CrossMark

PAPER IN FOREFRONT

Two-dimensional mass defect matrix plots for mapping genealogical links in mixtures of lignin depolymerisation products

Yulin Qi[1] · Rolf Hempelmann[2] · Dietrich A. Volmer[1]

Abstract Lignin is the second most abundant natural biopolymer, and lignin wastes are therefore potentially significant sources for renewable chemicals such as fuel compounds, as alternatives to fossil fuels. Waste valorisation of lignin is currently limited to a few applications such as in the pulp industry, however, because of the lack of effective extraction and characterisation methods for the chemically highly complex mixtures after decomposition. Here, we have implemented high resolution mass spectrometry and developed two-dimensional mass defect matrix plots as a data visualisation tool, similar to the Kendrick mass defect plots implemented in fields such as petroleomics. These 2D matrix plots greatly simplified the highly convoluted lignin mass spectral data acquired from Fourier transform ion cyclotron resonance (FTICR)–mass spectrometry, and the derived metrics provided confident peak assignments and strongly improved structural mapping of lignin decomposition product series from the various linkages within the lignin polymer after electrochemical decomposition.

✉ Dietrich A. Volmer
Dietrich.Volmer@mx.uni-saarland.de

[1] Institute of Bioanalytical Chemistry, Saarland University, Campus B2.2, 66123 Saarbrücken, Germany

[2] Institute of Physical Chemistry, Saarland University, Campus B 22, 66123 Saarbrücken, Germany

Introduction

The massive worldwide consumption of petroleum and the limited future availability of fossil resources have prompted the development of alternative biofuels. Storable liquid biofuels greatly reduce greenhouse gas emissions in comparison to conventional fossil fuels [1, 2]. While traditional biofuels are mostly produced from vital food supplies such as sugar cane and corn starch, alternative sources—in particular agricultural and municipal wastes—are now increasingly important to recover valuable chemicals contained in the chemically complex wastes. Importantly, these chemicals have significant value beyond simple energy generation.

Lignin is a major component of the cell wall of woody plants and the second most abundant natural biopolymer after cellulose [3]. Studies have shown that the chemical content of lignin is similar to that of fossil fuels [4]. Unfortunately, commercial applications of lignin have remained limited and focus mainly on low value products (e.g. raw materials for chemicals; dispersants for cement dyes; or dust suppression agents for roads) [2, 5], leaving lignin highly underutilised as alternative to conventional biofuels and source of high value chemicals [6].

Structurally, lignins are cross-linked macromolecules with molecular weights of up to 10,000 Da [7], which are mostly formed via free radical coupling of three basic hydroxyphenylpropanoid monolignols: coumaryl alcohol, coniferyl alcohol and sinapyl alcohol (Fig. 1). Catalysed by oxidative enzymes, electrons are delocalised in these monolignols, providing unpaired electron density at positions 1, 4, 5 and 8. The 8-position is the most favoured among the various linkages (Fig. S1, Electronic Supplementary Material, ESM) and the dominant bond in the lignin polymer is therefore the 8-O-4 linkage, comprising approximately 50 % of the linkages in the woody plants [8].
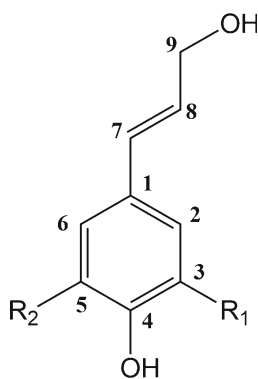
🖄 Springer

Fig. 1 Structures of the three main lignin monolignols with numbering scheme for the carbon positions: coumaryl alcohol ($R_1 = R_2 = H$), coniferyl alcohol ($R_1 = OCH_3$, $R_2 = H$), and sinapyl alcohol ($R_1 = R_2 = OCH_3$)

Because lignin polymers are complex heterogeneous materials, the full characterisation of the individual components at the molecular level is required to assess and salvage the desired high value chemicals from these polymers. Various degradation, extraction and catalysis methods have been developed to explore the lignin content [9, 10]. Unfortunately, the different methodologies often result in different estimates of lignin products, even from the same sample [11], as a result of lignin's inherent complex structural motifs and variable relative abundances within the material. Fortunately, modern instrumental analytical techniques have improved the structural identification process. For example, nuclear magnetic resonance (NMR) was utilised to analyse lignin from plant cell wall [12, 13], which was helpful to describe the different building blocks and bond types of lignin compounds. NMR requires relatively large amounts of analyte material from purified or concentrated samples, however. As a result, NMR experiments of lignin only revealed average structural features of the bulk mixtures, while information on low abundant, uncharacterised building blocks could not be determined [14, 15].

