Mass spectrometry-based methods for the advanced characterization and structural analysis of lignin: A review

Dane R. Letourneau  |  Dietrich A. Volmer

Department of Chemistry, Humboldt University Berlin, Berlin, Germany

Correspondence
Dietrich Volmer, Department of Chemistry, Humboldt University Berlin, D-12489 Berlin, Germany. Email: Dietrich.Volmer@hu-berlin.de

Abstract
Lignin is currently one of the most promising biologically derived resources, due to its abundance and application in biofuels, materials and conversion to value aromatic chemicals. The need to better characterize and understand this complex biopolymer has led to the development of many different analytical approaches, several of which involve mass spectrometry and subsequent data analysis. This review surveys the most important analytical methods for lignin involving mass spectrometry, first looking at methods involving gas chromatography, liquid chromatography and then continuing with more contemporary methods such as matrix assisted laser desorption ionization and time-of-flight-secondary ion mass spectrometry. Following that will be techniques that directly ionize lignin mixtures—without chromatographic separation—using softer atmospheric ionization techniques that leave the lignin oligomers intact. Finally, ultra-high resolution mass analyzers such as FT-ICR have enabled lignin analysis without major sample preparation and chromatography steps. Concurrent with an increase in the resolution of mass spectrometers, there have been a wealth of complementary data analyses and visualization methods that have allowed researchers to probe deeper into the “lignome” than ever before. These approaches extract trends such as compound series and even important analytical information about lignin substructures without performing lignin degradation either chemically or during MS analysis. These innovative methods are paving the way for a more comprehensive understanding of this complex biopolymer.

Abbreviations: 2D-HSQC-NMR, two-dimensional heteronuclear single quantum coherence nuclear magnetic resonance spectroscopy; APCI, atmospheric pressure chemical ionization; API, atmospheric pressure ionization; APPI, atmospheric pressure photoionization; CID, collision induced dissociation; DART, direct analysis in real time; DBE, double bond equivalents; DESI, desorption electrospray ionization; DFRC, derivatization followed by reductive cleavage; EFR, electron paramagnetic resonance spectroscopy; ESI, electrospray Ionization; FT-ICR, Fourier transform ion cyclotron resonance mass analyzer; GC-MS, Gas chromatography-mass spectrometry; GHG, Greenhouse gas; HPLC, high performance liquid chromatography; HRMS, high resolution mass spectrometry; HSQC, heteronuclear single quantum coherence; ICR, Fourier transform ion cyclotron resonance; LC-MS, liquid chromatography-mass spectrometry; LDI, laser desorption/ionization; LO, lignin oligomer; LRMS, low resolution mass spectrometry; MALDI, matrix assisted laser desorption/ionization; MRM, multiple reaction monitoring; MSI, mass spectrometry imaging; MS/MS, tandem mass spectrometry; MWL, milled wood lignin; NALDI, nano-assisted laser desorption/ionization; NMR, nuclear magnetic resonance spectroscopy; PDA, principal component analysis; PML, process modified lignin; Py-GC-MS, pyrolysis-gas chromatography-mass spectrometry; Py-MBMS, pyrolysis-molecular beam mass spectrometry; QQQ, triple quadrupole mass analyzer; RDB, ring double bond equivalent; SEC, size exclusion chromatography; SIFT, selected ion flow tube; SIM, selected ion monitoring; SIMS, secondary ion mass spectrometry; SNMS, secondary neutral mass spectrometry; SPE, solid phase extraction; S, G, and H, syringyl, guaiacyl and p-hydroxylphene lignin subunits; TGA, thermogravimetric analysis; TMAH, tetramethylammonium hydroxide; TOF, time of flight; UHPLC, ultra high performance liquid chromatography; VRL, virgin released lignin; VUVPI, vacuum ultraviolet photoionization.
understanding of this important biopolymer, as we seek more sustainable solutions for our human species’ energy and materials needs.

KEYWORDS
analytical chemistry, biofuels, HRMS, lignin, mass spectrometry

1 | INTRODUCTION

The urgent need for energy alternatives is driving governments, industries, and researchers to seek new and innovative ideas to sustainably meet humanity’s energy needs while drastically reducing our collective CO₂ emissions (Sun et al., 2018). Biofuels and biomaterials are an important part of this solution, as they offer renewable and carbon-neutral energy alternatives to fossil fuels (Ragauskas et al., 2006). Indeed, “if biomass is produced at a sustainable rate, carbon dioxide generated from the combustion of biomass fuels balances the carbon dioxide removed from the atmospheric pool by photosynthesis in the course of biomass production. Thus, the bioenergy cycle does not add carbon dioxide to the atmosphere and will not contribute to global warming” (Agblevor et al., 1994). It has been estimated that biomass could provide around 25% of global energy requirements (Briens et al., 2008), with nuclear, solar, hydrogen and wind energies potentially contributing the rest (Ragauskas et al., 2006). In terms of sustainability, a 2020 study of 27 EU member countries showed that in the period from 1990 to 2017, CO₂ emissions declined with an increase in wood biomass energy consumption (Sulaiman et al., 2020). Wood currently contributes more than 50% of total renewable energy consumed in the EU and this study recommends that all member countries increase the share of wood biomass energy in their energy mix to reduce CO₂ emissions (Sulaiman et al., 2020). Figure 1A demonstrates the sustainable lifecycle of biofuels and biomaterials.

Biomass is primarily composed of cellulose and hemicellulose, bound together in what is referred to as the “lignocellulosic biomass” (Figure 1B). For biofuels, these are mostly derived from agricultural waste, forest products, aquatic biomass, and herbaceous energy crops (Ragauskas et al., 2006), after they undergo various gasification, liquefaction, or pyrolysis processes to become usable biofuels.

1.1 | Lignin

Lignin is the second most abundant biopolymer after cellulose and forms 15%–30% of lignocellulosic biomass by weight and 40% by energy (Perlack et al., 2005). Lignin comprises the single largest source of aromatic compounds on Earth (Sun et al., 2018) and has therefore been of great interest because of the need for aromatics in both the fuel and chemical industries. Indeed, “lignin

FIGURE 1 (A) The fully integrated agro-biofuel-biomaterial-biopower cycle for sustainable technologies (reprinted with permission, Ragauskas et al., 2006). (B) Schematic representation of the location and structure of lignin in lignocellulosic material (reprinted with permission, Zakzeski et al., 2010) [Color figure can be viewed at wileyonlinelibrary.com]
valorization has the potential to completely replace fossil-based aromatic polymers” (Prothmann et al., 2020). However, there are significant challenges in utilizing this polymer, including finding environmentally friendly and cost-effective solutions to depolymerize and liberate it from the lignocellulosic matrix. In the past, a lack of understanding led to much of it going to waste or being used without modification in inefficient combustion processes (Gosselink et al., 2004). Currently, the primary challenge is finding the right processing conditions/catalyst to deliver high product yields in an energy- and material-efficient manner (Sun et al., 2018). The complex, nonrepeating structure of lignin makes this a significant challenge.

Lignin is made up of various linkages of three basic monolignol subunits, syringyl alcohol, coniferyl alcohol, and p-coumaryl alcohol, which become the characteristic syringyl (S), guaiacyl (G), and p-hydroxyphenyl (H) groups when linked together to form the lignin structure (Figure 2A). The proportion of these monomers in any given lignin sample is often characteristic of its source, making measurement of H/G/S ratios an analytical priority. Hardwood lignin (e.g., eucalyptus, poplar, birch) is primarily made of G and S with small quantities of H subunits. Softwood lignin (e.g., spruce, pine) is composed primarily of G units with small quantities of H (Liao et al., 2020). Other less abundant monomers have been identified, including acylated monolignols such as acetates, p-coumarates and p-hydroxybenzoates, and compounds with a 5-hydroxyguaiacyl structure such as 3-methoxycatechol, 5-vinyl-3-methoxycatechol, and 5-propenyl-3-methoxycatechol (del Río & Gutiérrez, et al., 2007). Linkages between monomers can be represented either by carbon number or Greek letters (e.g.,

![Figure 2](image-url)

**Figure 2** (A) Generic lignin units and (B) main interunit linkages (reprinted with permission, Richel et al., 2012 under the CC BY-NC 3.0 license)
Given that notations can vary in the literature, both will be used in this review. Common linkages between lignin monomers include β-O-4 (alkyl-aryl ethers, representing 50% of lignin linkages), 4-O-5 (diaryl ether), β-β (resinol), β-1 (1,2-diaryl-propane), β-5 (phenylcoumaran), and 5-5 bonds between diaryl inter-lignin units. However, there are over 12 different interunit C–C and C–O–C linkages possible (Wood & Kellogg, 1988). Examples of these linkages are shown in Figure 2B.

A lignin structure is formed when these monomer units are joined together at “random” (i.e., based only on concentration of subunits and environmental conditions) through oxidative, phenolic coupling reactions via radicals generated by peroxidase-H2O2 (Ralph et al., 2004). In the lignification process, the lignin macromolecule grows as the various monomers couple to the growing (phenolic) end of lignin polymer. Ultimately then, lignin has no one defined structure and the structural definition of lignin continues to change as our analytical technologies improve and our understanding of lignin’s biosynthetic pathways increases. Recently, Ralph et al. (2019) have expanded and refined the structure of lignin, including proposing a broader definition for lignin monomers. There are now a large number of phenolics that radically couple and cross-couple combinatorically to build the racemic lignin polymer and therefore could be considered lignin “monomers.” Ralph’s group produced an in-depth figure to help visualize this “updated” understanding of the lignin polymer by drawing representative 20-mer examples for three types of wood (Figure 3).

How exactly lignin exists in its native form as part of the lignocellulosic biomass is still unknown. For some time, it has been assumed that lignin exists as a cross-linked 3D network of subunits (Liao et al., 2020). However, milled wood lignin (MWL) has been shown to be a linear oligomer (Crestini et al., 2011) and Ralph et al. have argued that lignins may be less branched and more linear than commonly thought (Yue et al., 2016), so the degree to which lignin is truly crosslinked in its native form is still up for debate.

**Figure 3** Lignin model structures. Model lignin 20-mers are shown for: (A) a gymnosperm/softwood, (B) an angiosperm/dicot/hardwood, and (C) A acommelinid) monocot. A structure is colored by the type of interunit linkage, in which it is involved; where one unit is involved in two types of units, these are shown as split structures (reprinted with permission, Ralph et al., 2019) [Color figure can be viewed at wileyonlinelibrary.com]
Finally, it is worth considering some definitions to clear up confusion when approaching the literature on this topic. The field of “lignomics” has been defined as comprising the profiling of all phenolics, such as phenylpropanoids, lignans, and lignin oligomers (LOs), for which the biosynthesis and/or regulation is connected with lignin biosynthesis (Morreel & Kim, et al., 2010). It should be stated early on that in lignomics—and therefore in this review—lignin is often referred to in the singular. However, although pertaining to several common structural subunits and linkages, lignin “differs among plant species, among plant parts and even within plant cell walls” (Wood & Kellogg, 1988). Perhaps a better term for lignin and lignin-like/lignin-derived compounds is the “lignome” (Morreel & Dima, et al., 2010; Morreel & Kim, et al., 2010; Qi et al., 2016b)—“representing the ensemble of all the biosynthetic phenolics, metabolites and (neo) lignan biosynthetic pathways and their derivatives, as well as the lignin oligomers” (Albishi et al., 2019). This term has come into more popularity recently and certainly does a better job at representing the diversity of structures and products that are often simply referred to as “lignin.” This review will use both terms, but “lignome” is certainly preferred.

1.2 Analytical methods without mass spectrometry

The inherent complexity of the lignin polymer has prompted the development of numerous analytical methods to meet the challenge of understanding this material to the fullest degree. This review will not attempt to summarize all these techniques, rather it will highlight a few standout methodologies in this short section and focus almost entirely on MS methods.

Nuclear magnetic resonance (NMR) spectroscopy has given significant insight into the structure and composition of lignin. Various forms of NMR have been employed, including $^1$H, $^{13}$C, and $^{31}$P NMR and 2D HSQC NMR among others (Constant et al., 2016; Liao et al., 2020).

Fourier transform infrared (FTIR) spectroscopy is an inexpensive, sensitive technique that can provide valuable information on the structure and properties of natural polymers (Kačuráková & Wilson, 2001). It can be used to estimate lignin S/G ratios by measuring the intensities of the bands around 1327 cm$^{-1}$ (S units) and 1271 cm$^{-1}$ (G units) (del Río & Gutiérrez, et al., 2007). FTIR has also been used to observe changes in lignin functional groups during processes such as pulping (Sa’don et al., 2017).

Gel permeation chromatography (GPC) can give useful information on the average molecular weight distribution information for lignins and other polymers (Himmel et al., 1990). In GPC, there are no chemical interactions with a stationary or mobile phase; rather, analytes are separated based on their hydrodynamic volume (Beri et al., 2000). However, calibration is usually based on a polydisperse polymer such as polystyrene or poly(methyl methacrylate) and there are no specific standards yet available for lignin, making molar mass calculations questionable (Constant et al., 2016).

Although all of these methods have played an important part in our understanding of lignin, they are unable to provide exact information on the substrates and functional groups present in lignin and often miss less abundant components (Qi et al., 2020). MS has been the most important analytical technology in advancing our understanding of lignin, although significant challenges remain—for example, “the lack of standard lignin samples of reproducible quality that would be suitable to benchmark novel methodologies” (Sun et al., 2018). Issues such as this (along with potential solutions) will be explored in the course of this review.

1.3 Analytical methods with mass spectrometry

MS has played an integral part in the development of methodologies for the analysis and understanding of complex, natural organic matter (NOM)-type molecules such as lignin. Its utility ranges from simple detection of lignin mono-, di-, and trimers at the backend of a chromatographic separation, such as in many gas chromatography-mass spectrometry (GC-MS)-based methods (Guillén & Ibarogitia, 1999), to the advanced structural elucidation of large LOs available from the high-resolution data generated by Fourier transform ion cyclotron resonance (FT-ICR) or orbitrap instruments (Qi & Volmer, 2019a). As technologies and mass resolution have improved, these methods have become more instrumental in focus and, most recently, increasingly data-focused as new techniques become available to identify patterns and useful information from full scan high resolution data with no prior chromatography, sample preparation, or targeted analyses required.

