

METABOLOMIC PROFILING OF LIGNOCELLULOSIC BIOMASS PROCESS STREAMS

By

Afrand Kamali Sarvestani

A DISSERTATION

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

Chemistry—Doctor of Philosophy

2016

ABSTRACT

METABOLOMIC PROFILING OF LIGNOCELLULOSIC BIOMASS PROCESS STREAMS

By

Afrand Kamali Sarvestani

Practical advances in conversion of lignocellulosic biomass to bioethanol require improvement of current knowledge about the composition and chemical diversity of lignin. This research aimed to perform mass spectrometry-based metabolomic profiling of grasses to fill in these knowledge gaps. Six grasses, and the hardwood poplar, were investigated including: corn stover, wheat straw, rice straw, switchgrass, *Miscanthus*, and sorghum. Methanolic extracts of untreated biomass contained intact lignin constituents that were profiled using ultrahigh performance liquid chromatography-time-of-flight mass spectrometry (UHPLC-TOF MS). Most extractives exhibited molecular masses consistent with phenolic compounds. Derivatives of the flavonoid triclin were prominent in extracts of all grasses but were not detected in poplar. Multiplexing of non-selective collision-induced dissociation (CID) in UHPLC-MS analyses provided evidence that more than 90% of triclin is incorporated in compounds other than the small number of abundant triclin derivatives. Wide-mass window tandem mass spectrometry (MS/MS) analyses provided further evidence for incorporation of triclin, *p*-coumarates, and other monolignols up to 4 kDa molecular mass.

Mass spectrometric profiling of grass extractives also revealed abundant phenolic acid esters of glycerol, with 1-*p*-coumaroyl-3-feruloylglycerol the most abundant in corn stover. Profiles yielded evidence that these esters undergo dehydrodimerization via reactions of to form diferulate esters. The presence of these novel compounds demonstrates that glycerol esters

undergo oligomerization similar to classical monolignols. In addition, extracts contained conjugates of triclin with 1-*p*-coumaroyl-3-feruloylglycerol, and provide further evidence of.

Extracts of corn stover also contained numerous oligomers of the ester sinapyl *p*-coumarate. Using UHPLC-HRMS, MS/MS, and 1D and 2D NMR spectra, the most abundant isomer is formed by 8-8 coupling of two sinapyl *p*-coumarate groups followed by formation of a tetrahydrofuran (THF) core with oxygen bridging between position-7 carbons of each sinapyl group. Larger phenolic constructs containing this compound in their core were also annotated using UHPLC-MS/MS that were formed by addition of either more oxidized sinapyl *p*-coumarate groups or oxidized coniferyl alcohols. Compounds of this category were detected up to m/z 2000 in agreement with wide-mass window MS/MS analyses.

High molecular weight lignins may have low solubility, and to explore the range of lignin composition, γ -valerolactone (GVL) was evaluated as a solvent for electrospray ionization (ESI)–MS and as a mobile phase component for reversed phase liquid chromatography. GVL yielded simpler ESI spectra of phenolic substances due to reduced adduct multiplicity for each molecule. GVL in the mobile phase resulted in faster elution relative to methanol, while chromatographic resolution for major extractives of corn stover was retained. GVL also exhibited differential retention selectivity for lignin compounds compared to methanol which offers benefits for separation and analysis of polymers and large natural products.

Taken together, these findings suggest diversity of lignin components in grasses is extensive despite a limited range of lignin precursors. Recognition of the nature of triclin and *p*-coumarate ester derivatives provides a foundation for novel strategies for deconstructing lignins and converting biomass to renewable fuels and chemical feedstocks.

**Copyright by
AFRAND KAMALI SARVESTANI
2016**

TO MOM AND DAD

ACKNOWLEDGMENTS

Firstly, this work was not possible without the wisdom, guidance, and support of my advisor, Professor A. Daniel Jones. Dan's knowledge and mentorship are priceless to me. During my graduate studies, Dan always found a way to respond to my questions, despite his having so many other problems to contend with. Dan constantly challenged my mind to find answers and understand the nature of problems. I greatly admire the patience that Dan displayed while guiding me. In addition to the many aspects of science and research that I learned from Dan, I also learned a lot about 'real world' problems and issues, such as how to prepare for careers in industries and how to face the politics ruling the science and industries. My experience in Dan's group was the best introduction and preparation for a professional scientist's life one could ask for. However, now I am leaving his lab, I am hopeful that Dan will continue to provide me with advice for major future professional decisions that I will have to make over the course of my career. Thank you Dan for everything you have done for me and for everything I have learned from you.

I would like to convey my gratitude to Professor Babak Borhan, because he was always a support to my career and professional life ever since we met, and I am sure his support will continue into the future too. Thank you Babak for having your office and home's door open to me literally all the time anytime I needed to see you. Thank you for your scientific contributions, including your contributions in your role as a graduate committee member. Also, thanks for the non-scientific advice you provided relating to other aspects of my life, and thanks for all the

amazing cuisines you made and allowed me to enjoy in your office as if I were one of your own students.

I would also like to thank the other members of my graduate committee: Professor Merlin Bruening and Professor Gary Blanchard. Thank you for your guidance and constructive criticism during my comprehensive exam and my final dissertation. I also want to thank both of you for the courses you offered that I have attended, CEM 834 and CEM 835 ('Advanced Analytical Chemistry').

I want to thank all past and current members of the Jones group at Michigan State University for their cooperation, collaborations, and consultations. I would especially like to thank Xiaoxiao Liu for her major help and contribution on the acquisition of NMR data. I want to thank Dr. Zhenzhen Wang, Dr. Chen Zhang, Dr. Sujana Pradhan, Cindy Kaeser, Kristen Reese, Fanny Chu, Steven Hurney, Dr. Prabodha Ekanayaka, Dr. Banibrata Ghosh, and Dr. Siobhan Shay. I want to thank Dr. Chao Li for incredible advice and support relating to the development of my professional career prior to and after graduation. I am thankful to my Great Lakes Bioenergy Research Center (GLBRC) collaborators. I would especially like to thank the members of the Biomass Conversion Research Lab (BCRL), Professor Bruce Dale, Dr. Leonardo Sousa, Dr. Venkatesh Balan, James Humpula, and Christa Gunawan for all their help and contributions toward this research.

I also would like to convey my gratitude for the support I have received from the RTSF Mass Spectrometry & Metabolomics Core at Michigan State University. My especial thanks go to my friend Dr. Scott Smith for his constant contribution in my research challenges and the

tea/coffee times we had; thanks buddy! I also thank Lijun Chen and Dr. Tony Schillmiller for all the help they provided.

I also want to thank my other mass spectrometrists friends at Michigan State University, Dr. Cassie Phaner, Dr. Li Cui, Dr. Xiao Zhu, Dr. Shuai Nie, and Dr. Todd Lydic. Thank you for the camaraderie and your support!

I would also like to thank the Department of Chemistry at Michigan State University and GLBRC. These institutions funded my stipend and my research, respectively, throughout the duration of my career at MSU.

Many friends have made my career in Michigan State University memorable. My very special thanks go to Dr. Ramin Vismeh, who has been my friend, brother, colleague, roommate, and secret-keeper, who has provided me with his full support through some of the most challenging days of my life in Michigan. I am also thankful to have known Mersedeh, Hamideh, Bardia, Zahra, Farid Radifa, and Tate for their great companionship in happy and sad moments of my career at MSU. I also want to thank Maryam, Atefeh, Afra, Roozbeh, Meisam, Nastaran, Faezeh, and Omid. Thank you all for being such good friends.

To my wonderful companion, Dr. Juliana Lopes: thank you for all of the emotional, social, and scientific support you provided for me. You inspire me and make my life beautiful. I love you!

Lastly, and yet very importantly, I want to convey my gratitude for my family. More than anyone else, I want to thank my two greatest teachers, my parents, who raised their kids with love and who taught us a love for learning. Thank you Baba Rahim; it breaks my heart and brings tears to my eyes as I write this, that you could not see me finishing this work and my PhD graduation. Thank you Maman Aali for all of the love you give me and the enthusiasm for learning that you encouraged me to have. Thank you to my big brothers, Amir and Iman, for being two of my earliest teachers; you two ignited the excitement for science in my mind even prior going to elementary school. Thanks to my only sister, Andisheh, for watching out for me and for teaching me a lot about life. Thank you to Aram and Arash, my artist brothers, with all of your unique compassion and enthusiasm that you have provided to me. And, thank you to my little brother, Ali, for all of your curiosity and sweetness. I also want to thank all of my in-laws and my extended family. The path I have taken to the present has definitely been affected by my large family; exposure to my educated family since I was a kid has contributed to my interest in science. Thank you all!

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiii
KEY TO ABBREVIATIONS.....	xviii
Chapter One: The Necessity of Improved Strategies for Analysis of Lignocellulosic Biomass for Development of Renewable Liquid Fuels..... 1	
1.1 Introduction: Energy, carbon resource, environment, and sustainability: where does the need for bioethanol come from?.....	2
1.2 Bioethanol and its current sources.	3
1.3 Food vs. Fuel.....	4
1.4 Grass cell wall and lignocellulosic materials.....	5
1.5 Lignin.....	7
1.6 Pretreatment and effect of its products on ethanol production.	10
1.7 Analysis of lignin and its degradation products.....	19
REFERENCES	23
Chapter Two: A metabolomic investigation of diversity of tricin incorporation and pretreatment transformations in lignin extractives of grasses..... 28	
2.1 Abstract	29
2.2 Introduction.....	30
2.3 Experimental	31
2.4 Results and Discussion	33
2.4.1 Untargeted metabolomic profiling of extracts of untreated biomass.....	33
2.4.2 Profiles of tricin derivatives in extracts of untreated biomass	37
2.4.3 Mass balance of tricin derivatives measured using UHPLC-MS	42
2.4.4 Exploration of tricin conjugate diversity using tandem mass spectrometry (MS/MS)....	43
2.4.5 Tricin derivatives after extractive ammonia pretreatment	47
2.4.6 Incorporation of tricin into flavonolignans through the action of oxidative enzymes.....	48
2.5 Conclusions.....	52
REFERENCES	56
Chapter Three: A metabolomics investigation into whether phenolic acid esters are incorporated into lignin in monocot grasses 59	
3.1 Abstract	60
3.2 Introduction.....	61
3.3 Experimental	66
3.4 Results and Discussion: Pathways of Incorporation of Phenolic Acids into Grass Lignin	68
3.4.1 Hydroxycinnamoyl (<i>p</i> -Coumaroyl and Feruloyl) Glycerols.....	68
3.4.2 Hydroxycinnamoyl (<i>p</i> -coumaroyl and feruloyl) in Tricin Derivatives.....	79
3.4.3 Couplings of Sinapyl <i>p</i> -Coumarates to Make Larger Lignin Molecules	80

3.5 Conclusions	97
REFERENCES	99

Chapter Four: Mass Spectrometric Analysis of Polymers and Biomass Extractives using Liquid Chromatography-Time-of-Flight Mass Spectrometry with γ -Valerolactone as a Renewable Mobile Phase Component

4.1 Abstract	104
4.2 Introduction	105
4.3 Experimental	112
4.4 Results and Discussion	113
4.5 Conclusions	121
REFERENCES	123

Chapter Five: Concluding Remarks.....	126
REFERENCES	131

LIST OF TABLES

Table 1.1. List of compounds produced from different biomass sources upon pretreatment.....	12
Table 2.1. Major peaks in the base peak intensity UHPLC-MS chromatograms of extractives from different biomass sources shown in Figure 2.1. Tricin-containing substances are highlighted in bold text. Levels in biomass were estimated from extracted ion chromatogram peak areas by using the molar response for triclin as the response factor for all substances. Nomenclature for triclin compounds follows the convention as proposed in.....	36
Table 2.2. Annotations for UHPLC-MS peaks for untreated (designated as ‘u’) and AFEX-treated corn stover (designated ‘t’) from Figure 2, focusing on putative triclin derivatives. Abbreviations follow the convention of Lan et al. Plant Physiol. 2016. Compounds containing amino groups in place of hydroxyls in the monolignol portions are designated with ‘(NH ₂)’ in the structure abbreviation. Relative mass defect (RMD) values reflect fractional hydrogen content. XIC peak areas are for [M-H] ⁺ ions.....	38
Table 3.1. NMR Shift Assignments for 1- <i>p</i> -coumaroyl-3-feruloylglycerol.....	71
Table 3.2. ¹ H NMR and ¹³ C NMR assignments of the most abundant isomer of compound bis-sinapyl <i>p</i> -coumarate in maize with formula C ₄₀ H ₄₀ O ₁₃ (position numbers are shown in Figure 3.11).....	84
Table 4.1. Physicochemical properties of three common RPLC solvents [14] and β-valerolactone (GVL).....	110

LIST OF FIGURES

Figure 1.1. Ethanol Fuel production and import in the first decade of 2000's.....	4
Figure 1.2. Schematic view of plant cell wall components.....	6
Figure 1.3. Schematic demonstration of how pretreatment makes polysaccharides more accessible.....	7
Figure 1.4. Structure of major lignin monomers. $R_1=R_2=H$: <i>p</i> -coumaryl alcohol, $R_1=H$ $R_2=OCH_3$: coniferyl alcohol, and $R_1=R_2=OCH_3$: sinapyl alcohol.....	8
Figure 1.5. Panel A: Numbering system used for monolignols, Panel B: delocalization of radical at diverse positions in an oxidized monolignol, Panel C: examples of the different linkages that could form from coupling via different locations of the unpaired electron.....	9
Figure 1.6. Classification of lignin units based on the number and position of their aromatic methoxy groups.....	10
Figure 1.7. Schematic view of derivatization of lignin followed by reductive cleavage.....	20
Figure 2.1 Base peak ion (BPI) abundance UHPLC/TOF MS chromatograms generated in negative-ion mode for methanol extracts of (A) poplar, (B) sorghum, (C) corn stover, (D) wheat straw, (E) rice straw, (F) <i>Miscanthus</i> , and (G) switchgrass. Labeled peaks are annotated in Table 2.1. Base peak abundances corresponding to the 100% level for each chromatogram are included below the name of each biomass source.....	35
Figure 2.2. UHPLC-MS Extracted ion chromatograms for deprotonated triclin fragments at m/z 329 at elevated collision voltage for extracts of untreated (A) poplar, (B) sorghum, (C) corn stover, (D) wheat straw, (E) rice straw, (F) <i>Miscanthus</i> , and (G) switchgrass. Peaks and retention times corresponding to triclin (T) and the two guaiacylglycerol triclin (GGT) isomers are indicated. The peak in the poplar chromatogram highlighted with an asterisk is not triclin.....	40
Figure 2.3. Product ion MS/MS spectra generated in negative-ion mode for $[M-H]^-$ ions from (A) guaiacylglyceroltriclin (GGT1) from an extract of untreated corn stover and (B) the analogous compound differing by replacement of a hydroxyl group by an amino group in extracts of EA-treated corn stover. Product ions highlighted in red (color version) are attributed to the flavonoid portion of each molecule, and those highlighted in blue are attributed to the monolignols portion.....	41
Figure 2.4. UHPLC-MS Extracted ion chromatogram for m/z 329.066 ± 0.005 obtained for methanolic extract of untreated corn stover in negative-ion mode. The top panel shows the	

integrated peak area for triclin, and the bottom panel shows integration of all signal at this m/z eluting from 12-25 minutes.....43

Figure 2.5. High-mass region of the negative-ion mass spectrum obtained by averaging all mass spectra from UHPLC-MS analysis of a methanolic extract of untreated corn stover over the retention time region of 15-25 minutes. Peak averaging was performed using a bin width of 0.05 m/z . Spectra were generated at the lowest collision potential (5 V). The magnified inset of the range from m/z 1350-1360 demonstrates multiple resolved peaks at every nominal mass.....45

Figure 2.6. Histogram showing the frequency of ions in the averaged mass spectrum shown in Figure 2.5, sorted by the absolute signal ($\log_2(\text{number of ion counts})$). An arrow points to the $[M-H]^-$ signals for abundant molecules triclin and GGT isomers but do not appear because so few molecules had such high abundance.....46

Figure 2.7. Product ion abundances as a function of precursor ion m/z generated from wide-window (~ 60 Da) MS/MS spectra of a methanolic extract of untreated corn stover. Sample introduction employed flow injection analysis and negative-mode electrospray ionization. (A) Product ions m/z 329 (deprotonated triclin, red circles), 195 (monolignol G, blue squares), 165 (monolignol H, green line and black squares), and 225 (monolignol S, green line with yellow squares). (B) Product ions for phenolic acid anions m/z 163 (*p*-coumarate, turquoise), 193 (ferulate, violet), and 223 (sinapate, blue). Vertical axis scaling is the same for both panels.....49

Figure 2.8. UHPLC-MS extracted ion chromatograms for m/z 329.07 generated in negative-ion mode for methanolic extracts of (A) untreated corn stover and (B) AFEX-treated corn stover. Annotations of peaks are presented in Table 2.2.....50

Figure 2.9. Extracted ion UHPLC-MS chromatograms for triclin and its mono- and di-lignol conjugates (m/z 329.07) in (A) AFEX-treated corn stover and (B) untreated corn stover.....51

Figure 2.10. Extracted ion UHPLC-MS chromatograms for combined signals of m/z 567.15, 671.17, and 879.25 corresponding to esterified monolignol conjugates of triclin for extracts of (A) AFEX-treated corn stover and (B) untreated corn stover. The two chromatograms share a common vertical scale (100% = ion counts). Products containing benzylic amino groups are detected at the extracted ion masses, but correspond to substances containing one heavy isotope. As a result, their signals are approximately one-third of the monoisotopic ion signals.....52

Figure 2.11. UHPLC-MS extracted ion chromatograms for m/z 525 generated in negative-ion mode for control reactions (A-C) and incubations of triclin and coniferyl alcohol with laccase enzyme.....54

Figure 2.12. UHPLC-MS extracted ion chromatograms for m/z 341 (positive-ion mode; $[M+H-H_2O]^+$) for detection of dehydromer of coniferyl alcohol in (A) control, no coniferyl alcohol, (B) coniferyl alcohol plus laccase, (C) coniferyl alcohol without laccase, and (D) coniferyl alcohol plus triclin and laccase.....55

Figure 3.1. A model representation of hemicellulose arabinoxylan chains showing diferulate crosslinks between arabinoxylan chains (highlighted in red)	63
Figure 3.2. Results from UHPLC-MS profiling of methanolic extract of untreated corn stover. Extracted ion chromatograms (XICs) at 40 V collision potential for (A) ferulate fragment ion at m/z 193.051 (B) <i>p</i> -coumarate fragment ion at m/z 163.040, (c), triclin fragment ion at m/z 329.067, and (D) base peak ion (BPI) chromatogram at the lowest collision potential (5 V).....	69
Figure 3.3. MS/MS spectrum of product ions of m/z 413.124 from 1- <i>p</i> -coumaroyl-3-feruloylglycerol in a methanolic extract of corn stover, with proposed assignments of product ions and neutral mass losses.....	70
Figure 3.4. MS/MS product ion spectrum of m/z 237.07 (1- <i>p</i> -coumaroylglycerol) from a methanolic extract of untreated corn stover, with assignments of product ions and neutral mass losses.....	74
Figure 3.5. Narrow mass window UHPLC-MS extracted ion chromatograms of a methanolic extract of untreated corn stover showing: (A) m/z 443.135 ± 0.05 (diferuloyl glycerol), (B) m/z 413.124 ± 0.05 (<i>p</i> -coumaroyl-3-feruloylglycerol), and (C) m/z 383.114 ± 0.05 (di- <i>p</i> -coumaroyl glycerol).....	76
Figure 3.6. Extracted ion UHPLC-MS chromatogram peak areas for 1- <i>p</i> -coumaroyl-3-feruloylglycerol, diferuloyl glycerol, and di- <i>p</i> -coumaroyl glycerol for methanolic extracts of grasses and poplar. Note the logarithmic scale for the vertical axis. PAEGs are abundant in grasses but insignificant in poplar.....	77
Figure 3.7. (A) Oxidative coupling of two 1- <i>p</i> -coumaroyl-3-feruloylglycerol to yield the 8- <i>O</i> -4 diferulate-linked dimer at m/z 825.24 and (B) UHPLC-MS/MS chromatogram for m/z 825.24 for a methanolic extract of corn stover showing peaks attributed to multiple isomeric forms.....	77
Figure 3.8. Annotation of one of the bis-(<i>p</i> -coumaroylferuloylglycerol) isomers of detected in an extract of corn stover (m/z 825.24) eluted at 17.9 minutes using high resolution MS/MS. The zoomed in outset shows the key fragments resulting in deduction of diferulate core in the center of the structure (shown in Figure 3.9)	81
Figure 3.9. A series of key fragment ions (starting from precursor at m/z 825 to diferulate-derived m/z 341) that are useful to deduce presence diferulate core at the center of the structure of one of the isomers of bis-(1- <i>p</i> -coumaroyl-3-feruloylglycerol), corresponding MS/MS spectrum is in the zoomed-in outset of Figure 3.8 showing detection of all these fragment ions.....	82
Figure 3.10. Product ion MS/MS spectrum of m/z 741 for a product in methanolic extract of corn stover, annotated as a conjugate of triclin with <i>p</i> -coumaroyl-feruloylglycerol. Peaks in the spectrum from m/z 100-230 are magnified by a factor of 20, and from m/z 350-750, magnified by a factor of 15.....	83

Figure 3.11. Structure of the most abundant isomer of the compound with formula $C_{40}H_{40}O_{13}$ named bis-(sinapyl <i>p</i> -coumarate)	85
Figure 3.12. Formation of fused resinol rings upon radical coupling of sinapyl alcohols as 8–8 bond.....	86
Figure 3.13. Radical coupling of two sinapyl <i>p</i> -coumarates followed by formation of 5-membered resinol ring.....	86
Figure 3.14. MS/MS product ion spectrum of the most abundant isomer of m/z 727, dimer of sinapyl <i>p</i> -coumarate.....	87
Figure 3.15. Representative structures of (S-H [•])-O and G ^O units.....	88
Figure 3.16. Panel A: UHPLC-MS/MS base peak intensity chromatogram for product ion scan of m/z 923.31 for a methanolic extract of corn stover. The isomers resulting in m/z 727.23 product ion are labeled with red * and the other peaks yielded product ions of m/z 741.23 or m/z 743.25. Panel B: MS/MS product ion spectrum (m/z 923.3) of the most abundant isomer labeled with a green arrow in panel A. Panel C: a proposed structure for the compounds fragmented in panel B: The moiety highlighted in blue represents the G ^O substructure, the only G ^O unit precursor (described in Chapter 1) of this compound.....	90
Figure 3.17. (A): UHPLC-MS/MS-base peak intensity (BPI) chromatogram of product ions of m/z 1099.36, measured for a methanolic extract of untreated corn stover in negative-ion mode; (B): MS/MS product ion spectrum of the most abundant peak in A (identified by arrow); (C): proposed annotation of the compound with formula $C_{60}H_{60}O_{20}$ in maize as a trimer of (S-H [•]); [bis-(S-H [•])]-(S-H [•])-O. Each color separates one (S-H [•]) unit.....	91
Figure 3.18. One of the possible representative reactions that involves addition of third (S-H [•]) to bis-(S-H [•]) core to form [bis-(S-H [•])]-(S-H [•])-O.....	92
Figure 3.19. MS/MS product ion spectrum of m/z 1471.5 from a methanolic corn stover extract, annotated to be tetramer of (S-H [•]) which in its simplest form here shown as [bis-(S-H [•])]-(S-H [•])-O-(S-H [•])-O.....	94
Figure 3.20. MS/MS product ion spectrum of m/z 1295.43 of the most abundant isomer from corn stover annotated as [bis-(S-H [•])]-(S-H [•])-O- G ^O	94
Figure 3.21. Pattern of addition of (S-H [•])-O and G ^O units in compounds containing bis- (S-H [•]) core in a single point spectrum of UHPLC-MS of corn stover extract.....	95
Figure 4.1. Relative Solvent Parameters (normalized to highest value in each category) of water, methanol, acetonitrile, and GVL.....	111
Figure 4.2. Repeating unit of poly(4-vinylphenol)	112

Figure 4.3. (Left panel) Negative-ion MALDI mass spectrum of small PVP solution in GVL and (Right panel) negative-ion electrospray ionization mass spectrum of the same solution.....114

Figure 4.4. Flow-injection analysis (FIA)–MS of large PVP polymer obtained using electrospray ionization in negative-ion mode. Red circles show the zoomed-in view of the corresponding regions in the spectrum.....115

Figure 4.5. Comparison of negative ESI mass spectra of small PVP at 5 mg/mL in GVL (top) and in methanol (bottom).....118

Figure 4.6. UHPLC-MS of small PVP polymer in methanol and GVL. BPI chromatograms (top two panels). Asterisk and double-asterisk signs show the time points chosen to show extracted ion chromatograms (XICs). Two middle panels demonstrate XIC of DP=4+phenyl PVP as eluted by gradient of methanol and GVL in water, and the bottom two panels show XIC of DP=19+benzoate PVP as eluted by methanol and GVL.....119

Figure 4.7. BPI chromatograms of methanol extract of UTCS analyzed by UHPLC-MS using methanol (top panel) and GVL (bottom panel) in solvent gradient. Similar letters at different chromatograms denote identical compounds within each chromatogram. Peak labeled A: m/z 539.152, B: m/z 121.033 (benzoic acid), C: m/z 163.044 (*p*-coumaric acid), D: m/z 637.146, E: m/z 329.069 (tricin), E and F: m/z 525.137 (GGT), H: m/z 413.128 (1-*p*-coumaroyl-3-feruloylglycerol), I: m/z 567.154 (acetyl GGT), J: m/z 701.192 (O-9-(*p*-coumaroyl)syringyl glyceryl tricin), K: m/z 727.238 (bis-(sinapyl *p*-coumarate) L: m/z 671.180 (O-9-(*p*-coumaroyl)guaiacylglyceryl tricin)120

KEY TO ABBREVIATIONS

UHPLC	Ultra High Pressure Liquid Chromatography
TOF	Time-of-Flight
MS	Mass Spectrometry
CID	Collision Induced Dissociation
NMR	Nuclear Magnetic Resonance
THF	Tetrahydrofuran
GVL	γ -Valerolactone
ESI	Electrospray Ionization
MTBE	Methyl Tertiary Butyl Ether
EPA	Environmental Protection Agency
AFEX	Ammonia Fiber Expansion
UV	Ultraviolet
DFRC	Derivatization Followed by Reductive Cleavage
GC	Gas Chromatography
MALDI	Matrix Assisted Laser Desorption/Ionization
SEC	Size Exclusion Chromatography
EA	Extractive Ammonia Fiber Expansion
HPLC	High Pressure Liquid Chromatography
WWB	Wet Weight Basis
ASE	Accelerated Solvent Extraction
QTOF	Quadrupole Time-of-Flight

GGT	Guaiacylglyceryltricin
XIC	Extracted Ion Chromatogram
RMD	Relative Mass Defect
LC	Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
TIC	Total Ion Chromatogram
BPI	Base Peak Ion-Chromatogram
HNMR	Hydrogen NMR
HSQC	Heteronuclear Single Quantum Coherence
HMBC	Heteronuclear Multiple Bond Correlation
COSY	Correlation Spectroscopy
PAEG	Phenolic Acid Esters of Glycerol
PVP	poly(4-vinylphenol)
ASL	Acid Soluble Lignin
RPLC	Reversed Phase Liquid Chromatography
LA	Levulinic Acid
DHB	2,5-Dihydroxybenzoic Acid
FIA	Flow Injection Analysis
DP	Degree of Polymerization

**Chapter One: The Necessity of Improved Strategies for Analysis of Lignocellulosic Biomass
for Development of Renewable Liquid Fuels**

1.1 Introduction: Energy, carbon resource, environment, and sustainability: where does the need for bioethanol come from?

Fossil fuel resources are limited and will eventually become scarcer. The growing worldwide human population, now exceeding 7 billion, is expected to place greater demands on all sources of energy including fossil fuels. Meanwhile, political conflicts and economic challenges can be expected to pose risks to safe extraction, delivery, and storage of fossil fuels. In addition, fossil fuels remain important sources of organic carbon feedstocks for numerous industries[1]. The extensive use and combustion of fossil fuels is believed responsible for growth in levels of atmospheric carbon dioxide that threaten to drive worldwide changes in climate [2]. These factors are driving growing interest in developing an economically competitive generation of liquid transportation fuels and chemical feedstocks from renewable resources.

Production of biofuels from plants, which derive nearly all of their carbon from photosynthetic carbon fixation, offers potential as a renewable alternative to fossil fuels. Plants and photosynthetic microbes provide the most promising routes of converting carbon from its oxidized atmospheric form CO₂ to organic reduced forms, relying on sunlight and inorganic nutrients as drivers of carbon fixation.

The past decade has witnessed revived interest in developing liquid transportation fuels, particularly bioethanol, from renewable sources. Bioethanol has great potential for large-scale biofuel production because advanced engineering technologies yield efficient conversion of monosaccharides to ethanol via fermentation. Bioethanol plus biodiesel provided 2.7% of the world transportation energy sector in 2010 [3]. The expectation is that this contribution could reach 27% by 2050, according to international energy agency (<https://www.iea.org/topics/renewables/subtopics/bioenergy/>).

1.2 Bioethanol and its current sources.

Most bioethanol is produced from sugar cane and corn (food sources) [4-5] and is already blended with gasoline in the United States and even more extensively in Brazil. To keep supplies abundant and end product prices low, governments have subsidized bioethanol produced from both sugar cane and corn [6]. However, diverting food products into biofuel production is unlikely to be sustainable when foods are scarce, and the choice between food and fuel presents a dilemma with often tragic outcomes [6].

Another driving force for blending ethanol in gasoline is the excessive leakage of oxygenating additive methyl tertiary butyl ether (MTBE), which is added to modulate fuel combustion characteristics and address air pollution concerns, into underground water resources. Ethanol provides an environmentally-healthy replacement for MTBE. In 2005, the US Congress passed the Energy Policy Act that removed the oxygenate requirement for reformulated gasoline (RFG). At the same time, Congress also instituted a renewable fuel standard. Required levels of ethanol that must be blended in fuel have been systematically hiked over the past years, and according to the 2017 mandate of the US Environmental Protection Agency (EPA), 18 billion gallons of ethanol per year will be blended into gasoline in the USA. Groups opposing ethanol fuel claim agricultural production of ethanol precursors will take over the majority of cornfields in the USA (<http://www.ewg.org/research/ethanols-broken-promise/emissions-land-use-change>). It is however undeniable that ethanol fuel production has risen dramatically over the past decade. Figure 1.1 shows the domestic production and import of ethanol fuel during the first decade of 21st century in USA according to a report by the Renewable Fuel Association [7]. According to the same report, production of ethanol fuel in Brazil and the U.S. combined accounted for 87% of global production during 2011.

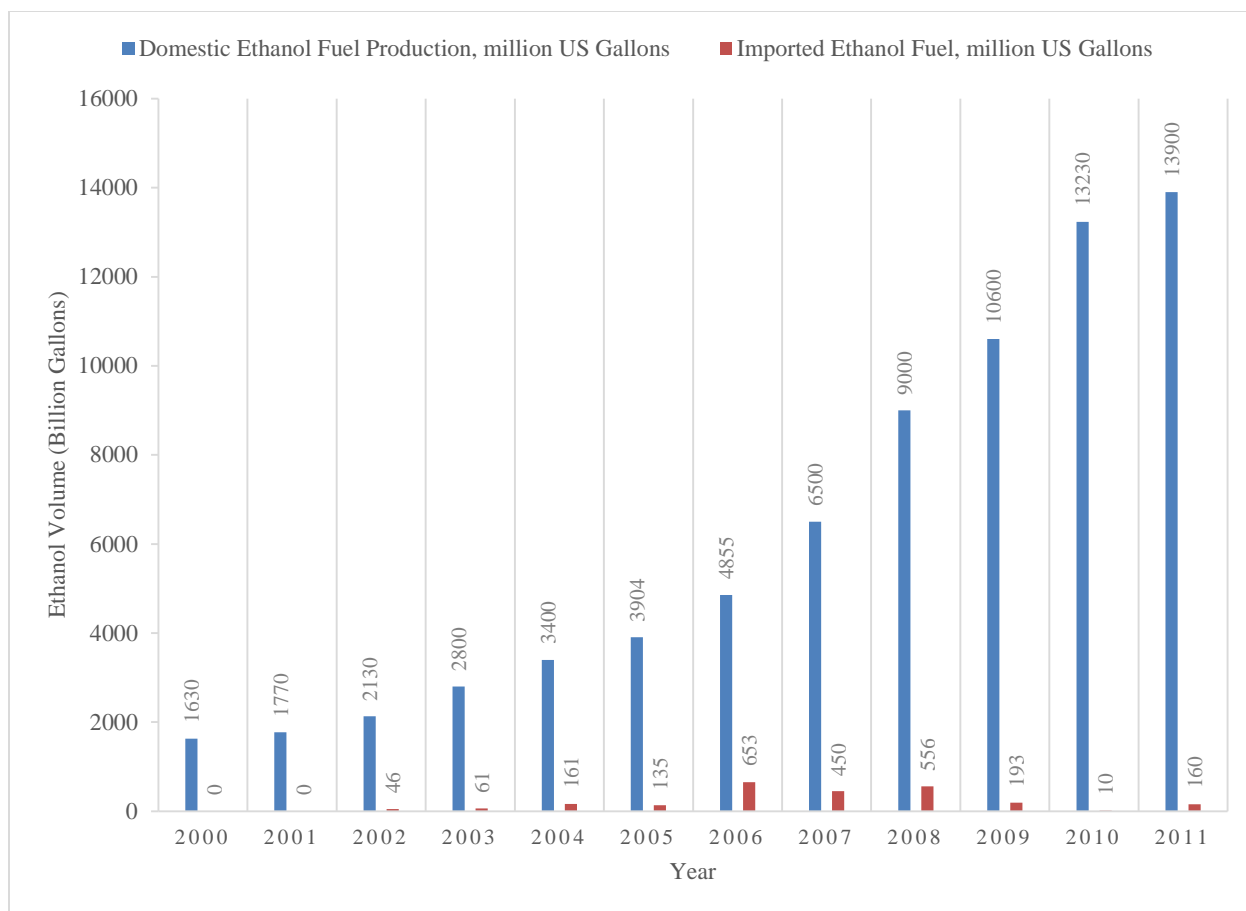


Figure 1.1. Ethanol Fuel production and import in the first decade of 2000's

1.3 Food vs. Fuel

Increases in food prices accompanied by conflicts of interest towards fossil fuel industry have made the conversion of foods to bioethanol a politically-charged subject. As a matter of fact, conflict between US oil and corn businesses has continued [8] Aside from current high levels of US oil and natural gas production that have surpassed Saudi Arabia and Russia, other factors including pressure to increase food prices, and plowing forests into corn fields are the main arguments made by opposing parties to influence government policies. That is why the US government has increased the demands for conversion of non-food sources such as lignocellulosic materials into transportation fuels [9-10]. Use of fast-replenishing non-food sources including grasses presents one of the most promising solutions because grasses grow

faster and convert more atmospheric carbon dioxide into biomass, and eventually to fuel, faster than trees grown on the same area of land.

1.4 Grass cell wall and lignocellulosic materials

In order to make biofuel from renewable materials independent of food resources such as grains, it is essential to focus on cellulose, hemicellulose and lignin, the main source of all three being plant cell walls. In contrast to animals, plants cannot run away from predators and pathogens, particularly microorganisms including bacteria and fungi. Plants also cannot move from one place to another in order to reach fresh sources of water. As a result, a major tool used by plants to protect themselves from microbial pathogens and to retain water inside cells involves production of a digestion-resistant network of biopolymers that confer rigidity and hydrophobicity on the exterior of cells. The biomass of plants consists primarily of their cell walls, which on average are composed of cellulose (30% to 50%), hemicellulose (20% to 40%) and lignin (15% to 30%) while 5% to 30% of cell wall might be made up of other components including proteins and minerals [11]. Cellulose is a polymer of β -(1-4) linked D-glucose units and is the main source of fermentable sugars in plant cell walls. In contrast, hemicellulose is made from 5- and 6-carbon sugars (primarily the 5-carbon sugars xylose and arabinose, termed arabinoxylans) with diverse chemical modifications. Hemicellulose can also be considered as a source of potentially fermentable sugars. Lignin is the naturally occurring polymer of a diverse set of phenolic units that by cross-linking to hemicellulose and cellulose fibrils confer rigidity and hydrophobicity. These together make the majority of cell wall structure and a schematic view of them is presented in Figure 1.2 [12].

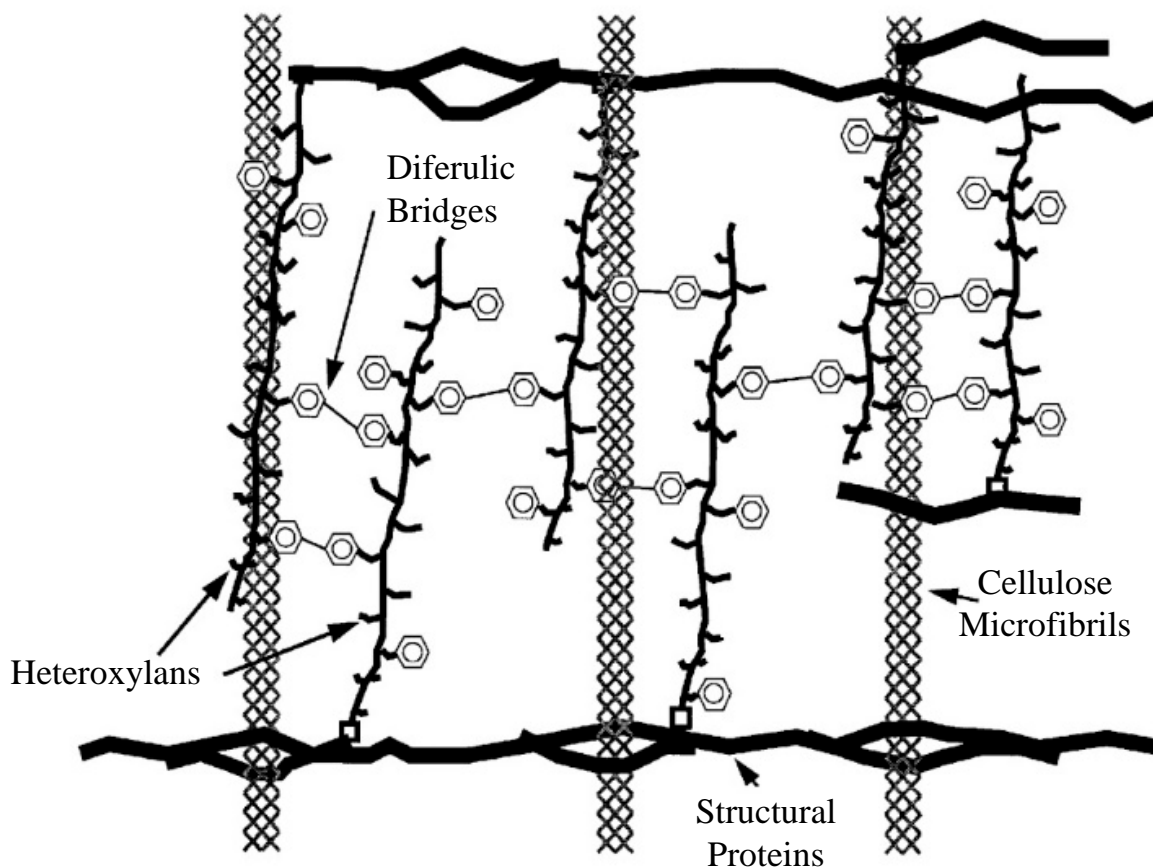


Figure 1.2. Schematic view of plant cell wall components.

Cellulose and hemicellulose are converted to ethanol after hydrolysis to monomeric sugars and fermentation [13], just as is the case for starch and food-sources of ethanol. With current technologies, the cellulose component of biomass is converted (treatment) to bioethanol quantitatively and almost completely, but conversions of hemicellulose to fermentable sugars are often less efficient. Lignocellulosic biomass also contains non-carbohydrate groups, also known as lignin, the presence of which represents a substantial fraction of biomass carbon. Lignin slows conversion of cell wall carbohydrate to fermentable sugars, ultimately decreasing conversion efficiency either by inhibiting hydrolase activity or fermentation [14]. Lignin also provides physical barriers and reduces accessibility of cell wall sugar units to hydrolytic

enzymes, as a result a thermochemical pretreatment before hydrolysis/fermentation is required to make glycopolymers more accessible to enzymes [15]. Figure 1.3 illustrates the process of pretreatment. Pretreatment involves both physical and chemical transformations [15]. One method of choice for pretreatment is ammonium fiber expansion (AFEX) which is more promising than other methods such as acid or other base treatments due to the efficiency of recycling ammonia. AFEX pretreatment has shown to improve efficiency of ethanol production from lignocellulosic biomass [16].

