



Green  
Chemistry

**Compositional Analysis of Organosolv Poplar Lignin by  
Using High-Performance Liquid Chromatography/High-  
resolution Multi-stage Tandem Mass Spectrometry**

Journal:	<i>Green Chemistry</i>
Manuscript ID	GC-ART-10-2020-003398.R1
Article Type:	Paper
Date Submitted by the Author:	10-Dec-2020
Complete List of Authors:	Zhang, Jifa; Department of Chemistry, Purdue University Jiang, Yuan; Department of Chemistry, Purdue University Easterling, Leah F; Department of Chemistry, Purdue University Anster, Anton; The University of Tennessee Knoxville Li, Wanru; Department of Chemistry, Purdue University Alzarini, Kawthar; Purdue University System, Chemistry Dong, Xueming; Department of Chemistry, Purdue University Bozell, Joseph; University of Tennessee, Kenttamaa, Hilka; Department of Chemistry, Purdue University

SCHOLARONE™  
Manuscripts

## **Compositional Analysis of Organosolv Poplar Lignin by Using High-Performance Liquid Chromatography/High-resolution Multi-stage Tandem Mass Spectrometry**

Jifa Zhang,<sup>a</sup> Yuan Jiang,<sup>a</sup> Leah F. Easterling,<sup>a</sup> Anton Anstner,<sup>b</sup> Wanru Li,<sup>a</sup> Kawthar Z. Alzarieni,<sup>a</sup> Xueming Dong,<sup>a</sup> Joseph Bozell,<sup>b</sup> and Hilikka I. Kenttämäa<sup>\*a</sup>

<sup>a</sup> Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, Indiana 47907, United States

<sup>b</sup> University of Tennessee, Center for Renewable Carbon, Knoxville, Tennessee, United States

\*Corresponding author: Hilikka I. Kenttämäa

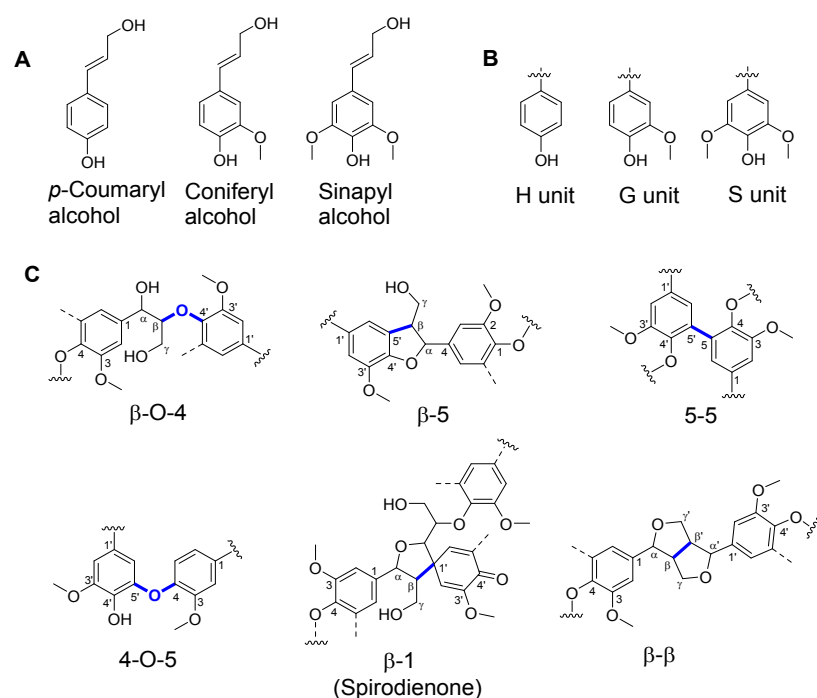
E-mail: [hilkka@purdue.edu](mailto:hilkka@purdue.edu)

## Abstract

Organosolv treatment is an efficient and environmentally friendly process to degrade lignin into small compounds. The capability of characterizing the individual compounds in the complex mixtures formed upon organosolv treatment is essential for the optimization of the further lignin conversion processes and for the rational genetic engineering of plants used to produce lignin in order to improve lignin properties. In this study, an organosolv poplar lignin sample was initially analyzed by high-resolution mass spectrometry coupled with negative-ion mode electrospray ionization ((-)ESI HRMS). Lignin monomers and dimers were found to constitute the majority of the compounds in the organosolv lignin sample. Larger lignin oligomers, such as trimers and tetramers, and some nonlignin compounds, were also detected. A high-performance liquid chromatograph/linear quadrupole ion trap/orbitrap mass spectrometer capable of multi-stage high-resolution tandem mass spectrometry experiments (HRMS<sup>n</sup>), equipped with an (-)ESI source (HPLC/(-)ESI HRMS<sup>n</sup>), was employed to separate the unknown compounds in the organosolv mixture and to obtain structural information for the deprotonated compounds via collision-activated dissociation (CAD) HRMS<sup>n</sup> experiments. To improve the understanding of the CAD behavior of deprotonated lignin-related compounds, 16 deprotonated model compounds with different functionalities and linkage types were examined. This approach enabled the assignment of likely structures for several lignin monomers, dimers, trimers, and tetramers, and some non-lignin compounds, most likely fatty acids. Based on the proposed structures, compounds in the organosolv lignin sample contain  $\beta$ -O-4, 5-5,  $\beta$ -5, and possibly also 4-O-5 linkages. Most compounds contain G- and S-monomeric units although a small amount of H-units were also detected.

## Introduction

The depletion of fossil fuels, compounded by the deterioration of the global environment caused by their production and extensive usage, has prompted the search for alternative renewable sources for energy and materials.<sup>1-7</sup> Lignocellulosic biomass holds great potential to fulfill this goal due to its high abundance, easy accessibility, and renewability.<sup>2,3</sup> It is composed of three major components: cellulose, hemicellulose, and lignin.<sup>2</sup> Lignin is the most abundant aromatic biopolymer in nature and therefore an attractive source for aromatic compounds. Lignin has an extremely complex, three-dimensional structure as it is formed via radical coupling reactions of three major monolignols, *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (**Figure 1**).<sup>8-16</sup> These monolignols correspond to *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively, in lignin. These basic units are cross-linked together by different inter-unit linkages, among which the  $\beta$ -O-4 is the most abundant.<sup>15,16</sup> Other linkage types include  $\beta$ -5, 5-5, 4-O-5,  $\beta$ -1 and  $\beta$ - $\beta$  linkages (**Figure 1**).<sup>15,16</sup>



**Figure 1.** A. The three most abundant monolignols of lignin. B. The three most common monomeric units in lignin. C. Some of the common linkages in lignin (highlighted in blue).

Many efforts have been dedicated to converting lignin into biofuel or valuable chemicals.<sup>2,4-6,17,18</sup> Organosolv treatment is an environmentally friendly and efficient process to isolate and degrade lignin into small compounds.<sup>19-22</sup> During the organosolv treatment, many of the labile  $\beta$ -O-4 bonds in lignin are cleaved. The extent to which they are cleaved is determined by the exact conditions used.<sup>6</sup> Further, various reactions can take place at the aliphatic hydroxyl groups: they can be oxidized to aldehydes, ketones, or carboxylic acids or converted to ether or ester functionalities.<sup>23</sup> Repolymerization reactions among the produced monomers can also occur to produce bigger oligomers.<sup>24-26</sup> Therefore, a very complex mixture of organic compounds is usually generated. The structures of most of the unknown compounds in these mixtures are unknown.<sup>23,27</sup> This presents a problem as a prerequisite for the optimization of further upgrading processes is the

ability to characterize the compounds in these complex mixtures. Furthermore, this is also a prerequisite for rational genetic engineering of plants in order to improve the plant raw materials for industrial usage.<sup>28</sup> Modifications of lignin content and composition via genetic engineering could have great economic and environmental benefits.<sup>28,29</sup>

Although many analytical techniques are available for the analysis of mixtures generated upon degradation of lignin, including nuclear magnetic resonance (NMR) spectroscopy, Fourier-transform infrared resonance (FT-IR) spectroscopy, and gel permeation chromatography (GPC), most of these techniques only provide bulk information for the mixtures.<sup>8,19,21,30-33</sup> Analytical pyrolysis coupled to gas chromatography with mass spectrometric detection (py-GC/MS) can be used to attempt to analyze some of the individual compounds in the lignin degradation mixtures.<sup>34,35</sup> However, only volatile lignin-related compounds can be examined by this method and their structures may have been changed during pyrolysis. Identification of the individual compounds in the lignin degradation products, therefore, requires other methods.

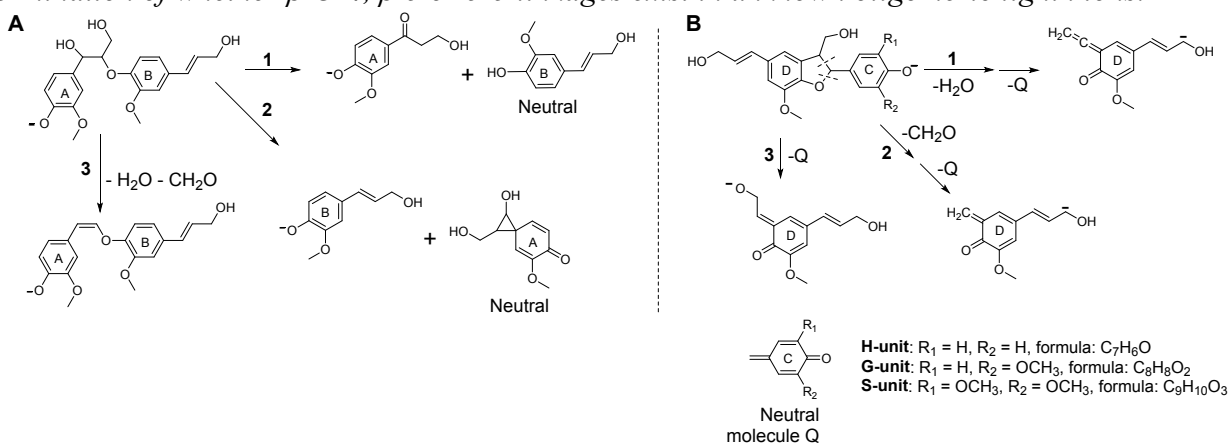
HPLC/high-resolution multi-stage tandem mass spectrometry (HRMS<sup>n</sup>) based on collision-activated dissociation (CAD) is a powerful method for the separation and structural elucidation of organic compounds in complex mixtures. Negative-ion mode electrospray ionization is widely used for the ionization of lignin-related compounds for mass spectrometric analysis<sup>23,27,36</sup> as many lignin-related compounds have a relatively acidic phenol group. Upon (-)ESI, the phenol group often gets deprotonated to afford phenoxide anions. Quantum chemical calculations at the M06-2X/6-311++G(d,p) level of theory carried out on a deprotonated  $\beta$ -O-4 type lignin dimer in the gas phase indicate<sup>37</sup> that the energy required to remove a proton from the phenol group is about 15 kcal mol<sup>-1</sup> less than removal from an aliphatic hydroxyl group (in spite of stabilizing hydrogen-bonding interactions), suggesting that deprotonation predominantly occurs on the phenol moiety in the gas phase.

Previous HRMS<sup>n</sup> studies based on (-)ESI and CAD experiments have provided some structural information for deprotonated unknown compounds in organosolv lignin samples.<sup>23,27</sup> The limitations of these studies mostly arise from insufficient knowledge of the CAD behavior of deprotonated lignin-related model compounds.<sup>23,27</sup> In a (-)ESI HRMS<sup>n</sup> study, the most important fragmentation pathways and likely mechanisms were elucidated for deprotonated oligomeric lignin  $\beta$ -O-4 model compounds.<sup>37</sup> Their fragmentation can be categorized into three major pathways (for a dimer, see Scheme 1A): 1) elimination of the charge-remote unit(s) (B in Scheme 1A) via the cleavage of the C-O bond to form fragment ions containing unit A; 2) elimination of the charged end unit A via the cleavage of a C $\beta$ -O bond to form fragment ions containing unit B; and 3) elimination of water and formaldehyde from the aliphatic hydroxyl groups in the side chain. Upon CAD, the formation of the two fragment ions of pathways 1 and 2 involve the cleavage of the  $\beta$ -O-4 linkage (**Scheme 1A**).<sup>37-39</sup> Observation of these fragment ions for an unknown oligomeric ion implies that the unknown analyte contains a  $\beta$ -O-4 linkage.

On the other hand, a previous fragmentation study of one deprotonated  $\beta$ -5 type lignin dimer model compound by using the (-)APCI HRMS<sup>n</sup> technique suggested that deprotonated  $\beta$ -5 type lignin dimers can lose their charged end unit A as a quinone methide-like neutral molecule (Q; see Scheme 1B) upon CAD in all three major fragmentation pathways.<sup>38</sup> Pathway 1 involves the elimination of water followed by the elimination of the neutral molecule Q while pathway 2 involves the elimination of formaldehyde followed by the elimination of Q. Pathway 3 only involves the elimination of Q.<sup>38</sup> Based on the proposed pathways, the eliminated quinone methide-like neutral molecule Q has the formula C<sub>7</sub>H<sub>6</sub>O, C<sub>8</sub>H<sub>8</sub>O<sub>2</sub>, and C<sub>9</sub>H<sub>10</sub>O<sub>3</sub> for the H-, G-, and S-monomeric units, respectively (**Scheme 1B**). Thus, the elimination of the charged end unit A as

a neutral molecule Q with a formula of  $C_7H_6O$ ,  $C_8H_8O_2$ , or  $C_9H_{10}O_3$  may be indicative of the presence of a  $\beta$ -5 linkage in an unknown ion.