Mass spectrometry (MS) provides important complementary information to NMR, which has been readily demonstrated for samples such as crude oil. Full mass scan spectral analysis of crude oil samples using ultra high resolution Fourier transform ion cyclotron resonance - mass spectrometry (FTICR–MS) generated thousands of m/z features per sample, with relative abundances stretching over several orders of magnitude [16, 17]. From these mass spectra, the individual compounds' accurate masses and elemental formulae were readily determined [18–20]. But even though each peak represented a chemically distinct compound, their structures mostly remained unknown. Lignin exhibits a similarly complex set of components, which has been coined the "lignome" [21]. The lignome is defined as the complement of all phenolics, which are regulated as part of the lignin biosynthesis [22]. Consequently, lignomics analysis of products formed from lignin decomposition and structural interpretation of lignin oligomers are similarly complicated as the petroleomics analysis of crude oils, requiring significant efforts for liquid chromatography (LC) separation, tandem mass spectrometry (MS/MS) and high resolution mass spectrometry (HRMS) analyses [22–27]. A very promising approach was presented by Morreel et al., who successfully elucidated the gas-phase fragmentation behaviour of lignome-associated molecules using comparative MS^n analyses [22]. The authors were able to assign characteristic fragmentation reactions for the essential lignin-bonding types and obtained diagnostic ions for the characteristic aromatic units, providing a foundation for detailed characterisation of the lignome and for MS-based sequencing of lignin oligomers. Owen et al. developed an LC method for lignin degradation products, which was able to analyse lignin-related model compounds in negative ion electrospray ionisation (ESI) mode, giving elemental compositions from accurate mass measurements by FTICR, followed by detailed MS^3 experiments for structural information. The authors demonstrated the utility of their method using a mixture of lignin degradation products [27].

Unfortunately, these and other HRMS full-scan and MS/MS structural protocols are not well suited for obtaining detailed identification of thousands of individual components in heterogeneous mixtures from lignin decomposition. Therefore, data simplification and classification tools are required to reduce the complexity and to visualise the hidden information in the complex mass spectra, to better characterise and classify the samples.

For the study of natural organic matter (NOM), various graphical and statistical methods have been developed to simplify high resolution mass spectra (HRMS). For example, Kendrick mass defect (KMD) analysis has allowed the alignment of thousands of peaks across the m/z range according to their homologous structural units [28]. The van Krevelen diagram helps in sorting the same classes of compounds in specific locations within the diagram and makes it possible to investigate possible links between molecules [29, 30]. Double bond equivalence (DBE) and carbon number distribution plots have been utilised to examine the structural features of natural organic matter in samples from different origins [31].

In this work, lignin was depolymerised by electrochemical decomposition [32] and the products measured by FTICR–MS to obtain detailed insight into the composition of the chemically complex degradation mixtures.

In recent work, we demonstrated the proof-of-concept of using HRMS for these mixtures and investigated appropriate ionisation techniques for analysis [33]. We also applied KMD and van Krevelen plots to interrogate the full-scan mass spectra, and observed a first glance of the enormous complexity of the mass spectral data sets. Here, we are significantly expanding the investigation by implementing much more advanced concepts for data mining of the lignin decomposition products. From the mass spectral raw data, more than 5000 elemental compositions were assigned in a single full-scan mass spectrum. As a result of isobaric inferences, however, most of the low abundant species were impossible to isolate for detailed MS/MS structural analyses, even though they were clearly visible in the high resolution mass spectra. In order to process the bulk of the features contained in the data sets, the concept of two-dimensional (2D) fractional mass filtering, derived from KMD plots, was utilised as a visualisation tool to allow meaningful interpretation of the observed signals and compounds. The 2D matrix plots provided systematic line-ups of the different lignin linkages using structure-specific metrics. The procedure greatly simplified data interpretation, because candidate m/z values and chemical structures were readily deduced from the genealogical links between products and their formation mechanisms, rather than unsystematically assigned chemical formulae in conventional elemental composition analysis. Starting from the low m/z region—consisting mainly of monomers—the core structures of lignin were readily identified via MS/MS; the higher oligomer structures originated from the same linkages and were therefore quickly visualised according to the mass defect base applied in the matrix plots. Structures of higher molecular weights—but lower abundances in the samples—could then be predicted on the basis of this information, and the full-scan data sets specifically mined for them. For validation, the proposed structures were confirmed by detailed collision-induced dissociation (CID) experiments.

Experimental

Reagents and chemicals

Methanol, ammonium hydroxide and the alkali lignin standards were purchased from Sigma-Aldrich (Steinheim, Germany). The electrochemical degradation process was performed as described previously [32, 33], yielding 20 % solid phase (w/w, based on mass of lignin precursor). The powder was dissolved in water/methanol/ammonium hydroxide (50:50:1 v/v/v) prior to MS analysis. Organic-free water was generated by a Millipore (Bedford, MA, USA) Direct-Q8 purification system.

High resolution mass spectrometry

Samples were ionised using electrospray ionisation (ESI) in negative ion mode. Mass spectra were recorded using a 7-Tesla FTICR–MS instrument (Bruker, Bremen, Germany) equipped with an Infinity cell [34]. For each spectrum, 40 individual transients were collected and co-added to the enhance signal-to-noise ratio (S/N) [35]. In MS/MS mode, precursor ions were isolated first in the quadrupole and externally accumulated for 0.1–1 s. For CID, 5–30 V collision voltage was applied.

Data processing

The acquired mass spectra were loaded into Bruker DataAnalysis 4.2 software for data interpretation. The full-scan mass spectra were internally calibrated using a series of homologous compounds throughout the m/z range from 121 to 731. Elemental formulae were assigned to the peaks inside the calibrated m/z range, with the following tolerances to filter the lignin compounds: composition was restricted to C, H and O; DBE ranged from 4 to 25; H/C ratios from 0.6 to 2, and O/C ratios from 0.2 to 0.7 were considered; the acceptable mass error was ±1.5 ppm for singly charged ions. The assigned mass values were imported into an Excel spreadsheet to rescale the m/z regions of interest.

