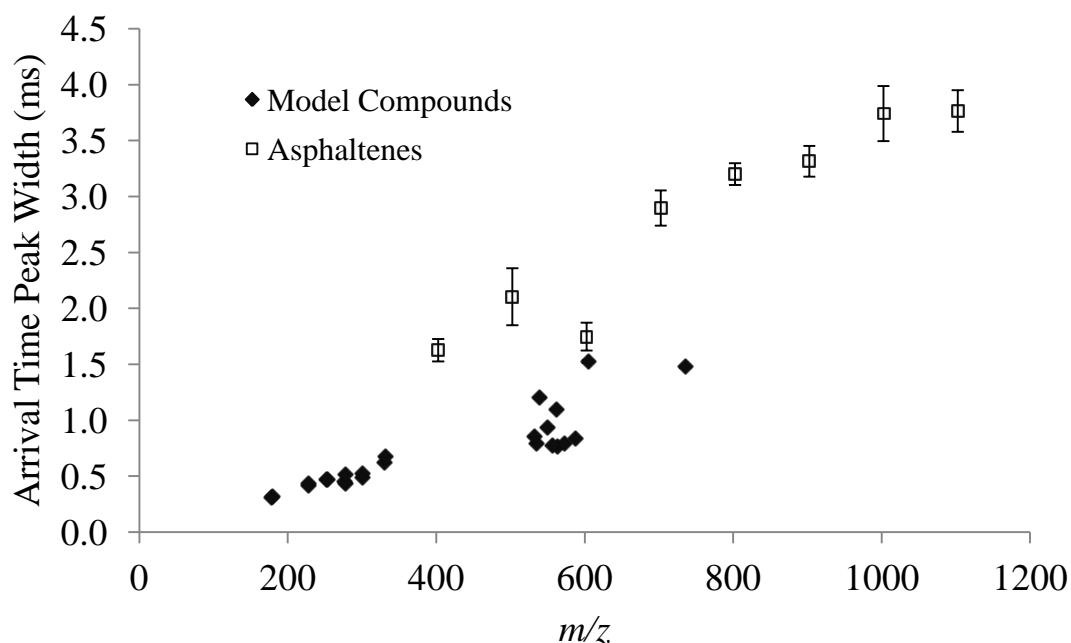


Figure 4.9. Average peak widths at half height for asphaltene and model compound ions as a function of m/z . See caption for Figure 4.8 for more details of peak width measurements.

would be observed for a combination of these two compounds, indicating a higher degree of structural diversity than that of PmPhP and rubrene. That is, the peak width distribution of model compounds in this m/z region provides further evidence for the presence of structural diversity of asphaltenes in a given m/z window.

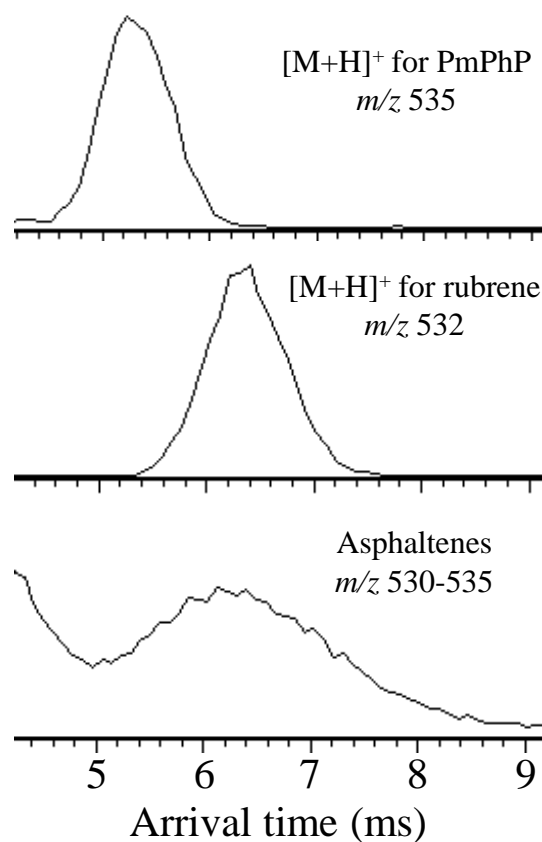


Figure 4.10. Arrival time distributions for model compounds (PmPhP and rubrene) and ions at similar m/z in an asphaltene sample.

Considerations in Asphaltene Aggregation

Whereas solution-phase aggregation of asphaltene molecules in crude oil presents a problem to the oil industry resulting in clogged pipelines and equipment, gas-phase aggregation of asphaltenes has made it difficult for mass spectrometrists to determine

accurate asphaltene molecular weight distributions for the last several decades. As discussed previously in this chapter, gas-phase aggregation has been previously observed in IM-MS experiments. Aggregation could result from multiple ions of the same compound or different compounds forming multimers (e.g., dimers, trimers, etc.). Preliminary investigations of gas-phase formation of multimers for the model compounds used in this work were carried out to determine the types of structures that might form aggregates in the gas phase.

Dimer formation of all combinations of chol-thio, THP, PmPhP, PThP, and P-2,5-PMP were investigated by mixing each possible pair and spraying together. Under the same ESI and IMS conditions utilized for collision-cross-section and peak-width measurements, dimers were observed for one archipelago-type structure (i.e., P-2,5-PMP), two island-type structures (i.e., THP and violanthrone-79), and one mixed island-archipelago pair (i.e., chol-thio and P-2,5-PMP). Sodium-adducted dimers, $[2M+Na]^+$, were observed for all four dimers and the protonated dimer, $[2M+H]^+$, was observed for P-2,5-PMP and chol-thio/P-2,5-PMP. No other dimers of mixed model compounds were observed. After inducing dimer formation by decreasing ESI parameters (i.e., source temperature, 80 °C; desolvation gas flow rate, 175 L/hr; desolvation temperature, 150 °C; and cone gas flow rate, 0 L/hr) and increasing solution concentration (by a factor of 10), chol-thio also showed sodium-adducted dimer formation.

The type and degree of dimer formation among these compounds can be indicative of the type of aggregation observed for asphaltene ions. Collision cross sections for these dimers fall within the asphaltene CCS distributions, as shown in Figure 4.11. These data confirm that ion signal in this mass range may also be composed of

dimers or other multimers in addition to individual molecules. For the model compounds investigated in this work, aggregates larger than dimers were either not observed or corresponding signal was very weak.

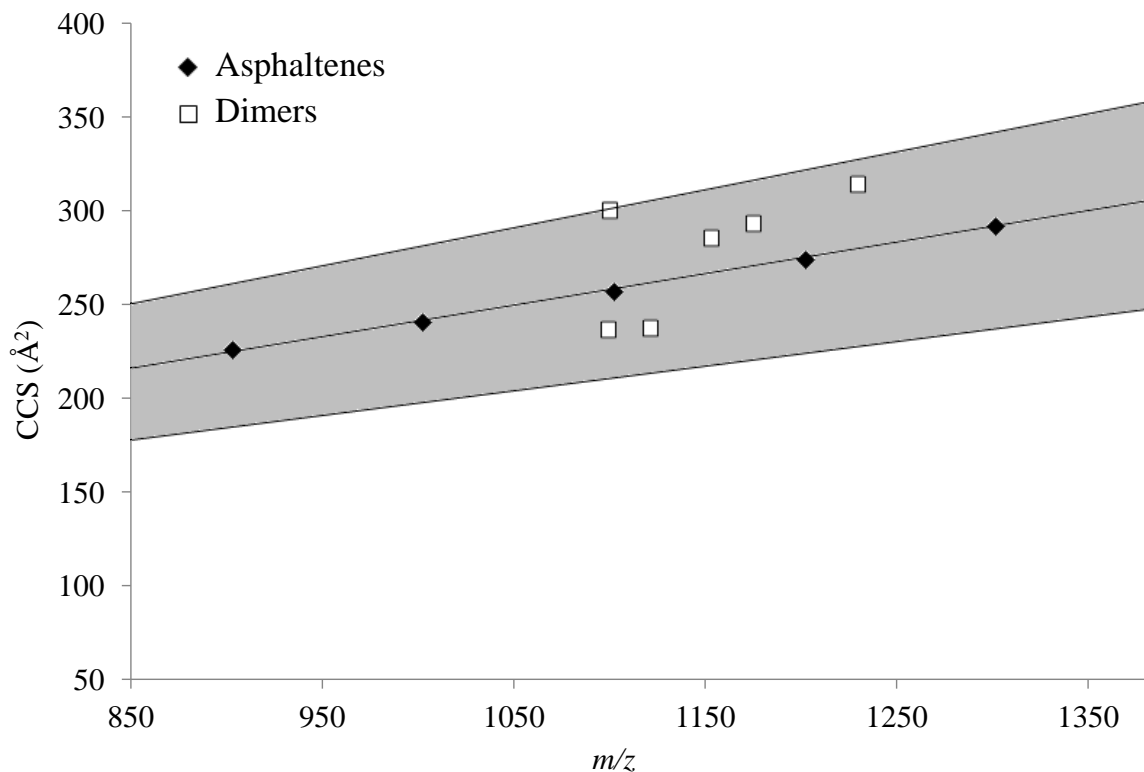


Figure 4.11. Comparison of collision cross sections for asphaltenes (average for peaks in a 5 Da window) and observed model compound dimers (see text for details). Asphaltene CCSs and trendlines are shown as in Figure 4.4.

It is difficult to pinpoint structural features dictating the degree of dimer formation. Even for the limited number of model compounds used in this study, those forming dimers represent a diverse set of compounds. Dimers were observed for island-island, archipelago-archipelago and island-archipelago type interactions; compounds with and without heteroatoms; compounds with and without a multi-ring PAH core; and compounds with and without alkyl substituents. It is important to note that intensity of

dimer signal was often observed between 15% and 100% of the monomer signal indicating that ion signal from dimers in asphaltenes may be expected to contribute significantly to asphaltene ion signal, when prevalent.

Conclusions

The results presented in this chapter have demonstrated how interpretation of IM-MS data can elucidate structural information in petroleomic samples. Comparison of collision cross sections and peak widths from IM-MS experiments provides informative data on asphaltene structure type and diversity. These data represent the first comparative evaluation of CCS between representative archipelago and island-type model structures and asphaltenes. We have demonstrated that both structural types are possible; however, of those investigated in this work, archipelago structures were more similar in CCS to average asphaltene distribution. Collision cross sections for observed model compound dimers appear within the thresholds for asphaltene ions, indicating that asphaltene aggregates and monomers are likely to appear in the same region of IM-MS spectra.

CHAPTER FIVE

Future Directions

One conclusion that may be drawn from the data described in this work is that improvements in resolution for both mobility and mass dimensions would be beneficial to the developed methodologies. Since the onset of these projects, the next generation Synapt ion mobility – mass spectrometer (G2) has been released. The G2 provides ~ 10x improvement in mobility resolving power and 2-4x improvement in MS resolving power. Moreover, high data-rate acquisition technologies have been incorporated into the electronics, improving data sampling. Such advances would significantly improve resolution of isomers and isobars present in these types of complex samples and result in improved mass accuracy, leading to a higher number of formula assignments for unknown compounds. Additionally, MS^c experiments with inclusion of ion mobility separations (only a feature of the G2) would enable time-alignment of fragments both in the LC and IM dimensions, offering enhanced matching of fragment and parent ions.

To exhaustively identify biomass pretreatment degradation products, the restrictions imposed for data collection and processing imposed in Chapter 3 will need to be removed. In the current work, data was collected for negative ions up to 500 m/z in an undiluted, extracted pretreatment hydrolysate in order to establish the methodology. The high mass range observed in Chapter 2 would indicate that the automated methodology should also be applied to higher m/z ions in unextracted hydrolysates. Analysis of unextracted hydrolysates would provide a more representative understanding of

hydrolysate composition. With an increased m/z range and analysis of an unextracted hydrolysate, KMD analysis will likely be useful in identifying larger structural groups (e.g., identification of carbohydrates via “C₆H₁₀O₆” and “C₅H₈O₅” units) and homologous series with more than two or three members. Intense ions observed in the undiluted sample (beyond the dynamic range of the detector) were excluded from the results in Chapter 3 due to poor mass accuracy. Including analysis of a diluted sample should allow for accurate identification of these excluded ions. Additionally, considering the prevalence of nitrogen-containing compounds in AFEX pretreatment hydrolysates, positive electrospray ionization mode could be employed for detection of nitrogen-containing compounds that may not ionize well in negative ion mode.

As indicated in Chapter 3, many compounds in the pretreatment hydrolysates eluted in the 30-35 minute region of the chromatographic run. To simplify mass spectra and combat low ionization efficiency for analytes in this region due to the presence of so many co-eluting species, an extended run time and decreased gradient slope is recommended. However, when run time was extended by 40 minutes with a corresponding decrease in gradient slope, abundant ion signal in the last half of the run can still be seen (Figure 5.1). The number of peaks in this region provides further evidence of the importance of an “x-omics” approach in identification of biomass degradation products.

The method for identification of pretreatment products presented in Chapter 3 does not have a built-in means to determine product origin (i.e., cellulose, hemicellulose, lignin, pretreatment process, etc.). However, application of the method to pretreated cellulose, hemicellulose and lignin might assist in identification of degradation products

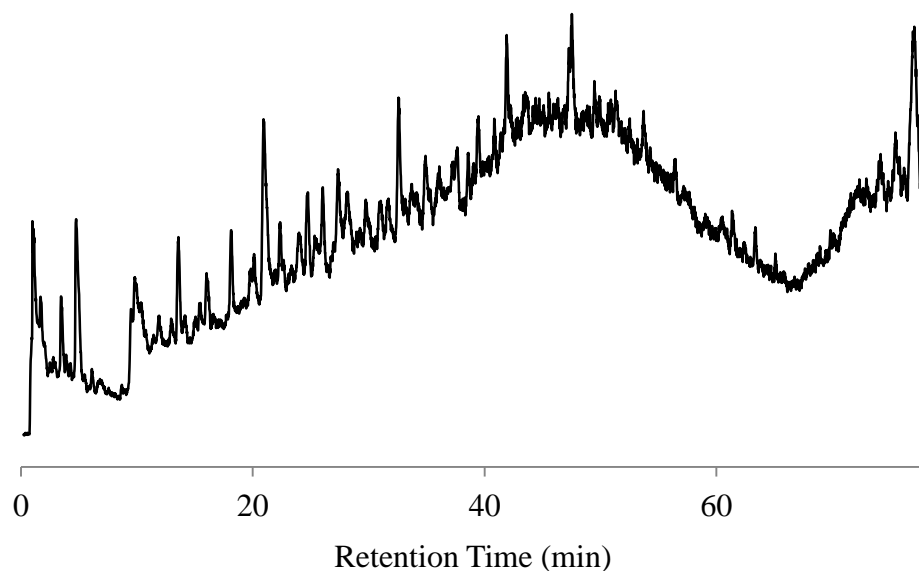


Figure 5.1. UPLC-HR-TOF-MS chromatogram of dilute-acid-pretreated hydrolysate of corn stover with an extended run time (90 min vs 51 min) relative to the method presented in Chapter 3.

from each of these biomass components. Figure 5.2 shows pretreated xylan (hemicellulose-like), inulin (polysaccharide), cellulose (glucose polysaccharide), and avicel (microcrystalline cellulose), which clearly show an abundance of degradation products. Analysis of these and similar samples should supply a catalog of possible degradation products, generated from a specific precursor, which can then be compared to pretreated biomass.

