

Elucidation of the complex molecular structure of wheat straw lignin polymer by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry

Joseph H. Banoub^{1*}, Bouchra Benjelloun-Mlayah², Fabio Ziarelli³, Nicolas Joly⁴ and Michel Delmas^{2,5}

¹Department of Chemistry, Memorial University of Newfoundland, and Fisheries and Oceans Canada, Science Branch, Special Projects, P.O. Box 5667, St John's, Newfoundland, A1C 5X1, Canada

²Compagnie Industrielle de la Matière Végétale (CIMV), 134 Rue Danton, 92300 Levallois-Perret, France

³Université P. Cézanne Aix-Marseille III, Faculté des Sciences et Techniques, Avenue de l'Escadrille Normandie-Niemen, 13397 Marseille, France

⁴Laboratoire de Physico-Chimie des Interfaces et Applications FRE CNRS 2485, Site de Béthune, Fédération Chevreul FR CNRS 2638, Université d'Artois, IUT de Béthune, Département Chimie, BP819, 62408 Béthune cedex, France

⁵École Nationale Supérieure des Ingénieurs en Arts Chimiques et Technologiques, Institut National Polytechnique de Toulouse, 118 route de Narbonne, 31077 Toulouse Cedex 4, France

Received 23 April 2007; Revised 21 June 2007; Accepted 26 June 2007

Wheat straw lignin was extracted using the novel CIMV procedure which selectively separates the cellulose, hemicelluloses and lignin. Solid-state ¹³C NMR experiments using cross polarization/magic angle spinning (CP/MAS) were carried out on the extracted wheat straw lignin and some structural indices were revealed. Atmospheric pressure photoionization mass spectrometry (APPI-MS) has proven to be a powerful analytical tool capable of ionizing small to large lignin oligomers, which cannot be ionized efficiently by atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI). The APPI mass spectra of the extracted wheat straw lignin were recorded in the positive and negative ion modes. Positive ion mode APPI-MS indicated the exact presence of 39 specific oligomeric ions. Negative ion APPI-MS indicated the additional presence of at least 18 specific oligomeric ions. The structural characterization of this novel and complete series of 57 specific related oligomers was achieved by calculating the exact molecular masses measured by high-resolution quadrupole time-of-flight mass spectrometry (QqToF-MS). Some oligomeric species photoionized in both the positive and negative ion modes to form the respective protonated and deprotonated molecules. Low-energy collision-induced dissociation tandem mass spectrometric analyses performed with a QqToF-MS/MS hybrid instrument provided unique dissociation patterns of the complete series of novel precursor ions. These MS/MS analyses provided diagnostic product ions, which enabled us to determine the exact molecular structures and arrangement of the selected 57 different related ionic species. Copyright © 2007 John Wiley & Sons, Ltd.

The cell walls of woods are mainly composed of cellulose, hemicelluloses, lignin and small amounts of proteins and inorganic components, all in different proportions, varying from hardwood to softwood species.^{1–3}

Cellulose, which is the most abundant natural polymer substance known, is composed of a β -D-(1 → 4)-glucopyranosyl disaccharide repeating unit forming a homogeneous polymer, with a degree of polymerization.⁴ The hemicelluloses

are a group of branched heteropolysaccharides, with an average degree of polymerization.⁵

The polymeric structure of lignin as described by Freudenberg was based on various phenylpropenyl alcohol monomeric units (also called lignols or phenylpropanoid units) attached through both ether and carbon–carbon bonds.⁶ More than a century of lignin research has produced an enormous amount of experimental results which are to a

*Correspondence to: J. H. Banoub, Department of Chemistry, Memorial University of Newfoundland and Fisheries and Oceans Canada, Science Branch, Special Projects, P.O. Box 5667, St John's, Newfoundland, A1C 5X1, Canada.

E-mail: banoubjo@dfo-mpo.gc.ca

Contract/grant sponsors: CIMV Company, Paris, France; Natural Sciences and Engineering Research Council of Canada.

large extent unrelated, and difficult to reproduce. Therefore, the chemistry of lignin has remained in a state of obscurity,^{6–10} despite the fact that lignin, after cellulose, is the second most abundant organic substance on our planet.⁹

In the paper industry, obtaining cellulose fibers from vegetable matter is accomplished by degrading large amounts of lignins and hemicelluloses by strong acid treatment, making them soluble in aqueous media.^{10,11}

Numerous studies with radioactive carbon have confirmed that three cinnamyl alcohols, namely, *p*-hydroxycinnamyl alcohol [or cumaryl alcohol (H)], coniferyl alcohol [or guaiacyl alcohol (G)], and sinapyl alcohol [or syringyl alcohol (S)], are the primary precursors and building units of all lignins (Scheme 1).^{6,16} It has been established that the monomer lignol precursor in conifers is predominantly coniferyl alcohol, whereas in deciduous trees it is sinapyl alcohol.^{12–14}

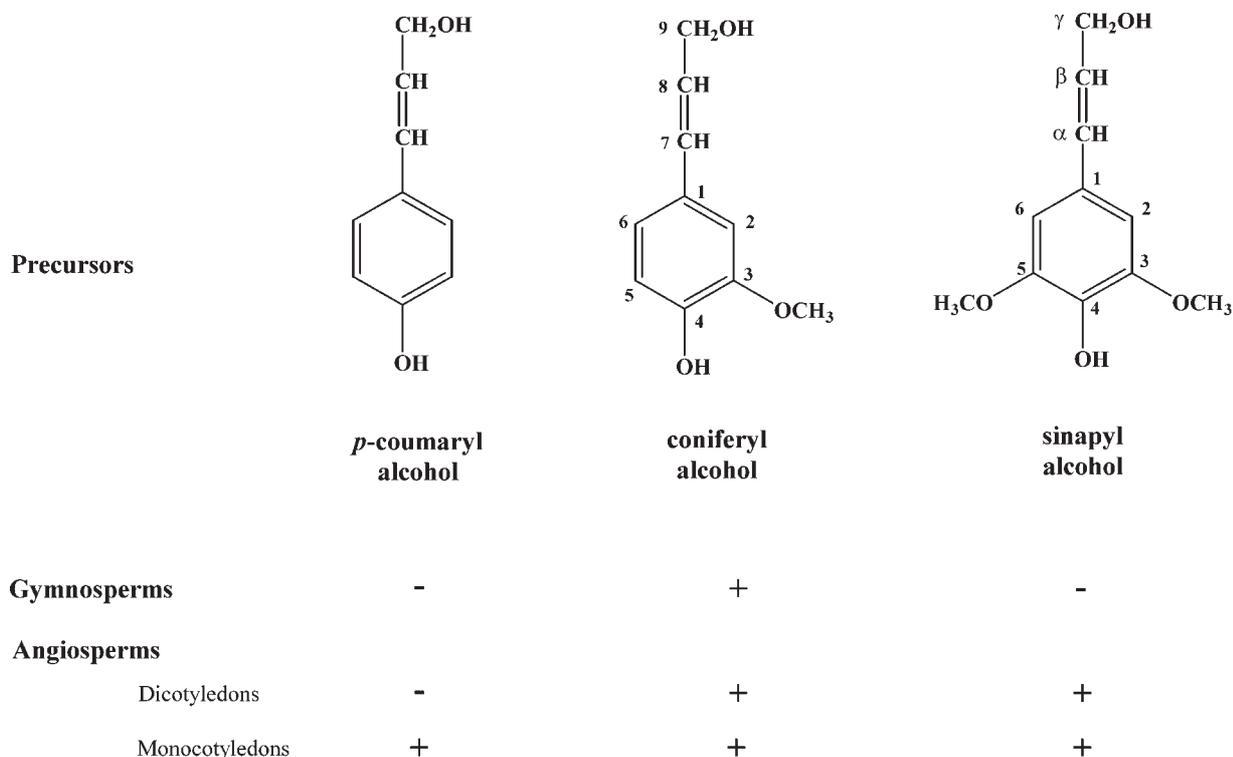
It was suggested by Freudenberg that lignin biosynthesis was initiated by an enzyme-catalyzed phenol dehydrogenation of mixtures of the phenylpropenyl alcohols. This was followed by a copolymerization of the resonance-stabilized phenoxy radicals, which were obtained from the monomer lignol, to form the growing lignin macromolecule.¹⁵

The building up of the lignin macromolecule by plants is accomplished through complicated biological and biochemical processes, which are believed to occur through random reactions without any ordered genetic control.^{12–14}

Novel biochemical studies have involved either the identification of new alternative pathways in lignin biosyn-

thesis, and/or attempts to genetically manipulate lignin structures based on these proposed pathways.^{11,17–19} The prevailing view of the biosynthesis of lignins, that they were produced by a random free-radical polymerization and coupling, has been challenged and a new model of polymerization of monolignols by redox-shuttle-mediated enzyme oxidation has been proposed.²⁰ Recently, the first understanding of regiochemical and stereochemical control of monolignol coupling in lignin biosynthesis was established, involving the participation of a new class of enzymes called the dirigent proteins.²¹ Although this latter model represents a significant departure from accepted views on the biogenesis of lignin, it provides a plausible rationale for patterns of structure distribution in native lignins. A more recent hypothesis suggests that the intracellular regulation of the structure of lignins occurs through the provision of templates for their spatial organization, prior to the development of covalent linkages during radical coupling reactions.²² The most plausible templates are the domain of the polysaccharide matrix and their interior surfaces formed through the interactions between cellulose and the hemicelluloses.^{22,23}

Although many chemical, mechanical and enzymatic methods have been developed for the isolation of lignin from vegetable fibers, none are without the risk of causing structural changes during isolation. These methods can either fundamentally change the native structure of lignin to afford modified lignins, or release small parts of relatively unchanged short oligomers, which are termed lignans.^{24–26} It is accepted that these methods produce fragments which



Scheme 1. The primary structures of the constituent lignol precursors (the '+' sign means that the lignol is present, whereas the '-' sign means the lignol is absent in the studied vegetable matter).

'hypothetically' reflect the structural sequence of the original polymer. Unfortunately, this type of analysis is extremely labor-intensive and, regrettably, gives no clear answer as to the molecular makeup of the original polymer.^{24–26} For this reason, the chemical structure of the lignin macromolecule remains unsolved.

Attempts to determine the molecular weights of lignins are still generally carried out by size-exclusion chromatography (SEC) despite the fact that their absolute molar mass values are difficult to determine because of the strong tendency of lignins to form aggregates in solution.²⁷

Mass spectrometry, using electron ionization and chemical ionization, has been used for the study and characterization of the derivatized lignol monomer constituents released by either reductive cleavage or by pyrolysis. In this case, only the monomeric and, to a small extent, dimeric products were identified by comparison of gas chromatographic retention times and mass spectra using authentic samples.^{28–31}

The last two decades have witnessed a systematic effort using novel soft MS ionization techniques to completely unravel the architecture of lignin molecules. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF-MS) has been used to determine the molecular mass of the lignin obtained from birch *Betula alba*.³² Electrospray ionization mass spectrometry (ESI-MS) was also used for the structural identification of a series of dioxin-extracted lignins and the *Eucalyptus globulus* lignins, and for the study of a new *in vitro* biosynthesis of lignin.^{32–34}

Recently, we have reported on the partial structural characterization of wheat straw lignin using atmospheric pressure chemical ionization (APCI) tandem mass spectrometry (MS/MS) and MALDI-ToF-MS.³⁵

The use of atmospheric pressure photoionization (APPI) has grown rapidly over the last five years and it is now an important analytical tool for the practising mass spectrometrists.^{36–38} The popularity of APPI is due to its success in allowing the ionization of non-polar compounds, weak acids, and halogenated organic compounds, which are not readily ionized by ESI- and MALDI-MS. APPI leads to cleaner spectra and is less susceptible to ion suppression.^{36–38} Unlike ESI and APCI, photoionization is not based on charge affinity, which makes it the ideal tool for ionizing non-polar compounds. The principal mechanism for the photoionization of a molecule M is photon absorption and electron ejection to form the molecular ion $M^{+\bullet}$. Because the photon energy typically is just above the ionization energy, the molecular radical ion is less susceptible to dissociation. In the presence of protic solvents the molecular radical cation $M^{+\bullet}$ can attract a hydrogen atom during gas-phase ionization to produce the protonated molecule $[M+H]^+$ if M has a high proton affinity.^{36–38} APPI-MS makes it possible to analyze complex mixtures by direct injection into the source, without resorting to liquid chromatography (LC) to separate the components.^{36–38}

We report herein the complete, precise and complex molecular structural identification of wheat straw lignin polymer by means of APPI-MS and low-energy collision-induced dissociation (CID)-APPI-MS/MS using a QqToF-MS/MS hybrid instrument.