Results and discussion

Electrochemical decomposition of alkali lignin was performed as described previously [32]. In the current study, negative ion electrospray ionisation (ESI) was implemented for MS analysis for two reasons: (1) the lignin decomposition compounds contain hydroxyl (–OH) groups, which were readily amenable to deprotonation to give [M−H]⁻ analyte ions; (2) ESI minimised sample consumption, which was particularly important for the detailed MS/MS acquisition routines performed here. Even though atmospheric pressure photoionisation (APPI) has been shown to provide slightly higher compound coverage in our preliminary experiments [33], the amount of sample required for APPI (or atmospheric pressure chemical ionization, APCI) was significantly higher, which was deemed unpractical in this study.

General appearance of mass spectra and peak assignments

The electrochemically degraded lignin sample was initially measured in the full-scan mode of the FTICR instrument.

The negative ion ESI–FTICR measurements revealed thousands of peaks in the mass spectrum with calculated resolutions ($m/\triangle m_{50\%}$, FWHM) of greater than 400,000 at m/z approx. 300 (Fig. 2). Comparable complexities and peak densities have been observed in environmental samples such as crude oil, asphaltenes and humic acids [36–39]. To demonstrate the density of signals in the full-scan spectra, a region of 1 m/z unit was further expanded in Fig. 2, exhibiting a total of 19 assigned features in this small segment.

Throughout the spectrum, odd nominal m/z values dominated over even mass numbers. The limited nitrogen content in the lignin samples caused this bias, as zero N atoms in the elemental composition automatically gives an odd m/z ion according to the nitrogen rule. Moreover, many of the peaks at even m/z ratios were ^{13}C [M+1] isotope species originating from odd m/z ions; they were readily confirmed by adding the mass of a neutron to the odd m/z species. A set of assignment criteria (viz. compositions restricted to C, H and O; DBE between 4 to 25; H/C ratios from 0.6 to 2; O/C ratios from 0.2 to 0.7; mass error max. ±1.5 ppm for singly charged ions) were then applied as a filter to exclude non-lignin signals, yielding approx. 2300 lignin-related compounds, which were further analysed. Even though high resolution FTICR–MS greatly helped in the unambiguous assignment of elemental compositions for these compounds, it was not possible to fully characterise their individual chemical structures, as detailed MS/MS analysis of each of these 2300 signals was not feasible. For this reason, post data processing methods were developed to reduce the complexity of the full-scan data, to enable rapid identification of important candidate components within the mixture for further analysis. In addition, linked m/z series
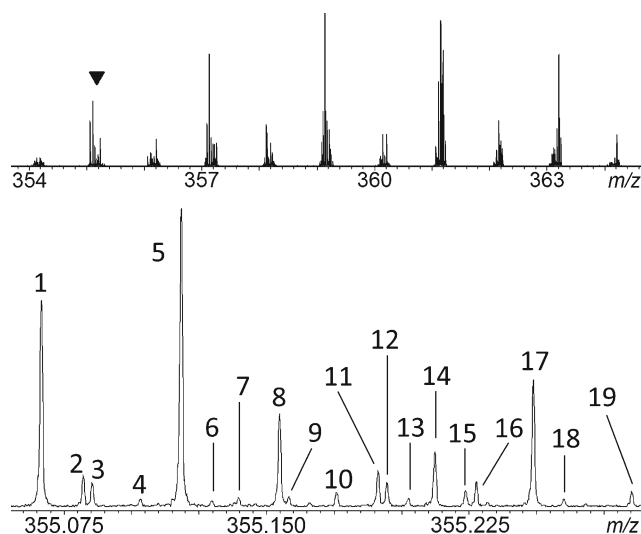

Fig. 2 FTICR–MS full-scan analysis of lignin sample after electrochemical decomposition: (*top*) detailed view of mass segment m/z 354–364; (*bottom*) further mass scale-expansion of the region m/z 355.0–355.3 (marked with *inverted triangle* symbol in *top trace*; chemical formulae of these signals can be found in Table S1, ESM)

were identified using this approach, from similarities and genealogical connectivities between species. These similarities and connections can illustrate different degradation pathways such as oxidation, demethylation, dehydration, dehydrogenation and aromatic ring cleavages.

Two-dimensional matrix network

The Kendrick mass defect (KMD) [40] plot is useful approach to identify compound classes in crude oil samples [28]. In the original application of the KMD, the m/z scale was converted from 14.01565 u to 14 u using the repeating CH_2 unit:

$$\text{Kendrick}\, m/z = \text{IUPAC}\, m/z \times (14/14.01565) \qquad (1)$$

It follows that compounds with the same number of heteroatoms and DBE, but different number of CH_2 units, will have identical Kendrick mass defects:

$$\text{Kendrick mass defect}$$
$$= \text{nominal Kendrick}\, m/z - \text{exact Kendrick}\, m/z \qquad (2)$$

This rescaling process makes it possible to quickly recognise compounds of the same chemical class and type—but different extent of alkylation (CH_2)—from the complex data sets. Importantly, KMD is not limited to CH_2 groups; any functional group of interest can be used as the base for KMD:

$$\text{Modified Kendrick}\, m/z = m/z \qquad (3)$$
$$\times (\text{nominal}\, m/z\, \text{of base unit} / \text{exact}\, m/z\, \text{of base unit})$$

That is, KMD is applicable to any base unit or even higher-order mass defect analysis [33, 41, 42].