Chemometric modeling of chromatographic data (i.e., HPLC with either UV or conductivity detection) was recently used in our laboratory to identify chromatographic retention times most strongly correlated with compounds responsible for microbial inhibition.¹⁵⁰ The utility of this method was limited by detector selectivity, insufficient chromatographic resolution in certain regions of the chromatogram and the inability of either detector to provide both structural and compositional information. A similar chemometric approach developed using the UPLC-HRMS method described in Chapter 3

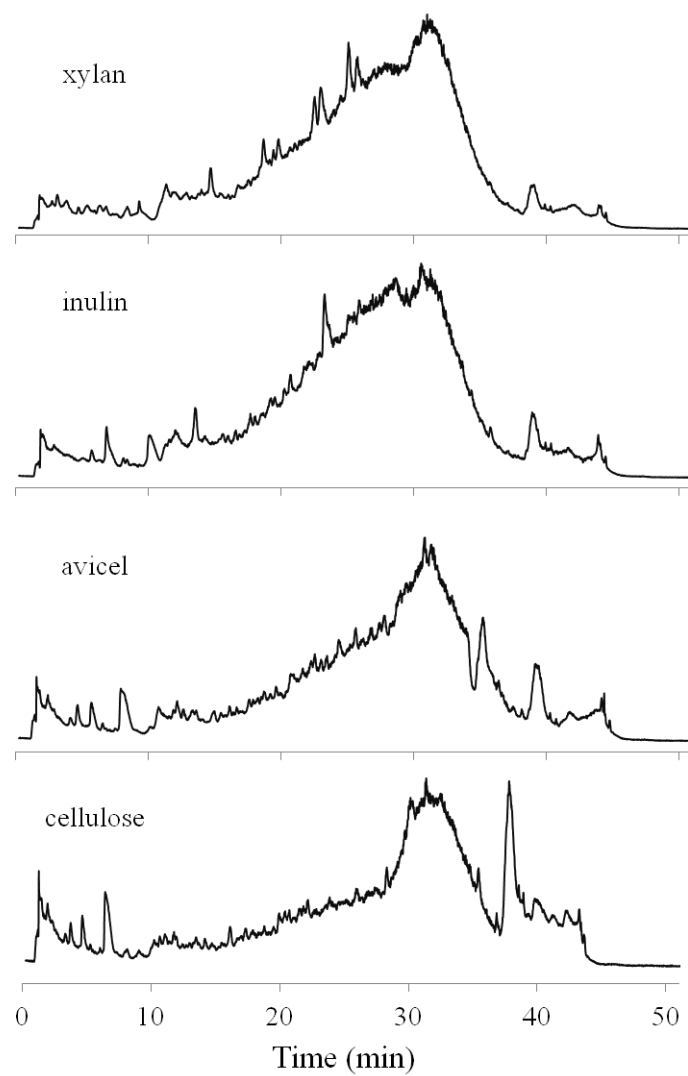


Figure 5.2. Chromatograms of MTBE-extracts of 0.7% H_2SO_4 -pretreated xylan, inulin, avicel, and cellulose.

would provide improvements in chromatographic resolution and a more universal detector capable of determining molecular formula via exact mass and structure via CID. Alternatively, with the increased peak capacity (i.e., higher resolving power in both the IM and HRMS dimensions, see Figure 5.3 for example) of the Synapt G2, chemometric modeling of IM-HRMS data could provide a technique capable of obtaining the same information (correlated with drift times rather than retention times) but with analysis time of mere seconds or minutes.

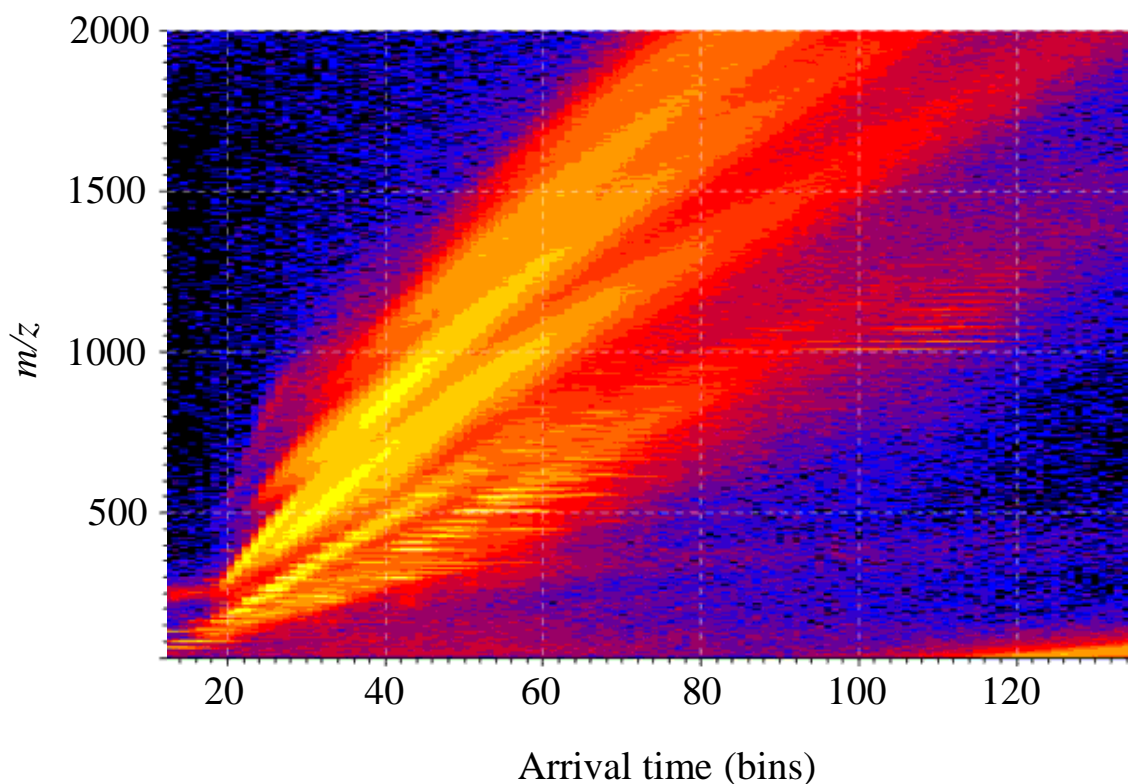


Figure 5.3. IM-MS spectra obtained with a next generation Synapt (G2) for an unextracted, dilute- acid pretreated hydrolysate of corn stover. Peak capacity is increased by a factor of ~10 relative to instrumentation used in this dissertation work.

APPENDICES

APPENDIX A

Collision-Cross-Section Measurements

Lignocellulosic Analytes (Chapter 2)

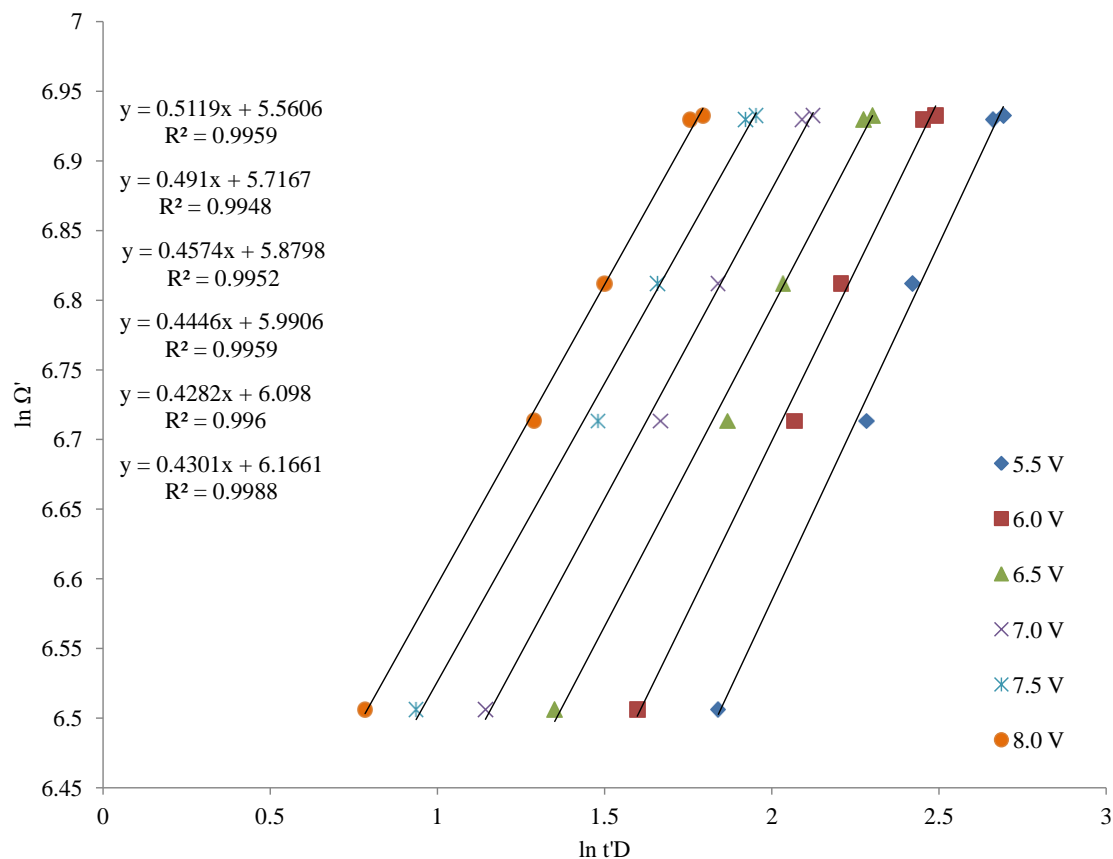


Figure A.1. Plot of $\ln(\Omega') = X \times \ln(t'D)$ used as the first step in collision-cross-section measurement of the lignocellulosic analytes described in Chapter 2. $[M - H]^-$ ions generated from raffinose, melezitose, and α -cyclodextrin standards, as well as fragment ions of m/z 221 and 323 resulting from collision-induced dissociation of α -cyclodextrin and melezitose, respectively, were used as calibrants. Ω' and $t'D$ are calculated as follows: $\Omega' = \Omega / [\text{charge} \times (1/\mu)^{1/2}]$ and $t'D = t_D - [1.41(\sqrt{(m/z)})/1000]$, where Ω is collision cross section of the calibrant, charge is the charge on the calibrant ion, μ is the reduced mass of the analyte ion and neutral gas molecule, and t_D is arrival time (ms).

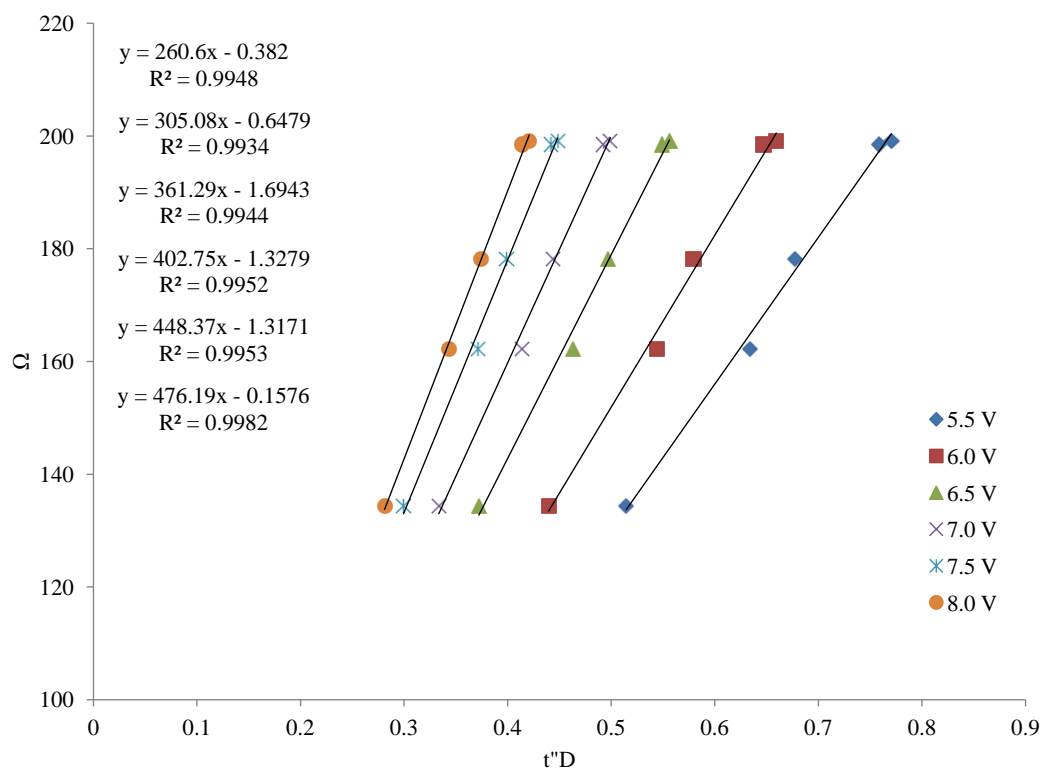


Figure A.2. Calibration plot used for collision-cross-section measurement of the lignocellulosic analytes, described in Chapter 2. The corrected arrival time, $t''D$, is calculated by $t''D = t'_D{}^X * \text{charge} \times (1/\mu)^{1/2}$, where X is the slope from the plot in Figure A.1.

Table A.1. Collision cross sections (CCSs) for $[M - H]^-$ ions at exact masses corresponding to aromatic acid and aldehyde degradation products standards at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, except for those with a CCS below 115 \AA^2 or above 122.5 \AA^2 . CCSs at 6.5 V were excluded for those below 115 \AA^2 due to field-strength-dependent effects present. For ions with CCS above 122.5 \AA^2 , CCSs at 5.5 V were excluded. Excluded data are denoted by *.

Compound	<i>m/z</i>	Collision cross section (Å ²)					
		5.5 V	5.8 V	6.0 V	6.5 V	Avg	RSD
malonic acid	103	91.0	92.3	94.5*	102.2*	92.6	N/A
lactic acid	89	98.3	98.8	100.1	106.6*	99.1	0.9
maleic acid	115	93.6	94.3	95.4	102.1*	94.4	1.0
cis-aconitic acid	173	111.5	108.8	109.5	110.4*	109.9	1.3
methylmalonic acid	117	98.4	97.7	99.1	103.5*	98.4	0.7
succinic acid	117	101.1	99.4	101.0	105.7*	100.5	0.9
fumaric acid	115	101.6	100.4	101.6	106.0*	101.2	0.7
trans-aconitic acid	173	111.6	109.5	109.2	111.7*	110.5	1.2
levulinic acid	115	116.7	117.0	118.1	121.9	118.4	2.0
glutaric acid	131	107.6	106.6	107.0	108.8*	107.1	0.5
itaconic acid	129	104.1	102.7	103.0	106.9*	103.3	0.7
2-hydroxy-3-methylbutyric acid	117	121.5	121.1	119.5	124.8	121.7	1.8
2-furoic acid	111	95.0	94.6	96.7	103.0*	95.4	1.2
gallic acid	169	114.3	113.3	113.2	115.2*	113.6	0.8
adipic acid	145	123.1	120.2	121.7	124.7	121.6	1.2
3,4-dihydroxybenzoic acid	153	115.2	116.3	115.2	120.4	116.8	2.1
3,5-dihydroxybenzoic acid	153	121.2	119.7	118.9	120.0	120.0	0.8
2,5-dihydroxybenzoic acid	153	112.2	110.7	110.0	111.7*	111.1	0.9
3,4-dihydroxybenzaldehyde	137	115.7	115.1	116.2	119.6	116.7	1.7
4-hydroxybenzoic acid	137	107.5	106.9	106.6	108.9*	107.0	0.4
4-hydroxybenzaldehyde	121	109.9	108.9	109.4	113.7*	109.4	0.5
vanillic acid	167	130.4*	128.0	128.1	128.4	128.2	0.9
homovanillic acid	181	134.2*	131.0	130.7	130.1	130.6	0.3
4-hydroxyacetophenone	135	121.2	120.9	120.9	123.7	121.0	1.1
caffeic acid	179	126.0*	123.2	123.3	124.4	123.6	1.1
syringic acid	197	144.5*	139.6	139.3	139.2	139.4	0.2
salicylic acid	137	106.4	105.4	105.5	108.5*	105.7	0.5
vanillin	151	116.5	114.2	114.5	115.8	115.3	0.9
benzoic acid	121	108.7	106.3	107.0	109.7*	107.3	1.2
syringaldehyde	181	129.0*	126.2	125.8	126.0	126.0	1.2
p-coumaric acid	163	121.9	120.5	120.6	121.8	121.2	0.6
ferulic acid	193	137.0*	135.2	132.8	134.9	134.3	1.0
sinapic acid	223	147.5*	143.0	142.6	141.9	142.5	0.4
3-hydroxy-4-methoxycinnamic acid	193	149.9*	144.3	143.8	143.7	143.9	0.2
4-hydroxycoumarin	161	116.9	115.4	115.5	116.6	116.1	0.7
o-toluic acid	135	117.9	115.3	114.8	116.4	116.1	1.2
p-toluic acid	135	119.4	116.7	116.4	117.7	117.6	1.2

Table A.2. Collision cross sections (CCSs) for $[M - H]^-$ ions at exact masses corresponding to aromatic acid and aldehyde degradation products in corn stover hydrolysate at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, except for those with a CCS below 115 \AA^2 or above 122.5 \AA^2 . CCSs at 6.5 V were excluded for those below 115 \AA^2 due to field-strength-dependent effects present. For ions with CCS above 122.5 \AA^2 , CCSs at 5.5 V were excluded. Excluded data are denoted by *.