While this review is not exhaustive in mentioning every individual combination of techniques that may be considered a unique “method,” it will attempt to summarize the most important developments and how they have advanced our understanding of the complex mixture known as the lignome. An organizational overview
of the general MS techniques surveyed here is presented in Table 1.

2 | GAS CHROMATOGRAPHY-MS METHODS

2.1 | Introduction to GC-MS methods for lignin

Low-resolution GC-MS has been used to identify lignin components from a number of processes. However, early studies were unable to conclusively identify many of these components because of the lack of comparable spectra in the literature. An example is an early GC-MS analysis of lignin monomers, dimers and trimers in the brown layer left on the wall of a receptacle by a liquid smoke flavoring (Guillén & Ibargoitia, 1999). In that study, GC was undertaken with a fused silica capillary column coated with a nonpolar stationary phase and He as carrier gas. Analytes were subjected to EI at 70 eV. The researchers positively identified several lignin monomers (phenol, guaiacol, and syringol derivatives and alkyl-aryl ethers) by comparing their results to a database. However, they were unable to conclusively identify any dimers or trimers because there were no spectra available for comparison in the database. They were, however, able to define certain characteristics for these dimer and trimer groups, including the presence of typical mass fragments of phenol, guaiacol, and syringol derivatives and the presence of biphenyl, methylene-bis-phenyl, 9,10-dihydrophenanthrene, and phenanthrene structures.

Various lignin degradation and derivatization methods have been utilized to break down and/or derivatize the complex lignin structure to give components more amenable to GC-MS analysis. For example, cupric oxide (CuO) degradations have been used for this purpose for nearly 40 years (Hedges & Ertel, 1982). In one example, GC-MS preceded by a cupric oxide (CuO) degradation was used to analyze alkali-lignin from a paper mill effluent decolorized with two Streptomyces strains (Hemández, 2001). The mild conditions of the CuO degradation allowed for cleavage of β-aryl-ether bonds without altering the propyl sidechain of the lignin moiety. After degradation and derivatization with trimethylsilyl (TMS) ether, analysis was conducted via GC-MS on an ion trap MS. Ethylvanilline was used as internal standard and quantifications were performed by response factors for a variety of standard lignin monomer compounds. This method allowed the identification of products derived from H, G, and S units as well as cinnamic, p-coumaric and ferulic acids (see Table 2). The resulting molar H/G/S relationship permitted the determination of the level of oxidation of the lignin after it had undergone bacterial decolorization and demonstrated the utility of Streptomyces as a bioremediation agent for oxidizing recalcitrant materials such as lignin.

Thioacidolysis is another significant lignin degradation/derivatization method of note. Using a BF3 etherate reagent for degradation and TMS for derivatization, researchers demonstrated they could effectively split arylglycerol-ether (β-O-4) bonds present in lignin, without the associated reduction in monomer yield from condensation reactions which frequently occur in a conventional acidolysis degradation (Lapierre et al., 1985). In addition, the resulting GC-MS chromatograms of the TMS derivatives are much simpler, with improved reproducibility. Since its development, the method has been used to look at the relative distribution of carbon and diaryl ether condensed bonds in spruce lignins (Lapierre et al., 1991), the lignin composition in normal and compression wood (with both ethanethiol and methanethiol as thioacidolysis reagents) (Önnerud, 2003), the differences in lignin structure between primary and secondary cell walls of hybrid aspen cell cultures (Christiernin et al., 2005) and for the rapid analysis of a new lignin monomer (tricin, an O-methylated flavone) recently identified in many grass species (Chen Fang, et al., 2021).

2.2 | Modifications and improvements

Desiring to improve upon thioacidolysis-based methods to achieve a simpler, more selective degradation of β-O-4 linkages, Lu and Ralph (1997) developed a derivatization method followed by reductive cleavage, or “DFRC.” Not only did this method allow them to avoid the laborious wet chemistry/harsh solvents and optimization required in previous degradation methods, it also demonstrated higher monomer yields (92%–97%) as compared to acidolysis and thioacidolysis techniques with comparable reproducibility (Lu & Ralph, 1997). Since their development, DFRC-GC-MS techniques have been used for the reproducible measurement of isolated lignins in softwood, hardwood, grass, and dicots (Lu & Ralph, 1998), the characterization of dietary fiber lignins from fruits and vegetables (Bunzel et al., 2005) and the quantitation of monolignols and low-abundance monolignol conjugates (Regner et al., 2018), although some have questioned the completeness of DFRC aryl ether cleavage and monomer yield (Holtman et al., 2003).

In more complex or “dirty” matrices, additional techniques are required to distinguish lignin components from coeluting matrix components. For example, further
<table>
<thead>
<tr>
<th>Method</th>
<th>Sample preparation</th>
<th>Ionization method</th>
<th>Mass analyzer(s)</th>
<th>Selected applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC-MS</td>
<td>Derivatization (e.g., TMS), degradation (e.g., thioacidolysis)</td>
<td>EI</td>
<td>Ion trap</td>
<td>Low MW lignin fragments—mono-, di-, trimers (Guillén &amp; Ibargoitia, 1999)</td>
</tr>
<tr>
<td>GC-MS/MS</td>
<td>Derivatization (e.g., TMS)</td>
<td>EI</td>
<td>Ion trap</td>
<td>Improved understanding of lignin pyrolysis products (Larson et al., 2018)</td>
</tr>
<tr>
<td>Py-GC-MS</td>
<td>Minimal</td>
<td>EI</td>
<td>Ion trap</td>
<td>H/G/S ratio of lignin (Martin, Saiz-Jimenez &amp; Gonzalez-Vila, 1979)</td>
</tr>
<tr>
<td>SPE-GC-MS</td>
<td>SPE extraction</td>
<td>EI</td>
<td>Single quad</td>
<td>Low MW lignin fragments in more complex matrices (Guggenberger et al., 2019)</td>
</tr>
<tr>
<td>SIFT-MS</td>
<td>None</td>
<td>CI</td>
<td>Single quad</td>
<td>Small molecule byproducts of lignin degradation processes (Gibson et al., 2015)</td>
</tr>
<tr>
<td>LC-MS and LC-MS/MS</td>
<td>Minimal</td>
<td>API</td>
<td>Triple quad</td>
<td>Analysis of larger, intact lignin subunits and structures (Dier &amp; Rauber, et al., 2017)</td>
</tr>
<tr>
<td>2D LC-MS</td>
<td>Minimal</td>
<td>API</td>
<td>Triple quad</td>
<td>“Fingerprinting” of lignin samples from various processes (Le Masle et al., 2014)</td>
</tr>
<tr>
<td>MALDI-MS</td>
<td>MALDI plate preparation</td>
<td>MALDI</td>
<td>TOF</td>
<td>Intact lignin oligomers up to 3 kDa (Kosyakov et al., 2018)</td>
</tr>
<tr>
<td>TOF-SIMS</td>
<td>None</td>
<td>SIMS</td>
<td>TOF</td>
<td>Detailed information about the surface of a solid sample (Saito, Kato &amp; Takamori, et al., 2005)</td>
</tr>
<tr>
<td>HRMS</td>
<td>Minimal/none</td>
<td>API</td>
<td>FT-ICR, orbitrap</td>
<td>Data mining for molecular formulas, series identification (Qi et al., 2020; Qi &amp; Volmer, 2019a); Computer modeling of lignin (Terrell et al., 2020)</td>
</tr>
</tbody>
</table>
improvements in the sensitivity and reproducibility of DFRC-GC-MS technique in complex, low-lignin samples were demonstrated using a modified sample cleanup and a stable-isotope dilution approach (Schäfer et al., 2015). The researchers synthesized deuterated DFRC products as internal standards and used a polymeric reversed phase solid phase extraction (SPE) for sample cleanup, as well as selective ion monitoring (SIM) mode in MS, to achieve quantitative lignin compositions in fibers from pear, radish, carrot, and asparagus samples.

Optimizations to the CuO oxidation-GC-MS method (vide supra), including SPE cleanup with a polymeric sorbent instead of traditional liquid-liquid extraction and optimized sample conditions for analyzing samples of dissolved organic matter (DOM), soil and sediment have been proposed (Kaiser & Benner, 2012). Following SPE cleanup, samples were derivatized with BSTFA/TMCS before GC-MS analysis. The MS was operated in full-scan and SIM modes, with SIM used for quantification; trans-cinnamic acid (CiAD) and ethyl-vanillin (EVAL) were used as internal standards. Reproducible results free of coeluting species for monomeric lignin phenols in complex matrices were achieved, including peat soil and DOM (Figure 4).

Recently, a method for quantifying volatiles from technical lignins by SPE-GC-MS was established by performing headspace sampling to achieve quantitative results (Guggenberger et al., 2019). First, 10 ml headspace vials containing the sample were equilibrated at 40°C in an agitator, into which a SPME fiber was placed for 30 min.

**TABLE 2** Concentration (mmol/100 mg sample) of the different lignin units, molar H/G/S relationship and cinnamic acids content determined by CuO degradation of the alkali-lignin (AL) obtained from untreated effluent (control) and effluent decolorized by *Streptomyces avermitilis* and *Sarcoptes scabies* (reprinted with permission, Hemández, 2001)

<table>
<thead>
<tr>
<th></th>
<th>Concentration (units)</th>
<th>Molar ratio, H/G/S</th>
<th>Cinnamic acid content (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>G</td>
<td>S</td>
</tr>
<tr>
<td>AL-control</td>
<td>0.96</td>
<td>3.31</td>
<td>4.2</td>
</tr>
<tr>
<td>AL-<em>S. avermitilis</em></td>
<td>2.03</td>
<td>3.32</td>
<td>3.99</td>
</tr>
<tr>
<td>AL-<em>S. scabies</em></td>
<td>1.61</td>
<td>2.7</td>
<td>3.83</td>
</tr>
</tbody>
</table>

Abbreviation: AL, alkali-lignin.

**FIGURE 4** Selected chromatograms of (A) a 40 pmol/μl standard mixture, (B) peat soil (wt %OC = 42.8), and (C) DOM extract from the Arctic Ocean (reprinted with permission, Kaiser & Benner, 2012). DOM, dissolved organic matter
Directly after extraction, the fiber was desorbed in the split/splitless injector and GC-MS was performed in both SIM and full scan modes. SPME extraction avoided interferences such as matrix components (salts, water, polymers) and eliminated most sample preparation. The method was able to quantify guaiacol and DMDS, two volatiles from technical lignins, in a range from ng to several μg. Depending on the analyte amount, recovery ranged from 89% to 123% for DMDS and 90%–105% for guaiacol.

Another method of resolving coeluting species and matrix interferences when analyzing complex samples is to utilize tandem MS to increase specificity. As an example, an analysis of lignin-derived phenols in several complex matrices, including standard reference materials and ocean DOM, was performed using GC-MS/MS (Louchouarn et al., 2010) after CuO oxidation and TMS derivatization. Ionization was performed with 70 eV EI and analysis was conducted with an ion trap MS in full scan or MRM mode. In MRM mode, acidic lignin components were easily identified by loss of the acid group (m/z 44), as were components with an aldehyde functional group (m/z 29) such as vanillin and syringaldehyde. Overall, a vast improvement in sensitivity and selectivity was achieved when using MRM mode.

2.3 Pyrolysis-GS-MS (Py-GC-MS)

2.3.1 Introduction to Py-GC-MS

Pyrolysis is a very useful tool for degrading biopolymers and natural organic materials such as humic acids or lignins into their fundamental structural subunits (Martin et al., 1977; Martin, Saiz-Jimenez, & Cert, 1979), either for better analytical understanding of their composition or as an industrial scale process to directly convert lignin to high-value aromatics, hydrocarbons and phenols (J. Zhu et al., 2020). Thermal activation of a biopolymer via pyrolysis allows the breaking of covalent bonds and the composition of the degradation products depends on the size of the polymer and available resonance structures (Sjöström & Reunanen, 1990). Following a pyrolyzing step with separation and analysis via GC-MS gives researchers a quick and powerful tool for determining important properties of lignin such as relative proportions of H/G/S subunits (Martin, Saiz-Jimenez, & Gonzalez-Vila, 1979), along with the potential to quantify these subunits (Faix et al., 1987).

Advantages to this method include the vast amount of literature supporting its efficacy, its relative inexpensiveness when compared to methods involving costly high-resolution instruments such as FT-ICR-MS and the ability to run samples without first separating lignin from the lignocellulosic matrix or any attached carbohydrate units (Faix et al., 1987). Therefore, Py-GC-MS (alongside the thioacidolysis and DFRC GC-MS methods reviewed earlier) has become a standard method for determining H/G/S lignin ratios in many materials, including agricultural and industrial residues (Sequeiros & Labidi, 2017), Cardoon stalks (Lourenço et al., 2015), beech wood (Hansen et al., 2016), wheat straw (Yang et al., 2010), Eucalyptus wood (Gonzalez-Vila et al., 1999), kenaf (Kuroda et al., 2002; Mazumder et al., 2005), and various other nonwoody plants (del Río & Gutiérrez, et al., 2007). Results from Py-GC-MS based methods are also quite comparable to those from other more established wet chemical methods such as the Klason lignin content (van Erven et al., 2017; Fahmi et al., 2007).

The primary disadvantage of Py-GC-MS-based methods is their inability to measure anything larger than lignin monomers or dimers, due to the high degree of degradation of the lignin structure during pyrolysis. However, they remain highly relevant analytical methods in lignin analysis due to how well they perform their specialized task. For example, Py-GC-MS has been a key analytical technology in discovering additional, less abundant monomer subunits of lignin such as 5-hydroxyguaiacyl units (del Río & Martínez, et al., 2007), further deducing the mechanisms of how these subunits come together to form the lignin structure in vivo (Guo & Wang, 2015) and how their ratios and composition change when modified by biological processes (Laskar et al., 2013).

2.3.2 Modifications and improvements

There have been many improvements to Py-GC-MS-based methods to optimize the analysis of lignin. For example, the temperature of pyrolysis has a significant effect on the distributions of lignin pyrolysis products. It has been shown that the higher the pyrolysis temperature, the more demethoxylation, demethylation, and alkylation reactions occur simultaneously, leading to more alkylphenol and polyhydroxybenzene products (Jiang et al., 2010). In this study, a pyrolysis temperature of 600°C produced the highest yield of monophenolic compounds. At 800°C, decarboxylation reactions started occurring. This is all valuable information for utilization of lignin, advantageous when tuning the Py-GC-MS method to specific applications. Other studies have used this information on temperature dependence to more effectively study the interaction of certain enzymes with lignin in delignification processes. For example, a sequential pyrolysis method, performed on various
fractions of a lignin sample from 320°C to 800°C, revealed two potential linkages between the enzyme of interest (HBT) and the lignin matrix (Kleen et al., 2003).