The first KMD model used here for lignin decomposition products was based on the simple coumaryl alcohol, with complex oligomers formed via methoxylation [OCH_3] reactions, and various phenyl [C_6H_4O] linkages [32]. Therefore, these two structural units were chosen as KMD bases, to rescale the m/z axis of the mass spectra. The first modified KMD plot shown here used [OCH_2] as horizontally rescaled x-axis (corresponding to methoxylation), while the phenol core [C_6H_4O] was chosen for the vertical y-axis, to line-up compounds belonging to phenyl-specific linkages. As illustrated in Fig. 3, each data point in this 2D mass defect matrix diagram represented a specific component in the mass spectrum of the lignin decomposition mixture, and structurally related compounds were therefore quickly identified for further interrogation. To demonstrate the benefits of this 2D strategy, Fig. 3 illustrates a small region that was further expanded from the 2D plot. In this magnified view, three series of data points are highlighted with red, green and brown colours. The measured m/z value for the first (red) data point was m/z 163.0389. This precursor ion was then further analysed by CID experiments and revealed the structure to be coumaryl acid (the CID spectrum
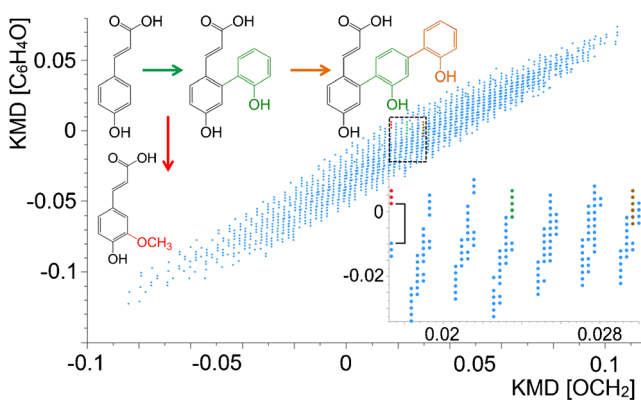
Fig. 3 2D mass defect matrix plot for a lignin sample after decomposition. *Blue data points* represent features in the KMD plot and correspond to degradation products from the sample. The *squared area* is enlarged (*inset*) and proposed core structures of the three compound species (*red*, *green*, *brown*) are highlighted in the expanded plot

is shown in Fig. S2, ESM). The two following, related red data points on the vertical *y*-axis indicated that up to two methoxy (OCH_3) groups can be attached to the compound. To confirm their identities, CID analyses were performed again. Although the absolute methoxylation sites could not be determined from these MS/MS experiments, it was clear that methoxylation occurred on the phenyl core of coumaryl acid and the two compounds corresponding to the red colour points were likely to be coniferyl and sinapyl acid (Fig. S2, ESM).

Similarly, on the *x*-axis, the green and brown data points were both aligned with the coumaryl alcohol series horizontally, which indicated that they shared the same core structure. CID experiments revealed that the structures of these compounds were linked by phenyl additions to the coumaryl acid core (Fig. 3, top).

In addition, it was interesting to note that there was a distinct gap with no signals in the vertical direction of Fig. 3, below the three red points belonging to the coumaryl acid series discussed above, followed by three further data points (shown by a bracket in Fig. 3). CID analysis of these additional three compounds was inconsistent with the coumaryl series, however. This phenomenon highlighted that constitutional isomers likely existed in the aligned compound series, and the gap between the points was an indication for the occurrence of such isomers of structurally different ion species.

Kattner et al. pointed out that for *m/z* values greater than 600, and element restrictions to C, H, O, N, P and S and a tolerance mass window of 1 ppm, more than 15 reasonable molecular formulae can be calculated [43]. Therefore, determining the correct elemental formula for lignin components from the many possible assignments is a challenging task during data processing. After plotting the data points using the rescaled axes in the 2D matrix diagram, however, structurally related compounds are lined up and readily identifiable

from the thousands of different signals. Consequently, the chosen approach improved the confidence for unambiguous peak assignments, because the elemental formulae were determined by using the homologous species as a basis with strict elemental restrictions. With increasing *m/z* value, the number of the possible elemental formulae increased exponentially, but the 2D approach permitted determination of theoretical *m/z* and homologous chemical structures from genealogical links, rather than unsystematically assigning elemental compositions in a shotgun manner, thus significantly improving trust in the peak assignments.