EXPERIMENTAL

Sample preparation

The wheat straw lignin was extracted using the novel CIMV procedure which selectively separates the cellulose, hemicellulose and lignin, and allows the destructuring of the vegetable matter at atmospheric pressure by a catalyst-solvent system of formic acid/acetic acid/water to produce, after precipitation and dissolution in dioxan, a white lignin.^{39,40} Approximately 0.1 mg of the purified lignin was dissolved in either 1 mL dioxan or 1 mL dioxan/methanol/chloroform (1:1:1) for MS analysis.

APPI-MS

Mass spectrometry was performed using an Applied Biosystems (Foster City, CA, USA) API QSTAR XL MS/MS quadrupole orthogonal time-of-flight (QqToF)-MS/MS hybrid instrument. APPI was performed with a PhotoSpray ion source (Applied Biosystems) operated at 1300 V at a temperature of 400°C, with all acquisitions performed in the positive ion mode. Samples were infused into the mass spectrometer with an integrated Harvard syringe pump at a rate of 0.1 mL/min. The auxiliary nebulizer gas pressure setting was fixed at 25 psi, and the nebulizer gas pressure at 74 psi. The curtain gas pressure was set at 30 psi. The declustering potential (DP) was set at +100 eV. The focus potential (FP) was adjusted to +100 V. Toluene was selected as the dopant for its ability to undergo trouble-free photoionization at 8.83 eV. The eluent was composed of methanol/chloroform (1:1). No modifier was used to enhance ion production.

The mass calibration of the ToF analyzer in the positive ion mode was performed with the PhotoSpray ion source, using 1,2,3,5-tetra-O-acetyl- β -D-ribofuranose and checking for the exact masses of the $[M+H]^+$ ion $[C_{13}H_{19}O_9]$ at m/z 320.1107 and the $[M+H-AcOH]^+$ ion $[C_{12}H_{15}O_7]$ at m/z 271.0808. Calibration for higher masses was performed with hexa-O-acetyl- β -D-lactopyranose and checking for the $[M+H]^+$ ion $[C_{28}H_{37}O_{19}]$ at m/z 677.1929. Calibration of the ToF analyzer in the negative ion mode was performed with the PhotoSpray ion source, using β -D-lactopyranose and checking for the exact mass of the $[M-H]^-$ ion $[C_{12}H_{11}O_{11}]$ at m/z 331.0302.

The APPI mass spectra were also recorded with higher DP values varying from 120 to 150 V.

Low-energy CID-MS/MS

Product ion spectra were obtained on the same instrument as described above. Nitrogen was used as the collision gas for MS/MS analyses with collision energies varying between 10 and 35 eV. Collision energy (CE) and CID gas conditions were adjusted in each acquisition such that the precursor ion remained abundant in the product ion spectra. We have used in general a 1 m/z unit resolution for the MS/MS selection of the precursor ion for simplification of the analysis.

In addition, re-confirmation of the various established fragmentation routes was effected by conducting a series of APPI in-source collision-induced experiments (APPI-CID-QqToF-MS/MS) on the same Applied Biosystems API QSTAR XL instrument. Fragmentation in the atmospheric

pressure/vacuum interface was achieved by ramping the DP from 102 to 150 V, until the characteristic product ions were observed in the spectra. These ions were then selected as precursors for the subsequent MS/MS acquisitions. Fragmentation conditions were the same as in the MS/MS experiments described above.

Solid-state NMR spectroscopy

Solid-state ^{13}C NMR experiments were performed on a Bruker AVANCE-400 Wide Bore NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany) operating at a ^{13}C resonance frequency of 106 MHz and using a commercial solid probe H/X cross polarization/magic angle spinning (CP/MAS) instrument (Bruker BioSpin). About 90 mg of sample was placed in a zirconium dioxide rotor with an outer diameter of 4 mm which was spun at 10 kHz at the magic angle.^{40–42} The cross polarization (CP) technique was applied with a ramped ^1H -pulse starting at 100% power and decreasing to 50% power. This procedure was used during the contact time of 2 ms, in order to circumvent the Hartmann-Hahn mismatches.⁴⁴ The experiments were performed at ambient temperature and 16K scans were accumulated using a delay of 2 s. The ^{13}C chemical shifts were referenced to tetramethylsilane and calibrated with the glycine carbonyl signal, set at 176.5 ppm.^{41–43}

RESULTS AND DISCUSSION

Solid state ^{13}C NMR of the wheat straw lignin using the CP/MAS method

Extensive 1D and 2D ^1H and ^{13}C NMR experiments have been carried out in the liquid state on dimeric and oligomeric lignans and, despite the rather random and heterogeneous nature of native lignins and/or synthetic lignins, some of their characteristic structural indices have been revealed by these spectra.^{43–46} It is known that isolated lignins from wood are normally insoluble; however, there has been a recent trend to dissolve them in dimethyl sulphoxide with *N*-imidazole.⁴⁴

NMR practitioners know that increasing molecular weights broaden resonances due to reduced relaxation times; therefore, the ^{13}C NMR spectra of lignins appear broader and more featureless than, for example, those of proteins or complex carbohydrates. Solid-state ^{13}C NMR with CP/MAS has proven invaluable in the study of complex organic solids. This approach renders the question of solubility irrelevant and eliminates the structural uncertainties associated with dissolution.^{43,44} Figure 1(A) shows the high-resolution solid-state ^{13}C NMR spectrum of the native wheat straw lignin acquired using the CP/MAS method.^{41–44} The tentative assignments of the major broad resonances were based on the program NMRPredict⁴⁷ version 3.2.2 and on lignin resonances reported in the literature.^{44–47} It is logical to infer that each reported resonance is the normal mathematical addition of numerous signals that coalesce to form such broad signals. Therefore, we assigned the characteristic signal at 55.6 ppm to the methoxyl groups of the lignin. The region between 125 and 160 ppm was assigned specifically to the aromatic carbons of coniferyl

units of lignin. The signal at 153.1 ppm was assigned to the cyclic C-5', C-3 and C-5 (of the first residue) of the coniferyl constituents containing a carboxyl group at C-4. The resonance at 147.5 ppm was assigned to C-3' and to the C-4' of the constituent coniferyl (G) units attached to the ether oxygen of the furan-like rings. The signal at 132.9 ppm was assigned to the C-1 and C-6 of coniferyl units. The signal at 114.7 ppm was attributed to the C-7 (also called the α -C carbon), whereas the signal at 104.3 ppm was attributed to C-8 (also called the β -C carbon) of the coniferyl unit. The resonances at 73.7 and 66.8 ppm were assigned to the γ -carbons (C-9 and C-9') bearing a primary hydroxyl group.^{44–46} The signal at 182.8 ppm was assigned to aliphatic aldehydes and aromatic ketones. The signal at 171.0 ppm was attributed to an aromatic carboxyl group. Finally, the resonance at 161.8 ppm was attributed to a carboxyl group present in an α -position to an aliphatic double bond.^{44–47}

The deconvoluted solid-state CP/MAS ^{13}C NMR spectrum of the native wheat straw lignin between 100 and 190 ppm is shown in Fig. 1(B). The deconvolution of this spectrum allowed us to calculate the individual integration of each distinct ^{13}C resonance. This analysis revealed that the ratio between the total integration of the resonance signal at 153.1 and that at 147.5 ppm displayed absolute integration values of 7521607 and 2971757, representing a ratio of 147.5/153.1 to 2.5.^{44,47} This ratio indicates that we have approximately 2.5 times more polycondensed phenylcoumaran units containing the coniferyl furan-like oligomer (also called the β -C-5' units) than the coniferyl units that do not carry the C-4 carboxyl group; and do not contain the cyclic furan-like ring; that is normal open-chain coniferyl oligomers (also called the α -O-4' units).

Positive mode APPI-MS of the wheat straw lignin

In our earlier study, the partial structure of the constituent dimer of the wheat straw lignin was unambiguously assigned by measuring the negative and the positive ion APCI mass spectra and conducting low-energy CID-MS/MS experiments using a conventional quadrupole-hexapole-quadrupole tandem mass spectrometer.³⁵ The protonated molecule of the phenylcoumaran derivative $\text{C}_{18}\text{H}_{19}\text{O}_6$ 2 ($[\text{M}+\text{H}]^+$ at m/z 331.11) was determined to be the protonated 4-carboxyl-(7 \rightarrow 4')-coniferyl ether-(8 \rightarrow 5')-(3'-methoxybenzene) unit, hence forming the cyclic ether dimer constituent unit of all the phenylcoumaran derivatives reported herein (see Scheme 2). Please note that the C-7 and C-8 positions are also known as the α - and β -positions.^{11,12,17} We also characterized the protonated trimer $\text{C}_{29}\text{H}_{33}\text{O}_8$ ($[\text{M}+\text{H}]^+$ at m/z 509.11), containing the (8 \rightarrow 5')-(3'-methoxycoumaran) unit.³⁵ Attempts to identify higher oligomeric lignin fragments failed, even when a scan range of m/z 600–1200 was used. It is noteworthy that identical results were obtained using ESI-MS, thus indicating the limitations of both APCI and ESI with this type of polymer.³⁵ We assumed that the native lignin polymer was extremely chemically reactive and, hence, degraded/reacted to yield smaller oligomers during acidic gas-phase ionization.^{14,35}

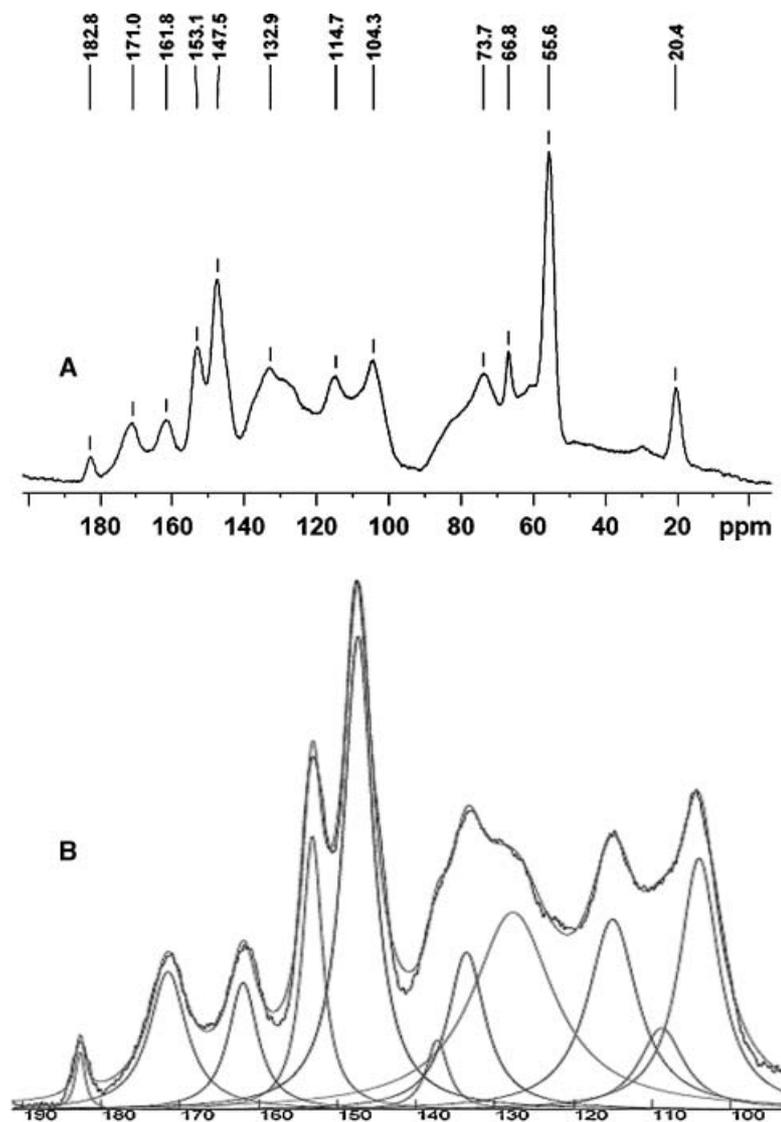


Figure 1. (A) Solid-state CP/MAS ^{13}C NMR spectrum of the extracted wheat straw lignin. (B) Deconvoluted solid-state ^{13}C NMR allowing the calculation of the integration of the ^{13}C resonances (inset of 100–190 ppm).