The use of tetramethylammonium hydroxide (TMAH) as an alkylating agent can improve volatilization and therefore detection for some polar compounds by GC (Challinor, 1989). In the pyrolysis of lignin, TMAH causes cleavage of the β-aryl ether bonds, helping to retain key functional groups (Fu & Lucia, 2004) and improve identification of the primary three fundamental lignin subunits (Camarero et al., 1999). However, other derivatizing agents are possible. One study utilized hexamethyldisilazane (HMDS) instead of TMAH (Tamburini et al., 2016) to derivatize lignin. This was helpful in avoiding the confusion caused when TMAH methylates lignin phenolic groups, turning them into methoxy groups (the fundamental unit of difference between the S, G, and H subunits). Silylation also served to protect lignin alcohol functionalities from secondary pyrolytic reactions. With these improvements, the researchers were able to measure 60 distinct compounds in their archeological wood samples and ascertain the degree of natural degradation of these samples by comparing the ratios of the pyrolysis products (including monomers, long-chain, monomers, benzyl, acids, alcohols, and demethylated/demethoxylated compounds).

With certain zeolite catalysists, the yield of aromatic and phenol type compounds from lignin in a Py-GC-MS experiment can be significantly improved (Kumar et al., 2019). In a recent study focused on industrial applications of Py-GC-MS, three different zeolites (Y-zeolite, mordenite and ZSM-5) were tested, showing that the final distribution of pyrolysis products depended on the shape, pore size and acidity of the zeolites as well as the pyrolysis temperature (Kumar et al., 2019). Specifically, ZSM-5 and mordenite improved yield of aromatic monomers, whereas Y-zeolite favored aromatic dimers. Indeed, catalysts such as these could prove useful in industrial processes aiming to “tune” pyrolysis to generate specific lignin products of high utility or value.

Briefly, though we will go into detail about various data-processing techniques later in Section 6, there is a rich variety of unique information about a sample contained even in low-resolution full-scan mass spectra such as that derived from typical Py-GC-MS analyses. Tools derived from the “omics” field, such as chemoinformatic analysis or other chemoinformatic data processing techniques, can prove useful in deconvoluting this information and recognizing the unique “fingerprint” of various components in complex mixtures such as lignin. One recent study used Py-GC-MS in tandem with multivariate statistical methods to classify 15 species of Cactaceae based on the composition of the lignocellulosic matrix in their spines (Reyes-Rivera et al., 2020). After preparing and milling the spine samples, the researchers performed Py-GC-MS experiments according to their previously published method (Reyes-Rivera et al., 2018). The acquired spectra were then deconvoluted and the peaks aligned and identified (by comparison to the Massbank of North America and NIST databases). Several multivariate statistical methods were applied to analyze the data, including principal component (PCA) and hierarchical clustering (HCA) analyses. In 2729 mass spectra, the researchers were able to identify 451 compounds, including carbohydrates and lignin derivatives, in a variety of compound classes (ketones, esters, alcohols, furans, and hydrocarbons). Abundances of these 451 compounds were assessed for each species of Cactaceae and relatively good agreement was found between the composition of the spines and the taxonomy of the groups (Py-GC-MS abundance patterns from the spines were very similar between species of the same genus and different from the species of other genera). Figure 5 shows a PCA plot obtained from these data, with colors representing G and H lignin as well as catechols (C).

This study, in addition to successfully demonstrating chemotaxonomic techniques derived from Py-GC-MS data, illustrated just how much information is available from even low-resolution MS data if the right tools are applied. Of course, there will be more information on this subject when dealing with the high-resolution methods later in Section 6.

### 2.3.3 Insight into pyrolysis mechanisms

A modification of the Py-GC-MS technique, termed pyrolysis molecular beam MS (Py-MBMS), was developed to sample pyrolyzing systems at various stages and conditions (temperature, residence time) of the process (Evans & Milne, 1987) and therefore better understand the molecular pathways undergone by biomass during pyrolysis (Hoover et al., 2002). The technique works by immersing a conical sampling probe in the region of the pyrolyzing sample. A small sonic orifice in the probe lets in gases, vapors, and particles, which are then quenched and introduced to the mass spectrometer as a modulated molecular-beam (Milne & Solty, 1983). The resulting Py-MBMS spectra depend on the pyrolysis temperature, the residence time and the physical state of the material pyrolyzed (Penning et al., 2014) and thus only provide a relative measure of pyrolyzed biomass components such as phenylpropanoids and carbohydrates. However, the true utility of this technique is its ability to deliver high-throughput analysis without GC separation (Li et al., 2017). One study employing a high-throughput
Py-MBMS screening enabled the researchers to rapidly determine abundances of G and S lignin in a large population of bioenergy grasses (Penning et al., 2014). Similarly, lignin pyrolysis mechanisms can be further investigated through the in situ capturing of primary volatiles. In desiring to gain more information about lignin pyrolysis behaviors and avoid some of the pitfalls of GC (including the extensive fragmentation caused by EI), a recent study utilized an in situ pyrolysis-double ionization time-of-flight MS method combined with electron paramagnetic resonance (EPR) spectroscopy (J. Zhu et al., 2020). Two ionization sources, a soft vacuum ultraviolet photoionization (VUVPI, 10.6 eV) and electron ionization (EI, 70 eV), enabled the detection of both primary organic volatiles as well as gas products with high ionization energies (such as CO, CO₂, H₂O, CH₄) in a TOF analyzer. EPR spectroscopy was used to detect stable radical intermediates, known to play a key role in the degradation of lignin during pyrolysis. Results from this analysis showed that lignin depolymerization at 100–300°C released G-type subunits, primarily due to cleavage of the β-O-4 linkage. These subunits then underwent cleavage of O-CH₃, C₆-HOCH₃, and C₆-HOH bonds to produce biphenolic hydroxyl compounds, phenols, and aromatic hydrocarbons. The EPR analysis revealed that radical species increased with temperature and were mainly made up of o-methoxy and hydroxyl substituted phenoxy radicals and carbon-centered aromatic radicals.

Coelution of components on the GC column can obscure some key lignin components of potential interest. 2D GC-MS offers one solution to this problem by employing a second column with different dimensions and stationary phase from the first to better resolve similar components. Recently, 2D GC was used in combination with a TOF-MS to investigate the pyrolysis characteristics of guaiacol (G) lignin from milled wood (Wang et al., 2019). The group had previously argued that MWL best represents the natural structure of lignin in biomass as compared to more typically available alkali and organosolv lignins (Wang et al., 2015). Indeed, MWL has already been the subject of several one-dimensional GC-MS studies, with noted limitations such as coelution of key components unresolvable by TOF-MS. Wang et al. therefore wanted to add a second dimension of GC separation to better understand their sample. In their recent study using HP-5 and DB-1 columns entrained to their TOF-MS, they were able to detect over 300 compounds in their sample. An example of a 3D peaks diagram for the MWL sample pyrolyzed at 550°C is visible in Figure 6.
The authors then proceeded to look at how the composition of pyrolysis products changed at different pyrolysis temperatures, observing that the decomposition of G-lignin at 450°C favored the formation of guaiacol phenols, especially 2-methoxyphenol and 4-methoxy-3-methylphenol. At higher temperatures, the abundance of G phenols decreased, as phenols increased until becoming the primary pyrolysis product from 650°C to 750°C. Aromatic hydrocarbons were not observed from 450°C to 550°C, but were generated from 650°C to 750°C. In this example, 2D GC-MS proved to be a very useful tool in discerning how pyrolysis products changed with temperature.

There also exist some creative techniques for insight into pyrolysis mechanisms through hybrid instrumentation. One such example is the GC-MS method with atmospheric pressure chemical ionization (APCI) developed by Larson et al. to investigate the structure of pyrolyzates from pyrolyzing Kraft alkali lignin. They connected their pyrolyzer to a GC column, on the back end of which they attached a GC-APCI source to interface the GC with their TOF-MS instrument (Larson et al., 2018). Their APCI source ionized lignin pyrolysis products as molecular ions (M⁺), protonated molecules ([M+H]+) and/or ammonium ion adducts ([M+NH₄]⁺). They were also able to perform tandem MS (MS/MS) experiments via collision induced dissociation (CID). The resulting MS/MS spectra were then compared to an experimental MS/MS database (CSI:FingerID, matched against >95 million molecular structures in PubChem library), enabling positive identification (59 out of 72) of the chromatographic peaks in their CID spectra. Despite challenges in determining exact matches for structurally similar compounds, their method could certainly serve to complement the traditional GC-EI-MS analysis which usually follows pyrolysis, offering more insights into these complex mixtures for situations where the EI-MS library is not sufficient, such as volatile pyrolysis products or low abundance natural products (Larson et al., 2018).

2.3.4 Thermogravimetric analysis

Naron et al. noticed some deficiencies in traditional Py-GC-MS methods, including acknowledgment that many dimeric and trimeric products of pyrolysis were not detectable with GC (due to size or lack of volatility), therefore biasing the method towards monomers with single phenolic ring. They attempted to limit secondary reactions and capture more of the actual phenolic lignin structures through their thermogravimetric analysis-thermal desorption-GC-MS method (Naron et al., 2017). Lignin samples were pyrolyzed over a range of temperatures from 30°C to 600°C with a low heating rate in an Ar atmosphere with a thermo-gravimetric analyzer. Volatiles emitted from the TGA were then captured with stainless steel, conditioned thermo-desorption (TD) tubes containing porous polymer sorbents. Trapped volatiles were then released offline via a thermal desorption system into a GC-MS system. Quantification of 5.5–12.9 wt.% of the dry lignin sample with 26 identified aromatic products.

![Figure 6](https://example.com/figure6.png)
compounds was achieved versus fast pyrolysis studies, which only managed to capture 3.6–6.8 wt.% with 21 compounds.

Recently, Chen et al. augmented their thermogravimetric MS of lignin pyrolysis products with CaO and K₂HPO₄·3H₂O in an attempt to understand how these catalysts could change the resulting pyrolysis products. Results showed that K₂HPO₄·3H₂O helped to form more aromatic ring products and catalyzed the demethylation of oxyphenols to form phenol, whereas addition of CaO reduced CO₂ emissions in the first pyrolysis stage and lowered the temperature of CO₂ emissions in the second stage (Chen et al., 2020). Used together, the two catalysts reduced peak temperature and weight loss rate, proving their potential use in bulk pyrolysis biomass conversion processes.

Another significant study into pyrolysis behaviors used the power of on-line photoionization MS along with thermogravimetry to study lignin and lignite copyrolysis (Zhou et al., 2019). Lignite is a low-rank coal with low calorific value, high water content (25%–65%) and low sulfur content (Yu et al., 2013). It bears some similarities to Kraft or alkali lignin, due to its high oxygen content, lower calorific value and high aromaticity. There is interest in processes that coconvert lignite and lignin to higher value products, as there are positive synergistic effects including the reduction of GHG emissions and SOₓ, NOₓ species (Zhou et al., 2019). With their method, the authors immediately noticed significantly different pyrolysis product distributions for Kraft lignin (catechol, guaiacol, and methyl guaiacol) and lignite (phenols and alkylphenols such as cresol) when measured separately. When copyrolyzed, however, there was a significant and advantageous synergistic effect close to the temperature of the maximum loss rate, where the production of volatiles and single-ring aromatics were increased and the total char yield was decreased. The blend ratio of the Kraft lignin and the lignite also affected the resulting products, with a 2:1 ratio of KL/LI promoting the production of guaiacols and syringols. The blend ratio, however, did not seem to affect the yield of phenol, cresol, and aromatics. In this case, being able to observe these changes in real-time was a significant advantage in the decision to use a photoionization-based MS method.

Thermogravimetric analysis, though quite useful for monitoring the physical properties of a substance undergoing a heating process, cannot determine the exact products or material evolved from these processes or resolve thermally overlapping events (Kamruddin et al., 2003). Evolved gas analysis (EGA) is a technique which can fill this gap in understanding, providing information on liberated product gases from these processes. As an example, one technique used EGA-MS to obtain information on the thermal complexity of barks from broad-leaved tree species and was able to measure several of their main chemical constituents (suberin, polysaccharides, lignin, tannins, and extractives) (Ribeichini et al., 2015). Another study used EGA to look at lignin oxidation and depolymerization in archeological wood, identifying three “zones” of thermal degradation, in which lignin, cellulose, and hemicellulose would variously exhibit thermal degradation (Tamburini et al., 2015). Differences in chemical composition during thermalization were indicative of woods of varying age and chemical alteration/degradation. For specific applications, it seems EGA-MS can provide information inaccessible through conventional thermogravimetric analyses and therefore improve our understanding of the gaseous products evolved during pyrolysis processes.