Examination of CAD of five deprotonated oligomeric lignin model compounds with a 5-5 linkage by using (-)ESI HRMS<sup>n</sup> demonstrated that the 5-5 linkage stays intact upon CAD.<sup>39</sup> For deprotonated compounds with two or more linkages, the characteristic dissociation behaviors of the  $\beta$ -O-4 and  $\beta$ -5 linkages can be identified based on the dissociation patterns shown in **Scheme 1**. The dimeric 5-5 unit in deprotonated trimers or bigger oligomers does not fragment further upon CAD.<sup>39</sup> *These unique dissociation behaviors identified for model compounds should enable the determination of whether  $\beta$ -O-4,  $\beta$ -5 or 5-5 linkages exist in unknown oligomeric lignin ions.*



**Scheme 1.** Fragmentation pathways reported for deprotonated lignin model compounds containing a  $\beta$ -O-4 linkage<sup>37-39</sup> (left, part A) and for a compound containing a  $\beta$ -5 linkage<sup>38</sup> (right, part B). Based on our unpublished data, the structures indicated for the ionic products may be incorrect for part B.

During organosolv treatment, various reactions can take place on the aliphatic hydroxyl groups of lignin monomers and oligomers, such as oxidation to a keto, aldehyde or carboxylic acid functionality or conversion to an ether or ester functionality.<sup>23,40</sup> These new functionalities can make the fragmentation of the ionized modified compounds different from that of their native forms. The characteristic fragmentation patterns of some simple deprotonated lignin model compounds with an aldehyde, carboxylic acid, or ester functionality and compounds with two functionalities in the side chain have been investigated using the (-)ESI HRMS<sup>n</sup> technique.<sup>23,36</sup> However, the fragmentation patterns of several relevant, simple lignin-related model compounds (e.g., alcohols, ketones, and catechols), and some model compounds with two functionalities in the side chain (e.g.,  $\beta$ -keto esters), have not been examined.<sup>23,36</sup> Further, while the majority of lignin is composed of the H-, G- and S-monomers discussed above, other minor monomeric units can also be found, such as catechol and 5-hydroxyl G, which can be considered as hydroxylated H- and G-monomers, respectively.<sup>28,41-43</sup> The fragmentation patterns of these deprotonated monomers have not been reported. Therefore, the fragmentation behavior of deprotonated lignin model compounds with hydroxy, aldehyde, keto or catechol functionalities, compounds with two functionalities in the side chain, and some other compounds not previously studied were investigated here to be able to more accurately interpret the fragmentation behavior observed for deprotonated unknown lignin-related compounds. Also, deprotonated 5-5 type lignin dimer model compounds, deprotonated 2,2'-biphenol and deprotonated 5,5'-bisvanillin containing an aldehyde

functionality, were studied to better understand the dissociation of deprotonated compounds with 5-5 linkages and different functional groups in the side chains.

In this study, the (-)ESI HRMS technique was utilized to obtain chemical composition information on a poplar organosolv lignin sample. Further, a HPLC/(-)ESI HRMS<sup>n</sup> (wherein n refers to the number of ion isolation steps) method based on CAD was employed to provide structural information for many of the unknown degradation products in the organosolv lignin sample. The previously published diagnostic fragmentation patterns determined for deprotonated model compounds with  $\beta$ -O-4,<sup>37</sup>  $\beta$ -5,<sup>38</sup> and 5-5 linkages,<sup>39</sup> as well as the fragmentation patterns of the model compound discussed above, were utilized to assign likely structures for the major unknown compounds in the organosolv lignin mixture.

## Experimental section

### Materials

Unless otherwise stated, all commercially available chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) and were used as received. LC/MS grade water, methanol, and acetonitrile were purchased from Fisher Scientific (Pittsburgh, PA, USA).

Ethyl 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanoate and 5,5'-bisvanillin were synthesized in house. The details of the syntheses and the NMR spectra can be found in supporting information (SI). The organosolv poplar lignin sample was supplied by Dr. Bozell of the University of Tennessee Center for Renewable Carbon obtained using using a previously published method.<sup>40</sup> The organosolv poplar lignin sample was generated at a temperature of 160 °C, an H<sub>2</sub>SO<sub>4</sub> concentration of 0.1 M, and a reaction time of 120 min.

### Instrumentation

All experiments were performed using a Thermo Scientific linear quadrupole ion trap (LQIT) mass spectrometer coupled with a high-resolution orbitrap mass analyzer. The instrument was equipped with an electrospray ionization (ESI) source that was operated in negative ion mode. The LTQ Tune Plus interface and Xcalibur 2.1 software were used to operate the instrument. The linear quadrupole ion trap was filled with 2 mTorr of helium. Nominal pressure of approximately  $1.5 \times 10^{-11}$  Torr was maintained in the Orbitrap vacuum manifold, as read by a cold ion gauge (IKR 270, Pfeiffer Vacuum, Asslar, Germany). A resolution of 60,000 (FWHM at  $m/z$  200) was used for all high-resolution measurements. For (-)ESI HRMS analysis, a solution of the organosolv poplar lignin sample was prepared at a concentration of 2 mg/mL in acetonitrile and water (1:1, v/v, 1 mL). The solution was introduced into the mass spectrometer via a syringe drive at a flow rate of 10  $\mu\text{L min}^{-1}$ . To facilitate the formation of a stable ESI spray, the sample solution was combined with acetonitrile and water (50/50 (v/v)) eluting from a Finnigan Surveyor MS Pump Plus via a tee connector at a flow rate of 100  $\mu\text{L min}^{-1}$ . The resulting mixture was introduced into the ESI source and the analytes were ionized using (-)ESI. A N<sub>2</sub> sheath gas flow rate of 30 (arbitrary units), N<sub>2</sub> auxiliary gas flow rate of 10 (arbitrary units), spray voltage of 3.5 kV, ion transfer capillary temperature of 275 °C, and capillary voltage -10 V were utilized.

For CAD studies, each of the 16 lignin model compounds was prepared at a concentration of approximately 1 mM in LC/MS grade methanol. Each solution was introduced into the mass spectrometer via a syringe drive at a flow rate of 10  $\mu\text{L min}^{-1}$ . The analytes were ionized by (-)ESI

and the analyte ions were transferred into the linear quadrupole ion trap of the mass spectrometer. An isolation width of 2  $m/z$  units was used to isolate the ions of interest for CAD. A normalized collision energy of 35, a  $q$  value of 0.25, and an activation time of 30 ms were used to induce fragmentation of the isolated ions in the ion trap. The most abundant fragment ions were isolated and subjected to CAD in the ion trap in MS<sup>3</sup> experiments. This process was repeated until no more fragmentation could be observed (up to MS<sup>5</sup> experiments). All the ionized analytes and fragment ions were transferred into the high-resolution orbitrap mass analyzer for high-resolution analysis. The reproducibility of the relative abundances of the fragment ions was  $\pm 15\%$ .

HPLC separation was performed using a Surveyor Plus HPLC equipped with a quaternary pump, a photodiode array (PDA) detector, and an autosampler. An organosolv lignin solution was prepared at a concentration of 10 mg/mL in acetonitrile and water (1:1, v/v) and 10  $\mu$ L of the solution was loaded onto a reversed-phase Zorbax SB Phenyl column (4.6  $\times$  250 mm, particle size 5  $\mu$ m, Agilent Technologies, Santa Clara, CA, USA) by using the partial loop injection mode. The mobile phase solvents used were water (A) and acetonitrile (B) at a flow rate of 600  $\mu$ L/min. Both solvents were doped with 0.1 % (v/v) of formic acid. The nonlinear gradient used was as follows: 0 min, 70% A, 30% B; 6.0 min, 70% A, 30% B; 36.0 min, 20% A, 80% B; 36.1 min, 5% A, 95% B; 45.0 min, 5% A, 95% B; 45.1 min, 70% A, 30% B; 50.0 min, 70% A, 30% B. The column temperature was kept at 40 °C and the sample tray was refrigerated at 4 °C. 400  $\mu$ L of a wash solvent of 50:50 (v/v) water:methanol was injected between each sample injection to prevent sample carryover between injections. After HPLC separation, the analytes were ionized by using (-)ESI under the following conditions: N<sub>2</sub> sheath gas flow rate of 60 (arbitrary units), N<sub>2</sub> auxiliary gas flow rate of 20 (arbitrary units), spray voltage of 3.5 kV, ion transfer capillary temperature of 275 °C, and capillary voltage of -10 V. The data-dependent scan functionality of the Xcalibur software was used to perform the CAD MS<sup>n</sup> analysis. The most abundant ions were isolated and subjected to CAD. Their most abundant fragment ions were then isolated and subjected to CAD. This process was repeated until MS<sup>4</sup> stage of ion isolation/fragmentation was reached. All the analyte and fragment ions were analyzed by using the high-resolution orbitrap mass analyzer. The reproducibility of the relative abundances of the ions was  $\pm 15\%$ .

## Results and discussion

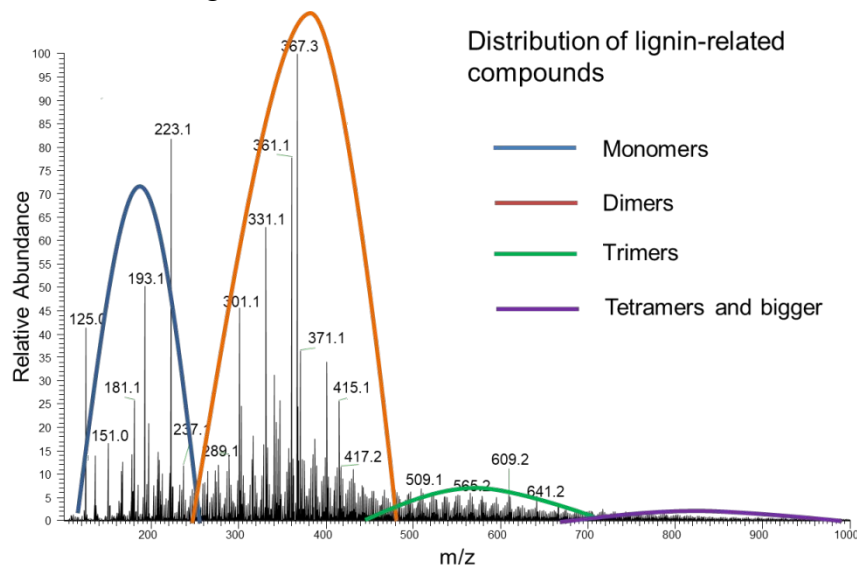
General compositional information was obtained for the organosolv lignin sample by using (-)ESI HRMS. HPLC/(-)ESI MS<sup>n</sup> experiments based on CAD were carried out to provide structural information for the individual unknown compounds in the sample. In order to better understand the dissociation behavior of deprotonated lignin degradation products, CAD of 16 deprotonated lignin-related model compounds was examined.

### High-resolution MS analysis of the organosolv poplar lignin sample

Accurate mass measurements (enabled by high resolution) revealed the elemental compositions of the ionized analytes and their degree of unsaturation and/or the number of rings in them (ring and double bond equivalence (RDBE) =  $a + 1 - b/2$  for compounds not containing nitrogen or halogen atoms, where  $a$  = the total number of carbon atoms and  $b$  = the total number of hydrogen atoms). Lignin-related compounds contain almost exclusively C, H, and O atoms. Based on the RDBE values, lignin-related ions can be classified into several subcategories. Deprotonated lignin monomers have an RDBE value of at least 4.5.<sup>23</sup> This is due to the fact that they contain one benzene ring with an RDBE value of 4. Deprotonation of these compounds increases the RDBE by 0.5. Also, any double or triple bonds in the side chains will increase the RDBE values of deprotonated analytes. Therefore, the RDBE values of deprotonated dimers



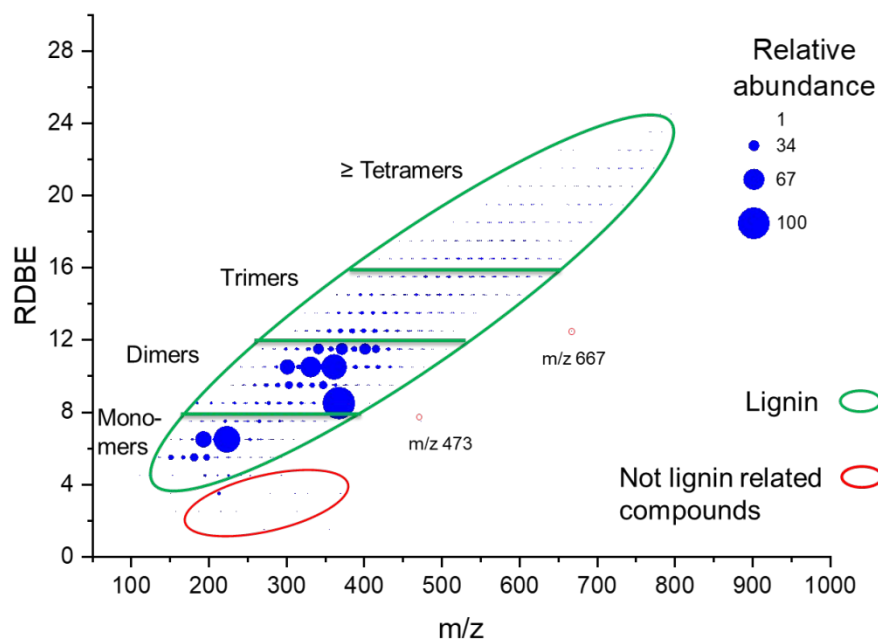
(containing two aromatic rings) are equal to or greater than 8.5.<sup>23</sup> Similarly, the RDBE values of deprotonated trimers and tetramers are at least 12.5 and 16.5, respectively. The  $m/z$  values of deprotonated lignin-related compounds correlate with their RDBE values. Therefore, the RDBE and  $m/z$  values were used to distinguish the degree of polymerization of the analytes in the organosolv lignin sample. As seen in a high-resolution mass spectrum (**Figure 2**), lignin monomers, dimers, trimers, tetramers and some bigger oligomers were observed, with monomers and dimers being most abundant.