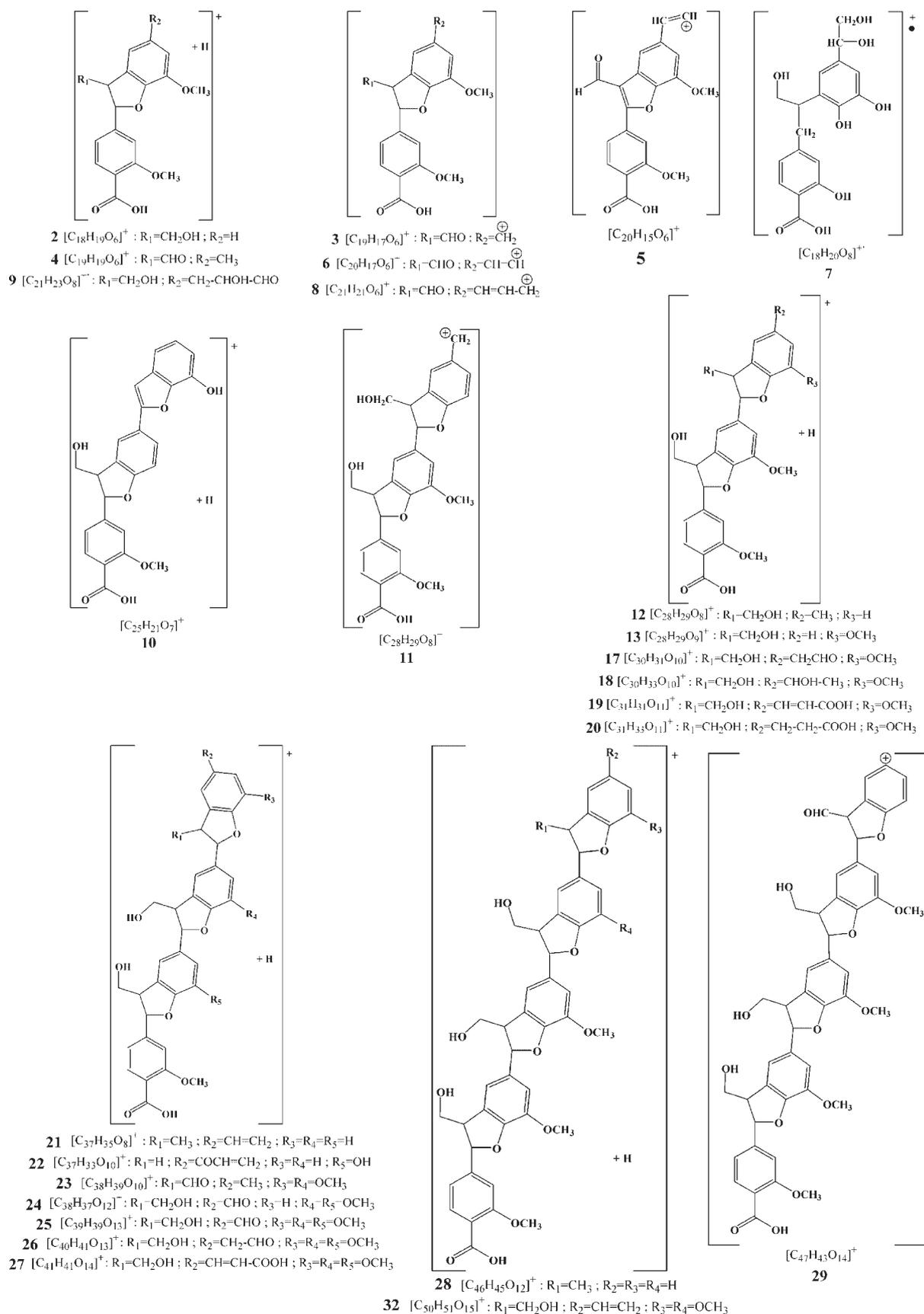
It was recently reported that APPI-MS produces a molecular radical ion $[\text{M}]^{+\bullet}$ or the protonated molecule $[\text{M}+\text{H}]^+$, with sufficiently high mass resolving power to directly identify the elemental composition (chemical formula) of virtually any organic ion.^{36–38} The characteristic ions in the APPI mass spectra of the extract from wheat straw lignin, recorded in the positive ion mode with scan ranges of m/z 300–700 and 500–900, and a DP of 100 V, are presented in Table 1. The table demonstrates a high number of chemically distinct ions, indicating that the lignin polymer is, indisputably, a very complex heterogeneous mixture. For such a composition, the analytical chemist will try to separate it by conventional wet chemical separation methods (chromatography, crystallization, etc.); unfortunately, to no avail in this case.^{14,24–26,31,36} Our original attempts, using a variety of high-performance liquid chromatography (HPLC) separation methods with different columns and solvent systems, failed and afforded only traces of the ion at m/z 331 which seem to survive such treatment. This major inadequacy in

structural analysis probably results from the extreme reactive chemical nature and instability of the constituent species forming this native polymer.^{14,24–26,31,32,35} The APPI mass spectra indicated the presence of a plethora of ions, from which we were capable of characterizing 39 specific oligomeric ions. It is crucial to mention that the protonated molecule at m/z 331.11 was the most abundant ion when the mass spectrum was recorded over a mass range of m/z 100–900. The structural characterization of the exact molecular masses was supported by the measured masses obtained by high-resolution QqToF-MS (see Table 1).

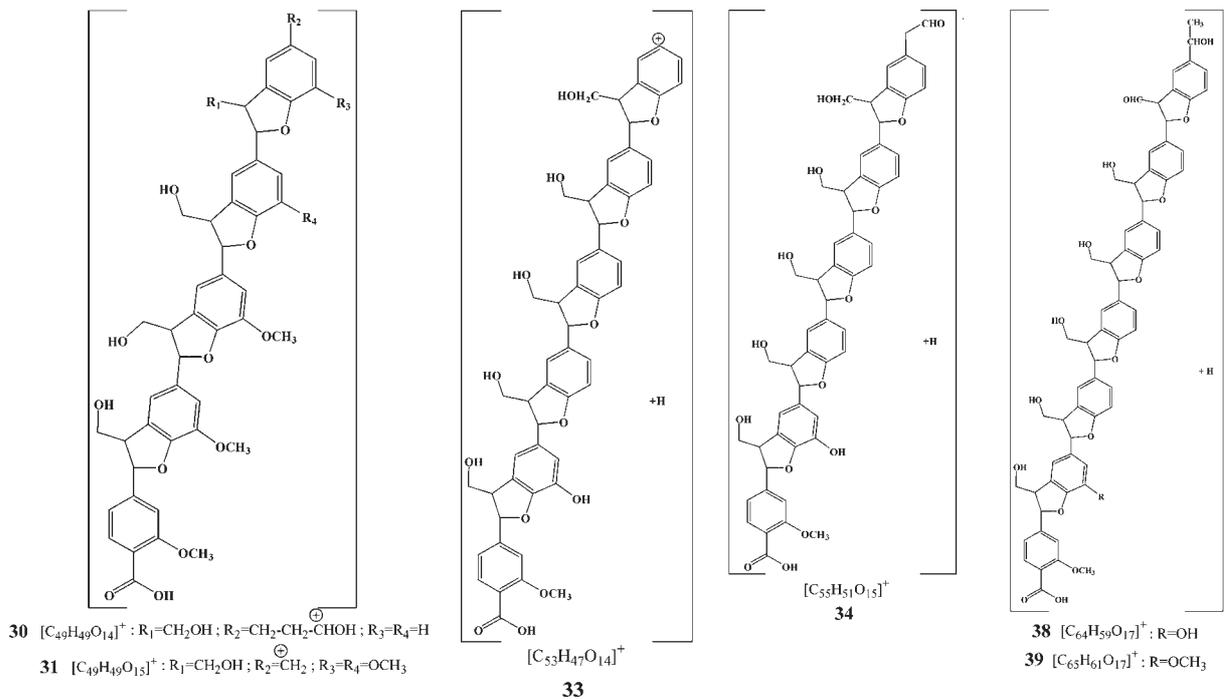
Positive mode APPI-QqToF-CID-MS/MS analyses of the specific oligomeric ions obtained from wheat straw lignin

The separation of this series of complex oligomers was successfully achieved by low-energy CID-MS/MS on the selected ions, using the QqToF-MS/MS hybrid instrument. It

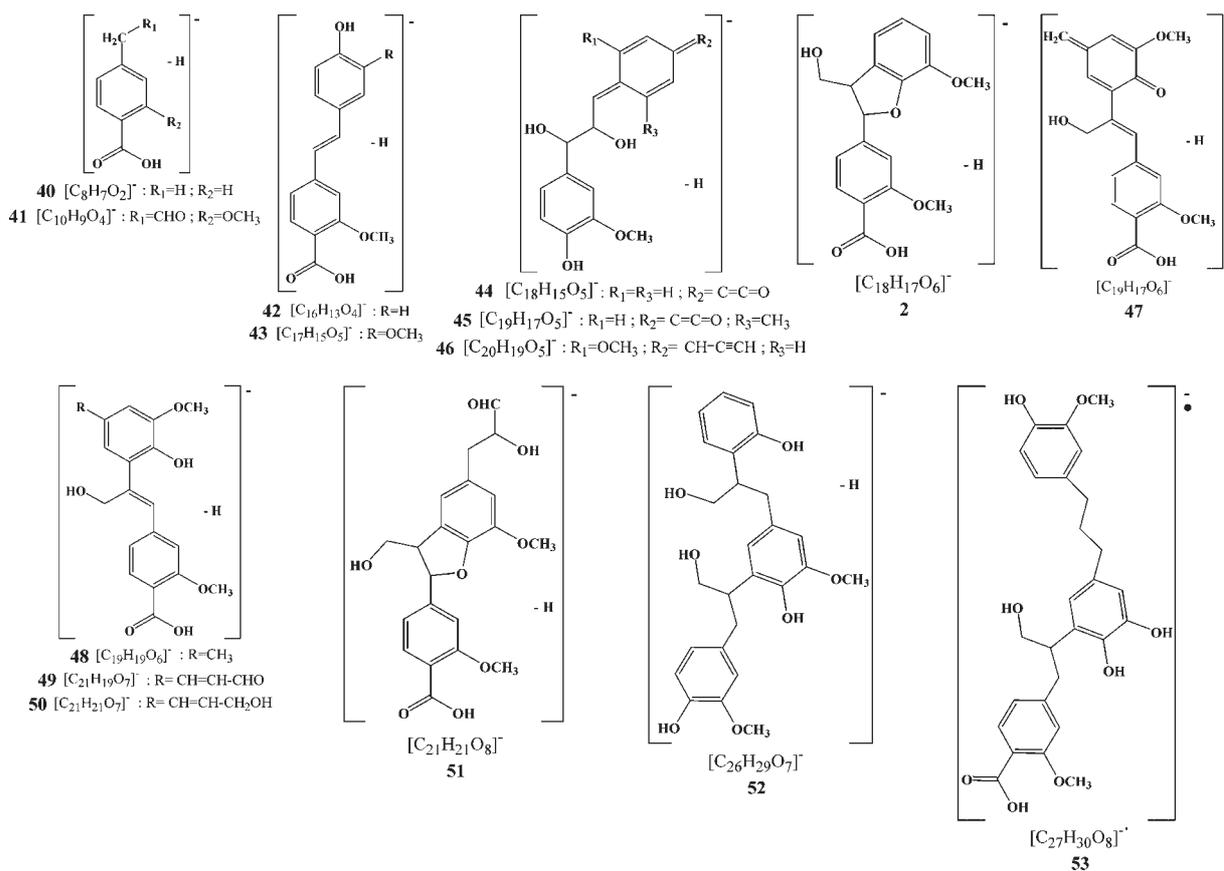
Positive oligomer ions



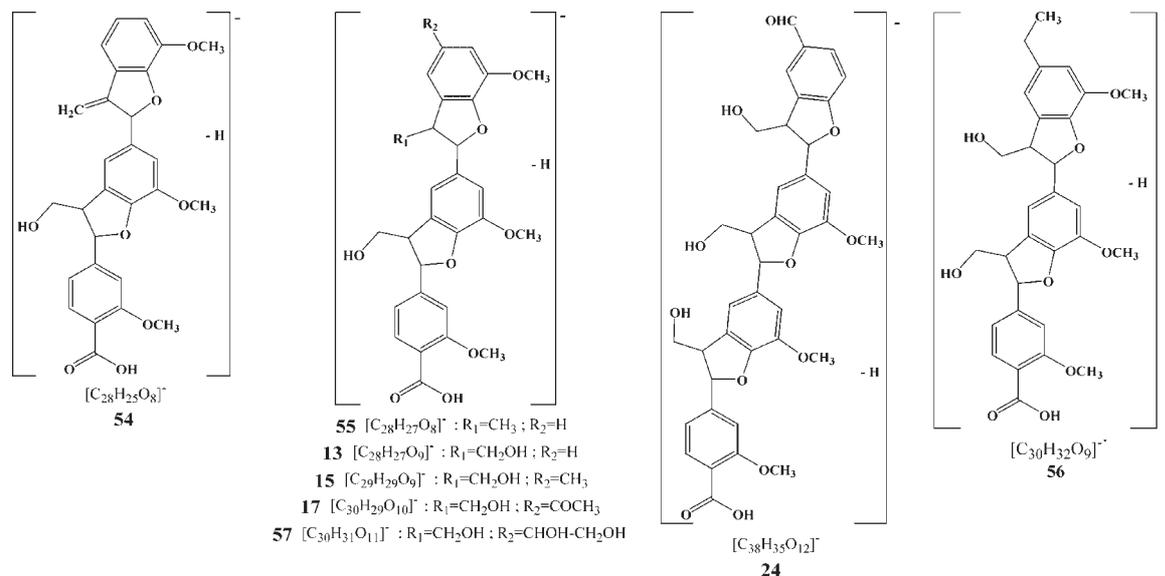
Scheme 2. Tentative structures of the 57 different ions isolated obtained from the wheat straw lignin, recorded by positive and negative ion APPI-QqToF-MS.