The 2D matrix plot can be further extended to other building blocks of lignin oligomers, to provide a tailored decomposition analysis of the complex mass spectra for specific connectivities. For example, the KMD bases can be modified to probe the various other lignin linkages. Typical linkages, such as dibenzyl ether and 8-O-4 linkages, were also interrogated in this study, the results of which are summarised in Fig. 4. Starting from the basic core structures in the low *m/z* region, the metrics not only improved the structural overview of the spectrum of lignin decomposition products but also eased the constraints for the mass accuracies needed for peak assignments in the complex mass spectra. In the mixtures, it was always difficult to pre-isolate compounds of higher molecular weights, but lower abundances for MS/MS
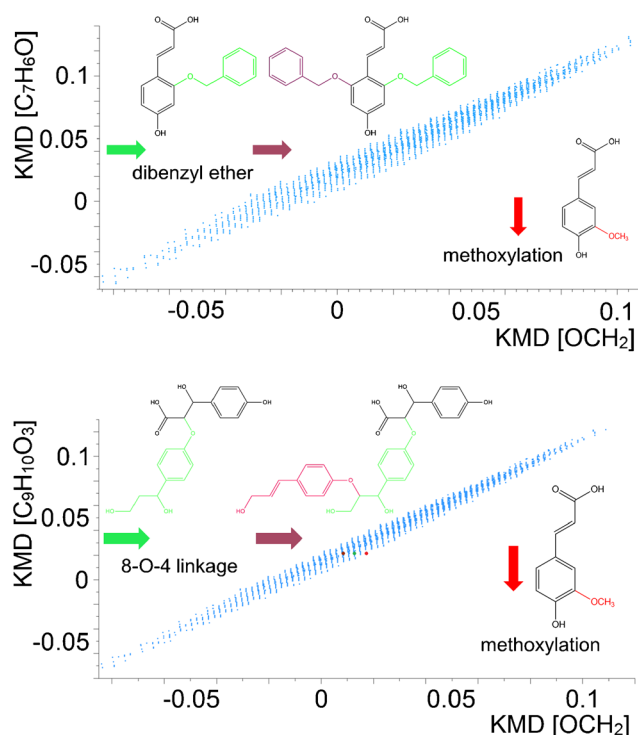

Fig. 4 Two-dimensional mass defect matrix plot for lignin sample components. *Top*: [OCH_2] versus [C_7H_6O]; *bottom*: [OCH_2] versus [$C_9H_{10}O_3$]. The *insets* show the basic structures of coumaryl acid (*black*), methoxylation (*red*), dibenzyl ether and 8-O-4 linkages (*green* and *brown*)

experiments, as a result of interferences from isobaric components. The 2D matrix plots offered a complementary approach to predict the possible candidate structures. Figure 4 shows an illustrative example for the component with measured m/z at approx. 495 (red colour). The 2D matrix plot demonstrated that this data point lined up with coumaryl acid (brown data point), with two $C_9H_{10}O_3$ units separating the two components. Hence, its elemental formula was determined to be $C_{27}H_{28}O_9$. In addition, its molecular structure was also predicted according to the KMD base applied, after attaching two building blocks via the 8-O-4 linkage of coumaryl acid. Subsequent CID experiments (Fig. 5) verified that the proposed structure was indeed correct; a sole chemical structure was assigned to $C_{27}H_{28}O_9$.

Abundance comparisons for selected species

As shown in the previous section, the horizontal and vertical axes of the 2D matrix plots align the compounds according to the specific KMD base used for data interrogation. In addition to these parallel axes, the slopes of the slanted lines also have diagnostic potential. Reactions that involve a loss or gain of a specific elemental formula of structural-related compounds can be identified from these trend lines, as, theoretically, each reaction pathway has its own trend line with its characteristic slope and intercept.

It is apparent from Fig. 3 that numerous trend lines in the 2D matrix plot can be readily distinguished, and particular chemical decomposition reactions can be selected for further analysis. The trend lines representing the hydroxylation and methoxylation reactions for two biphenyl compounds were selected as examples here. Their relative peak abundances were compared and the results shown in Fig. 6. From the plot,
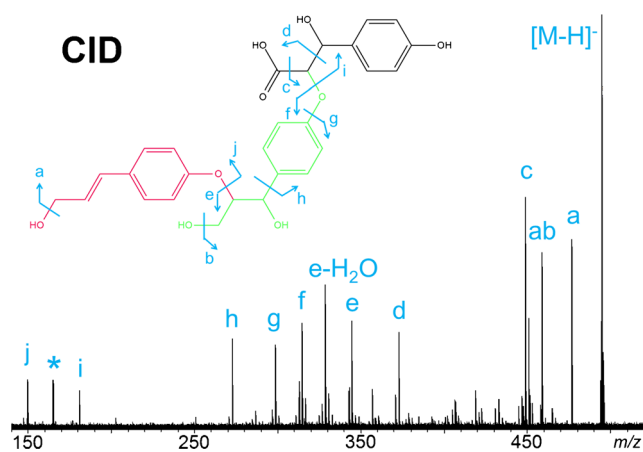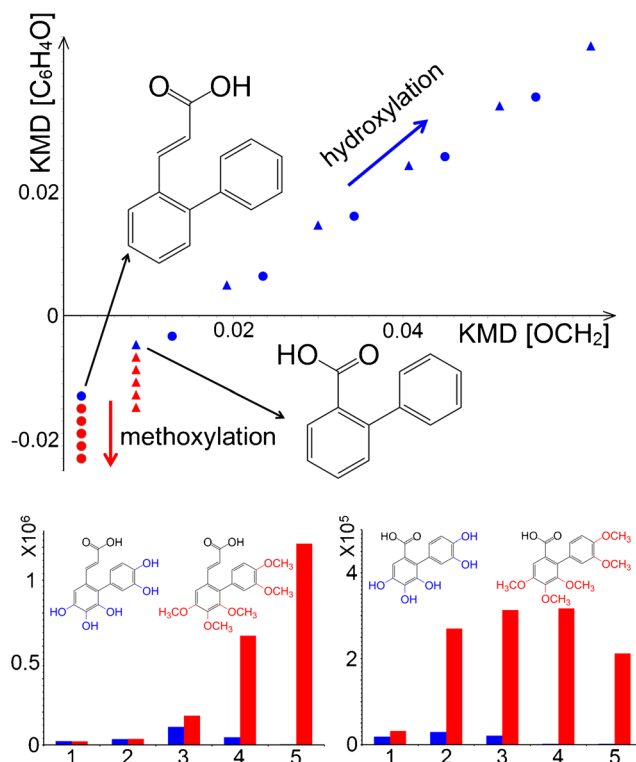