Compound	m/z	Collision cross section (\AA^2)						RSD
		5.5 V	5.8 V	6.0 V	6.5 V	Avg		
lactic acid	89	107.4	106.3	107.7	113.3*	107.1	0.7	
malonic acid	103	99.2	98.9	100.4	107.1*	99.5	0.8	
2-furoic acid	111	101.6	100.8	101.5	107.1*	101.3	0.4	
maleic acid	115							
fumaric acid	115	107.5	106.2	107.4	112.2*	107.0	0.6	
levulinic acid	115							
methylmalonic acid	117							
succinic acid	117	107.3	106.1	106.3	110.0*	106.6	0.6	
2-hydroxy-3-methylbutyric acid	117							
4-hydroxybenzaldehyde	121	109.9	108.9	109.4	112.2*	106.6	0.6	
benzoic acid	121							
itaconic acid	129	110.7	108.3	108.8	113.0*	109.2	1.2	
glutaric acid	131	109.1	107.1	107.0	110.8*	108.5	1.7	
4-hydroxyacetophenone	135							
o-toluic acid	135	113.1	110.9	111.7	113.3*	111.9	1.0	
p-toluic acid	135							
3,4-dihydroxybenzaldehyde	137							
4-hydroxybenzoic acid	137	110.5	109.2	109.5	112.9*	109.8	0.6	
salicylic acid	137							
adipic acid	145	115.2	113.2	113.4	113.8*	113.9	0.8	
vanillin	151	117.4	114.2	114.1	117.2	115.7	1.6	
3,4-dihydroxybenzoic acid	153							
3,5-dihydroxybenzoic acid	153	114.0	112.0	112.1	115.2*	112.7	1.4	
2,5-dihydroxybenzoic acid	153							
4-hydroxycoumarin	161	118.4	115.7	115.5	116.6	116.6	1.1	
p-coumaric acid	163	123.2	119.1	120.3	120.1	120.7	1.4	
vanillic acid	167	119.1	117.6	116.0	116.3	117.2	1.2	
gallic acid	169	117.7	116.1	117.4	118.0	117.3	0.7	

Table A.2. Collision cross sections (CCSs) for $[M - H]^-$ ions at exact masses corresponding to aromatic acid and aldehyd degradation products in corn stover hydrolysate at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, expect for those with a CCS below 115 \AA^2 or above 122.5 \AA^2 . CCSs at 6.5 V were excluded for those below 115 \AA^2 due to field-strength-dependent effects present. For ions with CCS above 122.5 \AA^2 , CCSs at 5.5 V were excluded. Excluded data are denoted by *. (cont'd)

Compound	<i>m/z</i>	Collision cross section (Å ²)					
		5.5 V	5.8 V	6.0 V	6.5 V	Avg	RSD
cis-aconitic acid	173	115.1	113.1	113.0	114.2*	113.7	0.9
trans-aconitic acid	173						
caffeic acid	179	125.2*	123.0	123.0	123.2	123.0	0.1
homovanillic acid	181	130.7*	127.0	127.1	127.6	127.3	0.3
syringaldehyde	181						
ferulic acid	193	134.8*	130.5	131.5	130.3	130.7	0.5
3-hydroxy-4-methoxycinnamic acid	193						
syringic acid	197	133.4*	130.3	131.3	131.7	131.1	0.5
sinapic acid	223	132.8*	129.7	130.3	130.7	130.3	0.4

Table A.3. Collision cross sections (CCSs) for $[M - H]^-$ ions at exact masses corresponding to aromatic acid and alehyde degradation products in sorghum hydrolysate at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, expect for those with a CCS below 115 \AA^2 or above 122.5 \AA^2 . CCSs at 6.5 V were excluded for those below 115 \AA^2 due to field-strength-dependent effects present. For ions with CCS above 122.5 \AA^2 , CCSs at 5.5 V were excluded. Excluded data are denoted by *.

Compound	<i>m/z</i>	Collision cross section (Å ²)					
		5.5 V	5.8 V	6.0 V	6.5 V	Avg	RSD
lactic acid	89				-		
malonic acid	103				-		
2-furoic acid	111				-		
maleic acid	115						
fumaric acid	115	116.2	115.5	114.9	120.0	116.7	2.0
levulinic acid	115						
methylmalonic acid	117						
succinic acid	117				-		
2-hydroxy-3-methylbutyric acid	117						
4-hydroxybenzaldehyde	121	114.2	113.5	114.4	119.4*	114.0	0.4
benzoic acid	121						

Table A.3. Collision cross sections (CCSs) for $[M - H]^-$ ions at exact masses corresponding to aromatic acid and aldehyde degradation products in sorghum hydrolysate at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, except for those with a CCS below 115 \AA^2 or above 122.5 \AA^2 . CCSs at 6.5 V were excluded for those below 115 \AA^2 due to field-strength-dependent effects present. For ions with CCS above 122.5 \AA^2 , CCSs at 5.5 V were excluded. Excluded data are denoted by *. (cont'd)

Compound	<i>m/z</i>	Collision cross section (Å ²)						RSD
		5.5 V	5.8 V	6.0 V	6.5 V	Avg		
itaconic acid	129	113.5	112.8	113.7	116.9*	113.3	0.4	
glutaric acid	131	117.5	115.7	116.7	118.6	117.1	1.0	
4-hydroxyacetophenone	135							
o-toluic acid	135	117.2	115.3	116.4	118.3	116.8	1.1	
p-toluic acid	135							
3,4-dihydroxybenzaldehyde	137							
4-hydroxybenzoic acid	137	112.9	113.7	113.1	118.1*	113.2	0.4	
salicylic acid	137							
adipic acid	145	125.6*	124.2	123.1	124.7	124.0	0.7	
vanillin	151	120.0	117.0	117.6	119.0	118.4	1.1	
3,4-dihydroxybenzoic acid	153							
3,5-dihydroxybenzoic acid	153	116.0	115.8	115.2	117.1	116.0	0.7	
2,5-dihydroxybenzoic acid	153							
4-hydroxycoumarin	161	119.7	119.2	120.7	122.0	120.4	1.0	
p-coumaric acid	163	124.4*	121.8	122.0	123.6	124.0	0.7	
vanillic acid	167	122.9	121.6	120.4	123.4	122.1	1.1	
gallic acid	169				-			
cis-aconitic acid	173							
trans-aconitic acid	173	131.1*	129.0	128.5	129.7	129.1	0.9	
caffeic acid	179	128.4*	124.9	125.4	124.4	124.9	0.4	
homovanillic acid	181							
syringaldehyde	181	130.7*	127.3	126.7	127.6	127.2	0.4	
ferulic acid	193							
3-hydroxy-4-methoxycinnamic acid	193	135.9*	132.9	132.8	131.9	132.5	0.4	
syringic acid	197				-			
sinapic acid	223	147.5*	144.1	143.7	143.8	144.2	0.1	

Table A.4. Collision cross sections (CCSs) for $[M - H]^-$ ions of carbohydrate standards at six ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, except for those with a CCS below 121 \AA^2 or above 227 \AA^2 . CCS at higher wave heights (7.0-8.0 V) for those below 120 \AA^2 were excluded due to field-strength-dependent effects present at higher wave heights. Excluded data are denoted by *. For ions with CCS above 215 \AA^2 , arrival time peaks were not observed in the arrival time window (0-18 ms) at smaller wave heights, indicated by “-”.

Compounds	m/z	Collision cross section (\AA^2)						Avg	RSD
		5.5 V	6.0 V	6.5 V	7.0 V	7.5 V	8.0 V		
glucose	179	118.7	117.8	118.3	120.1*	124.2*	129.7*	118.3	0.4
maltose	341	165.4	166.1	166.5	166.4	166.6	165.1	166.0	0.4
maltotriose	503	211.0	211.5	210.0	210.2	210.0	211.1	210.6	0.3
isomaltotriose	503	206.2	205.8	204.4	204.7	204.8	205.6	205.3	0.3
raffinose	503	200.6	200.7	199.5	200.0	199.9	200.5	200.2	0.2
melezitose	503	197.4	197.0	197.0	197.1	197.2	197.3	197.1	0.1
maltotetraose	665	-	230.0	226.4	226.0	226.1	227.4	227.2	0.7
stachyose	665	-	229.4	227.1	226.8	227.0	228.4	227.7	0.5
maltopentaose	827	-	-	258.4	258.5	257.0	258.8	258.2	0.3
xylose	149	109.5	109.8	111.9	114.4*	122.4*	130.1*	110.4	1.2
xylobiose	281	150.5	150.8	150.9	150.3	150.9	148.7	150.3	0.5
xylotetraose	413	215.2	215.5	213.5	214.2	214.3	215.4	214.7	0.4

Table A.5. Collision cross sections (CCSs) for negatively charged ions at exact masses corresponding to carbohydrates in corn stover hydrolysate at six ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights, except for those with a CCS below 130 \AA^2 or above 227 \AA^2 . CCS at higher wave heights for smaller ions (6.5-8.0 V for those below 114 \AA^2 , 7.0-8.0 V for those between 114 and 121 \AA^2 and 7.5-8.0 for those between 121 and 130 \AA^2) were excluded due to field-strength-dependent effects present at higher wave heights. Excluded data are denoted by *. For ions with CCS above 215 \AA^2 , arrival time peaks were not observed in the arrival time window (0-18 ms) at smaller wave heights, indicated by “-”.

Carbohydrate Type	DP unit	Ion	Collision Cross Section (\AA^2)								
			m/z	5.5 V	6.0 V	6.5 V	7.0 V	7.5 V	8.0 V	Avg	RSD
6-carbon	1	$[\text{M}-\text{H}]^-$	179	120.4	120.6	120.8	122.1*	127.6*	132.2*	120.6	0.2
		$[\text{M}+\text{Cl}]^-$	215	135.0	135.9	137.4	136.8	136.5	137.8	136.6	0.7
		$[\text{M}+\text{HSO}_4]^-$	277	143.5	144.6	146.0	144.5	145.8	143.7	144.7	0.7
	2	$[\text{M}-\text{H}]^-$	341	167.0	167.9	168.6	168.2	168.3	167.6	167.9	0.3
		$[\text{M}+\text{Cl}]^-$	377	185.2	185.4	184.6	185.4	185.9	185.8	185.4	0.3
		$[\text{M}+\text{HSO}_4]^-$	439	184.3	185.3	184.6	185.5	186.2	186.3	185.4	0.4
	3	$[\text{M}-\text{H}]^-$	503	210.4	210.8	209.2	209.8	210.0	210.5	210.1	0.3
		$[\text{M}+\text{Cl}]^-$	539	-	216.6	215.0	213.4	214.7	216.1	215.2	0.6
		$[\text{M}+\text{HSO}_4]^-$	601	-	218.1	215.9	216.4	216.1	217.8	216.9	0.5
	4	$[\text{M}-\text{H}]^-$	665	-	-	235.3	234.7	235.9	236.8	235.7	0.4
	5	$[\text{M}-\text{H}]^-$	827	-	-	-	256.4	255.4	257.9	256.6	0.5
5-carbon	1	$[\text{M}-\text{H}]^-$	149	113.2	114.3	117.3*	119.7*	127.1*	134.0*	113.8	0.0
		$[\text{M}+\text{Cl}]^-$	185	126.1	127.0	127.7	127.6	130.6*	134.4*	127.1	0.6
		$[\text{M}+\text{HSO}_4]^-$	247	146.1	146.5	146.9	145.3	144.9	143.4	145.5	0.9
	2	$[\text{M}-\text{H}]^-$	281	150.5	150.8	151.5	151.7	150.9	149.7	150.8	0.5
		$[\text{M}+\text{Cl}]^-$	317	157.2	158.0	158.2	159.1	159.6	157.5	158.3	0.6
		$[\text{M}+\text{HSO}_4]^-$	379	167.9	169.1	170.0	169.4	170.2	170.3	169.5	0.5
	3	$[\text{M}-\text{H}]^-$	413	190.7	191.2	190.3	191.1	191.4	192.3	191.2	0.3
		$[\text{M}+\text{Cl}]^-$	449	194.2	194.6	194.2	193.7	194.4	194.5	194.3	0.2
		$[\text{M}+\text{HSO}_4]^-$	511	202.4	202.8	201.9	201.8	202.6	203.0	202.3	0.2
	4	$[\text{M}-\text{H}]^-$	545	212.9	213.2	211.1	211.6	212.7	212.5	212.3	0.4
		$[\text{M}+\text{Cl}]^-$	581	-	220.9	219.1	219.1	219.2	220.9	219.8	0.4
		$[\text{M}+\text{HSO}_4]^-$	643	-	221.8	219.3	220.3	219.7	222.0	220.6	0.6

Petroleomic Analytes (Chapter 4)

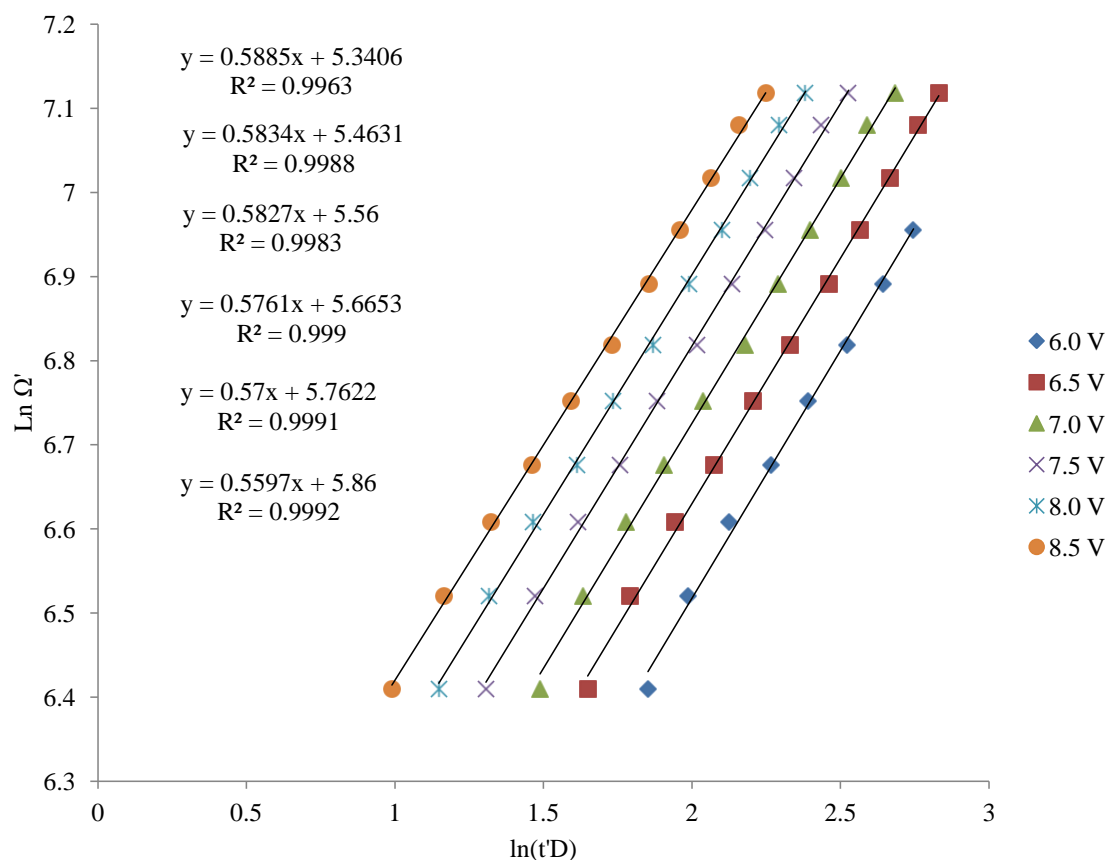


Figure A.3. Plot of $\ln(\Omega') = X \times \ln(t'D)$ used as the first step in collision-cross-section measurement of the model compounds and petroleomic analytes described in Chapter 4. $[M + H]^+$ ions generated from polypropylene glycol (PPG) standards (DP 5-15) were used as calibrants. Ω' and $t'D$ are calculated as follows: $\Omega' = \Omega / [\text{charge} \times (1/\mu)^{1/2}]$ and $t'D = t_D - [1.41(\sqrt{m/z})/1000]$, where Ω is collision cross section of the calibrant, charge is the charge on the calibrant ion, μ is the reduced mass of the analyte ion and neutral gas molecule, and t_D is arrival time (ms).