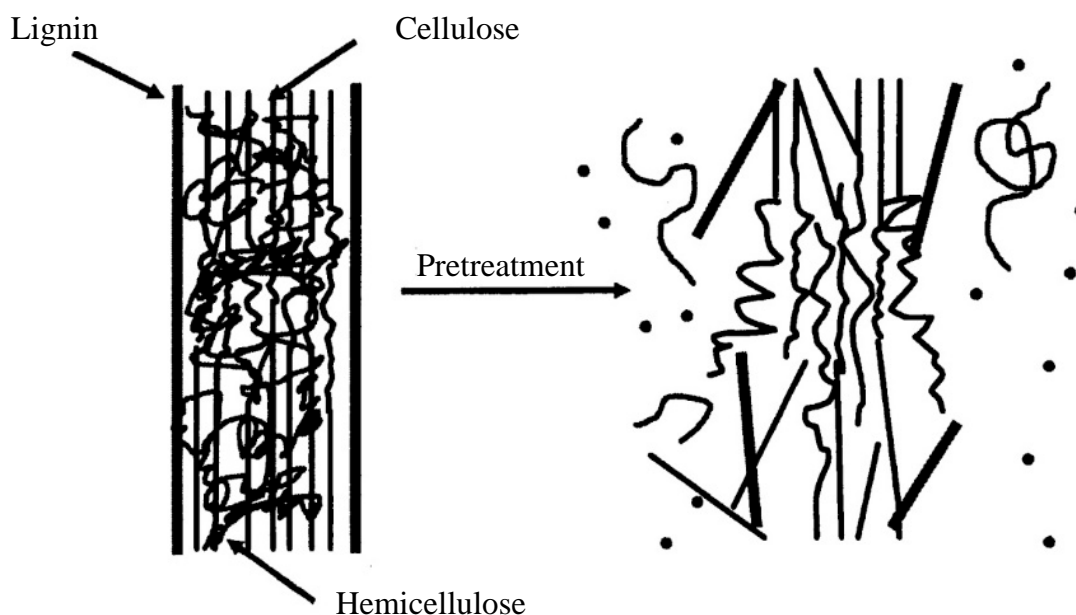


Figure 1.3. Schematic demonstration of how pretreatment makes polysaccharides more accessible

1.5 Lignin.

Lignin is the phenolic polymer that builds up in plants' secondary cell walls. It is known to be responsible for generation of wood and woody materials in trees and grasses. The conventional wisdom has been that lignin is formed by radical propagation reactions of three monomers called monolignols (or hydroxycinnamyl alcohol monomers): *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, shown in Figure 1.4 [17]. It is believed that lignin can

form large super-molecular structures that all together as a phenolic macromolecule confers rigidity and hydrophobicity to plant cell walls. Besides these monomers there are other phenolic compounds that have been characterized as lignin monomers that are derived from or are similar to these main monomers. Examples are the corresponding carboxylic acids (*p*-coumaric acid, ferulic acid, and sinapic acid), which are oxidized forms of the monolignols, and caffeic acid (3,4-dihydroxycinnamic acid) [18].

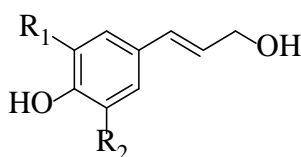


Figure 1.4. Structure of major lignin monomers. $R_1=R_2=H$: *p*-coumaryl alcohol, $R_1=H$, $R_2=OCH_3$: coniferyl alcohol, and $R_1=R_2=OCH_3$: sinapyl alcohol

Monolignol oxidation leads to delocalized radicals that can couple to each other through a variety of positions. This chemistry provides irregularity and diversity of structures in plant cell walls that makes lignin resistant to digestion by microorganisms. Radical coupling is often accompanied by ring closure to form 5-membered ring furans (phenyl coumarans) or 2 fused 5-membered rings, (pinoresinols), which adds yet more structural diversity to lignin. Delocalization of radical electrons on a representative monolignol, coniferyl alcohol, and some examples of lignol couplings are shown in Figure 1.5.

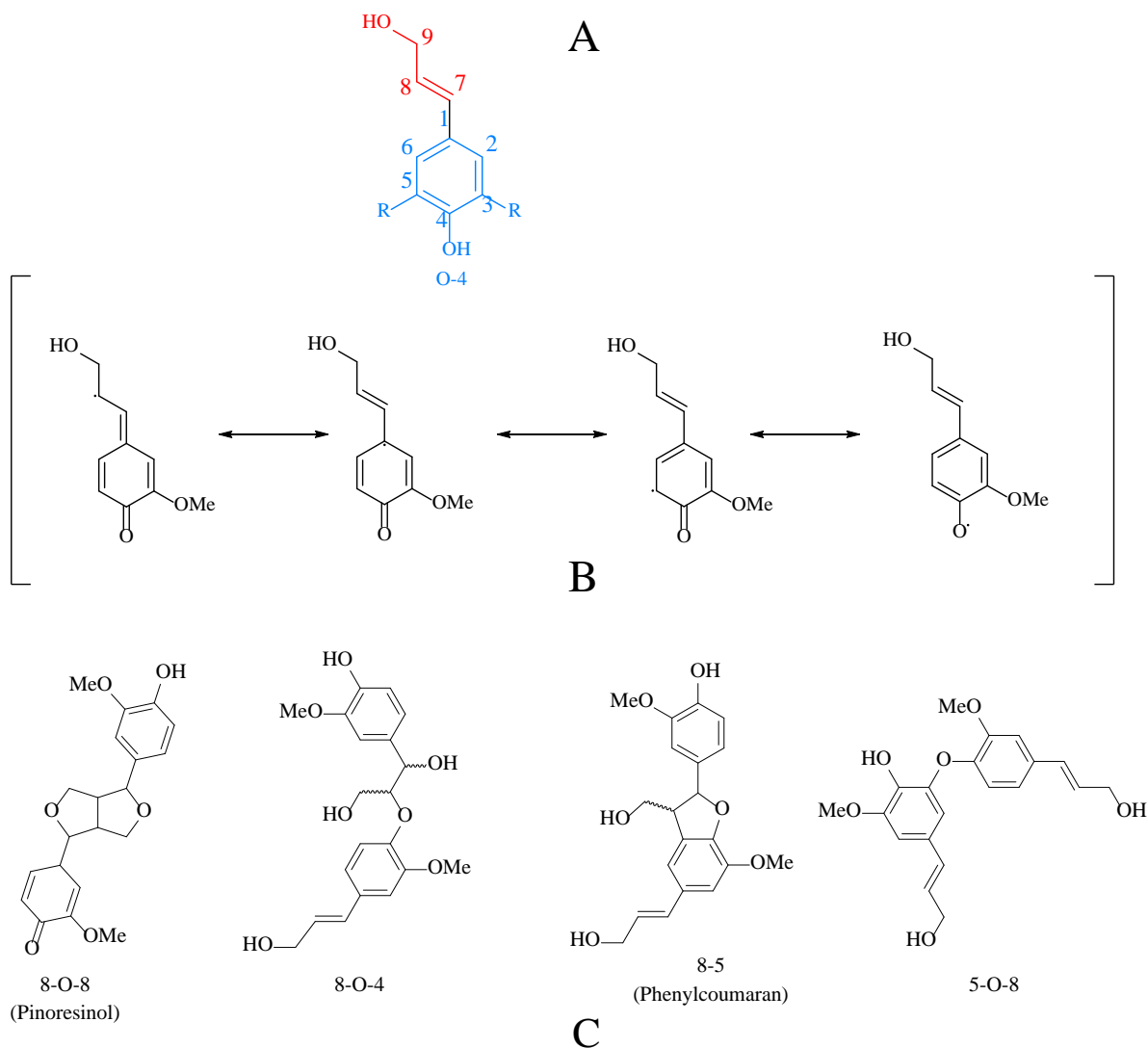


Figure 1.5. Panel A: Numbering system used for monolignols, Panel B: delocalization of radical at diverse positions in an oxidized monolignol, Panel C: examples of the different linkages that could form from coupling via different locations of the unpaired electron.

While radical coupling of monomers theoretically is favored on certain positions on aromatic or aliphatic moieties of monolignols, observation of theoretically-unfavored products which never form by *in vitro* radical polymerization suggests coupling of dimers preformed by other mechanisms is also an essential step in lignin formation [18].

Regardless of the modifications on the aliphatic (alcohol vs. carboxylic acid vs. aldehyde) the number of methoxy groups on the aromatic moiety changes the number of possible cross-links that lead to formation of rings, hence affecting the physical properties of the wood [18]. As

a result another categorization system for phenolic groups within lignin is based on their aromatic parts. Those units containing only one hydroxyl on the aromatic ring (such as *p*-coumaryl alcohol or *p*-coumaric acid) are called H units which is short for *p*-hydroxyphenyl, those with one methoxy group added on the aromatic ring are called G units which is short for guaiacyl, and those with two methoxy groups on the phenyl ring are called S units which is short for syringyl. That categorization is shown in Figure 1.6 [17 , 19]

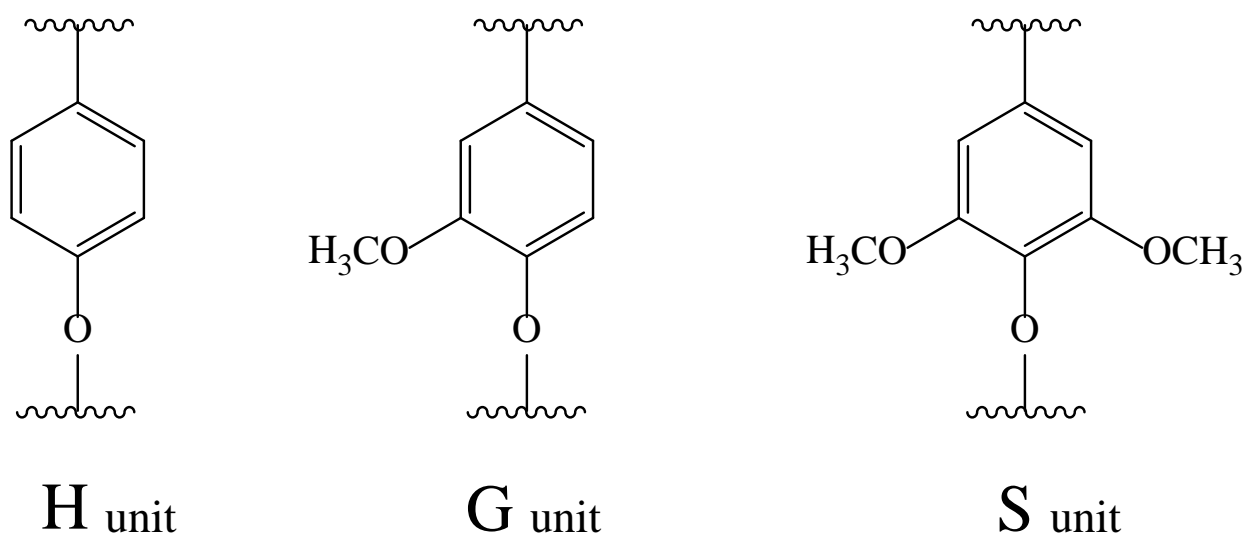


Figure 1.6. Classification of lignin units based on the number and position of their aromatic methoxy groups.

Ratios of these substructures (H:G:S) have been well characterized for different sub-categories in the plant kingdom. For example lignin in angiosperms (hardwoods) primarily consists of G and S units, gymnosperm (softwood) lignin consists of more G units and lesser amounts of H, and grasses (monocots) have roughly equal proportion of G:S units and more H units than dicots (trees) [20].

1.6 Pretreatment and effect of its products on ethanol production.

As mentioned before, in order to make the polysaccharide portion of biomass more accessible for enzymatic digestion and fermentation, a pretreatment process is usually utilized.

Among these methods are treatments with heat, acid, and bases [15 , 21] (add ref Kumar 2009). Perhaps the oldest used pretreatment process on biomass is use of calcium salts including limestone and calcium sulfite in paper pulping industry [22]. However, classic biomass pretreatment reagents including calcium sulfite or sulfuric acid [23] are too expensive for their use in generation of sustainable and economically-competitive biofuels [24]. Newer pretreatment methods have approached ethanol production with the goal of sustainability and recyclability of the reagents, and have shown considerable advantages toward production of ethanol. As examples, ammonia fiber expansion (AFEX) [25] and treatment with acidified γ -valerolactone (GVL) [26] have been proposed as promising options.

Pretreatment processes are often accompanied by partial degradation of cell wall components leading to formation of compounds that are not originally a part of the plant cell wall [15 , 27]. Degradation products of some widely-used pretreatment methods inhibit conversion of biomass polysaccharides to monosaccharides and/or their fermentation to bioethanol [28-29]. In order to select or engineer improved biomass resources with the best ethanol production potential while minimizing formation of degradation products, it is crucial to first characterize the degradation products of pretreatment of different biomass sources using different pretreatment methods [30]. A comprehensive collection of degradation products of different biomass feedstocks using different pretreatment methods have been collected by Ramin Vismeh of Michigan State University [31] and these are shown in Table 1.1.

From an examination of these biomass degradation products, one may conclude that many are common across different plants, or at least the class of compounds is shared among degradation products of different plants. Similarities of the aromatic products, which are presumed to be derived from lignin degradation, call for a more comprehensive and more

detailed approach to analysis of lignin from these biomass sources. Before this, it is first useful to consider what methodologies have been used for analysis of lignin and its degradation products in prior reports.

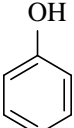
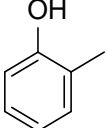
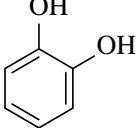
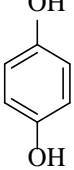
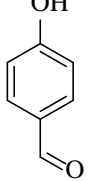
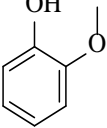
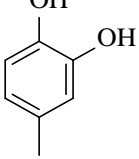
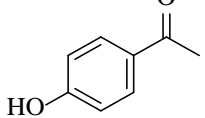
Compound	Structure	M.W. (Da)	Biomass Source
Phenol		94	Wheat straw
2-Methylphenol (cresol)		108	Willow
1,2 Benzenediol (catechol)		110	Willow / Spruce
Hydroquinone		110	Spruce
4-Hydroxybenzaldehyde		122	Wheat straw / Willow
2-Methoxyphenol (guaiacol)		124	Wheat straw / Willow
4-methyl-benzene-1,2- diol		124	Willow
<i>p</i> -hydroxyacetophenone		136	Wheat straw

Table 1.1 List of compounds produced from different biomass sources upon pretreatment.

Table 1.1 (cont'd)

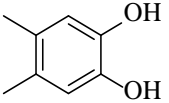
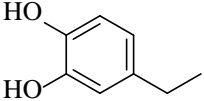
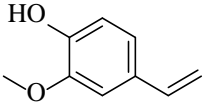
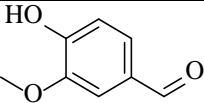
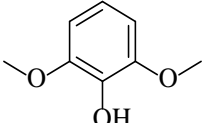
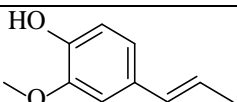
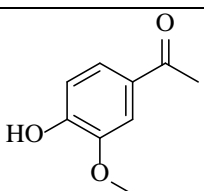
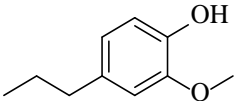
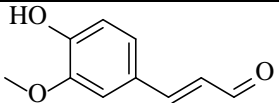
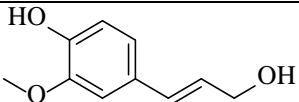
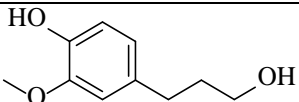
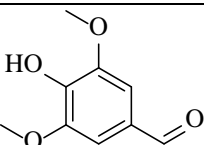
4,5 dimethyl benzene 1,2-diol		138	Willow
4-ethyl catechol		138	Willow
2-Methoxy-4-vinylphenol		150	Corn Stover
Vanillin		152	Wheat straw / sawdust feedstock/spruce / poplar / corn stover
2,6-Dimethoxyphenol (syringol)		154	Wheat straw / corn Stover
2-Methoxy-4-propenylphenol		164	Willow
acetoguaiacone		166	Wheat straw / spruce
2-methoxy-4-propylphenol		166	Willow
Coniferyl aldehyde		178	Spruce
Coniferyl alcohol		180	
Dihydroconiferyl alcohol		182	red oak wood / spruce
Syringaldehyde		182	Wheat straw / sawdust feedstock / poplar

Table 1.1 (cont'd)

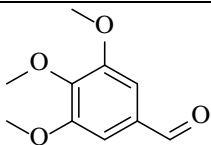
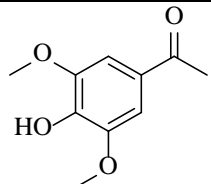
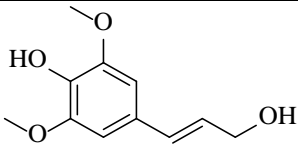
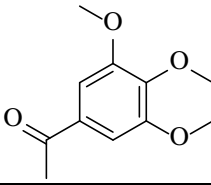
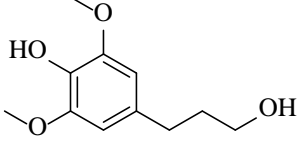
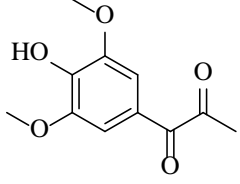
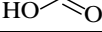
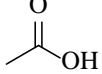
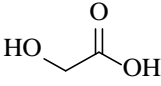
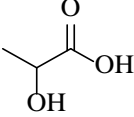
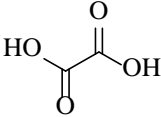
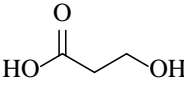
3,4,5-Trimethoxybenzaldehyde		196	Wheat straw
Acetosyringone		196	Wheat straw
Sinapyl alcohol		210	
3,4,5-Trimethoxyacetophenone /		210	Wheat straw
Dihydrosinapyl alcohol		212	Red oak wood
Syringoyl methyl ketone		224	Red oak wood
Formic acid		46	Wheat straw
Acetic acid		60	Red oak / poplar/wheat straw / corn stover
Hydroxyacetic (glycolic) acid		76	
Lactic acid		90	
Oxalic acid		90	
3-Hydroxypropanoic acid		90	Poplar

Table 1.1 (cont'd)

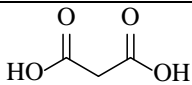
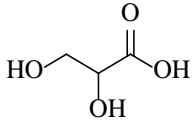
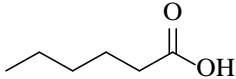
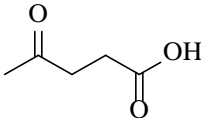
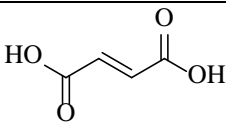
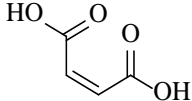
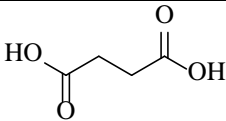
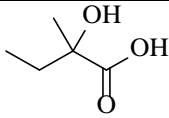
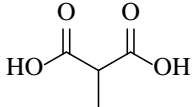
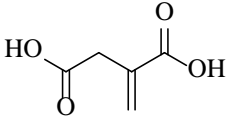
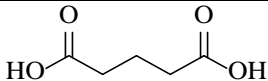
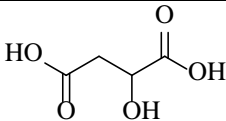
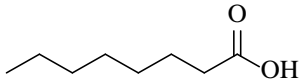
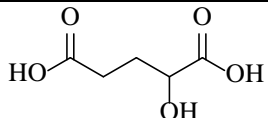
Propanedioic (malonic) acid		104	Corn stover
2,3-dihydroxypropanoic Acid		106	Poplar
caproic (hexanoic) acid		116	Red oak
4-Oxopentanoic (levulinic) acid		116	Poplar
(E)-Butenedioic (fumaric) acid		116	Corn stover
(Z)-Butenedioic (maleic) acid		116	Corn stover
Succinic acid		118	Poplar
2-Methyl-2-hydroxybutanoic acid		118	Poplar / corn stover
Methylpropanedioic (methyl malonic) acid		118	Poplar / corn stover
Methylidenebutanedioic acid (itaconic) acid		130	Corn Stover
Pentanedioic (glutaric) acid		132	Poplar
Malic acid		134	poplar
Caprylic (octanoic) acid		144	Red oak wood
2-Hydroxypentanedioic Acid		148	Poplar

Table 1.1 (cont'd)

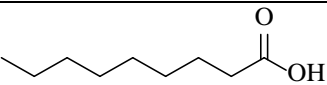
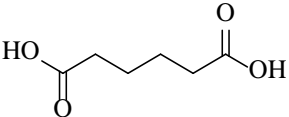
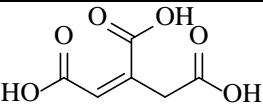
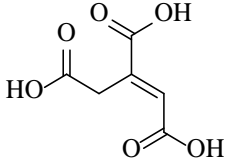
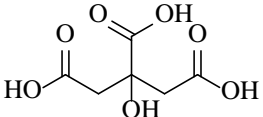
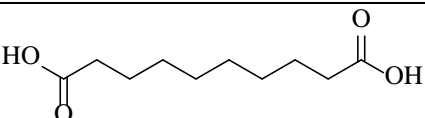
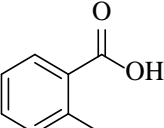
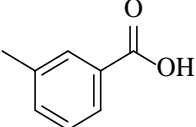
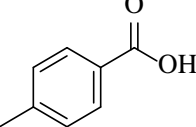
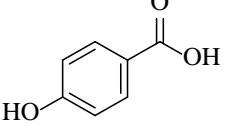
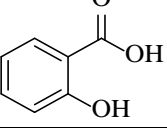
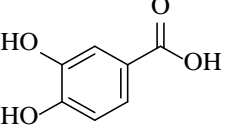
Pelargonic (nonanoic) acid		158	Red oak wood
Hexanedioic (adipic) acid		146	Poplar / corn stover
<i>cis</i> -Aconitic acid		174	Corn Stover
<i>trans</i> -Aconitic acid		174	Corn stover
Citric acid		192	
1,8-Octanedicarboxylic acid		202	Poplar
<i>o</i> -Toluic acid		136	
<i>m</i> -Toluic acid		136	
<i>p</i> -Toluic acid		136	
4-Hydroxybenzoic acid		138	Willow / spruce / wheat straw / poplar / corn stover
2-Hydroxybenzoic (salicylic) acid		138	Corn stover
3,4-Dihydroxybenzoic (protocatechuic) acid		154	Sawdust feedstock / willow/poplar

Table 1.1 (cont'd)

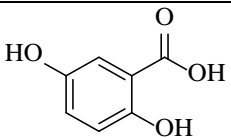
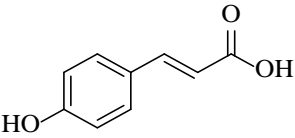
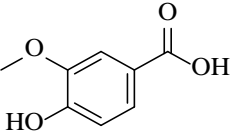
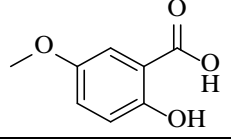
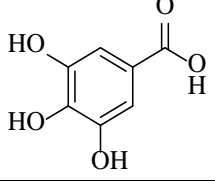
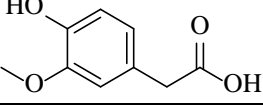
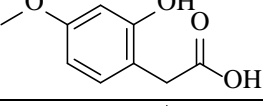
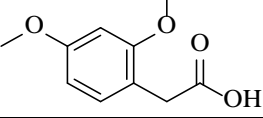
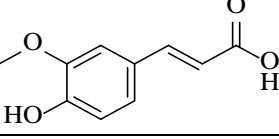
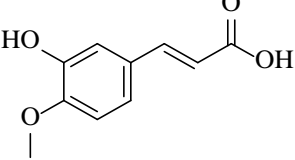
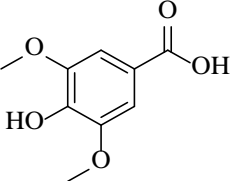
2,5-Dihydroxybenzoic acid (gentisic) acid		154	Willow / poplar
<i>p</i> -Coumaric acid		164	Corn stover
3-Hydroxy-4-methoxybenzoic acid (vanillic) acid		168	Willow / spruce / wheat straw / poplar
2-Hydroxy-5-methoxybenzoic acid		168	Poplar
Gallic acid		170	Sawdust feedstock / corn Stover
2-(4-Hydroxy-3-methoxyphenyl)-acetic acid (homovanillic acid)		182	Spruce / poplar
2-(2-hydroxy-4-methoxyphenyl)acetic acid		182	Poplar
2-(2,4-dimethoxyphenyl)acetic acid		182	Poplar
ferulic acid		194	Poplar/corn stover
(<i>E</i>)-3-(3-hydroxy-4-methoxyphenyl)acrylic acid		194	Poplar
4-Hydroxy-3,5-dimethoxybenzoic acid (syringic) acid		198	Willow /wheat straw /poplar

Table 1.1 (cont'd)

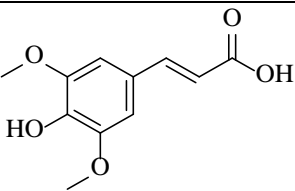
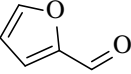
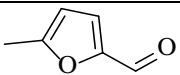
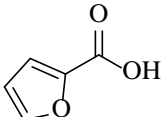
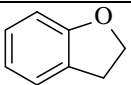
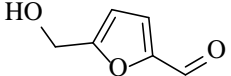
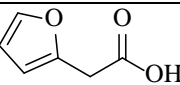
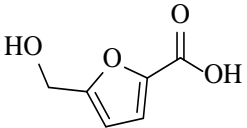
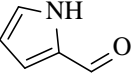
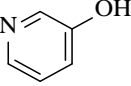
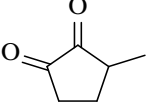
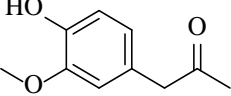
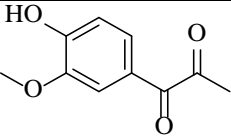
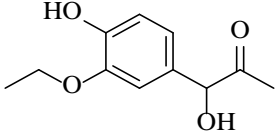
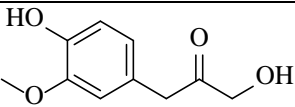
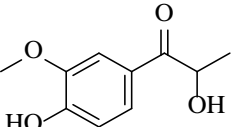
Sinapic acid		224	
furfural		96	Poplar / sawdust feedstock / wheat straw
5-Methyl-2-furfural		110	Corn stover
furoic acid		112	Wheat straw
2,3-Dihydrobenzofuran		120	Corn stover
5-hydroxymethylfurfural		126	sawdust feedstock / corn stover
2-Furanacetic acid		126	poplar
5-(hydroxymethyl)furan-2-carboxylic acid		142	poplar
Pyrrole-2-carboxaldehyde		95	Corn stover
3-Hydroxypyridine		95	Corn stover
3-Methyl-1,2-cyclopentanedione		112	Corn stover
1-(4-Hydroxy-3-methoxyphenyl)propan-2-one		180	Spruce
1-(4-Hydroxy-3-methoxyphenyl)propane-1,2-dione		194	Spruce
1-Hydroxy-1-(4-hydroxy-3-ethoxyphenyl)propan-2-one		196	Spruce / poplar

Table 1.1 (cont'd)

1-Hydroxy-3-(4-hydroxy-3-methoxyphenyl)propan-2-one		196	Spruce
2-Hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-1-one		196	Spruce / poplar

1.7 Analysis of lignin and its degradation products.

Due to the random nature of lignin and large molecular size that this random polymer can reach, investigation of lignin relied upon breaking it down to small molecules, particularly those small enough to be volatilized for analysis by gas chromatography. The usually entails conversion to molecules with 0-2 benzene rings which here for the purpose of ease of use are called non-lignols, monolignols, and dilignols respectively. Classic methods of analysis of lignin included, since the 1950's and 60's, infrared and NMR spectroscopies, followed by UV/visible and gas chromatography, both of which are often performed after derivatization [32-39].

Analysis of intact un-derivatized lignin molecules have since been characterized using NMR [40] and different mass spectrometry techniques [41-42] or both methods [43], sometimes after fractionation by size exclusion chromatography [44]. Among these methods, the coupling of ultrahigh performance liquid chromatography (UHPLC) followed by high resolution MS and MS/MS provides perhaps the newest approach for qualitative and quantitative lignin analysis. For example, Morreel et al. [42 , 45] have developed a sequencing approach based on UHPLC/MS/MS analysis of synthesized model compounds and subsequent analysis of solvent extracted lignin from poplar and grasses.

The most extensive and comprehensive investigations of lignin structure using NMR have been performed during the past two decades by the research group of Professor John Ralph,

now of the University of Wisconsin. Their findings revealed groundbreaking information regarding the type of linkages as well as ratio of different lignin units in different hardwood, softwood, and grass species [46-47]. In addition their research has led to elucidation of major pathways of incorporation of lignin monomers including *p*-coumaric acid into larger molecules [48] and more recently showed the incorporation of flavone triclin in lignin from monocot grasses [49]. One of the mostly used lignin derivatization methods of lignin analysis, derivatization followed by reductive cleavage (DFRC) has also been developed by the Ralph research group [36 , 50]. This method uses acetyl bromide to brominate the benzylic position (carbon 7) of lignin units, then applies zinc to reduce and cleave lignin units followed by acetylation to convert the products to more volatile forms that can be analyzed using GC-MS. This method has resulted in quantitative measurement of H, G, and S units in lignin and their ratios in different biomass sources. A schematic of DFRC is shown in figure 1.7.

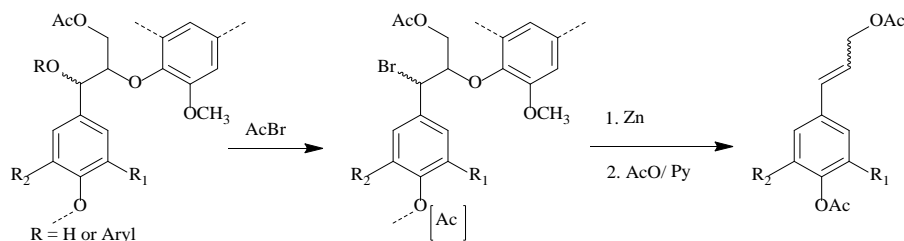


Figure 1.7. Schematic view of derivatization of lignin followed by reductive cleavage [50].

Another successful method for derivatization and analysis of lignin is based on forming thioesters of all three aliphatic carbons of lignin units via nucleophilic displacement of alkyl aryl ether bonds with a thiol. Derivatized units are then subjected to GC-MS analysis [51] through a process called thioacidolysis [52]. Prior to development of DFRC, thioacidolysis was the main GC-MS method used for elucidation of lignin monomeric units and their ratios. However, it has been shown that in some cases DFRC can reveal certain linkages that thioacidolysis is unable to cleave and measure, such as 8-O- ether linkages [53].

Despite advancements in analysis of intact lignin molecules, recent reports have demonstrated that widely-used analysis methods often only elucidate ratios of different phenylpropanoid monomeric unit ratios and linkage positions, and are often unsuccessful in discovering novel and unknown lignin units [49], perhaps because these building blocks lack volatility needed for GC separation. For instance, signals in NMR spectra of lignin can be misinterpreted as other structurally-similar lignin units, as shown by Banoub et al. [54].

In addition, limited concise information is available about how the size or molecular mass of intact lignin molecules in each biomass source. The main efforts to establish lignin molecular size have relied either on mass spectra that detect intact molecular ions [54] and/or on separation methods, most notably size exclusion chromatography [55]. While MS methods like matrix assisted laser desorption ionization (MALDI) show limited sensitivity due to low abundance of each individual large molecule, SEC cannot reveal as much chemical details as MS can, and results may be inaccurate when intermolecular forces drive significant noncovalent associations. While size exclusion chromatography has shown average lignin molecular masses up to 40 kDa, MALDI-MS, which usually breaks noncovalent associations, has only shown molecular ions for lignin in the range of 1-3 kDa [56-57].

In Chapter 2 of this dissertation, a new mass spectrometry approach has been used that expands the depth of analysis of large lignin molecules from grasses, and contrasts the findings with lignin from the hardwood tree poplar. This strategy has been applied to probe chemical diversity of lignin molecules and assess similarity of different grass biomass sources (corn stover, wheat straw, rice straw, *Miscanthus*, sorghum, and switchgrass) with emphasis on chemistry that incorporates the flavonoid tricetin into grass lignin.

Chapter 3 of this dissertation explores grass lignin in yet more depth with the theme of discovering new common crosslinkers in grass lignin and demonstrating presence of the same new linkers in large lignin molecules. In this chapter, results are presented that document high levels of incorporation of phenolic acid esters of glycerol, particularly esters of *p*-coumaric and ferulic acids, in grass lignin molecules via chemistry similar to incorporation of conventional monolignols.

Chapter 4 of this dissertation describes another approach to extract and analyze large lignin molecules utilizing the new lignin solubilization method employed by Luterbacher et al [26] that dissolves biomass quantitatively in the powerful solvent GVL in the presence of acid. In this chapter a new method for UHPLC separation of lignin material is demonstrated with use of γ -valerolactone (GVL) as the mobile phase for UHPLC followed by electrospray ionization (ESI) MS.

Chapter 5 reviews all the findings in this dissertation to make a more comprehensive picture of grass lignin with the perspective of explaining their fundamental differences from hardwood lignin. The goal in this chapter is to summarize the properties of grass lignin that remain unknown with the hope that these newly described properties can be used in bio-engineering of more digestible grass biomass which is the main side product of food production and also grown in marginal lands.

REFERENCES

REFERENCES

- (1.) Dijkstra, D. J.; Langstein, G., *Polymer International* **2012**, 61 (1), 6-8.
- (2.) Azar, C.; Lindgren, K.; Larson, E.; Mollersten, K., *Climatic Change* **2006**, 74 (1-3), 47-79.
- (3.) Limayem, A.; Ricke, S. C., *Progress in Energy and Combustion Science* **2012**, 38 (4), 449-467.
- (4.) Kim, S.; Dale, B. E., *Biomass & Bioenergy* **2004**, 26 (4), 361-375.
- (5.) Jambo, S. A.; Abdulla, R.; Azhar, S. H. M.; Marbawi, H.; Gansau, J. A.; Ravindra, P., *Renewable & Sustainable Energy Reviews* **2016**, 65, 756-769.
- (6.) Sanderson, K., *Nature* **2006**, 444 (7120), 673-676.
- (7.) Renewable Fuels, A., *Accelerating industry innovation : 2012 ethanol industry outlook*. Renewable Fuels Association: Washington, D.C., 2012.
- (8.) Gelles, D. The standoff between big oil and big corn. <http://www.nytimes.com/2016/09/18/business/energy-environment/the-standoff-between-big-oil-and-big-corn.html>.
- (9.) Johnson, J., *Chemical & Engineering News* **2013**, 91 (5), 20-20.
- (10.) Agency, E. P. Proposed Renewable Fuel Standards for 2017, and the Biomass-Based Diesel Volume for 2018. EPA-HQ-OAR-2016-0004. . <https://www.epa.gov/renewable-fuel-standard-program/proposed-renewable-fuel-standards-2017-and-biomass-based-diesel>.
- (11.) Pauly, M.; Keegstra, K., *Plant Journal* **2008**, 54 (4), 559-568.
- (12.) Saha, B. C., *Journal of Industrial Microbiology & Biotechnology* **2003**, 30 (5), 279-291.
- (13.) Gray, K. A.; Zhao, L. S.; Emptage, M., *Current Opinion in Chemical Biology* **2006**, 10 (2), 141-146.
- (14.) Sun, Y.; Cheng, J. Y., *Bioresource Technology* **2002**, 83 (1), 1-11.
- (15.) Mosier, N.; Wyman, C.; Dale, B.; Elander, R.; Lee, Y. Y.; Holtzapple, M.; Ladisch, M., *Bioresource Technology* **2005**, 96 (6), 673-686.
- (16.) Holtzapple, M. T.; Lundeen, J. E.; Sturgis, R.; Lewis, J. E.; Dale, B. E., *Applied Biochemistry and Biotechnology* **1992**, 34-5, 5-21.

- (17.) Freudenberg, K., *Nature* **1959**, 183 (4669), 1152-1155.
- (18.) Boerjan, W.; Ralph, J.; Baucher, M., *Annual Review of Plant Biology* **2003**, 54, 519-546.
- (19.) Freudenberg, K., *Science* **1965**, 148 (3670), 595.
- (20.) Baucher, M.; Monties, B.; Van Montagu, M.; Boerjan, W., *Critical Reviews in Plant Sciences* **1998**, 17 (2), 125-197.
- (21.) Hendriks, A.; Zeeman, G., *Bioresource Technology* **2009**, 100 (1), 10-18.
- (22.) Springer, E. L.; McSweeney, J. D., *Tappi Journal* **1986**, 69 (4), 129-130.
- (23.) Karagoz, S.; Tay, T.; Ucar, S.; Erdem, M., *Bioresource Technology* **2008**, 99 (14), 6214-6222.
- (24.) Bals, B.; Wedding, C.; Balan, V.; Sendich, E.; Dale, B., *Bioresource Technology* **2011**, 102 (2), 1277-1283.
- (25.) Bals, B.; Dale, B.; Balan, V., *Energy & Fuels* **2006**, 20 (6), 2732-2736.
- (26.) Luterbacher, J. S.; Rand, J. M.; Alonso, D. M.; Han, J.; Youngquist, J. T.; Maravelias, C. T.; Pfleger, B. F.; Dumesic, J. A., *Science* **2014**, 343 (6168), 277-280.
- (27.) Palmqvist, E.; Hahn-Hagerdal, B., *Bioresource Technology* **2000**, 74 (1), 25-33.
- (28.) Klinke, H. B.; Ahring, B. K.; Schmidt, A. S.; Thomsen, A. B., *Bioresource Technology* **2002**, 82 (1), 15-26.
- (29.) Klinke, H. B.; Thomsen, A. B.; Ahring, B. K., *Applied Microbiology and Biotechnology* **2004**, 66 (1), 10-26.
- (30.) Chundawat, S. P. S.; Vismeh, R.; Sharma, L. N.; Humpala, J. F.; Sousa, L. D.; Chambliss, C. K.; Jones, A. D.; Balan, V.; Dale, B. E., *Bioresource Technology* **2010**, 101 (21), 8429-8438.
- (31.) Vismeh, R. Multifaceted Metabolomics Approaches for Characterization of Lignocellulosic Biomass Degradation Products formed during Ammonia Fiber Expansion Pretreatment. Doctoral Dissertation, Michigan State University, East Lansing, MI, USA, 2012.
- (32.) Bolker, H. I., *Nature* **1963**, 197 (486), 489.
- (33.) Gagnaire, D.; Robert, D., *Bulletin De La Societe Chimique De France* **1968**, (2), 781.

- (34.) Gagnaire, D.; Robert, D.; Vignon, M.; Vottero, P., *European Polymer Journal* **1971**, 7 (7), 965.
- (35.) Kristersson, P.; Lundquist, K.; Strand, A., *Wood Science and Technology* **1980**, 14 (4), 297-300.
- (36.) Lu, F.; Ralph, J., *Abstracts of Papers of the American Chemical Society* **1996**, 211, 110.
- (37.) Polcin, J.; Rapson, W. H., *Pulp and Paper Magazine of Canada* **1969**, 70 (24), 99.
- (38.) Lin, S. Y., *Svensk Papperstidning-Nordisk Cellulosa* **1982**, 85 (18), R162-R171.
- (39.) Ghaffar, S. H.; Fan, M. Z., *Biomass & Bioenergy* **2013**, 57, 264-279.
- (40.) Gerasimowicz, W. V.; Hicks, K. B.; Pfeffer, P. E., *Macromolecules* **1984**, 17 (12), 2597-2603.
- (41.) Tokareva, E. N.; Pranovich, A. V.; Holmbom, B. R., *Wood Science and Technology* **2011**, 45 (4), 767-785.
- (42.) Morreel, K.; Dima, O.; Kim, H.; Lu, F. C.; Niculaes, C.; Vanholme, R.; Dauwe, R.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., *Plant Physiology* **2010**, 153 (4), 1464-1478.
- (43.) Navarrete, P.; Pizzi, A.; Pasch, H.; Delmotte, L., *Journal of Adhesion Science and Technology* **2012**, 26 (8-9), 1069-1082.
- (44.) Jacobs, A.; Dahlman, O., *Nordic Pulp & Paper Research Journal* **2000**, 15 (2), 120-127.
- (45.) Morreel, K.; Kim, H.; Lu, F. C.; Dima, O.; Akiyama, T.; Vanholme, R.; Niculaes, C.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., *Analytical Chemistry* **2010**, 82 (19), 8095-8105.
- (46.) Ralph, J.; Helm, R. F., *Journal of Agricultural and Food Chemistry* **1991**, 39 (4), 705-709.
- (47.) Ralph, J., *Magnetic Resonance in Chemistry* **1993**, 31 (4), 357-363.
- (48.) Ralph, J.; Hatfield, R. D.; Quideau, S.; Helm, R. F.; Grabber, J. H.; Jung, H. J. G., *Journal of the American Chemical Society* **1994**, 116 (21), 9448-9456.
- (49.) del Rio, J. C.; Prinsen, P.; Rencoret, J.; Nieto, L.; Jimenez-Barbero, J.; Ralph, J.; Martinez, A. T.; Gutierrez, A., *Journal of Agricultural and Food Chemistry* **2012**, 60 (14), 3619-3634.