### 2.3.5 Internal standardization with ¹³C lignin

Matrix effects and variations in system performance have greatly contributed to the inaccuracy of many studies attempting the quantitation of lignin and lignin subunits via Py-GC-MS (van Erven et al., 2019), rendering most attempts semi-quantitative at best. This “quantitative” gap in the literature is quite pronounced and there is a general lack of the use of internal standards or relative response factors for various pyrolysis products. However, van Erven et al. (2017) have recently performed some very impressive work to achieve analytical quantitation of lignin in Py-GC-MS analyses by using ¹³C lignin as an internal standard. They were able to obtain nonlabeled (98.9 at. % ¹²C) and uniformly ¹³C-labeled (97.7 at. % ¹³C) spring wheat plants produced under identical conditions in custom growth chambers. After isolating the lignin from the plants via an earlier method (Björkman, 1956), they analyzed for carbohydrate content and composition, protein content and ash content via previously published procedures and also assessed lignin purity via NMR. Samples were then pyrolyzed and immediately analyzed via GC-MS coupled to a single-quad MS. The researchers then undertook quantitation by first establishing relative response factors for 21 of the 46 identified pyrolysis products using authentic standards. Following this, quantification was first attempted for reconstituted biomass model systems (where both ¹²C and ¹³C lignin had been added to a standard along with cellulose and 50:50 (v/v) EtOH:CHCl₃) and then for real lignin samples (mixed with a quantity of ¹³C-IS solution). Using
SIM mode, they monitored the two most abundant fragments of each of their previously established 21 pyrolysis products and quantified based on peak areas using the following equation, where \( i \) refers to the pyrolysis product being quantified, \( A \) is area, RRF is relative response factor, \( m_{IS} \) is the amount of IS, \( m_{sample} \) is the amount of sample (μg) and PIS is a correction factor for the purity of the IS (van Erven et al., 2017):

\[
\text{Lignin content (w/w)\%} = \frac{\sum_{i=1}^{46} A_{IS}^{13C} \times m_{IS} \times P_{IS}}{\sum_{i=1}^{46} A_{IS}^{13C} \times m_{sample} \times 1.057} \times 100.
\]

This quantitation method was remarkably effective, enabling the researchers to quantify \(^{12}\text{C}\) lignin in their reconstituted biomass model systems on the basis of the \(^{13}\text{C}\) internal standard with excellent linearity (\( R^2 > 0.998 \)) and reproducibility (RSD < 1.5%). When applying this method to real-world lignin samples, the same method was used and compared against the classical Klason lignin method. Results showed good accuracy (samples deviated a maximum of 5% from the Klason method) and reproducibility (RSD < 7%).

Therefore, the internal standard served to correct for matrix effects and made genuine, reproducible quantitation possible for both total lignin content and subunit content in common poaceous biomass sources. Although highly effective, the method is limited by the necessity of finding and/or custom engineering \(^{13}\text{C}\) labeled lignin from identical or at least very similar botanical sources to the desired \(^{12}\text{C}\) lignin to be analyzed (van Erven et al., 2019).

A follow-up study expanded on this landmark method by using well-characterized, isolated lignin from uniformly \(^{13}\text{C}\)-labeled willow and Douglas fir (DF) wood as internal standards (van Erven et al., 2019). This enabled expansion of the method from grass lignin samples to include hardwoods and softwoods. In addition, the authors employed HRMS (as discussed later in Section 6) to further improve sensitivity and accuracy. With the reduced background noise in the HRMS, they were able to operate in full scan instead of SIM mode and therefore extended the method to a potentially infinite number of subunit quantitation experiments in a single MS run. The method fared well in both studies on biomass model systems and actual lignocellulosic biomass samples. Figure 7 displays their results from actual lignin samples from several botanical origins and compares them directly to the Klason lignin content, with an average relative deviation of approximately 1% and good linearity (\( R^2 = 0.963 \)). This impressive work demonstrates that with the use of the appropriate internal standard (i.e., coming from the same source as the desired lignin sample to be analyzed), Py-GC-MS can be a benchmark quantitative method for lignin quantification and structural characterization in grasses, hardwoods and softwoods.
### Table 3

Identified compounds in the processed lignin samples (reprinted from Prothmann et al., 2017 under the CC-BY-4.0 license)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Compound</th>
<th>Retention time (min)</th>
<th>Calculated m/z (([\text{M-H}]^-))</th>
<th>Measured m/z (([\text{M-H}]^-))</th>
<th>Base peak intensity</th>
<th>MS² fragmentation</th>
</tr>
</thead>
</table>
| A Guaiacol 0.51 123.0446 123.0446 4.3E5 123 → 108 CH₃ Vanillin 1.27 151.0395 151.0397 5.1E5 151 → 136 CH₃ acetovanillone 1.27 165.0552 165.0556 1.1E6 165 → 150 CH₃ p-Hydroxybenzaldehyde 2.51 121.0290 121.0293 3.3E5 121 → 92 CHO p-Hydroxyacetophenone 2.55 135.0446 135.0445 4.8E5 135 → 120 CH₃ p-Hydroxyacetophenone 2.55 135.0446 135.0445 4.8E5 135 → 120 CH₃ Vanillic acid 3.29 167.0345 167.0356 2.2E6 167 → 152 CH₃ p-Hydroxybenzoic acid 4.05 137.0239 137.0251 2.5E5 137 → 93 CO₂ Vanillin 1.27 151.0395 151.0399 2.3E5 151 → 136 CH₃ acetovanillone 1.27 165.0552 165.0563 1.7E5 165 → 150 CH₃ p-Hydroxybenzaldehyde 2.49 121.0290 121.0293 3.3E5 121 → 92 CHO p-Hydroxyacetophenone 2.53 135.0446 135.0447 1.6E5 135 → 120 CH₃ p-Hydroxyacetophenone 2.53 135.0446 135.0447 1.6E5 135 → 120 CH₃ Vanillic acid 3.29 167.0345 167.0349 2.3E6 167 → 152 CH₃ p-Hydroxybenzoic acid 4.05 137.0239 137.0250 2.4E5 137 → 93 CO₂ Guaiacol 0.50 123.0446 123.0442 1.4E6 123 → 108 CH₃ Vanillin 1.26 151.0395 151.0391 3.0E5 151 → 136 CH₃ acetovanillone 1.26 165.0552 165.0553 1.7E6 165 → 150 CH₃ p-Hydroxybenzaldehyde 2.49 121.0290 121.0293 1.2E5 121 → 92 CHO p-Hydroxyacetophenone 2.53 135.0446 135.0447 8.2E5 135 → 120 CH₃ p-Hydroxyacetophenone 2.53 135.0446 135.0447 8.2E5 135 → 120 CH₃ Sinapaldehyde 2.14 207.0658 207.0645 1.6E4 207 → 192 CH₃ 207 → 177 2x CH₃ Syringic acid 3.50 197.0450 197.0446 3.4E6 197 → 123 2x CH₃ + CO₂ 197 → 95 2x CH₃ + CO₂ + CO Ferulic acid 3.60 193.0501 193.0490 1.5E4 193 → 134 CH₃ + CO₂ Sinapinic acid 3.81 223.0607 223.0603 3.7E6 223 → 193 2x CH₃ 223 → 149 2x CH₃ + CO₂ 223 → 121 2x CH₃ + CO₂ + 2x CH₂
3 | LIQUID CHROMATOGRAPHY-MS METHODS

3.1 Introduction to LC-MS methods for lignin

There is an extensive array of LC-MS methods employed in the analysis of lignin. One of the primary advantages of LC over GC when analyzing lignin and lignin degradation products is that there is no need for derivatization to make analytes more amenable to volatilization or to thermally stabilize them once in the gas phase. This minimizes sample preparation time and cost of analysis (Pecina et al., 1986) and allows for the analysis of much larger intact lignin subunits and structures (Dier & Rauber, et al., 2017). There are also an enormous variety of ionization methods available, ranging from electrospray ionization (ESI) to APCI, atmospheric pressure photoionization (APPI) and many more (Kubátová et al., 2020). More details on these ionization techniques are discussed separately in Section 5.

As an introductory example, Mokochinski et al. developed a simple protocol to determine lignin S/G ratio in plants by UHPLC-MS. Their environmentally friendly, NaOH hydrolysis procedure (95°C for 24 h) was followed by acid neutralization and organic extraction of lignin products with ethyl acetate (Mokochinski et al., 2015). This mixture was then ionized via negative ion mode ESI and analyzed with a tandem quadrupole MS. They identified three monolignols (H, G, and S) and hydrolysis products at 28 Da less than the precursor—4-hydroxybenzaldehyde (H'), vanillin (G'), and syringaldehyde (S') (see Table 3). They also compared their S/G ratio for a single species of *Eukalyptus* (Eukalyptus has very low H) as determined with conventional Py-GC-MS and found good agreement (3.49 vs. 3.87 ± 0.17).

There have been several attempts to improve LC stationary and mobile phases to better separate lignin monomers and oligomers. One such example is the use of ultrahigh performance supercritical fluid chromatography followed by TOF-MS analysis to improve separation (Prothmann et al., 2017). By using a supercritical CO₂ mobile phase, the researchers achieved equivalent or shorter analysis times than had been previously seen, but with significantly improved and well-scattered separation for lignin monomers (primarily a test set of 11 phenols obtained from lignin through alkaline CuO oxidation). Their method achieved separation of these compounds in 6 min, in comparison to a previously reported 15 min (Xin-Ping et al., 2014).

Table 3 lists a variety of identified compounds in these processed lignin samples. Ionic liquids are compelling solvents for use as chromatographic stationary phases due to their ability to effectively dissolve far more lignin material than other solvents. In one study, the Volmer group developed novel mixed-mode phosphonium-based ionic liquid stationary phases for chromatographic separation of complex mixtures of decomposed lignin (Dier & Rauber, et al., 2017). This was performed to separate isobaric compounds observed in earlier HRMS-based studies conducted without chromatographic separation. The use of ionic liquids in the LC stationary phase allowed the researchers to see various classes of compounds previously inseparable as isobars in traditional reversed phase chromatography. For example, the oxidized β-O-4 class of compounds, previously found to have a total of 12 m/z features (Dier et al., 2016) was separated into three classes of compounds: Two classes of aromatic aldehydes, eluting at 18.1–18.7 min and 19.2–24.3 min for the SilPrPhoOTf ionic liquid stationary phase, and one class of aromatic acids, eluting at 63.7–66.3 min with the same stationary phase (Dier & Rauber, et al., 2017).

The Le Masle group has been using nontraditional LC methods to perform highly effective separations of lignocellulosic biomass products. Many of these methods involve pairing with MS, including using centrifugal partition chromatography paired with LC-MS to probe aqueous biomass samples (Dubuis et al., 2019). In this method, the researchers were able to fractionate and identify carbohydrates, furans, carboxylic acids, and phenols from a total of 217 peaks in their 2D chromatography system paired with low resolution MS with positive and negative ESI ionization. The resulting 2D maps allowed them to “fingerprint” samples coming from various processes.

Recently, the group added a size exclusion chromatography dimension to their reversed-phase LC-HRMS method (Le Masle et al., 2014) for analysis of lignocellulosic biomass products of mass <1000 Da (Dubuis et al., 2020). Adding the second chromatographic dimension almost doubled the quantity of peaks observed for their model compounds as well as for two samples obtained from thermochemical and biochemical processes. High-resolution MS was performed with a FT-ICR instrument and enabled the researchers to precisely identify individual compounds by their exact mass (see Figure 8, where every point on the figure refers to the formula of one particular lignin compound). Several families of compounds were observed when the data was plotted as van Krevelen and DBE plots. In the van Krevelen diagrams, carbohydrates appear at H/C ratios >1.5 and O/C ratios >0. Oxygenated aromatics were noticed at the lowest H/C and O/C ratios (H/C <1.2; O/C <0.5) and lignin-carbohydrate complexes (LCC) at 1.1 < H/C < 1.3; 0.5 < O/C < 0.7 (hard to distinguish from oxygenated aromatics).
Therefore, the DBE plots were useful in separating these two compound classes, where LCC are easily disentangled from oxygenated aromatics, with DBE > 6 and mass > 275 Da. This method seems very promising for the untargeted detection of new compounds in lignin/lignocellulose samples.

3.2 Liquid chromatography-MS/MS

MS/MS methods can assist in identification of mixture components, through known fragmentation routes or by comparison to existing literature or databases. Indeed, MS/MS has been a key part of untangling the structure of lignin, as understanding how lignin “breaks apart” can give definitive clues to what linkages and subunits existed in the intact lignin polymer.

An excellent example of a comprehensive tandem-MS method for lignin monomers and oligomers is from Kiyota et al. (2012), who performed analysis of soluble lignin monomers and oligomers in sugarcane by UHPLC-MS/MS. They also developed their own library of LOs by creating synthetic lignin in vitro, first mixing different ratios of the three monomers and added peroxidases to imitate the natural polymerization process and following with UHPLC-MS/MS analysis. The results from their sugarcane lignin samples were then compared to this database for identification. Their instrumentation included UHPLC feeding into a triple-quadrupole (QQQ) MS via negative ESI. Observed MS/MS transitions for lignin monomers included the loss of a methyl group (15 u), loss of CO (28 u), or a water molecule (18 u). For dimers, there were a variety of MS/MS fragments comparable to literature, such as the characteristic m/z 221 ion from fragmentation of the deprotonated (8−5) linked G−G dimer (the 4-aliphatic end), or the m/z 151 ion from the (8−8) linked G−G dimer (the 8-phenolic end). Figure 9 shows some of these noted fragmentations. In addition, the researchers noted similar transitions were noted for other combinations of dimers, trimers, and tetramers.

UHPLC-MS/MS was successfully combined with alkaline nitrobenzene oxidation for the determination of
lignin monomers in wheat straw (Zheng et al., 2017). First, a wheat straw sample was oxidized via the above process, producing phenolic aldehydes such as p-hydroxybenzaldehyde, vanillin, and syringaldehyde. Given that the molar ratio of these aldehydes was known to correspond to the relative amounts of the uncondensed H, G, and S units in the unmodified lignin sample (Min et al., 2014), the following UHPLC-MS/MS method was optimized for the aldehyde products, achieving separation in 6 min with good linearity ($R^2 > 0.997$). MRM was employed, allowing the observation and measurement of the three corresponding monomers simultaneously, without interferences.

A similar method looked at lignin phenols in more complex environmental samples with UHPLC-MS/MS (Yan & Kaiser, 2018). First, the researchers cleaned up their Brazos River methanol extract and humic acid reference material via SPE. Then they subjected the samples to separation by UHPLC with a C18 column. Ionization was ESI in both positive and negative ion modes, followed by detection with a QQQ instrument. Characteristic MRM transitions were selected for each compound, considering both specificity and intensity of the transition. In the analysis they also employed 13C-labeled surrogate standards to improve their quantitative results. With this method they were able to measure ultra-low lignin phenol concentrations (<100 fmol) with good reproducibility (RSDs < 10%) in environmental samples displaying high matrix interferences. For validation, they compared their results with a GC-EI-QQQ-MS method discussed earlier in this review (Kaiser & Benner, 2012), with relative average deviations <10% between two methods for individual and total lignin phenol yields.