**Figure 2.** High-resolution mass spectrum measured for the organosolv poplar sample by using (-)ESI HRMS. The approximate distributions of lignin-related monomeric, dimeric, trimeric, and tetrameric as well as bigger compounds are shown.

In addition to lignin-related compounds, carbohydrates, lipids, and other not lignin-related compounds may also exist in organosolv lignin samples.<sup>23</sup> To better visualize the distribution of the different types of compounds in the sample, a bubble chart (**Figure 3**) was constructed that shows the relationships of the  $m/z$  value and RDBE value for the ionized analytes. The size of each bubble represents the relative abundance of the ionized analyte (**Figure 3**). The majority of the analyte ions are positioned closely together to form a big “continent” of lignin-related ions (circled in green) while a few analyte ions that are not related to lignin are visible as small “offshore islands” (circled in red). Determination of the elemental compositions of the analyte ions revealed that the  $C/O$  ratio of the lignin-related compounds in the “continent” was 1.5-4.5. Other, nonlignin compounds, such as lipids, fatty acids and their esters, often have a low degree of unsaturation ( $RDBE < 4.5$ ) due to the lack of benzene rings. As can be seen in **Figure 3**, most of the analytes within the three red circles have  $RDBE$  values of less than 4.5, indicating that these analytes are not lignin-related compounds. Only a few compounds are located at the regions where the  $RDBE$  value is greater than 4.5. Ions of  $m/z$  473 have the chemical composition  $C_{29}H_{45}O_5$  and a  $C/O$  ratio of 5.8, which is much greater than that of lignin-related compounds ( $C/O$  ratio 1.5-4.5). The large  $C/O$  ratio and the observed losses of water and  $CO_2$  upon CAD of these ions suggest that they are likely to correspond to a fatty acid with a high degree of unsaturation. On the other hand, the  $C/O$  ratio of carbohydrates is generally  $\sim 1$ .<sup>23</sup> Potential oxidation reactions at the hydroxyl groups of the carbohydrates will increase the  $RDBE$  values but the  $C/O$  ratio will stay relatively unaffected. Ions of  $m/z$  667 have the chemical composition of  $C_{23}H_{23}O_{23}$  and a  $C/O$  ratio of 1, indicating that

they are ionized carbohydrates. Thus, the C/O ratios in addition to the RDBE values can be used to categorize the analytes into different compound types (**Figure 3**).

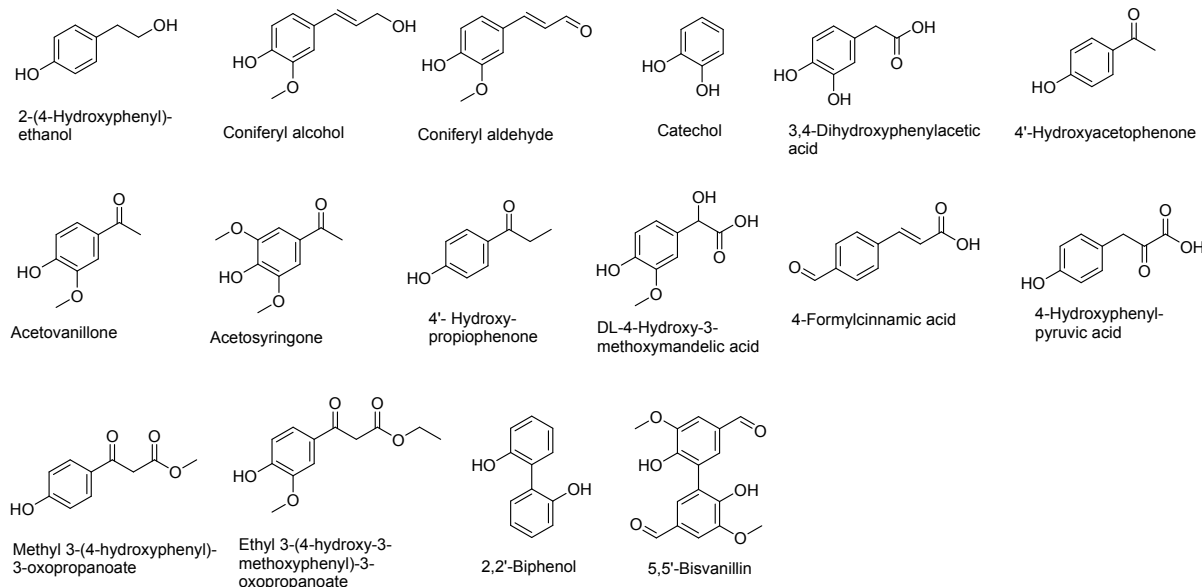


**Figure 3.** Chemical composition of the organosolv poplar lignin sample based on (-)ESI HRMS analysis. Lignin-related compounds and nonlignin compounds (most likely fatty acids) are shown inside green and red circles, respectively. For lignin-related compounds, green lines were drawn to divide them into different categories based on their RDBE.

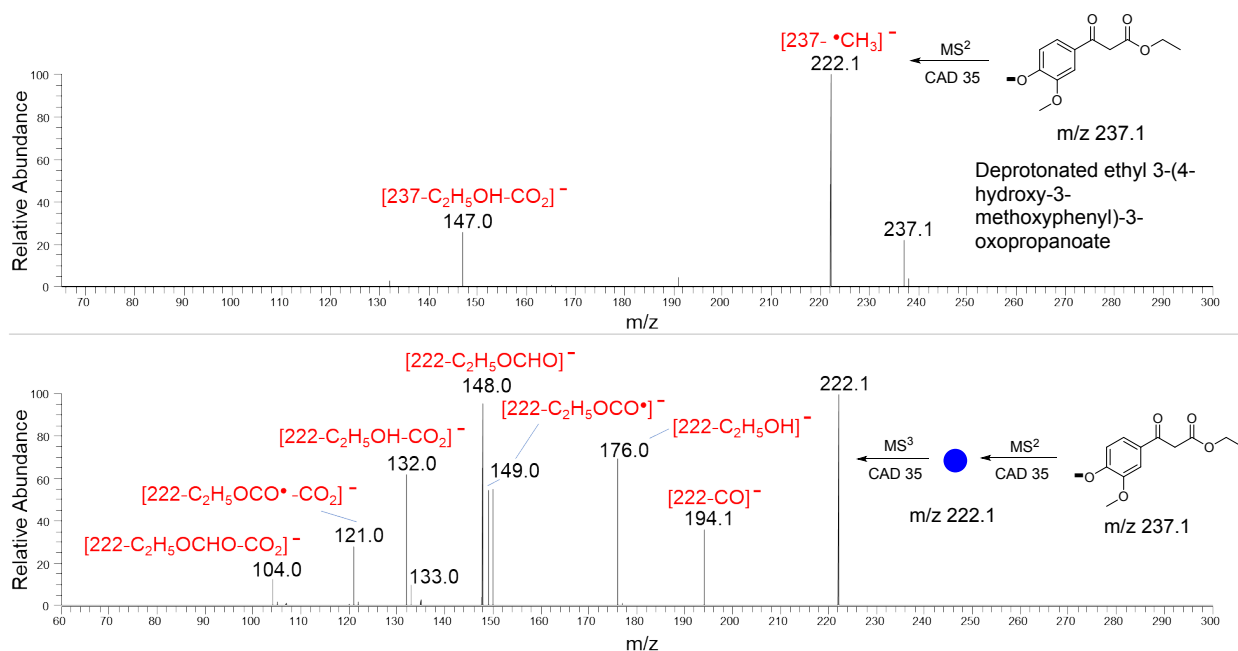
### HPLC/HRMS<sup>n</sup> analysis of the organosolv poplar lignin sample

To structurally characterize individual unknown compounds in the organosolv lignin sample, an HPLC/(-)ESI HRMS<sup>n</sup> method based on CAD of isolated analyte ions was employed. Structural elucidation of unknown lignin degradation products requires a fundamental understanding of how deprotonated lignin-related compounds fragment upon CAD. Previous studies elucidated the dissociation behavior of lignin model compounds with a  $\beta$ -O-4,<sup>37-39</sup> a  $\beta$ -5,<sup>38</sup> or a 5-5 linkage.<sup>39</sup> Dissociation studies on simple lignin model compounds, however, have only included compounds with aldehyde, carboxylic acid, and ester functionalities and a few compounds with two functionalities in the side chain.<sup>23,36</sup> Additional deprotonated lignin model compounds, such as compounds with a hydroxy- or keto-functionality, as well as catechols, were studied here to be able to more accurately interpret the fragmentation behavior observed for deprotonated unknown lignin-related compounds. Therefore, 16 deprotonated lignin-related model compounds, including catechols and compounds with hydroxy and keto functionalities, were examined.

**Model compound studies.** HRMS<sup>n</sup> experiments based on CAD were carried out for all of the ionized model compounds, their fragment ions, their fragment ions, and so on until no further fragmentation was observed (up to MS<sup>5</sup>). The results are shown in **Table S1** and are summarized below. The MS<sup>2</sup> and MS<sup>3</sup> spectra of deprotonated ethyl 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanoate are shown as an example (**Figure 5**).



**Figure 4.** Model compounds studied.



**Figure 5.** CAD MS<sup>2</sup> and MS<sup>3</sup> spectra of deprotonated ethyl 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanoate at a normalized collision energy of 35.

Deprotonated lignin model compounds with methoxy groups readily lose methyl radicals and this is often the first observed fragmentation event. For deprotonated lignin-related compounds with several methoxy groups, losses of several methyl radicals were often observed upon the initial consecutive fragmentation events (**Table S1**). Upon further CAD, elimination of one or more CO molecules were often observed. Occasionally, CO<sub>2</sub> was lost after the loss of methyl radical(s).<sup>36</sup> The fragmentations of deprotonated lignin model compounds are discussed in detail below.

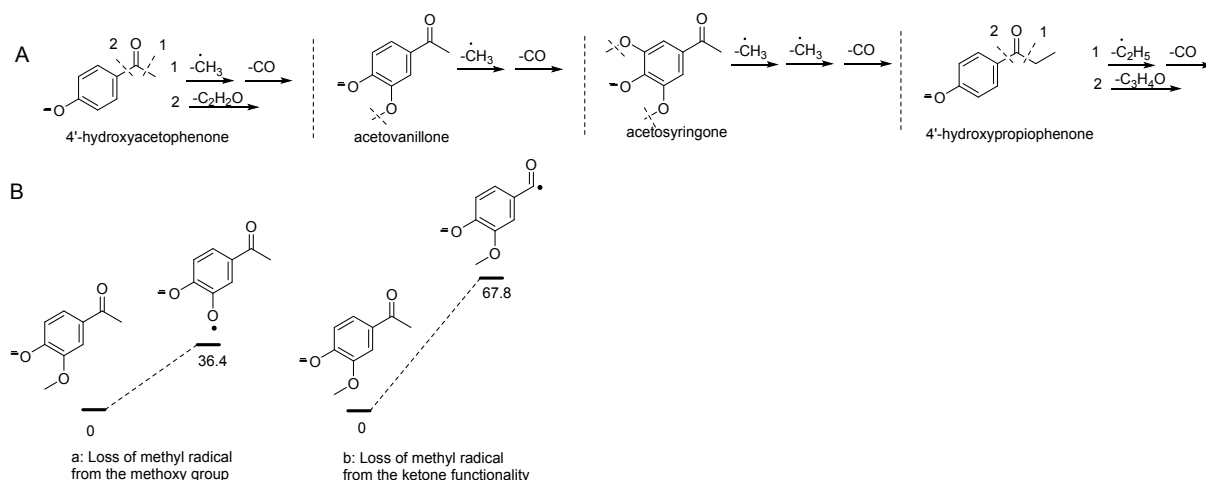
*Hydroxy.* The deprotonated lignin model compounds containing an aliphatic hydroxy group predominantly lost water upon CAD, as is evident from the data measured for deprotonated coniferyl alcohol and 2-(4-hydroxyphenyl)ethanol (**Table S1**). Formaldehyde loss is also possible if a primary hydroxy group is attached to an alkyl chain, which is evident from the data measured for deprotonated 2-(4-hydroxyphenyl)ethanol (**Table S1**).

*Aldehyde.* According to a previous study, deprotonated lignin-related aldehydes lack specific fragmentation patterns unique to the aldehyde functionality.<sup>36</sup> Some of the deprotonated model compounds studied here, namely, deprotonated coniferyl aldehyde and 5,5'-bisvanillin, showed elimination of an HCO• radical (**Table S1**), which likely involves the aldehyde functionality.<sup>36</sup> However, based on previously published studies, many other deprotonated aldehydes, such as vanillin, syringaldehyde, and sinapaldehyde, do not lose HCO• upon CAD.<sup>36</sup> Therefore, deprotonated aldehydes are not straightforward to identify as they fragment similarly to many other types of deprotonated compounds, such as ketones.