Negative oligomer ions



Scheme 2. (Continued).



Scheme 2. (Continued).

Table 1. Characteristic ions obtained from the APPI-MS analysis of the wheat straw lignin recorded in the positive ion mode scanning m/z 300–700 with a DP = 100 V

Characteristic ions	Cpd #	Calculated m/z	Observed m/z	Abundance %	Difference (ppm)
$[C_{19}H_{17}O_4]^+$	1	309.1127	309.1100	7.69	8.7
$[C_{18}H_{19}O_6]^+$	2	331.1182	331.1172	84.60	3.0
$[C_{19}H_{17}O_6]^+$	3	341.1025	341.0999	63.50	7.6
$[C_{19}H_{19}O_6]^+$	4	343.1182	343.1166	100.00 ^a	4.6
$[C_{20}H_{15}O_6]^+$	5	351.0869	351.0859	9.87 ^a	2.8
$[C_{20}H_{17}O_6]^+$	6	353.1025	353.1019	9.21 ^a	1.7
$[C_{18}H_{20}O_8]^+$	7	364.1158	364.1149	62.44	2.4
$[C_{21}H_{21}O_6]^+$	8	369.1338	369.1329	100.00	2.4
$[C_{21}H_{23}O_8]^+$	9	403.1393	403.1363	50.06	7.4
$[C_{25}H_{21}O_7]^+$	10	433.1287	433.1282	17.38	-1.1
$[C_{28}H_{27}O_8]^+$	11	491.1706	491.1709	63.33 ^a	-0.6
$[C_{28}H_{29}O_8]^+$	12	493.1862	493.1861	18.18 ^a	0.2
$[C_{28}H_{29}O_9]^+$	13	509.1890	509.1881	68.18 ^a	1.8
$[C_{29}H_{29}O_9]^+$	14	521.1812	521.1811	16.66 ^a	0.2
$[C_{29}H_{31}O_9]^+$	15	523.1968	523.1967	22.11	0.2
$[C_{29}H_{31}O_{10}]^+$	16	539.1917	539.1916	13.63	0.2
$[C_{30}H_{31}O_{10}]^+$	17	551.1917	551.1909	22.72 ^a	1.5
$[C_{30}H_{33}O_{10}]^+$	18	553.2074	553.2068	4.91	1.1
$[C_{31}H_{31}O_{11}]^+$	19	579.1866	579.1859	9.16	1.2
$[C_{31}H_{33}O_{11}]^+$	20	581.2023	581.2022	4.62 ^a	0.2
$[C_{37}H_{35}O_8]^+$	21	607.2332	607.2322	12.50	1.6
$[C_{37}H_{33}O_{10}]^+$	22	637.2074	637.2066	9.86	1.2
$[C_{38}H_{39}O_{10}]^+$	23	655.2543	655.2418	5.92	1.9
$[C_{38}H_{37}O_{12}]^+$	24	685.2285	685.2271	9.25	2.0
$[C_{39}H_{39}O_{13}]^+$	25	715.2547	715.2545	24.23 ^b	0.3
$[C_{40}H_{41}O_{13}]^+$	26	729.2848	729.2847	26.32 ^b	0.1
$[C_{41}H_{41}O_{14}]^+$	27	757.2496	757.2495	9.42 ^b	0.1
$[C_{46}H_{45}O_{12}]^+$	28	789.2911	789.2898	16.45 ^b	1.6
$[C_{47}H_{43}O_{14}]^+$	29	831.2653	831.2651	22.41 ^b	0.2
$[C_{49}H_{49}O_{14}]^+$	30	861.3122	861.3120	16.92 ^b	0.2
$[C_{49}H_{49}O_{15}]^+$	31	877.3071	877.3059	18.92 ^b	1.4
$[C_{50}H_{51}O_{15}]^+$	32	891.3228	891.3226	11.14 ^b	0.2
$[C_{53}H_{47}O_{14}]^+$	33	907.2966	907.2965	10.12 ^b	0.1
$[C_{55}H_{51}O_{15}]^+$	34	951.3228	951.3226	10.11 ^b	0.2
$[C_{57}H_{49}O_{15}]^+$	35	973.3071	973.3069	12.11 ^b	2.0
$[C_{59}H_{52}O_{14}]^+$	36	984.3357	984.3355	15.16 ^b	0.2
$[C_{64}H_{59}O_{16}]^+$	37	1083.3803	1083.3800	9.98 ^b	0.3
$[C_{64}H_{59}O_{17}]^+$	38	1099.3752	1099.3749	8.12 ^b	0.3
$[C_{65}H_{61}O_{17}]^+$	39	1113.3909	1113.3907	10.17 ^b	0.2

^a Recorded with a DP = 120V.^b Recorded with a mass range of m/z 700–1200.

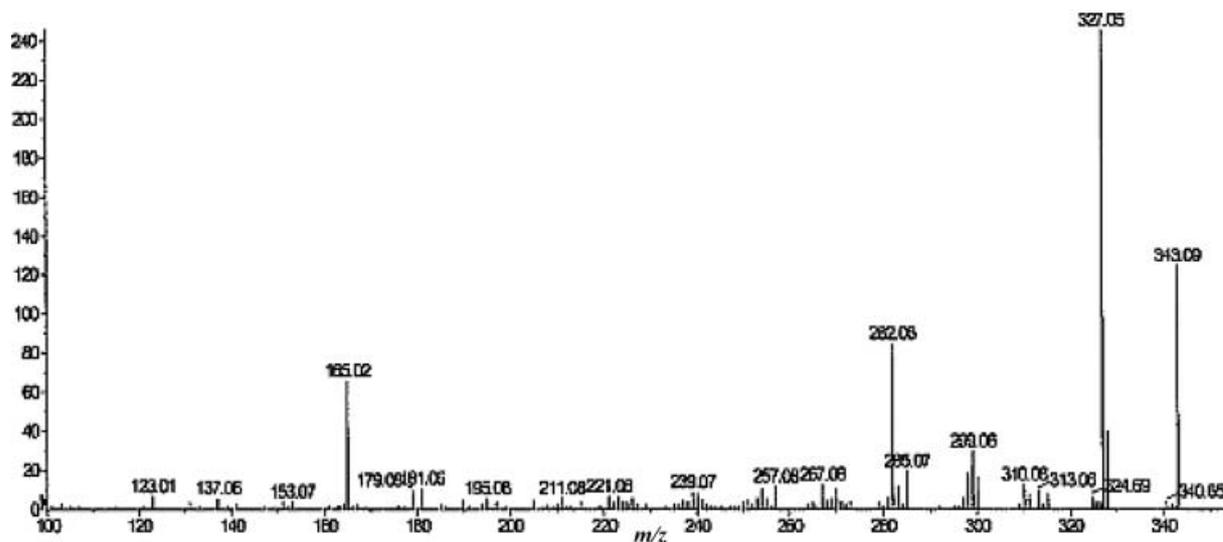


Figure 2. APPI-CID-QqToF-MS/MS of the protonated phenylcoumaran derivative **4** at m/z 343.09.

is well established that novel structural information, such as the sequences of the oligomeric units of the native polymer of wheat lignin, can be obtained by MS/MS. These MS/MS analyses will help reveal the precise structures and arrangement of the selected ions. Consequently, in the present MS/MS lignin analyses, the obtained unique dissociation patterns provided different related diagnostic product ions, which in all cases were those of related substructures.

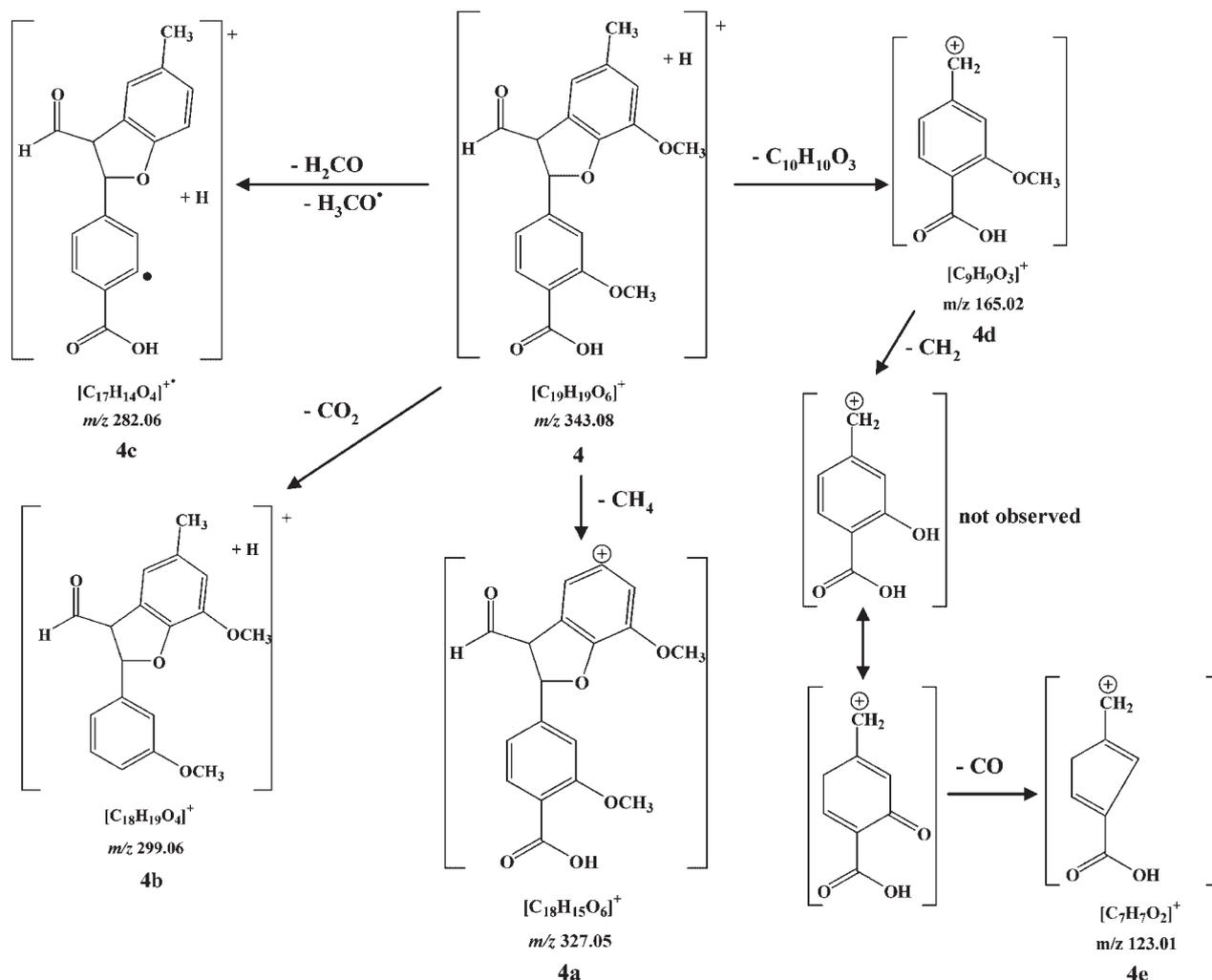
In Scheme 2, we tentatively reveal the molecular structures of the most characteristic different compounds, isolated from the 39 specific ions obtained. These ions correspond to a series of related oligomers, obtained from the photoionized wheat straw lignin (note that not all the structures are represented). Please also observe that the majority of reported structures in this rationale are formed from assorted *hydrocoumaran* structural unit combinations. The coding of the compounds in Scheme 2 is based on the assigned numbers **1** to **39**. All the ions analyzed were protonated molecules, except for the dimers **7** and **9**, which were identified as positive radical ions and the non-protonated cations **3**, **5**, **8**, **11**, **30**, **31** and **33**.