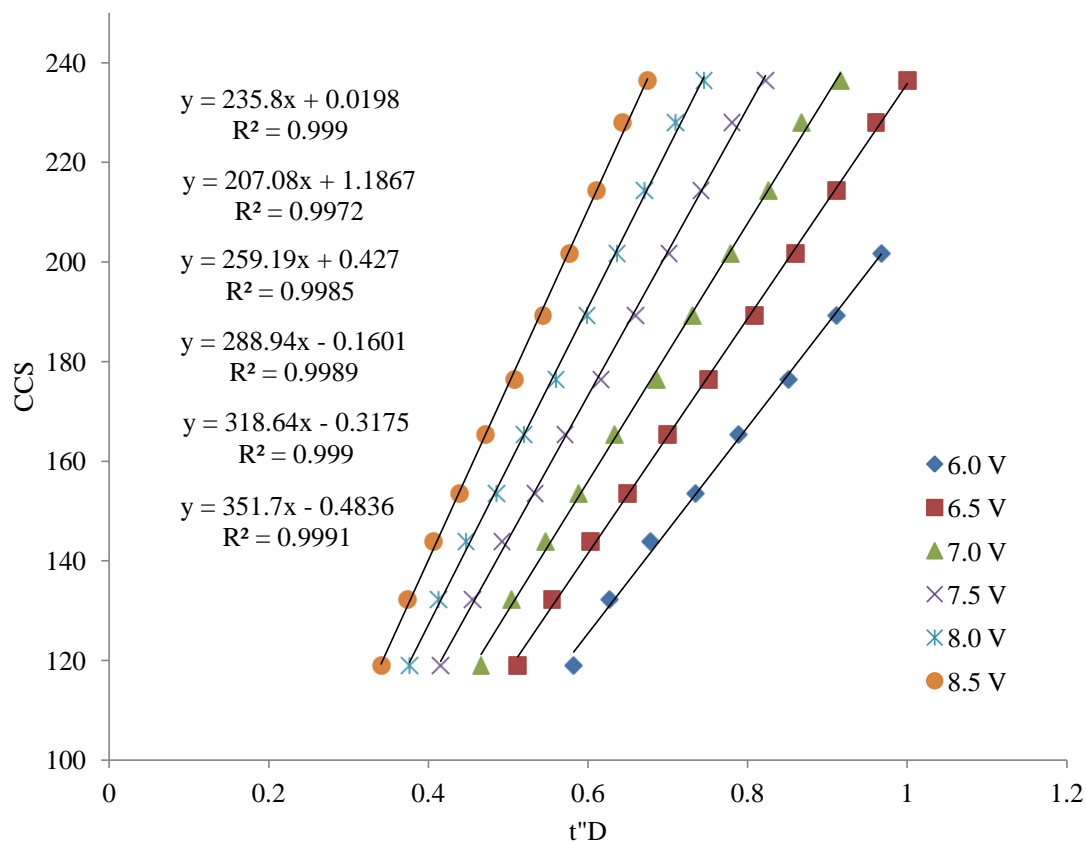


Figure A.4. Calibration plot used for collision-cross-section measurement of the model compounds and petroleomic analytes, described in Chapter 4. The corrected arrival time, t''_D , is calculated by $t''_D = t'_D{}^X * \text{charge} \times (1/\mu)^{1/2}$, where X is the slope from the plot in Figure A.3.

Table A.6. Collision cross sections (CCSs) for model compound ions at six ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights measured. CCS for wave heights that were not measured are marked (-).

Compounds	<i>m/z</i>	ion	Collision Cross Sections (Å ²)							AVG	RSD
			6.0 V	6.5 V	7.0 V	7.5 V	8.0 V	8.5 V			
PThP	563.2	[M+Na] ⁺	142.4	144.5	143.5	143.4	145.4	-	143.8	0.8	
P-2,5-pyridine	550.2	[M+H] ⁺	151.6	152.5	151.1	151.5	153.2	-	152.0	0.6	
P-2,5-pyridine	572.2	[M+Na] ⁺	147.0	147.2	147.1	147.2	148.1	-	147.3	0.3	
P-2,5-pyridine	588.2	[M+K] ⁺	150.6	152.6	150.9	151.0	152.6	-	151.5	0.7	
Chol-Thio	604.5	[M+H] ⁺	205.0	204.4	203.9	203.5	206.6	-	204.7	0.6	
PmPhP	534.3	M ⁺	144.0	144.6	143.6	144.5	145.0	-	144.3	0.4	
PmPhP	535.3	[M+H] ⁺	146.4	146.9	146.1	145.5	147.4	-	146.5	0.5	
PmPhP	557.3	[M+Na] ⁺	146.7	147.2	146.4	146.8	147.6	-	146.9	0.3	
THP	539.5	[M+H] ⁺	-	-	193.2	193.5	-	-	193.3	0.1	
THP	561.5	[M+Na] ⁺	181.5	181.2	181.2	182.1	182.8	-	181.8	0.4	
naphthalene	128.1	M ⁺	60.9	61.0	63.2	-	-	-	60.9	0.1	
phenanthrene	178.1	M ⁺	72.6	70.9	69.0	-	-	-	70.8	2.5	
anthracene	178.1	M ⁺	72.9	70.0	69.0	-	-	-	70.7	2.9	
anthracene	179.1	[M+H] ⁺	73.6	71.7	70.3	-	-	-	71.8	2.3	
chrysene	228.1	M ⁺	84.8	83.4	81.4	-	-	-	83.2	2.1	
triphenylene	228.1	M ⁺	84.1	82.0	80.7	-	-	-	82.2	2.1	
pyrene	202.1	M ⁺	77.5	76.4	74.1	-	-	-	76.0	2.2	
pyrene	203.1	[M+H] ⁺	79.2	77.4	75.5	-	-	-	77.4	2.4	
benz[b]anthracene	228.1	M ⁺	85.3	84.5	82.4	-	-	-	84.1	1.7	
pentacene	278.1	M ⁺	99.5	98.8	97.4	-	-	-	98.6	1.1	
benzo[e]pyrene	252.1	M ⁺	89.3	87.3	85.9	-	-	-	87.5	2.0	
benzo[e]pyrene	253.1	[M+H] ⁺	90.4	88.7	87.5	-	-	-	88.9	1.7	

Table A.6. Collision cross sections (CCSs) for model compound ions at six ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights measured. CCS for wave heights that were not measured are marked (-).
(cont.'d)

Compounds	m/z	ion	Collision Cross Sections (\AA^2)							
			6.0 V	6.5 V	7.0 V	7.5 V	8.0 V	8.5 V	AVG	RSD
benzo[ghi]perylene	276.1	M^+	93.5	92.9	91.4	89.5	-	-	91.8	2.0
benzo[ghi]perylene	277.1	$[\text{M}+\text{H}]^+$	95.6	94.4	93.0	91.3	-	-	93.6	2.0
coronene	300.1	M^+	99.0	98.4	97.1	96.0	-	-	97.6	1.4
coronene	301.1	$[\text{M}+\text{H}]^+$	100.2	99.9	98.7	97.1	-	-	99.0	1.4
9,10-diphenylanthracene	330.2	M^+	119.6	119.3	118.7	117.9	-	-	118.9	0.6
9,10-diphenylanthracene	331.2	$[\text{M}+\text{H}]^+$	117.9	117.9	117.2	117.1	-	-	117.5	0.4
rubrene	532.2	M^+	162.2	162.9	162.9	162.2	-	-	162.6	0.2
violanthrone-79	712.4	M^+	-	191.2	193.6	192.9	196.6	-	192.6	0.6
violanthrone-79	735.4	$[\text{M}+\text{Na}]^+$	-	220.4	220.8	221.0	223.3	-	221.4	0.6
THP dimer	1100.0	$[\text{2M}+\text{Na}]^+$	-	-	-	-	-	300.9	300.9	N/A
P-2,5-pyridine dimer	1099.5	$[\text{2M}+\text{H}]^+$	-	-	239.4	239.4	241.1	-	240.0	0.4
P-2,5-pyridine dimer	1121.5	$[\text{2M}+\text{Na}]^+$	-	-	239.7	239.9	242.1	-	240.6	0.6
chol-thio+P-2,5-pyridine	1175.6	$[\text{2M}+\text{Na}]^+$	-	-	-	-	294.8	292.1	293.5	N/A
chol-thio+P-2,5-pyridine	1153.7	$[\text{2M}+\text{H}]^+$	-	-	-	286.7	285.8	285.0	285.8	0.3
chol-thio dimer	1229.9	$[\text{2M}+\text{Na}]^+$	-	-	-	-	-	314.7	314.7	N/A
violanthrone dimer	1447.8	$[\text{2M}+\text{Na}]^+$	-	-	-	-	-	329.3	329.3	N/A

Table A.7. Collision cross sections (CCSs) for asphaltene ions at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights measured. Arrival time peaks were not observed in the arrival time window (0-18 ms) at smaller wave heights for higher m/z ions, indicated by “-”.

Asphaltene m/z range		Collision Cross Sections (\AA^2)					
		7.0 V	7.5 V	8.0 V	8.5 V	AVG	RSD
400-405	apex	137.1	135.8	139.3	141.3	138.4	1.8
	boundary 1	118.0	117.8	123.0	123.5	120.6	2.5
	boundary 2	154.5	152.1	154.4	157.6	154.6	1.5
500-505	apex	156.0	157.2	159.5	161.8	158.6	1.6
	boundary 1	130.1	132.3	135.5	141.0	134.7	3.5
	boundary 2	179.1	179.5	181.1	180.8	180.1	0.5
600-605	apex	172.3	173.9	174.2	177.5	174.5	1.2
	boundary 1	137.8	143.7	145.1	147.0	143.4	2.8
	boundary 2	202.5	201.1	200.0	204.5	202.0	1.0
700-705	apex	192.0	192.7	193.9	196.5	193.8	1.0
	boundary 1	153.6	157.1	161.4	164.3	159.1	3.0
	boundary 2	225.5	224.1	222.9	225.3	224.4	0.5
800-805	apex	208.2	208.5	209.8	213.6	210.0	1.2
	boundary 1	166.5	167.6	169.9	178.0	170.5	3.0
	boundary 2	244.6	244.1	244.7	245.3	244.7	0.2
900-905	apex	224.5	225.1	226.4	230.2	226.5	1.1
	boundary 1	174.6	183.5	185.0	190.7	183.4	3.6
	boundary 2	267.3	261.6	262.8	265.2	264.3	1.0
1000-1005	apex	-	239.2	240.2	244.4	241.3	1.2
	boundary 1	-	193.5	195.5	201.0	196.7	2.0
	boundary 2	-	279.1	279.4	282.8	280.4	0.7

Table A.7. Collision cross sections (CCSs) for asphaltene ions at four ion mobility wave heights. Averages and relative standard deviations are reported over all wave heights measured. Arrival time peaks were not observed in the arrival time window (0-18 ms) at smaller wave heights for higher m/z ions, indicated by “-”. (cont.’d)

Asphaltene m/z range		Collision Cross Sections (\AA^2)					
		7.0 V	7.5 V	8.0 V	8.5 V	AVG	RSD
1100-1105	apex	-	256.8	256.3	258.8	257.3	0.5
	boundary 1	-	-	212.2	212.6	212.4	0.1
	boundary 2	-	-	295.4	299.6	297.5	1.0
1200-1205	apex	-	-	271.3	278.1	274.7	1.8
	boundary 1	-	-	219.7	220.6	220.1	0.3
	boundary 2	-	-	316.5	328.0	322.2	2.5
1300-1305	apex	-	-	289.8	295.2	292.5	1.3
	boundary 1	-	-	-	242.2	242.2	N/A
	boundary 2	-	-	-	341.9	341.9	N/A

APPENDIX B

Additional DIESI-IM-HRMS Data

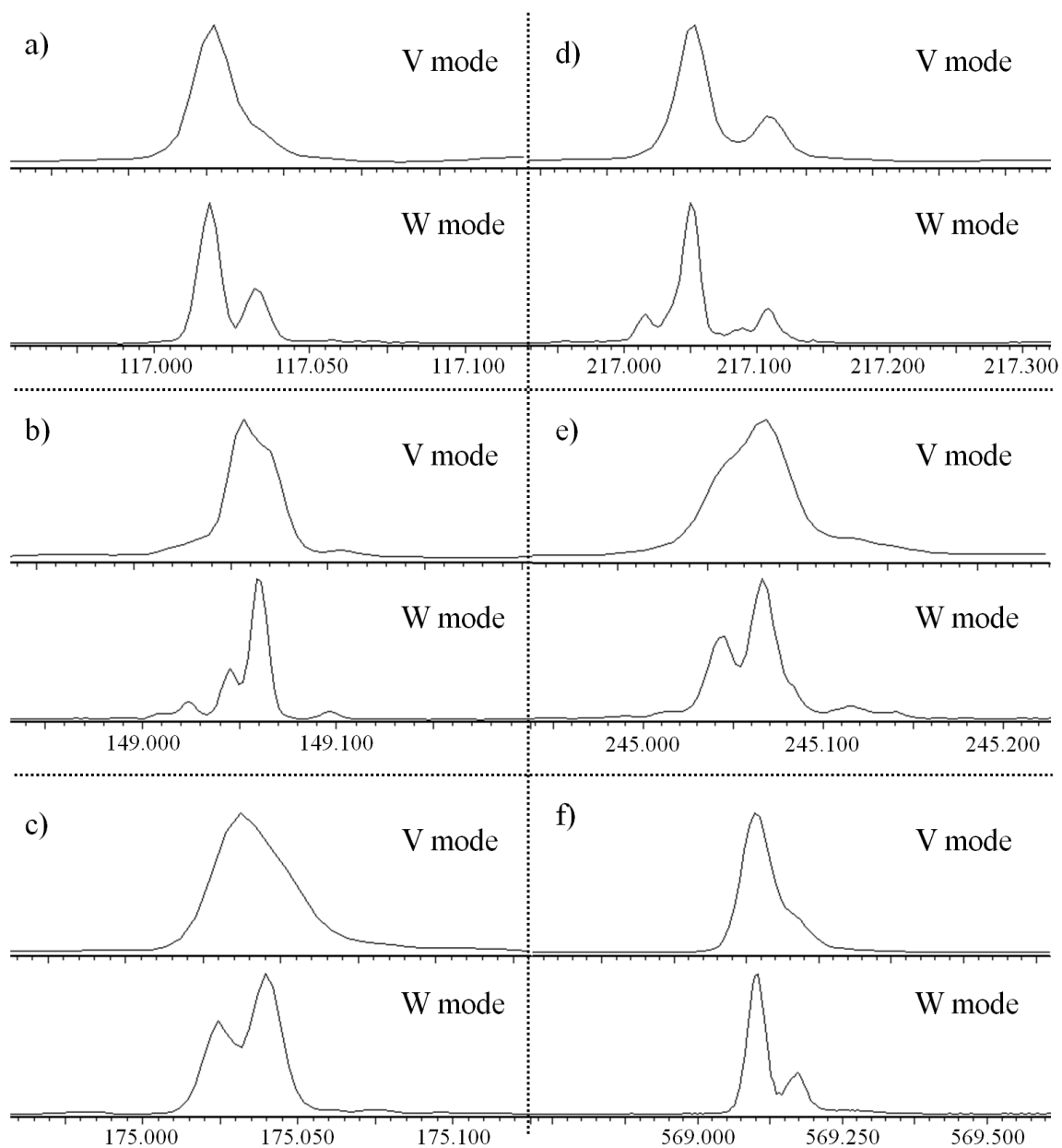


Figure B.1. Examples of the resolving power differences between V and W mode for several ions in a corn stover hydrolysate.