*Carboxylic Acid.* Deprotonation of lignin-related compounds with a carboxylic acid functionality in the alkyl side chain can occur at either the phenol or the carboxylic acid functionality, depending on the ESI solvents used.<sup>44-46</sup> Based on literature, the phenol is more likely to be deprotonated upon ESI when the aprotic solvent acetonitrile is used.<sup>44-46</sup> Therefore, all deprotonated model compounds studied here are shown to be deprotonated at the phenol site even though deprotonation at the carboxylic acid moiety may also occur for some of them. Deprotonated lignin model compounds containing a carboxylic acid functionality have been reported to readily lose CO<sub>2</sub> upon CAD.<sup>36</sup> Indeed, all the deprotonated carboxylic acids studied here, namely, 3,4-dihydroxyphenylacetic acid, 4-hydroxy-3-methoxymandelic acid, 4-formylcinnamic acid, and 4-hydroxyphenyl-pyruvic acid, showed the loss of CO<sub>2</sub> when subjected to CAD in the MS<sup>2</sup> experiments. Therefore, the loss of CO<sub>2</sub> may indicate the presence of a carboxylic acid functionality in an unknown deprotonated compound. This loss was usually observed in the MS<sup>2</sup> experiments but sometimes other fragmentations were more competitive. For example, deprotonated 2-ethoxy-3-(4-hydroxyphenyl)propionic acid and *p*-hydroxyphenyllactic acid both showed the loss of CO<sub>2</sub> in the MS<sup>3</sup> experiments rather than in the MS<sup>2</sup> experiments.<sup>36</sup>

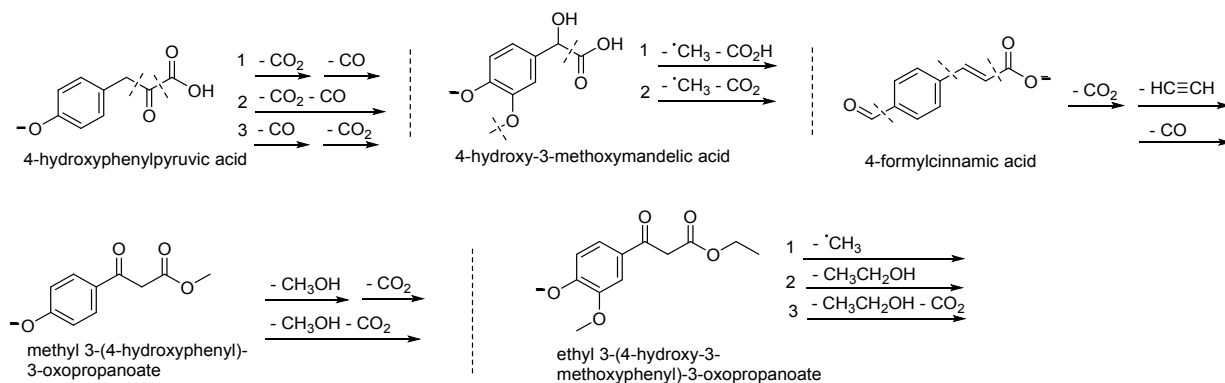
*Catechol.* Upon CAD, competitive losses of CO and H<sub>2</sub>O from deprotonated 3,4-dihydroxyphenylacetic acid were observed after the initial loss of the carboxylic acid moiety as CO<sub>2</sub> (**Table S1**). Deprotonated catechol also showed competitive losses of CO and H<sub>2</sub>O upon CAD (**Table S1**). Competition of these two pathways may indicate the presence of a catechol moiety in an unknown ion.

*Keto.* Deprotonated lignin model compounds with a keto functionality can fragment via cleavages of bonds between the carbonyl functionality and the two  $\alpha$ -carbons (**Figure 6A** and **Table S1**). These fragmentations may be indicative of the presence of a keto functionality in an unknown ion. However, other dissociation pathways, such as methyl radical losses from the methoxy groups, can be more competitive than these dissociation pathways. Quantum chemical calculations at the M06-2X/6-311++G(d,p) level of theory on deprotonated acetovanillone indicate that the loss of a methyl radical from the methoxy group adjacent to the deprotonated site is energetically more favorable than from the carbonyl functionality, as expected (**Figure 6**). Therefore, other dissociation pathways may render fragmentations at the keto functionalities uncompetitive. Due to the existence of competitive dissociation pathways, deprotonated lignin model compounds with a keto functionality may not be straightforward to identify.



**Figure 6.** A) Pathways for the loss of small neutral molecules from deprotonated lignin model compounds with a keto functionality. B) Calculated Gibbs free energy surfaces for the loss of a methyl radical from the methoxy group and from the keto-functionality in deprotonated acetovanillone at the M06-2X/6-311++G(d,p) level of theory. The free energies are shown in kcal/mol.

*Model compounds with two functionalities in the side chain.* The fragmentation patterns of deprotonated model compounds with two functional groups in the side chain largely depend on the type of the functional groups. The aliphatic carboxylic acid functionality is often lost in the first CAD event as  $\text{CO}_2$ , as observed for deprotonated 4-hydroxyphenylpyruvic acid, 4-hydroxy-3-methoxymandelic acid, and 4-formylcinnamic acid (**Figure 7** and **Table S1**).  $\text{CO}_2$  can also be lost together with another small molecule, such as  $\text{CO}$  or  $\cdot\text{CH}_3$ , as is evident for deprotonated 4-hydroxyphenylpyruvic acid and 4-hydroxy-3-methoxymandelic acid, respectively. After the loss of  $\text{CO}_2$  from deprotonated 4-formylcinnamic acid,  $\text{CO}$  can be lost from the aldehyde functionality. On the other hand, deprotonated  $\beta$ -keto esters, the methyl 3-(4-hydroxyphenyl)-3-oxopropanoate and ethyl 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanoate, readily lose their alcohol moiety upon CAD. After this, a  $\text{CO}_2$  loss was observed for both deprotonated  $\beta$ -keto esters upon CAD (**Figure 7** and **Table S1**), which must involve a rearrangement. Similarly, deprotonated vanillin has been reported to eliminate  $\text{CO}_2$  after elimination of  $\cdot\text{CH}_3$ , and based on quantum chemical calculations, a mechanism was delineated for this complex rearrangement.<sup>47</sup>



**Figure 7.** Pathways for the loss of small molecules upon CAD of deprotonated model compounds with two functionalities in the side chain.

*Dimeric lignin model compounds with an aldehyde functionality and a 5-5 linkage.* As mentioned above, the 5-5 linkage is quite recalcitrant and can stay intact after lignin degradation via an organosolv process.<sup>39,48</sup> Similarly, this linkage does not break upon CAD of deprotonated lignin model compounds.<sup>39</sup> To better understand the fragmentations of simple deprotonated 5-5 type lignin model compounds, 2,2'-biphenol (H lignin dimer with no functionalities in the side chains) and 5,5'-bisvanillin (G lignin dimer with aldehyde functionalities in the side chains) were studied. Both ions lost water upon the first CAD experiment (**Table S1**). Due to the existence of two aldehyde functionalities and two methoxy groups in deprotonated 5,5'-bisvanillin, losses of  $\bullet\text{CH}_3$ , CO, and  $\bullet\text{CHO}$ , amongst others, were also observed (**Table S1**). Thus, no losses of neutral molecules bigger than a monomer were observed. The exclusive observance of small molecule losses for deprotonated lignin dimers is indicative of the presence of a 5-5 linkage.<sup>39</sup>

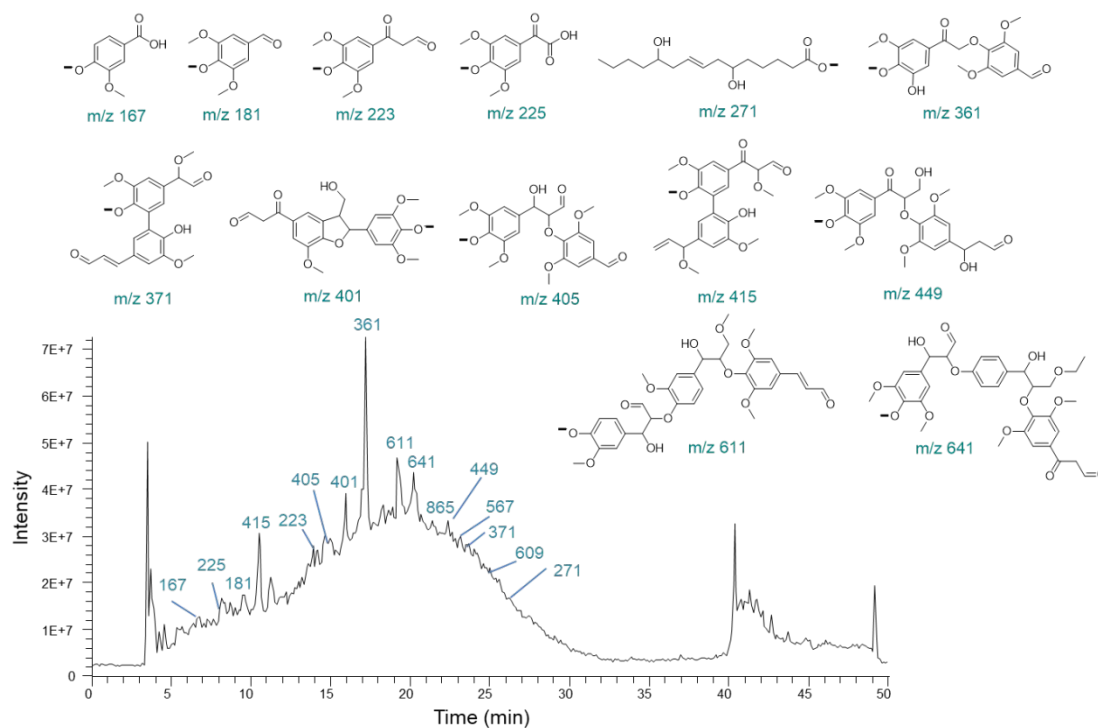
**Organosolv lignin sample.** The organosolv lignin sample was subjected to HPLC (for details of the chromatographic method, see the experimental section and SI) and the eluting compounds were ionized by (-)ESI in the linear quadrupole ion trap mass spectrometer. The deprotonated analytes were transferred into the ion trap and subjected to CAD in HRMS<sup>n</sup> experiments to obtain structural information. The knowledge obtained from the literature and in the model compound studies discussed above, combined with the CAD HRMS<sup>n</sup> spectra measured for deprotonated unknown compounds, were used to elucidate the likely structures for the major unknown compounds in the organosolv sample.<sup>23,36,38,49-51</sup>

Structures were proposed for most of the unknown analyte ions detected based on the retention times of the corresponding neutral analytes, fragmentation patterns of the ionized analytes, the elemental compositions of the ionized analytes and their fragment ions, and their RDBE values. The previously published fragmentation patterns diagnostic for the  $\beta$ -O-4,  $\beta$ -5, and 5-5 linkages were used to identify linkages in the unknown ions.<sup>37-39</sup> Deprotonated lignin model compounds with other linkage types, such as 4-O-5 and  $\beta$ -1 linkages, which have a low abundance in plants, were not specifically examined. However, in cases for which the  $\beta$ -O-4,  $\beta$ -5, and 5-5 linkages could be ruled out, other linkages, such as 4-O-5 or  $\beta$ -1, are likely present and were considered during structural elucidation.

When elucidating the structures of the lignin-related degradation products, the number of consecutive carbons in the side-chains was assumed not to exceed three because lignin is formed upon polymerization of the phenylpropanoid monolignols that contain three consecutive carbons in their side chains, and polymerization via the formation of linkages such as  $\beta$ -O-4,  $\beta$ -5, 5-5, and others does not increase the number of the consecutive carbons in the side chains.

The  $m/z$  values and possible structures of some of the deprotonated analytes are provided in the total ion current (TIC) chromatogram shown in **Error! Reference source not found.** The CAD results, HPLC retention times and the proposed structures of some analytes are listed in

**Table 1.** All the remaining information is provided in **Table S2**. Structures were proposed for 62 compounds that generated adequate fragmentations upon CAD for structural elucidation. The structural elucidation of some representative compounds in different types of analyte classes is discussed below.



**Figure 8.** HPLC total ion chromatogram obtained for the organosolv poplar sample analyzed using the HPLC/(-)ESI HRMS<sup>n</sup> method. The  $m/z$  values and proposed structures are shown for some of the major ionized analytes. One of the several possible isomeric structures is shown for the ions of  $m/z$  271 that likely correspond to an ionized fatty acid.

*Structural elucidation of a monomeric lignin-related compound.* The unknown analyte ions of  $m/z$  193 ( $C_{10}H_9O_4$ , RDBE = 6.5,

**Table 1**) are used here as an example. The RDBE value of 6.5 indicates that these ions contain only one benzene ring (RDBE = 4) and possibly two double bonds (RDBE = 1 for each). Upon CAD of these analyte ions, loss of  $\bullet CH_3$  was observed in the MS<sup>2</sup> experiment, suggesting that a methoxy group bound to the benzene ring is present. The methoxy group is most likely located at C3 because most lignins are composed of the three basic monolignols (**Figure 1**). After the loss of  $\bullet CH_3$ , losses of  $H_2C=C=O$ , CO, 2CO, or  $H_2C=C=O + CO$  were observed in the MS<sup>3</sup> experiment. The eliminated CO is likely to originate from the aromatic ring, just as in the case of deprotonated vanillin after elimination of  $\bullet CH_3$ .<sup>47</sup> The eliminated  $H_2C=C=O$  suggests the presence of a keto functionality in the side chain. Consideration of all this information leads to the conclusion that the unknown analyte ions of  $m/z$  193 likely correspond to deprotonated 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanal. Also, cleavages of the two  $\alpha$ -carbons of the keto group results in the elimination of  $H_2C=C=O$  and  $H_2C=C=O + CO$  to form product ions of  $m/z$  136 and 108, respectively.