- (50.) Lu, F. C.; Ralph, J., *Journal of Agricultural and Food Chemistry* **1997**, 45 (7), 2590-2592.
- (51.) Lapierre, C.; Monties, B.; Rolando, C., *Journal of Wood Chemistry and Technology* **1985**, 5 (2), 277-292.
- (52.) Rolando, C.; Monties, B.; Lapierre, C., Thioacidolysis. In *Methods in Lignin Chemistry*, Lin, S. Y.; Dence, C. W., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1992; pp 334-349.
- (53.) Grabber, J. H.; Quideau, S.; Ralph, J., *Phytochemistry* **1996**, 43 (6), 1189-1194.
- (54.) Banoub, J.; Delmas, G. H.; Joly, N.; Mackenzie, G.; Cachet, N.; Benjelloun-Mlayah, B.; Delmas, M., *Journal of Mass Spectrometry* **2015**, 50 (1), 5-48.
- (55.) Baumberger, S.; Abaecherli, A.; Fasching, M.; Gellerstedt, G.; Gosselink, R.; Hortling, B.; Li, J.; Saake, B.; de Jong, E., *Holzforschung* **2007**, 61 (4), 459-468.
- (56.) Kosyakov, D. S.; Ul'yanovskii, N. V.; Sorokina, E. A.; Gorbova, N. S., *Journal of Analytical Chemistry* **2014**, 69 (14), 1344-1350.
- (57.) Bayerbach, R.; Nguyen, V. D.; Schurr, U.; Meier, D., *Journal of Analytical and Applied Pyrolysis* **2006**, 77 (2), 95-101.

Chapter Two: A metabolomic investigation of diversity of tricin incorporation and pretreatment transformations in lignin extractives of grasses

2.1 Abstract

Chemical transformations that lead to deconstruction of lignin have roles with great potential importance to efforts to exploit lignin as a source of renewable energy and chemicals. Recent reports have raised awareness that the flavonol triclin is an important constituent in lignin from monocots (grasses). In an effort to deepen probes into the chemistry of triclin in lignin, a metabolomics approach based on ultrahigh performance liquid chromatography/mass spectrometry (UHPLC/MS) and a non-selective multiplexed collision-induced dissociation (CID) method generated fragments of ionized molecules. These analyses generated evidence that the major triclin derivatives identified to date represent a minority of triclin derivatives in all grass extractives, with numerous triclin derivatives of low individual abundances contributing to mass balance considerations regarding triclin levels in lignin. Wide window MS/MS spectra provided evidence for triclin incorporation into molecules at least as large as 4 kDa, and yielded quantitative indications that proportions of building blocks of lignin vary with molecular mass. Ammonia fiber expansion (AFEX) and extractive AFEX (EA) pretreatments led to small decreases in levels of most individual triclin derivatives, most notably those containing labile acetate and *p*-coumarate esters. However, acetate esters of a triclin oligolignol did survive but only if the β -hydroxyl group was converted to an amino group. These findings suggest that the β -hydroxyl group serves as an important determinant of rates of ester group hydrolysis/ammonolysis during AFEX pretreatment.

2.2 Introduction

Conversion to lignocellulosic biomass to simple sugars and liquid fuels is hindered by the recalcitrance of plant cell walls to enzymatic conversions to fermentable sugars. Numerous pretreatment strategies have been developed to improve yields including processing with acids, alkali, oxidants, assorted solvents, and various combinations of these. One of these processes, ammonia fiber expansion (AFEX), drives physical and chemical changes that cleave crosslinkers from hemicelluloses and improve digestibility by glycolytic enzymes[1-2]. A newer version of this process, called extractive-AFEX (EA), uses liquid ammonia to cleave crosslinkers and separate substantial amounts of lignin and fermentation inhibitors from ammonia-insoluble cell wall glycopolymers[3]. A variety of approaches including EA, extraction with ionic liquids, and acid-treatment in the renewable solvent γ -valerolactone (GVL) show promise for solubilizing lignin and aiding separation from other cell wall degradation products including oligosaccharides[4-5].

Lignin makes up substantial amounts (~ 13-31% dry weight) of carbon in monocot grasses including corn stover, switchgrass, *Miscanthus*, sorghum, and rice and wheat straw that have been touted as renewable bioenergy crops[6-11]. Application of advanced technologies to convert lignin to valuable chemical precursors and fuels has potential to improve biorefinery economics, but detailed understanding of how lignin chemistry contributes to bioprocessing recalcitrance has remained elusive owing to its chemical complexity. Much understanding of the structure and chemistry of lignin has been based on conclusions that lignin consists of a random polymer derived from the monolignols *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol[12]. In addition, recent research

has demonstrated from NMR spectra that the flavonoid tricetin (5,7-dihydroxy-2-(4-hydroxy-3,5-dimethoxyphenyl)-4*H*-chromen-4-one) has substantial abundance in monocot lignins[**13-15**], and proposed that tricetin should be considered an important lignin monomer and nucleation site for lignin growth.

Though 2D NMR spectra, molecular mass based fractionation using size exclusion chromatography, and wet chemical degradation methods continue to provide vital information about the structural moieties in lignins, there remain unanswered questions regarding the diversity of substructure connectivity that can influence conversion of lignin to biofuels and renewable chemical feedstocks, particularly for larger lignin constituents. In this report we describe results of mass spectrometric profiling of extractives from several monocot species touted as lignocellulosic bioenergy crops. Our goal has been to develop and apply analytical approaches that can illuminate the diversity of monocot lignin chemistry in ways that are not evident from NMR and chemical degradation methods so that this knowledge might be used to deconstruct monocot lignins in biorefineries for their utilization as chemical feedstocks.

2.3 Experimental

Materials. HPLC grade methanol and HPLC grade hexanes were purchased from Sigma-Aldrich and JT Baker respectively. Corn stover (Pioneer 36H56) was harvested in September 2009 in Wisconsin (USA) and oven dried at 60 °C down to about 10% moisture on wet weight basis (WWB). Rice straw was obtained from California (USA) and air-dried to approximately 7.92% moisture content (WWB). The two above mentioned materials were further passed through a 5 mm screen installed in a Christy hammer mill (Christison Scientific LTD, England) and stored at 4 °C in heat-sealed bags

prior to utilization. *Miscanthus x giganteus*, harvested in spring 2005, was a generous gift from Professor Steven P. Long, University of Illinois, Urbana–Champaign. Cave-in-Rock (CIR) switchgrass (an upland variety) was grown at Michigan State University (East Lansing, MI, USA) and harvested in October 2008. Wheat straw, harvested in 2014, was a generous gift from MBI International, who purchased the air-dried material from a farm located in Webberville, MI. *Miscanthus*, switchgrass and wheat straw were milled using a JT-6 Homoloid mill from the Fitzpatrick Co. with a 3.175 mm diameter sieve, before storage at 4 °C in zip-lock bags. Whole forage sorghum was received from Florida (USA) and milled using a Wiley mill equipped with a 5-mm diameter sieve, prior to storage at 4 °C in zip-lock bags.

Each biomass feedstock was further milled through a 1 mm mesh using a Foss Cyclotec™ 1093 mill (Foss, Denmark). Accelerated solvent extraction (ASE) was performed using a Dionex ASE-200 system. Stainless steel ASE cartridges (33-mL) were packed with 5 g of milled biomass and extracted 3X with 28 mL of hexane to remove lipids, followed by 3X extraction with 28 mL of methanol. Each extraction was performed at 100° C and 1500 psi. Methanol extracts were then pooled and diluted 10X in methanol before analysis using UHPLC-MS. Crude EA extractives of each biomass were produced as previously described in the literature[16]. Tricin standard was purchased from ChromaDex (Irvine, CA).

UHPLC-MS analyses were performed using a Waters G2-S QToF mass spectrometer equipped with an Acquity pump, model 2777C autosampler, and Acquity Column Manager. Separations were performed using a fused core Supelco Ascentis Express C18 column (100 x 2.1 mm; 2.7 µm particles). Gradient elution was performed

using solvent A (0.1% aq. formic acid) and solvent B (methanol) at a total flow rate of 0.3 mL/min. Linear gradient conditions (A/B) were: initial, hold to 1.0 min (99/1), followed by increase in B to (1/99) at 30 min. Electrospray ionization was used for all analyses. Nonselective multiplexed CID mass spectra [17-18] were acquired in both positive and negative ion modes by quasi-simultaneous switching of the collision cell voltage through 5 different values (5, 25, 40, 55, and 80 V), accumulating transients for 0.1 seconds per function. Mass spectra were acquired in centroid peak mode for each function, and leucine-enkephalin was introduced as a lock spray reference, with automatic mass correction. UHPLC-MS data were processed using Waters MarkerLynx XS software, which performed extracted ion chromatogram peak detection, retention time alignment, and peak integration. Quantification of tricin flavonolignans was performed by analysis of tricin standard solutions prepared at 0.01, 0.05, 0.1, 0.5, 1.0, and 5.0 μ M concentrations, and linear calibration curve generation. All flavonolignan concentrations were calculated based on the assumption that their molar response factors were identical to tricin.

2.4 Results and Discussion

2.4.1 Untargeted metabolomic profiling of extracts of untreated biomass

Profiling of methanol extracts of an assortment of monocots and the hardwood poplar using ultrahigh performance liquid chromatography/time-of-flight mass spectrometry (UHPLC/TOF MS) yielded evidence of a diverse assortment of substances from each biomass source (Figure 2.1). The negative-ion mode used for spectrum acquisition yielded evidence of a variety of polyphenols that were annotated using accurate molecular and fragment mass measurements and relative mass defect calculations, which yield information about the fractional hydrogen content of molecules

[18-19]. Though the different biomass materials differ in chemical complexity, some compounds were observed in extracts of all grasses (Figure 2.1B-G) but not poplar, which yielded a chemical profile (Figure 2.1A) distinct from the monocots. Among the most abundant compounds in nearly all grasses, but not poplar, were derivatives of the flavonol tricetin ((5,7,4'-trihydroxy-3',5'-dimethoxyflavone, Table 2.1), previously recognized to be abundant as a structural component of wheat lignin[13], and was recently shown to be an important component of lignin in monocots[14-15]. These compounds include guaiacylglyceryltricetin (GGT; also abbreviated as T-(4-O- β)-G) [15] in which tricetin is conjugated to a monolignol derived from coniferyl alcohol (G monomer), often in two chromatographically-resolved diastereomeric forms as expected for *erythro*- and *threo*-isomers (Figure 2.1, peaks *b*₆, *b*₇, *c*₇, *c*₈, *d*₃, *d*₅, *e*₂, *e*₃, *f*₈, *f*₉, *g*₂, and *g*₃). NMR spectra of GGT isolated from corn stover matched ¹H and ¹³C spectral data for GGT isolated from the grass *Hyparrhenia hirta* [20]; (Table 2.2).

Among the investigated grasses, *Miscanthus* extracts yielded the greatest number of compounds uncommon to the other grasses. The most abundant compound common to all grasses was the flavonoid tricetin, detected in negative-ion mode as [M-H]⁻ at *m/z* 329.07. This compound's identity was confirmed from accurate mass measurement of this ion and product ions (fragments) generated using MS/MS, and coelution with authentic tricetin standard.

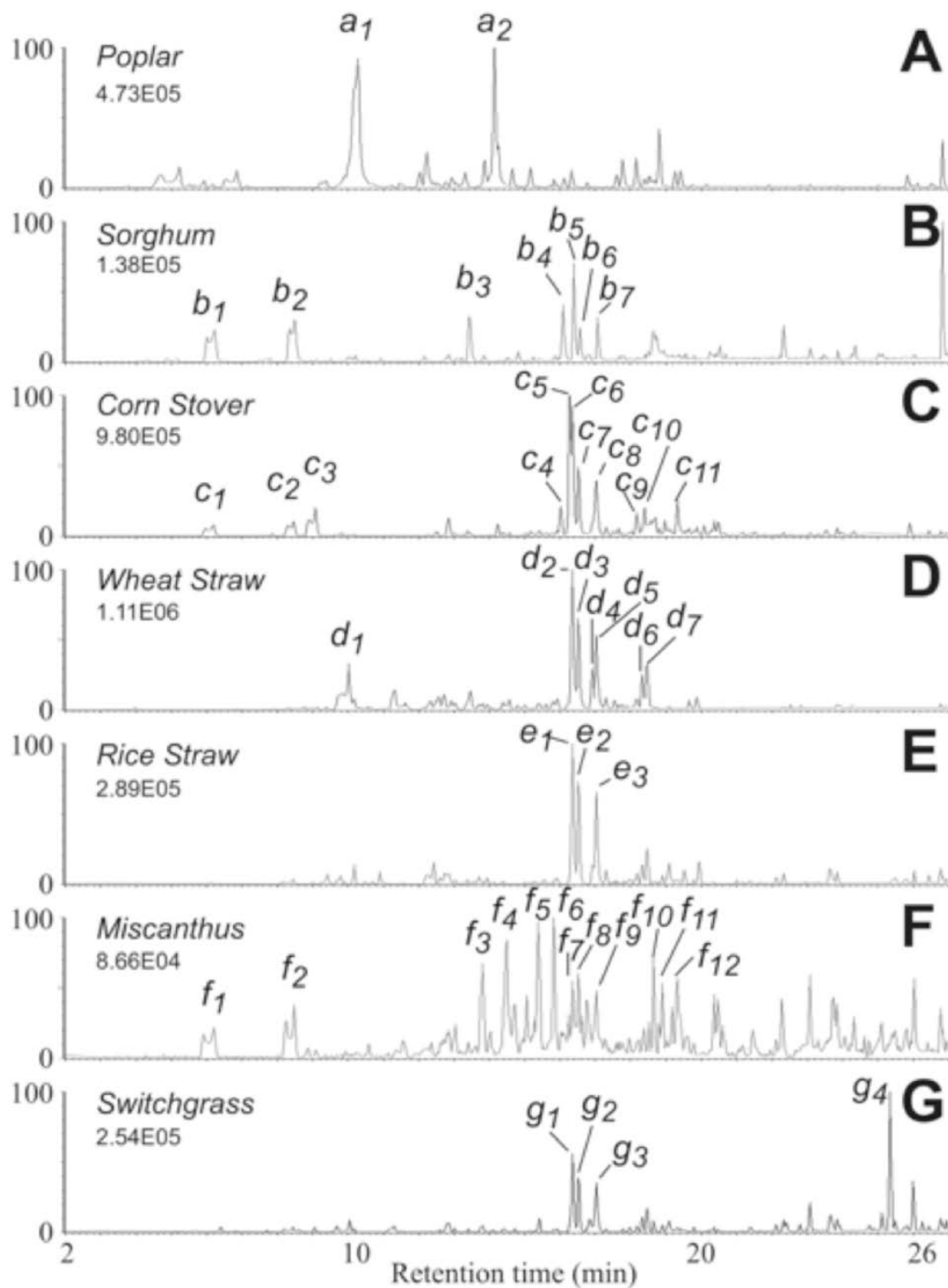


Figure 2.1 Base peak ion (BPI) abundance UHPLC/TOF MS chromatograms generated in negative-ion mode for methanol extracts of (A) poplar, (B) sorghum, (C) corn stover, (D) wheat straw, (E) rice straw, (F) *Miscanthus*, and (G) switchgrass. Labeled peaks are annotated in Table 2.1. Base peak abundances corresponding to the 100% level for each chromatogram are included below the name of each biomass source.

Biomass source	Peak Label	Compound Annotation	µg/g biomass
Poplar (<i>Populus sp.</i>)	a1	Salicortin	103
	a2	2'- Benzoylsalicortin	85
Sorghum (<i>Sorghum sp.</i>)	b1	Benzoic acid	2
	b2	<i>p</i> -Coumaric acid	15
	b3	Coniferaldehyde	8
	b4	Apigenin	15
	b5	Tricin	26
	b6	Guaicylglyceryltricin (T-(4-O-β¹)-G	16
	b7	Guaicylglyceryltricin (T-(4-O-β²)-G	18
Corn stover (<i>Zea mays</i>)	c1	Benzoic acid	3
	c2	<i>p</i> -Coumaric acid	4
	c3	<i>p</i> -Coumaroyl glycerol	15
	c4	1- <i>p</i> -Coumaroyl-2-feruloyl glycerol	68
	c5	1- <i>p</i> -Coumaroyl-3-feruloyl glycerol	344
	c6	Tricin	268
	c7	Guaicylglyceryltricin (T-(4-O-β¹)-G	220
	c8	Guaicylglyceryltricin (T-(4-O-β²)-G	197
	c9	Guaicylglyceryltricin acetate (T-(4-O-β¹)-G')	69
	c10	Guaicylglyceryltricin acetate (T-(4-O-β²)-G')	179
	c11	Sinapyl <i>p</i> -coumarate oxidized dimer C ₄₀ H ₄₀ O ₁₃	155
Wheat straw (<i>Triticum aestivum</i>)	d1	Corymboside	122
	d2	Tricin	349
	d3	Guaicylglyceryltricin (T-(4-O-β¹)-G	96
	d4	<i>p</i>-Hydroxyphenylglyceryltricin (T-(β-O-4)-H)	132
	d5	Guaicylglyceryltricin (T-(4-O-β²)-G	318
	d6/d7	Guaicylglyceryltricin methyl ether (T-(4-O-β)-G^{αOMe}); 2 isomers	316
Rice straw (<i>Oryza sativa</i>)	e1	Tricin	86
	e2	Guaicylglyceryltricin (T-(4-O-β¹)-G	96
	e3	Guaicylglyceryltricin (T-(4-O-β²)-G	88

Table 2.1. Major peaks in the base peak intensity UHPLC-MS chromatograms of extractives from different biomass sources shown in Figure 2.1. Tricin-containing substances are highlighted in bold text. Levels in biomass were estimated from extracted ion chromatogram peak areas by using the molar response for triclin as the response factor for all substances. Nomenclature for triclin compounds follows the convention as proposed in [15].

Table 2.1. (cont'd)

Miscanthus (<i>Miscanthus</i> <i>x giganteus</i>)	<i>f1</i>	Benzoic acid	3
	<i>f2</i>	<i>p</i> -Coumaric acid	3
	<i>f3</i>	C ₃₁ H ₃₆ O ₁₁ isomer 1	44
	<i>f4</i>	C ₃₁ H ₃₄ O ₁₁ isomer 1	44
	<i>f5</i>	Syringoresinol, bis-guaiacylglyceryl ether isomer 1	64
	<i>f6</i>	Syringoresinol, bis-guaiacylglyceryl ether isomer 2	70
	<i>f7</i>	Tricin	14
	<i>f8</i>	Guaicylglyceryltricin (T-(4-O-β-D)-G	22
	<i>f9</i>	Guaicylglyceryltricin (T-(4-O-βe)-G	19
	<i>f10</i>	C ₄₀ H ₄₂ O ₁₃ , isomer 1	25
	<i>f11</i>	C ₄₀ H ₄₂ O ₁₃ , isomer 2	30
	<i>f12</i>	Sinapyl <i>p</i> -coumarate oxidized dimer C ₄₀ H ₄₀ O ₁₃	57
Switchgrass (<i>Panicum</i> <i>virgatum</i>)	<i>g1</i>	Tricin	35
	<i>g2</i>	Guaicylglyceryltricin (T-(4-O-β-D)-G	45
	<i>g3</i>	Guaicylglyceryltricin (T-(4-O-βe)-G	42
	<i>g4</i>	Saponin 1014	199

It is also noteworthy that the compound extracted from switchgrass with the greatest peak area (Figure 2.1G, peak *g4*), is suggested to be a steroidal glycoside saponin based upon exact mass analysis of molecular and fragment ions generated using nonselective multiplexed CID. Annotation as a steroidal glycoside was based in part on formation of a fragment ion at m/z 413.31 in positive-ion mode analysis, consistent with a formula of C₂₇H₄₁O₃⁺, corresponding to a steroidal core. Some saponins exhibit antagonistic activity toward yeast[21], and the high abundance of saponins in switchgrass extract suggests they may be responsible for recalcitrance of some switchgrass cultivars to hydrolysis and subsequent fermentation[22].

2.4.2 Profiles of triclin derivatives in extracts of untreated biomass


To profile the diversity of triclin derivatives in biomass extractives, multiplexed non-selective CID was performed to form fragment ions using a data-independent protocol, and UPLC-MS extracted ion chromatograms (XICs) were generated for the

tricin fragment ion at m/z 329.066 using a relatively gentle collision cell potential of 20 V (Figure 2.2).

Label	Compound annotation	Proposed Formula of neutral molecule	Ret. Time (min)	m/z	Theoretical m/z	Δm (ppm)	RMD (ppm)	XIC Peak Area
u1	Unknown	C28H30O17	12.5	637.1406	637.1405	0.2	221	3437
u2	T-(4-O- β)-G-(4-O- β)-G	C37H38O15	15.6	721.2113	721.2132	-2.6	296	1524
u3	T-(4-O- β)-G-(4-O- β)-G	C37H38O15	15.8	721.2114	721.2132	-2.5	296	1007
	Unknown	C31H34O11	15.8	581.2008	581.2023	-2.6	348	1516
u4	T-(4-O- β)-G-(4-O- β)-G	C37H38O15	15.9	721.2122	721.2132	-1.4	296	2934
u5	Tricin	C17H14O7	16.4	329.0664	329.0661	0.9	201	62269
	T-(4-O- β)-G-(4-O- β)-G	C37H38O15	16.3	721.2124	721.2132	-1.1	296	4387
	<i>p</i> -Hydroxyphenylglyceryl triclin (T-(4-O- β)-G-H)	C26H24O10	16.4	495.1286	495.1289	-0.6	260	6970
u6	Guaicylglyceryltriclin (T-(4-O- β)-G)	C27H26O11	16.5	525.1395	525.1397	-0.4	266	34897
u7	<i>p</i> -Hydroxyphenylglyceryl triclin (T-(4-O- β)-G-H)	C26H24O10	16.9	495.1278	495.1289	-2.2	260	7123
u8	Unknown	C47H42O16	17.0	861.2436	861.2395	4.8	278	11568
	Guaicylglyceryltriclin (T-(4-O- β)-G)	C27H26O11	17.0	525.1397	525.1397	0.0	266	31313

Table 2.2 Annotations for UHPLC-MS peaks for untreated (designated as ‘u’) and AFEX-treated corn stover (designated ‘t’) from Figure 2, focusing on putative triclin derivatives. Abbreviations follow the convention of Lan et al. Plant Physiol. 2016. Compounds containing amino groups in place of hydroxyls in the monolignol portions are designated with ‘(NH₂)’ in the structure abbreviation. Relative mass defect (RMD) values reflect fractional hydrogen content. XIC peak areas are for [M-H]⁺ ions.

Table 2.2. (cont'd)

u9	Unknown	C27H24O1 1	17.3	523.124 4	523.124 0	0.8	237	4374
	Unknown	C47H42O1 6	17.3	861.244 0	861.239 5	5.2	278	2853
u10	Guaicylglyceryltricin acetate (T-(4-O- β -  -G')	C29H28O1 2	18.2	567.150 0	567.150 3	-0.5	265	1162 3
u11	Guaicylglyceryltricin acetate (T-(4-O- β e)-G')	C29H28O1 2	18.4	567.150 0	567.150 3	-0.5	265	1492 9
u12	T-(4-O- β)-G''	C36H32O1 3	18.7	671.175 2	671.176 5	-1.9	263	1341 4
		C39H34O1 5	18.6	741.180 8	741.181 9	-1.5	245	6227
	T-(4-O- β)-S''	C37H34O1 4	18.6	701.187 7	701.187 0	1.0	267	9557
u13	Unknown	C30H30O1 2	19.7	581.165 2	581.165 9	-1.2	285	3929
u14	Unknown	C30H30O1 2	19.9	581.164 9	581.165 9	-1.7	285	4129
	Unknown	C59H50O1 3	20.0	965.320 3	965.317 3	3.1	329	870

About 10-20 major peaks in each extract of monocot biomass yielded this characteristic fragment ion, and these were annotated as various conjugates of triclin with mono- and oligo-lignols, often esterified by phenolic acids, based on accurate mass measurements and the presence of other characteristic fragment ions and neutral mass losses. Many triclin derivatives observed in maize extracts were described in a recent publication[15]. Though thousands of ions, distinguished by m/z and retention time, were detected in the extract of untreated poplar, none of these exhibited other fragment ions characteristic of triclin derivatives, and the major chromatographic peaks in the poplar extract are attributed to salicortin derivatives based on exact pseudomolecular and

fragment ion masses. Annotations of these substances are presented in Table 2.1. Extracts of all six grasses contained two isomeric substances that were consistently the most abundant of the tricin conjugates, and often the most abundant compounds in individual extracts based on UPLC/TOF MS peak areas (Table 2.1). Based on molecular and fragment masses generated using UPLC/TOF MS in negative-ion mode, these were annotated as isomers of the flavonolignan guaiacylglyceryltricin (GGT).

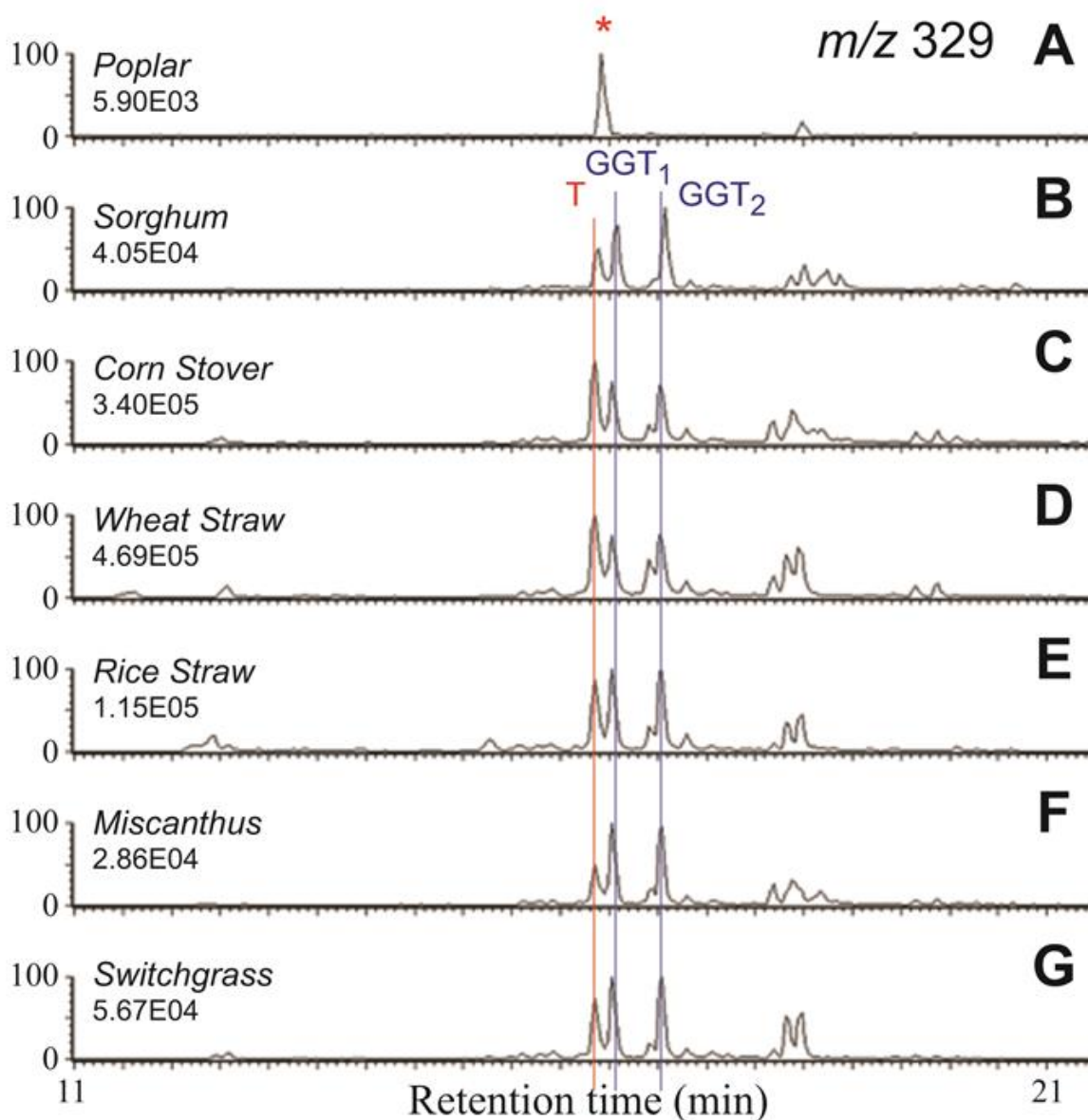


Figure 2.2. UHPLC-MS Extracted ion chromatograms for deprotonated tricin fragments at m/z 329 at elevated collision voltage for extracts of untreated (A) poplar, (B) sorghum,

Figure 2.2. (cont'd)

(C) corn stover, (D) wheat straw, (E) rice straw, (F) Miscanthus, and (G) switchgrass. Peaks and retention times corresponding to tricetin (T) and the two guaiacylglycerol tricetin (GGT) isomers are indicated. The peak in the poplar chromatogram highlighted with an asterisk is not tricetin.

We propose this to match a structure reported in 2012 (based on[13]) with tricetin linked to a monolignol through an ether bond between the 4'-position of tricetin and the β -position of the guaiacylglycerol group (Figure 2.3A). The MS/MS product ion spectra of the two isomers were indistinguishable, consistent with *erythro*- and *threo*- isomers that differ by stereochemical configuration at the α - and β -positions. Characteristic ions derived from the guaiacylglycerol (m/z 195, 165, and 150) and tricetin (m/z 329, 314, and 299) moieties are highlighted in Figure 2.3A.

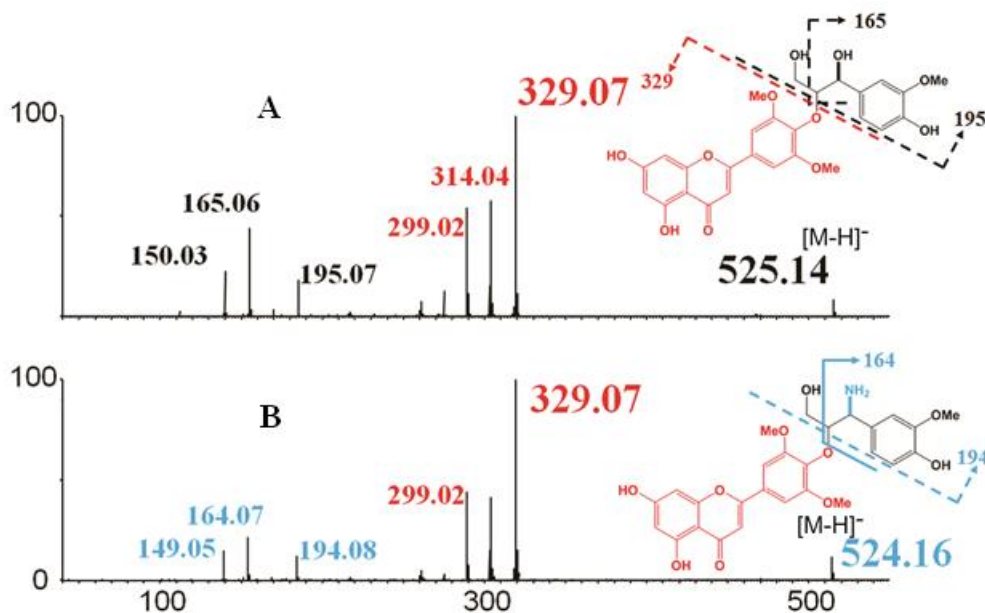


Figure 2.3. Product ion MS/MS spectra generated in negative-ion mode for $[M-H]^-$ ions from (A) guaiacylglyceryltricetin (GGT1) from an extract of untreated corn stover and (B) the analogous compound differing by replacement of a hydroxyl group by an amino group in extracts of EA-treated corn stover. Product ions highlighted in red (color version) are attributed to the flavonoid portion of each molecule, and those highlighted in blue are attributed to the monolignol portion.

2.4.3 Mass balance of tricin derivatives measured using UHPLC-MS

To estimate levels of tricin derivatives in biomass extractives, extracted ion chromatograms for $[M-H]^-$ ions were integrated, and absolute concentrations were calculated for all compounds using the absolute response factor for an authentic tricin standard and normalized to sample mass (Table 2.1). Levels of tricin in monocot extractives ranged from 14 $\mu\text{g/g}$ levels in most *Miscanthus* to 349 $\mu\text{g/g}$ in extracts of untreated wheat straw, and levels in the corresponding AFEX-treated materials were similar for each source. Combined amounts of the two tricin-monolignol conjugate GGT isomers were similar to tricin levels in all cases, but abundances of most other individual tricin analogs were approximately 10-fold lower. Combined totals of the major tricin conjugates consistently reached no more than 0.1% of biomass, about an order of magnitude below the 1.5% of lignin estimated using thioacidolysis in a recent report[15]. The combination of these results suggested that the major tricin-containing substances might only account for a small fraction of the total.

All negative-ion MS/MS spectra of tricin derivatives generated in this investigation and in a recent paper[15] have shown abundant product ions at m/z 329 (corresponding to deprotonated tricin). We postulated that the diversity of tricin derivatives would be revealed in narrow mass window extracted ion chromatograms for m/z 329.066 ± 0.005 at elevated collision energies, and an example from an extract of untreated corn stover is presented in Figure 2.4. Though the chromatogram displays prominent peaks attributed to the major tricin and GGT isomers, the signal remains elevated well above baseline from 12-24 minutes, suggesting a diverse population of unresolved tricin derivatives. Comparison of the peak areas for the tricin peak and the

total m/z 329 signal established that triclin itself accounted for only 8% of the total signal in corn stover. Similar ratios were observed for extracts of all monocots described in this report, and integration of the total m/z 329 signal yields an estimate that accounts for about 1% of corn stover lignin, a result more consistent with amounts measured by thioacidolysis. We conclude that the majority of triclin derivatives in monocot lignins exist in a diverse set of chromatographically unresolved compounds, each of which has low abundance relative to triclin or GGT.

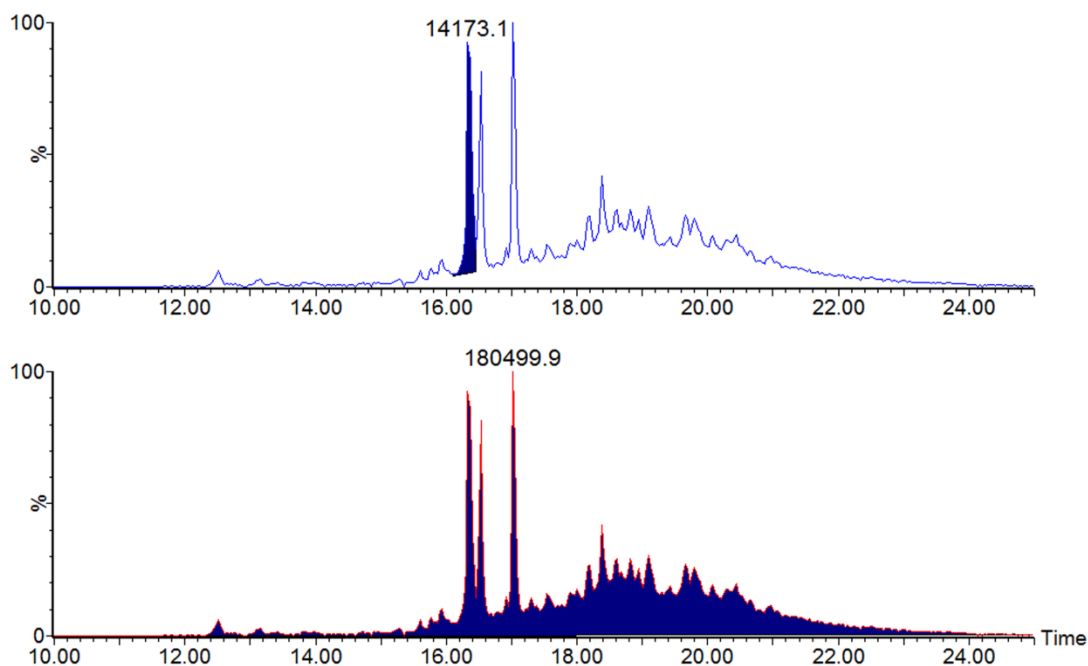


Figure 2.4. UHPLC-MS Extracted ion chromatogram for m/z 329.066 \pm 0.005 obtained for methanolic extract of untreated corn stover in negative-ion mode. The top panel shows the integrated peak area for triclin, and the bottom panel shows integration of all signal at this m/z eluting from 12-25 minutes.

2.4.4 Exploration of triclin conjugate diversity using tandem mass spectrometry (MS/MS)

Two mass spectrometric analyses of corn stover extract provided evidence of remarkable diversity in phenolic metabolites. In the first case, we averaged all mass

spectra in the UHPLC-MS data files from the region encompassing elution times of 12-24 minutes in the chromatogram. The resulting mass spectrum showed multiple peaks at every nominal (integer) m/z value (Figure 2.5) across the entire mass range plus additional peaks at half-integer values that are attributed to larger doubly-charged compounds. More than 35000 ions were reported in the spectrum peak list, which was limited to m/z 50-1500. At least 6000 of these exhibited relative mass defects (RMDs)[17-18], a measure of fractional hydrogen content, in the range of 180-450 ppm that are consistent with flavonol or monolignol precursors. Though we recognize that a single molecule usually is detected in multiple ionized forms including isotopologues resulting from natural abundance of stable isotopes, it must also be considered that many lignin constituents exist in multiple isomeric forms. We propose that the mass spectra suggest a diversity of phenolic constituents on the order of at least 5000-10000 distinct chemical forms in extracts of each monocot investigated in this work.

One may wonder whether most of the detected ions are present in significant quantities, and to address this question, a histogram of abundances of the detected ions in an extract of untreated corn stover was generated (Figure 2.6). Detected ion signals span 5 orders of magnitude in range, and though the most abundant extractives (e.g. tricin and GGT) are about 1000-fold more abundant than the median, each bin around the median abundance contains several thousand ions. The results suggest that the combined levels of low abundance substances account for a significant fraction of extracted substances, and that a focus on large peaks in LC/MS chromatograms will likely neglect the contributions of an extensive pool of compounds that individually are present at much lower levels.

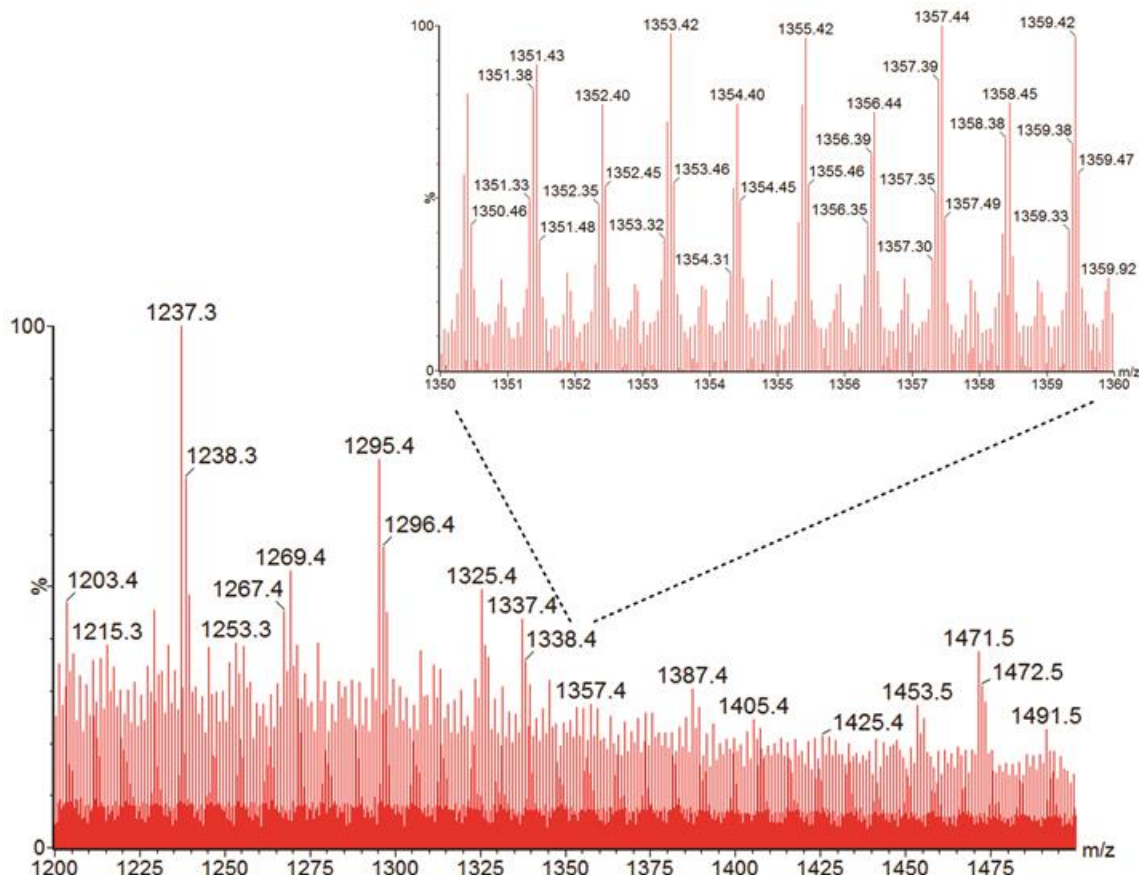


Figure 2.5. High-mass region of the negative-ion mass spectrum obtained by averaging all mass spectra from UHPLC-MS analysis of a methanolic extract of untreated corn stover over the retention time region of 15-25 minutes. Peak averaging was performed using a bin width of 0.05 m/z . Spectra were generated at the lowest collision potential (5 V). The magnified inset of the range from m/z 1350-1360 demonstrates multiple resolved peaks at every nominal mass.