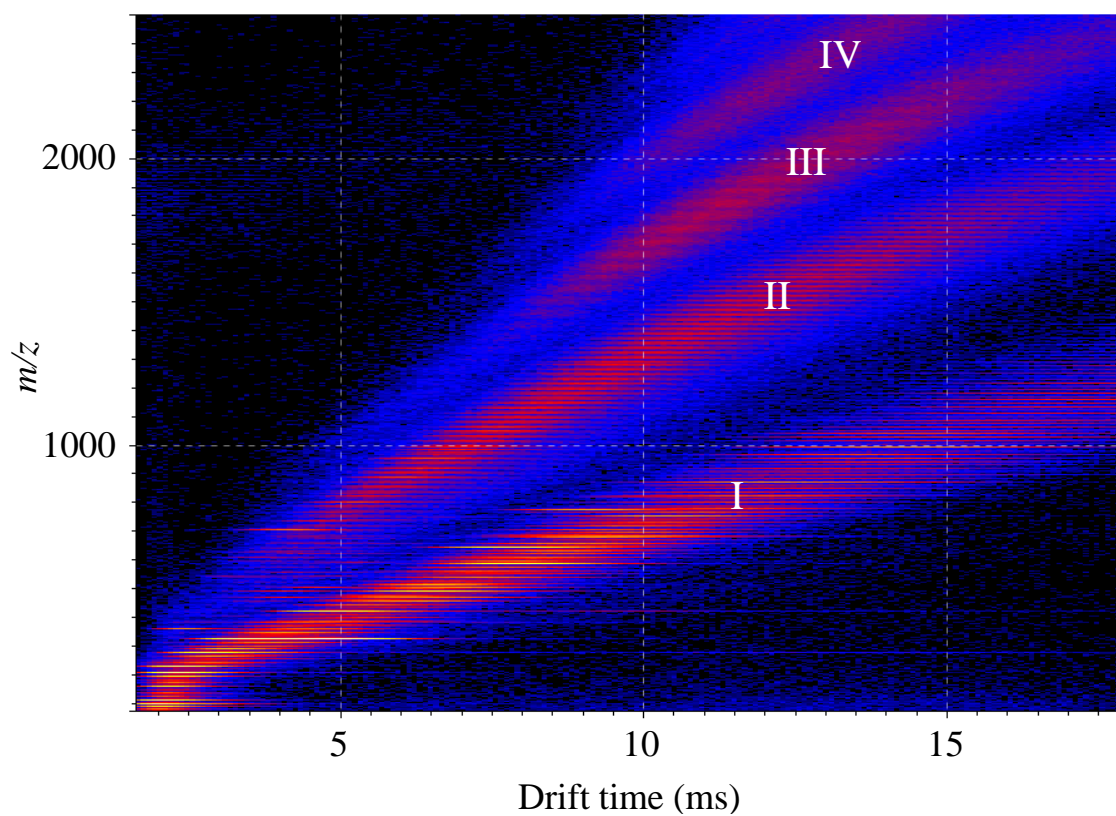


Figure B.2. DIESI-IM-MS plot of m/z vs. arrival time for negative ions resulting from a 1:10 dilution of dilute-acid-pretreated corn stover hydrolysate with four distinct trendlines (I, II, III, and IV). Trendline one contains singly charged biomass analytes. Trendlines II-IV consist of multiply charged ions.

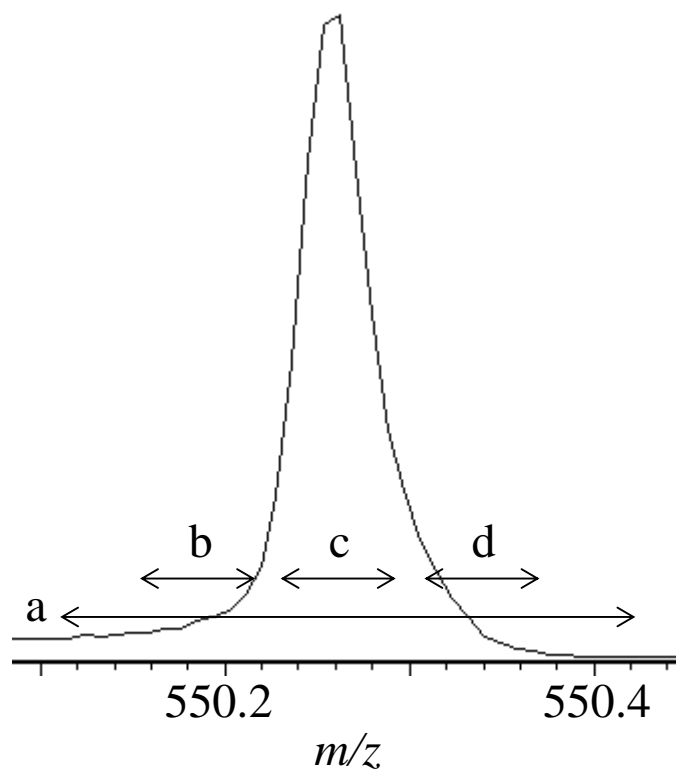


Figure B.3. M/z peak for the $[M+H]^+$ ion of P-2,5-PMP. The TOF detector was operating in dead time during the acquisition of most of this peak. Integration of the areas indicated (a-d) resulted in the arrival time distributions shown in Figure B.4. Artifact arrival time peaks (Figure B.4 a,c-d) are observed for intense ions such as this one.

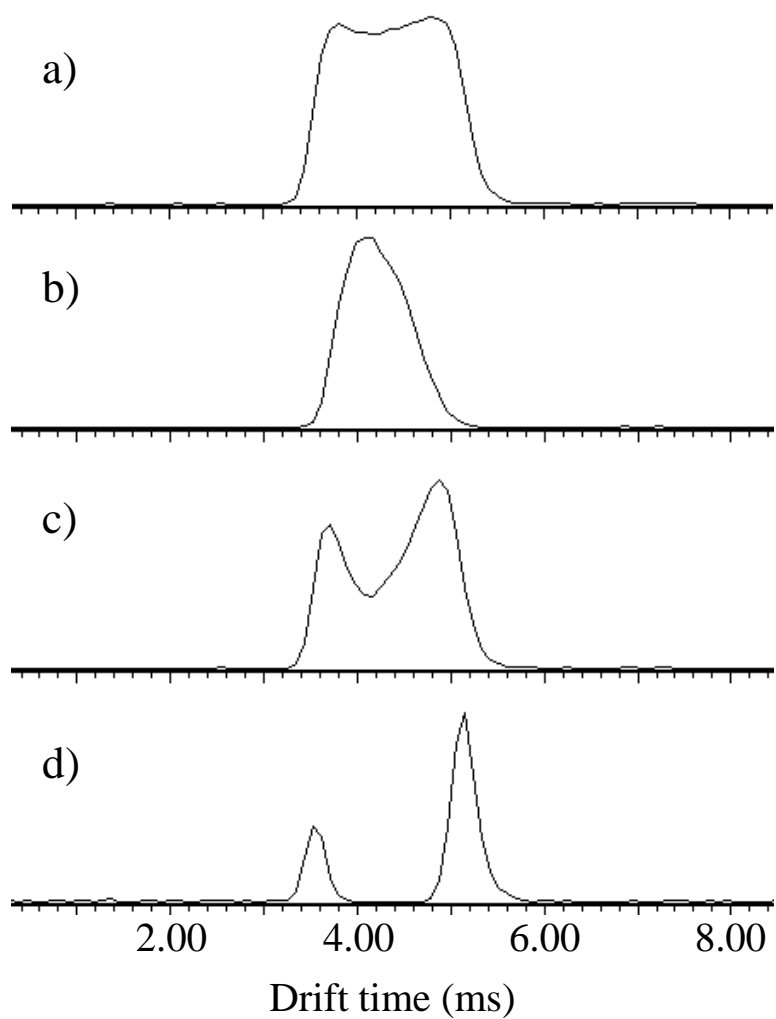


Figure B.4. Arrival time distributions (ATDs) for the peak segments denoted in Figure B.3. The ATD across the whole m/z peak (a) shows two unresolved arrival time peaks. The ATDs across the middle (c) and right side (d) of the peaks show two artifact peaks. Only the left side of the m/z peak (b), where ion signal is less intense and the detector is not yet operating in dead time, corresponds to the real arrival time distribution.

APPENDIX C

Optimization and Development of the UPLC Method Presented in Chapter Three

Table C.1. UPLC gradient elution profile for time-optimized separation of 35 degradation products with a Waters BEH C18 column (5.0 cm x 2.1 cm; particle size = 1.7 μ m) at a flow rate of 0.350 mL/min, injection volume of 5.0 μ L, and column temperature of 35 $^{\circ}$ C.

Time (min)	Mobile Phase Composition	
	0.01% formic acid in water (%A)	0.01% formic acid in methanol (%B)
0.00	99	1.0
3.00	99	1.0
4.00	80	20
6.00	50	50
6.10	2.0	98
8.00	2.0	98
8.10	99	1.0
10.00	99	1.0

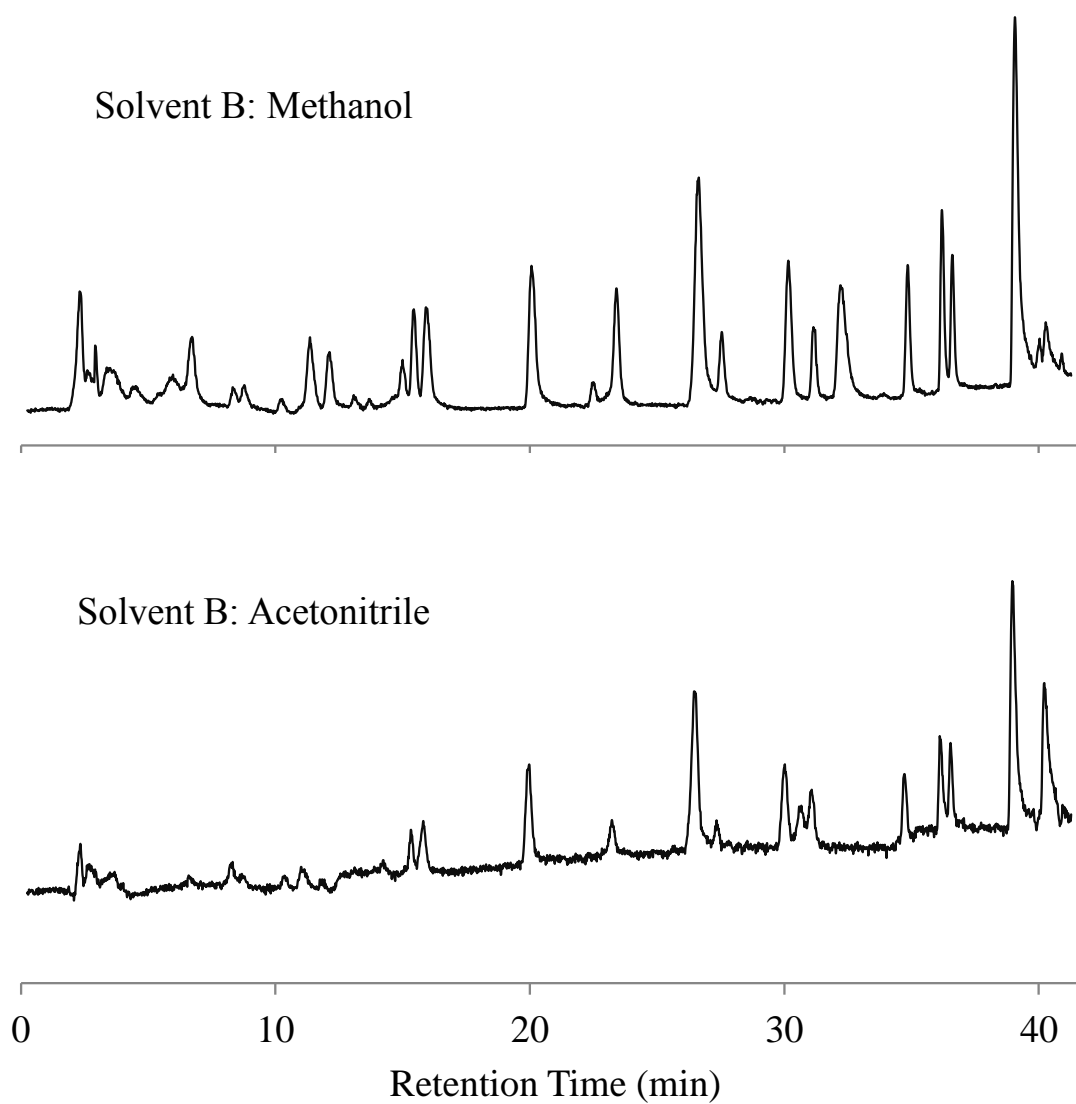


Figure C.1. UPLC chromatograms of a mixture of 35 standard degradation products obtained with different “B” solvents: methanol (top) and acetonitrile (bottom), each containing 0.01% formic acid. 0.01% aqueous formic acid was solvent A in both cases. Note the higher background signal for acetonitrile.

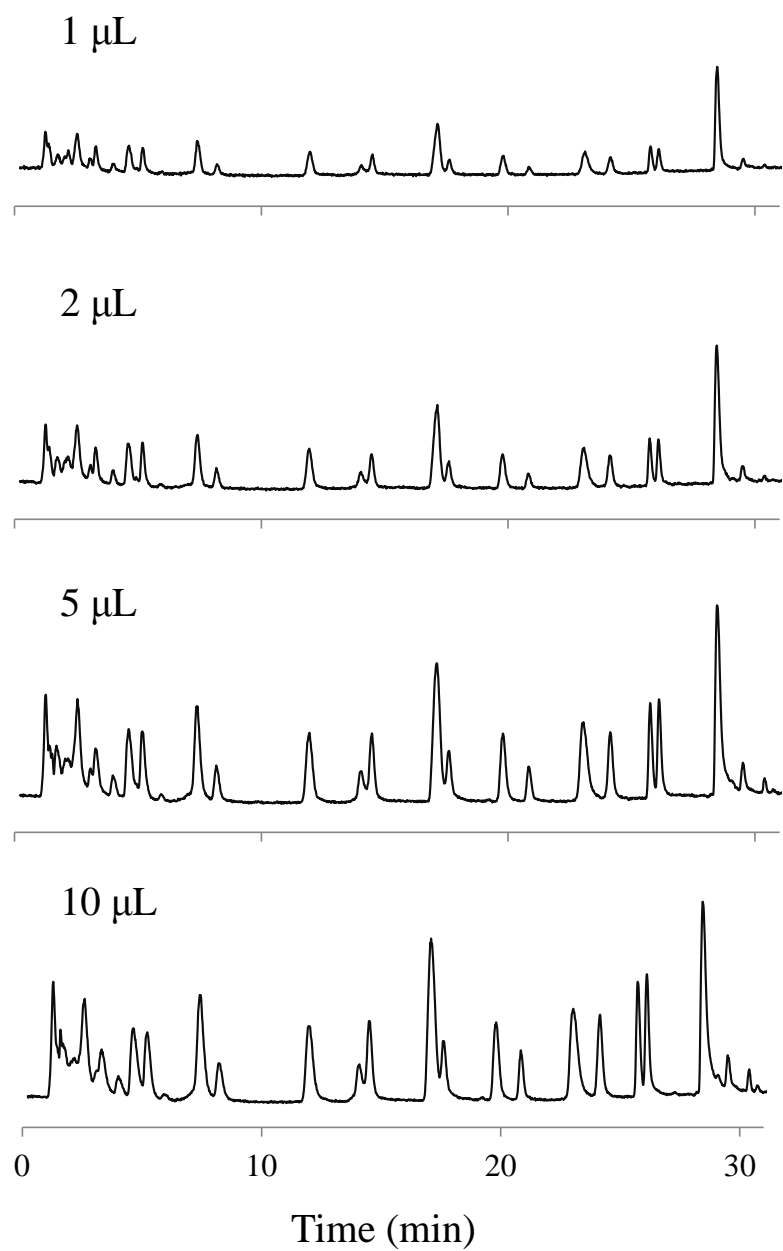


Figure C.2. UPLC chromatograms of 1-, 2-, 5-, and 10- µL injections of a 5 mg/L mixture of 35 standard degradation products. The 2- and 5-µL injections were determined to be more optimal than the other injection volumes.