Fig. 6 *Top* Visualisation of hydroxylation (*blue*) and methoxylation (*red*) reaction trend lines for two biphenyl compounds in the 2D matrix plot. The *circular points* represent acrylic acid-substituted biphenyls; *triangular points* correspond to biphenyl-2-carboxylic acid derivatives. *Bottom* Abundances of biphenyl compounds after hydroxylation (*blue*) or methoxylation (*red*); the *x-axis* represents the number of –OH or –OCH$_3$ groups added to the compounds. Detailed compound formulae and their peak areas can be found in Table S4 (ESM)

it is evident that up to five functional groups were added to both compounds. In addition, methoxylation was favoured over hydroxylation for these two selected lignin decomposition products, as the methoxylation product abundances were several orders of magnitude higher than those of hydroxylation products. On the other hand, detailed comparison of these two compound series also demonstrated that the total number of products with attached acrylic acid residue in the 2-position was about fivefold higher than products with attached biphenyl-2-carboxylic groups; this was probably due to the longer side chain of acrylic acid, which provided higher spatial flexibility to the benzene ring for the additional reactions. Probing the possible decomposition pathways via trend lines is a typical analysis from the van Krevelen diagram [28]. Here, this function was greatly enhanced by plotting the data in two dimensions, by allowing the trend line approach to be applied to the 2D matrix plots of lignin products. Of course, given the complexity of the investigated samples, it is always possible that a data point within these plots consists of a mixture of constitutional isomers with different chemical structures.

Fig. 5 CID–MS/MS spectrum of the deprotonated [M−H]⁻ ion of $C_{27}H_{28}O_9$, with fragments labelled in the proposed molecular structure. (Note *asterisk* labels an FTICR artefact and corresponds to the third harmonics of the precursor ion at m/z 495. The detailed peak assignment is summarised in Table S3 (ESM)

While the shown strategy is not a replacement for detailed MS/MS experiments, the 2D matrix plots developed here nevertheless provide a simple filtering and data simplification strategy to reduce the complexity of the data sets, and to allow for selection of relevant precursor structures from the full-scan mass spectra for further study.

Conclusions

High resolution mass spectrometry is widely applied in the analysis of complex samples, such as crude oil, asphaltenes, humic acids and lignin. Recent advances in HRMS have strongly increased the amount of information that can be gathered from a single mass spectrum, including detailed compositional information from spectra that contain thousands of signals. The primary full-scan mass spectrum initially only provides the elemental compositions, and great manual data processing efforts are often required to extract the required information from the data, to achieve the ultimate goal of structural identification of each of the components in the mixture. While liquid chromatography can be used to fractionate complex samples to achieve a certain degree of separation, the enormous complexity of some samples, such as those investigated here, requires further measures to aid the structural elucidation process. Because it is not feasible to conduct detailed MS/MS analyses for each of the m/z features in the sample, various post-processing methods have been developed. KMD plots are widely adopted in petroleomics studies to visually separate compounds with the same heteroatom content (e.g. O, S, N etc.) on the basis of mass defect. Lignin decomposition products, however, are composed of three elements only, viz. C, H and O. Therefore, KMD plots were modified in this study to allow classification of the different internal linkages of the lignin oligomers. The necessary rescaling of the m/z domain can be easily performed via most open source MS tools such as mMASS [44], or simply by using Excel spreadsheets. The strategy shown here quickly highlighted linkage-specific components, provided reasonable prediction of their structures and identified possible reaction pathways for further investigation. The method therefore readily facilitated peak assignment in HRMS spectra of lignin decomposition products. Importantly, the 2D matrix mapping strategy can be equally applied to other complex samples containing molecules with small numbers of heteroatoms, such as natural organic matter, asphaltenes, humic acids and naphthenic acids.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Sticklen MB. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat Rev Genet. 2008;9:433–43. doi:10.1038/nrg2336.

2. Simmons BA, Loqué D, Ralph J. Advances in modifying lignin for enhanced biofuel production. Curr Opin Plant Biol. 2010;13:313–20. doi:10.1016/j.pbi.2010.03.001.

3. Calvo-Flores FG, Dobado JA. Lignin as renewable raw material. ChemSusChem. 2010;3:1227–35. doi:10.1002/cssc.201000157.

4. Mohan D, Pittman CU, Steele PH. Pyrolysis of wood/biomass for bio-oil: a critical review. Energy Fuel. 2006;20:848–89. doi:10.1021/ef0502397.