Structural features of individual ions may be interrogated by generating MS/MS product ion spectra, but it is often infeasible to generate MS/MS spectra for each nominal mass during a single LC/MS analysis. Furthermore, the limited mass resolution of the quadrupole precursor ion filter ensures that selecting any nominal mass window of 1 m/z width will still transmit numerous isobaric ions that have the same nominal mass but different elemental formulas (Figure 2.5 inset). Even if these factors could be addressed, the discussion above demonstrates that levels of individual substances may be too low to generate sufficient signal in product ion spectra to characterize an individual compound.

Given these limitations, we chose to introduce biomass extracts into the mass spectrometer by continuous infusion, and to use wide mass window ($\sim 60\ m/z$) for precursor ion selection. Ion source voltages were elevated to minimize non-covalent dimer and oligomer ions, and MS/MS product ion spectra were generated using 13 precursor windows ranging from m/z 926 to 3934.

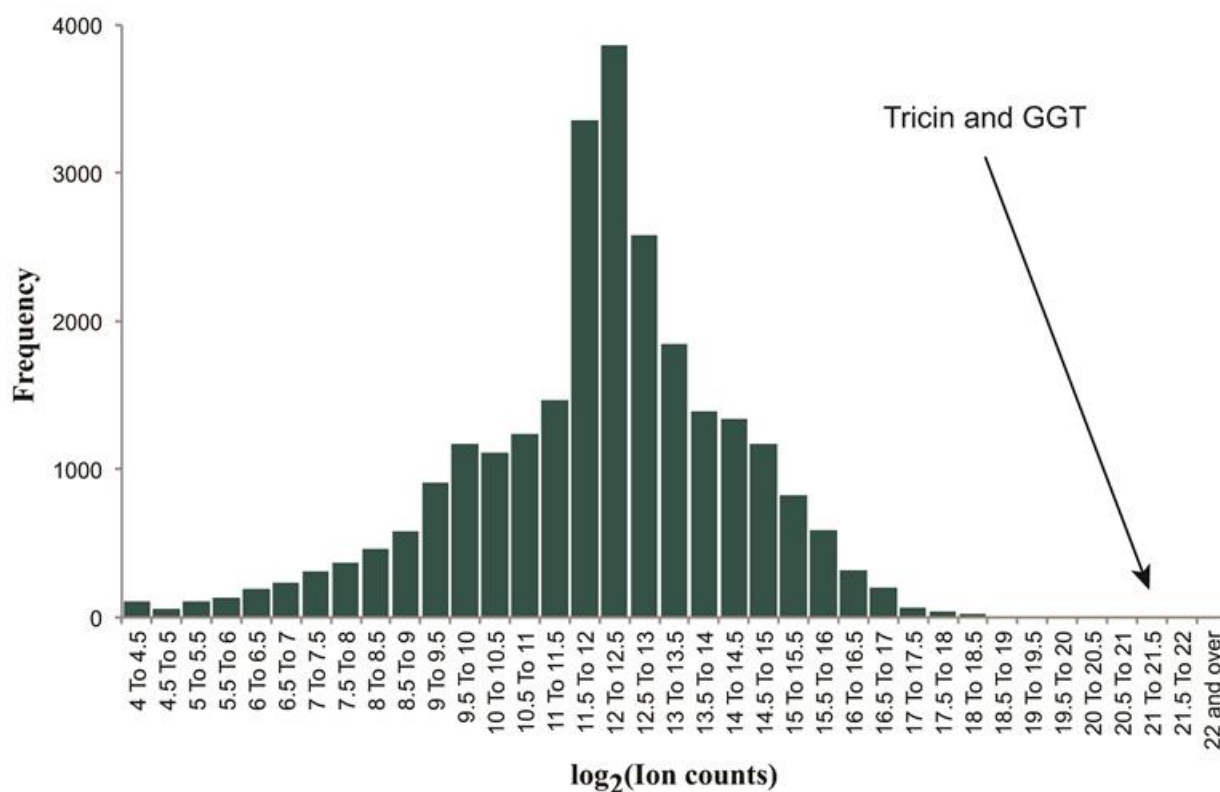


Figure 2.6. Histogram showing the frequency of ions in the averaged mass spectrum shown in Figure 2.5, sorted by the absolute signal ($\log_2(\text{number of ion counts})$). An arrow points to the $[M-H]^-$ signals for abundant molecules triclin and GGT isomers but do not appear because so few molecules had such high abundance.

In all MS/MS spectra, the characteristic triclin fragment ion was not only observed, but it was the most abundant ion for all precursor masses up to m/z 1500 (Figure 2.7) though its absolute and relative abundance declined gradually toward higher m/z values,

where it became less abundant than monolignol fragments of m/z 195 and 165 (G and H units). Fragments attributed to *p*-coumarate (m/z 163) were also abundant at lower masses and declined in parallel with the triclin fragment, and the corresponding ferulate fragment (m/z 193) increased in absolute abundances up to m/z ~3000. The observation of triclin as a significant fragment ion that decreases in abundance at higher molecular masses is consistent with the notion that triclin serves as an end group in lignin formation.

2.4.5 Triclin derivatives after extractive ammonia pretreatment

EA pretreatment of corn stover led to substantial changes in the LC/MS profiles of extractives (Figure 2.8). A new set of chromatographic peaks emerged after pretreatment, some of which are attributed to ammoniated analogues of the compounds characterized in untreated biomass extracts. The most abundant (peak t_5) is a compound detected in negative-ion mode as m/z 524.15, eluting at 14.07 minutes. The product ion MS/MS spectrum shows similarity to GGT (also known as T-(4-O- β)-G according to recently proposed nomenclature[15]), but the 1 Da lower molecular mass suggests replacement of a hydroxyl group by an amino group. Fragment masses at m/z 149, 164, and 194 are consistent with the amino group located on the monolignol portion, most likely at the more reactive benzylic position. It was also anticipated that AFEX treatment would drive hydrolysis and ammonolysis of esters, and examination of extracted ion chromatograms of extractives for triclin and its mono- and di-lignol conjugates from untreated and AFEX-treated corn stover document disappearance of GGT acetate (T-(4-O- β)-G'; Figures 2.9 and 2.10), but about half of the acetate esters survived pretreatment, but only with an amino group replacing the hydroxyl at the benzylic α -position.

In contrast, the *p*-coumarate ester of GGT (T-(4-O- β)-G"; retention time 18.70 min, both isomers coeluting) proved more resistant to the AFEX process, with about half the original amount remaining after pretreatment, with most of the remainder retaining the *p*-coumarate ester but with amino substitution at the benzylic α -position.

The unexpected survival of the benzylic amine esters suggests the amino group stabilizes the ester toward ammonolysis or hydrolysis, and makes the case that the hydroxyl groups in the α -position accelerate lysis of the ester bonds relative to the effect of amino substitution at the same position. Annotation of putative triclin-derived substances observed in extracts of EA-treated corn stover are compared to those in extracts of untreated corn stover in Table 2.2.

2.4.6 Incorporation of triclin into flavonolignans through the action of oxidative enzymes

The abundance of triclin-containing flavonolignans in grasses and the low abundance of oligolignols without triclin incorporation suggested that triclin might play important roles in the formation of lignin in grasses. To test whether common oxidative enzymes catalyse incorporation of triclin into flavonolignans and perhaps higher oligomers, incubations were performed with combinations of triclin and coniferyl alcohol with laccase and peroxidase. Reaction products were assessed at multiple reaction time points using UHPLC TOF MS.

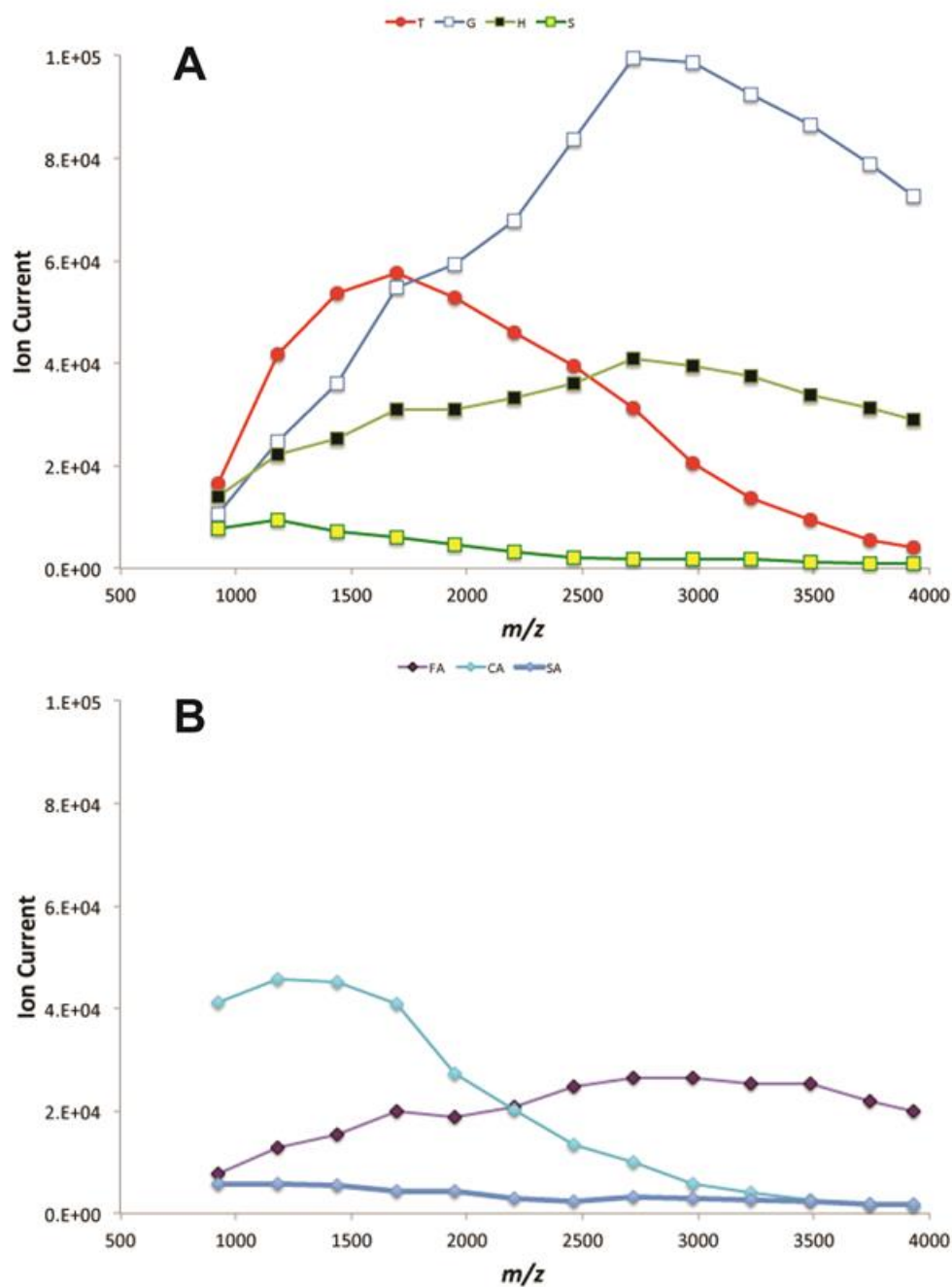


Figure 2.7. Product ion abundances as a function of precursor ion m/z generated from wide-window (~ 60 Da) MS/MS spectra of a methanolic extract of untreated corn stover. Sample introduction employed flow injection analysis and negative-mode electrospray ionization. (A) Product ions m/z 329 (deprotonated tricin, red circles), 195 (monolignol G, blue squares), 165 (monolignol H, green line and black squares), and 225 (monolignol S, green line with yellow squares). (B) Product ions for phenolic acid anions m/z 163 (*p*-coumarate, turquoise), 193 (ferulate, violet), and 223 (sinapate, blue). Vertical axis scaling is the same for both panels.

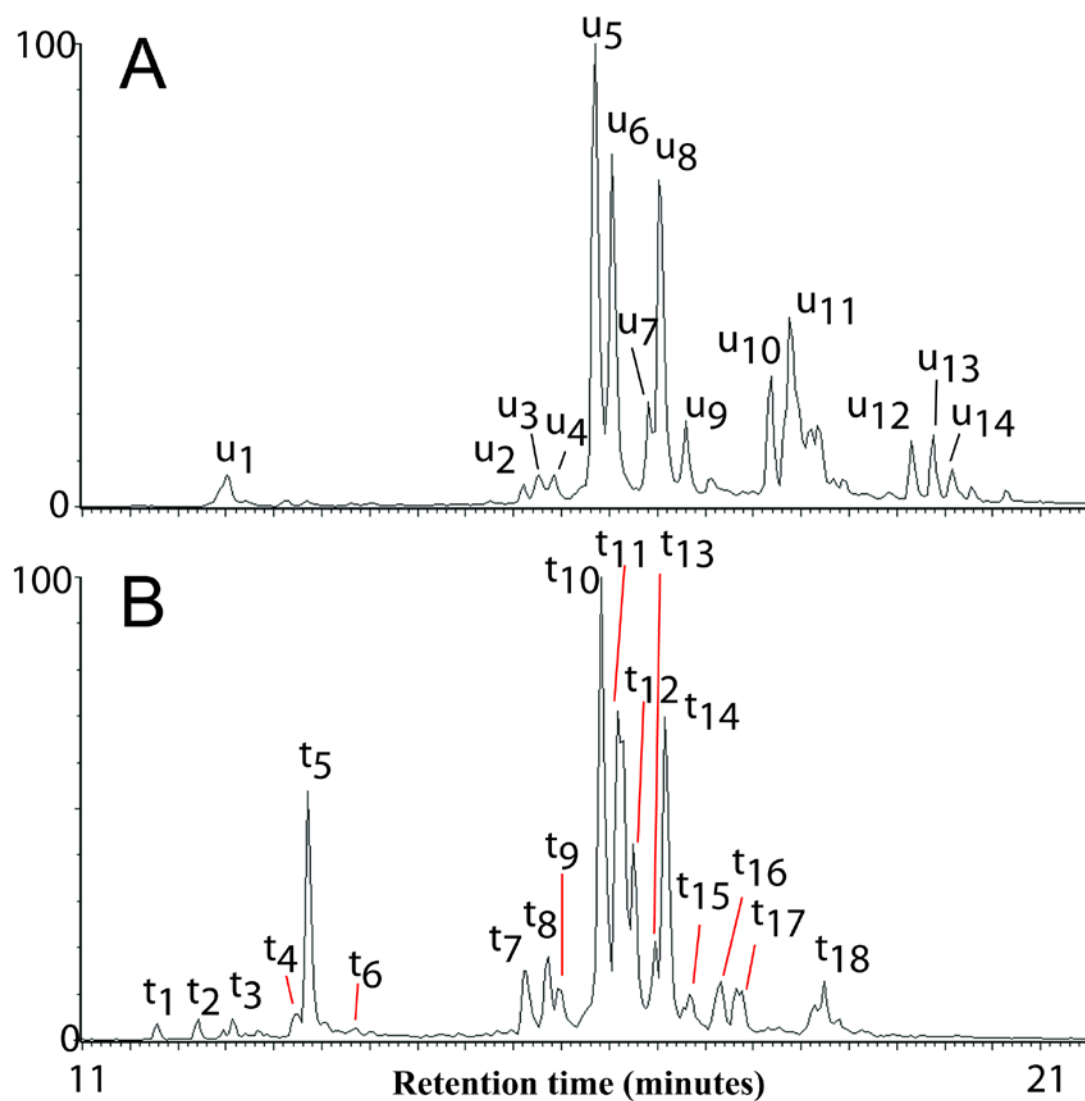


Figure 2.8. UHPLC-MS extracted ion chromatograms for m/z 329.07 generated in negative-ion mode for methanolic extracts of (A) untreated corn stover and (B) AFEX-treated corn stover. Annotations of peaks are presented in Table 2.2.

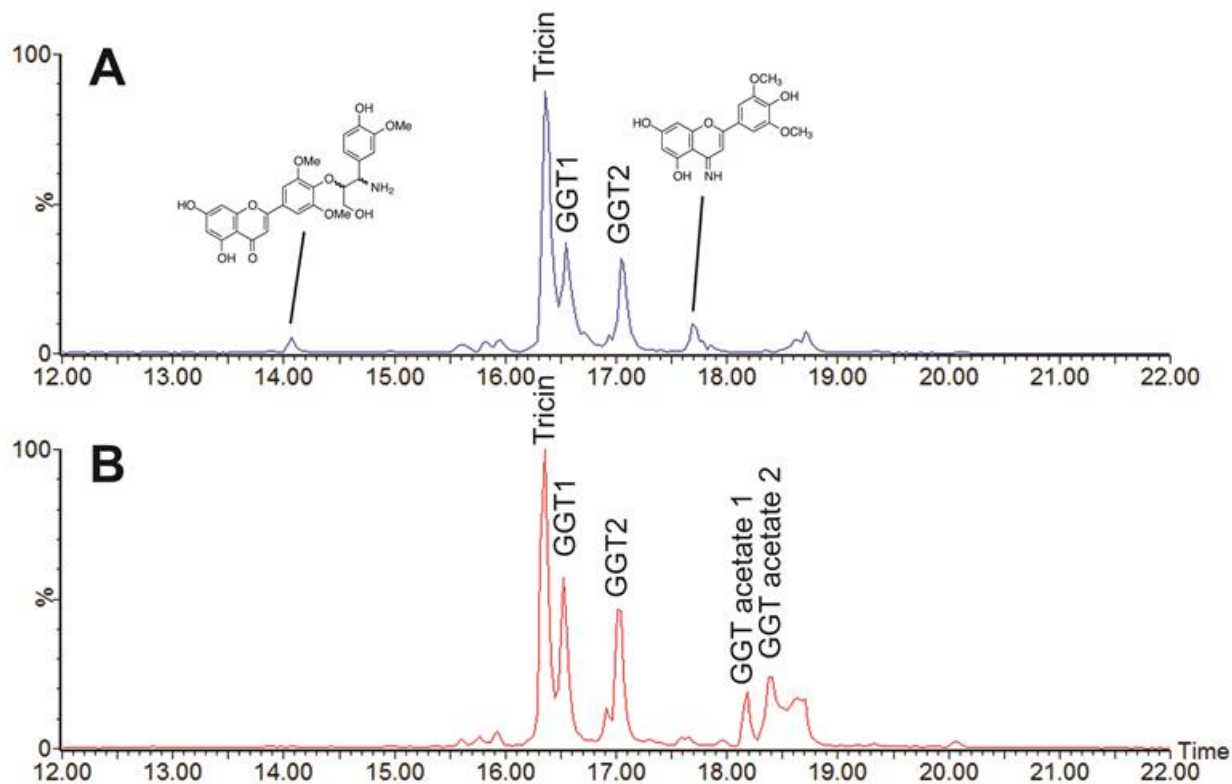


Figure 2.9. Extracted ion UHPLC-MS chromatograms for triclin and its mono- and di-lignol conjugates (m/z 329.07) in (A) AFEX-treated corn stover and (B) untreated corn stover.

Incubation of laccase with an equimolar mixture of coniferyl alcohol and triclin yielded two peaks that matched the two common GGT isomers in exact mass and retention times (Figure 2.11). Yields of GGT dropped off with increasing incubation time and had completely disappeared by 2 h incubation. No evidence for slightly larger oligomers or oxidation products was observed in the UHPLC TOF MS analyses, suggesting that products either had higher molecular masses than the mass spectrometer recorded or that the products became insoluble and were not eluted using liquid chromatography. Incubation of coniferyl alcohol alone with laccase yielded multiple isomers of the dehydro-dimer product as expected, detected by UHPLC TOF MS as m/z 341 in positive-ion mode (Figure 2.12B), but addition of triclin to the mixture caused more than a 10-fold decrease in yields of this product (Figure 2.12D).

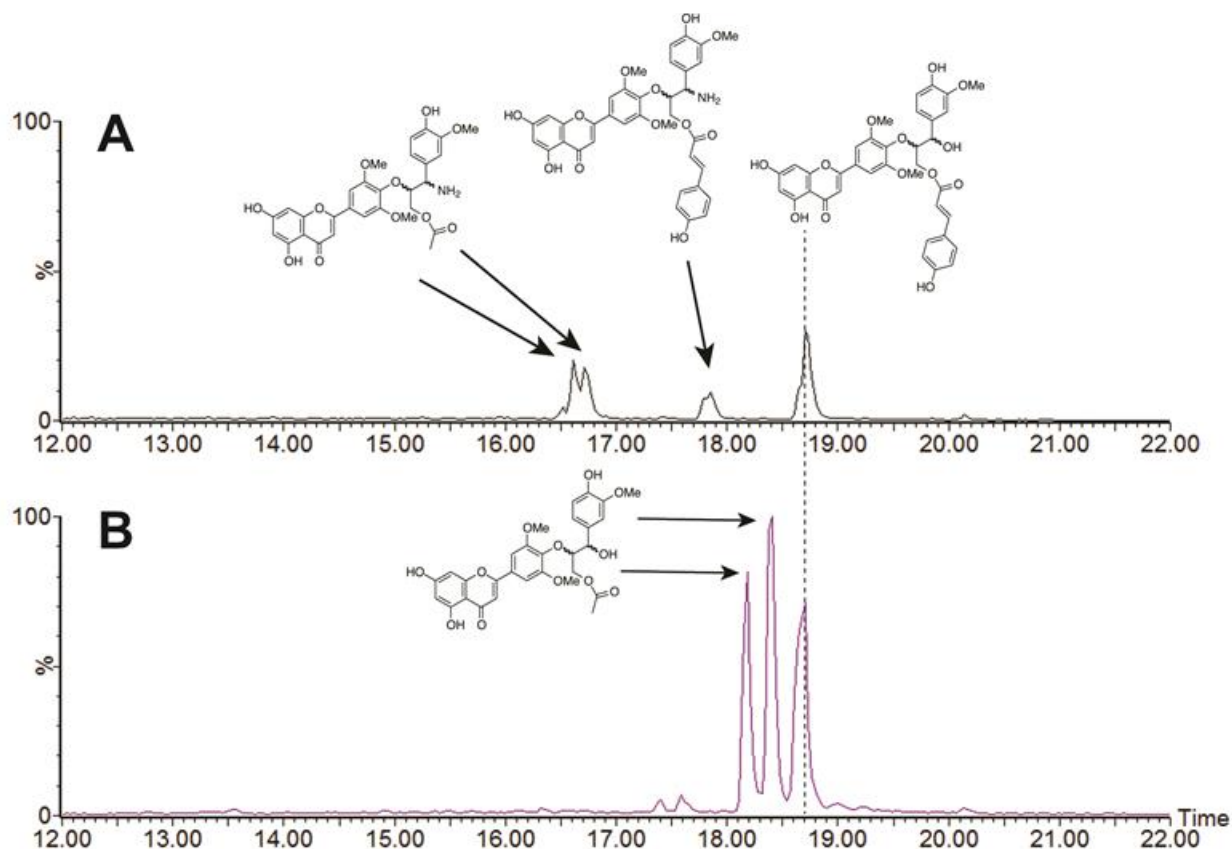


Figure 2.10. Extracted ion UHPLC-MS chromatograms for combined signals of m/z 567.15, 671.17, and 879.25 corresponding to esterified monolignol conjugates of tricetin for extracts of (A) AFEX-treated corn stover and (B) untreated corn stover. The two chromatograms share a common vertical scale (100% = ion counts). Products containing benzylic amino groups are detected at the extracted ion masses, but correspond to substances containing one heavy isotope. As a result, their signals are approximately one-third of the monoisotopic ion signals.

2.5 Conclusions

Tricin and its conjugates are among the most abundant extractive compounds in untreated and AFEX-treated monocots, exceeded in corn stover by phenolic esters of glycerol and by hydroxycinnamoyl amides following AFEX treatments. All grasses investigated show tricetin forms conjugates with mono-, di-, and tri-lignols and phenylpropanoid acids, and that suggest it is incorporated into higher molecular mass lignin oligomers. Mass spectrometric analysis indicates that a remarkably diverse suite of

tricin-containing substances up to 4 kDa was detected in methanolic extracts. Untargeted profiling revealed that chromatographically unresolved triclin-containing substances account for more triclin than the major triclin mono- and oligo-lignol conjugates, and the wide window MS/MS spectra suggest quantitative differences in composition that depend on molecular masses of individual lignin components or fractions.

We propose that these mass spectrometric approaches offer useful advantages relative to size exclusion chromatography for investigations into how lignin chemistry changes as a function of molecular mass, thermochemical treatments, and perhaps metabolic engineering of lignin in plants. AFEX and EA pretreatments incorporate nitrogen into a variety of compounds, most notably as an α -amino group in place of a benzylic alcohol of conjugated monolignols. Replacement of the α -hydroxyls by amino groups appears to stabilize γ -position esters to hydrolysis and ammonolysis. Ammonia pretreatments also incorporate nitrogen into flavonoid cores, including ammoniated triclin derivatives that are probably imines.

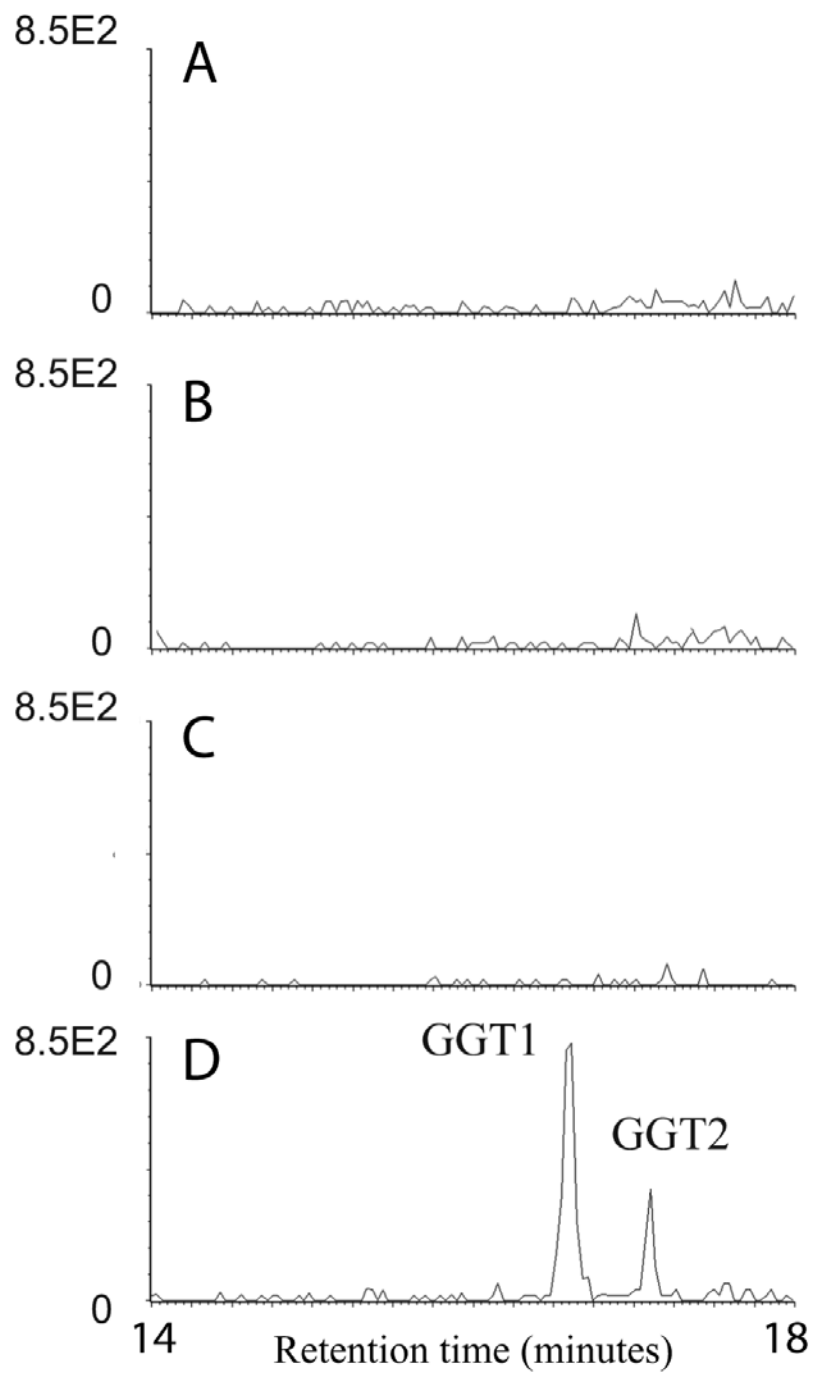


Figure 2.11. UHPLC-MS extracted ion chromatograms for m/z 525 generated in negative-ion mode for control reactions (A-C) and incubations of triclin and coniferyl alcohol with laccase enzyme.

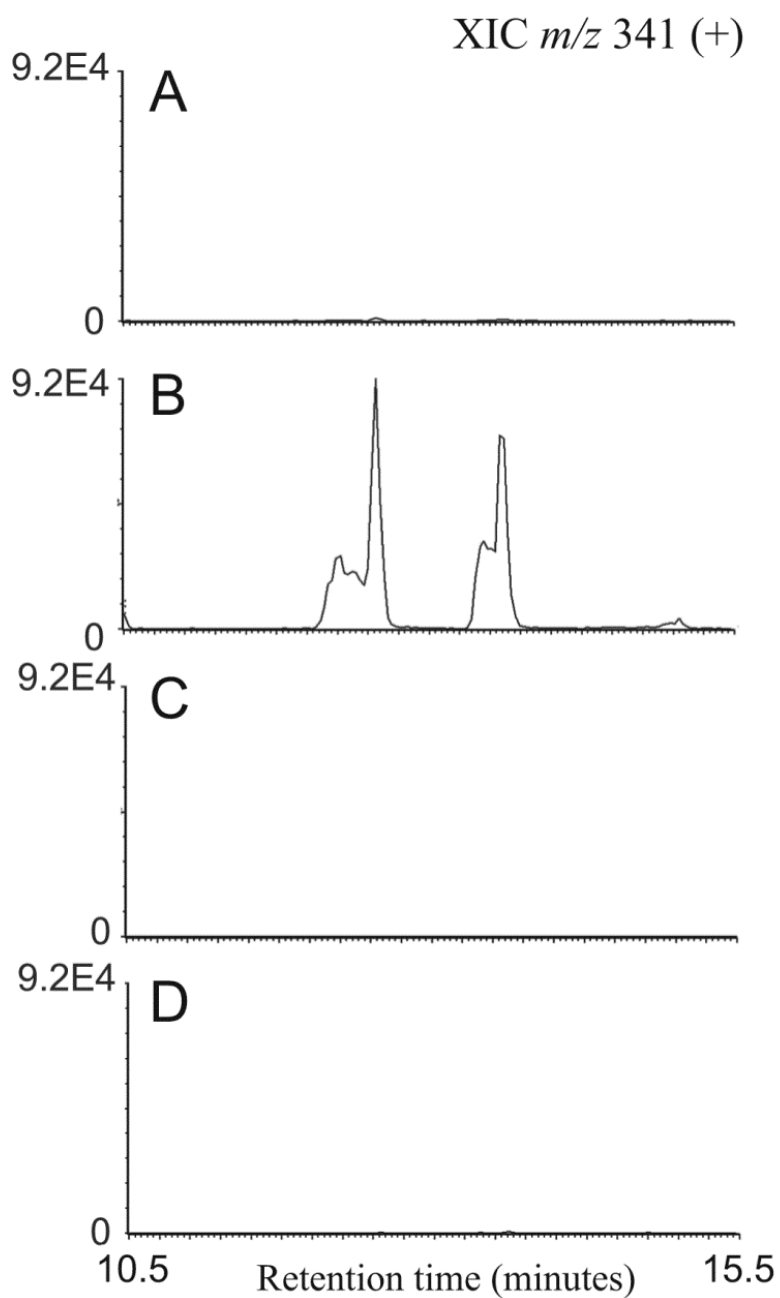


Figure 2.12. UHPLC-MS extracted ion chromatograms for m/z 341 (positive-ion mode; $[M+H-H_2O]^+$) for detection of dehydrodimer of coniferyl alcohol in (A) control, no coniferyl alcohol, (B) coniferyl alcohol plus laccase, (C) coniferyl alcohol without laccase, and (D) coniferyl alcohol plus triclin and laccase.

REFERENCES

REFERENCES

- (1.) Balan, V.; Bals, B.; Chundawat, S. P. S.; Marshall, D.; Dale, B. E., Lignocellulosic Biomass Pretreatment Using AFEX. In *Biofuels: Methods and Protocols*, Mielenz, J. R., Ed. Humana Press Inc: Totowa, 2009; Vol. 581, pp 61-77.
- (2.) Chundawat, S. P. S.; Vismeh, R.; Sharma, L. N.; Humpala, J. F.; da Costa Sousa, L.; Chambliss, C. K.; Jones, A. D.; Balan, V.; Dale, B. E., *Bioresource Technology* **2010**, *101* (21), 8429-8438.
- (3.) da Costa Sousa, L.; Chundawat, S.; Bokade, V.; Foston, M.; Azarpira, A.; Dale, B. E.; Balan, V., Fractionation and characterization of lignin extractives from E-AFEX(TM) pretreatment process. In *12th AIChE Annual Meeting*, Pittsburgh, PA, 2012.
- (4.) Alonso, D. M.; Wettstein, S. G.; Mellmer, M. A.; Gurbuz, E. I.; Dumesic, J. A., *Energy & Environmental Science* **2013**, *6* (1), 76-80.
- (5.) Luterbacher, J. S.; Rand, J. M.; Alonso, D. M.; Han, J.; Youngquist, J. T.; Maravelias, C. T.; Pfleger, B. F.; Dumesic, J. A., *Science* **2014**, *343* (6168), 277-280.
- (6.) Guo, G. L.; Hsu, D. C.; Chen, W. H.; Chen, W. H.; Hwang, W. S., *Enzyme and Microbial Technology* **2009**, *45* (2), 80-87.
- (7.) Kumar, R.; Mago, G.; Balan, V.; Wyman, C. E., *Bioresource Technology* **2009**, *100* (17), 3948-3962.
- (8.) Li, C. L.; Knierim, B.; Manisseri, C.; Arora, R.; Scheller, H. V.; Auer, M.; Vogel, K. P.; Simmons, B. A.; Singh, S., *Bioresource Technology* **2010**, *101* (13), 4900-4906.
- (9.) Sambusiti, C.; Ficara, E.; Malpei, F.; Steyer, J. P.; Carrere, H., *Energy* **2013**, *55*, 449-456.
- (10.) She, D. A.; Xu, F.; Geng, Z. C.; Sun, R. C.; Jones, G. L.; Baird, M. S., *Industrial Crops and Products* **2010**, *32* (1), 21-28.
- (11.) Ververis, C.; Georghiou, K.; Christodoulakis, N.; Santas, P.; Santas, R., *Industrial Crops and Products* **2004**, *19* (3), 245-254.
- (12.) Freudenberg, K., *Science* **1965**, *148* (3670), 595.
- (13.) del Rio, J. C.; Rencoret, J.; Prinsen, P.; Martinez, A. T.; Ralph, J.; Gutierrez, A., *J Agric Food Chem* **2012**, *60* (23), 5922-35.

- (14.) Lan, W.; Lu, F.; Regner, M.; Zhu, Y.; Rencoret, J.; Ralph, S. A.; Zakai, U. I.; Morreel, K.; Boerjan, W.; Ralph, J., *Plant Physiol* **2015**, *167* (4), 1284-95.
- (15.) Lan, W.; Morreel, K.; Lu, F. C.; Rencoret, J.; del Rio, J. C.; Voorend, W.; Vermeris, W.; Boerjan, W.; Ralph, J., *Plant Physiology* **2016**, *171* (2), 810-820.
- (16.) Sousa, L. D.; Foston, M.; Bokade, V.; Azarpira, A.; Lu, F. C.; Ragauskas, A. J.; Ralph, J.; Dale, B.; Balan, V., *Green Chemistry* **2016**, *18* (15), 4205-4215.
- (17.) Gu, L. P.; Jones, A. D.; Last, R. L., *Plant Journal* **2010**, *61* (4), 579-590.
- (18.) Stagliano, M. C.; DeKeyser, J. G.; Omiecinski, C. J.; Jones, A. D., *Rapid Communications in Mass Spectrometry* **2010**, *24* (24), 3578-3584.
- (19.) Ekanayaka, E. A. P.; Celiz, M. D.; Jones, A. D., *Plant Physiology* **2015**, *167*, 1221-1232.
- (20.) Bouaziz, M.; Veitch, N. C.; Grayer, R. J.; Simmonds, M. S. J.; Damak, M., *Phytochemistry* **2002**, *60* (5), 515-520.
- (21.) Simons, V.; Morrissey, J. P.; Latijnhouwers, M.; Csukai, M.; Cleaver, A.; Yarrow, C.; Osbourn, A., *Antimicrobial Agents and Chemotherapy* **2006**, *50* (8), 2732-2740.
- (22.) Kim, Y.; Mosier, N. S.; Ladisch, M. R.; Pallapolu, V. R.; Lee, Y. Y.; Garlock, R.; Balan, V.; Dale, B. E.; Donohoe, B. S.; Vinzant, T. B.; Elander, R. T.; Falls, M.; Sierra, R.; Holtzaple, M. T.; Shi, J.; Ebrik, M. A.; Redmond, T.; Yang, B.; Wyman, C. E.; Warner, R. E., *Bioresource Technology* **2011**, *102* (24), 11089-11096.

**Chapter Three: A metabolomics investigation into whether phenolic acid esters are
incorporated into lignin in monocot grasses**

3.1 Abstract

Conversion of lignocellulosic biomass to fuel suffers from limited knowledge about phenolic structures of grass lignin. A central goal of this study has been to establish roles of common phytochemical hydroxycinnamic acids, namely *p*-coumaric acid and ferulic acid, in the formation of grass lignin. This investigation relied on ultrahigh pressure liquid chromatography-high resolution mass spectrometry (UHPLC-HRMS) and high resolution tandem MS (HRMS/MS) strategies, and to a lesser yet very important extent, 1D and 2D nuclear magnetic resonance (NMR) spectroscopy to characterize molecular structures of purified fractions containing the more abundant compounds. Detection and characterization of a variety of glycerol esterified to one or two hydroxycinnamic acids in all grasses studied suggests an unreported role for this class of compounds in lignin of monocot grasses. The most abundant component of methanol extract of corn stover was identified as 1-*p*-coumaroyl-3-feruloylglycerol using NMR structural assignments of the purified compound. MS/MS annotation of multiple detected dimers of this compound suggests modification of diferulates, the famous lignin-hemicellulose crosslinkers with *p*-coumaroylglycerol rather than exclusive esterification to carbohydrate polymers. The discovery of covalent linkage of 1-*p*-coumaroyl-3-feruloylglycerol to tricetin, the flavonoid recently investigated for its role in formation of lignin formation in grasses, was suggested by MS/MS annotation. This suggests a deeper involvement of glycerol esters in lignification. Another abundant component detected in methanol extracts of all studied grasses, was annotated as a dimer of sinapyl *p*-coumarate, formed in part by ring closure to form a tetrahydrofuran (THF) derivative. Detection and HRMS/MS annotation of multiple additional compounds detected in corn stover extracts suggested addition of one or more oxidized sinapyl *p*-coumarates and/or guaiacylglycerol units to this abundant unit to produce large lignin

molecules. HRMS/MS annotation of these compounds suggests involvement of *p*-coumarate units in coupling of different units to form higher molecular mass lignin constituents. This finding expands the role of *p*-coumarate units beyond what was known before this report.