The low-energy APPI-QqToF-MS/MS analysis of the protonated molecule $[M+H]^+$ of the 6-aldehydro-4-carboxyl-(7 \rightarrow 4')-ether-3-methoxylconiferyl-(8 \rightarrow 5')-(1'-methyl-3'-methoxyltoluene) unit **4** (or phenylcoumaran derivative) at m/z 343.09 was carried out and is shown in Fig. 2. This precursor protonated molecule was chosen to introduce, in this study, the concept of the presence of the five-membered intermediate furan-like ring of the coumaran unit, formed by the (8 \rightarrow 5')-covalent bond and the C-7–C-4' ether linkage, between the contiguous first and second coniferyl units. The product ion scan of m/z 343.09 $[C_{19}H_{19}O_6]^+$ afforded the product ion **4a** at m/z 327.05 by elimination of a molecule of methane. The precursor ion also lost a molecule of carbon dioxide to afford the product ion **4b** at m/z 299.06. The radical product ion **4c** at m/z 282.06 can be created from the precursor ion, by the consecutive losses of formaldehyde and a methoxyl radical. Finally, the product ion **4d** at m/z 165.02

was produced from the precursor ion following the consecutive cleavages of the C-7–C-8 covalent bond of the coniferyl aldehyde, and the ether link formed between C-7 and the oxygen atom located on C-4' of the second coniferyl unit (see Scheme 3). This latter product ion then generates the ion at m/z 123.01 by the consecutive losses of a methylene radical, followed by the loss of a carbonyl group involving a ring contraction. The elimination of either a methylene radical or perhaps a methylene molecule containing an sp^2 carbon, although surprising, is not unprecedented as we have previously reported such an elimination, and have confirmed it, by performing the precursor ion analysis of the intermediate product ion obtained, during the elucidation of the molecular structure of lipid A isolated from a Gram-negative bacterium.⁴⁸

We have chosen, as the next example, the APPI-QqToF-CID-MS/MS of the open, demethylated, dimeric, radical cation $[C_{18}H_{20}O_8]^+$ **7** at m/z 364.23 to indicate the fragmentation pattern of an acyclic compound (i.e. the α -O-4', which does not contain the furan-like cyclic unit), which can be viewed as a precursor of the constituent β -C-5' cyclic dimer of lignin (see Fig. 3). It can also be formed during the extraction process, by a demethylation reaction, followed by the unzipping of the β -C-5' cyclic dimer constituent of lignin. The product ion scan of this radical ion **7** at m/z 364.23 gave a series of product ions, as shown in Fig. 3. The product ion **7a** at m/z 347.23 is formed by loss of a hydroxyl radical. Decarboxylation of the precursor radical ion produces the radical ion **7b** at m/z 320.24. Ion **7** also eliminates an ethylene diol radical (61 Da) to afford the product ion **7c** at m/z 303.21, which can then lose carbon dioxide to produce the ion **7d** at m/z 275.22. The **7c** ion may eliminate either one or two molecules of water to form the ions **7e** and **7g** at m/z 285.20 and 267.22, respectively. The **7e** ion eliminates a carbonyl group followed by ring contraction, to produce the product ion **7f** at m/z 257.22 (see Scheme 4).

The product ion scan of a complete protonated radical-dimeric unit $[C_{21}H_{23}O_8]^+$, the 4-carboxylconiferyl-(8 \rightarrow 5')-[9-hydroxy(coniferyl aldehyde)] **9**, is shown in Fig. 4. The



protonated molecule at m/z 403.12 loses a methyl radical, to produce the product ion **9a** at m/z 388.10. The protonated molecule may also eliminate either carbon monoxide or formaldehyde to give the product ions **9b** and **9c** at m/z 375.14

and 373.13, respectively. Cleavage of the aliphatic chain of the $C-7'-C-8'$ covalent bond of the precursor ion affords the ion **9d** at m/z 343.08. A homologous cleavage of the $C-1'-C-7'$ covalent bond affords the partial dimeric unit **9e** at m/z

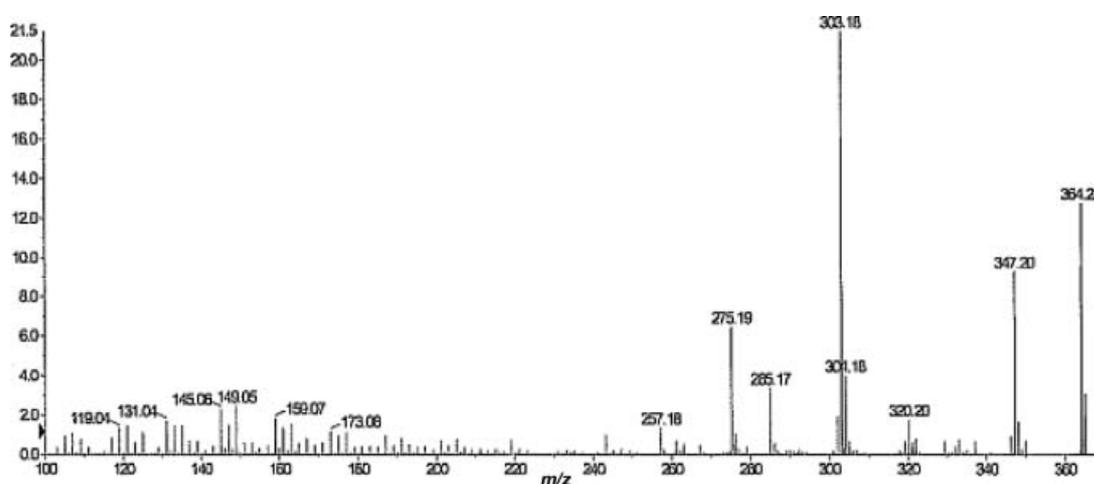
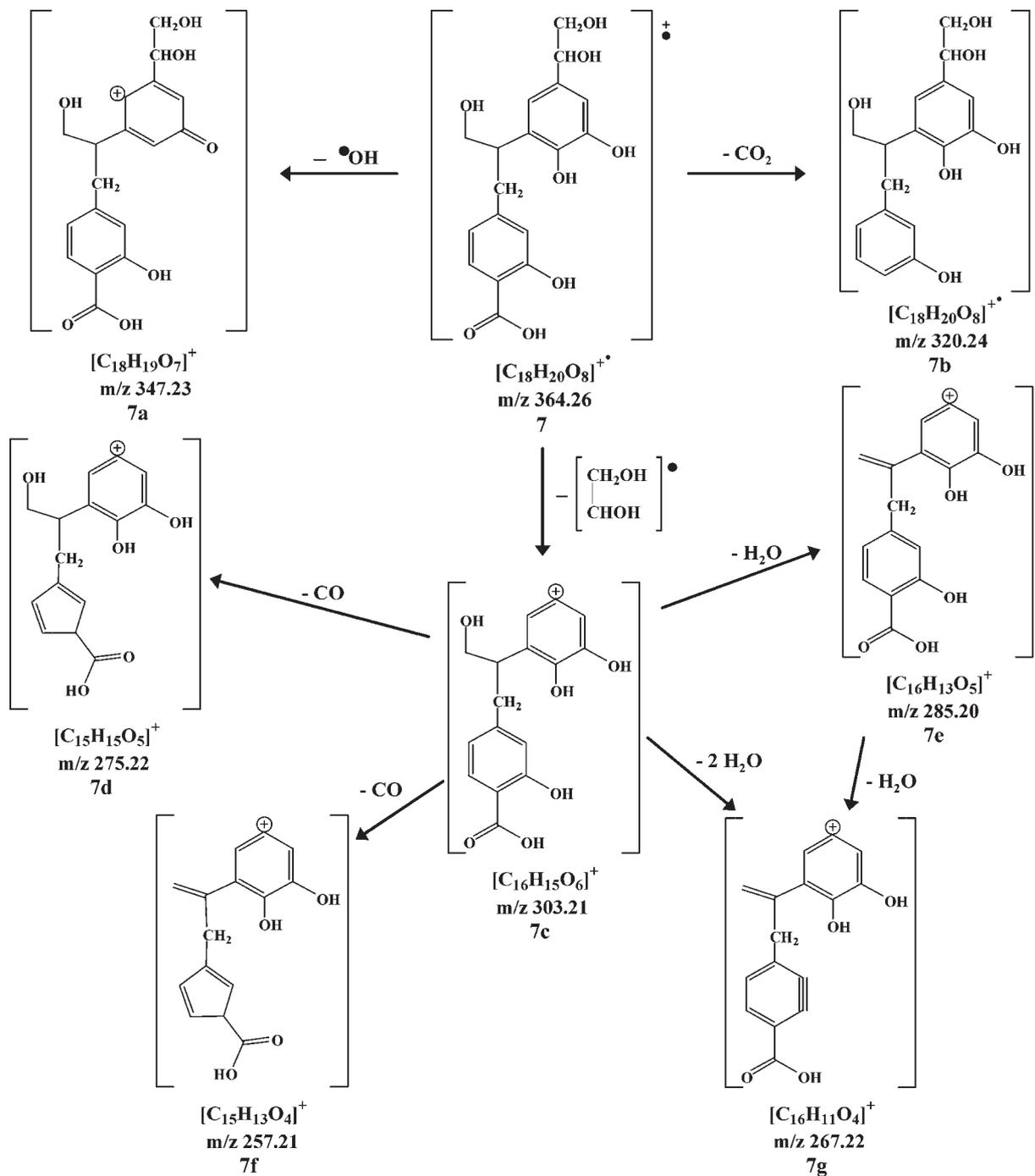


Figure 3. APPI-CID-QqToF-MS/MS of the open, demethylated, dimeric, radical cation $[C_{18}H_{20}O_8]^{\bullet+}$ **7** at m/z 364.23.



Scheme 4. Tentative breakdown processes in the MS/MS of the demethylated, dimeric-radical cation $[C_{18}H_{20}O_8]^{\bullet+}$ **7** at m/z 364.26.

331.08, which has been previously reported using APCI- and ESI-MS/MS.³⁷ Please note that the APPI product ion scan of the partial dimeric unit **9e** at m/z 331.08 is obviously different from the MS/MS spectra recorded by APCI and ESI.³⁷ The product ions **9f** and **9g** at m/z 315.05 and 301.10 have been, respectively, produced by the elimination of a molecule of methane and a molecule of formaldehyde from the product ion **9e**.³⁷ Homolytic cleavage of the C-8-C-5' covalent bond of the coniferyl aldehyde and the C-7-O-4' formed from coniferyl and the ether oxygen situated on C-4' creates the product ion **9h** at m/z 181.05. This latter ion affords the

product ions **9i**, **9j** and **9k**, respectively, at m/z 167.07, 153.07 and 137.07. The tentative fragmentation routes of the product ion scan of the protonated molecule **9** are shown in Scheme 5.

The APPI-QqToF-CID-MS/MS spectrum of the protonated trimeric dicarboxylic acid $[C_{31}H_{33}O_{11}]^+$ **20** at m/z 581.20 is shown in Fig. 5. The protonated molecule at m/z 581.20 loses a molecule of water to form the ion **20a** at m/z 563.14. The precursor ion **20** also eliminates a molecule of ketene to create the ion **20b** at m/z 539.14. Loss of a molecule of acetic acid, from the precursor ion, by cleavage of the C-8''-C-9'' covalent bond, affords the product ion **20c** at m/z 521.11. Elimination