The Kenttämaa group has produced some excellent work looking at fundamental gas phase ion chemistry in the fragmentation and tandem MS analysis of lignin and lignin-related compounds. For example, Amundson et al. (2011) used negative ESI and multiple CID steps to identify and count oxygen functionalities in aromatic analytes related to lignin. The authors looked at fragmentations of 40 small aromatic lignin-related analytes, conducting MS/MS experiments up to MS$^3$ using a linear ion trap MS. They were able to identify and count various

---

**FIGURE 9** Dimers of G units showing the three main links. Their characteristic fragmentation patterns are indicated by red lines. (I) The 8--O--4, $\beta$-aryl ether link yields fragment ions A$^-$ ($m/z$ 195) and B$^-$ ($m/z$ 179). (II) The 8--5, phenylcoumaran link yields fragment ion B$^-$ ($m/z$ 221). (III) The 8--8, resinol link yields fragment ion A$^-$ ($m/z$ 151) (reprinted with permission, Kiyota et al., 2012) [Color figure can be viewed at wileyonlinelibrary.com]
oxygen-containing functionalities for their individual analytes (carboxylic acid, hydroxymethyl, acetyl, keto, aldehyde, methoxy, aliphatic hydroxy, and nitro groups) and then applied their method to LC-MS/MS analysis for a mixture of four of the compounds. This proved the applicability of using higher-order MS² analyses to better understand the composition of small lignin-related compounds using even a simple low-resolution, ion trap MS. A follow up study by Marcum et al. (2016) continued this study, looking at CID spectra of 34 model lignin degradation products via both experimental and computational approaches. MS/MS experiments up to MS⁶ were performed to characterize these fragments and the fragmentation patterns allowed the researchers to clearly distinguish carboxylic acid, aldehyde, ester, and phenol groups. Applying this knowledge to an organosolv lignin sample allowed them to identify the presence of specific functionalities and their combinations in lignin components via a single HPLC-MSⁿ experiment.

Later, Zhu et al. expanded on these fundamental approaches by including ion-molecule reactions with diethylmethoxyborane (DEMB) to identify phenol functionalities in deprotonated monomeric and dimeric lignin degradation products via tandem MS. The group began by analyzing model lignin-like compounds containing phenol, carboxylic acid, and other functionalities via negative ESI ionization with an ion trap MS, including NaOH as a dopant (Zhu et al., 2016). All analytes were deprotonated to form [M-H⁻] ions and then allowed to react with DEMB in an external reagent mixing manifold, forming DEMB adduct ions ([M-H + DEMB]⁻), some also having lost a methanol molecule ([M-H + DEMB-MeOH]⁻). Only model compounds without a strong electron withdrawing substituent in the ortho- or para-position failed to exhibit this substitution. By including this gas-phase ion chemistry in their method, the researchers were able to differentiate between deprotonated isomeric compounds with phenol and carboxylic acid, aldehyde, carboxylic acid ester, or nitro functionalities. Their method was then applied to the analysis of a complex biomass degradation mixture by HPLC, enabling the analysis of an entire class of analytes in a complex mixture using a single HPLC run. Certainly, it seems that ion-molecule reactions can play an important role in methods attempting the high-throughput screening of lignin degradation product mixtures, as elegantly demonstrated here.

As the group's approach became more articulated, they expanded to begin sequencing LOs using their detailed understanding of CID gas phase ion chemistry. Sheng et al. (2017) demonstrated an CID MSⁿ procedure for sequencing LOs based on a study of synthetic model compounds, synthesizing seven oligomeric compounds with β-O-4 and 5-5 linkages (ranging from a dimer to an octamer), representing G-type lignin. They then conducted CID experiments, looking at MS² and MS³ spectra to observe fragmentation patterns. This technique, alongside measuring the elemental compositions of the most abundant fragment ions, allowed them to differentiate between key functional groups and ions.

4 | IMAGING METHODS

4.1 | Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)

MALDI has played an important role in the analytical understanding of lignin via mass spectrometry in recent decades. MALDI is able to ionize and measure larger intact fragments of the lignin structure than were previously possible (Banoub & Delmas, 2003). In contrast to the Py-GC-MS-based techniques that preceded it, early lignin studies with MALDI were able to ionize intact oligomers, from trimers of ~600 Da to nonamers of ~1800 Da (Metzger et al., 1992). In addition, MALDI can achieve spatial resolution within sample cross sections with a relatively simple sample preparation, allowing for rapid qualitative assessment of the localization of lignins in various samples. When paired with a mass analyzer such as TOF, a wealth of sample information—including average molecular weights, types and quantities of repeating units and end-group determination—can be deduced for a lignin sample from a MALDI analysis (Richel et al., 2012).

One study used MALDI to assess the quantity and type of lignin in two Eucalyptus species (Araújo et al., 2014). First, the researchers identified silica as an ideal MALDI matrix for lignin, as it was able to ionize soluble lignin structures with no added chemical noise and was significantly cheaper than conventional MALDI matrices. Samples were cut to ~1.5 mm thick by hand and dusted with silica powder. Analysis was conducted with a MALDI linear trap instrument. Identification of oligomers was achieved by comparison to a library of MS/MS data (Kiyota et al., 2012), as shown in Table 4. With their method, the researchers were able to identify 22 out of the 24 compounds in this library, representing the primary soluble lignin monomers and oligomers.

Another advantage of MALDI is the ability to look at samples in two dimensions (MS-based imaging). Figure 10 displays photographs of Eucalyptus sections followed by images generated from MALDI imaging software (Araújo et al., 2014). The software was able to assign a value corresponding to a single compound in
each stem cut from the sum of the intensity of each pixel. A relative quantification was then performed by dividing these values by the sum of the all values obtained in the same cut (de Oliveira et al., 2013). As a result, it was possible to clearly visualize the localization of each of the lignin S, G, and H units in the plant cross-sections.

Numerous improvements have been made to MALDI-based techniques to optimize for the analysis of lignin. Ionization modes, either positive or negative, can reveal different aspects of the lignin structure. For example, using a MALDI matrix of α-cyano-4-hydroxycinnamic acid/α-cyclodextrin, Richel et al. noted that positive ion mode was highly efficient for low molecular weight lignin components, emphasizing 8-S bonds between phenylpropane units particularly at m/z 331 and 314. Negative ion mode, however, displayed good efficiency for higher molecular weight ammonia lignins, emphasizing β-O-4 aryl ether linkages (Richel et al., 2012).

In addition to ionization mode, the matrix employed has a marked effect on ionization of the selected analyte. For MALDI, 2,5-dihydroxybenzoic acid (DHB) is quite a common choice, but often alternative matrices are employed, including 2,5-dihydroxyacetophenone (DHAP) or α-cyano-4-hydroxycinnamic acid (CHCA) (Bowman et al., 2019). In their MALDI study of LOs, Bowman et al. used cationization techniques to form [M + Na+] and [M + Li+] ions and noted that using a DHAP matrix and lithium cationization together offered an improved average signal intensity for model lignin dimers. Their optimized matrix also resulted in less complex and more reproducible positive ion spectra.

Matrix effects and interferences can significantly complicate MALDI spectra and make interpretation difficult and thus techniques are often developed with the aim of mitigating these interferences as much as possible. Nano-assisted laser desorption/ionization time-of-flight MS (NALDI-TOF-MS) can help in eliminating matrix effects and improving ionization efficiency (Yoshioka et al., 2012). NALDI is a matrix-free MALDI-type method that utilizes a nanostructured silicon-based target plate with an LDI instrument. Figure 11 demonstrates a comparison of spectra for a lignin sample obtained with both MALDI and NALDI, which have different ionization profiles emphasizing different components of the complex mixture. In particular, the intense peaks in the lower part of the NALDI-TOF mass spectra represent additions of monolignol units and could be used for structural analysis.

The Kosyakov group has made significant advances in the optimization of MALDI techniques and sample preparation conditions for the analysis of lignin. In one study, they looked into the effect of six common MALDI matrices used for analysis of polymers, in addition to investigating three different methods of mixing the matrix and sample components: s-m-s, representing applying the lignin solution before and after the application of the matrix onto the target; m-s, representing the application of the lignin after the matrix and m+s representing spotting the pre-mixed matrix and sample (Kosyakov et al., 2014). Table 5 displays the results of this optimization study. The most successful matrices were α-cyanohydroxycinnamic, ferulic and DHBs and 2,4.

### Table 4 Relative pixel quantification of oligomers in *Eucalyptus globulus* and *Eucalyptus grandis* (reprinted with permission, Araújo et al., 2014)

<table>
<thead>
<tr>
<th>Parent ion (m/z)</th>
<th>Product ions (m/z)</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>149</td>
<td>131, 103</td>
<td>p-Coumaroyl alcohol (H)</td>
</tr>
<tr>
<td>179</td>
<td>164, 161</td>
<td>Coniferyl alcohol (G)</td>
</tr>
<tr>
<td>209</td>
<td>191, 176, 161</td>
<td>Cinapyl alcohol (S)</td>
</tr>
<tr>
<td>297</td>
<td>279, 191, 121</td>
<td>H(8-5)H</td>
</tr>
<tr>
<td>267</td>
<td>149, 121</td>
<td>H(8-8)H</td>
</tr>
<tr>
<td>327</td>
<td>313, 309, 191</td>
<td>G(8-5)H</td>
</tr>
<tr>
<td>357</td>
<td>339, 221, 191</td>
<td>G(8-5)G</td>
</tr>
<tr>
<td>375</td>
<td>327, 195, 191</td>
<td>G(8-5)G</td>
</tr>
<tr>
<td>387</td>
<td>369, 221, 191</td>
<td>S(8-5)G</td>
</tr>
<tr>
<td>405</td>
<td>375, 225, 179</td>
<td>S(β-O-4)G</td>
</tr>
<tr>
<td>417</td>
<td>181, 166</td>
<td>S(8-8)S</td>
</tr>
<tr>
<td>435</td>
<td>225, 209, 195</td>
<td>S(β-O-4)S</td>
</tr>
<tr>
<td>493</td>
<td>313, 297, 191</td>
<td>H(8-5)H(β-O-4)G</td>
</tr>
<tr>
<td>535</td>
<td>517, 505, 221</td>
<td>G(8-5)G(8-5)G</td>
</tr>
<tr>
<td>553</td>
<td>357, 339, 195</td>
<td>G(β-O-4) G(8-5)G</td>
</tr>
<tr>
<td>571</td>
<td>391, 375, 179</td>
<td>G(β-O-4)G(β-O-4)G</td>
</tr>
<tr>
<td>583</td>
<td>387, 369, 195</td>
<td>G(β-O-4)S(8-5)G</td>
</tr>
<tr>
<td>601</td>
<td>357, 339, 225</td>
<td>S(β-O-4)G(8-5)G</td>
</tr>
<tr>
<td>535</td>
<td>387, 209, 195</td>
<td>G(β-O-4)G(β-O-4)S</td>
</tr>
<tr>
<td>613</td>
<td>387, 369, 225</td>
<td>S(β-O-4)S(8-5)G</td>
</tr>
<tr>
<td>643</td>
<td>417, 195, 181</td>
<td>G(β-O-4)S(8-8)S</td>
</tr>
<tr>
<td>661</td>
<td>435, 225, 209</td>
<td>S(β-O-4)S(β-O-4)S</td>
</tr>
<tr>
<td>851</td>
<td>627, 433, 417</td>
<td>S(β-O-4)S(8-8)S</td>
</tr>
</tbody>
</table>
Figure 10  Phloroglucinol staining and mapping of sinapyl (m/z 209), coniferyl (m/z 179), and p-coumaroyl (m/z 149) alcohols in freshly hand-cut sections of *Eucalyptus* stems. Different colors were selected for each unit for illustrative reasons (reprinted with permission, Araújo et al., 2014) [Color figure can be viewed at wileyonlinelibrary.com]

Figure 11  Mass spectra of isolated lignin obtained from *Pinus densiflora* (linear-positive ion mode). Measurement by (A) MALDI-TOF MS and (B) NALDI-TOF-MS (reprinted with permission, Yoshioka et al., 2012)
6-trihydroxyacetophenone using the s-m-s sample application procedure.

The group is also well known for pioneering the use of ionic liquid matrices in MALDI analysis of lignin. They published a study after noting poor ionization efficiencies with traditional MALDI matrices and proposed a variety of ionic liquid matrices (Kosyakov et al., 2018). With this innovation, they achieved impressive signal intensities and reduced matrix interferences and saw intact, singly charged LOs up to 3 kDa in size.

The Volmer group has also investigated a variety of MALDI matrices for the analysis of lignin degradation products (Qi & Volmer, 2019b). They noticed that the selection of matrix had a significant influence on selectivity (particularly for sulfur and nitrogen-containing species) and ionization efficiency. Direct LDI (no matrix) was found to be well-suited for ionizing lignin containing only C, H, and O atoms; DHB seemed to select for longer chain alkane species; DCTB revealed high nitrogen content, some of which was believed to come from the matrix itself; and CHCA was found to have a severe matrix effect, making it unsuitable for lignin analysis. The study effectively demonstrated that, as for many other MALDI analytes, it is necessary to tune the matrix to the sample to achieve optimal results.

Banoub et al. have made significant progress in using MALDI and MS/MS methods to perform preliminary sequencing of LOs. In a 2015 review paper, the group made some novel proposals for the structure and sequence of lignin (Banoub et al., 2015). In summary, a popular understanding is that native lignin polymers exist as macromolecules, linked to cellulose fibers, in the lignocellulosic matrix. Contrary to this, Banoub et al. proposed that native lignin polymers instead exist as linear, related oligomers of different lengths, which are then covalently linked in crisscross patterns to cellulose and hemicellulose fibers. They also suggested that achieving more accurate structural determination for lignin in its native form requires analyzing so-called virgin released lignin (VRL), which has not been purified or chemically transformed apart from chemical hydrolysis and/or enzymatic hydrolysis to liberate it from the glycolignin complex. Conversely, they defined processed modified lignin (PML) as lignin that has experienced more chemical transformations of the lignin biopolymer resulting from, for example, the Kraft or Organosolv lignin processes. While PML no longer represent lignin in its native form, they argue that VRL come closest to representing lignin in its native form. Thus, attempting to sequence VRL oligomers represents a significant step closer to understanding the true nature of the lignin polymer in vivo.

Recently, the group performed a top-down sequencing of LOs extracted from date palm wood via MALDI-TOF analysis (Albishi et al., 2019). They identified six novel VRL oligomer molecules, describing their structure in detail via never-before-reported CID-MS/MS fragmentations. The authors were able to sequence the oligomers based on the well-known H, G, and S subunits, as well as a recently proposed C subunit (Tobimatsu et al., 2013), which is a catechyl lignin homopolymer derived solely from caffeoyl alcohol. Based on these four subunits, they were able to fully describe the following oligomers—HG dimer, HHSSG pentamer, HHHGSG hexamer, HHHSSG hexamer, HGGSSG hexamer, and CGGSSG hexamer.