5. Torney F, Moeller L, Scarpa A, Wang K. Genetic engineering approaches to improve bioethanol production from maize. Curr Opin Biotechnol. 2007;18:193–9. doi:10.1016/j.copbio.2007.03.006.

6. Margeot A, Hahn-Hagerdal B, Edlund M, Slade R, Monot F. New improvements for lignocellulosic ethanol. Curr Opin Biotechnol. 2009;20:372–80. doi:10.1016/j.copbio.2009.05.009.

7. Tolbert A, Akinosho H, Khunsupat R, Naskar AK, Ragauskas AJ. Characterization and analysis of the molecular weight of lignin for biorefining studies. Biofuels Bioprod Biorefin. 2014;8:836–56. doi:10.1002/bbb.1500.

8. Hatfield R. Lignin formation in plants. The dilemma of linkage specificity. Plant Physiol. 2001;126:1351–7. doi:10.1104/pp.126.4.1351.

9. Hanson SK, Baker RT, Gordon JC, Scott BL, Thorn DL. Aerobic oxidation of lignin models using a base metal vanadium catalyst. Inorg Chem. 2010;49:5611–8. doi:10.1021/ic100528n.

10. Hasegawa I, Inoue Y, Muranaka Y, Yasukawa T, Mae K. Selective production of organic acids and depolymerization of lignin by hydrothermal oxidation with diluted hydrogen peroxide. Energy Fuel. 2011;25:791–6. doi:10.1021/ef101477d.

11. Brinkmann K, Blaschke L, Polle A. Comparison of different methods for lignin determination as a basis for calibration of near-infrared reflectance spectroscopy and implications of lignoproteins. J Chem Ecol. 2002;28:2483–501. doi:10.1023/A:1021484002582.

12. Lu F, Ralph J. Non-degradative dissolution and acetylation of ball-milled plant cell walls: high-resolution solution-state NMR. Plant J. 2003;35:535–44. doi:10.1046/j.1365-313X.2003.01817.x.

13. Bozell JJ, O'Lenick CJ, Warwick S. Biomass fractionation for the biorefinery: heteronuclear multiple quantum coherence-nuclear magnetic resonance investigation of lignin isolated from solvent fractionation of switchgrass. J Agric Food Chem. 2011;59:9232–42. doi:10.1021/jf201850b.

14. Capanema EA, Balakshin MY, Kadla JF. A comprehensive approach for quantitative lignin characterization by NMR spectroscopy. J Agric Food Chem. 2004;52:1850–60. doi:10.1021/jf035282b.

15. El Mansouri N-E, Vilaseca F, Salvadó J. Structural changes in organosolv lignin during its reaction in an alkaline medium. J Appl Polym Sci. 2012;126:E214–21. doi:10.1002/app.36628.

16. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom

Rev. 1998;17:1–35. doi:10.1002/(SICI)1098-2787(1998) 17:1<1::AID-MAS1>3.0.CO;2-K.

17. Amster IJ. Fourier transform mass spectrometry. J Mass Spectrom. 1996;31:1325–37. doi:10.1002/(SICI)1096-9888(199612) 31:12<1325::AID-JMS453>3.0.CO;2-W.

18. Headley JV, Peru KM, Barrow MP. Advances in mass spectrometric characterization of naphthenic acids fraction compounds in oil sands environmental samples and crude oil-a review. Mass Spectrom Rev. 2015;35:311–28. doi:10.1002/mas.21472.

19. Cho Y, Ahmed A, Islam A, Kim S. Developments in FT-ICR MS instrumentation, ionization techniques, and data interpretation methods for petroleomics. Mass Spectrom Rev. 2015;34:248–63. doi:10.1002/mas.21438.

20. Hughey CA, Rodgers RP, Marshall AG. Resolution of 11 000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil. Anal Chem. 2002;74:4145–9. doi:10.1021/ac020146b.

21. Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, et al. Metabolic engineering of novel lignin in biomass crops. New Phytol. 2012;196:978–1000. doi:10.1111/j.1469-8137.2012.04337.x.

22. Morreel K, Kim H, Lu F, Dima O, Akiyama T, Vanholme R, et al. Mass spectrometry-based fragmentation as an identification tool in lignomics. Anal Chem. 2010;82:8095–105. doi:10.1021/ac100968g.

23. Morreel K, Dima O, Kim H, Lu F, Niculaes C, Vanholme R, et al. Mass spectrometry-based sequencing of lignin oligomers. Plant Physiol. 2010;153:1464–78. doi:10.1104/pp.110.156489.

24. Kiyota E, Mazzafera P, Sawaya ACHF. Analysis of soluble lignin in sugarcane by ultrahigh performance liquid chromatography-tandem mass spectrometry with a do-it-yourself oligomer database. Anal Chem. 2012;84:7015–20. doi:10.1021/ac301112y.

25. Jarrell TM, Marcum CL, Sheng H, Owen BC, O'Lenick CJ, Maraun H, et al. Characterization of organosolv switchgrass lignin by using high performance liquid chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-ion mode electrospray ionization. Green Chem. 2014;16:2713–27. doi:10.1039/C3GC42355G.

26. Banoub J, Delmas G-H, Joly N, Mackenzie G, Cachet N, Benjelloun-Mlayah B, et al. A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. J Mass Spectrom. 2015;50:5–48. doi:10.1002/jms.3541.