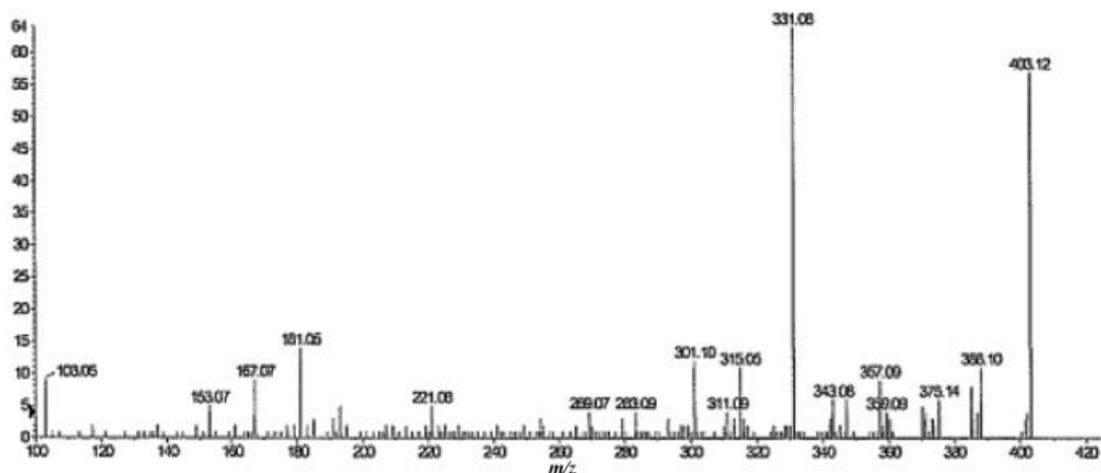
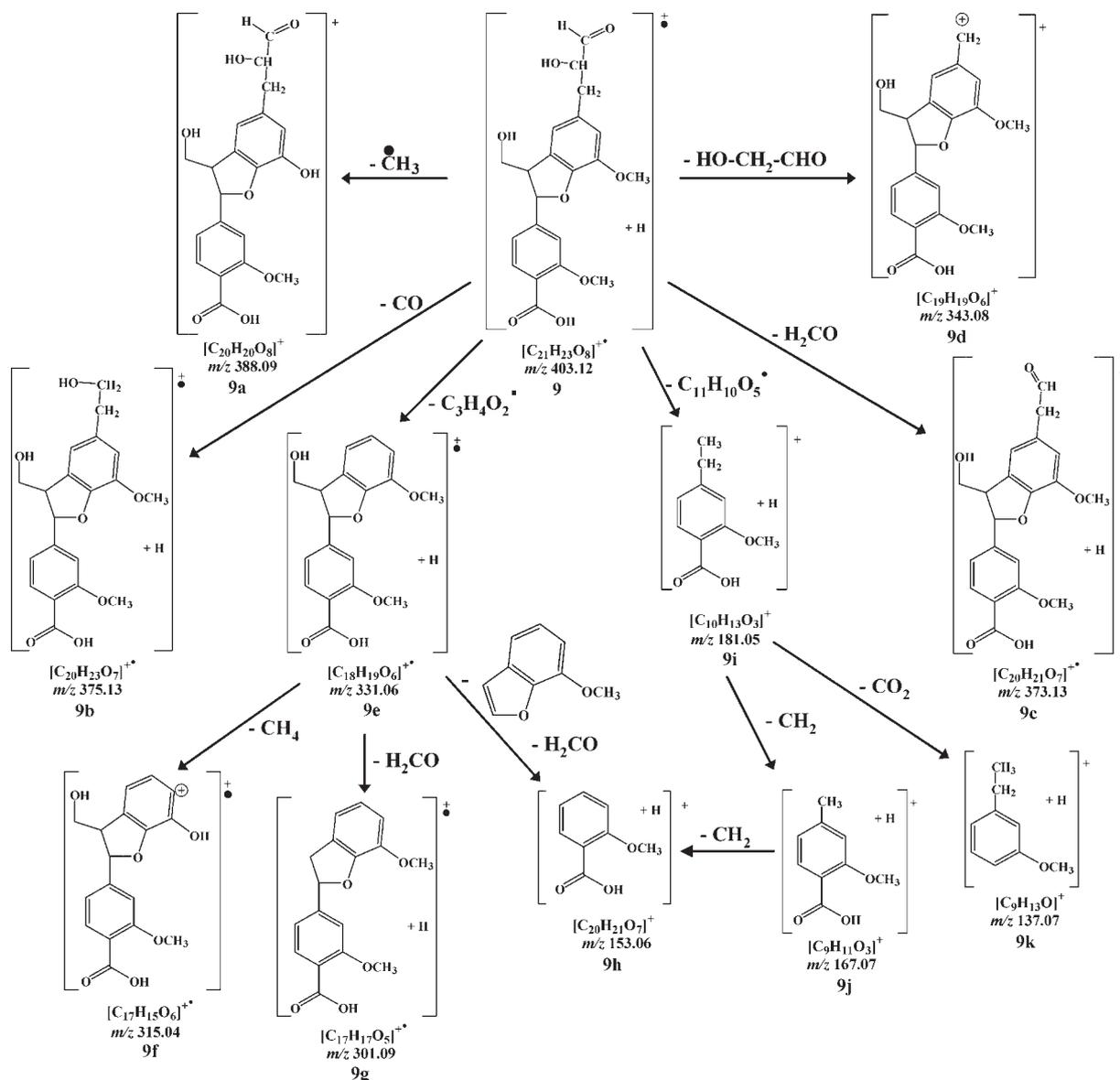


Figure 4. The product ion scan of the protonated radical-dimeric unit $[C_{21}H_{23}O_8]^+$ **9** at m/z 403.12.



Scheme 5. Tentative breakdown processes in the MS/MS of the protonated radical-dimeric unit $[C_{21}H_{23}O_8]^+$ **9** at m/z 403.12.

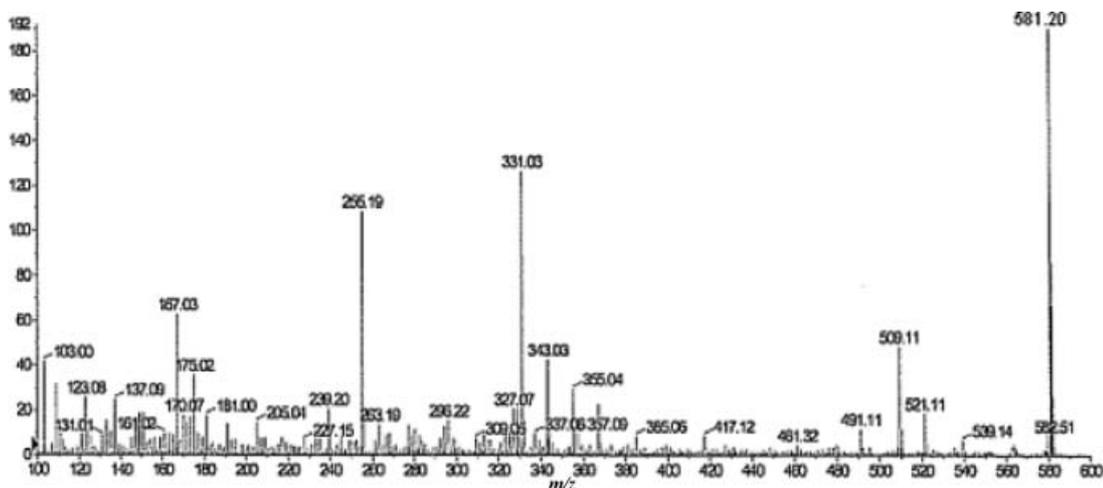


Figure 5. The product ion scan of the protonated trimeric dicarboxylic acid $[C_{31}H_{33}O_{11}]^+$ **20** at m/z 581.20.

of propenoic acid (72 Da), by cleavage of the C-1''–C-7''' covalent bond, affords the product ion **20d** at m/z 509.12, which is identical to the protonated molecule $[C_{28}H_{29}O_9]^+$ of the trimeric ionic species **13**. Elimination of the terminal residue, containing the aliphatic carboxylic acid of the trimeric precursor ion, by consecutive covalent bond breakages, affords the series of product ions **20e**, **20f** and **20g** at m/z 367.06, 355.06 and 343.06, respectively. Finally, cleavage of the C-1'–C-7''' covalent bond of the second and third coniferyl residues of the precursor ion affords the product ion **20h** at m/z 331.06, which is identical to the protonated molecule $[C_{18}H_{19}O_6]^+$ **2**. Once more, the tentative fragmentation routes of the product ion scan of the protonated molecule **20** are shown in Scheme 6.

We deliberately selected the next example of MS/MS analysis to exemplify the presence of protonated molecule **22**, which represents a tetrameric protonated molecule of mixed coniferyl and *p*-hydroxycinnamyl (cumaryl) residues. We therefore cautiously propose that this mixed oligomeric unit may exist, either in the native extracted lignin, or as a degradation product obtained during the AVIDEL extraction process. The APPI-QqToF-CID-MS/MS of the protonated molecule $[C_{37}H_{33}O_{10}]^+$ at m/z 637.20 afforded two major product ions at m/z 581.20 and 525.16. The ion **22a** at m/z 581.20 was formed by the loss of a molecule of propenal (56 Da). This latter product ion forms the ion **22b** at m/z 525.15 by the consecutive losses of molecules of acetylene and formaldehyde. For the sake of brevity, the product ion scan and the tentative break-down processes and the CID-QqToF-MS/MS spectrum of the protonated molecule **4** are available, as an Addendum, upon request.⁴⁹

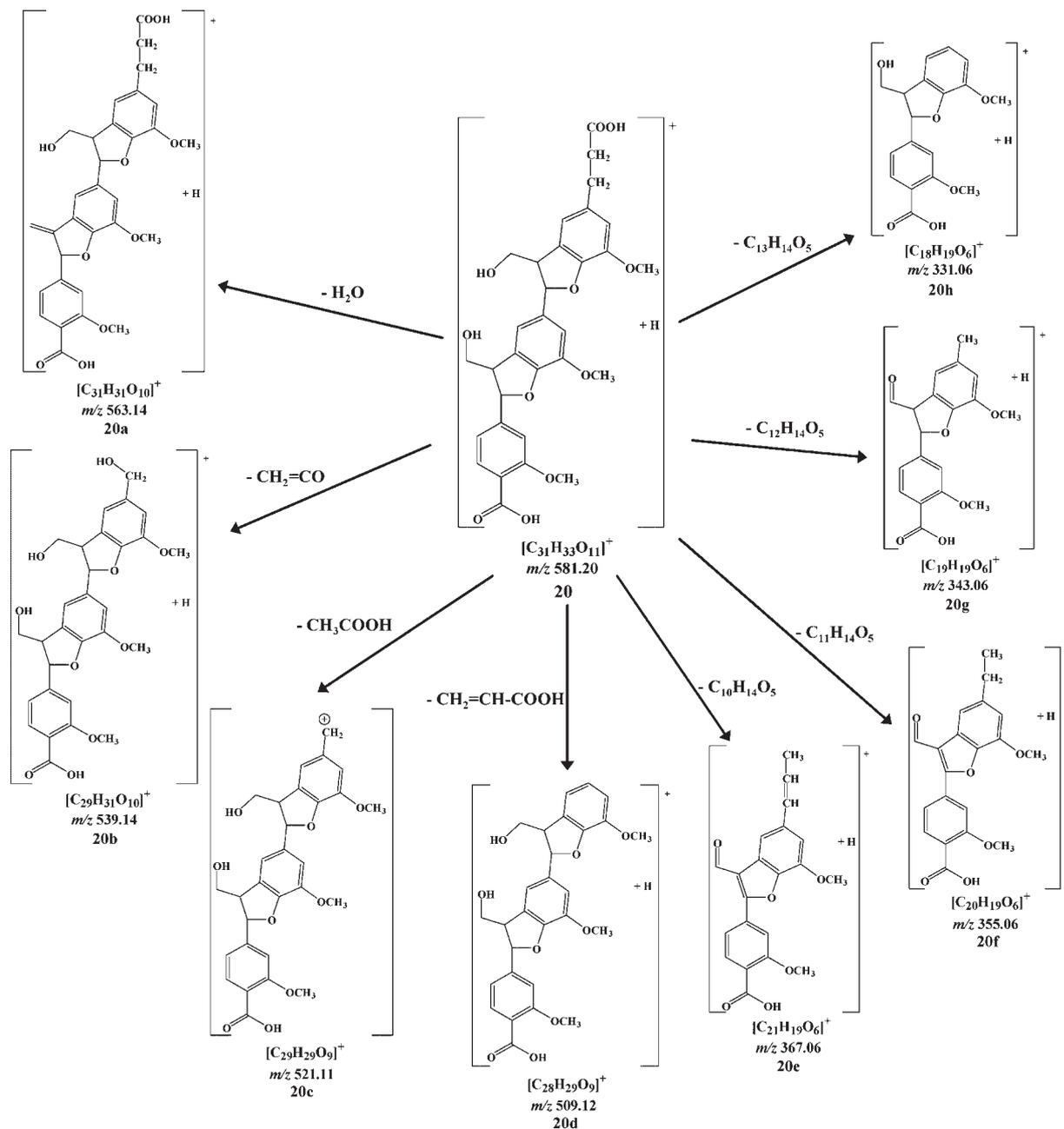
The APPI-QqToF-CID-MS/MS of the protonated tetrameric unit **25** $[C_{39}H_{39}O_{13}]^+$ at m/z 715.19 is shown in Fig. 6. The protonated molecule **25** $[C_{39}H_{39}O_{13}]^+$ at m/z 715.19 afforded the product ion **25a** at m/z 697.15 by loss of a molecule of water. Fission of the C-1'–C-7' covalent bond of the precursor ion affords the protonated molecule **25c** at m/z 331.06, which is identical with the protonated molecule $[C_{18}H_{19}O_6]^+$ **2**. The concomitant cleavages of the C-7'–C-8' covalent bond and the C-7'–O-4' ether bond of the precursor

ion afford the product ion **25d** at m/z 343.06. The product ion **25b** at m/z 643.18 is formed from the precursor ion by the consecutive losses of carbon monoxide and carbon dioxide. This latter product ion generated the product ion **25c** at m/z 591.12, as shown in the general breakdown pattern depicted in Scheme 7.

APPI-QqToF-CID-MS/MS of the protonated tetrameric oligomeric precursor $[C_{40}H_{41}O_{13}]^+$ **26** at m/z 729.21 afforded a series of product ions at m/z 669.17, 657.16, 639.16 and 167.02. The product ion scan and the tentative breakdown processes and the CID-QqToF-MS/MS spectrum of the protonated molecule **4** are available, as an Addendum, upon request.⁴⁹

For the higher molecular species **33–39** the APPI mass spectra were recorded with a mass range of m/z 700–1100 to increase the signal intensities of these specific ions. The MS/MS analyses were conducted with acquisitions varying from 60–100 counts.