3.2 Introduction

Grasses make one of the most adaptive families of plants, growing in a variety of climates from steppes to deserts and from savannahs to mountains. Humans use grasses in numerous ways, from food production to furnishing turf for recreation and sport arenas. Grasses include >11,000 species and habitat all around the planet [1]. Although a recent report suggests grasses were present on Earth around 100 million years ago [2], it is generally accepted that grasses diverged from other monocots between 50 to 60 million years ago [3-4], making them a younger group of plants relative to many dicots or conifers [5]. Grass physical properties have evolved to aid their survival in a variety of climates. For instance, most grasses possess flexible cell wall fibers helping them to survive in windy or dry climates. One important component that contributes to the physical properties of plant cell walls is lignin, which is derived from oligomerization of phenolic metabolites. As was demonstrated in Chapter 2, grass lignification is triggered by oxidation of a flavone molecule, tricetin, followed by oligomerization by its reaction with monolignols; a different lignification process that does not involve tricetin produces lignin in hardwood dicots. However, current knowledge about grass lignin does not support the differences in physical properties of lignin of grasses compared to hardwoods. For example, studying grass lignin has relied on using the same methodologies to establish amounts of monomeric units in hardwood lignin including revelation of its monolignols ratio and/or linkage types usually with the help of synthesized model compounds [6]. This approach, however, does

not address differences in the chemical details that translate into different physical properties such as rigidity of grasses versus hardwoods.

Much of the secondary cell wall in grasses consists of hemicellulose, composed of arabinoxylans, which are polymers of β -D-(1,4)-linked xylose units with side chains of arabinosyl groups attached at every 2-3 xylose units. Some of the arabinose moieties are esterified by ferulate (4-hydroxy-3-methoxycinnamic acid), a phenolic acid [7-8]. Esters of the hydroxycinnamic acid family including *p*-coumarates (4-hydroxycinnamic acid) and/or ferulates (here called phenolic acids) have been reported to be bound to hemicelluloses of multiple monocots (including grasses) including barley straw[9], maize [10], maize and wheat bran [11], bamboo shoots [12-14], and extractives including sugar cane bagasse[15]. Secondary cell wall hemicelluloses exhibit resistance to digestion and this resistance has been attributed in part to crosslinking of hemicellulose oligosaccharides [16]. Oxidation of ferulate esters to form dehydrodiferulate crosslinkers (generally termed diferulates) has been reported in maize [17]. These crosslinkers come in a variety of linkage forms (8-*O*-4, 4-*O*-5, 5-5, 8-8), with the nomenclature describing the positions on the ferulate group involved in crosslinking.

It is expected that diferulate crosslinks must be hydrolyzed during acid or alkali pretreatment of biomass for optimal conversion of hemicelluloses to fermentable sugars. Qualitative and semi-quantitative analyses of diferulate linkages in corn before and after alkali pretreatment of biomass via ammonium fiber expansion (AFEX) have been reported recently [18]. Figure 3.1 shows a schematic representation of hemicellulose arabinoxylans and diferulate crosslinks as well as non-linker modifications (e.g. acetylation) on the oligosaccharides.

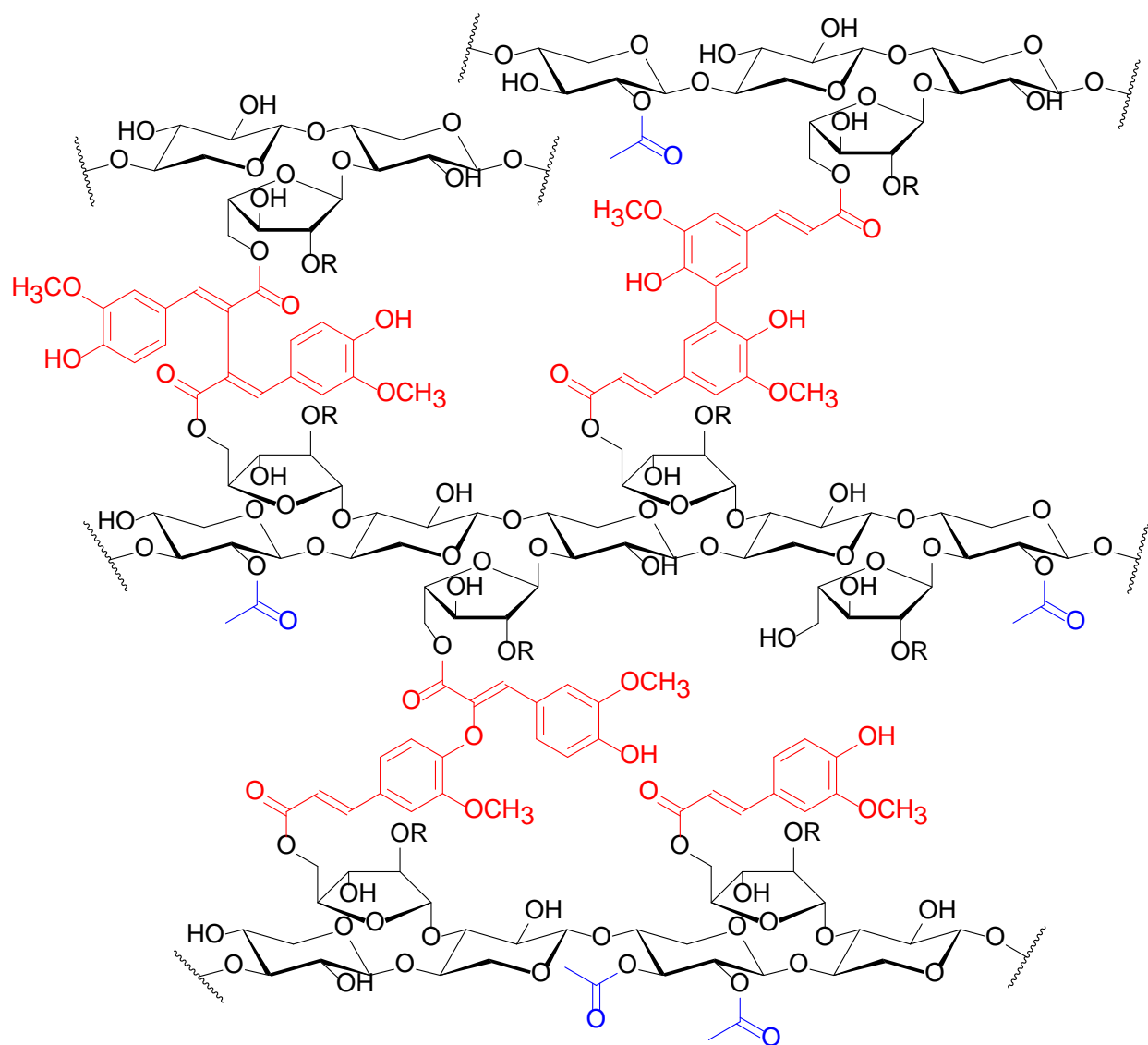


Figure 3.1. A model representation of hemicellulose arabinoxylan chains showing diferulate crosslinks between arabinoxylan chains (highlighted in red) [18].

Interestingly, reports of the presence of hydroxycinnamic acids (such as *p*-coumaric acid and ferulic acid) and their esters have been limited to grasses [19-20] and other monocots including bamboo [12-14]. This is coincident to the association of triclin in lignin of grasses and not hardwood (Chapter 2). However, as it will be discussed in this chapter, association of phenolic acids in biomass goes beyond serving just as esterified groups to saccharides and diferulate cross-linkers in hemicellulose. Whether phenolic esters and crosslinkers are linked only to the arabinoxylan polymer as has been proposed, or are incorporated through covalent

linkages to lignin has remained largely unexplored. Some important knowledge gaps regarding these crosslinkers center around how they are incorporated into higher molecular mass substances in plant cell walls and how they contribute to cell wall mechanical properties.

Presence of free hydroxycinnamic acids has not been reported to be in high levels in extracts of grasses. However, pretreatment of lignocellulosic biomass from grasses with dilute acid [21], alkali [22], or ammonia [21 , 23] releases substantial amounts free phenolic acids as well as amides (in case of ammonia treatment). Notable examples are *p*-coumaroyl and feruloyl amides released by ammonia pretreatment. For instance, Chundawat et al. [21] reported free *p*-coumaric acid and ferulic acid to be 161 and 35 µg/g of biomass in untreated corn stover respectively, while after pre-treatment with AFEX those values increased to 1,080 and 103 µg/g respectively. Upon treatment of corn stover with dilute sulfuric acid, levels of these compounds reached 1,837 and 1,314 µg/g respectively. In the same manner, diferulates (and their amides) are released as free molecules after hydrolytic (or ammonolytic) pretreatment [18]. In the latter case, diferulates have been detected with zero, one or both carboxyl groups converted to amides formed by ammonolysis.

In parallel to findings of association of triclin with grass lignin (described in Chapter 2) other small molecules containing phenolic acids (less than or around 1 kDa) were also abundant in the grass biomass, not including diferulates. Some of those compounds were characterized in Chapter 2 and some recent reports [24] because they were also triclin-containing substances. The high abundances of these compounds in biomass extracts raises questions about how these substances are beneficial to the plant, and whether these benefits derive from their incorporation into lignin.

Esters of hydroxycinnamates (including *p*-coumarate and ferulate) have long been recognized as abundant components of grass cell walls. While the role of ferulates in forming crosslinks between arabinoxylan glycopolymers is well established [25-27], the importance of *p*-coumarate esters is less well understood [28], in part because oligomers of *p*-coumarates have not been observed in significant quantities. Most *p*-coumarate in corn cell walls has been attributed to an esterified form, 9-sinapyl *p*-coumarate [28-30], and it has been proposed that the *p*-coumaroyl unit in sinapyl *p*-coumarate does not have a direct role in radical coupling reactions [28] owing to its lower reactivity.

Being certain that phenolic acid derivatives exhibit great abundance in many monocots, this chapter presents a deeper approach to profile and identify novel phenolic acid derivatives in grass lignin. As mentioned above, many reports have shown presence of *p*-coumarates in lignin based on signals in NMR spectra, however whether *p*-coumarate plays a role in cross-linking (similar to ferulates) has remained unclear. A key aim of this research has been to use metabolomic analyses to establish roles of phenolic acid precursors to higher molecular mass constituents in monocot cell walls. Another carbohydrate reported to be esterified to phenolic acids in grasses is glycerol [31-33]. In this chapter employment of an untargeted approach (multiplexed CID, discussed in Chapter 2) has been applied to drive discoveries regarding the presence and importance of phenolic acids esterified to glycerol.

Finally, esterification of *p*-coumarate to monolignols will be shown to find its way into larger lignin molecules (>1kDa). Here, UHPLC, accurate high resolution MS, and MS/MS are applied to answer questions about how these precursors are incorporated into lignin. Here we focus on extractives from corn stover while comparing the major findings with other grass biomass sources.

3.3 Experimental

Materials. HPLC grade methanol and HPLC grade hexanes were purchased from Sigma-Aldrich and JT Baker respectively. Untreated biomass samples were provided by the laboratory of Professor Bruce Dale of Michigan State University. Each biomass source was milled, and particles passing through a 1 mm mesh were collected using a sieve. Accelerated solvent extraction (ASE) was performed using a Dionex ASE system. Stainless steel ASE cartridges (33-mL) were packed with 5 g of milled biomass and extracted 3X with 28 mL of hexane to remove lipids, followed by 3X extraction with 28 mL of methanol. Each extraction was performed at 100° C and 1500 psi. Methanol extracts were then pooled and diluted 10X in methanol before analysis using UHPLC-MS. Crude EA extractives of each biomass were produced by the Dale laboratory. Tricin standard was purchased from ChromaDex (Irvine, CA).

UHPLC-MS analyses were performed using a Waters G2-S QToF mass spectrometer equipped with an Acquity pump, model 2777C autosampler, and Acquity Column Manager. Separations were performed using a fused core Supelco Ascentis Express C18 column (100 x 2.1 mm; 2.7 μ m particles). Gradient elution was performed using solvent A (0.1% aq. formic acid) and solvent B (methanol) at a total flow rate of 0.3 mL/min. Linear gradient conditions (A/B) were: initial, hold to 1.0 min (99/1), followed by increase in B to (1/99) at 30 min. Electrospray ionization was used for all analyses. Nonselective multiplexed CID mass spectra [34] were acquired in both positive and negative ion modes by quasi-simultaneous switching of the collision cell voltage through 5 different values (5, 25, 40, 55, and 80 V), accumulating transients for 0.1 seconds per function. Mass spectra were acquired in centroid peak mode for each function, and

leucine-enkephalin was introduced as a lock spray reference, with automatic mass correction. UHPLC-MS data were processed using Waters MarkerLynx XS software, which performed extracted ion chromatogram peak detection, retention time alignment, and peak integration. Quantification of tricin flavonolignans was performed by analysis of tricin standard solutions prepared at 0.01, 0.05, 0.1, 0.5, 1.0, and 5.0 μM concentrations, and linear calibration curve generation. All flavonolignan concentrations were calculated based on the assumption that their molar response factors were identical to tricin.

Profiling of methanolic extracts of untreated corn stover was performed using UHPLC-MS (Figure 3.2) using negative-ion mode multiplexed collision-induced dissociation in 5 collision energy functions to generate fragment ion masses [34]. Using this method, compounds elute from the UHPLC column and are ionized, and these ions are accelerated and transported without mass selection into a region inside the mass spectrometer where the gas pressure is sufficient ($\sim 10^{-3}$ mbar) to ensure that each ion undergoes collision with multiple gas molecules. These collisions convert translational energy into internal vibrational energy, which results in formation of fragment ions if sufficient energy is deposited into molecular vibrations. All ions are then transported to the mass analyzer, where they are separated by m/z , and the number of ions at each m/z is counted. The resulting ion counts are digitized and stored, and various chromatograms, e.g. either the Total Ion Count (TIC) of all ions, or extracted ion chromatogram (XIC) of a single or limited range of ions of user-selected m/z values, can be calculated and displayed.

3.4 Results and Discussion: Pathways of Incorporation of Phenolic Acids into Grass Lignin

3.4.1 Hydroxycinnamoyl (*p*-Coumaroyl and Feruloyl) Glycerols

The strongest signal in the UHPLC-MS profile of corn stover extractives was detected in negative ion mode at m/z 413 eluting at 16.29 min, with a less abundant signal eluting at 16.03 min (Figure 3.2.D). Mass spectra of these peaks at the lowest collision energy (5 V) contained ions at m/z 413.124, annotated as $[M-H]^-$ ions and consistent with two isomers of neutral formula $C_{22}H_{22}O_8$ (theoretical m/z 413.12419). Mass spectra generated at 40 V collision potential exhibited ions at m/z 163 and 193, suggestive of *p*-coumarate and ferulate groups, and the MS/MS product ion spectra for m/z 413 (Figure 3.3) confirmed that these fragments were generated from both of the two resolved isomers.

The strongest signal in the UHPLC-MS profile of corn stover extractives was detected in negative ion mode at m/z 413 eluting at 16.29 min, with a less abundant signal eluting at 16.03 min (Figure 3.2.D). Mass spectra of these peaks at the lowest collision energy (5 V) contained ions at m/z 413.124, annotated as $[M-H]^-$ ions and consistent with two isomers of neutral formula $C_{22}H_{22}O_8$. Mass spectra generated at 40 V collision potential exhibited ions at m/z 163 and 193, suggestive of *p*-coumarate and ferulate groups, and the MS/MS product ion spectra for m/z 413 (Figure 3.3) confirmed that these fragments were generated from both of the two resolved isomers.

Purification of the more abundant compound by semi-preparative HPLC yielded a substance whose structure was determined by NMR spectra obtained in HNMR, HSQC, HMBC, and COSY (Table 3.1) to be 1-*p*-coumaroyl-3-feruloylglycerol, in agreement with a previous report that suggested presence of this compound in another monocot, *Tillandsia streptocarpa*, [35] using NMR and in maize using MS/MS [33].

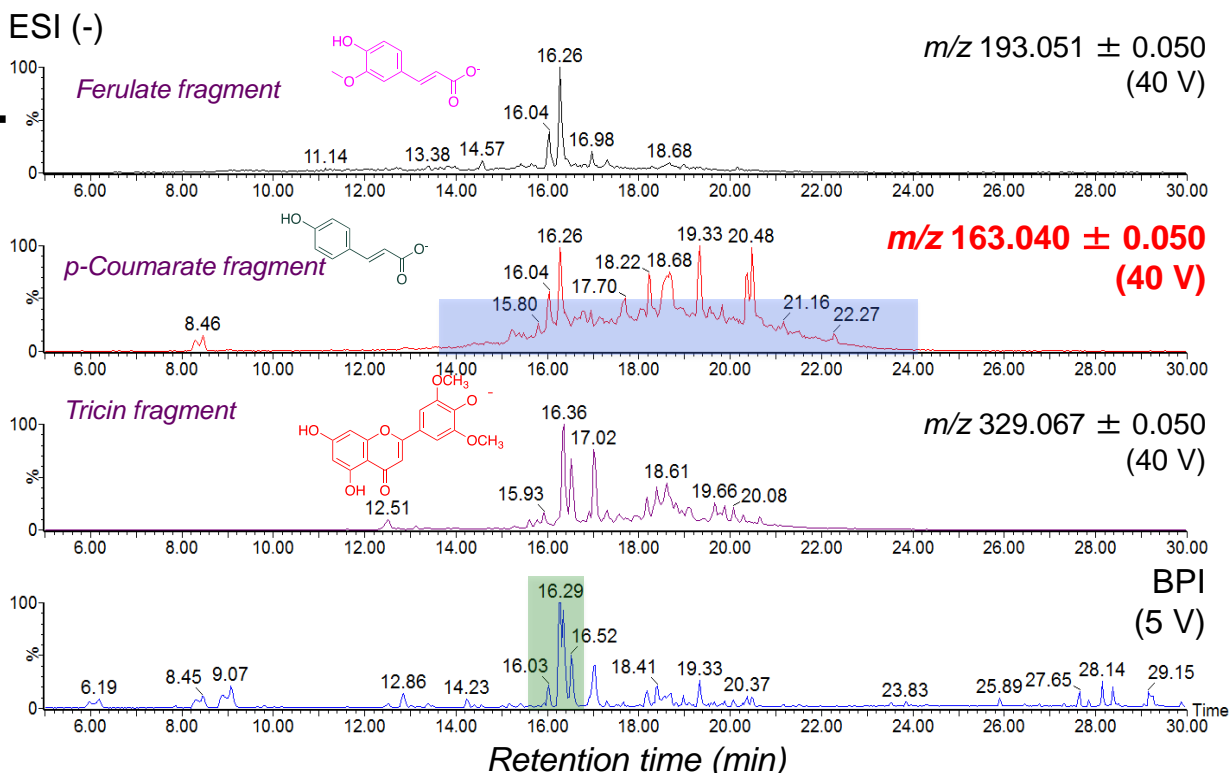


Figure 3.2. Results from UHPLC-MS profiling of methanolic extract of untreated corn stover. Extracted ion chromatograms (XICs) at 40 V collision potential for (A) ferulate fragment ion at m/z 193.051 (B) *p*-coumarate fragment ion at m/z 163.040, (c), triclin fragment ion at m/z 329.067, and (D) base peak ion (BPI) chromatogram at the lowest collision potential (5 V).

Extracted ion chromatograms (XICs) were generated for ions at m/z 193.051 (ferulate fragment ion, Figure. 3.2.A), 163.04 (*p*-coumarate fragment ion, Figure 3.2.B), and 329.067 (triclin fragment ion, Figure. 3.2.C).

Consistent with what has been reported about triclin incorporation into monocot lignins [24 , 36], numerous peaks in the XIC (Figure 3.2.C) display spectra at elevated collision energies that contain ions at m/z 329, 314, and 299. It is worth noting that the XIC for m/z 329 does not return to the baseline across the retention time range of 16-22 minutes, even though individual chromatographic peaks are narrow. The results in Chapter 2 revealed triclin incorporation across a wide range of molecular masses, and the chromatographic results suggest that many are not

resolved by the chromatographic separation despite the use of an ultrahigh performance fused core column.

In similar fashion, perhaps the most striking observation revealed that the XIC for *p*-coumarate fragment (m/z 163) formed at elevated collision potential (Figure 3.2.B) remained well above the baseline over the retention range of 14-24 minutes, and the elevated unresolved signal reached levels greater than 20% of the signal of the most abundant individual compound that displayed this signal. Integration of the individual peaks and the total signal suggested that the majority of *p*-coumarate derivatives detected in this manner can be attributed to unresolved substances. We interpret this result as evidence that a large number of unresolved *p*-coumarate esters elute in this range of retention times, perhaps hundreds to thousands of individual chemical forms.

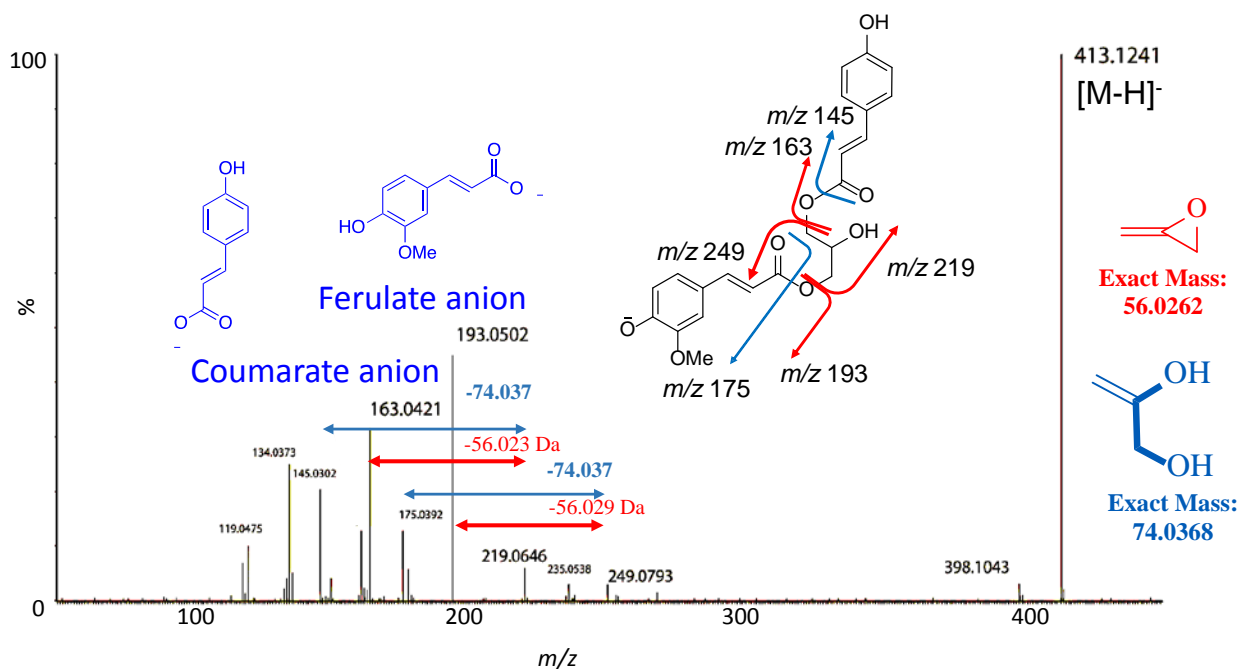


Figure 3.3. MS/MS spectrum of product ions of m/z 413.124 from 1-*p*-coumaroyl-3-feruloylglycerol in a methanolic extract of corn stover, with proposed assignments of product ions and neutral mass losses.

Position	δ_H	δ_C
1	4.28 (2, d, 5.0 Hz)	66.4
2	4.16 (1, m)	68.6
3	4.28 (2, d, 5.0 Hz)	66.4
1'	-	127.7
2'	7.16 (1, s)	112.0
3'	-	149.4
4'	-	148.9
5'	6.8 (1,d, 8 Hz)	116.1
6'	7.07 (1,d,8 Hz)	124.4
7'	7.67 (1, d, 16 Hz)	147.4
8'	6.36 (1, d, 16 Hz)	115.3
9'	-	167.5
1''	-	127.2
2''	7.45 (1, d, 8 Hz)	131.3
3''	6.79 (1, d, 8 Hz)	115.4
4''	-	162.7
5''	6.79 (1, d, 8 Hz)	115.4
6''	7.45 (1, d, 8 Hz)	131.3
7''	7.68 (1, d, 16 Hz)	147.4
8''	6.40 (1, d, 16 Hz)	115.3
9''	-	167.5
OCH₃	3.88 (3, s)	56.5

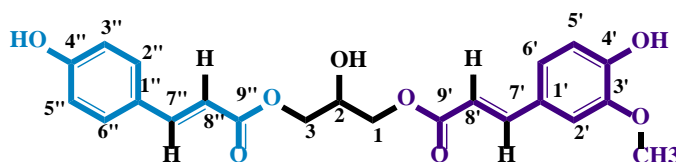


Table 3.1. NMR Shift Assignments for 1-*p*-coumaroyl-3-feruloylglycerol.

Observation of phenolic acid carboxylate fragments was not limited to *p*-coumarate. The extracted ion chromatogram for ferulate fragment ion (m/z 193.051, Figure 3.2.A) also shows an elevated baseline but at a lower level relative to the most abundant signal (from 1-*p*-coumaroyl-3-feruloylglycerol, eluting at 16.26 min), suggesting that ferulate ester diversity is less than for *p*-coumarate esters and tricinn derivatives or that ferulate derivatives undergo further crosslinking

that precludes formation of the m/z 193 fragment ion following collisional activation. This finding is consistent with the dominance of *p*-coumarate esters in maize lignin as assessed using NMR spectroscopy [28-29, 37].

The high abundance of phenolic acid esters of glycerol (PAEGs) and the striking diversity of *p*-coumarate derivatives in corn stover extractives raises the possibility that phenolic acid esters, including glycerol esters, might be incorporated into lignin. Though earlier investigations suggested incorporation of *p*-coumaroyl esters of monolignol alcohols (*e.g.* syringyl alcohol)[29] into lignin, investigations have largely ignored incorporation of phenolic acid esters of glycerol or other polyols except for a paper by Grabber and co-workers in 2010 [38]. Common lignin degradation reactions often employ conditions likely to cleave phenolic acid esters, and regions of lignin NMR spectra with resonances attributed to glycerol overlap other polyhydroxylated phenylpropanoids. Purification and elucidation of structures of the enormous number of *p*-coumarate esters is not feasible, and the LC/MS profiles (Figure 3.2) suggest that individual forms may be numerous but present in multiple isomeric forms, with each of low abundance. Despite the prolific efforts of the Ralph group to synthesize lignin oligomers [24], it is not clear that they have incorporated PAEGs into synthetic oligomers. As a result, [39]reliance on mass spectrometry for structure annotation of lignin constituents becomes essential, and more information is needed about how lignin constituents containing esters fragment upon collision-induced dissociation.

To serve as the foundation for understanding collision-induced dissociation (CID) of *p*-coumarate esters, we began by evaluating the behavior of 1-*p*-coumaroyl-3-feruloylglycerol upon CID by generating MS/MS product ion spectra of $[M-H]^-$ in negative ion mode (Figure 3.3). Accurate measurements of product ion masses revealed product ions 56.02 Da heavier than

each of the abundant *p*-coumarate and ferulate carboxylate fragments (m/z 163 and 193), corresponding to C_3H_4O (doubly dehydrated glycerol). A similar finding was observed in MS/MS product ion spectra of glycerol mono *p*-coumarate (Figure 3.4) and glycerol monoferulate, which exhibited product ions 74.037 Da heavier than the carboxylate anions, corresponding to singly dehydrated glycerol ($C_3H_6O_2$) remaining attached to the phenolic acid. It is important to note that the structures corresponding to the neutral mass losses within Figures 3.3 and 3.4 cannot be deduced with certainty from MS/MS spectra alone, and are shown to depict the portions of the glycerol ester lost upon CID. Since the 56 and 74 Da losses are neutral fragments, they are not detected in the MS/MS spectra. Both eluting isomers of 1-*p*-coumaroyl-3-feruloylglycerol (1,2- vs. 1,3- substitutions) yielded indistinguishable MS/MS spectra, so the position of esterification in mono *p*-coumaroylglycerol was not distinguished with certainty using MS/MS alone. A few reports have relied solely on MS/MS data to suggest presence of this compound in different monocots including sorghum [39-40], maize [33], and *Anans* [41].

Glycerol esterified by feruloyl or *p*-coumaroyl groups or both has been reported previously in extracts of grasses [35 , 39 , 42]. Xiong and coworkers characterized *p*-coumaroyl *trans*-feruloyl glycerol in rhizomes of the aquatic grass *Sparganium* [42]. *p*-Coumaroyl-feruloylglycerol has been assigned in sorghum from MS/MS spectra [40] and in maize where its levels were elevated following biotic stress [33]. Glycerol esterified by both *p*-coumaroyl and feruloyl groups have been reported in maize [33] with compound annotation based on HRMS/MS also reported in other monocots using NMR characterization [35]. This report is the first to use NMR to confirm assignment of this compound in maize.

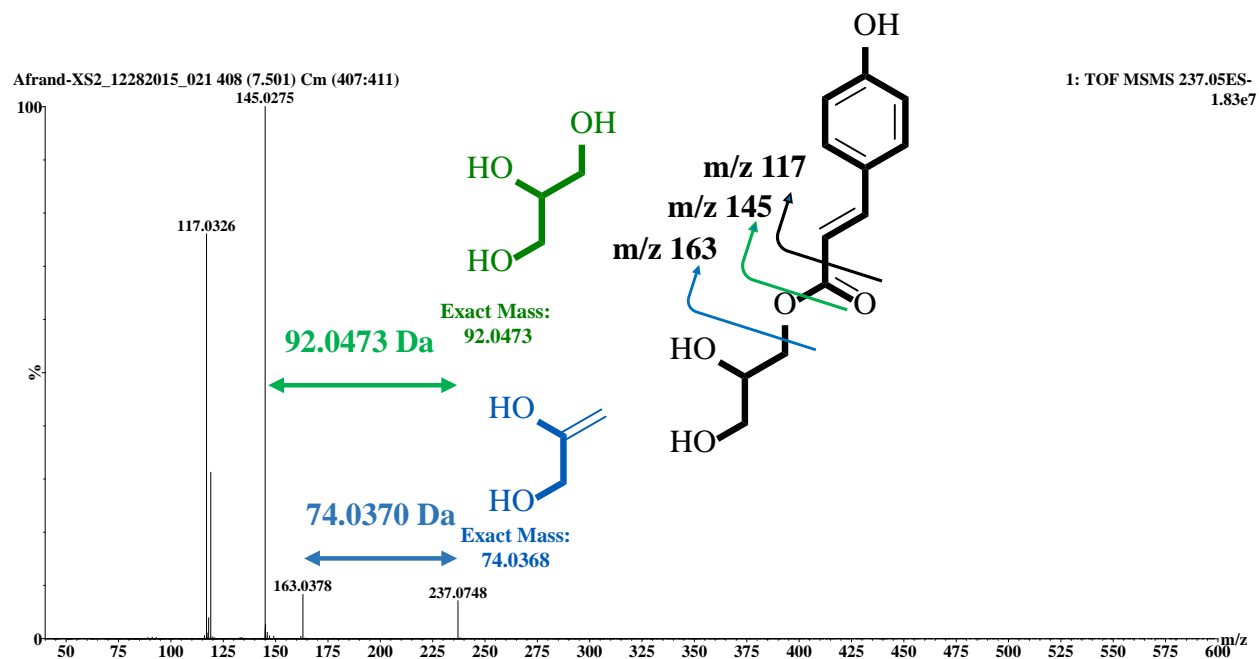


Figure 3.4. MS/MS product ion spectrum of m/z 237.07 (1-*p*-coumaroylglycerol) from a methanolic extract of untreated corn stover, with assignments of product ions and neutral mass losses.

Multiple forms of phenolic acid esters of glycerol (PAEGs) (different positional isomers and phenolic acid groups) were present in extracts of all the monocot plants examined, based on XICs of m/z values corresponding to their $[M-H]^-$ ions. (Figure 3.5). The two 1-*p*-coumaroyl-3-feruloylglycerol isomers were about an order of magnitude more abundant in corn stover extracts than other grasses, and these PAEGs were below detectable levels in extracts of the hardwood poplar (Figure 3.6).

The universal presence of PAEGs in extracts of all monocots investigated in this study, combined with similarities in structure of the phenylpropanoid acid moieties to classical lignin monomers, suggests these esters might become incorporated as components of lignin through chemical transformations similar to traditional monolignols including oxidative dehydrodimerization. Evidence for formation of such dimers was evident during analysis of a corn stover extract from an extracted ion chromatogram for m/z 825.24, which corresponds to

[M-H]⁻ of the dehydrodimers formed by coupling of two molecules of 1-*p*-coumaroyl-3-feruloylglycerol (Figure 3.7). The XIC shows evidence for at least 20 isomers, as expected given multiple diferulate forms [23] and the presence of both 1,2- and 1,3-substituted *p*-coumaroyl-feruloyl glycerol. It cannot be ruled out that coupling might occur through the *p*-coumarate ester group, though it is anticipated that the additional methoxy group in feruloyl esters should confer greater reactivity. General understanding in the scientific community about diferulates is that they serve as crosslinkers between hemicellulose strands or between hemicellulose and lignin strands [23].

Annotation of one of the more abundant isomers of *p*-coumaroyl feruloyl glycerol is putatively assigned as the 8-*O*-4 linked dimer because this is usually the most abundant diferulate form [17], was supported from its MS/MS spectrum (Figure 3.8) similarities to diferulate MS/MS spectra [23]. Since the exact location of the linkage between phenolic acids and glycerol is not assigned in all isomers of this compound, any isomer of it will be referred as bis-(*p*-coumaroylferuloylglycerol) (without assignment of specific positions of esterification or crosslinking).

All peaks shown in the LC-MS/MS BPI chromatogram of *m/z* 825 precursor ion shown in right panel of Figure 3.7 result in the following product ions that support their assignments as *p*-coumaroylferuloylglycerol dimers: *m/z* 413 ([monomer-H]⁻), 415 (2H+[monomer-H]⁻), and 397 ([monomer-H]⁻-O) . With this evidence that glycerol esters of ferulate undergo dimerization, two crucial questions emerge: (1) are such molecules involved in cross-linking in monocot cell walls similar to what is observed in diferulate crosslinking of arabinoxylans previously [23] and (2) does this contribute to the physical properties of secondary cell walls conferred by lignin and influence digestability? If the answers to these question are yes, such findings might suggest

potential for engineering designer lignins in plants by engineering higher levels of feruloylglycerols, metabolic engineering of plants with new cross-linkers with longer and/or more flexible substructures, or incorporation of diferulates in larger phenolic structures.

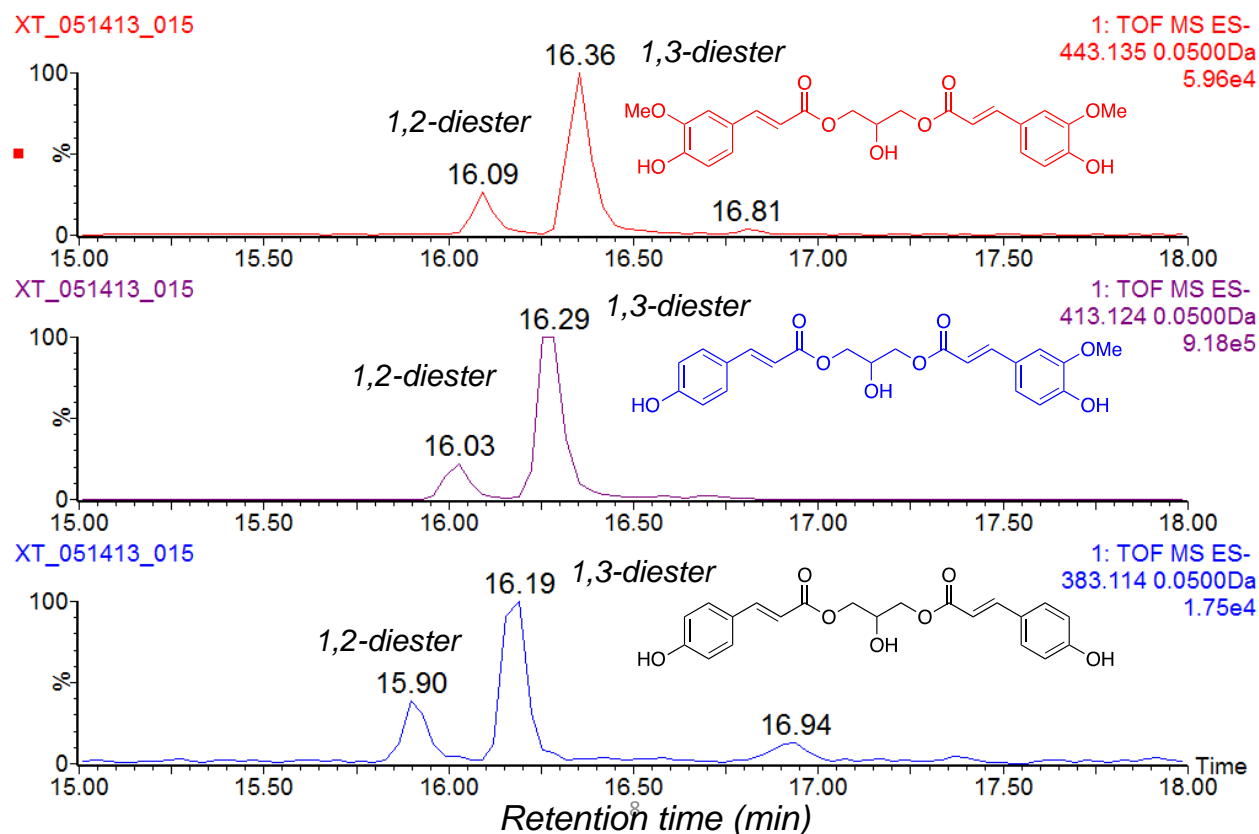


Figure 3.5. Narrow mass window UHPLC-MS extracted ion chromatograms of a methanolic extract of untreated corn stover showing: (A) m/z 443.135 \pm 0.05 (diferuloyl glycerol), (B) m/z 413.124 \pm 0.05 (*p*-coumaroyl-3-feruloylglycerol), and (C) m/z 383.114 \pm 0.05 (di-*p*-coumaroyl glycerol).

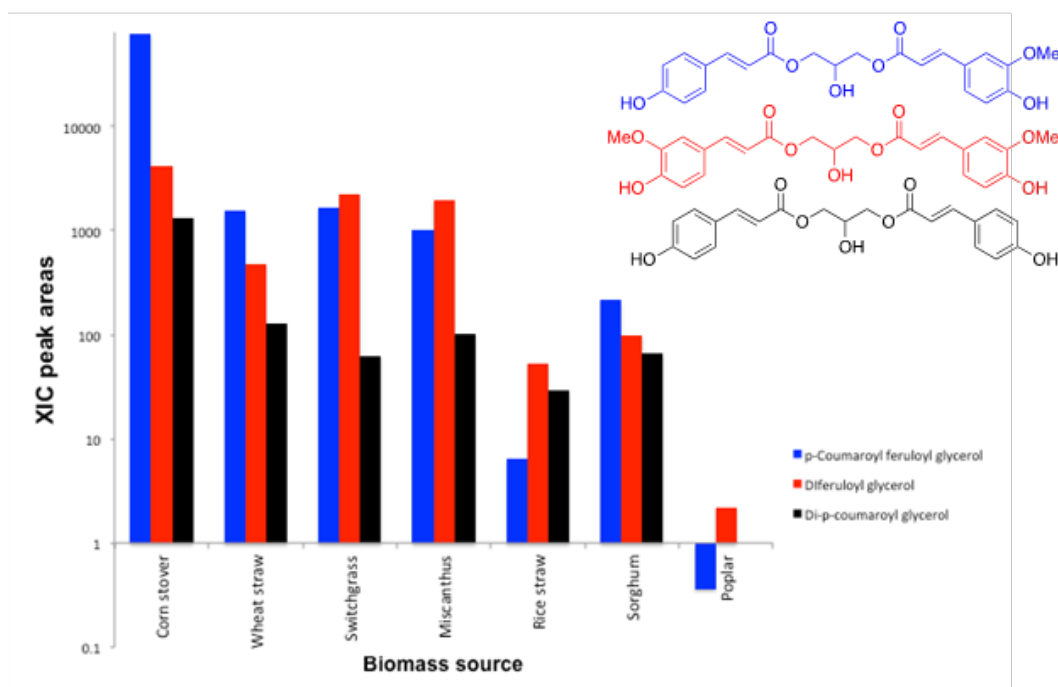


Figure 3.6. Extracted ion UHPLC-MS chromatogram peak areas for 1-*p*-coumaroyl-3-feruloylglycerol, diferuloyl glycerol, and di-*p*-coumaroyl glycerol for methanolic extracts of grasses and poplar. Note the logarithmic scale for the vertical axis. PAEGs are abundant in grasses but insignificant in poplar.