Most recently, the Banoub group conducted an in-depth MALDI analysis of a French Oak lignin sample. Using MALDI-TOF-MS in negative ion mode and high-energy collision CID-TOF-TOF-MS/MS, they were able to identify a novel series of lignin and tricin derivatives attached to carbohydrate and shikimic acid moieties (Mikhael et al., 2020). They used the group’s top-down MALDI-MS sequencing method to fully describe these twenty novel compounds, observing VRL rich in syringol moieties, syringyl lignin units, and tricin derivatives. Their analysis revealed more complexity and diversity of lignin- and lignin-linked components than previously observed for such a sample, leading the researchers to suggest that as these techniques improve, we can no longer depend on calculating the ratio of the traditional H/G/S monolignols by pyrolysis GC/EI-MS for structural determination of lignins (Mikhael et al., 2020).

**TABLE 5** Effect of matrix and method of sample application onto the target on the intensity of signals in MALDI mass spectra of lignin (matrix-to-sample ratio 100:1) (reprinted with permission, Kosyakov et al., 2014)

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Signal intensity in the ranges m/z 850–2000/2000–4000/4000–8500</th>
<th>mV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s-m-s</td>
<td>m-s</td>
</tr>
<tr>
<td>SA</td>
<td>–/-*</td>
<td>–/-</td>
</tr>
<tr>
<td>DHB</td>
<td>248/61</td>
<td>206/50</td>
</tr>
<tr>
<td>CHCA</td>
<td>245/53</td>
<td>173/41</td>
</tr>
<tr>
<td>THAP</td>
<td>241/49</td>
<td>145/33</td>
</tr>
<tr>
<td>HPA</td>
<td>107/-</td>
<td>63/20</td>
</tr>
<tr>
<td>FA</td>
<td>250/44</td>
<td>182/39</td>
</tr>
<tr>
<td>THAP + CHCA</td>
<td>258/54</td>
<td>154/32</td>
</tr>
<tr>
<td>THAP + DHB</td>
<td>245/64</td>
<td>212/62</td>
</tr>
<tr>
<td>THAP + CHCA + DHB</td>
<td>234/79</td>
<td>172/70</td>
</tr>
</tbody>
</table>

*No signal or the signal is masked by the matrix.*
4.2 Time-of-flight-secondary ion MS (TOF-SIMS)

TOF-SIMS is another important imaging technique used to assess complex, biologically derived materials. MALDI and SIMS imaging are often seen as complementary techniques, where MALDI is able to measure high-mass molecules such as peptides, proteins and oligosaccharides and TOF-SIMS is well suited to analyzing lipids, unaltered surfaces and extremely fine structures with \( m/z < 1500 \) (Gross, 2017).

Saito et al. have made important contributions towards the use of TOF-SIMS to better characterize and understand lignin. In looking to first identify the characteristic secondary ions of the lignin polymer, they employed a 15 keV, 2 nA current primary beam \(^{69}\text{Ga}^+\) liquid metal ion source and were able to achieve a resolution of 2000–3000 at \( m/z \) 23 (Saito, Kato & Tsuji, et al., 2005). With this, they were able to quickly identify guaiacyl (\( m/z \) 137 and 151) and syringyl rings (\( m/z \) 167 and 181)—the basic G and S units of lignin. Figure 12 shows their spectra of two MWL with the characteristic ions for each of G and S type monomers.

Following this study, the group proceeded to study linkages between the monomers in lignin dimers using their method. In this case, the researchers wanted to know which interunit linkages were disrupted to give rise to the characteristic SIMS ions they had previously noted for lignin and additionally, if C–O–C and C–C linkages were ruptured (Saito, Kato, Takamori, et al., 2005). Using previously established protocols, they synthesized a variety of lignin dimers representing the most common linkages within lignin (\( \beta-O-4'\), 8-1', 8-5', 8-8' and 5-5') and proceeded to analyze them using their TOF-SIMS method (Saito, Kato, Tsuji, et al., 2005). They found that the characteristic ions at \( m/z \) 137 (\( C_6-C_1 \) benzyl ion, \([C_8H_9O_2]^+\)) and 151 (\( C_6-C_1 \) benzoyl ion, \([C_8H_7O_3]^+\)) arise from breaking of the \( \beta-O-4'\), \( \beta-1'\), \( \beta-8'\), and \( \beta-5'\) interunit linkages, but not 5-5'. Direct Ga primary ion bombardment was primarily responsible for breaking the \( \beta-O-4'\) and 8-1' linkages. The researchers also noticed adduct ions, including \([M+13]^+\) and \([M+CH]^+\), thought to arise from the combination of the molecules with their stable fragments.

After establishing the fundamentals of the TOF-SIMS technique for the characterization of lignin, the group applied their discoveries to demonstrate the power of the technique by directly mapping the morphological distribution of syringyl and guaiacyl lignin in the xylem of maple wood (Saito, Kaori, et al., 2012). They also compared their results (S/G ratios) with an established method (thioacidolysis) to assess the applicability of their method for quantitation. Though an earlier study by Zhou et al. (2011) had established the possibility of S/G monomer quantitation using TOF-SIMS, their results were only able to demonstrate a qualitative distribution.

**Figure 12** Positive TOF-SIMS spectra of (A) pine MWL and (B) beech MWL. The figure shows the tentative structures of the main secondary ions, each of which has a guaiacyl (G) or syringyl (S) ring (reprinted with permission, Saito, Kato & Tsuji, et al., 2005)
of these monomers in wood cross sections. Figure 13 demonstrates the power of the TOF-SIMS method in assessing spatially resolved concentrations. From the figure, it is possible to see that the G unit is primarily located in and around the vessel walls, ray parenchyma cells and bark, while the S unit is more concentrated in the fiber walls. Their results from TOF-SIMS also compared well to a thioacidolysis method when determining the lignin S/G ratio in several contiguous growth rings, thus validating the utility of their method for lignin quantitation.

Secondary neutral MS (SNMS) is a modification of SIMS, which generates desorbed neutral analyte molecules that must be ionized before analysis. The proposed advantage over SIMS is that neutrals will not experience as much environmental loss before being analyzed as ions, potentially making this approach more quantitative than SIMS (Kollimer et al., 2003). For example, one study used this technique to characterize the fragmentation mechanisms of two lignin monomers, coniferyl, and sinapyl alcohol (Takahashi et al., 2011). The researchers used single-photon, near-threshold ionization using wavelength-tunable synchrotron vacuum-ultraviolet (VUV) radiation to ionize their neutrals and then performed detection with a TOF-MS analyzer. They also performed SIMS experiments for comparison, noticing that the VUV-SNMS spectra were qualitatively simpler than SIMS spectra, showing mostly a few strong clear characteristic peaks at \( m/z \) 124, 137, 151, and 180. In particular, the strength of the parent mass at \( m/z \) 180 is a very useful feature. However, a clear disadvantage of their VUV-SNMS technique was its lack of sensitivity due to the low flux of ionizing photons, something the SIMS technique did not experience. Regardless, both SIMS and SNMS were useful in characterizing ionization energies, fragmentation mechanisms and characteristic mass markers for lignin model compounds.

There has been an effort to expand the library of secondary ions that distinguish lignin and polysaccharides in SIMS analyses, particularly necessary when analyzing wood (Goacher et al., 2011). Recognizing that a majority of ions in TOF-SIMS spectra of wood samples fall below \( m/z \) 100, Goacher et al. focused on identifying interferences and characteristic ions in this mass range. They managed to find 38 additional ions to create their expanded library and confirmed their results by observing localized lignin-derived peaks in the cross-section TOF-SIMS images that were concentrated in the middle lamella and cell corners, as well as characteristic polysaccharide peaks concentrated in cell walls. They later applied their TOF-SIMS method to studying the action of bacterial laccase-mediated systems on both hardwood and softwood samples (Goacher Robyn et al., 2018).

Bacterial degradation of lignin via laccase enzymes represent important processes naturally and in application to industrial delignification (Singh et al., 2017). TOF-SIMS is an excellent technique for visualizing these changes, as it allows for direct analysis of the lignocellulose matrix with very minimal or no sample preparation. The Master group had previously characterized the action of lignin peroxidase and manganese peroxidase on ground wood samples (MacDonald et al., 2016), but their recent work represents a more direct analytical approach. Here, they examined peak ratios for the three characteristic S, G, and H lignin monomers to determine how the bacterial enzymes modified their sample compositions (Goacher Robyn et al., 2018). They also observed the effect of three enzyme mediators (ABTS, gallic acid and sinapic acid) on these same peak ratios. Their observations led them to conclude that both enzymes were effective at modifying their wood samples and they were able to distinguish between lignin modification (by monitoring the increase and decrease of specific lignin moieties) and delignification (by monitoring the polysaccharide peak fraction, indicating net lignin loss or gain at the sample surface).

5 | ATMOSPHERIC PRESSURE IONIZATION-LOW RESOLUTION MS METHODS WITHOUT CHROMATOGRAPHY

5.1 | Introduction to API techniques

API techniques are soft enough to allow the molecular ions of large biomolecules and NOM-type structures, such as lignin, to be present in the mass spectrum. In all API techniques, a charge is transferred to the analyte via an ionized dopant or electrolyte (Kubátová et al., 2020). These low energy interactions promote soft ionization such as proton addition or removal, or molecular adduct ions. Figure 14 from the recent review by Kubátová et al. summarizes these techniques nicely.

Depending on the mixture being analyzed, each ionization technique will have different ionization efficiencies for different components in the mixture and reveal different aspects of the mixture being analyzed. This section will focus on introducing ESI, APCI, and APPI techniques individually and highlighting important literature contributions employing these ionization methods for the analysis of lignin.
5.2 | ESI-low resolution MS

ESI is a very soft ionization technique, which allows for ions very large masses (up to $10^6$ Da) to be analyzed without disintegration (Heck & van den Heuvel, 2004). This is of high relevance for lignin, as larger LOs are otherwise disintegrated in other methods (i.e., EI, MALDI). ESI is therefore an ideal ionization method for the analysis of lignin by MS. For example, a recent study used well-characterized $\beta$-O-4’ lignin model compounds to optimize several parameters of their ESI setup, including ionization mode (positive or negative), addition of lithium, addition of polar aprotic “supercharging” solvents (modifiers) such as dimethyl sulfoxide and dimethylformamide and the role of nonpolar groups in affecting ionization response (Asare & Lynn, 2019). Their optimized method involved adding a nonpolar group onto a $\beta$-O-4’ lignin compound, increasing the lithium cationization ESI response in positive ion mode.

FIGURE 13  Simultaneous imaging of lignin in a transverse section including the bark and several growth rings (rings I–V) of maple by TOF-SIMS. Positive images of (A) total ion, (B, C) guaiacyl (G) units at $m/z$ 137 and 151, (D, E) syringyl (S) units at $m/z$ 167 and 181. Each image comprises a series of 18 square images ($400 \times 400 \mu m^2$). Scale bar = 500 µm. (F) Extracted areas by ROI function for vessels (V1–V20) and fibers (F1–F31). (G) Color scale bar: light color corresponds to a high intensity; dark color corresponds to a low intensity. (H) A transverse section mounted on a TOF-SIMS sample holder, showing a measurement area illustrated in total ion image (A). Scale bar = 1 cm (reprinted with permission, Saito, Kaori et al., 2012). F, fiber; P, (arrowheads), ray parenchyma cell; V, vessel [Color figure can be viewed at wileyonlinelibrary.com]
The power of ESI lies in its ability to give intact ion species for large lignin fragments, making it highly useful for sequencing LOs. Finding the sequential order of lignin structural units and interunit linkages (the primary order) was the goal of a study by Evtuguin and Amado. They used negative ESI paired with a QTOF analyzer (resolution set to ~10,000) and harnessed the power of MS/MS to probe the structure of lignin (Evtuguin & Amado, 2003). First, they synthesized lignin dimeric model compounds of \( \beta \)-aryl glycerol, pinoresinol, syringaresinol and dehydrodiconiferyl alcohol and used their MS method to investigate fragmentation patterns. Then, they analyzed a previously well characterized lignin sample (Eucalyptus globulus) which was known to be composed of 85 mol% syringyl and 13 mol% guaiacyl linked by \( \beta \)-O-4', 8-8', 4-O-5', 8-5', and 5-5' linkages, occurring at known mol% frequencies. They were able to identify signals centered around \( m/z \) 340–420, 640, 870, 1070, and 1300 (lignin dimers, trimers, tetramers, pentamers, and hexamers, respectively). Dominant peaks differed from each other by 226 and 196 \( u \), the mass of S and G units, respectively. This enabled the researchers to infer structures for these molecules, as seen in Figure 15.

Boes et al. have performed some interesting work demonstrating the capabilities of ESI-QTOF to perform both qualitative and quantitative analyses of lignin compounds and LCC without extensive sample preparation. Their method involved ionization dopants, targeted MS/MS and an internal standard to correct for ion suppression and similar effects from the complexity of the mixtures. In one study, they achieved quantitation of oxygen-rich eugenol in diesel fuel from 300 to 2500 ng/ml with sufficient linearity (\( R^2 = 0.97 \pm 0.01 \)) and excellent accuracy (percent error = 0% ± 5%) (Boes & Roberts, et al., 2018). In another study, they were able to better characterize the poorly understood LCC in a biorefinery pretreatment stream complexes, finding molecules ranging in mass from 326–714 Da and exhibiting both carbohydrate (glucose and xylose units) and lignin (high lignin-like unsaturation) characteristics (Boes & Narron, et al., 2018). Figure 16 displays one such structure, with identification of all observed fragments.