As mentioned earlier in this manuscript, comparable APPI-QqToF-CID-MS/MS analyses were conducted on all the 39 different protonated molecules presented in Scheme 2, confirming their proposed structures. These findings established the universal comparable fragmentation patterns of this series of oligomers of the wheat straw lignin. We have shown that in all the product ion scans of this series of selected protonated molecules, the main dissociation reactions occurred by eliminations of small molecules such as carbon dioxide and formic acid, which were indicative of the presence of the carboxyl group on the C-4 of the first coniferyl residue. In addition, we have noted eliminations of formaldehyde molecules from the respective methoxylated group present on the coniferyl residues. We also noted the losses of methane molecules and methyl radicals from the selected precursor ions, consequently affording the cyclic ketone group on the respective coniferyl unit. The loss of the carbonyl group generally occurred by a ring contraction, to afford the five-membered rings in the respective coniferyl units. It is imperative to mention that the majority of the product ion scans of this series of protonated molecules indicated the presence of the intermediate five-membered



Scheme 6. Tentative breakdown processes in the MS/MS of the protonated trimeric dicarboxylic acid $[C_{31}H_{33}O_{11}]^+$ **20** at m/z 581.20.

furan-like ring in the oligomers, formed by the C-8-C-5' covalent bond and the C-7-O-4' ether linkage between every other contiguous di-coniferyl unit.

Negative mode APPI-MS of the wheat straw lignin

The APPI mass spectra of the extracted wheat straw lignin were recorded in the negative ion mode in the range of m/z 100–1000. Please note that Table 2 shows the characteristic ions obtained from the APPI-MS analyses which were recorded over a scan range of m/z 350–600 with two different declustering potentials. These spectra indicated the presence of at least 26 additional specific deprotonated oligomers. We have identified at least 6 oligomeric species that photo-

ionized in both the positive and negative ion modes producing the respective protonated and deprotonated molecules. The identities of the novel 18 elucidated oligomers species **40** to **57** are highlighted in Scheme 2.

Once more, the structural characterization of the oligomers was based on the exact molecular masses measured by high-resolution QqToF-MS. The differences between the calculated and the observed molecular weight values ranged between ± 1.1 and 8 ppm for molecules with molecular masses smaller than 400; this is considered as an acceptable difference for the observed molecular weights.⁵⁰ For ions greater than m/z 400, the difference for the observed masses was below ± 0 to 2 ppm.⁵⁰ We considered as many structures as possible that fitted the experimental values, and in the

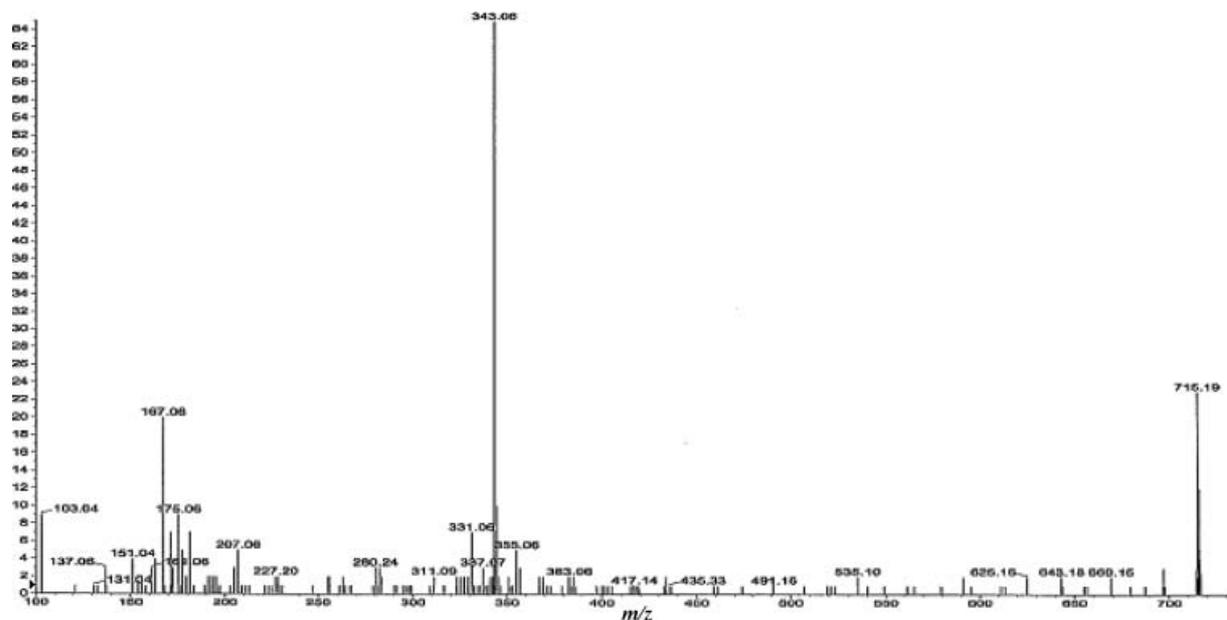


Figure 6. APPI-QqToF-CID-MS/MS of the protonated tetrameric unit $[C_{39}H_{39}O_{13}]^+$ **25** at m/z 715.19.

Addendum,⁴⁹ we also take into account the results of the combustion analysis of the extracted wheat lignin: Found: C 58.4%, H 5.4%, O 35.0%, which corresponds to a molecular formula of $C_{4.9}H_{5.5}O_{2.2}$. The candidate formulae that we are proposing are those that fit the tentative structures, obtained

after studying their fragmentation patterns by APPI-MS/MS. In the APPI mass spectra we observed two deprotonated monomers **40** and **41**; six open-chain deprotonated dimers **42**, **43**, **47**, **48**, **49** and **50** containing the C-4 carboxyl group; three open-chain deprotonated dimers **44**, **45** and **46**

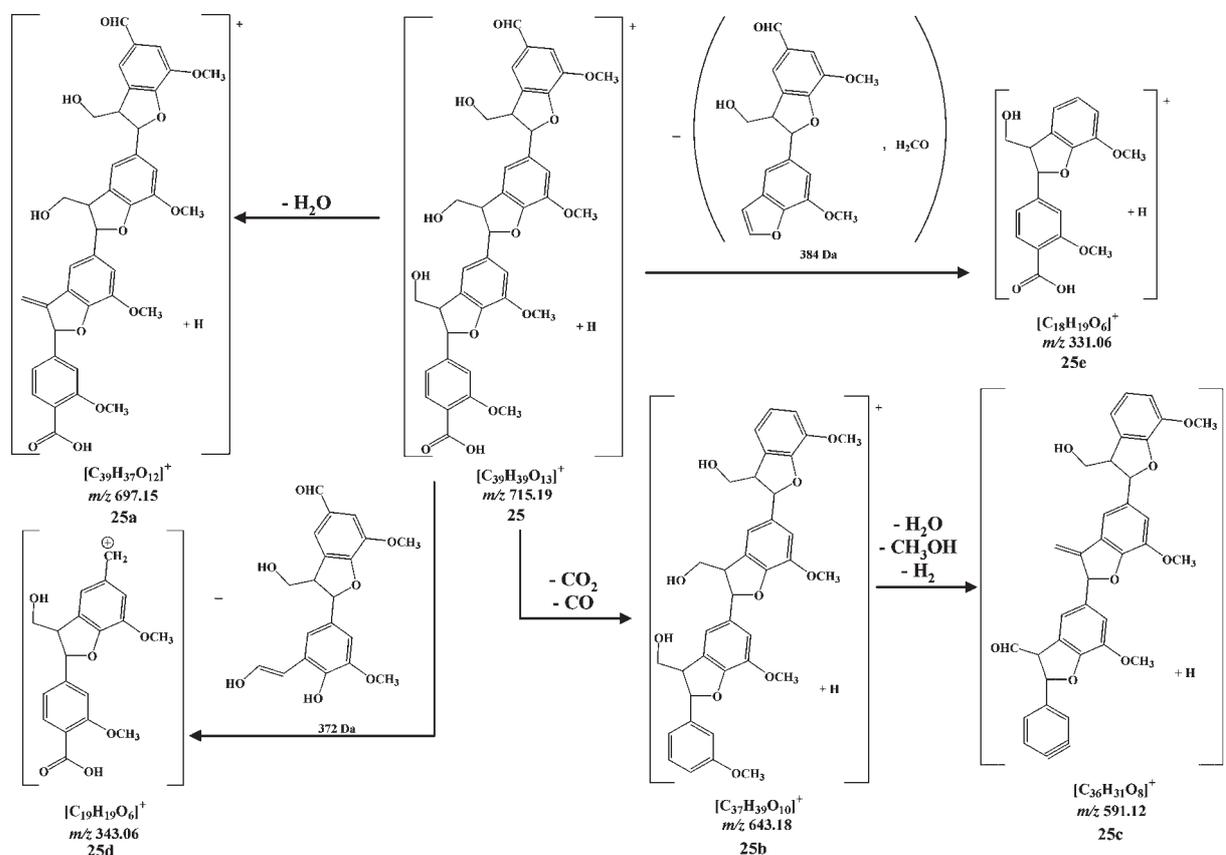


Table 2. Characteristic ions obtained from the APPI-MS analysis of the wheat straw lignin recorded in the negative ion mode scanning m/z 300–700 with a DP = 100 V

Characteristic ions	Cpd #	Calculated m/z	Observed m/z	Abundance (%)	Difference (ppm)
$[\text{C}_{10}\text{H}_9\text{O}_4]^-$	41	193.0501	193.0509	72.14 ^a	-4.1
$[\text{C}_{16}\text{H}_{13}\text{O}_4]^-$	42	269.0814	269.0799	51.61 ^a	5.5
$[\text{C}_{17}\text{H}_{15}\text{O}_5]^-$	43	299.0919	299.0920	48.28 ^a	-0.3
$[\text{C}_{18}\text{H}_{15}\text{O}_5]^-$	44	311.0919	311.0920	61.12 ^a	-0.3
$[\text{C}_{19}\text{H}_{17}\text{O}_5]^-$	45	325.1076	325.1075	50.12 ^a	0.3
$[\text{C}_{18}\text{H}_{17}\text{O}_6]^-$	2	329.1025	329.1020	100.00 ^a	1.5
$[\text{C}_{20}\text{H}_{19}\text{O}_5]^-$	46	339.1232	339.1233	14.90 ^a	-0.3
$[\text{C}_{19}\text{H}_{17}\text{O}_6]^-$	47	341.1025	341.1013	40.62 ^a	3.5
$[\text{C}_{19}\text{H}_{19}\text{O}_6]^-$	48	343.1182	343.1181	45.71 ^a	0.29
$[\text{C}_{21}\text{H}_{19}\text{O}_7]^-$	49	383.1131	383.1130	50.00	0.3
$[\text{C}_{21}\text{H}_{21}\text{O}_7]^-$	50	385.1076	385.1075	64.52	0.3
$[\text{C}_{21}\text{H}_{21}\text{O}_8]^-$	51	401.1236	401.1235	38.77	0.2
$[\text{C}_{26}\text{H}_{29}\text{O}_7]^-$	52	453.1913	453.1914	21.40	-0.2
$[\text{C}_{27}\text{H}_{30}\text{O}_8]^-$	53	482.1941	482.1940	12.91	0.2
$[\text{C}_{28}\text{H}_{25}\text{O}_8]^-$	54	489.1549	489.1550	35.82	0.2
$[\text{C}_{28}\text{H}_{27}\text{O}_8]^-$	55	491.1706	491.1706	36.73	0.0
$[\text{C}_{28}\text{H}_{27}\text{O}_9]^-$	13	507.1655	507.1654	67.34	0.2
$[\text{C}_{29}\text{H}_{29}\text{O}_9]^-$	15	521.1812	521.1811	36.15	0.2
$[\text{C}_{30}\text{H}_{32}\text{O}_9]^-$	56	536.2046	536.2045	22.50	0.2
$[\text{C}_{30}\text{H}_{29}\text{O}_{10}]^-$	17	549.1761	549.1760	26.63	0.2
$[\text{C}_{30}\text{H}_{31}\text{O}_{11}]^-$	57	567.1865	567.1864	15.12	0.2
$[\text{C}_{38}\text{H}_{35}\text{O}_{12}]^-$	24	683.2493	683.2493	30.12 ^b	0.0

^a Recorded with a mass range of m/z 100–400.