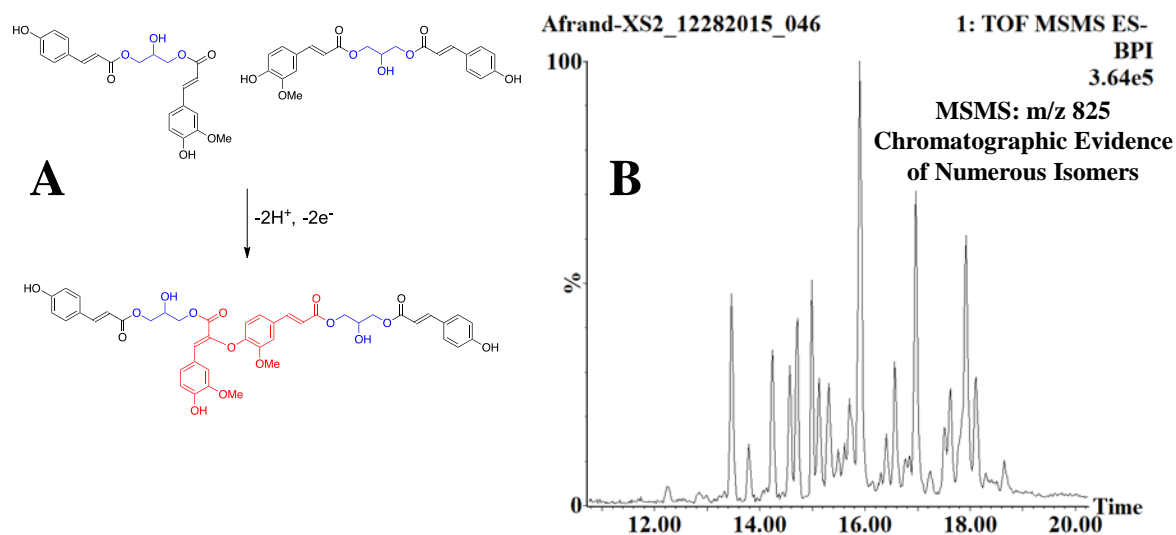


Figure 3.7. (A) Oxidative coupling of two 1-*p*-coumaroyl-3-feruloylglycerol to yield the 8-*O*-4 diferulate-linked dimer at *m/z* 825.24 and (B) UHPLC-MS/MS chromatogram for *m/z* 825.24 for a methanolic extract of corn stover showing peaks attributed to multiple isomeric forms.

In order to establish structures of a diferulate cores in bis-(*p*-coumaroylferuloylglycerol) (detected as $[M-H]^-$ as m/z 825, itself having multiple isomers) it is necessary to show connection of two ferulate groups with any possible diferulate topology [23] in at least one of the isomers of bis-(*p*-coumaroylferuloylglycerol). For more conclusive evidence of structure, it is desirable to purify or synthesize different isomers of bis-(*p*-coumaroylferuloylglycerol) and obtain NMR spectra of each isomer. However, these compounds are of low abundance compared to other abundant phenolic compounds present in methanol extracts of corn stover. Also, as discussed in Chapter 2 and Chapter 5, complexity of this extract and multiplicity of large yet low-abundance phenolic compounds that often co-elute together makes purification and/or synthesis of all individual >20 isomers of this compound cumbersome, time-consuming, and expensive.

A powerful tool to examine presence or absence of diferulate core in structure of bis-(*p*-coumaroylferuloylglycerol) isomers, yet without assignment of specific linkages is HRMS/MS. To do so, we rely on the precedent that oxidative dimerization of two ferulates would generate only limited combinations of molecular masses. MS/MS of different isomers should be evaluated, hunting for those mass combinations in characteristic fragment ions, examples of which are shown in Figure 3.9. The MS/MS product ion spectrum for $[M-H]^-$ of one of the abundant isomers of bis-(*p*-coumaroylferuloylglycerol) eluting at 17.9 minutes is shown in Figure 3.8. To contrast origins of ferulate groups from each individual *p*-coumaroylferuloylglycerol units, red and blue colors are used in structures represented in both Figure 3.8 and 3.9. Some abundant product ions derived from the m/z 825 precursor (bis-(*p*-coumaroylferuloylglycerol)) include m/z 413 (deprotonated *p*-coumaroylferuloylglycerol), m/z 163 for coumarate anion, and m/z 679 (neutral loss of 146 Da, “precursor minus coumaroyl” group) do not yield direct information that ferulate units are coupled to one another. However,

product ions that result from more extensive fragmentation including m/z 533, 515, 497, 441, and 341 provide strong evidence that two ferulates coupled to each other. Examples of corresponding fragments are shown in Figure 3.9.

This finding supports the hypothesis that diferulate links form between feruloyl esters of glycerol. We conclude that measurements of diferulates released by hydrolysis or ammonolysis may reflect more than crosslinks between arabinoxylan chains.

3.4.2 Hydroxycinnamoyl (*p*-coumaroyl and feruloyl) in Tricin Derivatives

Beyond formation of diferulate-linked glycerol esters, we hypothesized that PAEGs also undergo radical addition by initiators of lignification, including the flavonoid triclin. As shown by Lan et al. [24] and in Chapter 2 of this report, phenolic acids are bound to triclin derivatives, often through incorporation of esters such as sinapyl *p*-coumarate. An abundant example is T-(4-O- β t)-G", where G" is *p*-coumaroylguaiacylglyceryl unit and T stands for triclin.

Here, we hypothesize that hydroxycinnamoylglycerol units are also directly coupled to larger molecules including triclin via chemistry common to known lignin biosynthetic reactions. To test this hypothesis, XICs were generated for m/z 741.19, which corresponds to $[M-H]^-$ for the product expected from addition of triclin to *p*-coumaroyl feruloyl glycerols, expecting the initial product of radical addition is quenched by hydrogen abstraction to regenerate a carbon-carbon double bond. The MS/MS product ion spectrum for the XIC peak with strongest signal shows characteristic fragment ions that support the proposed structure (Figure 3.10). The most abundant product ions are m/z 329, 314, and 299 that are characteristic of the triclin group and losses of one and two methyl radicals. The basis for assignment of triclin bound to ferulate is that the expected ferulate carboxylate ion at m/z 193 was not detected while the unmodified *p*-coumarate carboxylate is clearly detected at m/z 163. Additional support for these conclusions comes from

the product ion at m/z 577.15, attributed to loss of neutral *p*-coumaric acid (minus H₂O). Also, the loss of neutral triciny ferulic acid leaves m/z 219 (C₁₂H₁₁O₄⁻) that is detected in the spectrum (though not labeled in Figure 3.10), and the fragment at m/z 503.11 corresponds to deprotonated triciny anhydroferulate (C₂₇H₁₉O₁₀⁻; theoretical m/z 503.10) support annotation as a triciny-substituted ferulate. This example demonstrates that PAEGs are incorporated into higher molecular mass lignin constituents via both mechanisms of dehydrodimerization and radical addition to phenolic acid esters.

3.4.3 Couplings of Sinapyl *p*-Coumarates to Make Larger Lignin Molecules

In addition to oligomerization reactions involving PAEGs, oxidative coupling of phenolic acid esters of monolignols provides an additional pathway for incorporation of *p*-coumarate esters into higher molecular mass lignins.

Though such couplings of monolignol esters of hydroxycinnamates have been established [30], it has usually been proposed that the chemistry involves reactions on the monolignol moieties. In the current study, the most abundant corn stover extractive not containing triciny or glycerol moieties is proposed to be a compound of formula C₄₀H₄₀O₁₃ (observed at m/z 727.24 in the negative-ion LC/MS data; theoretical m/z 727.23961). We propose that this forms by oxidative coupling of two molecules of sinapyl *p*-coumarate, in chemistry analogous to the coupling of PAEGs, to form multiple isomers of substituted tetrahydrofuran derivatives termed “bis-sinapyl *p*-coumarates”. One of these isomers was purified by HPLC and subjected to NMR analysis. This structure has not been reported to the best of our knowledge as a plant metabolite. Figure 3.11 shows the structure of the most abundant isomer of this compound assigned by NMR, and Table 3.2 lists the ¹H and ¹³C NMR shifts that support assignment of this compound, and these assignments were supported by COSY, HSQC, and HMBC spectra. Absolute

stereochemical assignments were not established owing to the small amounts of material available. Proposed structure here, is in consistent with the suggested presence of “*p*-coumaroylated syringyl” units in maize [29] revealed by thioacidolysis. However thioacidolysis does not retain the linkages reported here, hence the intact molecule has not been detected using thioacidolysis [29].

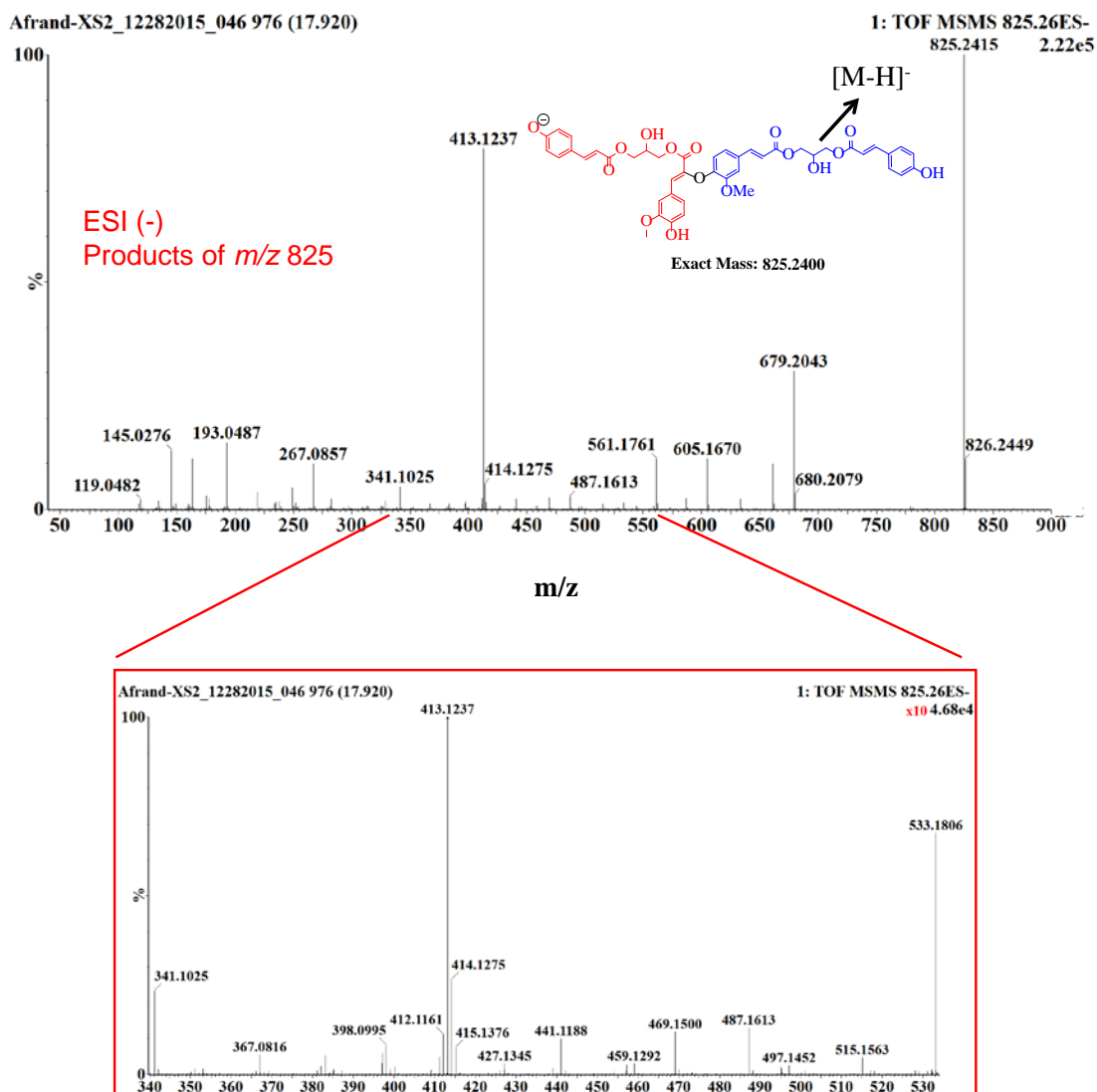


Figure 3.8. Annotation of one of the bis-(*p*-coumaroylferuloylglycerol) isomers of detected in an extract of corn stover (m/z 825.24) eluted at 17.9 minutes using high resolution MS/MS. The zoomed in inset shows the key fragments resulting in deduction of diferulate core in the center of the structure (shown in Figure 3.9).

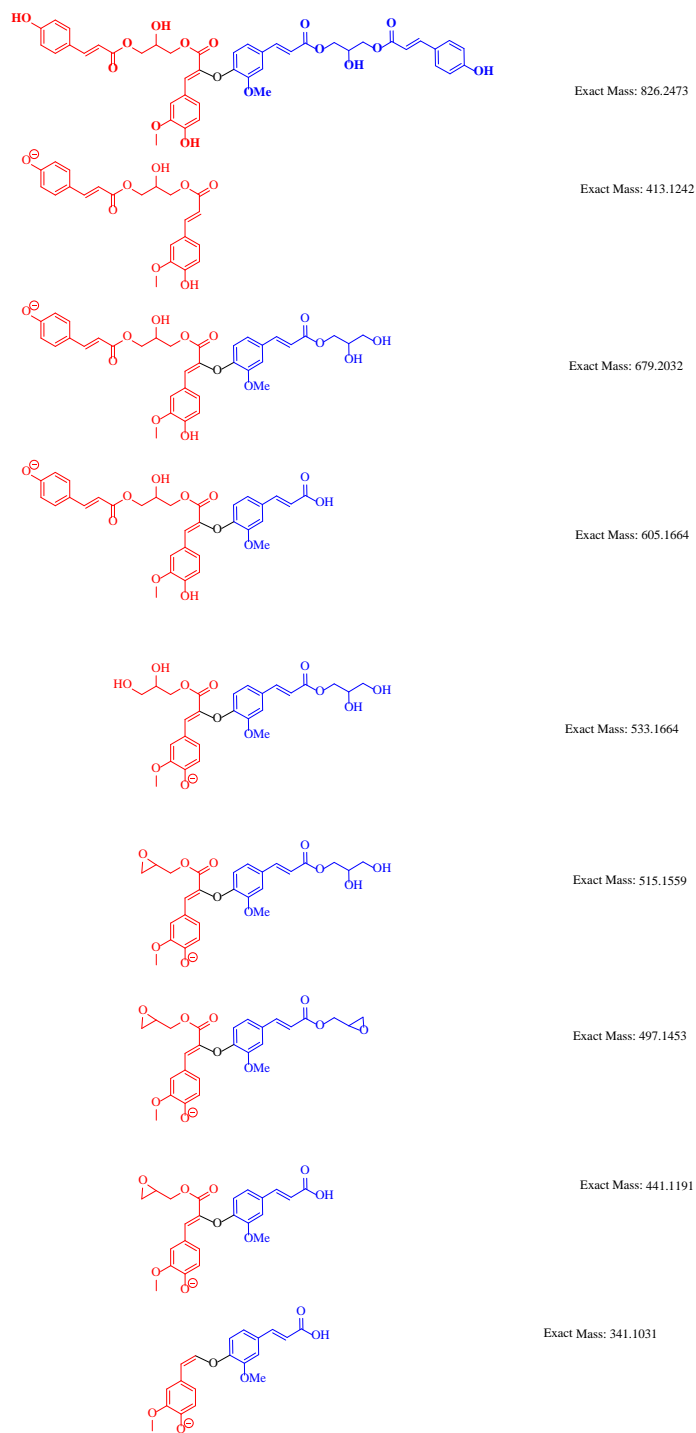


Figure 3.9. A series of key fragment ions (starting from precursor at m/z 825 to diferulate-derived m/z 341) that are useful to deduce presence diferulate core at the center of the structure of one of the isomers of bis-(1-*p*-coumaroyl-3-feruloylglycerol), corresponding MS/MS spectrum is in the zoomed-in outset of Figure 3.8 showing detection of all these fragment ions.

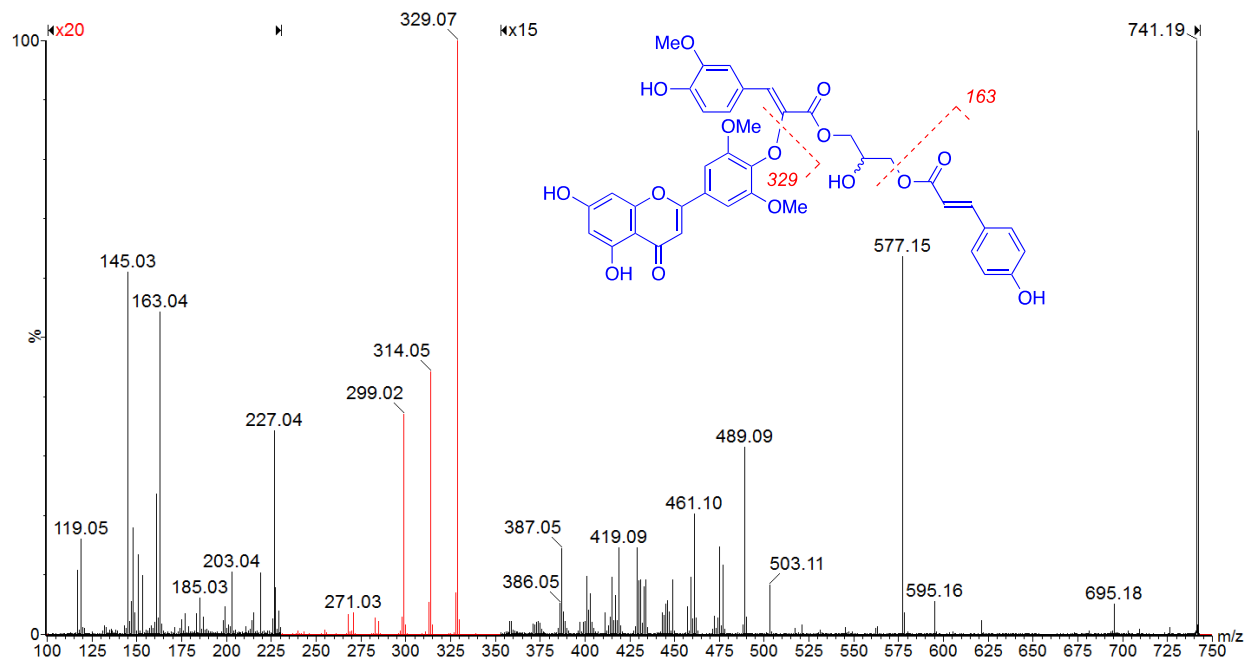


Figure 3.10. Product ion MS/MS spectrum of m/z 741 for a product in methanolic extract of corn stover, annotated as a conjugate of tricetin with *p*-coumaroyl-feruloylglycerol. Peaks in the spectrum from m/z 100-230 are magnified by a factor of 20, and from m/z 350-750, magnified by a factor of 15.

It is known that two sinapyl alcohol molecules undergo 8–8 coupling during radical lignification, of both C7 atoms to O-9 atoms from alcohol groups in sinapyl alcohol. this leads to formation of two fused oxygen-containing 5-membered rings (two THF rings) often referred as pinoresinol, mechanism of which is shown in Figure 3.12 [43-44].

In case of sinapyl *p*-coumarate (Figure 3.13), however, work up of radical reaction is slightly different because both the oxygen atoms attached to C9 are occupied via esterification to *p*-coumarates. Hence, once radical addition between the two C8 atoms of individual sinapyl groups occurs, the radical which is then stabilized on one the sinapyl C7 (benzylic) atoms does not have the option to cross-couple to O-9 of the other sinapyl group (and form THF ring) anymore. The most likely scenario for this benzylic radical in the oxidative environment is to add dioxygen to form a peroxy radical and then undergo cleavage of O-O bond to form aryl –O.

radical (labeled with red colored font in Figure 3.13). This radical that is now stabilized on the new oxygen atom can undergo a similar ring-enclosure, but in this case to C7 of the other sinapyl group, unlike ring closure that happens with free sinapyl alcohols where C7 of one cross-couples to O-9 of the other. As a result only one THF ring can form between two units of sinapyl *p*-coumarates. To the best of our knowledge cross-coupling of two C7 atoms of sinapyl alcohol, which is an alteration from the known lignin-crosslinking mechanisms enforced by esterification of C9 atoms to hydroxycinnamic acids, have not been reported before.

Position	δH	δC
1	-	133.4
2	6.75 (1, s)	103.6
3	-	149.5
4	-	136.6
5	-	149.5
6	6.75 (1, s)	103.6
7	5.03 (1,d, 8 Hz)	84.1
8	2.67 (1, m)	50.2
9	4.43 (2, dd, 13.5 Hz, 4 Hz)	63.1
O-CH3 (3 &5)	3.82 (6, s)	55.2
1'	-	133.4
2'	6.75 (1, s)	103.6
3'	-	149.5
4'	-	136.6
5'	-	149.5
6'	6.75 (1, s)	103.6
7'	5.03 (1,d, 8 Hz)	84.1
8'	2.67 (1, m)	50.2
9'	4.43 (2, dd, 13.5 Hz, 4 Hz)	63.1
O-CH3 (3'&5')	3.82 (6, s)	55.2
1''	-	129.8
2''	7.36 (1, d, 8Hz)	129.6
3''	6.77 (1,d, 8Hz)	115.2

Table 3.2. 1H NMR and ^{13}C NMR assignments of the most abundant isomer of compound bis-sinapyl *p*-coumarate in maize with formula $C_{40}H_{40}O_{13}$ (position numbers are shown in Figure 3.11).

Table 3.2 (Cont'd)

4''	-	159.9
5''	6.77 (1,d, 8Hz)	115.2
6''	7.36 (1, d, 8Hz)	129.6
7''	7.43 (1,d, 16Hz)	145.4
8''	6.23 (1,d, 16Hz)	113.2
9''	-	167.4
1'''	-	129.8
2'''	7.36 (1, d, 8Hz)	129.6
3'''	6.77 (1,d, 8Hz)	115.2
4'''	-	159.9
5'''	6.77 (1,d, 8Hz)	115.2
6'''	7.36 (1, d, 8Hz)	129.6
7'''	7.43 (1,d, 16Hz)	145.4
8'''	6.23 (1,d, 16Hz)	113.2
9'''	-	167.4

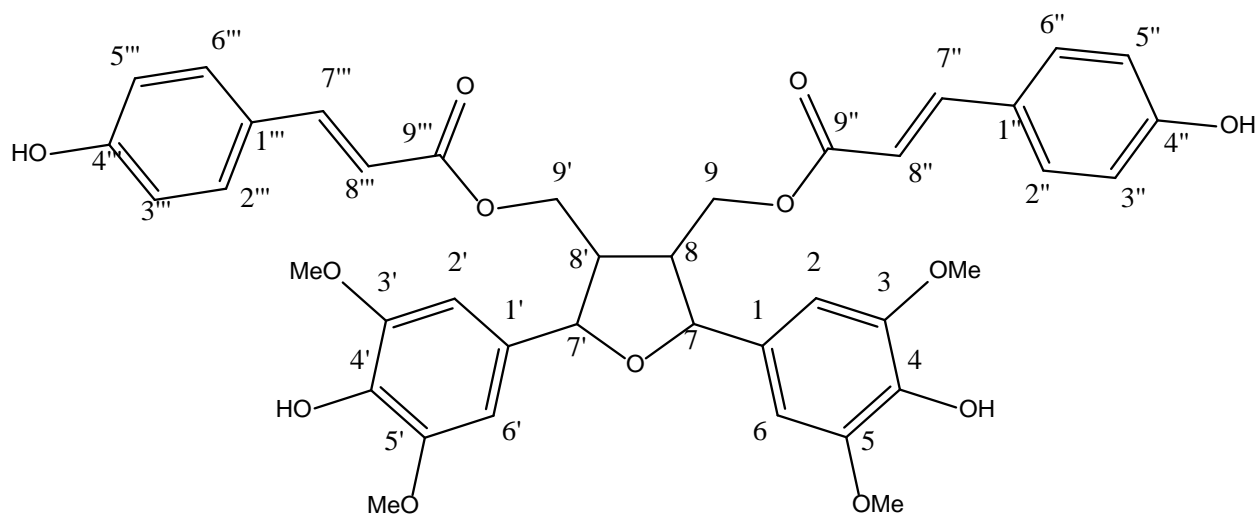


Figure 3.11. Structure of the most abundant isomer of the compound with formula $C_{40}H_{40}O_{13}$ named bis-(sinapyl *p*-coumarate).

Presence of even this single THF ring in sinapyl *p*-coumarate makes the dimer harder to fragment into equal units and negative mode MS/MS of all isomers of m/z 727 that results of *p*-coumarate loss scenario (Figure 3.14) including, m/z 145 (*p*-coumarate minus water), m/z 163 (*p*-coumarate anion), m/z 581 ([precursor – *p*-coumaryl] anion).

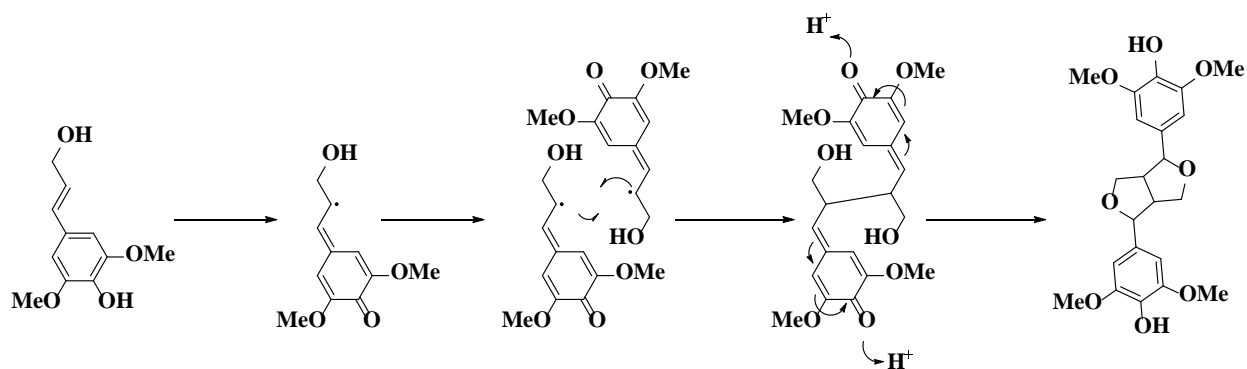


Figure 3.12. Formation of fused resinol rings upon radical coupling of sinapyl alcohols as 8–8 bond [44].

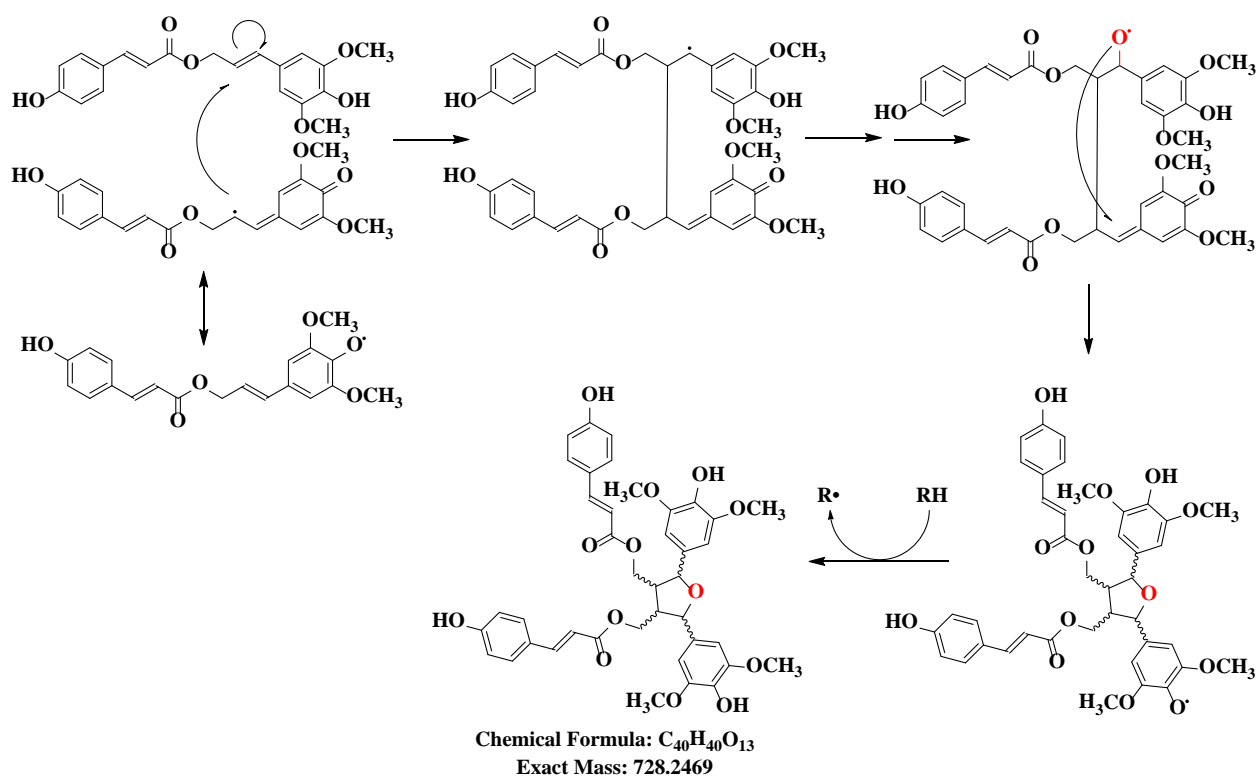


Figure 3.13. Radical coupling of two sinapyl *p*-coumarates followed by formation of 5-membered resinol ring.

This agrees with preservation of bis-sinapyl dimer in the core upon collision in the collision cell of mass spectrometer. This type of linkage expands the realm of *p*-coumarate esters into an even larger group of lignin molecules, as this dimer serves as the lignification platform for larger molecules each having multiple isomers. Following discussion will be focused on such

compounds. Examples include substances detected at m/z 923.3, 1099.3, m/z 1295, and m/z 1471 each having multiple isomers.

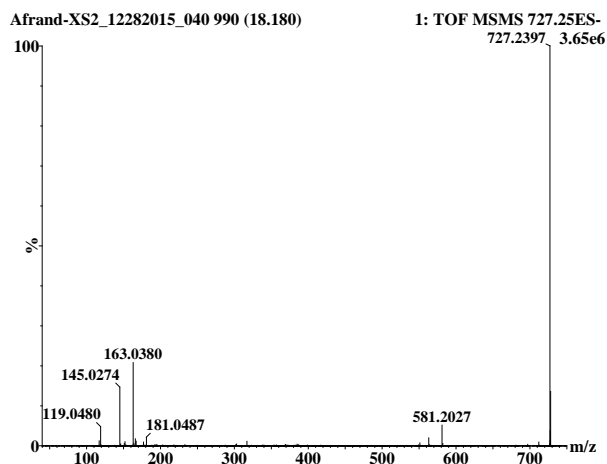


Figure 3.14. MS/MS product ion spectrum of the most abundant isomer of m/z 727, dimer of sinapyl *p*-coumarate.

Evidence for tentative annotation of each of these substances will be followed in this chapter. For the sake of ease of representation sinapyl *p*-coumarate (356 Da) will be shown as (S-H^{''}) and oxidized sinapyl *p*-coumarate (372 Da) will be shown as (S-H^{''})-O and guaiacylglycerol unit (196 Da) will be represented with G^O. Also from now on, bis-(sinapyl *p*-coumarate) with m/z 727 which assignment of its most abundant isomer is discussed above using NMR data will be abbreviated as bis-(S-H^{''}). Figure 3.15 demonstrates example identities of (S-H^{''})-O and G^O units.

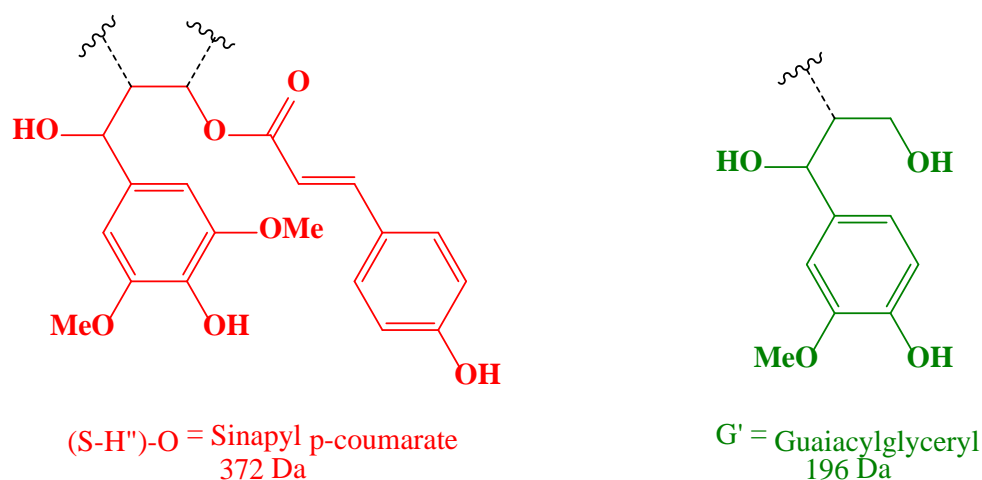


Figure 3.15. Representative structures of (S-H'')-O and G⁰ units.

Extracted ion UHPLC-MS chromatograms provide evidence that multiple isomeric compounds observed at m/z 923.31, annotated as $[M-H]^-$, are formed when bis-(S-H'') (728.2 Da for the neutral molecule) undergoes further addition to coniferyl alcohol and forms a guaiacylglycerol moiety. The MS/MS product ion spectrum and annotated structure of the most abundant isomer of this formula is shown in Figure 3.16 (panel B and C respectively). There are at least 20 chromatographic peaks (Figure 3.16 panel A) denoting separate resolved isomers for m/z 923.31, and MS/MS product ion spectra of all of them show either m/z 727.23 (bis-(S-H'')) or oxidized (m/z 741.25) or hydrated (m/z 743.23) versions. Formation of m/z 727.2 may be explained by a linkage such as 8-O-4 coupling to guaiacylglyceryl group (derived from addition to coniferyl alcohol) where the 8-carbon of the guaiacylglyceryl moiety is connected to oxygen at one of 4- (e.g. 4, 4', 4'', or 4''') carbon positions of bis-(S-H''), hence resulting in facile formation of the bis-(S-H'') fragment ion at m/z 727 upon collisional activation. One can employ a similar argument to explain MS/MS spectra of isomers that yield m/z 741.2 or 743.2 fragments, which are formed via alternative linkages such as 8-8, 5-O-4, 8-5, or 8-O-4 (from 8'', or 8''' of bis-(S-H'') to oxygen of 4-carbon of G⁰). Prominent formation in all isomers of m/z 923.31 of product ions at m/z 727.23, or 741.25, or 743.23, consistent with different linkage forms

mentioned above. In addition, formation of m/z 195.06 and neutral loss of 196 Da, both from G^O moieties provide evidence of G^O moieties attached to the $C_{40}H_{40}O_{13}$ bis-(S-H") cores to make metabolites with molecular mass 924 Da.

That products of more extensive fragmentation of m/z 923 below m/z 727 are only limited to *p*-coumarate ions (m/z 145, m/z 163) and G^O (m/z 195) supports presence of THF ring-containing bis-(S-H") in this structure. The major fragment observed in the MS/MS product ion spectrum of the major isomer of a compound detected in corn stover extract at m/z 1099.36 (deprotonated negative ion) is m/z 727. The 372 Da ($C_{20}H_{20}O_7$) difference between precursor and fragment suggests addition of a third oxidized (S-H")-O unit (356 Da, consistent with sinapyl *p*-coumarate + one oxygen = 372 Da) to the bis-(S-H") with formula of $C_{40}H_{40}O_{13}$. As a result, the compound detected at m/z 1099 is annotated as an oxidized trimer of (S-H") ($C_{60}H_{60}O_{20}$). Despite our failure to isolate sufficient amount of pure compound for NMR characterization, accurate HRMS/MS data guide annotation of this compound using lignin sequencing strategies described by Morreel et al. [44-45]. In the LC/MS/MS chromatogram generated for products of m/z 1099.36 (panel A), the MS/MS product ion spectrum of the most abundant isomer (panel B), and the proposed annotated structure of the compound with formula $C_{60}H_{60}O_{20}$ (panel C; theoretical m/z 1099.36052) are shown in Figure 3.17. The UHPLC-MS/MS base peak intensity product ion chromatogram of m/z 1099.3 shows 4 major isomers, all of which fragment into common ions similar to the example shown Figure 3.17 panel B; where product ions at m/z 163.04, 727.24, and 953.23 are detected as dominant fragments. The latter fragment is formed by neutral loss of dehydrated *p*-coumaric acid (146 Da)

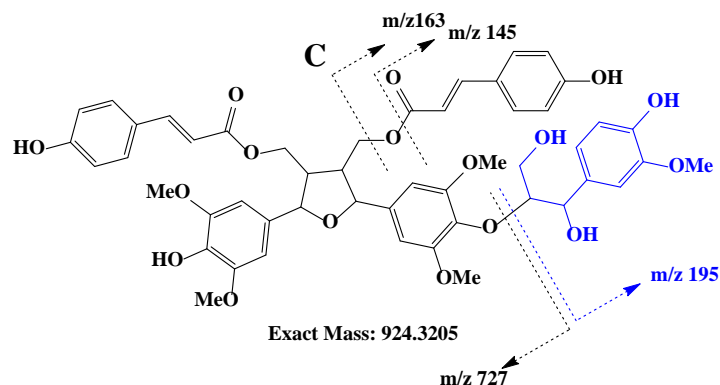
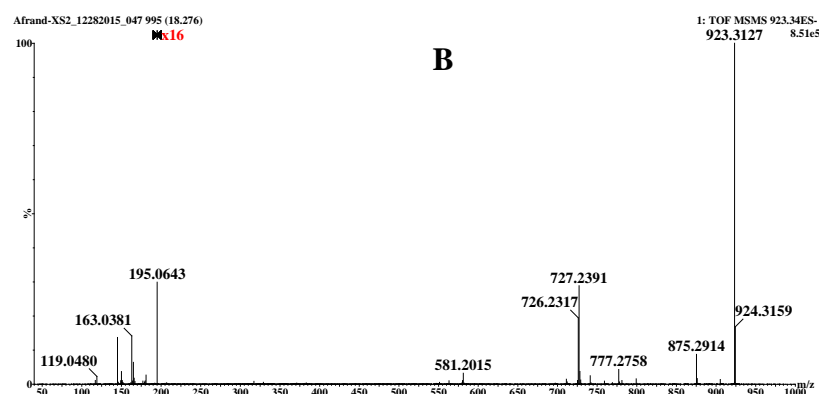
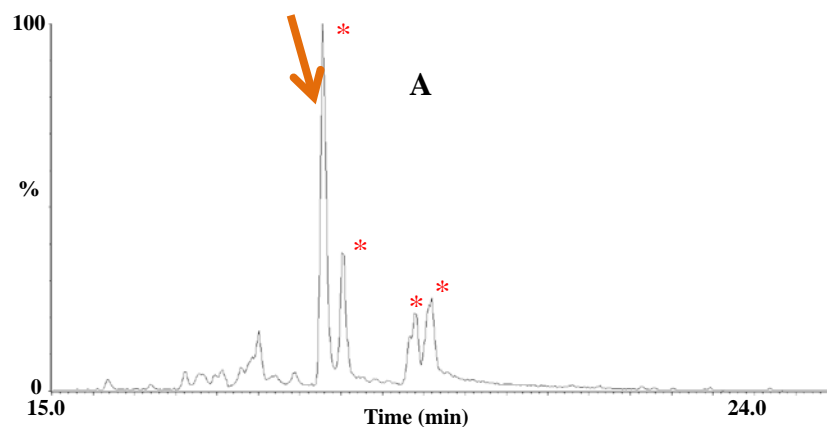


Figure 3.16. Panel A: UHPLC-MS/MS base peak intensity chromatogram for product ion scan of m/z 923.31 for a methanolic extract of corn stover. The isomers resulting in m/z 727.23 product ion are labeled with red * and the other peaks yielded product ions of m/z 741.23 or m/z 743.25.

Panel B: MS/MS product ion spectrum (m/z 923.3) of the most abundant isomer labeled with a green arrow in panel A. Panel C: a proposed structure for the compounds fragmented in panel B: The moiety highlighted in blue represents the G^0 substructure, the only G^0 unit precursor (described in Chapter 1) of this compound.