### 5.3 Atmospheric pressure chemical ionization-LRMS

APCI is, like ESI, a very gentle form of ionization that will likely leave analyte ions intact and thus is suitable for analysis of larger LOs. It also has some potential advantages over ESI including the ability to better ionize weakly polar molecules and a lower sensitivity to matrix effects (Kosyakov et al., 2016). One example is the structural elucidation of wheat straw lignin polymer with APCI and MALDI (Banoub & Delmas, 2003), in which lignin was first extracted from cellulose and hemicellulose via the Avidel procedure (a patented process for the production of pulp, lignins, sugars, and acetic acid by fractionation of lignocellulosic plant material in a formic acid/acetic acid medium) (Kham et al., 2003) and then ionized via APCI in positive and negative ion mode (Lam et al., 2001). Analysis was performed with a QQQ MS.
Figure 17 demonstrates a proposed fragmentation routes for product ion spectra of the deprotonated molecule \( [M - H]^- \) at \( m/z \) 329.

Unfortunately, this study failed to identify mass fragments larger than trimers. The researchers speculated that the fragile lignin polymer was not compatible with atmospheric pressure ionization, as it always broke down into dimeric and trimeric fragments. Other studies have discovered similar results—for example, Haupert et al. performed a characterization of lignin model compounds with both ESI and APCI ionization methods. Their conclusions regarding APCI were much the same as Banoub et al.—the fragmentation in APCI is too extensive to observe the complexity of larger lignin degradation products,

**FIGURE 15**  MS/MS spectra of lignin oligomers with \( m/z \) 643 (top image) and \( m/z \) 839 (bottom image) and correspondent inferred structures (reprinted with permission, Evtuguin & Amado, 2003)
but is perfectly functional and efficient for ionization of lignin model compounds (Haupert et al., 2012).

APCI, despite its limitations, can serve as a tool to better understand the behavior of lignin fragmentation in the gas phase. In general, interpreting MS fragmentation spectra requires an understanding of the dissociation pathways in the gas phase. Traditionally, performing CID experiments has been a standard method of fragmentation, but results depend on several variables (type of ionization source, the source settings, the analyzer, the collision gas, and the collision energy) and are therefore difficult to standardize and compare to each other (Morreel & Kim, et al., 2010). However, APCI ionization paired with an ion trap MS has been successfully used to elucidate the gas-phase fragmentation behavior of several standard compounds linked to the lignome (Morreel & Kim & et al., 2010). In one example, a series of lignin dimers (i.e., dilignols) were analyzed using APCI and the researchers were able to identify the β-O-4‘ linkage-associated β-aryl ethers and benzodioxanes, the 8-5‘ linkage-associate phenylcoumarans and the 8-8‘ linkage-associated resinsols from small characteristic neutral losses for each.

In a breakthrough study published the same year, the Morreel & Dima, et al. (2010) were also able to use their APCI/ion trap MS technique to perform the first known sequencing study of LOs. Based on their understanding of the fragmentation of LOs in the gas phase, they were able to not just sequence a series of oligomers, but also detect new subunits and linkages and obtain conditional frequencies for the different linkage types. Of 134 oligolignols (dimers to hexamers) present in the xylem of poplar trees, they were able to completely sequence 36 and additionally identified 10 previously unknown monomeric units including an arylglycerol end unit. They also employed a consistent shorthand for all identified units.

5.4 | Atmospheric pressure photoionization-LRMS

APPI was developed relatively recently in contrast to other API methods and was primarily seen as an alternative to APCI (Robb et al., 2000). Photon absorption from a laser light source causes electron ejection from the analyte to form the molecular ion M+•. If the analyte has a high proton affinity, it may also form the protonated ion [M+H]+ in the presence of a protic solvent. The unique photoionization mechanism is well suited for ionizing nonpolar compounds, weak acids and halogenated organic compounds, unlike ESI and APCI which rely on charge affinity (Banoub et al., 2007). However, it exhibits many of the same limitations as APCI, including a limited mass range and significant fragmentation, mostly restricting its use to evaluation of individual monomers and their linkages (Kubátová et al., 2020).

Despite its limited utility for general lignin analysis, there have been a few successful studies using APPI. The Banoub et al. (2007) used their APPI-QTOF-MS/MS method to investigate the structure of

**FIGURE 16** Fragmentation of SGRASS Cl-adduct ion 361.0600 elucidates potential structure of low molecular weight LCC (reprinted with permission, Boes & Narron, et al., 2018) [Color figure can be viewed at wileyonlinelibrary.com]
wheat straw lignin, identifying 39 specific oligomeric ions in positive ion mode and at least 18 in negative ion mode. They found that these oligomers were composed of linear polycondensed coniferyl units, primarily repeating phenylcoumaran units made of two di-coniferyl residues which are linked by a C-8–C-5′ covalent bond and the ether C-7–O-4′ linkage. They found that their method offered better ionization for thermally labile and reactive oligomers versus ESI or APCI methods and was able to be performed without any prior sample cleanup or chromatographic purification.

The Kosyakov group was also able to achieve impressive results with APPI for their spruce dioxane lignin samples, noticing improved signal intensities and lower sensitivity to contaminants over ESI and APCI (Kosyakov et al., 2016). Their method will be looked over in more detail in the following Section 6, as their detection involved the use of high-resolution MS, the next topic of review.

6 | HIGH RESOLUTION MS METHODS

6.1 | Introduction to HRMS visualization concepts for lignin analysis

In search of better visualization techniques for the high-resolution data resulting from analysis of complex mixtures such as petroleum, the Hughey et al. (2001) developed Kendrick mass defect (KMD) plots based on fundamental concepts introduced decades earlier. The Kendrick mass scale is based on adjusting the mass of CH₂ units to exactly 14 to facilitate the visualization of complex hydrocarbon molecules (Kendrick, 1963). The following equation demonstrates this:

\[
\text{Kendrick mass} = \text{IUPAC mass} \times \left( \frac{14}{14.01565} \right).
\]

The KMD is then defined as the nominal mass minus the Kendrick mass and then scaled up by 1000:
KMD = (nominal mass – Kendrick mass) \times 1000.

This means that all members of a homologous series of compounds with differing numbers of CH₂ units but the same constitution of heteroatoms and number of rings plus double bonds will have identical KMD values (Qian et al., 2001). The scale as defined this way is ideal for hydrocarbon analysis, but of course custom Kendrick scales can be developed to separate compounds based on the mass of any repeating unit. If the KMDs for multiple components of a complex mixture are plotted versus the nominal mass, the result is called a KMD plot (Figure 18 shows a KMD plot for a degraded lignin sample; HRMS data measured by FT-ICR-MS).

van Krevelen diagrams are another powerful visualization tool able to, for example, clearly separate biomolecules such as lipids, proteins, sugars, char, coal, cellulose, and lignin from each other in two-dimensional space (Podgorski et al., 2012). First, the elemental composition of a sample must be determined, which can be done directly from HRMS data. Then the O:C elemental ratio is plotted against the H:C ratio for every identified molecular formula in the sample. This will conveniently separate various classes of compounds of various oxygenation or saturations. An example for lignin is shown in Figure 19, which shows a spruce dioxane lignin sample as analyzed via APPI-HRMS (Kosyakov et al., 2017). It is possible to clearly observe changes in sample composition from the pure lignin sample (Figure 19A, top of the figure) to the sample after having undergone depolymerization (Figure 19B, bottom of the figure).

Another study used van Krevelen diagrams to observe the lignin components most easily degraded by a laccase produced by the terrestrial fungus *Trametes versicolor* (Echavarri-Bravo et al., 2019). The researchers used negative ion ESI coupled with their 12T FT-ICR MS for the characterization of enzymatic processing of commercial lignin and saw clear changes in elemental distribution and abundance before and after fungal degradation. In general, van Krevelen diagrams have found utility in many analyses and play an important role in deconvoluting HRMS spectra of complex lignin samples. More examples will be reviewed in upcoming sections, particularly Section 6.4 on novel data processing methods for HRMS.

### 6.2 Liquid chromatography-HRMS

In the past decade, LC-HRMS analyses have given significant insight into the structure and composition of lignin. Similar to the LC-LRMS methods examined previously in Section 3, these techniques can resolve isobars inseparable by mass analysis alone and with the added data density of HRMS, they can be powerful tools for digging deep into the lignome. One example is a study from the Kenttämaa group, who characterized organosolv switchgrass lignin with an HPLC-MS³/high resolution system (Jarrell et al., 2014). Figure 20 shows the total ion chromatogram followed by extracted ion chromatograms on the bottom, with clear separation between monomeric and dimeric lignin units, as well as lignin-carbohydrate complexes.
The study also mentioned an improvement in the negative ion mode ESI by adding a hydroxide dopant (Owen et al., 2012). This allowed for the deprotonation of phenolics without their fragmentation, which could then be assigned elemental compositions with analyzed via FT-ICR-MS. They were also able to perform tandem MS via CID in the ICR cell and get valuable structural information on the lignin molecules in addition to the elemental compositions. HPLC separation before MS allowed isobars to be resolved. The analytical setup was able to identify the 5–5 linkage in lignin dimers, a strong linkage that usually does not fragment and had not been observed in any other studies up until this point. Ultimately, the group was able to use this technique to assign formulas and identify common fragments for unknowns in their organosolv switchgrass lignin sample.

Another recent study employed a liquid–liquid extraction followed by MSn (up to MS⁷) with HRMS detection to study a wheat straw lignin sample (Reymond et al., 2020). First, the sample was separated into carbohydrates (AQ1), organic acids (AQ3), phenols (ORG3), and neutral products (ORG2) via a liquid-liquid extraction with water and MTBE. This helped reduce issues with coeluting compounds and ion suppression in the API source. These fractions were then subjected to HPLC-MSn-HRMS analysis, allowing the researchers to observe and elucidate the structure of mono, di, tri and tetra-aromatic lignin compounds as well as heavy (up to 600 g/mol) LCCs and carbohydrate oligomers with acid functionalities.

6.3 Atmospheric pressure ionization-HRMS

Many HRMS-based techniques without prior chromatography employ API sources. Much like the API-LRMS techniques reviewed earlier, these can produce varying results, and different lignin species will be observed depending on the technique used. The Kosyakov et al. (2016) performed a comparative study of negative ion mode API methods using an orbitrap mass analyzer. They found the most efficient method for ionization was APPI with an acetone dopant, at least for their spruce dioxane lignin sample. APPI improved signal intensities and displayed a lower sensitivity to contaminants over ESI and APCI, leading the researchers to claim that this ionization method could be considered the preferred means for studying lignin. Figure 21 shows comparison spectra of these three ionization methods.

Following studies from the group successfully applied their APPI-HRMS method to characterize products from the alkaline decomposition of hydrolysis lignin, revealing oligomers of up to ten aromatic rings with an average MW of 150 Da (Kosyakov et al., 2017) and the structure of grass lignins from nettle (Urtica dióica), revealing formation of lignins by the addition of guaiacyl- and syringylpropane units followed by etherification by p-coumaric, ferulic, and dihydroferulic acids (Pikovskoi et al., 2019).

A recent study surveyed a diversity of molecular lignin products by various ionization techniques and high-resolution MS, specifically FT-ICR-MS (Qi et al., 2020).

Figure 19 van Krevelen diagram for (A) spruce dioxane lignin and (B) products of depolymerization of hydrolysis lignin. The color of points corresponds to the relative intensity of peaks in the mass spectrum (white, 0.1%–1%; gray, 1%–10%; black, >10%) (reprinted with permission, Kosyakov et al., 2017)
In contrast to Kosyakov et al., Qi found that the greatest number of elemental formulae was achieved when using negative ion ESI. ESI in negative ion mode was also deemed the superior ionization source for lignin-like species (O/C 0.2–0.6 and H/C 0.7–1.5) and was able to detect some sulfur-containing species not seen in other ionization modes.

In general, ESI seems to be the preferred ionization source for lignin when performing HRMS studies, if only for the sheer volume of studies utilizing it and/or the convenience of having it preinstalled on many standard HRMS instruments. One recent study used ESI-TOF-MS as a tool for deconvoluting lignin mass spectra (Andrianova et al., 2018). The researchers found an optimized response for lignin model compounds (methoxy-substituted arenes and polyphenols) in positive ion mode. They also analyzed Kraft alkali lignin by the same method and were able to derive a significant amount of information from their TOF analyzer. However, high MW species often carried multiple charges and therefore required deconvolution with the following equation:

\[
\frac{m}{z} \cdot z - [(\text{atomic mass of charge carrier} \\
\text{(}^1\text{H or } ^{23}\text{Na}) - \text{electron rest mass}) \cdot z] = M, \text{ Da.}
\]

Positive ion ESI can also be quite useful for ionizing LOs, especially when using an ionization dopant such as lithium. One such study used this technique to analyze 13 lignin model oligomers containing \(\beta-O-4\) and \(4-O-\alpha\) linkages (Asare et al., 2019). Lithium chloride was used as a dopant to encourage positive ion formation in positive ion ESI. Accurate masses were obtained for all observed lithium adduct ions via an orbitrap mass analyzer, enabling identification of all fragments and proposals of some sequence-specific fragmentation pathways. Figure 22 shows the observed lithium-adduct fragments from one such model lignin dimer. This study illustrates that lignin may prefer to form stable adduct ions with hard cations relative to other molecules that readily accept a proton, and cation-adduct studies should not be overlooked when attempting ESI analysis of lignin and lignin fragmentation mechanisms.

Recently, the Zhang et al. (2020) employed high-resolution orbitrap MS to expand their understanding of the fragmentation mechanisms of deprotonated lignin model compounds in MS/MS. Although many groups have proposed mechanisms of gas-phase fragmentations for LOs and model compounds, most of these were not backed up with calculations and/or experimental data. This study aimed to rectify some of these issues by providing detailed CID data (MS\(^n\); \(n = 2–5\)) backed up by

![Figure 20](https://example.com/figure20.png)

**Figure 20** Top: HPLC total ion chromatogram obtained using ESI in the negative-ion mode for organosolv switchgrass mixture. Bottom: (A) Extracted ion chromatogram for deprotonated analytes with C:O ratios \(\geq 5:2\), \(m/z\) values from 90 up to 250 and double bond equivalence from 4 up to 7. (B) Extracted ion chromatogram for deprotonated analytes with C:O ratios \(\geq 5:2\), \(m/z\) values from 250 up to 400 and double bond equivalence \(>8\). (C) Extracted ion chromatogram for deprotonated analytes with C:O ratios \(\leq 2:1\), \(m/z\) values from 250 up to 400 and double bond equivalence \(\leq 6\). The ion current for additional low molecular weight analytes are also shown, including fatty acids and monosaccharides (reprinted with permission, Jarrell et al., 2014) [Color figure can be viewed at wileyonlinelibrary.com]
quantum chemical calculations. HRMS enabled confirmation of the structures of lignin fragments. The authors managed to identify and describe three major CID pathways for deprotonated lignin model compounds with β-O-4 linkages. This not only helps in rationalizing previously published CID data for β-O-4 lignin compounds but can indicate the existence of β-O-4 linkages in unknown deprotonated lignin compounds.