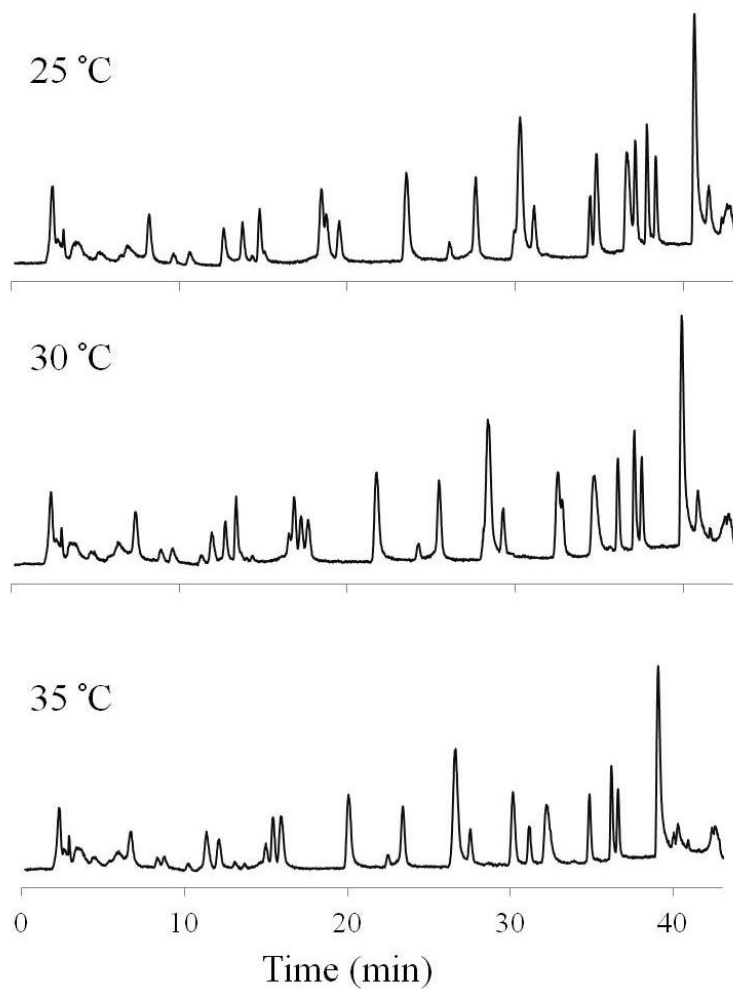


Figure C.3. UPLC chromatograms of a 5 mg/L mixture of 35 standard degradation products obtained at column temperatures of 25, 30, and 35 °C. A column temperature of 35 °C was determined to provide more separation than 25 and 30 °C.

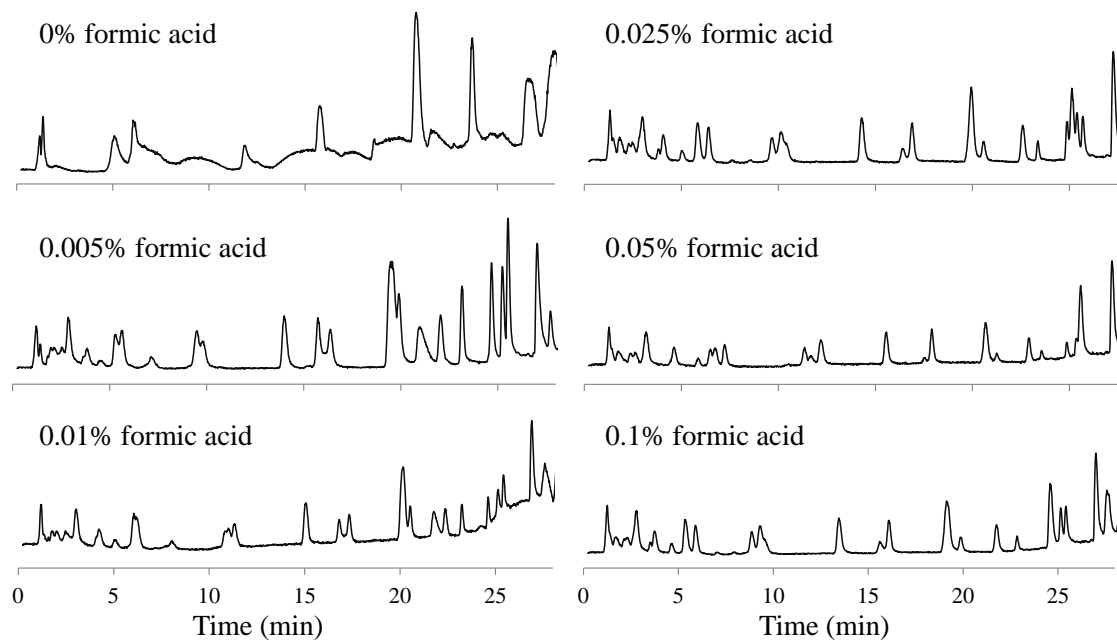


Figure C.4. UPLC chromatograms of a 5 mg/L mixture of 35 standard degradation products with varying percentages of formic acid in the mobile phase. The 0.005% and 0.01% concentrations gave the best separation of front-end isomers (e.g., fumaric and maleic acids).

APPENDIX D

Development and Optimization of MetaboLynx and MS^e Viewer Software

Complete List of MetaboLynx Method Parameters

False Positives Page

Mass matching window (Da): 0.05
Retention Time match by time window: 1.50 Da
Check for real positive with analyte to control
peak area ratio exceeding 5.0 : No
List of masses to be considered as false positives
(Da, mins): none
Add Control Sample masses (with times) to list
of masses: True
Exclude TOF Lock Mass: False, Monoisotopic
Lock Mass 0.0000
Exclude Isotopic Entries: True, Mass Window:
5.00Da
Report False Positives in the Browser: Yes

Elemental Composition Settings

Perform Elemental Composition Analysis on
Unexpected Metabolites: True
Elemental Composition Settings File:
C:\MassLynx\biomass.els
Element Counts Relative to Compound: None
MS Results Settings:
Reject Elemental Composition results with DBE
outside precursor range: False
Precursor DBE minus 0.0, plus 0.0
MS/MS Results Settings:
Reject Elemental Composition results with
element count greater than parent: False
Reject Elemental Composition results with DBE
outside parent range: False
Parent DBE minus 0.0, plus 0.0

MS Data Page

Start Time (min): 0.00
End Time (min): 42.00
Min Peak Separation: 0.00
Expected Mass Chromatograms
Create Chromatogram: True
Peak Detection
Enable smoothing: True
Window size (+/- scans): 3
Number of smooths: 2
Smoothing method: Savitzky Golay
ApexTrack Peak Integration: True

Peak to Peak Baseline Noise: 0.00 Automatic:
True
Peak Width at 5% Height (Mins): 0.00
Automatic: True
Baseline Start Threshold%: 0.00
Baseline End Threshold%: 0.00
Detect Shoulders: False
Relative height: False
Relative height: 1.50
Absolute height: False
Absolute height: 0.00
Relative area: False
Relative area: 2.00
Absolute area: False
Absolute area: 0.00
Use Mass Window 0.050: No
Mass Range Chromatograms for Unexpected
Metabolites
Create Chromatogram: True
Peak Detection
Enable smoothing: True
Window size (+/- scans): 3
Number of smooths: 2
Smoothing method: Savitzky Golay
ApexTrack Peak Integration: True
Peak to Peak Baseline Noise: 0.00 Automatic:
True
Peak Width at 5% Height (Mins): 0.00
Automatic: True
Baseline Start Threshold%: 0.00
Baseline End Threshold%: 0.00
Detect Shoulders: False
Relative height: False
Relative height: 1.50
Absolute height: False
Absolute height: 0.00
Relative area: False
Relative area: 2.00
Absolute area: False
Absolute area: 0.00
Full mass range: Yes
Mass Range: Low Mass (Da): 0.00 High Mass
(Da): 0.00 Window(amu): 1.00
Report all traces generated

PDA Data Page

Start Time (min): 0.00
End Time (min): 0.00
Offset Time (min): 0.00
TAC Chromatogram not selected
BPI Chromatogram not selected
Wavelengths
Chromatogram not selected
Fixed Wavelength Range
Chromatogram not selected

Analog Data Page

None selected

Instrument Adducts Page

Positive ion series
Adduct: 1.0078
Charges: 1
Multimers: 1
Desc.: M+H
Negative ion series
Adduct: -1.0078
Charges: 1
Multimers: 1
Desc.: M-H

Spectrum Page

Combine Spectrum:
Minimum peak separation (Da): 0.050
Background subtract: NO
Use peak top spectra
Check Metabolite Counts: NO
Metabolite Identification Criteria:
Mass Find Mode: Nearest
Mass Window: 0.050
Intensity Threshold Mode: Absolute
Threshold: 0.00

Advanced Settings Page

Spectrum Combine over Peak Top:
Number of Spectra to Combine: 3
Amount of spectra inside peak (start,end): 80%
Spectrum Masses Reported:
Filter Mode: Absolute
Value: 0
Fractional mass filter: YES, above 0.010000
below 0.200000

Pre-Process Data Page

Mass Measure: True
Data Type: TOF
Settings (MS+)
Background subtract: True
Polynomial order: 1
Below curve (%): 40.00

Smooth: True
Smooth window (channels) +/-: 0.75
Number of smooths: 2
Smoothing method: Mean
Min peak width at half height (channels): 4
Peak detect method: Top
Centroid top (%): 80.00
Mass measure stick type: Heights
Use TOF mass correction: False
Resolution: 5000.00
Np Multiplier: 1.000
Mass Window +/-: 1.000
Lock Mass: 0.0000
Settings (MS-)
Background subtract: True
Polynomial order: 1
Below curve (%): 40.00
Smooth: True
Smooth window (channels) +/-: 3.00
Number of smooths: 2
Smoothing method: Savitzky Golay
Min peak width at half height (channels): 4
Peak detect method: Centroid top (%)
Centroid top (%): 80.00
Mass measure stick type: Areas
Use TOF mass correction: True
Resolution: 9000.00
Np Multiplier: 0.700
Mass Window +/-: 0.100
Lock Mass: 112.9851
Isotopic Cluster Analysis: False
Use Mass Differences Only
RT Start Time (mins): 0.00
RT End Time (mins): 0.00
Mass Tolerance (amu): 0.05
Ratio Tolerance (%): 30.00
Intensity Threshold (%): 0.50
Use Second Mass Difference: False
First Mass Difference (amu): 0.0000
First Intensity Ratio: 0.00
Second Mass Difference (amu): 0.0000
Second Intensity Ratio: 0.00

MS/MS Settings Page

Use Metabolite for MS/MS if Analog or PDA
Trigger present: False
Automatically start MS/MS acquisition: False
MS/MS Settings File:
Compare MS/MS Spectra: False
Mass Window (amu): 0.50
Intensity Threshold (%): 5.00
Peak Detection
Enable smoothing: True
Window size (+/- scans): 1
Number of smooths: 1
Smoothing method: Savitzky Golay

ApexTrack Peak Integration: True
Peak to Peak Baseline Noise: 0.00 Automatic:
True
Peak Width at 5% Height (Mins): 0.00
Automatic: True
Baseline Start Threshold%: 0.00
Baseline End Threshold%: 0.00
Detect Shoulders: False
Relative height: False
Relative height: 1.50
Absolute height: False
Absolute height: 0.00

Relative area: False
Relative area: 2.00
Absolute area: False
Absolute area: 0.00
Perform Elemental Composition on MS/MS
Fragments: False
Elemental Composition Intensity Threshold (%):
5.0

Table D.1. Compound names, mass differences (from the C₃₃H₃₈ON₃S₃ “parent” mass), and expected retention times entered into the MetaboLynx method for the “expected metabolites”.

Analyte	Mass difference (from “parent” mass)	Retention Time (min)
malonic acid	-740.126	0.31
lactic acid	-754.105	0.69
succinic acid	-726.110	0.76
fumaric acid	-728.126	1.50
methylmalonic acid	-726.110	1.82
maleic acid	-728.126	1.85
gallic acid	-674.115	2.54
trans/cis-aconitic acid	-670.120	2.56
levulinic acid	-728.089	3.44
glutaric acid	-712.094	3.63
itaconic acid	-714.110	4.38
2-hydroxy-2-methylbutyric acid	-726.073	5.15
3,4-dihydroxybenzoic acid	-690.110	5.18
2-furoic acid	-732.120	5.60
3,4-dihydroxybenzaldehyde	-706.105	9.74
2,5-hydroxybenzoic acid	-690.110	9.76
4-hydroxybenzoic acid	-706.105	9.88
adipic acid	-698.079	10.05
4-hydroxybenzaldehyde	-722.100	13.56
vanillic acid	-676.094	15.93
caffeic acid	-664.094	16.90
vanillin	-692.089	20.00
4-hydroxyacetophenone	-708.084	20.06
syringic acid	-646.084	20.98
homovanillic acid	-662.079	21.12
salicylic acid	-706.105	21.12
4-hydroxycoumaric acid	-680.089	22.97
syringaldehyde	-663.079	24.54
benzoic acid	-722.100	25.32
sinapic acid	-620.068	27.61
ferulic acid	-650.079	27.62
3-hydroxy-4-methoxycinnamic acid	-650.079	29.80
4-hydroxycoumarin	-682.105	32.07
<i>o</i> -toluic acid	-708.084	32.71
<i>p</i> -toluic acid	-708.084	33.95

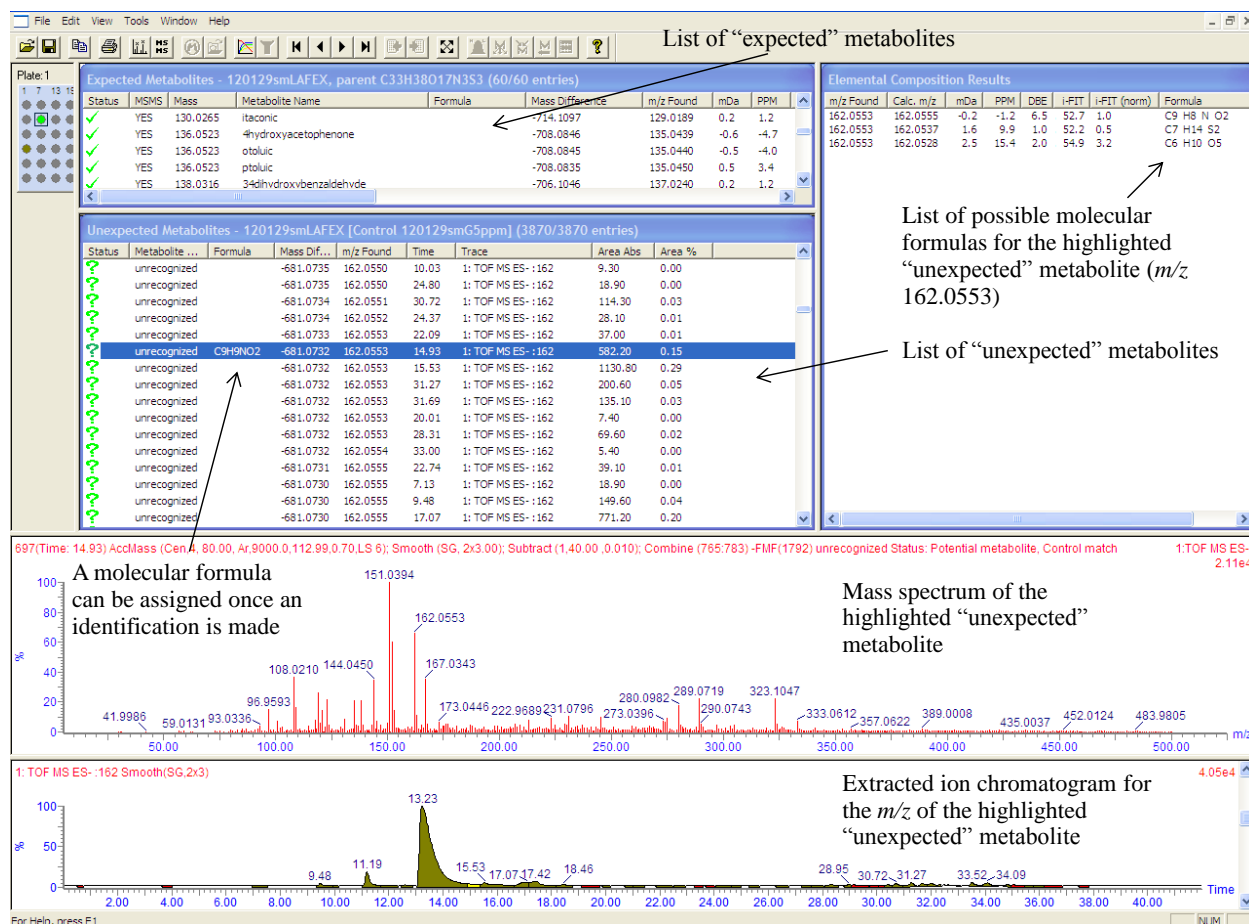


Figure D.1. Snapshot of the MetaboLynx software interface. Tables of data for the “expected” metabolites and unexpected metabolites can be found on the left side (top and middle, respectively). Elemental composition results for a selected unexpected metabolite are shown in the top right. Mass spectra and the extracted ion chromatogram for the selected m/z are seen in the bottom two windows.