*Structural elucidation of dimeric lignin-related compounds.* The unknown analyte ions of  $m/z$  405 ( $C_{20}H_{21}O_9$ , RDBE = 10.5, **Table 1**) are used here as an example of a deprotonated  $\beta$ -O-4 dimer. When subjected to CAD, these ions generated the most abundant fragment ions of  $m/z$  181 via the loss of a neutral molecule of MW of 224 Da and the second most abundant fragment ions of  $m/z$  223 via the loss of a neutral molecule of MW 182 Da. This fragmentation pattern matches that

shown in **Scheme 1A** for deprotonated lignin model compounds with a  $\beta$ -O-4 linkage. If the unknown analyte ions of  $m/z$  405 contained a  $\beta$ -O-4 linkage, a major fragment ion A (ions of  $m/z$  181) could be formed via the elimination of a neutral molecule B (MW 224 Da) and another major fragment ion B (ions of  $m/z$  223) could be formed via the elimination of a neutral molecule A (MW 182 Da). These findings indicate that the ions of  $m/z$  405 contain a  $\beta$ -O-4 linkage. When the most abundant fragment ions of  $m/z$  181 were isolated and subjected to CAD in an MS<sup>3</sup> experiment, the loss of  $\bullet\text{CH}_3$  was observed to generate ions of  $m/z$  166. These ions lost another  $\bullet\text{CH}_3$  in an MS<sup>4</sup> experiment. The elemental composition ( $\text{C}_9\text{H}_9\text{O}_4$ , RDBE = 5.5) and fragmentation behavior of the ions of  $m/z$  181 (

**Table 1**) suggest that they correspond to deprotonated syringaldehyde.<sup>36</sup> Based on the elemental composition of the unknown dimeric ions of  $m/z$  405, their fragment ions, their fragmentation pattern and the finding that the unknown ions contain a  $\beta$ -O-4 linkage, the structure shown in **Scheme 2** was proposed for the unknown ions of  $m/z$  405. Based on the proposed structure, deprotonated syringaldehyde of  $m/z$  181 was formed upon CAD via a pathway involving the elimination of the charged end unit A (**Scheme 2**). On the other hand, upon CAD, fragment ions of  $m/z$  223 were generated via a pathway involving the elimination of the charge-remote unit B (**Scheme 2**). The observation of these characteristic fragmentations associated with deprotonated compounds with a  $\beta$ -O-4 linkage supports the proposed structure for the unknown analyte ions of  $m/z$  405. Ions of  $m/z$  181 were likely formed via a previously proposed cyclopropane mechanism,<sup>37</sup> which requires that the carbonyl is not at the  $\text{C}\alpha$ -position. Otherwise, the elimination of the charged end unit A to form ions of  $m/z$  181 cannot occur. Therefore, the hydroxyl group was assigned to the  $\text{C}\alpha$ -position and the carbonyl group to the  $\text{C}\beta$ -position.

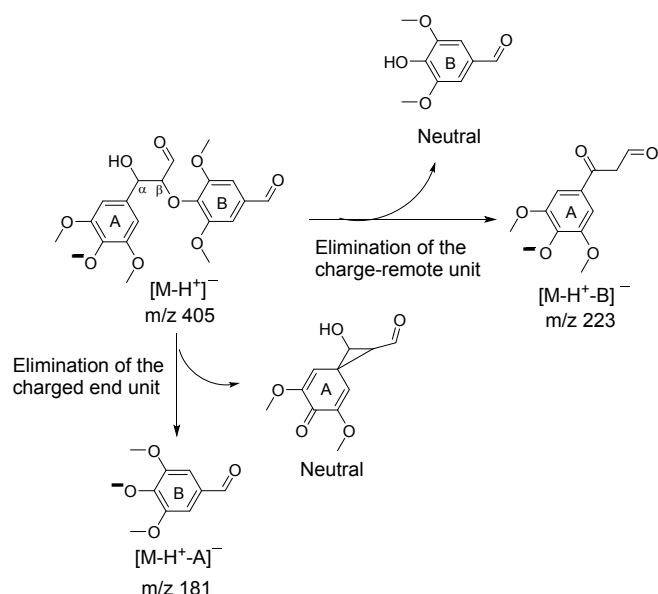
The unknown analyte ions of  $m/z$  401 ( $\text{C}_{21}\text{H}_{21}\text{O}_8$ , RDBE = 11.5,

**Table 1**) are used here as an example of a deprotonated  $\beta$ -5 dimer. Based on a previous publication, deprotonated  $\beta$ -5 dimeric model compounds fragment by elimination of the charged end unit as a quinone methide-like neutral molecule (Q, **Scheme 1B**).<sup>38</sup> This neutral molecule would have the formula  $\text{C}_7\text{H}_6\text{O}$ ,  $\text{C}_8\text{H}_8\text{O}_2$ , or  $\text{C}_9\text{H}_{10}\text{O}_3$  for the H-, G-, and S-units, respectively (**Scheme 1B**).<sup>38</sup> Upon CAD of the unknown analyte ions of  $m/z$  401, ions of  $m/z$  235 were formed via the loss of a neutral molecule Q with the formula  $\text{C}_9\text{H}_{10}\text{O}_3$ , which suggests that the ions of  $m/z$  401 may contain a  $\beta$ -5 linkage and an S-type monomeric unit. After further examination of the fragmentation of the ions of  $m/z$  401 (

**Table 1**), the ions of  $m/z$  401 were concluded not to contain a  $\beta$ -O-4 linkage because the characteristic fragmentations shown in **Scheme 1A** for  $\beta$ -O-4 type compounds were not observed.

Therefore, the ions of  $m/z$  401 were concluded to contain a  $\beta$ -5 linkage. Fragment ions of  $m/z$  235 fragmented by losing  $\bullet\text{CHO}$  to form ions of  $m/z$  206 in an MS<sup>2</sup> experiment, indicating that a formaldehyde functionality is present in the ions of  $m/z$  235. Ions of  $m/z$  235 also fragmented by losing  $\bullet\text{CH}_3$  to form ions of  $m/z$  220 in an MS<sup>2</sup> experiment, indicating that a methoxy group

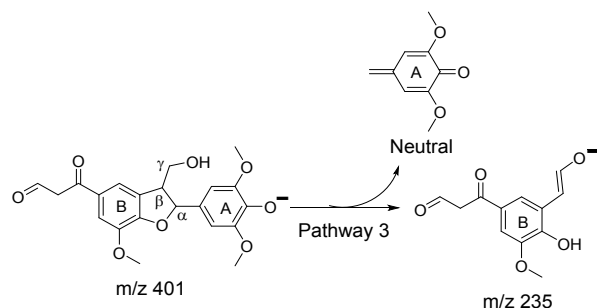




**Scheme 2.** The structure proposed for the analyte ions of  $m/z$  405 and pathways for the formation of their major fragmentation products.

exists in the ions of  $m/z$  235. When the elemental composition of the unknown ions of  $m/z$  401, their fragment ions, and their fragmentation pattern were considered, the structure shown in **Scheme 3** was proposed for these ions. According to the mechanism shown in **Scheme 1B**, the generation of the ions of  $m/z$  235 via the pathway involving the elimination of the charged end unit A requires the participation of the aliphatic hydroxyl group at the  $\gamma$ -position, which reveals that an aliphatic hydroxyl group is located at the  $\gamma$ -position. Apart from the formaldehyde functionality in the side chain of ring B, information on the elemental composition and RDBE value of the ions of  $m/z$  401 indicate that another carbonyl group exists in the side chain of ring B. The second carbonyl group is likely directly connected to ring B based on the fragmentation patterns of ions of  $m/z$  401 shown in

**Table 1.**



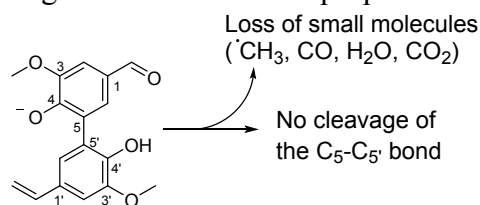
**Scheme 3.** Proposed structure of the analyte ions of  $m/z$  401 and the pathway for the generation of their major fragment ions of  $m/z$  235.

The unknown analyte ions of  $m/z$  299 ( $C_{17}H_{15}O_5$ , RDBE = 10.5,

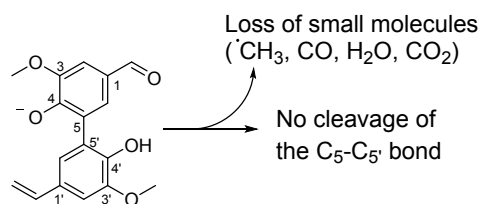
**Table 1**) are used here as an example of a deprotonated 5-5 dimer. These unknown dimeric ions showed the loss of  $\bullet CH_3$  upon CAD in an  $MS^2$  experiment, followed by loss of  $\bullet CH_3$  or both  $\bullet CH_3$

and  $\text{CO}_2$  upon CAD in an  $\text{MS}^3$  experiment. When the fragment ions generated in the  $\text{MS}^3$  experiment upon loss of  $\bullet\text{CH}_3$  were subjected to CAD, elimination of  $\text{CO}$  was observed ( $\text{MS}^4$ ). On the other hand, the fragment ions formed upon loss of both  $\bullet\text{CH}_3$  and  $\text{CO}_2$  fragmented via loss of  $\text{CO}_2$ ,  $\text{CO}_2$  and  $\text{CO}$ , or  $\text{H}_2\text{O}$  in an  $\text{MS}^4$  experiment (

**Table 1**). No neutral fragments bigger than a monomer were lost upon CAD, indicating that the ions of  $m/z$  299 contain a 5-5 linkage. Only two  $\bullet\text{CH}_3$  radicals were lost throughout the CAD experiments, indicating that two methoxy groups are present in the ions of  $m/z$  299. The fragmentations and proposed structure for ions of  $m/z$  299 are shown in



#### Scheme 4.



**Scheme 4.** Proposed structure of the analyte ions of  $m/z$  299. Only losses of small molecules were observed for this ion.

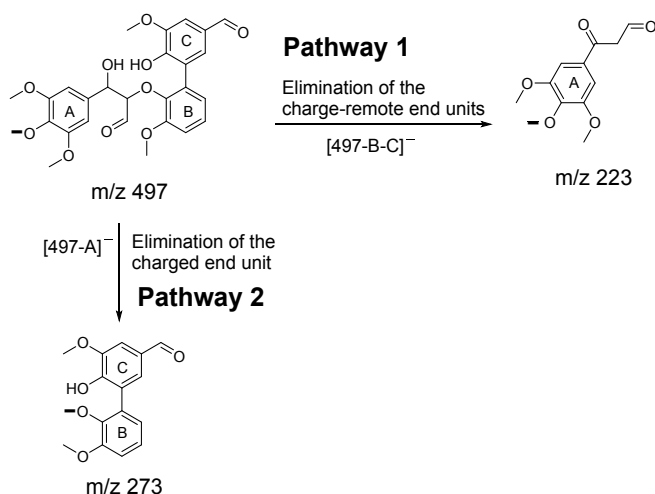
*Structural elucidation of trimeric and tetrameric lignin-related compounds.* Upon CAD of trimeric unknown analyte ions of  $m/z$  497 ( $C_{26}H_{25}O_{10}$ , RDBE = 14.5,

**Table 1**), fragment ions of  $m/z$  273 and 223 were formed via the elimination of neutral molecules of MW of 224 and 274 Da, respectively. The fragmentation pattern of deprotonated trimeric unknown analyte ions of  $m/z$  497 matches that shown in **Scheme 1A** for deprotonated lignin model compounds with a  $\beta$ -O-4 linkage. If the unknown analyte ions of  $m/z$  497 contained a  $\beta$ -O-4 linkage, a fragment ion A (ions of  $m/z$  273) could be formed via the elimination of a neutral molecule B (molecule of MW 224 Da) and a second major fragment ion B (ions of  $m/z$  223) could be formed via the elimination of a neutral molecule A (molecule of MW 274 Da). These findings indicate that the ions of  $m/z$  497 contain a  $\beta$ -O-4 linkage. The most abundant fragment ions of  $m/z$  273 lost  $\bullet CH_3$  to form fragment ions of  $m/z$  258 in an MS<sup>3</sup> experiment. These fragment ions fragmented by eliminating another  $\bullet CH_3$  in an MS<sup>4</sup> experiment (**Table S2**). No large neutral molecules were eliminated from the fragment ions of  $m/z$  273 upon CAD, indicating that the ions of  $m/z$  273 contain a 5-5 linkage. Based on the elemental composition of the unknown trimeric ions of  $m/z$  497, their fragment ions, their fragmentation pattern and the finding that they contain a  $\beta$ -O-4 and a 5-5 linkage, the structure shown in

**Scheme 5** was proposed for the unknown ions of  $m/z$  497. Based on the proposed structure, fragment ions of  $m/z$  273 are formed via the elimination of the charged end unit A as a neutral molecule and ions of  $m/z$  223 are formed via the elimination of the charge-remote end units B and C (