The weathering of lignin has been effectively studied via API-HRMS. Two studies from the Volmer group demonstrated the ability of FT-ICR-MS using negative ion mode ESI to effectively observe changes in lignin composition (as well as photooxidation products) during irradiation with UV light. Using HRMS and a variety of data mining and visualization tools (van Krevelen diagrams as well as KMD, DBE, and carbon number plots), they noticed extensive oxidation for classes of lignin compounds with high oxygen content and small molecules with a single aromatic ring, whereas larger compounds with multi-aromatic structures were protected from photo-degradation (Qi et al., 2016b). Compounds with DBE < 10 remained particularly stable during the degradation and were proposed as potential marker compounds for the type and source of lignin. A follow-up study improved upon this study by introducing a phase correction technique to improve spectral quality (Qi et al., 2017). By utilizing the absorption mode capability of the FT-ICR instrument (which typically collects data in the so-called “magnitude” mode), the researchers were able to reduce false-positive identifications and distinguish isobars (e.g., SH4 and C6) that were previously unresolvable. This study demonstrates the power of API-HRMS techniques to give advanced insight into changes in complex mixtures over time, especially with regard to the weathering and decomposition of biopolymers.

### 6.4 | Novel data processing methods for HRMS

#### 6.4.1 | Modified and two-dimensional KMD plots

In recent years, an increase in the resolution and general availability of high-resolution mass spectrometers has fueled a simultaneous interest in extracting more information from the rich datasets that result from HRMS analyses (Zhang et al., 2021). Indeed, it is perhaps impossible to harness the full power of HRMS without employing computational strategies to untangle all of trends observable in these datasets. This is especially true when dealing with complex, heterogenous mixtures such as NOM, petroleum, or in this case, lignin. With the highest mass resolution currently available, FT-ICR-MS instruments still stand at the forefront of these innovations and feature prominently in many of the studies reviewed here. One such study used FT-ICR-MS analysis to analyze steam-exploded wheat straw lignin, which contained components ranging from 100 to 100,000 g/mol (Dauria et al., 2012). The researchers noticed that the most intense peaks in their spectrum were all 44.026 u apart, equivalent to a C2H4O (methoxy) group. Upon plotting this data in KMD format, the researchers were able to observe these patterns as plot trendlines on the KMD plot, as seen in Figure 23.

HRMS has also been used to investigate the formation of black carbon-like and alicyclic aliphatic compounds by
hydroxyl radical initiated degradation of lignin (Waggoner et al., 2015). Although this study was more focused on the fate of lignin in aquatic systems (sediments, rivers and oceans), the researchers did make use of KMD plots, also noticing groups of compounds separated by \( m/z \) 44 (a carboxylic acid group), allowing them to suggest that decarboxylation is most likely responsible for the shift to lower O/C observed in their data.

However, both of these highly preliminary studies only scratch the surface of the plethora of information available in the mass defect space. Although many publications also use other visualization tools such as double bond equivalence plots (Bai et al., 2014; Smith et al., 2012) and van Krevelen diagrams (Kim et al., 2015; Kloekhorst et al., 2015; Xu et al., 2012) to interrogate their data, many are focused on the lignin depolymerization process and do not specifically aim to delve deeper into the intact lignome with these visualization techniques.

Recently, the Volmer group has made some novel contributions to graphical and statistical methods for interrogating HRMS data. They have proposed several strategies for constructing modified KMD plots that have enabled deeper understanding of the lignome from the interrogation of high-resolution MS data alone. Instead of using conventional KMD units based on the \(-\text{CH}_2\) moiety (ideal for hydrocarbon/petroleum analysis), base units can be modified to fit common structural themes in lignin. In fact, both axes of a KMD plot can be modified in this way, leading to what the group has termed “2D” mass defect plots. Figure 24 shows an example of such a plot, with horizontal trends in the data depicting the addition/loss of a phenyl group and vertical trends depicting the addition/loss of a methoxy group (Qi et al., 2016a).

The Volmer group has previously applied various mass defect techniques to illustrate trends in electrochemically degraded lignin mixtures (Dier et al., 2016). Their degradation process in ionic liquids and under \( \text{N}_2 \) atmosphere (Reichert et al., 2012) generated various consecutive series of molecules resulting from removal and transformation of \(-\text{CH}_2\), \(-\text{OH}\), phenyl, and other functional groups and subsequent formation of \(-\text{CHO}\) and \(-\text{COOH}\) groups. FT-ICR-MS analyses of their electrochemically degraded lignin were conducted with ESI, APCI, and APPI ionization sources. Raw data was processed to extract chemical compositions of measured \( m/z \) values with relative abundances \( >0.1\% \) (corresponding to \( S/N>4 \)) in the \( m/z \) ranging from 100 to 1000, and composition was restricted to C, H, and O with double bond equivalents (DBE) between 4 and 25 and H/C ratios from 0.5 to 2.5 for both odd and even electron ions. A van Krevelen plot, seen in Figure 25, demonstrates the
variety of compounds found in the electrochemically degraded lignin with the various ionization methods.

The measured $m/z$ ratios were then converted from the IUPAC $^{12}\text{C}$ scale to the standard Kendrick mass scale as well as a new mass scale representing the guaiacol repeating units ($\text{C}_7\text{H}_7\text{O}_2$):

$$m/z(\text{C}_7\text{H}_7\text{O}_2) = m/z(12\text{C}) \times 123/123.044605.$$

The results of this calculation plotted as a modified KMD revealed clear series of compounds with varying oxygen content. Movements vertically in these plots indicate differing numbers of CH$_2$ groups in the compound and movement left-to-right horizontally represents the addition of guaiacol units. Based on patterns from elemental compositions (i.e., DBE, oxygen content, and carbon number), the researchers were then able to assign compounds in the degraded lignins to various molecular classes, as seen in Figure 26.

In a follow-up study, the group applied similar visualization techniques to the analysis of lignin-based model resins (Dier & Fleckenstein, et al., 2017). Again, by modifying the Kendrick mass scales to suit the target analytes, they were able to achieve clear visualizations for components belonging to various compound classes. Figure 27 shows series of compounds in PF resol resins that differ by additions of cresol, phenol, and catechol units.

The Volmer group has also explored ambient ionization techniques in combination with HRMS to characterize lignin. Expanding on the ESI, APCI, and APPI techniques employed by Dier et al., they conducted a study focused on exploring DART, pyrolysis DART, and DESI for direct analysis of solid, powdered lignin samples (Crawford et al., 2017). Visualizing their results with several modified KMD plots, the researchers chose phenol [$\text{C}_6\text{H}_4\text{O}$] as a Kendrick base. By selecting a more complex base, they were able to better resolve components than when they were separated by the methoxy ($-\text{OCH}_2$) base unit and therefore dig deeper into the lignome. The authors were able to use these modified KMD plots to better understand the various species detected in each of the three ambient ionization techniques they employed, as well as conventional negative ion mode ESI.

6.4.2 Other statistical and computational advances

Kew et al. developed a highly useful software tool for the interactive visualization of HRMS data, used by a number of researchers mentioned in this section (e.g., Qi et al., 2020). The collection of scripts, written in the Python programming language, are fully customizable to any parameters desired by the researchers, from custom KMD units to selecting particular constraints for creating possible molecular formulas from HRMS peaks (Kew et al., 2017). Figure 28 displays an example of figures generated by this software, including both van Krevelen and DBE plots. This software is an excellent example of the innovative, custom solutions required by users of HRMS to truly get the most out of the HRMS data resulting from the analysis of complex mixtures such as lignin.
The most recent work from Turner’s group includes developing a nontargeted method for identification of phenolic compounds in complex technical lignin samples (Prothmann et al., 2020), based on ultrahigh performance supercritical fluid chromatography (UHPSFC)–HRMS with a combination of PCA, QDA, and KMD plots to perform discrimination of their HRMS data. Three different technical lignin samples were analyzed, viz. lignoboost Kraft lignin, sodium lignosulfonate lignin, and depolymerized Kraft lignin. Multiple MS^n experiments were used to classify compounds as either lignin monomers, dimers, trimers, or tetramers. Finally, four KMD–PCA–QDA models were constructed, one each for lignin monomers, dimers, trimers, and tetramers. For unknowns, the researchers were able to propose chemical structures based on determined chemical formula, RDB equivalent, and MS^3 fragmentation pathway (Figure 29).

This study demonstrated that a detailed nontargeted analysis of lignin is possible via HPLC-HRMS, requiring six chromatographic injections, one full-scan and five data-dependent neutral loss MS^3 experiments, with a runtime of 10 min each. LOs could be identified by a benzene loss in MS^3 and tentative isomers for several...
**FIGURE 28** Overview screenshot of the i-van Krevelen Main Page. Four subplots are shown, from top left clockwise, van Krevelen, centroid mass spectrum, DBE versus C number, and AI(mod) versus C number. The scatter plots have points sized per their relative abundance, while the color scales represent either m/z range (van Krevelen) or oxygen number (DBE vs. C number and AI(mod) vs. C number) (reprinted with permission, Kew et al., 2017) [Color figure can be viewed at wileyonlinelibrary.com]

**FIGURE 29** Proposed chemical structure and MS$^3$ fragmentation pathway for the tentative lignin monomer with m/z 231.0661 detected in the LKL sample (reprinted with permission, Prothmann et al., 2020 under the CC BY 4.0 license). LKL, lignoboost Kraft lignin
Phenolic compounds were identified through the novel KMD-PCA-QDA model developed by this group. Although more work remains to be done to probe the many unknowns in these complex samples, it seems that methods such as this, utilizing both chromatographic, mass spectrometric and statistical/computational spaces to achieve separation, will be instrumental in paving the way forward.

Recently, computer modeling has been used to stochastically generate oligomeric lignin structures for interpretation of MALDI-FT-ICR-MS results. In a recent study, researchers conducted MALDI-FT-ICR-MS analyses of two MWL samples, a hybrid poplar (HP) and DF (Terrell et al., 2020). Plotting the H/C ratio versus the nominal mass, they were able to see clear “clusters” of lignin fragments, each cluster differing by ~190 u (approximately one lignin monomer), as seen in Figure 30. From this data, they concluded that their observed HRMS peaks were made up of groups containing primarily three, four, and five aromatic rings (and for DF lignin, a small quantity containing six aromatic rings).

Taking a cue from the proposals made by the Volmer group in a study on data processing methods for HRMS spectra (Qi & Volmer, 2019a), Terrell et al. performed a variety of KMD calculations for the two lignin samples. Rather than selecting CH$_2$ as the base unit for the KMD analysis (as in conventional KMD plots targeting petroleum-type mixtures), the researchers used modified base units to visualize changes in common lignin functional groups: C$_6$H$_4$O, OCH$_2$, C$_3$H$_2$, O, and C$_3$H$_3$OH. Clear patterns directly emerged for LOs differing by only a methoxy group (vertical trends). The researchers then used the LigninBuilder software (Vermaas et al., 2019) to build molecular lignin models suitable for simulations. These models enabled them to propose 400 potential candidate structures for the lignin-derived oligomers in the high-resolution FT-ICR-MS data. Their model was informed by results from Py-GC-MS and NMR, and included simulated losses of hydroxyl, hydrogen, and methyl groups from the lignin structures.

### 7 | CONCLUSIONS

As a highly abundant and vastly under-utilized resource, lignin presents an exciting opportunity for new biofuels and as a source of value chemicals, even though its extremely complex and diverse nature presents a series of analytical challenges. This review attempted to give an overview of the most significant advances in MS-based analytical methods as applied to lignin (and more generally, the “lignome”) to date.

The discussion began with a survey of GC-MS-based methods, many of which helped pave the way for more contemporary techniques. Several of these GC-MS techniques are still considered to be “analytical standards” and continue to be used to this day, such as pyrolysis-GC-MS with internal standardization for determination of lignin monomer subunit composition. The following sections gave an overview of LC-MS and imaging-based methods, where softer ionization techniques such as API and MALDI play important roles in observing intact lignin fragments. Finally, the review looked at a variety of techniques using simply mass spectrometry without chromatographic (and often, few sample preparation or cleanup) steps involved beforehand. These ranged from pairing low-cost, low resolution mass spectrometers with soft ionization techniques, such as ESI, to HRMS-based...
methods, where a significant part of the analysis becomes the data-mining/processing after the MS experiment is finished.

In particular, studies utilizing a variety of techniques (including the power of elucidation through computer modeling and visualization) are helping make significant headway into the understanding and utilization of lignin by probing deeper than ever before into the structure and composition of this polymer. Only through the power of high resolution mass spectrometry is a wealth of data for the lignin polymer available without costly prior chromatographic and sample clean-up methods.

Looking towards the future, it is expected that higher mass resolving powers and better and more standardized (and specialized) data analysis tools will lead the way forward in probing further into the structures and patterns observable in HRMS datasets. These methods may be combined with ion mobility spectrometry in the future, to provide additional structure-specific separation power for lignin species. It is without that mass spectrometry, regardless of the specific techniques used, will continue to stand at the forefront of analytical progress towards the understanding of lignin.

ACKNOWLEDGMENTS

Dietrich A. Volmer acknowledges research support by the German Research Foundation (DVG VO 1355/4-3). Dane R. Letourneau is grateful for a supplementary fellowship from the SALSA School of Analytical Sciences Adlershof.

REFERENCES


Larson EA, Hutchinson CP, Lee YJ. Gas chromatography-tandem mass spectrometry of lignin pyrolyzates with dopant-assisted...


Penning BW, Sykes RW, Babcock NC, Dugard CK, Klimek JF, Cynara cardunculus L.) stalks. *Bioenergy Res.* 2005;21:77. Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/mas.21716 by ... on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License


**How to cite this article:** Letourneau DR, Volmer DA. Mass spectrometry-based methods for the advanced characterization and structural analysis of lignin: A review. *Mass Spec Rev.* 2023;42:144-188. [https://doi.org/10.1002/mas.21716](https://doi.org/10.1002/mas.21716)