27. Owen BC, Haupert LJ, Jarrell TM, Marcum CL, Parsell TH, Abu-Omar MM, et al. High-performance liquid chromatography/high-resolution multiple stage tandem mass spectrometry using negative-ion-mode hydroxide-doped electrospray ionization for the characterization of lignin degradation products. Anal Chem. 2012;84: 6000–7. doi:10.1021/ac300762y.

28. Hughey CA, Hendrickson CL, Rodgers RP, Marshall AG, Qian K. Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal Chem. 2001;73:4676–81. doi:10.1021/ac010560w.

29. Kim S, Kramer RW, Hatcher PG. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. Anal Chem. 2003;75:5336–44. doi:10.1021/ac034415p.

30. Gougeon RD, Lucio M, Frommberger M, Peyron D, Chassagne D, Alexandre H, et al. The chemodiversity of wines can reveal a metabologeography expression of cooperage oak wood. Proc Natl Acad Sci U S A. 2009;106:9174–9. doi:10.1073/pnas.0901100106.

31. Bae E, Yeo IJ, Jeong B, Shin Y, Shin K-H, Kim S. Study of double bond equivalents and the numbers of carbon and oxygen atom distribution of dissolved organic matter with negative-mode FT-ICR MS. Anal Chem. 2011;83:4193–9. doi:10.1021/ac200464q.

32. Reichert E, Wintringer R, Volmer DA, Hempelmann R. Electrocatalytic oxidative cleavage of lignin in a protic ionic liquid. Phys Chem Chem Phys. 2012;14:5214–21. doi:10.1039/c2cp23596j.

33. Dier TKF, Egele K, Fossog V, Hempelmann R, Volmer DA. Enhanced mass defect filtering to simplify and classify complex mixtures of lignin degradation products. Anal Chem. 2016;88: 1328–35. doi:10.1021/acs.analchem.5b03790.

34. Caravatti P, Allemann M. The "infinity cell": a new trapped-ion cell with radiofrequency covered trapping electrodes for fourier transform ion cyclotron resonance mass spectrometry. Org Mass Spectrom. 1991;26:514–8. doi:10.1002/oms.1210260527.

35. Qi Y, O'Connor PB. Data processing in Fourier transform ion cyclotron resonance mass spectrometry. Mass Spectrom Rev. 2014;33:333–52. doi:10.1002/mas.21414.

36. Rodgers RP, McKenna AM. Petroleum analysis. Anal Chem. 2011;83:4665–87. doi:10.1021/ac201080e.

37. Kim S, Kaplan LA, Hatcher PG. Biodegradable dissolved organic matter in a temperate and a tropical stream determined from ultrahigh resolution mass spectrometry. Limnol Oceanogr. 2006;51: 1054–63. doi:10.4319/lo.2006.51.2.1054.

38. Headley JV, Peru KM, Barrow MP. Mass spectrometric characterization of naphthenic acids in environmental samples: a review. Mass Spectrom Rev. 2009;28:121–34. doi:10.1002/mas.20185.

39. Grinhut T, Hertkorn N, Schmitt-Kopplin P, Hadar Y, Chen Y. Mechanisms of humic acids degradation by white rot fungi explored using 1H NMR spectroscopy and FTICR mass spectrometry. Environ Sci Technol. 2011;45:2748–54. doi:10.1021/es1036139.

40. Kendrick E. A mass scale based on $CH_2 = 14.0000$ for high resolution mass spectrometry of organic compounds. Anal Chem. 1963;35:2146–54. doi:10.1021/ac60206a048.

41. Roach PJ, Laskin J, Laskin A. Higher-order mass defect analysis for mass spectra of complex organic mixtures. Anal Chem. 2011;83: 4924–9. doi:10.1021/ac200654j.

42. Kilgour DPA, Mackay CL, Langridge-Smith PRR, O'Connor PB. Appropriate degree of trust: deriving confidence metrics for automatic peak assignment in high-resolution mass spectrometry. Anal Chem. 2012;84:7431–5. doi:10.1021/ac301339d.

43. Koch BP, Dittmar T, Witt M, Kattner G. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. Anal Chem. 2007;79:1758–63. doi:10.1021/ac061949s.

44. Strohalm M, Kavan D, Novák P, Volný M, Havlícek V. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal Chem. 2010;82:4648–51. doi:10.1021/ac100818g.

Yulin Qi is a postdoctoral scientist with Professor Dietrich Volmer at the Institute for Bioanalytical Chemistry at Saarland University in Germany. His research interests focus on Fourier transform ion cyclotron resonance mass spectrometry and its application to biological and biomedical studies.

Rolf Hempelmann is a Professor of Physical Chemistry in the Department of Chemistry at Saarland University in Saarbrücken, Germany. His current research focuses on nanoparticles, ionic liquids, colloidal chemistry, electrochemistry, sustainable chemistry and didactics of chemistry.

trometry and metabolomics.

Dietrich A. Volmer is Krupp Professor, Chair in Analytical Chemistry, and Director of the Institute of Bioanalytical Chemistry at the Department of Chemistry of Saarland University in Saarbrücken, Germany. Dietrich Volmer has also been Editor of *Rapid Communications in Mass Spectrometry* since 2004. His current research themes include bioanalytical chemistry, mass spectrometry, gas phase ion chemistry, imaging mass spec-