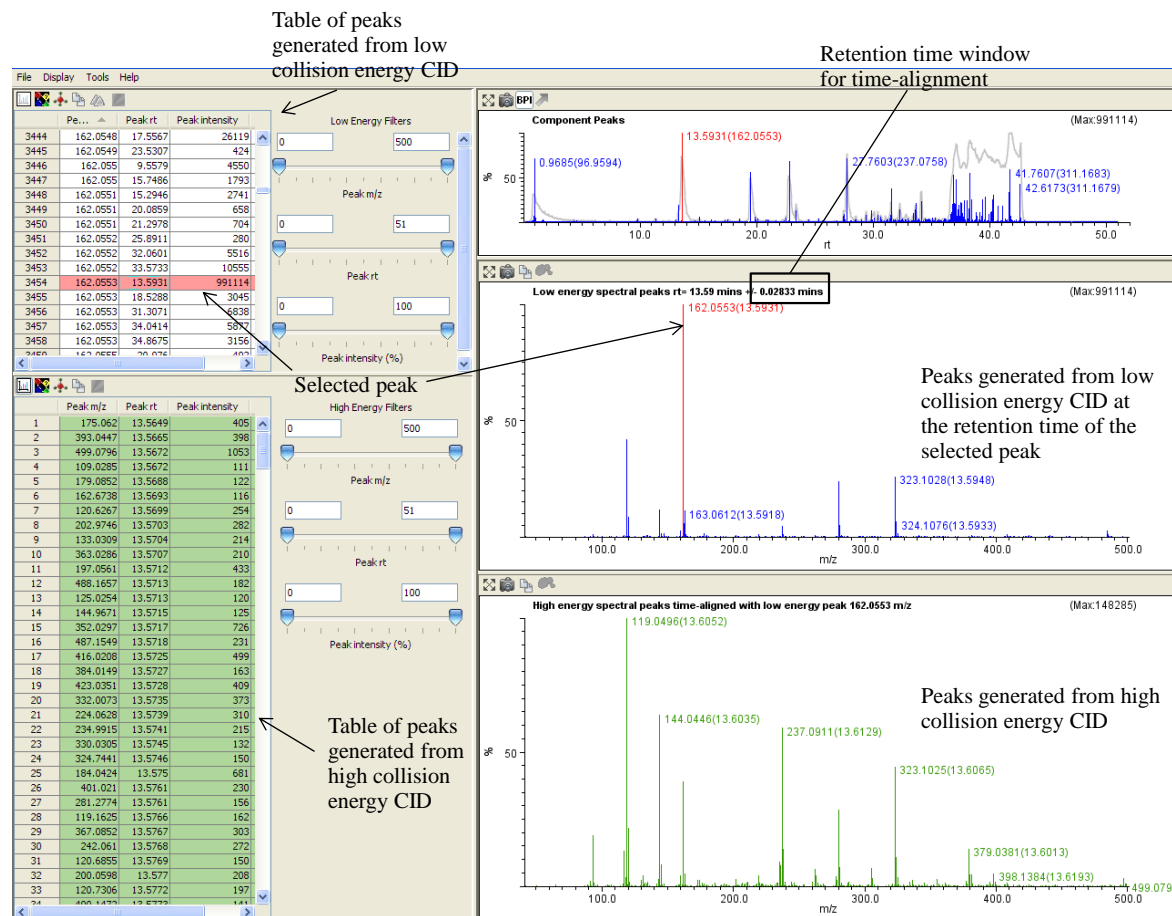


Figure D.2. Snapshot of the MS^e software interface. On the left side of the display, peaks found via low collision energy (top table) and high collision energy (bottom table) segments of the MS^e analysis are listed. In this example, m/z 162.0553 at retention time 13.5391 min is selected and the corresponding low energy and high energy spectra time aligned for a 0.02833-min window are shown on the right in the middle and bottom windows, respectively. The selected peak is also shown at its retention time in the chromatogram (top).

MetaboLynx Method Validation Experiments

Data from the 5 mg/L control sample (Figure 3.2) was first processed with the MetaboLynx method. Thirty-five standard analytes were successfully identified by MetaboLynx as “expected metabolites” in this sample. When this same sample was reprocessed as an “analyte” sample, no peaks were identified as “unexpected metabolites”, indicating that the software is excluding any non-targeted peaks found in the control sample from the analyte sample. (Referring back to Figure 3.3, this validation step confirmed that if your control and analyte samples were both composed of the same ions (e.g., X and Y), the software successfully identified Y in the control sample as a mass it should exclude from the analyte sample.) This enables variable background ion signal to be excluded from results.

REFERENCES

- (1) Keusch, G. T. *Am. J. Clin. Nutr.* **2006**, 83, 520S-522S.
- (2) Gubb, E.; Matthiesen, R. In *Introduction to Omics*; Matthiesen, R., Ed.; Methods in Molecular Biology; Humana Press: Clifton, NJ, 2010; Vol. 593, pp 1-23.
- (3) Lederberg, J.; Mccray, A. *The Scientist* **2001**, 17.
- (4) Konig, S. In *"Omics" Applications; Mass Spectrometry*; John Wiley & Sons, Inc: 2008; pp 243-252.
- (5) Current Topics in Biotechnology: X-omes & X-omics.
<http://www.argosbiotech.de/700/omics/Xomics.htm> (accessed January 2012).
- (6) Schneider, M. V.; Orchard, S. In *Omics Technologies, Data and Bioinformatics Principles; Bioinformatics for Omics Data*; 2011; Vol. 719, pp 3-30.
- (7) Prohaska, S. J.; Stadler, P. F. In *The Use and Abuse of -Omics; Bioinformatics for Omics Data* 2011; Vol. 719, pp 173-196.
- (8) Nunnally, B. K., Ed.; In *Analytical Techniques in DNA Sequencing*; Taylor & Francis Group: Boca Raton, FL, 2005; pp 237.
- (9) Gibson, G.; Muse, S. V., Eds.; In *A Primer of Genome Science*; Sinauer Associates, Inc.: Sunderland, MA, 2002; pp 345.
- (10) Whitford, D., Ed.; In *Proteomics: Structure and Function*; John Wiley & Sons: Hoboken, NJ, 2005; pp 528.
- (11) Palzkill, T., Ed.; In *Proteomics*; Kluwer Academic Publishers: Norwell, MA, 2002; pp 127.
- (12) Marchetti-Deschmann, M.; Allmaier, G. *J. Proteomics* **2011**, 74, 915-919.
- (13) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, 17, 676.
- (14) Rodgers, R. P.; McKenna, A. M. *Anal. Chem.* **2011**, 83, 4665-4687.
- (15) Mullins, O. C.; Sheu, E. Y.; Hammami, A.; Marshall, A. G. In *Asphaltenes, Heavy Oils, and Petroleomics*; Springer: New York, 2007; pp 690.

- (16) Marshall, A. G.; Rodgers, R. P. *Proc. Natl. Acad. Sci.* **2008**, *6*, 1-6.
- (17) Marshall, A. G.; Rodgers, R. P. *Acc. Chem. Res.* **2004**, *37*, 53-59.
- (18) Chen, S.; Mowery, R. A.; Scarlata, C. J.; Chambliss, C. K. *J. Agric. Food Chem.* **2007**, *55*, 5912-5918.
- (19) Sevcik, R. S.; Mowery, R. A.; Becker, C.; Chambliss, C. K. *J. Chromatogr. A* **2011**, *1218*, 1236-1243.
- (20) Persson, P.; Andersson, J.; Gorton, L.; Larsson, S.; Nilvebrant, N.; Jonsson, L. J. *J. Agric. Food Chem.* **2002**, *50*, 5318-5325.
- (21) Luo, C.; Brink, D. L.; Blanch, H. W. *Biomass Bioenergy* **2002**, *22*, 125-138.
- (22) Humpula, J. F.; Chundawat, S. P. S.; Vismeh, R.; Jones, A. D.; Balan, V.; Dale, B. E. *J. Chromatogr. B* **2011**, *879*, 1018-1022.
- (23) Fenske, J. J.; Griffin, D. A.; Penner, M. H. *J. Ind. Microbiol. Biotechnol.* **1998**, *20*, 364-368.
- (24) Chen, S.; Mowery, R. A.; Castleberry, V. A.; Van Walsum, G. P.; Chambliss, C. K. *J. Chromatogr., A* **2006**, *1104*, 54-61.
- (25) Chen, S.; Mowery, R. A.; Chambliss, C. K.; van Walsum, G. P. *Biotechnol. Bioeng.* **2007**, *98*, 1135-1145.
- (26) Sharma, L. N.; Becker, C.; Chambliss, C. K. In *Analytical Characterization of Fermentation Inhibitors in Biomass Pretreatment Samples Using Liquid Chromatography, UV-Visible Spectroscopy, and Tandem Mass Spectrometry*; Mielenz, J. R., Ed.; Biofuels: Methods and Protocols; Humana Press: Totowa, NJ, 2009; pp 125.
- (27) Chundawat, S. P. S.; Vismeh, R.; Sharma, L. N.; Humpula, J. F.; da Costa Sousa, L.; Chambliss, C. K.; Jones, A. D.; Balan, V.; Dale, B. E. *Bioresource Technol.* **2010**, *101*, 8429-8438.
- (28) Du, B.; Sharma, L. N.; Becker, C.; Chen, S.; Mowery, R. A.; van Walsum, G. P.; Chambliss, C. K. *Biotechnol. Bioeng.* **2010**, *107*, 430-440.
- (29) Zheng, R.; Zhang, H.; Zhao, J.; Lei, M.; Huang, H. *J. Chromatogr., A* **2011**, *1218*, 5319-5327.
- (30) Chen, Y.; Mori, M.; Pastusek, A. C.; Schug, K. A.; Dasgupta, P. K. *Anal. Chem.* **2011**, *83*, 1015-1021.

- (31) Sotgia, S.; Zinellu, A.; Pisanu, E.; Murgia, L.; Pinna, G. A.; Gaspa, L.; Deiana, L.; Carru, C. *Anal. Bioanal. Chem.* **2010**, 396, 2937-2941.
- (32) Bordeerat, N. K.; Georgieva, N. I.; Klapper, D. G.; Collins, L. B.; Cross, T. J.; Borchers, C. H.; Swenberg, J. A.; Boysen, G. *Proteomics* **2009**, 9, 3939-3944.
- (33) Hu, P.; Fang, L.; Jones, C. M.; Chess, E. K. *Carbohydr. Res.* **2011**, 346, 2268-2273.
- (34) Rainville, P. D.; Stumpf, C. L.; Shockcor, J. P.; Plumb, R. S.; Nicholson, J. K. *J. Proteome Res.* **2007**, 6, 552-558.
- (35) Liu, Z.; Cao, J.; He, Y.; Qiao, L.; Xu, C.; Lu, H.; Yang, P. *J. Proteome Res.* **2010**, 9, 227-236.
- (36) Griffin, J. L.; Nicholls, A. W. *Pharmacogenomics* **2006**, 7, 1095-1107.
- (37) Heuberger, A. L.; Lewis, M. R.; Chen, M.; Brick, M. A.; Leach, J. E.; Ryan, E. P. *PLoS ONE* **2010**, 5, e12915.
- (38) Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Anal. Chem.* **2006**, 78, 3289-3295.
- (39) Nagy, K.; Redeuil, K.; Williamson, G.; Rezzi, S.; Dionisi, F.; Longet, K.; Destailats, F.; Renouf, M. *J. Chromatogr., A* **2011**, 1218, 491-497.
- (40) Grata, E.; Boccard, J.; Guillarme, D.; Glauser, G.; Carrupt, P.; Farmer, E. E.; Wolfender, J.; Rudaz, S. *J. Chromatogr., B* **2008**, 871, 261-270.
- (41) Ross, K. L.; Tu, T. T.; Smith, S.; Dalluge, J. J. *Anal. Chem.* **2007**, 79, 4840-4844.
- (42) Ross, K. L.; Dalluge, J. J. *Anal. Chem.* **2009**, 81, 4021-4026.
- (43) Spáčil, Z.; Nováková, L.; Solich, P. *Talanta* **2008**, 76, 189-199.
- (44) Ehling, S.; Cole, S. *J. Agric. Food Chem.* **2011**, 59, 2229-2234.
- (45) Gruz, J.; Novák, O.; Strnad, M. *Food Chem.* **2008**, 111, 789-794.
- (46) Xiang, G.; Yang, L.; Zhang, X.; Yang, H.; Ren, H.; Miao, M. *Chromatographia* **2009**, 70, 1007-1010.
- (47) Lin, L.; Sun, J.; Chen, P.; Harnly, J. J. *J. Agric. Food Chem.* **2011**, 59, 12059-12072.
- (48) Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J.; Noreen, S.; Hill, C. A.; Thomas, C. L. P. *Analyst* **2004**, 129, 984-994.