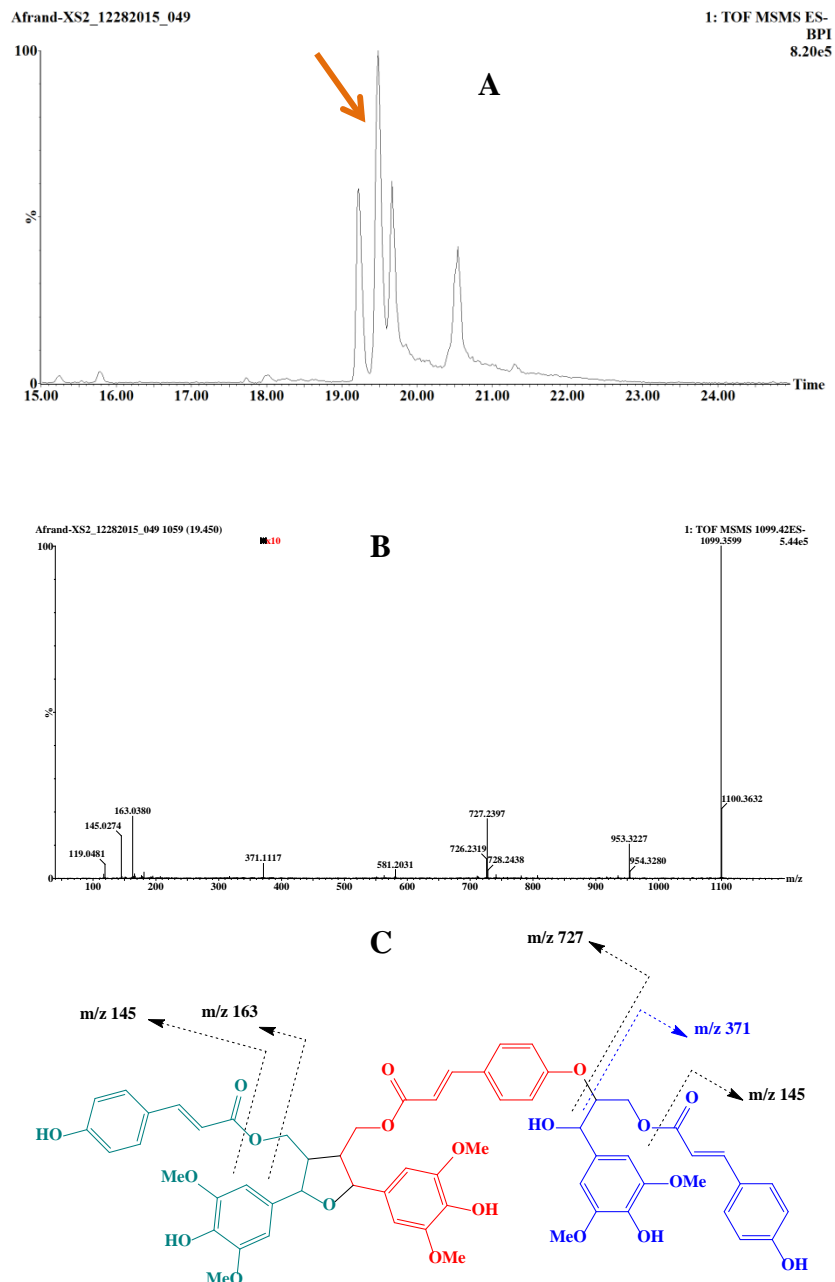


Figure 3.17. (A): UHPLC-MS/MS-base peak intensity (BPI) chromatogram of product ions of m/z 1099.36, measured for a methanolic extract of untreated corn stover in negative-ion mode; (B): MS/MS product ion spectrum of the most abundant peak in A (identified by arrow); (C): proposed annotation of the compound with formula $C_{60}H_{60}O_{20}$ in maize as a trimer of (S-H''); [bis-(S-H'')]- (S-H'')-O. Each color separates one (S-H'') unit.

Although this is consistent with the repeating building block again being the (S-H'') unit, the new linkage formed by the new (S-H'') is unlikely to involve the same THF ring closure

observed for bis-(sinapyl *p*-coumarate). The major evidence for this claim is observation of the key fragment of m/z 371 denoting the (S-H'')-O⁻ (356+16 Da, Figure 3.17 panel B).

This product ion was not detected in MS/MS of bis-(S-H'') where oxidative coupling of two sinapyl groups form the THF ring. This fragment ion lends support to annotation of a trimer, with one monomeric unit cleaved more easily, for example, than the ether bond of common 8–O–4 linkages; e.g. from C8 of sinapyl of the new (S-H'') unit to one of the O-4 location on pre-formed bis-(S-H'') core. It is already known from NMR data of bis-(S-H'') that coupling between the first two (S-H'') is followed by formation of stronger ring of THF which does not fragment to m/z 371 (Figure 3.14). Moreover, upon coupling of the third (S-H'') group, ring closure is not a process available as it was for formation of the (S-H'') dimer. In light of this evidence, it is concluded that the third (S-H'') unit most likely adds as the oxidized (S-H'') (372 Da) to form an 1100 Da molecule with formula C₆₀H₆₀O₂₀. This structure can be represented as [bis-(S-H'')]- (S-H'')-O. A representative reaction of such addition is shown in Figure 3.18.

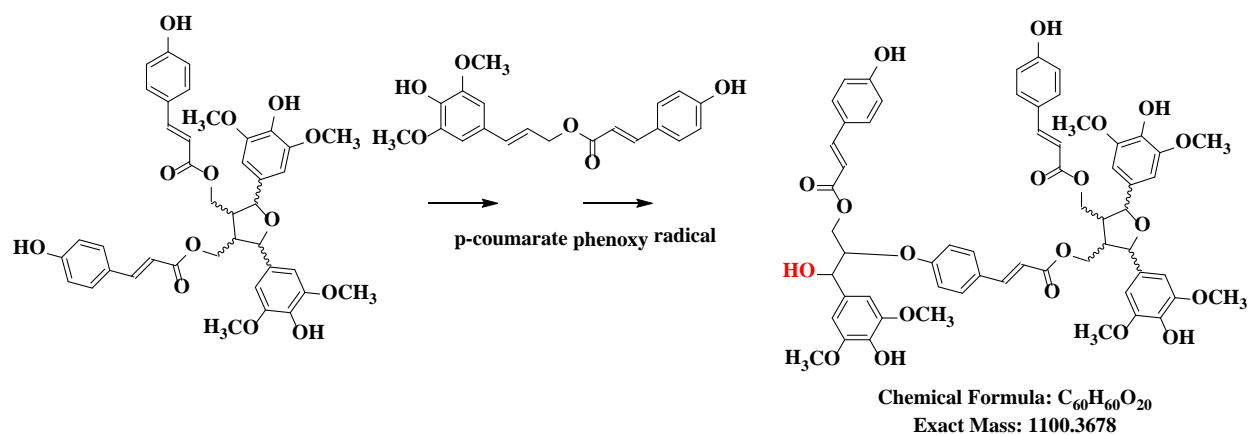


Figure 3.18. One of the possible representative reactions that involves addition of third (S-H'') to bis-(S-H'') core to form [bis-(S-H'')]- (S-H'')-O.

Another larger derivative of bis-(S-H'') that is studied here by using HRMS/MS, ionizes as m/z 1471.48 with molecular formula of C₈₀H₈₀O₂₇, which has 372 Da (C₂₀H₂₀O₇) higher than 1099.36 which suggest addition of two (S-H'')-O units to bis-(S-H'') core. The MS/MS product

ion spectrum of the most abundant isomer shows prominent fragments at m/z 163 (*p*-coumarate), m/z 371 ((S-H'')-O), m/z 727 (bis-(S-H'')), m/z 1099 ([bis-(S-H'')]- (S-H'')-O), those are shown in Figure 3.19. As a result, this compound is annotated as tetramer of sinapyl-*p*-coumarate or [bis-(S-H'')]- (S-H'')-O-(S-H'')-O. The same rationale that was used for MS/MS of m/z 1099 is applicable here too: Easy fragmentation of (S-H'')-O units as well as relatively more stable fragment at m/z 727 denoting THF ring-containing bis-(S-H'') core suggests addition of two (S-H'')-O units most likely via 8-*O*-4 ether linkage. This suggested annotation only represents the simplest structure with least steric effects imposed on the structure and of course assigning such structure needs NMR spectra. Since larger oligomers discussed here are all detected having multiple isomers at different retention times, it is concluded that there are multiple linkage locations and linkages types (e.g. 8-*O*-4, 8-8, 8-5, 5-*O*-4). However using MS/MS sequencing approach for lignin it can be deduced that this compound is made by four (S-H'') units.

Another molecule detected resulting in core unit of bis-(S-H''), m/z 727, in its MS/MS spectrum is the compound detected in corn stover extract at m/z 1295.43. This compound is only 196 larger than [bis-(S-H'')]- (S-H'')-O which also shows in MS/MS spectrum as m/z 1099. This mass increase suggests addition of a G^O unit to [bis-(S-H'')]- (S-H'')-O to form [bis-(S-H'')]- (S-H'')-O- G^O with formula C₇₀H₇₂O₂₄. The MS/MS spectrum for the most abundant isomer of this compound as well as simplest annotation for that based on the observed fragments, are presented in Figure 3.20.

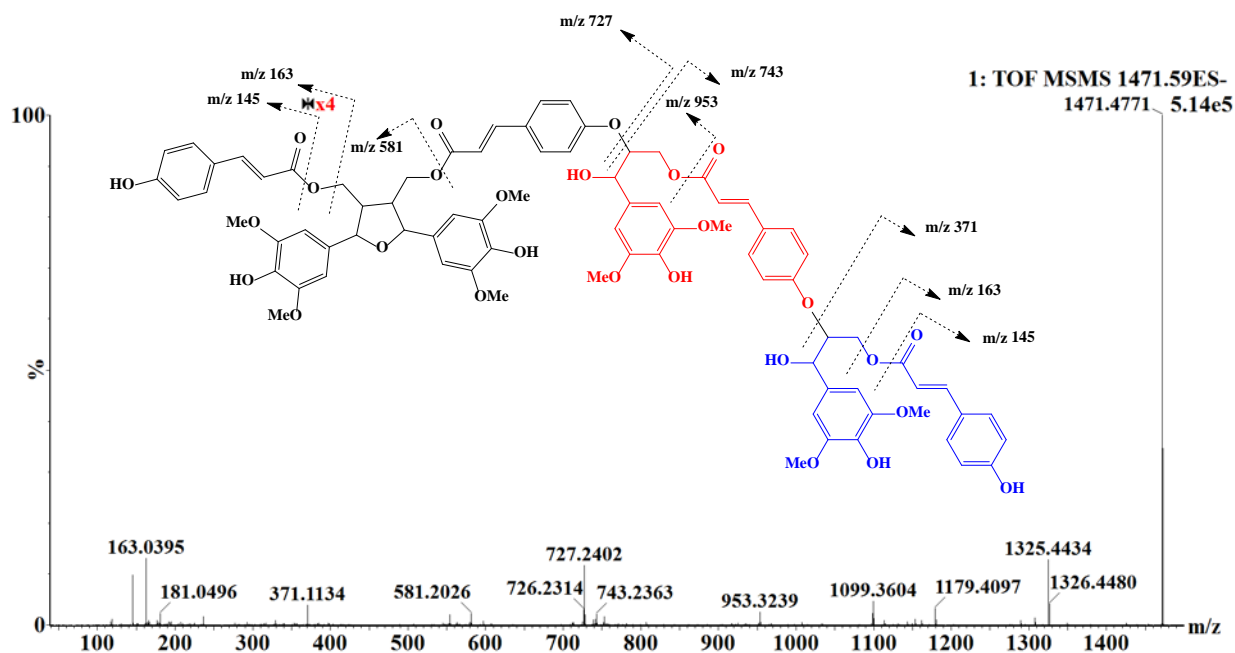


Figure 3.19. MS/MS product ion spectrum of m/z 1471.5 from a methanolic corn stover extract, annotated to be tetramer of (S-H^{''}) which in its simplest form here shown as [bis-(S-H^{''})]-(S-H^{''})-O-(S-H^{''})-O.

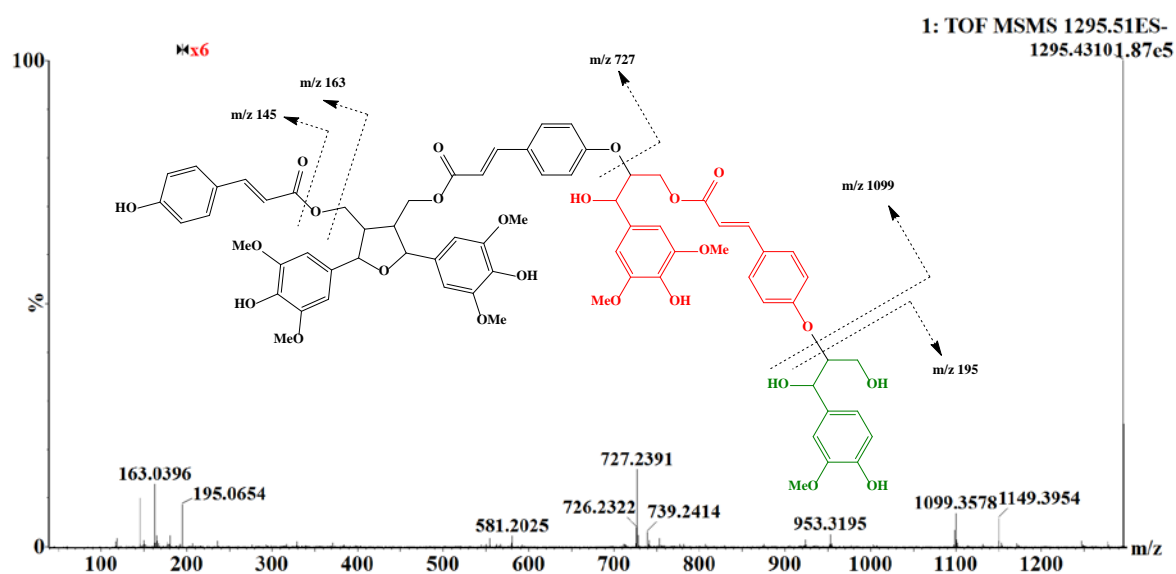


Figure 3.20. MS/MS product ion spectrum of m/z 1295.43 of the most abundant isomer from corn stover annotated as [bis-(S-H^{''})]-(S-H^{''})-O- G^O.

The above observations suggest that bis-(sinapyl *p*-coumarate), bis-(S-H^{''}), is a platform unit of phenolic family capable of reacting with more units of (S-H^{''})-O and/or G^O and make

larger phenolic compounds. Unfortunately with the current technologies, chromatographic separation of the enormous number of large isomeric phenolic molecules is not feasible to enable their purification. Moreover, as oligomer size grows, the number of isomers with similar physical properties and identical molecular masses grows owing to increases in the number of chiral centers. As a result, even LC/MS/MS analyses are likely to present multiple isomeric and isobaric compounds to the mass spectrometer at a single time. However, even in the mass ranges beyond this threshold, ions detected yield exact mass measurements that suggest formation of yet larger oligomers from common precursors. A great example is m/z 1843.61 in corn stover extract that is assigned a formula ($C_{100}H_{100}O_{34}$; theoretical m/z 1843.6023) suggests an oxidized pentamer of (S-H"). However it is noteworthy to point out that this compound and many other compounds discussed in this chapter elute at very close retention times. Figure 3.21 shows a single narrow retention time window where several ions discussed here elute including m/z 1099, m/z 1295, m/z 1461, and m/z 1843.

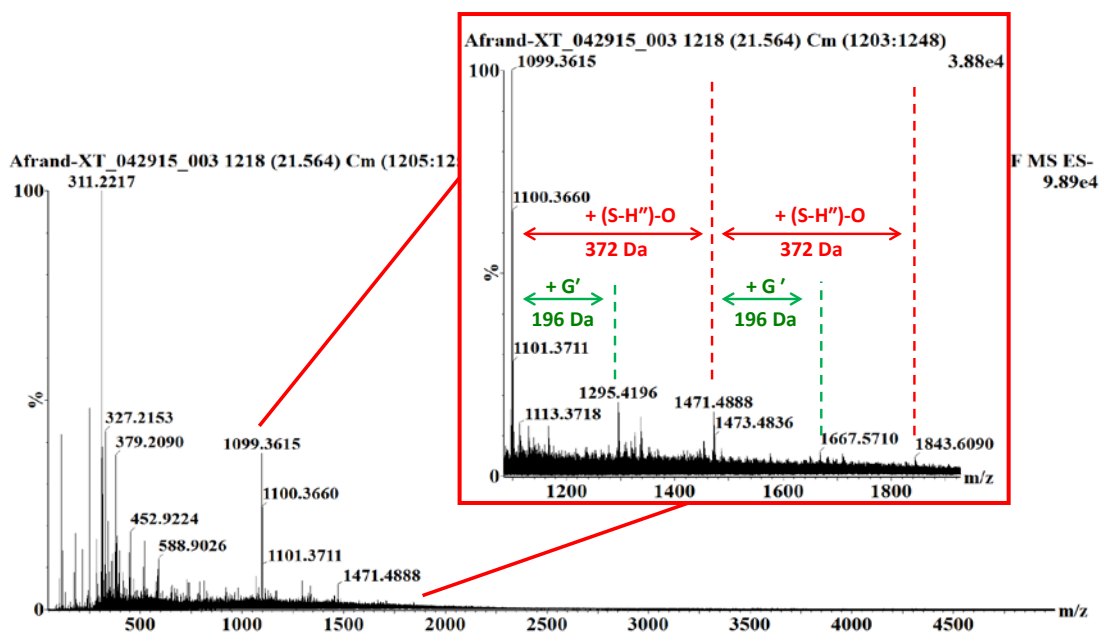


Figure 3.21. Pattern of addition of (S-H'')-O and G^O units in compounds containing bis- (S-H'') core in a single point spectrum of UHPLC-MS of corn stover extract.

As discussed above, these masses suggest the (S-H⁺) core serves as a nucleating point, modified by additions of 196 Da and/or 372 Da , with overlapping chromatographic retention that leads to their appearance in the spectrum on display.

Another important finding is that formation of compounds derived from sinapyl *p*-coumarate units, where one or more (S-H⁺)-O and/or G^O are added to a bis-(S-H⁺) core must undergo coupling involving at least one or perhaps more *p*-coumarate groups. If *p*-coumarate was not involved in coupling, these compounds must use only sinapyl (or guaiacylglyceryl units) to bind to O-4 of the sinapyl groups in the bis-(S-H⁺) core. First, such coupling at the phenolic hydroxyl (O-4) would be discouraged due to steric hindrance by methoxy groups on C3 and C5 of sinapyl units. Secondly, even if they occurred, MS/MS fragmentation of at least one of these molecules discussed above should have resulted in a fragment of three or more sinapyl units where all esterified *p*-coumarates would have been removed following collisional activation; so far no such ions containing three or more sinapyls or sinapyl- G^O without any *p*-coumarate ester have been detected. Lignin-type couplings of phenolic acids has not been reported before to the best of our knowledge. Many reports have demonstrated abundance of *p*-coumarate in grasses, but concluded that these units do not have any role in lignin coupling [28-30 , 37]. However, with example molecules discussed in this part, here we suggest that due to both steric forces on O-methyl of S and G units as well as lack of MS/MS evidence proving otherwise, *p*-coumarates do contribute in lignin coupling chemistry.

The abundance of these compounds serves to indicate that oligomerization reactions, largely heretofore considered to involve the classic lignin monomers, extend to include oxidative couplings of phenolic acid esters that should be labile to hydrolysis (or ammonolysis). We propose these couplings primarily occur through the less hindered *p*-coumaroyl moieties rather

than the more hindered syringyl groups, though the latter cannot be entirely ruled out. Formation of some of derivatives of bis-sinapyl *p*-coumarate is best explained by involvement of *p*-coumarate in radical addition and others can be formed by radical addition to syringyl groups.

3.5 Conclusions

In summary, it was shown that the chemical contributions of hydroxycinnamic acids in grass lignin have remained underappreciated until now. Feruloyl- and *p*-coumaroyl-containing molecules described here that were missed in previous reports and investigations are evidence that suggest a deeper chemistry in lignin of grasses that were not understood before.

The analyses described in this chapter document that 1-*p*-coumaroyl-3-feruloylglycerol was the most abundant compound in methanolic extract of corn stover lignin and is also abundant in other grass lignin extracts (Chapter 2). Dimers of this compound detected at *m/z* 825 can be represented as diferulates esterified to glyceryl-*p*-coumarate. This dimer was detected in multiple isomeric forms. However, whether this dimer is involved in lignin-hemicellulose cross-linking has yet to be determined. One might expect that thermochemical pretreatments may release substantial amounts of glycerol, which has usually not been measured because it does not ionize well using electrospray ionization, and requires derivatization for analysis by GC/MS.

The profiling of extractives in the investigations described in this Chapter provided evidence for addition of tricin to 1-*p*-coumaroyl-3-feruloylglycerol that are consistent with radical coupling reactions. These findings suggest that phenolic acid esters of glycerol have potential to be incorporated into higher molecular mass lignins.

p-Coumarate was shown to be incorporated into higher molecular mass substances via its esterification to sinapyl alcohol, which serves as a reactive moiety that participates in oligomerization. Prior to this report, incorporation of *p*-coumarate esters and sinapyl *p*-

coumarate in lignification has not been emphasized, and reactivity of *p*-coumarate in grass lignin was reported to be irrelevant to development of lignin structures [28 , 30]. Here it was shown that sinapyl *p*-coumarate can dimerize to form a platform for addition of one or more or other lignin units including guaiacylglyceryl (oxidized coniferyl alcohol), oxidized sinapyl *p*-coumarates, or combinations of both. The reactivity of *p*-coumarate groups is considered essential to lignin growth from these lignin nucleation sites, and the observation of higher molecular mass forms (to ~ 2 kDa) in this investigation supports this conclusion.

Deeper investigation of such large and low abundant molecules with multiple isomers, deserves more focus, including improved technologies for characterization, including those that do not require purification of milligram quantities of individual compounds as needed for conventional NMR analysis. Considering presence of multiple isobaric species at a given retention time, achieving such goal will remain a great challenge; and one major question would be: how much information is necessary to obtain useful structural information for the thousands of large constituents of lignin?

REFERENCES

REFERENCES

- (1.) Aliscioni, S.; Bell, H. L.; Besnard, G.; Christin, P. A.; Columbus, J. T.; Duvall, M. R.; Edwards, E. J.; Giussani, L.; Hasenstab-Lehman, K.; Hilu, K. W.; Hodkinson, T. R.; Ingram, A. L.; Kellogg, E. A.; Mashayekhi, S.; Morrone, O.; Osborne, C. P.; Salamin, N.; Schaefer, H.; Spriggs, E.; Smith, S. A.; Zuloaga, F.; Grass Phylogeny Working, G., II, *New Phytologist* **2012**, *193* (2), 304-312.
- (2.) Prasad, V.; Stromberg, C. A. E.; Leache, D.; Samant, B.; Patnaik, R.; Tang, L.; Mohabey, D. M.; Ge, S.; Sahni, A., *Nature Communications* **2011**, *2*.
- (3.) Piperno, D. R.; Sues, H. D., *Science* **2005**, *310* (5751), 1126-1128.
- (4.) Prasad, V.; Stromberg, C. A. E.; Alimohammadian, H.; Sahni, A., *Science* **2005**, *310* (5751), 1177-1180.
- (5.) Chaw, S. M.; Chang, C. C.; Chen, H. L.; Li, W. H., *Journal of Molecular Evolution* **2004**, *58* (4), 424-441.
- (6.) Grabber, J. H.; Ralph, J.; Hatfield, R. D., *Journal of Agricultural and Food Chemistry* **2002**, *50* (21), 6008-6016.
- (7.) Maslen, S. L.; Goubet, F.; Adam, A.; Dupree, P.; Stephens, E., *Carbohydrate Research* **2007**, *342* (5), 724-735.
- (8.) Smith, M. M.; Hartley, R. D., *Carbohydrate Research* **1983**, *118* (JUL), 65-80.
- (9.) Hatfield, R.; Jones, B.; Grabber, J.; Ralph, J., *Plant Physiology* **1995**, *108* (2), 125-125.
- (10.) Appeldoorn, M. M.; Kabel, M. A.; Van Eylen, D.; Gruppen, H.; Schols, H. A., *Journal of Agricultural and Food Chemistry* **2010**, *58* (21), 11294-11301.
- (11.) Bunzel, M.; Allerdings, E.; Ralph, J.; Steinhart, H., *Journal of Cereal Science* **2008**, *47* (1), 29-40.
- (12.) Ishii, T.; Hiroi, T., *Carbohydrate Research* **1990**, *196*, 175-183.
- (13.) Ishii, T.; Hiroi, T., *Carbohydrate Research* **1990**, *206* (2), 297-310.
- (14.) Ishii, T.; Hiroi, T.; Thomas, J. R., *Phytochemistry* **1990**, *29* (6), 1999-2003.
- (15.) Bunzel, M., *Phytochemistry Reviews* **2010**, *9* (1), 47-64.
- (16.) Lattimer, J. M.; Haub, M. D., *Nutrients* **2010**, *2* (12), 1266-1289.

- (17.) Allerdings, E.; Ralph, J.; Schatz, P. F.; Gniechwitz, D.; Steinhart, H.; Bunzel, M., *Phytochemistry* **2005**, *66* (1), 113-124.
- (18.) Vismeh, R. Multifaceted Metabolomics Approaches for Characterization of Lignocellulosic Biomass Degradation Products formed during Ammonia Fiber Expansion Pretreatment. Doctoral Dissertation, Michigan State University, East Lansing, MI, USA, 2012.
- (19.) Hatfield, R. D.; Ralph, J.; Grabber, J. H., *Journal of the Science of Food and Agriculture* **1999**, *79* (3), 403-407.
- (20.) Lam, T. B. T.; Kadoya, K.; Iiyama, K., *Phytochemistry* **2001**, *57* (6), 987-992.
- (21.) Chundawat, S. P. S.; Vismeh, R.; Sharma, L. N.; Humpala, J. F.; Sousa, L. D.; Chambliss, C. K.; Jones, A. D.; Balan, V.; Dale, B. E., *Bioresource Technology* **2010**, *101* (21), 8429-8438.
- (22.) Klinke, H. B.; Ahring, B. K.; Schmidt, A. S.; Thomsen, A. B., *Bioresource Technology* **2002**, *82* (1), 15-26.
- (23.) Vismeh, R.; Lu, F. C.; Chundawat, S. P. S.; Humpala, J. F.; Azarpira, A.; Balan, V.; Dale, B. E.; Ralph, J.; Jones, A. D., *Analyst* **2013**, *138* (21), 6683-6692.
- (24.) Lan, W.; Morreel, K.; Lu, F. C.; Rencoret, J.; del Rio, J. C.; Voorend, W.; Vermerris, W.; Boerjan, W.; Ralph, J., *Plant Physiology* **2016**, *171* (2), 810-820.
- (25.) Bunzel, M.; Ralph, J.; Lu, F.; Hatfield, R. D.; Steinhart, H., *Journal of Agricultural and Food Chemistry* **2004**, *52* (21), 6496-6502.
- (26.) Ishii, T., *Phytochemistry* **1991**, *30* (7), 2317-2320.
- (27.) Ralph, J.; Guillaumie, S.; Grabber, J. H.; Lapierre, C.; Barriere, Y., *Comptes Rendus Biologies* **2004**, *327* (5), 467-479.
- (28.) Hatfield, R.; Ralph, J.; Grabber, J. H., *Planta* **2008**, *228* (6), 919-928.
- (29.) Grabber, J. H.; Quideau, S.; Ralph, J., *Phytochemistry* **1996**, *43* (6), 1189-1194.
- (30.) Lu, F. C.; Ralph, J., *Journal of Agricultural and Food Chemistry* **1999**, *47* (5), 1988-1992.
- (31.) Compton, D. L.; Laszlo, J. A., *Biotechnology Letters* **2009**, *31* (6), 889-896.
- (32.) Shi, H. M.; Yang, H. S.; Zhang, X. W.; Sheng, Y.; Huang, H. Q.; Yu, L. L., *Journal of Agricultural and Food Chemistry* **2012**, *60* (40), 10041-10047.

- (33.) Balmer, D.; de Papajewski, D. V.; Planchamp, C.; Glauser, G.; Mauch-Mani, B., *Plant Journal* **2013**, 74 (2), 213-225.
- (34.) Stagliano, M. C.; DeKeyser, J. G.; Omiecinski, C. J.; Jones, A. D., *Rapid Communications in Mass Spectrometry* **2010**, 24 (24), 3578-3584.
- (35.) Delaporte, R. H.; Guzen, K. P.; Laverde, A.; dos Santos, A. R.; Sarragiotto, M. H., *Biochemical Systematics and Ecology* **2006**, 34 (7), 599-602.
- (36.) del Rio, J. C.; Prinsen, P.; Rencoret, J.; Nieto, L.; Jimenez-Barbero, J.; Ralph, J.; Martinez, A. T.; Gutierrez, A., *Journal of Agricultural and Food Chemistry* **2012**, 60 (14), 3619-3634.
- (37.) Ralph, J.; Hatfield, R. D.; Quideau, S.; Helm, R. F.; Grabber, J. H.; Jung, H. J. G., *Journal of the American Chemical Society* **1994**, 116 (21), 9448-9456.
- (38.) Grabber, J. H.; Schatz, P. F.; Kim, H.; Lu, F. C.; Ralph, J., *Bmc Plant Biology* **2010**, 10.
- (39.) Kang, J. G.; Price, W. E.; Ashton, J.; Tapsell, L. C.; Johnson, S., *Food Chemistry* **2016**, 211, 215-226.
- (40.) Svensson, L.; Sekwati-Monang, B.; Lutz, D. L.; Schieber, A.; Ganzle, M. G., *Journal of Agricultural and Food Chemistry* **2010**, 58 (16), 9214-9220.
- (41.) Ma, C.; Xiao, S. Y.; Li, Z. G.; Wang, W.; Du, L. J., *Journal of Chromatography A* **2007**, 1165 (1-2), 39-44.
- (42.) Xiong, Y.; Deng, K. Z.; Guo, Y. Q.; Gao, W. Y.; Zhang, T. J., *Archives of Pharmacal Research* **2009**, 32 (5), 717-720.
- (43.) Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W., *Plant Physiology* **2010**, 153 (3), 895-905.
- (44.) Morreel, K.; Kim, H.; Lu, F. C.; Dima, O.; Akiyama, T.; Vanholme, R.; Niculaes, C.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., *Analytical Chemistry* **2010**, 82 (19), 8095-8105.
- (45.) Morreel, K.; Dima, O.; Kim, H.; Lu, F. C.; Niculaes, C.; Vanholme, R.; Dauwe, R.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., *Plant Physiology* **2010**, 153 (4), 1464-1478.

**Chapter Four: Mass Spectrometric Analysis of Polymers and Biomass Extractives using
Liquid Chromatography-Time-of-Flight Mass Spectrometry with γ -Valerolactone as a
Renewable Mobile Phase Component**

4.1 Abstract

Analysis of large biopolymers including lignin remains challenging owing to their similar physical properties compared to other large molecules within each class. Though molecular mass and chromatographic fractionation provide powerful tools to discriminate similar substances, selection of solvents appropriate for solubilization and chromatographic separation of large molecules of lignin retains central importance. In this report, the novel solvent system γ -valerolactone (GVL) is evaluated for solubilizing lignin for separation and mass spectrometry. Since recently it was reported that GVL dissolves lignocellulosic biomass almost quantitatively (> 90% dissolved) in the presence of minimal concentrations (10 mM) of sulfuric acid, this investigation evaluated use of the same solvent for UHPLC-MS analysis of solubilized biomass. The rationale behind this investigation is that mysteries regarding biomass constituents insoluble in common HPLC solvents (water, acetonitrile, methanol) might yield to a stronger solvent system for dissolution and chromatographic separation. GVL presents physical properties that make it unusually attractive for this purpose, specifically its high dipole moment and low *cohesive energy density* (Hildebrand parameter). To assess its performance for analysis of high molecular mass substances that are better defined than natural lignins, its performance as mobile phase and for ionization in mass spectrometry was evaluated using a synthetic phenolic polymer, poly(4-vinylphenol) (PVP). Results were compared with those obtained using methanol in place of GVL. Ionization of phenolic polymer in GVL revealed simpler spectra due to minimal number of adduct types in GVL (mainly chloride adduct) in negative ion mode. Chromatography of PVP performed by application organic gradient in water using the same column revealed that some larger polymer molecules that are easily eluted by GVL do not even elute completely off the C18 reversed phase column when methanol is used in mobile phase. After initial studies with PVP,

GVL was used for UHPLC-MS of corn stover extracts using a reversed phase C18 column. , GVL demonstrated altered selectivity and retention of major phenolic components of corn stover extract compared to separations that used methanol in the mobile phase. Also, most constituents of corn stover extract rich in lignin eluted faster using GVL while chromatographic resolution was retained in acceptable level (baseline resolution for major components). Successful application of GVL as mobile phase for UHPLC separation and ionization of large polymer molecules, while offering a different selectivity and shorter retention times compared to methanol mobile phase, provides a new platform for UHPLC-MS analysis of large hydrophobic molecules that exhibit low solubility in common organic solvents employed for separations. Advantages of using GVL in UHPLC and MS should not overshadow its disadvantages. Its use may be largely limited to negative-ion mode because it yields abundant ions from protonation and formation of protonated dimer and oligomers that can suppress analyte ionization, and its low volatility may require elevated ion source temperatures for optimal performance. Despite these drawbacks, use of GVL for UHPLC-MS analyses offers alternative retention selectivity that should prove useful in specific applications.

4.2 Introduction

Characterization of oligomeric constituents of lignocellulosic biomass is an essential step towards utilizing them as sustainable sources of energy and chemical feedstocks. A major challenge arises from the complexity and randomness of these natural polymers, which derives from a diverse suite of monomers and combinatorial oligomerization chemistry. Lignin exhibits remarkable complexity among biopolymers and as was discussed in Chapter 2 and Chapter 3 of this dissertation, multiple factors play roles to construct phenolic oligomers revealed by mass

spectrometry to have components at every nominal molecular mass from several hundred to several thousand Daltons.

While choosing appropriate analytical methodologies and techniques for lignin analysis presents challenges, it is usually necessary to solubilize lignin components prior to analysis. Hence, a few steps should be taken to define and optimize lignin solubility. Dissolution and analysis of lignin have often been accompanied via its deconstruction in various ways including pyrolysis [1] and hydrolysis[2] as the main methods while less frequent biodegradation[3] or electrolysis[4] have also been employed. The most primitive approaches to classify lignin have been based on its reactions and solubilization behavior. Classic examples are Klason Lignin [5-6] and Acid Soluble Lignin (ASL) [7-8]. To obtain Klason lignin, biomass is treated with >70% sulfuric acid below room temperature (10°C) for 16 hr or for shorter time (two periods of 2 hr) at room temperature (20°C), and these conditions hydrolyze glycosidic bonds in polysaccharides. The mixtures are then diluted in water to decrease acid concentration to 4% acid w/w, which leads to precipitation of some phenolic compounds. The insoluble residue is called Klason lignin while the soluble phenolic compounds, usually measured using UV-visible spectrophotometry, are named acid soluble lignin [6]. A related procedure known as the Kraft process uses harsh alkaline conditions to remove lignin from cellulose, and is one the most historic and widely used methods in the paper pulping industry [9]. Here, aqueous solution of sodium hydroxide and sodium sulfide (together called white liquor) reacts with wood chips for two hours under pressure at 170°C. Hydroxide and sulfide anions react with lignin units causing hydrolysis and displacement of ester bonds, breaking down lignin into smaller units. Use of the reactive nucleophile sulfide to solubilize fragments here resembles the use of sulfur in thioacidolysis [10], aiding displacement of labile bonds and forming smaller substances with increased solubility.

Both methods of acidic and alkaline solubilization of lignin rely on chemical degradation of this polymer. This highlights the major paradox in investigation of lignin chemical structure: intact lignin structures are not easily isolated and solubilized to be studied individually. On the other hand, effective methods of analysis have presumed it necessary to change the original identity of these molecules through chemical transformation prior to most analyses.

Another way of lignin categorization is based on the fractions soluble in common organic solvents including ethanol, methanol, acetonitrile, acetone, and dioxane, and fractions derived this way have been called organosolv lignin [11-12]. Organic-soluble lignin, can account for as low as 0.5% to as much as 10% of biomass (our data shown in Chapters 2 and 3). Often minor amounts of sulfuric acid (1-10 mM) have been used along with organic solvent to catalyze hydrolysis of ester or glycosidic cross-links of lignin to cellulose, hemicellulose, or other lignin strands. Although this modification can increase recovery yield of lignin to 50% [12] again this rather small modification of solvent is anticipated to modify some aspects of lignin structure.

A new solvent system used for solubilization of lignin which itself is a renewable product of sulfuric acid-catalyzed polysaccharide degradation is γ -valerolactone (GVL) [13]. Use of this solvent along with 5-10 mM sulfuric acid dissolved >90% of total biomass in corn stover [13]. The key factor about GVL that has made it an interesting solvent for biomass is that sugars in lignocellulosic biomass can be converted to GVL during the process, making this low-boiling-point solvent a sustainable system for extraction of lignin from biomass [13].

Quantitative dissolution of almost all of corn stover biomass in acidified GVL provides a great probe for analysis of almost all phenolic and non-phenolic compounds present in corn stover. Back to the problem mentioned earlier, GVL extracts almost all biomass, and then this solution can be subjected to a variety of analysis methods. As was shown in Chapters 2 and 3,

UHPLC-MS and MS/MS provide powerful, and perhaps unique, approaches that reveal chemical information about molecular masses and functionality of individual large molecules.

The unusual success of GVL in solubilizing almost all acid-treated biomass also suggests that conventional solvent systems (water, acetonitrile, methanol, isopropanol, ethanol, dichloromethane, hexane, ethyl acetate, THF, etc.) used in HPLC gradients for separation prior to MS may not be effective for eluting some components.

In order to resolve and analyze large molecules of biomass that are solubilized in GVL, a suitable HPLC solvent system is required that can elute dissolved constituents from the column. We hypothesize that GVL itself should have utility in HPLC gradients for separation and analysis of large molecules that are not easily solubilized in conventional solvents.

In order to rationalize the use of GVL as mobile phase of HPLC, physicochemical properties of this solvent should be considered and compared with those of common HPLC solvents. Acetonitrile and methanol serve as common reversed phase liquid chromatography (RPLC) solvents that are used in gradients against water. Throughout Chapters 2 and 3 of this dissertation, methanol was the primary organic component because it was used for extraction of phenolic material from biomass. However, acetonitrile was also used in a limited number of experiments, and it was observed that retention times of all compounds were shorter than when methanol substituted. However, the order of elution of all major compounds remained the same. So, only a net shift of retention time to earlier times is seen when acetonitrile replaces methanol in water-organic gradients for RPLC. Acetonitrile is used as organic mobile phase in a majority of proteomics, lipidomics, and metabolomics HPLC and HPLC-MS analyses, and methanol is not used nearly as much.

The reason for using an organic solvent against water in reversed phase chromatography is to decrease partitioning of analyte components on/in lipophilic stationary phases such as C18. The conventional wisdom has been that RPLC gradients progress from a polar phase (water) to nonpolar phase (organic) to successively elute nonpolar substances. However, this is an oversimplification, and it is worth noting that dipole moments reported for both methanol (2.87 D) and acetonitrile (3.92 D) are significantly higher than the dipole moment of water (1.85 D) [14], so such gradients progress to solvents with higher dipole moments to elute less-polar compounds. This would appear to contradict the classic principal of “like-dissolves-like”. As a result, the choice of organic solvent for RPLC is often attributed to refractive index, which reflects the dielectric constant of the solvent and dipole-induced dipole intermolecular forces. An examination of solvent properties reported in Table 4.1, it is evident that neither refractive index nor the dielectric constants of acetonitrile and methanol differ largely enough to explain and rationalize why acetonitrile results in faster elution of compounds in RPLC.

In fact, an important reason why acetonitrile elutes many compounds faster than methanol on C18 lies in the lesser energy needed to separate acetonitrile molecules from each other relative to water or methanol; both of which have strong intermolecular hydrogen bonds.

In order to explain this contrast, another physicochemical parameter, known as the Hildebrand parameter [15], δ , is used to explain an important solvent property. The Hildebrand parameter is defined as square root of cohesive energy density [16]:

$$\delta = \sqrt{\frac{\Delta H_v - RT}{V_m}}$$

where ΔH_v is heat of vaporization, R is the universal gas constant, T is temperature and V_m is molar volume. The Hildebrand parameter reflects the energy needed to create a cavity in the solvent to accommodate a solute, and this drives much of chromatographic retention, particularly

in reversed phase separations. Table 4.1 and Figure 4.1 compare and contrast solvent properties of GVL with three other conventional HPLC solvents: water, methanol, and acetonitrile.