**Scheme 5).** The formation of ions of  $m/z$  273 via the cyclopropane mechanism<sup>37</sup> (**Scheme 1A**) requires that a hydroxyl group is located at the  $\alpha$ -position and a carbonyl group is located at the  $\beta$ -position in ions of  $m/z$  497. In addition to above structural details, based on the elemental composition of ions of  $m/z$  273, an aldehyde functionality should exist in these ions. This functionality could be attached to the 1-position of either ring B or ring C. The structure shown in

**Scheme 5** represents one of the two possible structures. Based on the proposed structure, two phenol groups are present in the unknown ions of  $m/z$  497 and either phenol group may get



**Scheme 5.** Structure and major fragmentation pathways proposed for the trimeric analyte ions of  $m/z$  497.

deprotonated upon ionization of the trimer. If the deprotonated phenol group is that attached to the C ring, fragment ions of  $m/z$  223 cannot be formed upon CAD because the charge on the phenol group of ring C cannot migrate to the unit A via the cyclopropane mechanism (**Scheme 1A**).<sup>37</sup> On

the other hand, if the phenol group on ring A is deprotonated, this ion could form both fragment ions of  $m/z$  273 and 223 upon CAD. Hence, the phenol group on ring A is shown as deprotonated in

### Scheme 5.

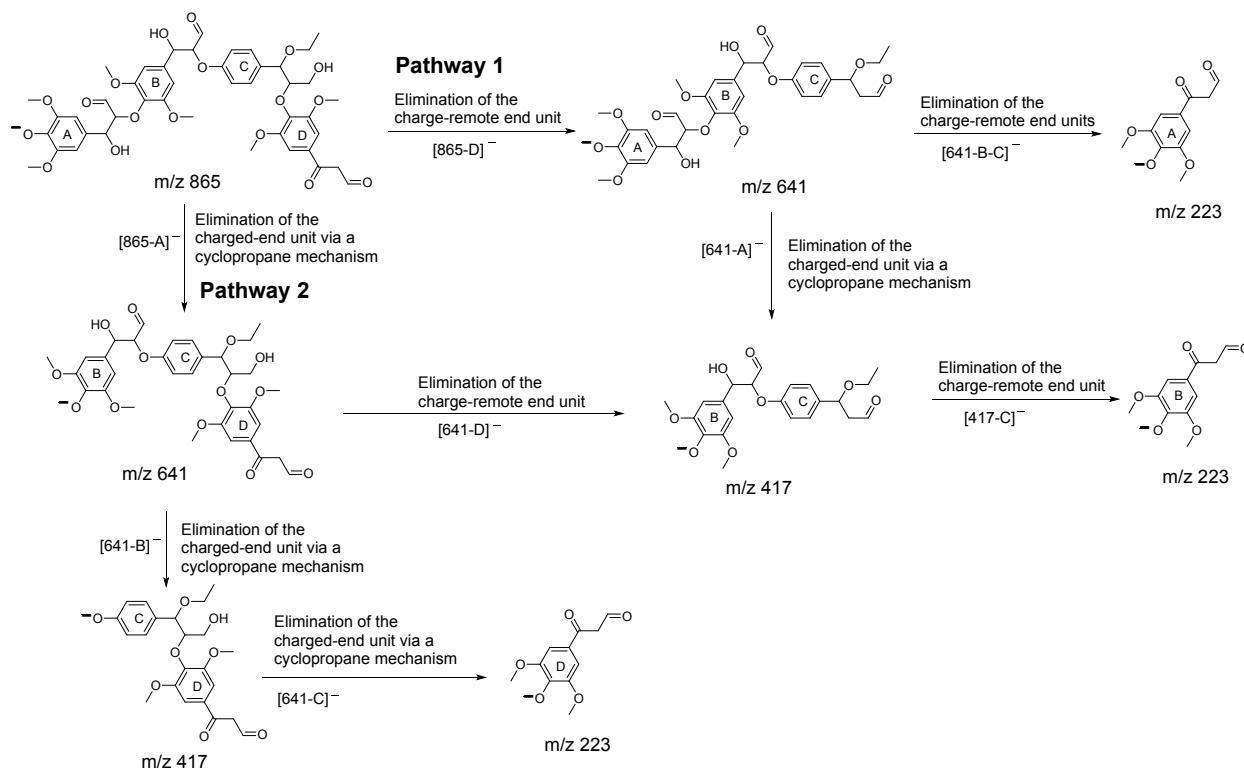
In contrast to the above example, ionized oligomeric compounds with no 5-5 linkages can lose more than two monomeric units due to the labile nature of  $\beta$ -O-4,  $\beta$ -5, and 4-O-5 linkages. This information was used to identify the unknown tetrameric analyte ions of  $m/z$  865 ( $C_{44}H_{49}O_{18}$ ,  $RDBE = 20.5$ ,

**Table 1**). Upon CAD, these ions generated the key fragment ions of  $m/z$  641. Fragment ions of  $m/z$  641 fragmented to form fragment ions of  $m/z$  417 and 223 via losses of neutral molecules of MW of 224 and 418 Da, respectively. If the unknown fragment ions of  $m/z$  641 contained a  $\beta$ -O-4 linkage, a fragment ion A (ions of  $m/z$  417) could be formed via the elimination of a neutral molecule B (molecule of MW 224 Da) and a second major fragment ion B (ions of  $m/z$  223) could be formed via the elimination of a neutral molecule A (molecule of MW 418 Da). These findings suggest that the fragment ions of  $m/z$  641 contain a  $\beta$ -O-4 linkage. After carefully examining the fragmentation patterns of ions of  $m/z$  865, all the monomeric units were concluded to be linked via  $\beta$ -O-4 linkages. Based on the elemental composition of the unknown tetrameric ions of  $m/z$  865, their fragment ions, their fragmentation pattern and the finding that they contain three  $\beta$ -O-4 linkages, the structure shown in **Scheme 6** was assigned for the unknown ions of  $m/z$  865. Further explanation for this structural assignment is provided below.

Based on the proposed structure of the unknown analyte ions of  $m/z$  865 and the pathways shown in **Scheme 1A**, the charge-remote end unit D was eliminated to form fragment ions of  $m/z$  641 (**Scheme 6**, pathway 1). The fragment ions of  $m/z$  641 fragment by eliminating the charged end unit A to generate fragment ions of  $m/z$  417, which dissociate to form ions of  $m/z$  223 via elimination of the charge-remote end unit C. Fragment ions of  $m/z$  641 can also eliminate the charge-remote end units B and C together to generate ions of  $m/z$  223. A second pathway for the generation of fragment ions of  $m/z$  641 involves the elimination of the charged-end unit A from the analyte ions of  $m/z$  865. The fragment ions of  $m/z$  641 can eliminate the charge-remote end unit D to form ions of  $m/z$  417. Ions of  $m/z$  641 can also eliminate the charged-end unit B to form fragment ions of  $m/z$  417, which can fragment by eliminating the charged end unit C to form fragment ions of  $m/z$  223. Alongside the losses of the monomeric or dimeric units, losses of small

molecules, such as  $\text{H}_2\text{O}$  and  $\text{CH}_2\text{O}$ , were also observed. The locations of the hydroxyl groups and the carbonyl groups in the unknown analyte ions of  $m/z$  865 were deduced based on the fact that the elimination of the charged end unit via the cyclopropane mechanism<sup>37</sup> requires that the  $\alpha$ -position of the  $\beta$ -O-4 linkage does not contain a keto functionality. The number of methoxy groups was determined based on the elemental compositions of the fragment ions. An ethyl ether functionality is proposed to be bound to the C unit based on the elemental compositions of the fragment ions. The ethyl ether functionality may be a residue formed from a lignin-carbohydrate complex. Based on the structure proposed for the tetrameric analyte ions of  $m/z$  865 (**Scheme 6**), the neutral analyte of MW 866 Da consists of three S-units and one H-unit, and thus is an S/H type lignin tetramer.

All the other structural assignments for lignin-related compounds were made in a similar manner as for the detailed examples discussed above.



**Scheme 6.** Pathways proposed for the production of the major fragment ions of the tetrameric analyte ions of  $m/z$  865.

*Structural elucidation of compounds not related to lignin.* Apart from lignin-related compounds, seven compounds not related to lignin were detected in the sample (

**Table 1** and **S2**). Six of these compounds have an RDBE value of less than that of a benzene ring (RDBE = 4). These six compounds are likely to be fatty acids based on their measured elemental compositions and fragmentation patterns (possible structures are shown in **Table 1** and **S2**). For example, deprotonated analyte ions of  $m/z$  473 have a higher C/O ratio (chemical formula:  $\text{C}_{32}\text{H}_{41}\text{O}_3$ , RDBE = 12.5, C/O ratio = 10.7, **Table S2**) than lignin-related compounds (C/O ratio in the range of 1.5-4.5). The identity of this analyte ion is still under investigation.

*An overview of the results.* A total of 62 deprotonated compounds generated adequate fragment ions upon CAD for structural elucidation. Among these compounds, 15 are lignin monomers, 27 are lignin dimers, 12 are lignin trimers, one is a lignin tetramer (ion of  $m/z$  857), and seven are not lignin-related compounds, most of them likely fatty acids. Lignin monomers and dimers (most with MW values of less than 550) account for the majority of the compounds detected. Oligomers larger than tetramers had very low abundances and did not generate enough fragment ions for structural elucidation.

Based on the proposed structures, compounds with various functionalities, such as hydroxy, aldehyde, keto, and carboxylic acid, were detected (

**Table 1** and **S2**). Lignin-related compounds with two functionalities in the side chains were also found in this organosolv lignin sample. Monomeric lignin-related analyte ions of  $m/z$  151, 167, 193, and one isomer in the isomeric group of ions of  $m/z$  209 correspond to G-lignin degradation products. Ions of  $m/z$  181 and one isomer in each of the isomeric groups of ions of  $m/z$  209, 223, 225, 227, 253, and 295 correspond to S-lignin degradation products. Two isomeric ions of  $m/z$  163 and ions of  $m/z$  221 correspond to H-lignin degradation products. Among the lignin dimers, 14 contain a 5-5 linkage, eight contain a  $\beta$ -O-4 linkage, four contain a  $\beta$ -5 linkage, and one likely contains a 4-O-5 linkage. The labile nature of the most common  $\beta$ -O-4 linkage in lignin likely contributed to the relatively low amount of compounds detected with this linkage compared to the 5-5 linkage.  $\beta$ -O-4, 5-5,  $\beta$ -5, and/or possibly 4-O-5 linkages were observed in the lignin trimers and tetramers. Despite the labile nature of the  $\beta$ -O-4 linkage, it is the most abundant linkage type in plants and thus some trimers and the detected tetramer contained only  $\beta$ -O-4 linkages. However, the abundances of these oligomers were low. No more than two of each of the 5-5,  $\beta$ -5, and 4-O-5 linkages (**Table S2**) were observed in the oligomers in this study.

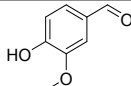
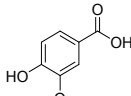
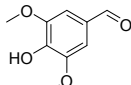
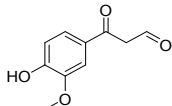
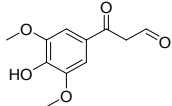
Two or more isomeric ions were successfully separated and characterized for each of the isomeric groups of ions of  $m/z$  163, 209, 343, 371, 385, 401, 473, and 609 (**Table S2**). Particularly, four different isomers of ions of  $m/z$  371 and three different isomers of ions of  $m/z$  385 (**Table S2**) were identified. Seven compounds were detected that were not lignin-related compounds. Most of them are likely to be fatty acids. By comparing the observed fragmentation patterns with those of authentic deprotonated model compounds (**Tables S1** and **S2**) or previously reported fragmentation data,<sup>36</sup> unknown analyte ions of  $m/z$  151, 167 ( $C_8 H_7 O_4$ ), 181, and 301 were identified as deprotonated vanillin, vanillic acid, syringaldehyde, and 5,5'-bisvanillin, respectively. The many unknown analytes, several corresponding to different isomeric structures, demonstrate that this organosolv lignin sample is very complex.

In order to perform a preliminary examination of the quantitative aspects of the above analytical method, an equimolar mixture containing 11 lignin model compounds, including several monomers, three dimers with different types of linkages, and a tetramer with only  $\beta$ -O-4 linkages, was studied (**Figures S1** and **S2**). The majority of the analytes (7 from 11) had a similar ionization efficiency (**Figures S1** and **S2**), suggesting that the relative abundances of their deprotonated forms reflect their relative molar amounts in the organosolv lignin sample. However, isoeugenol, a monomeric lignin model compound, was not detected, indicating that it was not efficiently ionized due to competition from the other analytes in the mixture. This may be due to the lack of electron-withdrawing substituents or internal hydrogen-bonding ability for isoeugenol. The  $\beta$ -O-4 (S-S) dimer also had a very low ionization efficiency for unknown reasons. On the other hand, the deprotonated 5-5 dimer, 2,2'-bisphenol, which bears two free phenolic hydroxyl groups, had an exceptionally high ionization efficiency, likely due to its ability to stabilize the negatively charged

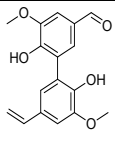
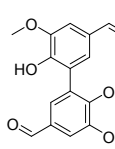
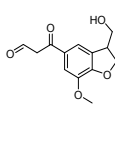
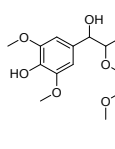
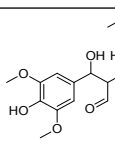
phenolic oxygen via internal hydrogen bonding. Similarly, the deprotonated  $\beta$ -O-4 (G-G-G-G) tetramer also had a high abundance in the MS spectrum (**Figure S1**), again likely due to the ability of forming an internal hydrogen bond to stabilize the negatively charged phenolic oxygen. This result is important as it indicates that the (-)ESI HRMS method used does not discriminate against the bigger analytes in the mixture.