- (49) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401-2414.
- (50) Liu, X.; Plasencia, M.; Ragg, S.; Valentine, S. J.; Clemmer, D. E. *Brief Funct. Genomic Proteomic* **2004**, *3*, 177-186.
- (51) McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. *Int. J. Mass Spectrom.* **2005**, *240*, 301-315.
- (52) Dwivedi, P.; Schultz, A. J.; Hill Jr., H. H. *Int. J. Mass Spectrom.* **2010**, *298*, 78-90.
- (53) Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. *Anal. Chem.* **2009**, *81*, 9941-9947.
- (54) Becker, C.; Fernandez-Lima, F. A.; Russell, D. H. *Spectroscopy* **2009**, *24*, 38-42.
- (55) Becker, C.; Qian, K.; Russell, D. H. *Anal. Chem.* **2008**, *80*, 8592-8597.
- (56) Ahmed, A.; Cho, Y. J.; No, M.; Koh, J.; Tomczyk, N.; Giles, K.; Yoo, J. S.; Kim, S. *Anal. Chem.* **2011**, *83*, 77-83.
- (57) Zhu, M.; Bendiak, B.; Clowers, B.; Hill Jr., H. H. *Anal. Bioanal. Chem.* **2009**, *394*, 1853-1867.
- (58) Clowers, B. H.; Dwivedi, P.; Steiner, W. E.; Hill Jr., H. H.; Bendiak, B. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 660-669.
- (59) Dwivedi, P.; Bendiak, B.; Clowers, B. H.; Hill Jr., H. H. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1163-1175.
- (60) Fenn, L. S.; McLean, J. A. *J. Phys. Chem. Chem. Phys.* **2011**, *13*, 2196-2205.
- (61) Vakhrushev, S. Y.; Langridge, J.; Campuzano, I.; Hughes, C.; Peter-Katalinic, J. *Anal. Chem.* **2008**, *80*, 2506-2513.
- (62) Ferrer, I.; Thurman, E. M., Eds.; In *Liquid Chromatography Time-of-Flight Mass Spectrometry: Principles, Tools, and Applications for Accurate Mass Analysis*. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2009; pp 261.
- (63) Sun, J.; Baker, A.; Chen, P. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2591-2602.
- (64) Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; H., H. H., Jr *J. Mass Spectrom.* **2008**, *43*, 1-22.
- (65) Waldron, K. W., Ed. In *Bioalcohol production: Biochemical conversion of lignocellulosic biomass*; CRC Press: Boca Raton, 2010; pp 496.

- (66) Kelley, S. S.; Rials, T. G.; Snell, R.; Groom, L. H.; Sluiter, A. *Wood Sci. Technol.* **2004**, *38*, 257-276.
- (67) Nkansah, K.; Dawson-Andoh, B. *J. Renewable Sustainable Energy* **2010**, *2*, No. 023103.
- (68) Kobayashi, T.; Kohn, B.; Holmes, L.; Faulkner, R.; Davis, M.; Maciel, G. E. *Energy Fuels* **2011**, *25*, 1790-1797.
- (69) Helm, R. F.; Jervis, J.; Ray, W. K.; Willoughby, N.; Irvin, B.; Hastie, J.; Schell, D. J.; Nagle, N. *J. Agric. Food Chem.* **2010**, *58*, 12642-12649.
- (70) Chen, S.; Mowery, R. A.; Castleberry, V. A.; van Walsum, G. P.; Chambliss, C. K. *J. Chromatogr. A* **2006**, *1104*, 54-61.
- (71) Guignard, C.; Jouve, L.; Bogéat-Triboulot, M. B.; Dreyer, E.; Hausman, J.; Hoffman, L. *J. Chromatogr. A* **2005**, *1085*, 137-142.
- (72) Becker, C.; Sharma, L. N.; Chambliss, C. K. In *Analytical monitoring of pretreatment and hydrolysis processes in lignocellulose-to-bioalcohol production*; Waldron, K. W., Ed.; Bioalcohol production: Biochemical conversion of lignocellulosic biomass; CRC Press: Boca Raton, 2010; pp 281.
- (73) Weston, D. J.; Bateman, R.; Wilson, I. D.; Wood, T. R.; Creaser, C. S. *Anal. Chem.* **2005**, *77*, 7572-7580.
- (74) Eiceman, G. A.; Blyth, D. A.; Shoff, D. B.; Snyder, A. P. *Anal. Chem.* **1990**, *62*, 1374-1379.
- (75) Dwivedi, P.; Wu, C.; Matz, L. M.; Clowers, B. H.; Siems, W. F.; Hill, J., H. H. *Anal. Chem.* **2006**, *78*, 8200-8206.
- (76) McCooeye, M.; Ding, L.; Gardner, G. J.; Fraser, C. A.; Lam, J.; Sturgeon, R. E.; Mester, Z. *Anal. Chem.* **2003**, *75*, 2538-2542.
- (77) Dwivedi, P.; Hill, J., H. H. *Int. J. Ion Mobil. Spec.* **2008**, *11*, 61-69.
- (78) Steiner, W. E.; Clowers, B. H.; Matz, L. M.; Siems, W. F.; Hill, J., H. H. *Anal. Chem.* **2002**, *74*, 4343-4352.
- (79) Ewing, R. G.; Atkinson, D. A.; Eiceman, G. A.; Ewing, G. J. *Talanta* **2001**, *54*, 515-529.
- (80) Asbury, G. R.; Wu, C.; Siems, W. F.; Hill Jr., H. H. *Anal. Chim. Acta* **2000**, *404*, 273-283.

- (81) Lee, D. S.; Wu, C.; Hill Jr., H. H. *J. Chromatogr. A* **1998**, 822, 1-9.
- (82) Liu, Y.; Clemmer, D. E. *Anal. Chem.* **1997**, 69, 2504-2509.
- (83) Um, B.; Aghazadeh, M.; van Walsum, G. P. A Microassay to Evaluate Inhibition of Enzymatic Hydrolysis Following Dilute Acid Pretreatment of Corn Stover. Presented at 32nd Symposium on Biotechnology for Fuels and Chemicals, April 2010, Poster 3-44.
- (84) Wolff, J.-C.; Eckers, C.; Sage, A. B.; Giles, K.; Bateman, R. *Anal. Chem.* **2001**, 73, 2605-2612.
- (85) de Laeter, J. R.; Böhlke, J. K.; de Bièvre, P.; Hidaka, H.; Peiser, H. S.; Rosman, K. J. R.; Taylor, P. D. P. *Pure Appl. Chem.* **2003**, 75, 683-800.
- (86) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. *Nature Protocols* **2008**, 3, 1139-1152.
- (87) Čmelík, R.; Chmelík, J. *Int. J. Mass Spectrom.* **2010**, 291, 33-40.
- (88) Hoaglund, C. S.; Valentine, S. J.; Clemmer, D. E. *Anal. Chem.* **1997**, 69, 4156-4161.
- (89) Taylor, V. F.; March, R. E.; Longerich, H. P.; Stadey, C. J. *Int J Mass Spectrom* **2005**, 243, 71-84.
- (90) 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 Physiol* **2006**, 141, 1205-1218.
- (91) Chen, S.; Mowery, R. A.; Sevcik, R. S.; Scarlata, C. J.; Chambliss, C. K. *J. Agric. Food Chem.* **2010**, 58, 3251-3258.
- (92) Klinke, H. B.; Ahring, B. K.; Schmidt, A. S.; Thomsen, A. B. *Bioresource Technol.* **2002**, 82, 15-26.
- (93) Sluiter, J. B.; Ruiz, R. O.; Scarlata, C. J.; Sluiter, A. D.; Templeton, D. W. *J. Agric. Food Chem.* **2010**, 58, 9043-9053.
- (94) Ruotolo, B. T.; Gillig, K. J.; Stone, E. G.; Russell, D. H. *J. Chromatogr. B* **2002**, 782, 385-392.
- (95) Wytenbach, T.; von Helden, G.; Bowers, M. T. *J. Am. Chem. Soc.* **1996**, 118, 8355-8364.
- (96) Lau, M. W.; Dale, B. E. *Proc. Natl. Acad. Sci.* **2009**, 106, 1368-1373.

- (97) Sanchez-Patan, F.; Monagas, M.; Moreno-Arribas, M. V.; Bartolome, B. *J. Agric. Food Chem.* **2011**, *59*, 2241-2247.
- (98) Deng, X.; Yang, H.; Li, J.; Song, Y.; Guo, D.; Luo, Y.; Du, X.; Bo, T. *J. Liq. Chromatogr. Rel. Technol.* **2011**, *34*, 2286-2303.
- (99) Guan, F.; Uboh, C. E.; Soma, L. R.; You, Y.; Liu, Y.; Li, X. *J. Mass Spectrom.* **2010**, *45*, 1270-1279.
- (100) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. *Anal. Chem.* **2003**, *75*, 3019-3030.
- (101) Vanderford, B. J.; Pearson, R. A.; Rexing, D. J.; Snyder, S. A. *Anal. Chem.* **2003**, *75*, 6265-6274.
- (102) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J. Mass Spectrom.* **2001**, *36*, 849-865.
- (103) Rodgers, R. P.; Schaub, T. M.; Marshall, A. G. *Anal. Chem.* **2005**, *77*, 20A-27A.
- (104) Corilo, Y. E.; Vaz, B. G.; Simas, R. C.; Lopes Nascimento, H. D.; Klitzke, C. F.; Pereira, R. C. L.; Bastos, W. L.; Santos Neto, E. V.; Rodgers, R. P.; Eberlin, M. N. *Anal. Chem.* **2010**, *82*, 3990-3996.
- (105) Chundawat, S. P. S.; Donohoe, B. S.; da, C. S.; Elder, T.; Agarwal, U. P.; Lu, F.; Ralph, J.; Himmel, M. E.; Balan, V.; Dale, B. E. *Energy Environ. Sci.* **2011**, *4*, 984.
- (106) Altgelt, K. H.; Boduszynski, M. M. *Energy Fuels* **1992**, *6*, 68-72.
- (107) Altgelt, K. H.; Boduszynski, M. M. In *Composition and Analysis of Heavy Petroleum Fractions*; Marcel Decker Inc: NY, 1994; pp 512.
- (108) Boduszynski, M. M. *Energy Fuels* **1987**, *1*, 2-11.
- (109) Boduszynski, M. M. *Energy Fuels* **1988**, *2*, 597-613.
- (110) Boduszynski, M. M.; Altgelt, K. H. *Energy Fuels* **1992**, *6*, 72-76.
- (111) Meyer, R. F.; Attanasi, E. D. *Heavy Oil and Natural Bitumen-Strategic Petroleum Resources*. Fact Sheet No. 70-03; U.S. Geological Survey, 2003, pp 1-6.
- (112) Qian, K.; Rodgers, R. P.; Hendrickson, C. L.; Emmett, C. L.; Marshall, A. G. *Energy Fuels* **2001**, *15*, 492-498.
- (113) Barre, L.; Simon, S.; Palermo, T. *Langmuir* **2008**, *24*, 3709-3717.

- (114) Klein, G. C.; Kim, S.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2006**, *20*, 1965-1972.
- (115) Klein, G. C.; Kim, S.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2006**, *20*, 1973-1979.
- (116) Xu, H.; Yu, D.; Que, G. *Fuel* **2005**, *84*, 647-652.
- (117) Abu-Khader, M. M.; Speight, J. G. *Oil Gas Sci. Technol.* **2008**, *62*, 715-722.
- (118) Gawel, I.; Bociarska, D.; Biskupski, P. *Applied Catalysis A* **2005**, *295*, 89-94.
- (119) Mullins, O. C. *SPE Journal* **2008**, *13*, 48-57.
- (120) Acevedo, S.; Gutierrez, L. B.; Negrin, G.; Pereira, J. C.; Mendez, B.; Delolme, F.; Dessalces, G.; Broseta, D. *Energy Fuels* **2005**, *19*, 1548-1560.
- (121) Suelves, I.; Islas, C. A.; Millan, M.; Galmes, C.; Carter, J. F.; Herod, A. A.; Kandiyoti, R. *Fuel* **2003**, *82*, 1-14.
- (122) Luo, P.; Gu, Y. *Fuel* **2007**, *86*, 1069-1078.
- (123) Muller, H.; Andersson, J. T.; Schrader, W. *Anal. Chem.* **2005**, *77*, 2536-2543.
- (124) Mullins, O. C.; Sheu, E. Y. In *Structure and Dynamics of Asphaltenes*; Plenum Press: New York, 1998; pp 314.
- (125) Merdrignac, I.; Espinat, D. *Oil Gas Sci. Technol.* **2007**, *62*, 7-32.
- (126) Speight, J. G. *Oil Gas Sci. and Technol.* **2004**, *59*, 467-477.
- (127) Hortal, A. R.; Hurtado, P.; Martinez-Haya, B.; Mullins, O. C. *Energy Fuels* **2007**, *21*, 2863-2868.
- (128) Becker, C.; Qian, K.; Russell, D. H. *Anal. Chem.* **2008**, *80*, 8592-8597.
- (129) Groenzin, H.; Mullins, O. C. *J. Phys. Chem. A* **1999**, *103*, 11237.
- (130) Pomerantz, A. E.; Hammond, M. R.; Morrow, A. L.; Mullins, O. C.; Zare, R. N. *J. Am. Chem. Soc.* **2008**, *130*, 7216-7217.
- (131) Schneider, M. H.; Ballard Andrews, A.; Mitra-Kirtley, S.; Mullins, O. C. *Energy Fuels* **2007**, *21*, 2875-2882.
- (132) Badre, S.; Goncalves, C. C.; Norinaga, K.; Gustavson, G.; Mullins, O. C. *Fuel* **2005**, *85*, 1-11.

- (133) Cunico, R. L.; Sheu, E. Y.; Mullins, O. C. *Petrol. Sci. Technol.* **2004**, 22, 787-798.
- (134) Qian, K.; Dechert, G. J.; Edwards, K. E. *Int. J. Mass Spectrom.* **2007**, 265, 230-236.
- (135) Scotti, R.; Montanari, L. Molecular Structure and Intermolecular Interaction of Asphaltenes by FT-IR, NMR and EPR. In *Structure and Dynamics of Asphaltenes*; Mullins, O. C., Sheu, A., Eds.; Plenum Press: New York, 1998; pp 79-114.
- (136) Bergmann, U.; Mullins, O. C. Carbon X-Ray Raman Spectroscopy of PAH's and Asphaltenes. In *Asphaltenes, Heavy Oils and Petroleomics* Mullins; O. C., Sheu, A., Hammami, A. and Marshall, A. G., Eds.; Springer: New York, 2007; pp 139-153.
- (137) Mitra-Kirtley, S.; Mullins, O. C.; Elp, J. V.; George, S. J.; Chen, J.; Cramer, S. P. *J. Am. Chem. Soc.* **1993**, 115, 252-258.
- (138) Hsu, C. S.; Qian, K.; Chen, Y. C. *Anal. Chim. Acta* **1992**, 264, 79-89.
- (139) Sharma, B. K.; Sharma, C. D.; Tyagi, O. S.; Bhagat, S. D.; Erhan, S. Z. *Petrol. Sci. Technol.* **2007**, 25, 121-139.
- (140) Qian, K.; Robbins, W. K.; Hughey, C. A.; Cooper, H. J.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2001**, 15, 1505-1511.
- (141) Jaffe, S. B.; Freund, H.; Olmstead, W. N. *Ind. Eng. Chem. Res.* **2005**, 44, 9840-9852.
- (142) Gallagher, R. T.; Balogh, M. P.; Davey, P.; Jackson, M. R.; Sinclair, I.; Southern, L. J. *Anal. Chem.* **2003**, 75, 973-977.
- (143) Gidden, J.; Wytenbach, T.; Jackson, A. T.; Scrivens, J. H.; Bowers, M. T. *J. Am. Chem. Soc.* **2000**, 122, 4692-4699.
- (144) Pinkston, D. S.; Duan, P.; Gallardo, V. A.; Habicht, S. C.; Tan, X.; Qian, K.; Gray, M.; Müllen, K.; Kenttämä, H.I. *Energy Fuels* **2009**, 23, 5564-5570.
- (145) Becker, C.; Fernandez-Lima, F. A.; Russell, D. H. Moving Towards Accurate Molecular Compositions of Petroleum Products Using Ion Mobility-Mass Spectrometry. Presented at 56th Annual ASMS Conference on Mass Spectrometry, Denver, CO, May 2008.
- (146) Mullins, O. C. *Annu. Rev. Anal. Chem.* **2011**, 4, 393-418.
- (147) Spangler, G. E. *Int. J. Mass Spectrom.* **2002**, 220, 399-418.
- (148) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, 119, 2987-2994.

- (149) Hsu, C. S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *J. Mass Spectrom.* **2011**, *46*, 337-343.
- (150) Hedayatifar, N. Chemometric Modeling of UV-Visible and LC-UV Data for Prediction of Hydrolysate Fermentability and Identification of Inhibitory Degradation Products, Baylor University, Waco, TX, 2011.