	Water	Methanol	Acetonitrile	GVL
Dipole moment (Debye)	1.85[14]	1.70[14]	3.93[14]	4.22
<i>N</i> (refractive index)	1.33[14]	1.33[14]	1.34[14]	1.43
δ , Hildebrand parameter ($\text{cal}^{1/2} \text{cm}^{-3/2}$)	23.5	14.2	11.9	9.6
ϵ (dielectric constant)	80.4[14]	33.6[14]	36.6[14]	36.9[14]

Table 4.1. Physicochemical properties of three common RPLC solvents [14] and γ -valerolactone (GVL).

In terms of solvent properties, GVL exhibits dielectric constant similar to acetonitrile (Figure 4.1), slightly higher dipole moment, and lower cohesive energy density. GVL contains one chiral carbon, but is presumed to exist as a mixture of enantiomers in common use. This chiral molecule which has molecular mass of 100 Da has dipole moment much greater than water while its Hildebrand parameter is smaller than water, acetonitrile, and methanol. Taken together, these values suggest GVL has great capacity to solubilize organic materials. As a matter of fact, close relatives of GVL such as γ -butyrolactone (simply known as butyrolactone) have long been used for paint stripping, which involves solubilization of paint polymers.

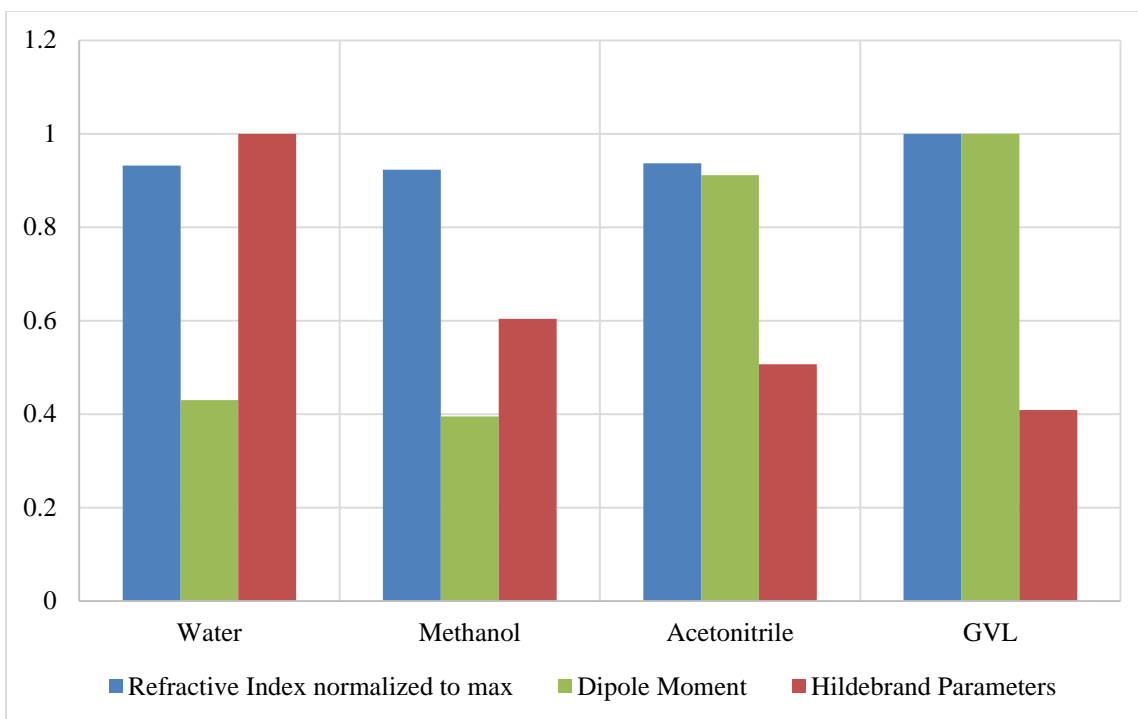


Figure 4.1. Relative Solvent Parameters (normalized to highest value in each category) of water, methanol, acetonitrile, and GVL.

After it was reported that GVL is a promising candidate solvent derived from sustainable and renewable fuel production[17], research has focused on conversion of lignocellulosic polysaccharides into this material. Recently, Alonso et al. [13 , 18] and Luterbacher et al. [19] demonstrated conversion of cellulose and hemicellulose into furfurals and levulinic acid (LA) and further conversion of LA into GVL. The process starts with initial presence of GVL and catalytic amounts of acid (5-10 mM sulfuric acid) as solvent and ends by formation of even more GVL, which in the crude mixture contains dissolved lignin and degradation products including furfural. In the reports mentioned above [13] quantitative dissolution of biomass in the GVL was reported.

Total dissolution of lignin in GVL without use of extensive chemical reagents provides an excellent opportunity to analyze intact or nearly intact molecules of lignin that have undergone minimal modification. Although it is expected that high pressure and temperature in

presence of dilute acid drive hydrolysis of many linkages including ester linkages of *p*-coumarate and ferulate units (discussed in Chapter 3), extensive deconstruction of C-C or C-O-C linkages that are formed by radical coupling of monolignols is not expected.

In this chapter, use of GVL as a UHPLC mobile phase and solvent for electrospray ionization (ESI)-MS will be evaluated.

In order to evaluate how phenolic compounds might be separated and ionized in a GVL solvent system, the synthetic polymer poly(4-vinylphenol) (or PVP), a compound with phenolic units at every vinyl monomer, was used to establish a model study set for lignin-like compounds. Different molecules of this polymer differ in multiplies of 120 Da which is the molecular mass of its monomer, 4-vinylphenol. Figure 4.2 shows structure of the PVP repeating unit. Solid PVP samples of two different molecular weight ranges were studied: one with ~4250 Da (distribution of 1500 - 7000 Da), called small-PVP, and the other one of ~25 kDa, called large-PVP.

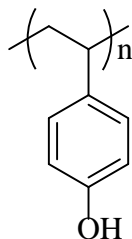


Figure 4.2. Repeating unit of poly(4-vinylphenol)

In this chapter, the performance and applicability of GVL as a solvent for reversed phase UHPLC and ESI-MS will be evaluated.

4.3 Experimental

γ -Valerolactone (GVL), Reagent Plus®, 99% was purchased from Sigma-Aldrich (SKU:V403). Methanol HPLC grade was purchased from (VWR Scientific). Water was used as purified to make resistance of 18.2 M Ω by MilliQ system (Millipore). Sulfuric acid and formic

acid were purchased from VWR Scientific. Poly(4-vinylphenol) (PVP) with 1500-7000 Da molecular mass was purchased from Polysciences, Inc. and PVP with ~25 kDa molecular mass was purchased from Sigma-Aldrich (SKU: 436224). 2,5-Dihydroxybenzoic acid (DHB) was purchased from Sigma-Aldrich. Corn stover biomass was provided by the laboratory of Professor Bruce Dale at Michigan State University

Flow injection mass spectrometry was performed using electrospray ionization (ESI) in negative-ion mode using a Waters Xevo-G2S quadrupole time-of-flight (Q-TOF) mass spectrometer while solutions were introduced at 5 μ L/min into the ion source.

UHPLC-MS analyses were performed using Waters Acquity UPLC system and the same QToF mass spectrometer mentioned above. Ascentis Express C18 reversed phase columns were purchased from Supelco with 2.7 μ m particle size 100Å pore size packing and 100 mm x 2.1 mm dimensions. A gradient of solvent B (methanol or GVL) from 1% (v/v) B in 0.15% aqueous formic acid (solvent A) was ramped from minute 1 to 99% (v/v) B over 30 minutes followed by a 4-minute hold at 99% (v/v) B, followed by re-equilibration to initial conditions for 2 minutes.

Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) analyses were performed in negative ion mode using a Shimadzu Axima cfr-plus mass spectrometer using DHB matrix at 1000:1 matrix: sample ratio and a cocktail of peptides and proteins was used as MS calibrant.

4.4 Results and Discussion

Solutions (5 mg/mL) of small and large PVP in methanol and in GVL were initially analyzed using flow-injection analysis mode (FIA) and electrospray ionization. Figure 4.3 compares and contrasts the ESI-MS and MALDI-MS of the same solution of small PVP in GVL.

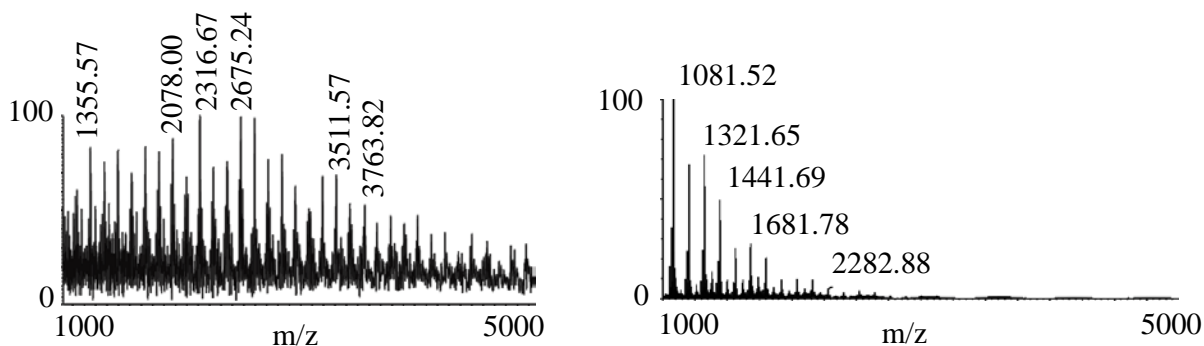


Figure 4.3. (Left panel) Negative-ion MALDI mass spectrum of small PVP solution in GVL and (Right panel) negative-ion electrospray ionization mass spectrum of the same solution.

Both ionization methods resulted in a series of peaks in their mass spectra separated by 120 Da, the mass of the monomeric unit. As it is seen in Figure 4.3, the ESI mass spectrum yielded a different mass distribution of PVP polymer ions, showing less signal at higher molecular mass. This disagreement is attributed to a decrease in ESI ionization efficiency as molecules become heavier. However, MALDI-MS displays evidence of a distribution extended to higher molecular masses than observed using ESI for the same sample. MALDI-MS clearly show the m/z values closer to the reported average of this sample (~4000 Da). Alternatively, the reported molecular masses, which were probably determined by size exclusion chromatography, may overestimate the true mass distribution.

The ESI mass spectrum for the large PVP is shown in Figure 4.4 where the ions from the ~25 kDa polymer are not resolved clearly and individually. Instead an almost continuous of signal that is elevated from baseline is detected. In some lower m/z regions some ions are resolved as isotopic peaks, with a triply charged ion at m/z 1349.66 corresponds to a molecule of ~4000 Da. The higher m/z regions of the spectrum are a continuous line with most of the signal

not resolved as specific isotopic contributions. As a result of these findings, the small PVP was chosen for further FIA-MS and UHPLC-MS studies.

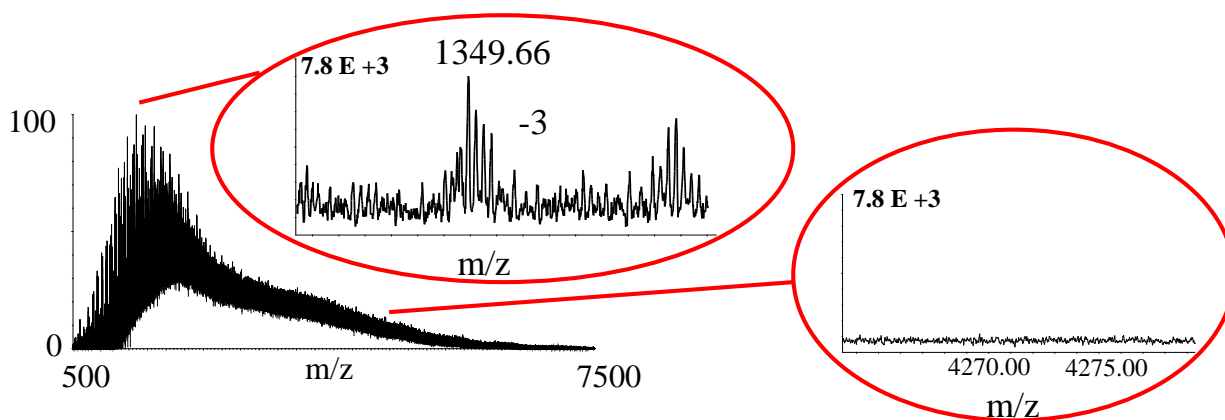


Figure 4.4. Flow-injection analysis (FIA)-MS of large PVP polymer obtained using electrospray ionization in negative-ion mode. Red circles show the zoomed-in view of the corresponding regions in the spectrum.

Negative-ion mode ESI ionization of the small PVP in GVL solution yielded a simpler spectrum than the spectrum observed using methanolic solutions of the same material (Figure 4.5). Despite the weak acidity of phenolic groups, the spectra show ions annotated as more abundant adducts (e.g. $[M+Cl]^-$ and $[M+formate]^-$ than $[M-H]^-$). This relative simplicity of the spectrum using GVL is attributed to its high boiling point (207°C). According to the charged residue model [20] for electrospray ionization, as the solvent evaporates from charged droplets and undergoes sequential Coulombic explosions, charged molecules are left behind in the unevaporated residue. When methanol is the solvent, its evaporation is expected to be faster than other droplet components, including residual formic acid and salts from the solvent delivery system. These have lower vapor pressures (higher boiling points) than methanol, and are therefore expected to be enriched in the droplet residue. However, GVL has a higher boiling point (and lower vapor pressure) than some of these constituents such as formic acid. For example, water and formic acid are expected to evaporate before GVL. As a result the only

adduct-forming component left in the droplet as GVL finally evaporates is chloride ion, which is slow to ionize as long as it exists in ionic form in the droplet (and not as HCl). For example, as it is shown in Figure 4.5, electrospray of methanol solution results in a mass spectrum displaying deprotonated PVP tetramer at m/z 481.23 and formate adduct of the same ion $[M+HCOO]^-$ at m/z 527.23 while the same molecule forms only $[M+Cl]^-$ adduct as electrosprayed in GVL solution (upper panel). The same pattern of differences is observed for the PVP pentamer and hexamer (m/z 637.28 and m/z 757.33 respectively). Singularity of ion type for individual species may provide a beneficial increase in sensitivity while making mass spectra simpler to interpret. The latter is especially important for analysis of complex polymers including lignin.

It is noteworthy that since GVL has a higher boiling point (207°C) than water, methanol, and other common solvents, the desolvation gas temperature, which is normally set to about 200°C was elevated to 600°C. After establishing that GVL supports electrospray ionization of phenolic compounds and polymers and showing that it results in simpler spectra, it was critical to assess whether or not GVL is suitable as a mobile phase for UHPLC separations.

Solutions of small PVP (50 mg/mL) were prepared in both methanol and GVL, and a gradient of each solvent with water was used to analyze the solution of PVP with UHPLC-MS in negative ion mode. Figure 4.6 shows the base-peak ion (BPI) chromatogram obtained as well as the extracted ion chromatogram (XIC) for two different oligomers of PVP (m/z 573.26 and m/z 2403.37) in methanol and GVL. Dashed lines on the chromatograms depict time-dependence of solvent B (methanol or GVL) in water as % v/v. In order to compare performance of the two solvents in separation and elution of compounds present in the polymer mixture it was useful to compare retention and peak shape of a given compound in two different solvents. A pair of different PVP oligomers possessing different degrees of polymerization (DP) of 4 (capped with

phenyl group) and 19 (capped with benzoate) was chosen for generation of extracted ion chromatograms (XICs) which are also shown in Figure 4.6. These two oligomers are labeled with asterisk and double-asterisk respectively in chromatograms obtained using gradient elution with GVL in the mobile phase. A comparison of BPI chromatograms of PVP polymer in methanol and GVL gradients (panel A and B of Figure 4.6 respectively) suggests that GVL elutes the polymer molecules with shorter retention times than elution using methanol in the same proportions. XICs of DP=4 oligomer (4 monomers capped with phenyl group, panel C and D) and DP=19 (19 monomers capped with benzoate, Panel E and F) of PVP suggest that usage of methanol as UHPLC solvent gradient did not elute larger polymer molecules from C18 column using this gradient, though it cannot be excluded that these might elute if the solvent composition were held at high organic for a longer time. This is evident in case of DP=19 polymer which had not finished eluting from the C18 column when solvent program reached the end of the highest methanol content. In contrast, all isomers of DP=19 are completely eluted when GVL was used in the mobile phase. This example shows that conventional solvents might be unable to resolve or even elute higher mass (or lower solubility) polymers from RPLC columns.

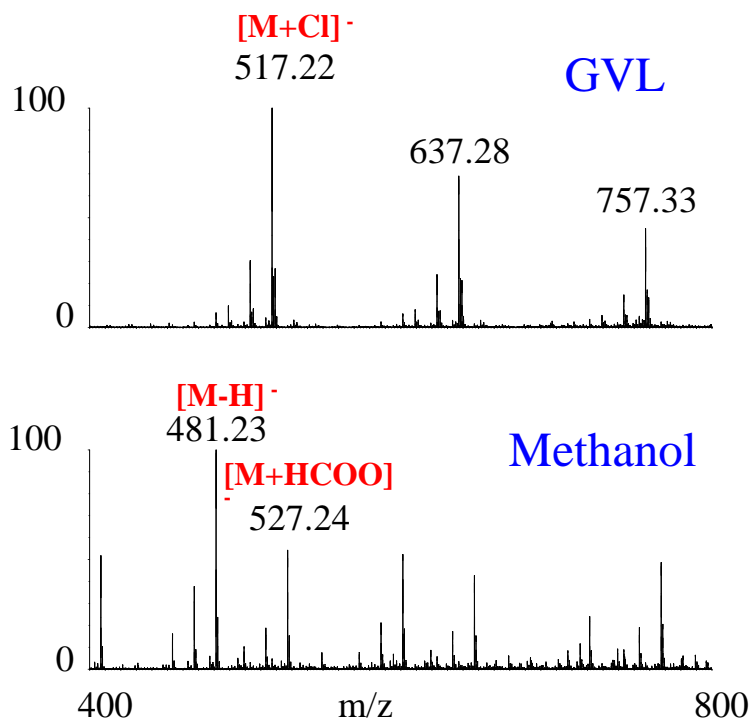


Figure 4.5. Comparison of negative ESI mass spectra of small PVP at 5 mg/mL in GVL (top) and in methanol (bottom).

The UHPLC-MS utilizing GVL was then applied in the profiling of untreated corn stover (UTCS) extracts. Methanol extracts of UTCS (roughly 10 mg/mL) were analyzed by UHPLC-MS using two different gradients that were used with PVP solution before; methanol in water and GVL in water. Results are shown in Figure 4.7 where some of the compounds studied in Chapters 2 and 3 of this report are labeled with letters A to K in red color within each BPI chromatogram obtained from methanol and GVL gradient. The asterisk-containing labels denote derivatives of the flavonoid triclin. Interestingly GVL not only elutes compounds faster, it also improves the chromatographic resolution yielding narrower peaks and taller peaks leading to improved sensitivity. An interesting feature displayed by GVL in comparison with methanol is the difference in retention selectivity towards the phenolic compounds.

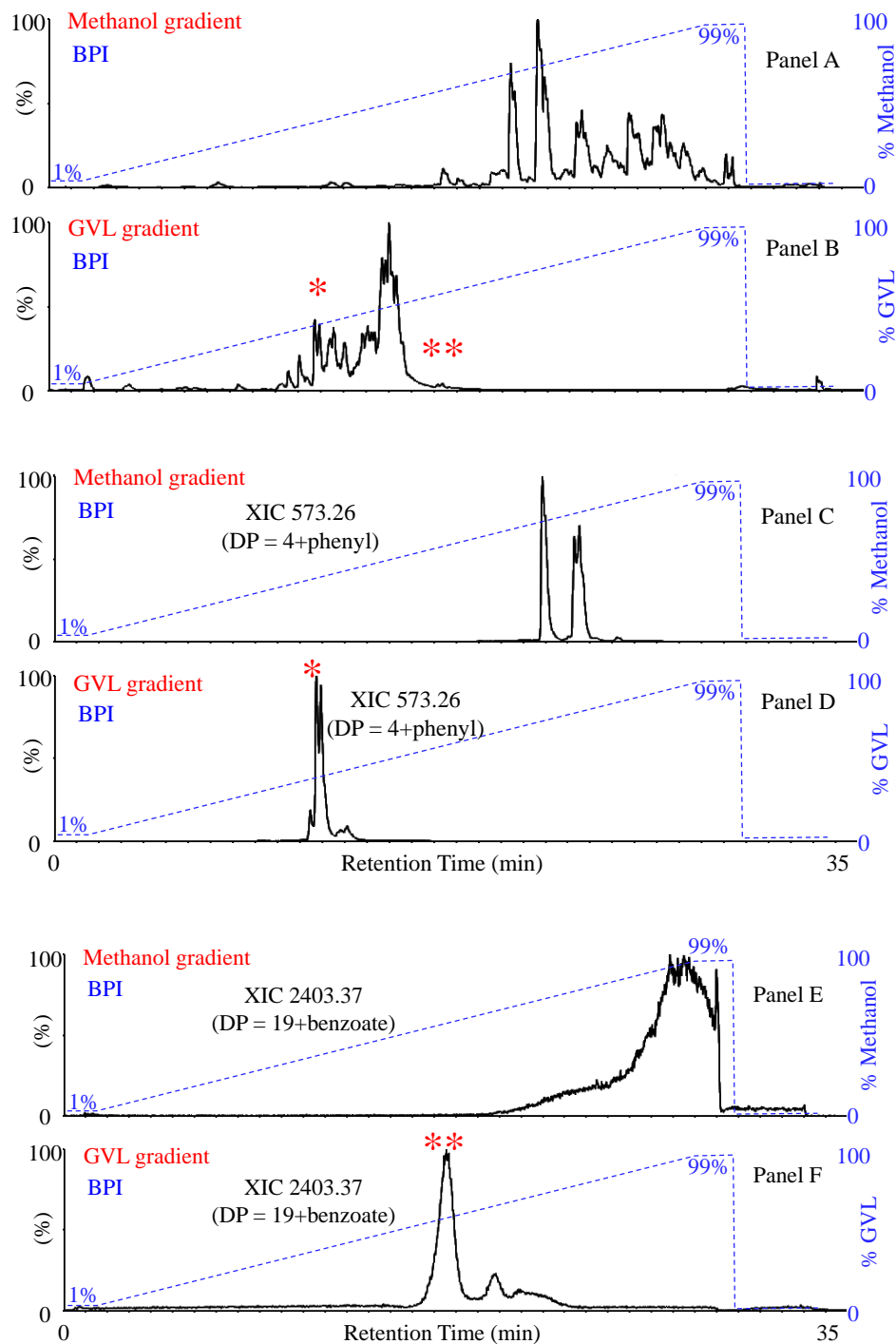


Figure 4.6. UHPLC-MS of small PVP polymer in methanol and GVL. BPI chromatograms (top two panels). Asterisk and double-asterisk signs show the time points chosen to show extracted ion chromatograms (XICs). Two middle panels demonstrate XIC of DP=4+phenyl PVP as eluted by gradient of methanol and GVL in water, and the bottom two panels show XIC of DP=19+benzoate PVP as eluted by methanol and GVL.

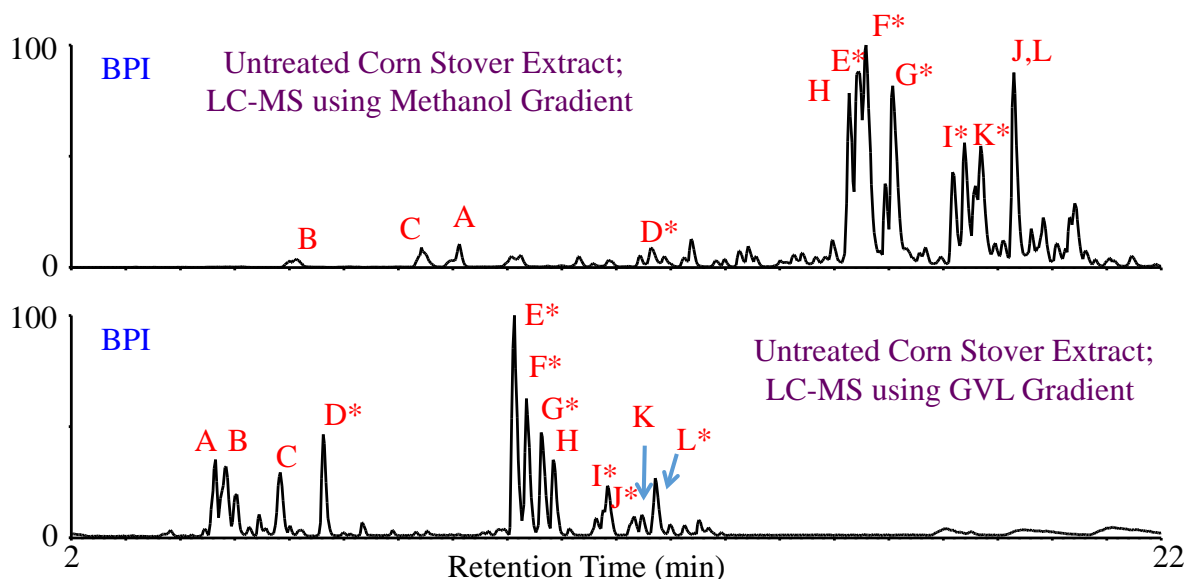


Figure 4.7. BPI chromatograms of methanol extract of UTCS analyzed by UHPLC-MS using methanol (top panel) and GVL (bottom panel) in solvent gradient. Similar letters at different chromatograms denote identical compounds within each chromatogram. Peak labeled A: m/z 539.152, B: m/z 121.033 (benzoic acid), C: m/z 163.044 (*p*-coumaric acid), D: m/z 637.146, E: m/z 329.069 (tricin), E and F: m/z 525.137 (GGT), H: m/z 413.128 (1-*p*-coumaroyl-3-feruloylglycerol), I: m/z 567.154 (acetyl GGT), J: m/z 701.192 (O-9-(*p*-coumaroyl)syringyl glyceryl tricin), K: m/z 727.238 (bis-(sinapyl *p*-coumarate) L: m/z 671.180 (O-9-(*p*-coumaroyl)guaiacylglyceryl tricin).

UHPLC separations using GVL in mobile phase retain good chromatographic resolution for compounds that are minimally resolved when methanol is used in UHPLC solvent gradients. Examples in Figure 4.7 include peaks labeled E, F, G, and H (tricin, two GGT isomers (discussed in Chapter 2, and 1-*p*-coumaroyl-3-feruloylglycerol discussed in Chapter 3) that are baseline-resolved using GVL while in methanol they overlap. Another great example appears in peaks labeled J and L, which differ by only one -O-CH₃ group and co-elute when methanol was used in gradient, however GVL clearly separates them with baseline level resolution.

The combination of all advantages of using GVL in HPLC mobile phase, including faster elution, sharper peaks, alternative selectivity, and great chromatographic resolution, suggests that this solvent can be used as an alternative for chromatographic separation of polymers, especially

phenolic polymers. Theoretically GVL should be applicable to HPLC separation modules hyphenated to spectroscopic detectors too, like UV and visible spectroscopy, providing that solvent absorbance does not interfere.

4.5 Conclusions

Use of GVL in LC mobile phases enables separation of a wider range of compounds than can be achieved with methanol owing to its solvation of semi-polar compounds with molecular masses beyond those normally separated using conventional solvents.

GVL also can be used as a solvent for LC/MS. Despite being an aprotic solvent, GVL supported ionization of phenolic compounds including polymers, but it is unclear whether trace amounts of water or dissolved salts are necessary for ionization. The high dipole moment of GVL leads to expectations that it will allow ionization of polar compounds other than phenolic substances, and has potential to serve as a universal ESI solvent. GVL resulted in simpler negative ion spectra of poly(4-vinylphenol) (PVP) than methanol.

It was also demonstrated here that GVL can be used with water in binary LC gradients to separate large polymer molecules of with C18 stationary phase. Example of DP=19 PVP demonstrated that how conventional solvents may only partially elute compounds from C18 stationary phases, yielding incomplete chromatographic separations.

A more interesting behavior of GVL distinct from methanol became explicit when extract of untreated corn stover were analyzed by comparing UHPLC separations using GVL and methanol (both as solvent B) in the mobile phase versus water (as solvent A) with otherwise identical solvent programming. As expected, GVL resulted in faster elution of compounds without sacrificing chromatographic resolution. Moreover, it was demonstrated that elution order of the components of corn stover extract exhibited marked differences in GVL compared to

methanol. Faster elution along with preservation of chromatographic resolution and altered retention selectivity make up a set of excellent options for a solvent needed for separation of large and small molecules followed by MS detection.

The use of GVL solvent system provides opportunities for separation and analysis of mixtures of large molecules that were not well-separated before. It is envisioned that GVL may find applications in a variety of research and industrial fields including, but not limited to, synthetic polymers industries, petrochemicals, and natural products.

As for limitations, it should be considered that GVL forms abundant $[M+H]^+$ ions in positive-ion mode electrospray ionization, and may suppress ionization of other compounds. Also due to its high boiling point, GVL may condense in ion sources, and may require elevated source temperatures or periodic source cleaning. These may keep applications of this solvent focused on samples that are not otherwise separated using gradients based on methanol or acetonitrile.

REFERENCES

REFERENCES

- (1.) Yang, H. P.; Yan, R.; Chen, H. P.; Lee, D. H.; Zheng, C. G., *Fuel* **2007**, 86 (12-13), 1781-1788.
- (2.) Marques, A. V.; Pereira, H.; Meier, D.; Faix, O., *Holzforschung* **1994**, 48, 43-50.
- (3.) Ahmad, M.; Taylor, C. R.; Pink, D.; Burton, K.; Eastwood, D.; Bending, G. D.; Bugg, T. D. H., *Molecular Biosystems* **2010**, 6 (5), 815-821.
- (4.) Ghatak, H. R., *Industrial Crops and Products* **2013**, 43, 738-744.
- (5.) Hatfield, R. D.; Jung, H. J. G.; Ralph, J.; Buxton, D. R.; Weimer, P. J., *Journal of the Science of Food and Agriculture* **1994**, 65 (1), 51-58.
- (6.) Wald, W. J.; Ritchie, P. F.; Purves, C. B., *Journal of the American Chemical Society* **1947**, 69 (6), 1371-1377.
- (7.) Yasuda, S.; Fukushima, K.; Kakehi, A., *Journal of Wood Science* **2001**, 47 (1), 69-72.
- (8.) Kaar, W. E.; Brink, D. L., *Journal of Wood Chemistry and Technology* **1991**, 11 (4), 465-477.
- (9.) Chakar, F. S.; Ragauskas, A. J., *Industrial Crops and Products* **2004**, 20 (2), 131-141.
- (10.) Rolando, C.; Monties, B.; Lapierre, C., Thioacidolysis. In *Methods in Lignin Chemistry*, Lin, S. Y.; Dence, C. W., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1992; pp 334-349.
- (11.) Lange, W.; Schweers, W., *Wood Science and Technology* **1980**, 14 (1), 1-7.
- (12.) El Hage, R.; Brosse, N.; Sannigrahi, P.; Ragauskas, A., *Polymer Degradation and Stability* **2010**, 95 (6), 997-1003.
- (13.) Luterbacher, J. S.; Rand, J. M.; Alonso, D. M.; Han, J.; Youngquist, J. T.; Maravelias, C. T.; Pfleger, B. F.; Dumesic, J. A., *Science* **2014**, 343 (6168), 277-280.
- (14.) CRC handbook of chemistry and physics. Chapman and Hall/CRCnetBASE: Boca Raton, FL, 1999; pp CD-ROMs.
- (15.) Hildebrand, J. H., *Chemical Reviews* **1949**, 44 (1), 37-45.
- (16.) Lee, S. H.; Lee, S. B., *Chemical Communications* **2005**, (27), 3469-3471.

- (17.) Horvath, I. T.; Mehdi, H.; Fabos, V.; Boda, L.; Mika, L. T., *Green Chemistry* **2008**, *10* (2), 238-242.
- (18.) Alonso, D. M.; Wettstein, S. G.; Mellmer, M. A.; Gurbuz, E. I.; Dumesic, J. A., *Energy & Environmental Science* **2013**, *6* (1), 76-80.
- (19.) Alonso, D. M.; Wettstein, S. G.; Dumesic, J. A., *Green Chemistry* **2013**, *15* (3), 584-595.
- (20.) Kebarle, P., *Journal of Mass Spectrometry* **2000**, *35* (7), 804-817.

Chapter Five: Concluding Remarks

The focus of this study included obtaining deeper molecular understanding of grass lignins. This need was driven by two facts. First, the majority of knowledge of phenolic and non-phenolic materials extracted from grasses used in biomass production streams had been largely limited to small water-soluble molecules. Earlier research by Vismeh et al. covered this extensively [1-2]. Second, most available methods for analysis of large lignin molecules either yield average-structure-outputs (NMR, SEC) including linkage types and lignin monomer compositions or they degrade large molecules to smaller units as a part of analytical strategies (DFRC [3-4] and thioacidolysis [5-6]) that converted lignin to small and water-soluble analytes. As a result of both criteria, limited information has been available regarding detailed molecular structures within lignin and their involvement in cross-linking to hemicellulose and cellulose.

Recent HPLC-MS and MS/MS analyses of lignin constituents have been used extensively by Morreel et al., [7-8] and revealed that extraction of grass lignin in conventional solvents including methanol, acetonitrile, acetone and dioxane, followed by MS analysis can be used to annotate individual molecules. This, in combination with NMR studies of purified and/or synthesized lignin compounds, offers a promising platform for studying larger phenolic molecular structures of lignin. However, this approach often relied on pre-knowledge of the subunits used in lignin biosynthesis, meaning that this approach retains a targeted approach focused on anticipated molecular structures.

This dissertation has presented an original untargeted UHPLC-MS and MSMS approach towards analysis of intact lignin molecules. Our initial efforts revealed large association of flavone tricetin into grass lignin. Conjugates of tricetin with monolignols and phenolic acids in other monocots had been already reported before this [9-10], but these compounds were viewed more as specialized metabolites than indicators of lignin chemistry. At the same time as this work was

initiated, del Rio et al., [11] reported of triclin as an abundant component of wheat straw lignin [11], and the earliest publications reporting characterization and/or assignment of triclin conjugates in corn stover and other grasses were presented by Lan et al. [12-13].

In Chapter 2 of this dissertation, it was demonstrated that triclin (as well as other lignin identities which were later discussed in Chapter 3) is incorporated into large phenolic structures up to at least 4 kDa. It was discussed in Chapter 2 that phenolic components of grass lignin form a material of remarkable chemical complexity as their molecular size increases, as evident from mass spectra showing peaks at every nominal mass measured. Abundance of individual larger lignin molecules decreases as their size increases, and fewer ions are detected for each m/z value with increasing mass. Moreover, the prevalence of isobaric and isomeric species also increase as the size increases, but the combined levels of low-abundant large molecules of lignin still represent a great fraction, probably the majority, of its total mass. By generating wide mass window MS/MS spectra, it was seen that fragments of triclin, *p*-coumarate, monolignols, and other fragment ions were common building blocks of larger lignin molecules.

In Chapter 3, a deeper evaluation of incorporation of phenolic acids into monocot lignins was performed. Hydroxycinnamic acids, namely *p*-coumaric acid and ferulic acid, were detected in structures of a large number of monocot extractive molecules. The most abundant compound in methanol extract of corn stover, that is also present at lower levels in other grasses, is 1-*p*-coumaroyl-3-feruloylglycerol. Additional glycerol-containing compounds detected in extracts of all grasses, but not hardwoods, are made by one or two esterifications of phenolic acids on glycerol. Detection of multiple isomers of dimers of *p*-coumaroylferuloylglycerol suggested diferulate crosslinking involves glycerol esters; an already-known set of cross-linkers between lignin and hemicellulose [2 , 14]. Lignin coupling of 1-*p*-coumaroyl-3-feruloylglycerol to triclin

was also detected in corn stover, suggesting that glycerol esters are subject to oligomerization reactions known previously for monolignol couplings.

Later in Chapter 3, polymerization of sinapyl *p*-coumarates and guaiacylglyceryl units on bis-(sinapyl *p*-coumarate) platform was demonstrated. This along with previous findings about glycerol derivatives opened up a new venue for deeper examination of phenolic acid contributions to the properties and chemistry of grass lignin. Such contribution may have been underestimated before [15].

In Chapter 4, the need for a stronger solvent system to analyze mixtures of large phenolic molecules with UHPLC-MS/MS was addressed. GVL allowed for reliable ionization that is simpler to interpret compared to ionization done in methanol solvent system. After demonstration of GVL's ability to separate large phenolic molecules in a gradient with water using reversed phase chromatography mode (C18 stationary phase), GVL demonstrated alternative retention selectivity relative to methanol for various phenolic constituents of corn stover extract. Use of GVL for liquid chromatography and ESI-MS and their combination as UHPLC-MS was reported here for the first time.

This approach described in this dissertation advocates for untargeted analysis. Though confirmation of structure by chemical synthesis has clear importance and utility, the enormous number and abundance of higher molecular mass lignin constituents probably precludes synthesis or purification of the thousands of lignin constituents. Mass spectrometric approaches have potential to reveal many details of lignin chemistry, and it is recommended that alternative strategies be developed, perhaps including ion mobility spectrometry, to annotate these substances. Despite the challenges of lignin chemical complexity, this work provided a reliable platform for interrogation of lignin structure, contribution of different monomers in its formation,

and detection of chemical transformations that occur during biomass pretreatments. It is hoped that such information will aid current and future efforts to optimize conversion of biomass to fuels and chemical feedstocks.

REFERENCES

REFERENCES

- (1.) Vismeh, R. multifaceted metabolomics approaches for characterization of lignocellulosic biomass degradation products formed during ammonia fiber expansion pretreatment. Michigan State University, East Lansing, MI, USA, 2012.
- (2.) Vismeh, R.; Lu, F. C.; Chundawat, S. P. S.; Humpala, J. F.; Azarpira, A.; Balan, V.; Dale, B. E.; Ralph, J.; Jones, A. D., *Analyst* **2013**, *138* (21), 6683-6692.
- (3.) Lu, F.; Ralph, J., *Abstracts of Papers of the American Chemical Society* **1996**, *211*, 110-CELL.
- (4.) Lu, F. C.; Ralph, J., *Journal of Agricultural and Food Chemistry* **1997**, *45* (7), 2590-2592.
- (5.) Rolando, C.; Monties, B.; Lapierre, C., Thioacidolysis. In *Methods in Lignin Chemistry*, Lin, S. Y.; Dence, C. W., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1992; pp 334-349.
- (6.) Lapierre, C.; Monties, B.; Rolando, C., *Journal of Wood Chemistry and Technology* **1985**, *5* (2), 277-292.
- (7.) Morreel, K.; Kim, H.; Lu, F. C.; Dima, O.; Akiyama, T.; Vanholme, R.; Niculaes, C.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., *Analytical Chemistry* **2010**, *82* (19), 8095-8105.
- (8.) Morreel, K.; Dima, O.; Kim, H.; Lu, F. C.; Niculaes, C.; Vanholme, R.; Dauwe, R.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., *Plant Physiology* **2010**, *153* (4), 1464-1478.
- (9.) Bouaziz, M.; Veitch, N. C.; Grayer, R. J.; Simmonds, M. S. J.; Damak, M., *Phytochemistry* **2002**, *60* (5), 515-520.
- (10.) Nakajima, Y.; Yun, Y. S.; Kunugi, A., *Tetrahedron* **2003**, *59* (40), 8011-8015.
- (11.) del Rio, J. C.; Rencoret, J.; Prinsen, P.; Martinez, A. T.; Ralph, J.; Gutierrez, A., *Journal of Agricultural and Food Chemistry* **2012**, *60* (23), 5922-5935.
- (12.) Lan, W.; Lu, F. C.; Regner, M.; Zhu, Y. M.; Rencoret, J.; Ralph, S. A.; Zakai, U. I.; Morreel, K.; Boerjan, W.; Ralph, J., *Plant Physiology* **2015**, *167* (4), 1284-U265.
- (13.) Lan, W.; Morreel, K.; Lu, F. C.; Rencoret, J.; del Rio, J. C.; Voorend, W.; Vermerris, W.; Boerjan, W.; Ralph, J., *Plant Physiology* **2016**, *171* (2), 810-820.

- (14.) Bunzel, M.; Allerdings, E.; Ralph, J.; Steinhart, H., *Journal of Cereal Science* **2008**, 47 (1), 29-40.
- (15.) Hatfield, R.; Ralph, J.; Grabber, J. H., *Planta* **2008**, 228 (6), 919-928.