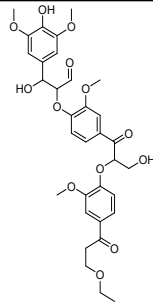
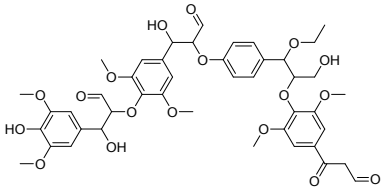
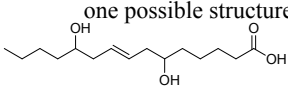
Last, a limitation of the method should be mentioned. The assignment of the likely structures for the compounds in the organosolv lignin sample based on their retention times and the elemental compositions and fragmentation patterns of their ions upon CAD is a very time-consuming process and requires a comprehensive understanding of the dissociation mechanisms of deprotonated lignin-related compounds, which may limit its application. It is noteworthy, however, that this method should be applicable as it is to other lignin degradation mixtures prepared by using other methods or prepared using a different type of biomass, as long as they contain similar compounds

**Table 1.** HPLC retention times (R.T.) of the compounds detected in the organosolv lignin sample,  $m/z$  values of the deprotonated analytes, their measured elemental compositions, their ionic and neutral fragments<sup>a</sup> formed upon CAD in MS<sup>2</sup> to MS<sup>4</sup> experiments and their relative abundances,<sup>b</sup> and the structures proposed for each analyte.

R.T. (min)	$m/z$	Elemental composition	RDBE	MS <sup>2</sup>	MS <sup>3</sup>	MS <sup>4</sup>	Proposed structure for the analyte
<b>Monomers</b>							
11.3	<b>151<sup>c</sup></b>	C <sub>8</sub> H <sub>7</sub> O <sub>3</sub>	5.5	151-CH <sub>3</sub> (136) 100%	136-CO (108) 30% 136-CO <sub>2</sub> (92) 100%		
6.6	<b>167<sup>c</sup></b>	C <sub>8</sub> H <sub>7</sub> O <sub>4</sub>	5.5	167-•CH <sub>3</sub> (152) 100% 167-CO <sub>2</sub> (123) 10%	152-CO (124) 100% 152-CO <sub>2</sub> (108) 20%		
9.8	<b>181<sup>c</sup></b>	C <sub>9</sub> H <sub>9</sub> O <sub>4</sub>	5.5	181-•CH <sub>3</sub> (166) 100%	166-•CH <sub>3</sub> (151) 100%	151-CO (123) 100%	
13.5	<b>193</b>	C <sub>10</sub> H <sub>9</sub> O <sub>4</sub>	6.5	193-•CH <sub>3</sub> (178) 100%	178-C <sub>2</sub> H <sub>2</sub> O (136) 100% 178-CO (150) 10% 178-C <sub>2</sub> H <sub>2</sub> O-CO (108) 5% 178-2CO (122) 3%	136-CO (108) 100%	
13.8	<b>223</b>	C <sub>11</sub> H <sub>11</sub> O <sub>5</sub>	6.5	223-•CH <sub>3</sub> (208) 100%	208-•CH <sub>3</sub> (193) 100% 208-•CH <sub>3</sub> -2CO (137) 70%	193-CO (165) 100% 193-2CO (137) 70% 193-CO-C <sub>2</sub> H <sub>2</sub> O (123) 20% 193-3CO (109) 20% 193-CO-CO <sub>2</sub> (121) 15% 193-CO <sub>2</sub> (149) 10%	



				193-2CO-C <sub>2</sub> H <sub>2</sub> O (95) 10%			
<b>Dimers</b>							
25.5	<b>299</b>	C <sub>17</sub> H <sub>15</sub> O <sub>5</sub>	10.5	299-•CH <sub>3</sub> (284) 100%	284-•CH <sub>3</sub> (269) 100%	269-CO (241) 100%	
					284-CH <sub>3</sub> -CO <sub>2</sub> (225) 10%	269-H <sub>2</sub> O (251) 15%	
						269-CO <sub>2</sub> (225) 10%	
						269-CO-CO <sub>2</sub> (197) 10%	
16.8	<b>301<sup>c</sup></b>	C <sub>16</sub> H <sub>13</sub> O <sub>6</sub>	10.5	301-•CH <sub>3</sub> (286) 100%	286-H <sub>2</sub> O (268) 100%	268-CO (240) 100%	
				301-•CH <sub>3</sub> -CO (258) 50%	286-•CHO (257) 90%	268-•CHO (239) 20%	
					286-H <sub>2</sub> O-CO (240) 60%	268-2CO (212) 10%	
					286-•OH (269) 50%		
					286-•OH-H <sub>2</sub> (267) 45%		
					286-H <sub>2</sub> O-•CHO (239) 35%		
					286-CO-•CHO (229) 20%		
					286-•CH <sub>3</sub> (271) 15%		
18.3	<b>401</b>	C <sub>21</sub> H <sub>21</sub> O <sub>8</sub>	11.5	401-H <sub>2</sub> O (383) 100%	383-CH <sub>3</sub> O• (352) 100%	352-•CH <sub>3</sub> (337) 100%	
				401-C <sub>9</sub> H <sub>10</sub> O <sub>3</sub> (235) 50%	383-•CH <sub>3</sub> (368) 85%		
				401-•CH <sub>3</sub> (386) 20%	383-CH <sub>2</sub> O (353) 60%		
				401-CH <sub>2</sub> O (371) 20%	383-CH <sub>2</sub> O-CO <sub>2</sub> (309) 20%		
				401-C <sub>2</sub> H <sub>4</sub> O <sub>3</sub> (325) 20%	383-•CH <sub>3</sub> -H <sub>2</sub> O (338) 10%		
				401-C <sub>9</sub> H <sub>10</sub> O <sub>3</sub> -•CH <sub>3</sub> -CO (192) 15%			
				401-C <sub>9</sub> H <sub>10</sub> O <sub>3</sub> -•CH <sub>3</sub> (220) 10%			
				401-C <sub>9</sub> H <sub>10</sub> O <sub>3</sub> -•CHO (206) 10%			
14.7	<b>405</b>	C <sub>20</sub> H <sub>21</sub> O <sub>9</sub>	10.5	405-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> (181) 100%	181-•CH <sub>3</sub> (166) 100%	166-•CH <sub>3</sub> (151) 100%	
				405-C <sub>9</sub> H <sub>10</sub> O <sub>4</sub> (223) 60%			
				405-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -•CH <sub>3</sub> (166) 30%			
				405-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -2 •CH <sub>3</sub> (151) 20%			
				405-CO (377) 10%			
				405-C <sub>9</sub> H <sub>10</sub> O <sub>4</sub> -•CH <sub>3</sub> (208) 10%			
<b>Trimers</b>							
21	<b>497</b>	C <sub>26</sub> H <sub>25</sub> O <sub>10</sub>	14.5	497-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> (273) 100%	273-•CH <sub>3</sub> (258) 100%	258-•CH <sub>3</sub> (243) 100%	
				497-C <sub>15</sub> H <sub>14</sub> O <sub>5</sub> (223) 65%			
				497-CH <sub>2</sub> O (467) 20%			
				497-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -•CH <sub>3</sub> (258) 20%			
				497-C <sub>15</sub> H <sub>14</sub> O <sub>5</sub> -•CH <sub>3</sub> (208) 10%			
19.1	<b>641</b>	C <sub>33</sub> H <sub>37</sub> O <sub>13</sub>	15.5	641-CH <sub>2</sub> O (611) 100%	611-CH <sub>2</sub> O (581) 100%	581-C <sub>11</sub> H <sub>14</sub> O <sub>4</sub> (371) 100%	
				641-C <sub>22</sub> H <sub>26</sub> O <sub>8</sub> (223) 80%	641-C <sub>22</sub> H <sub>26</sub> O <sub>8</sub> (223) 90%	581-CH <sub>3</sub> OH (549) 70%	

				641-H <sub>2</sub> O (623) 60%	611-H <sub>2</sub> O (593) 60%	581-C <sub>8</sub> H <sub>10</sub> O <sub>3</sub> (427) 40%	
				641-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> (417) 50%	611-CH <sub>3</sub> OH (579) 30%		
				641-H <sub>2</sub> O-CH <sub>2</sub> O (593) 30%	611-CH <sub>2</sub> O-CH <sub>3</sub> OH (549) 20%		
				641-CH <sub>2</sub> O-CH <sub>3</sub> OH (579) 30%	611-C <sub>12</sub> H <sub>16</sub> O <sub>5</sub> (371) 20%		
				641-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -CH <sub>2</sub> O (387) 15%	611-C <sub>14</sub> H <sub>16</sub> O <sub>5</sub> (347) 15%		
				641-C <sub>14</sub> H <sub>16</sub> O <sub>5</sub> (377) 15%	611-C <sub>10</sub> H <sub>14</sub> O <sub>4</sub> (401) 12%		
				647-C <sub>14</sub> H <sub>16</sub> O <sub>5</sub> -CH <sub>2</sub> O (347) 15%	611-C <sub>9</sub> H <sub>12</sub> O <sub>4</sub> (427) 10%		
				641-2CH <sub>2</sub> O (581) 10%			
<b>Tetramers</b>							
22.2	<b>865</b>	C <sub>44</sub> H <sub>49</sub> O <sub>18</sub>	20.5	865-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> (641) 100%	641-CH <sub>2</sub> O (611) 100%		
				865-CH <sub>2</sub> O (835) 50%	641-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> (417) 85%		
				865-H <sub>2</sub> O (847) 45%	641-C <sub>22</sub> H <sub>26</sub> O <sub>8</sub> (223) 80%		
				865-2H <sub>2</sub> O (829) 10%	641-H <sub>2</sub> O (623) 65%		
				865-H <sub>2</sub> O-CH <sub>2</sub> O (817) 10%			
				865-C <sub>2</sub> H <sub>6</sub> O <sub>2</sub> (803) 10%			
				865-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -CH <sub>2</sub> O (611) 10%			
				865-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -C <sub>11</sub> H <sub>14</sub> O <sub>5</sub> (415) 10%			
				865-C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> -C <sub>11</sub> H <sub>12</sub> O <sub>5</sub> (417) 5%			
<b>Fatty acid</b>							
26	<b>271</b>	C <sub>15</sub> H <sub>27</sub> O <sub>4</sub>	2.5	271-H <sub>2</sub> O-CO <sub>2</sub> (209) 100%		one possible structure: 	
				271-H <sub>2</sub> O (253) 40%			

<sup>a</sup>The elemental compositions of the fragment ions were measured and the elemental compositions of the neutral molecules lost were derived from the measured elemental compositions of the fragmenting ions and the fragment ions.

<sup>b</sup>The reproducibility of the relative abundances of the fragment ions was  $\pm 15\%$ . The fragment ions are shown in the order of their relative abundances. The information in the MS<sup>2</sup> to MS<sup>4</sup> columns is shown in the following order: precursor ion *m/z*, elemental composition of the neutral fragment lost upon CAD, *m/z* of the product ion (in parenthesis), and the relative abundance of the fragment ion.

<sup>c</sup>Verified by comparing with authentic model compounds (**Table 1** and **S1**) or previously published data on authentic model compounds.<sup>36</sup>

## Conclusions

An organosolv poplar lignin sample was analyzed by using (-)ESI HRMS and HPLC/(-)ESI HRMS<sup>n</sup> based on CAD. After HPLC separation, the analytes in the sample were ionized by (-)ESI. The elemental compositions as well as the degree of unsaturation (number of double bonds and rings or RDBE) were determined for each detected deprotonated analyte, which enabled the determination of their compound type. Many lignin-related compounds, as well as a few compounds not related to lignin, were detected, with lignin-related compounds being the most abundant. Lignin monomers, dimers, trimers, tetramers, and bigger oligomers were observed, with lignin monomers and dimers constituting the majority of the compounds in the mixture.

The CAD fragmentation behavior studied using HRMS<sup>n</sup> experiments on 16 deprotonated lignin-related model compounds provided insights into the fragmentation behavior of deprotonated unknown lignin-related compounds with various functionalities and linkage types. These results and previous studies<sup>37-39</sup> facilitated the structural elucidation of several unknown analyte ions. Structural information for 62 ionized unknown compounds derived from the organosolv lignin sample was obtained by using the HPLC/(-)ESI HRMS<sup>n</sup> experiments. Among these compounds, 15 were identified as lignin monomers, 27 as lignin dimers, 12 as lignin trimers, one as a tetramer, and seven were concluded not to be related to lignin. G-, H-, and S-lignin monomers were detected in the unknown analyte ions.  $\beta$ -O-4, 5-5,  $\beta$ -5, and possibly 4-O-5 linkages were identified to be present.

### Conflicts of interest

There are no conflicts to declare.

### Acknowledgements

The authors greatly appreciate the funding of this research by the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under award number DE-SC0000997.

### Supporting information

Materials, chemicals used, instrumental conditions, experimental procedures, HRMS<sup>n</sup> CAD spectra of ionized lignin model compounds, CAD HRMS<sup>n</sup> spectra of unknown deprotonated compounds in the organosolv lignin sample, synthesis procedures, NMR data for the synthesized compounds, and tables of Cartesian coordinates, electronic energies, zero-point vibrational energies, and 298 K thermal contributions for all computed structures.

### References

1. Yoshikawa, T.; Yagi, T.; Shinohara, S.; Fukunaga, T.; Nakasaka, Y.; Tago, T.; Masuda, T., Production of phenols from lignin via depolymerization and catalytic cracking. *Fuel Process. Technol.* **2013**, *108*, 69-75.
2. Alonso, D. M.; Wettstein, S. G.; Dumesic, J. A., Bimetallic catalysts for upgrading of biomass to fuels and chemicals. *Chem. Soc. Rev.* **2012**, *41* (24), 8075-8098.
3. Alonso, D. M.; Bond, J. Q.; Dumesic, J. A., Catalytic conversion of biomass to biofuels. *Green Chem.* **2010**, *12* (9), 1493-1513.
4. Mohan, D.; Pittman, C. U.; Steele, P. H., Pyrolysis of wood/biomass for bio-oil: a critical review. *Energy Fuels* **2006**, *20* (3), 848-889.
5. Huber, G. W.; Iborra, S.; Corma, A., Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem. Rev.* **2006**, *106* (9), 4044-4098.
6. Zakzeski, J.; Bruijninx, P. C.; Jongerius, A. L.; Weckhuysen, B. M., The catalytic valorization of lignin for the production of renewable chemicals. *Chem. Rev.* **2010**, *110* (6), 3552-3599.

7. Peralta-Yahya, P. P.; Zhang, F.; del Cardayre, S. B.; Keasling, J. D., Microbial engineering for the production of advanced biofuels. *Nature* **2012**, *488* (7411), 320-328.
8. Ghaffar, S. H.; Fan, M. Z., Structural analysis for lignin characteristics in biomass straw. *Biomass Bioenergy* **2013**, *57*, 264-279.
9. Freudenberg, K., Biosynthesis and constitution of lignin. *Nature* **1959**, *183* (4669), 1152-1155.
10. Heitner, C.; Dimmel, D.; Schmidt, J., *Lignin and Lignans: Advances in Chemistry*. CRC press: Boca Raton, FL, 2016.
11. Lewis, N. G.; Davin, L. B., The biochemical control of monolignol coupling and structure during lignan and lignin biosynthesis. In *Lignin and Lignan Biosynthesis*, American Chemical Society: Washington, DC, 1998; Vol. 697, pp 334-361.
12. Gang, D. R.; Costa, M. A.; Fujita, M.; Dinkova-Kostova, A. T.; Wang, H. B.; Burlat, V.; Martin, W.; Sarkanen, S.; Davin, L. B.; Lewis, N. G., Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis. *Chem. Biol.* **1999**, *6* (3), 143-51.
13. Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W., Lignin biosynthesis and structure. *Plant Physiol.* **2010**, *153* (3), 895-905.
14. Boerjan, W.; Ralph, J.; Baucher, M., Lignin biosynthesis. *Annu. Rev. Plant Biol.* **2003**, *54*, 519-546.
15. Weng, J. K.; Chapple, C., The origin and evolution of lignin biosynthesis. *New Phytol.* **2010**, *187* (2), 273-85.
16. Lewis, N. G.; Yamamoto, E., Lignin: occurrence, biogenesis and biodegradation. *Annu. Rev. Plant Biol.* **1990**, *41* (41), 455-496.
17. Staš, M.; Chudoba, J.; Kubička, D.; Pospíšil, M., Chemical characterization of pyrolysis bio-oil: application of orbitrap mass spectrometry. *Energy Fuels* **2015**, *29* (5), 3233-3240.
18. Das, L.; Kolar, P.; Sharma-Shivappa, R., Heterogeneous catalytic oxidation of lignin into value-added chemicals. *Biofuels* **2014**, *3* (2), 155-166.
19. Huijgen, W. J. J.; Telysheva, G.; Arshanitsa, A.; Gosselink, R. J. A.; de Wild, P. J., Characteristics of wheat straw lignins from ethanol-based organosolv treatment. *Ind. Crops Prod.* **2014**, *59*, 85-95.
20. Bozell, J. J.; Black, S. K.; Myers, M.; Cahill, D.; Miller, W. P.; Park, S., Solvent fractionation of renewable woody feedstocks: Organosolv generation of biorefinery process streams for the production of biobased chemicals. *Biomass Bioenergy* **2011**, *35* (10), 4197-4208.
21. Yáñez-S, M.; Matsuhira, B.; Nuñez, C.; Pan, S.; Hubbell, C. A.; Sannigrahi, P.; Ragauskas, A. J., Physicochemical characterization of ethanol organosolv lignin (EOL) from *Eucalyptus globulus*: effect of extraction conditions on the molecular structure. *Polym. Degrad. Stabil.* **2014**, *110*, 184-194.
22. Watkins, D.; Nuruddin, M.; Hosur, M.; Tcherbi-Narteh, A.; Jeelani, S., Extraction and characterization of lignin from different biomass resources. *J. Mater. Res. Technol.* **2015**, *4* (1), 26-32.
23. Jarrell, T. M.; Marcum, C. L.; Sheng, H.; Owen, B. C.; O'Lenick, C. J.; Maraun, H.; Bozell, J. J.; Kenttämä, H. I., Characterization of organosolv switchgrass lignin by using high performance liquid chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-ion mode electrospray ionization. *Green Chem.* **2014**, *16* (5), 2713-2727.

24. Ouyang, X.; Huang, X.; Hendriks, B. M. S.; Boot, M. D.; Hensen, E. J. M., Coupling organosolv fractionation and reductive depolymerization of woody biomass in a two-step catalytic process. *Green Chem.* **2018**, *20* (10), 2308-2319.
25. Li, J.; Henriksson, G.; Gellerstedt, G., Lignin depolymerization/repolymerization and its critical role for delignification of aspen wood by steam explosion. *Biores. Technol.* **2007**, *98* (16), 3061-3068.
26. Sun, Q.; Khunsupat, R.; Akato, K.; Tao, J.; Labbé, N.; Gallego, N. C.; Bozell, J. J.; Rials, T. G.; Tuskan, G. A.; Tschaplinski, T. J.; Naskar, A. K.; Pu, Y.; Ragauskas, A. J., A study of poplar organosolv lignin after melt rheology treatment as carbon fiber precursors. *Green Chem.* **2016**, *18* (18), 5015-5024.
27. Owen, B. C.; Hauptert, L. J.; Jarrell, T. M.; Marcum, C. L.; Parsell, T. H.; Abu-Omar, M. M.; Bozell, J. J.; Black, S. K.; Kenttämä, H. I., High-performance liquid chromatography/high-resolution multiple stage tandem mass spectrometry using negative-ion-mode hydroxide-doped electrospray ionization for the characterization of lignin degradation products. *Anal. Chem.* **2012**, *84* (14), 6000-6007.
28. Baucher, M.; Halpin, C.; Petit-Conil, M.; Boerjan, W., Lignin: genetic engineering and impact on pulping. *Crit. Rev. Biochem. Mol. Biol.* **2003**, *38* (4), 305-350.
29. Li, X.; Weng, J. K.; Chapple, C., Improvement of biomass through lignin modification. *Plant J.* **2008**, *54* (4), 569-581.
30. Hallac, B. B.; Pu, Y.; Ragauskas, A. J., Chemical transformations of *Buddleja davidii* lignin during ethanol organosolv pretreatment. *Energy Fuels* **2010**, *24* (4), 2723-2732.
31. El Hage, R.; Brosse, N.; Sannigrahi, P.; Ragauskas, A., Effects of process severity on the chemical structure of *Miscanthus* ethanol organosolv lignin. *Polym. Degrad. Stabil.* **2010**, *95* (6), 997-1003.
32. Nagy, M.; David, K.; Britovsek, G. J. P.; Ragauskas, A. J., Catalytic hydrogenolysis of ethanol organosolv lignin. *Holzforschung* **2009**, *63* (5), 513-520.
33. Xu, F.; Sun, J.-X.; Sun, R.; Fowler, P.; Baird, M. S., Comparative study of organosolv lignins from wheat straw. *Ind. Crops Prod.* **2006**, *23* (2), 180-193.
34. Ralph, J.; Hatfield, R. D., Pyrolysis-GC-MS characterization of rorage materials. *J. Agric. Food Chem.* **1991**, *39*, 1426-1437.
35. van Erven, G.; de Visser, R.; Merckx, D. W. H.; Strolenberg, W.; de Gijssel, P.; Gruppen, H.; Kabel, M. A., Quantification of lignin and its structural features in plant biomass using <sup>13</sup>C lignin as internal standard for pyrolysis-GC-SIM-MS. *Anal. Chem.* **2017**, *89* (20), 10907-10916.
36. Marcum, C. L.; Jarrell, T. M.; Zhu, H.; Owen, B. C.; Hauptert, L. J.; Easton, M.; Hosseinaei, O.; Bozell, J.; Nash, J. J.; Kenttämä, H. I., A fundamental tandem mass spectrometry study of the collision-activated dissociation of small deprotonated molecules related to lignin. *ChemSusChem* **2016**, *9* (24), 3513-3526.
37. Zhang, J.; Feng, E.; Li, W.; Sheng, H.; Milton, J. R.; Easterling, L. F.; Nash, J. J.; Kenttämä, H. I., Studies on the fragmentation mechanisms of deprotonated lignin model compounds in tandem mass spectrometry. *Anal. Chem.* **2020**, *92* (17), 11895-11903.
38. 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., Mass spectrometry-based fragmentation as an identification tool in lignomics. *Anal. Chem.* **2010**, *82* (19), 8095-8105.
39. Sheng, H.; Tang, W.; Gao, J.; Riedeman, J. S.; Li, G.; Jarrell, T. M.; Hurt, M. R.; Yang, L.; Murria, P.; Ma, X.; Nash, J. J.; Kenttämä, H. I., (-)ESI/MS<sup>n</sup> procedure for sequencing lignin

oligomers based on a study of synthetic model compounds with  $\beta$ -O-4 and 5-5 linkages. *Anal. Chem.* **2017**, *89* (24), 13089-13096.

40. Bozell, J. J.; O'Lenick, C. J.; Warwick, S., Biomass fractionation for the biorefinery: heteronuclear multiple quantum coherence-nuclear magnetic resonance investigation of lignin isolated from solvent fractionation of switchgrass. *J. Agric. Food Chem.* **2011**, *59* (17), 9232-9242.
41. Lapiere, C.; Pollet, B.; Petit-Conil, M.; Toval, G.; Romero, J.; Pilate, G.; Leplé, J.-C.; Boerjan, W.; Ferret, V.; De Nadai, V., Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiol.* **1999**, *119* (1), 153-164.
42. Van Doorselaere, J.; Baucher, M.; Chognot, E.; Chabbert, B.; Tollier, M. T.; Petit-Conil, M.; Leplé, J. C.; Pilate, G.; Cornu, D.; Monties, B., A novel lignin in poplar trees with a reduced caffeic acid/5-hydroxyferulic acid O-methyltransferase activity. *Plant J.* **1995**, *8* (6), 855-864.
43. Chen, F.; Tobimatsu, Y.; Havkin-Frenkel, D.; Dixon, R. A.; Ralph, J., A polymer of caffeyl alcohol in plant seeds. *Proc. Natl. Acad. Sci. U S A* **2012**, *109* (5), 1772-1777.
44. Tian, Z.; Wang, X. B.; Wang, L. S.; Kass, S. R., Are carboxyl groups the most acidic sites in amino acids? Gas-phase acidities, photoelectron spectra, and computations on tyrosine, *p*-hydroxybenzoic acid, and their conjugate bases. *J. Am. Chem. Soc.* **2008**, *131* (3), 1174-1181.
45. Steill, J. D.; Oomens, J., Gas-phase deprotonation of *p*-hydroxybenzoic acid investigated by IR spectroscopy: solution-phase structure is retained upon ESI. *J. Am. Chem. Soc.* **2009**, *131* (38), 13570-13571.
46. Schroder, D.; Budesinsky, M.; Roithova, J., Deprotonation of *p*-hydroxybenzoic acid: does electrospray ionization sample solution or gas-phase structures? *J. Am. Chem. Soc.* **2012**, *134* (38), 15897-905.
47. Marcum, C. L.; Riedeman, J. S.; Jarrell, T. M.; Nash, J. J.; Kenttämaa, H. I., Losses of CO and CO<sub>2</sub> upon collision-activated dissociation of substituted 2-methoxyphenoxides after methyl radical loss. *Int. J. Mass spectrom.* **2020**, *456*, 116397-116402.
48. Upton, B. M.; Kasko, A. M., Strategies for the conversion of lignin to high-value polymeric materials: review and perspective. *Chem. Rev.* **2016**, *116* (4), 2275-2306.
49. Demarque, D. P.; Crotti, A. E.; Vesseccchi, R.; Lopes, J. L.; Lopes, N. P., Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat. Prod. Rep.* **2016**, *33* (3), 432-55.
50. Li, S.; Lundquist, K., Acid reactions of lignin models of  $\beta$ -5 type. *Holzforschung* **1999**, *53*, 39-42.
51. Morreel, K.; Dima, O.; Kim, H.; Lu, F.; Niculaes, C.; Vanholme, R.; Dauwe, R.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., Mass spectrometry-based sequencing of lignin oligomers. *Plant Physiol.* **2010**, *153* (4), 1464-1478.