^b Recorded with a mass range of m/z 400–800.

containing the coniferyl C-4, C-7 and C-8 positions, bearing a hydroxyl group; two deprotonated cyclic dimers **2** and **51** containing the C-4 carboxylic group; seven deprotonated cyclic trimers **13**, **15**, **17**, **54**, **55**, **56** and **57**; and a deprotonated cyclic tetramer **24**, all of which contained the carboxyl group on the C-4 position of the first coniferyl unit. We also noted the presence of one open-chain deprotonated trimer **53** possessing a carboxyl group at the C-4 position of the first coniferyl unit. We would like to point out that the latter deprotonated molecule was an open-chain oligomer, formed by the C-8–C-5' covalent bond between contiguous di-coniferyl units which has only a hydroxyl group at positions C-4, C-4' and C-4'' (in which, obviously, the intermediate furan-like rings are missing). Similarly, we noted the presence of one open-chain deprotonated trimer **52** containing a hydroxyl group at the C-4 position of the first coniferyl unit.

Negative mode APPI-QqToF-CID-MS/MS analyses of the specific oligomeric ions obtained from wheat straw lignin

The APPI-QqToF-CID-MS/MS spectra of the deprotonated molecules **44**, **45** and **46**, respectively, at m/z 311.04, 325.04 and 339.05, are shown in Fig. 7. Please note that this series of anions contained the guaiacylglycerol unit (i.e. containing hydroxyl groups at C-7, C-8 and C-9 of the first coniferyl residue), which has been previously reported by Forss and Fremmer, as part of an ordered polymer called 'glycolignin'.⁵¹ The product ion scan of the deprotonated molecule **44** at m/z 311.04 shows the loss of a molecule of methane, with hydrogen transfer, to create the product ion **44a** at m/z 295.89. This latter ion eliminates a carbonyl group followed by a ring contraction to form the product ion **44b** at m/z 266.94. The

homolytic cleavage of the covalent bond C-8–C-9 of the deprotonated molecule affords the product ion $[\text{C}_9\text{H}_{11}\text{O}_4]^-$ **44c** at m/z 182.93. Similarly, the product ion scans of the deprotonated molecule **45** at m/z 325.04, and deprotonated molecule **46** at m/z 339.05, afford the same product ion $[\text{C}_9\text{H}_{11}\text{O}_4]^-$ **44c** at m/z 182.93, which is a diagnostic ion confirming the presence of the coniferyl glycerol or the guaiacylglycerol unit (fragmentation patterns are shown in the figures).

The product ion scans of the deprotonated molecules $[\text{C}_{19}\text{H}_{17}\text{O}_6]^-$ **47**, $[\text{C}_{19}\text{H}_{19}\text{O}_6]^-$ **48** and $[\text{C}_{21}\text{H}_{21}\text{O}_7]^-$ **50**, respectively, at m/z 340.92, 342.93 and 384.92, were recorded and the fragmentation patterns of the deprotonated molecules are discussed in the Addendum.⁴⁹

The product ion scan of the last deprotonated molecule **50** at m/z 384.92 exemplifies the characterization of the 4-carboxyl-7,8-dehydroconiferyl alcohol, covalently attached by the C-8–C-5' covalent bond to the following coniferyl alcohol residue (Fig. 8). In the product ion scan of this deprotonated molecule, we propose that this precursor ion exists in equilibrium between two tautomers, as indicated in Scheme 9. The elimination of the C-4'-keto group from the precursor ion, by a ring contraction, affords the product ion **50c** at m/z 356.91 containing the furan ring. The precursor ion also loses a molecule of methane, with a hydrogen-atom transfer, to create the product ion **50a** at m/z 369.90. This latter ion loses a molecule of acetaldehyde (44 Da) to afford the ion **50e** at m/z 325.93. This last ion loses a molecule of carbon dioxide to give the ion **50g** at m/z 281.97. A loss of a molecule of water from the precursor ion yields the ion **50b** at m/z 366.88. The precursor ion, also **50**, loses a molecule of 1-hydroxyethylene to generate the ion **50d** at m/z 340.95. Please note that this latter product ion is identical to

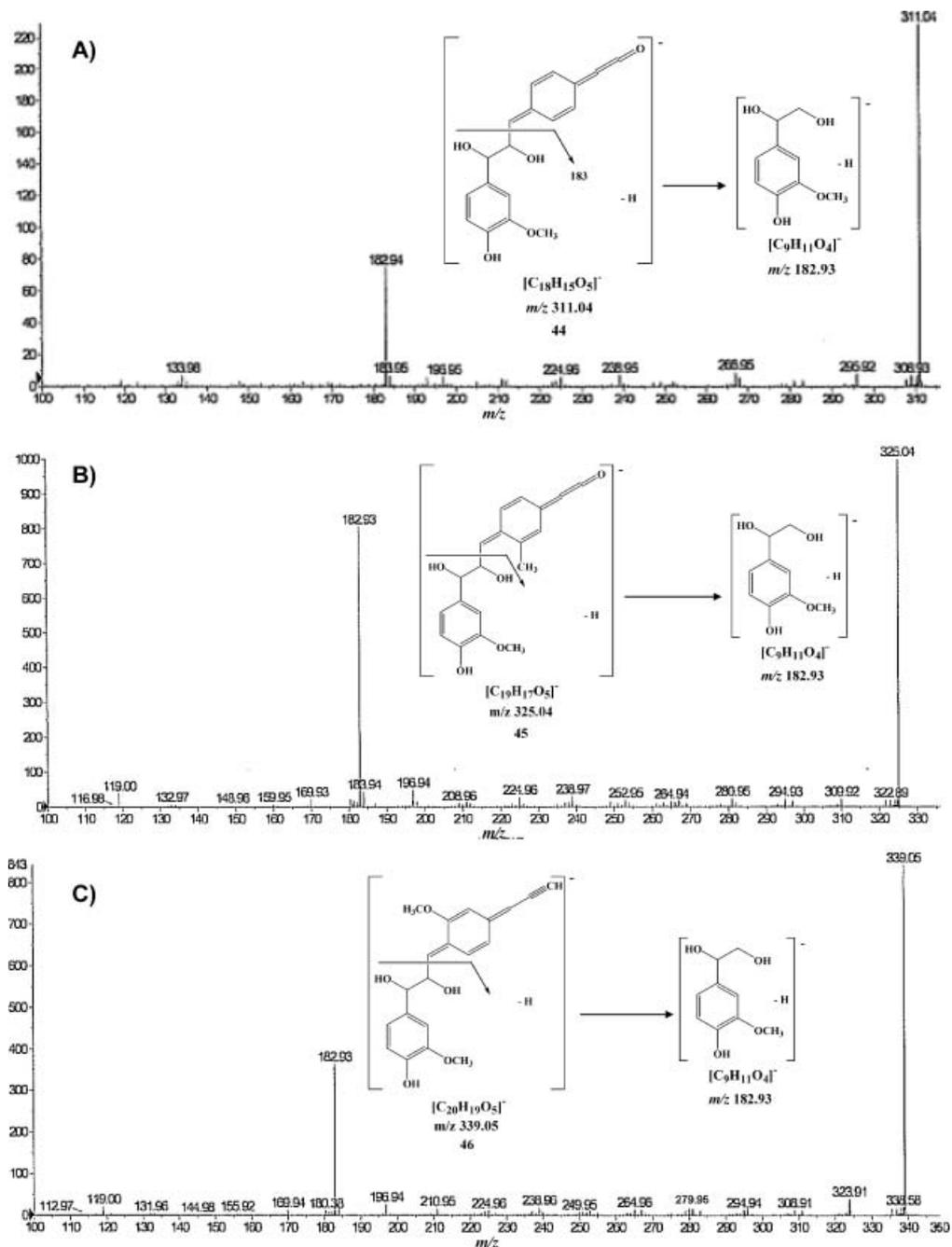


Figure 7. (A) CID-APPI-QqToF-MS/MS of deprotonated **44** at m/z 311.04. (B) CID-APPI-QqToF-MS/MS of deprotonated **45** at m/z 325.04. (C) CID-APPI-QqToF-MS/MS of deprotonated **46** at m/z 339.05.

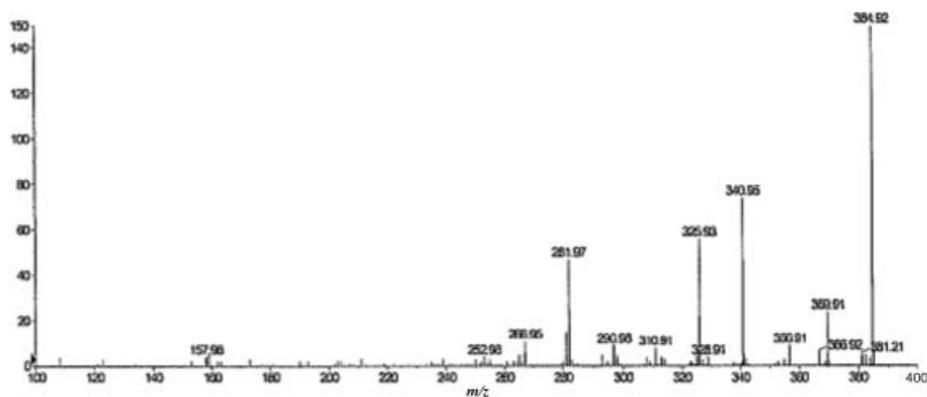
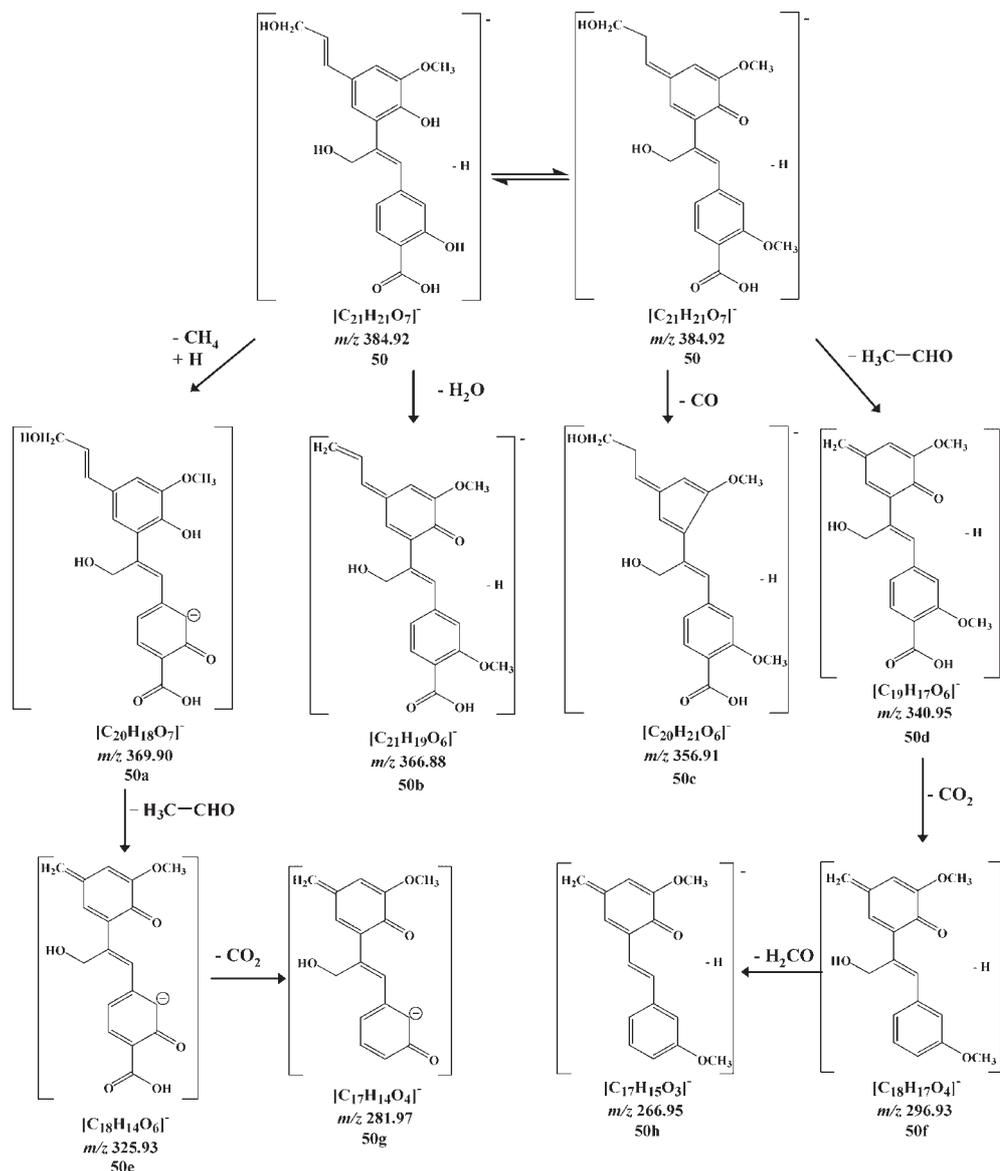


Figure 8. Product ion scan of deprotonated $[C_{21}H_{21}O_7]^-$ **50** at m/z 384.92.



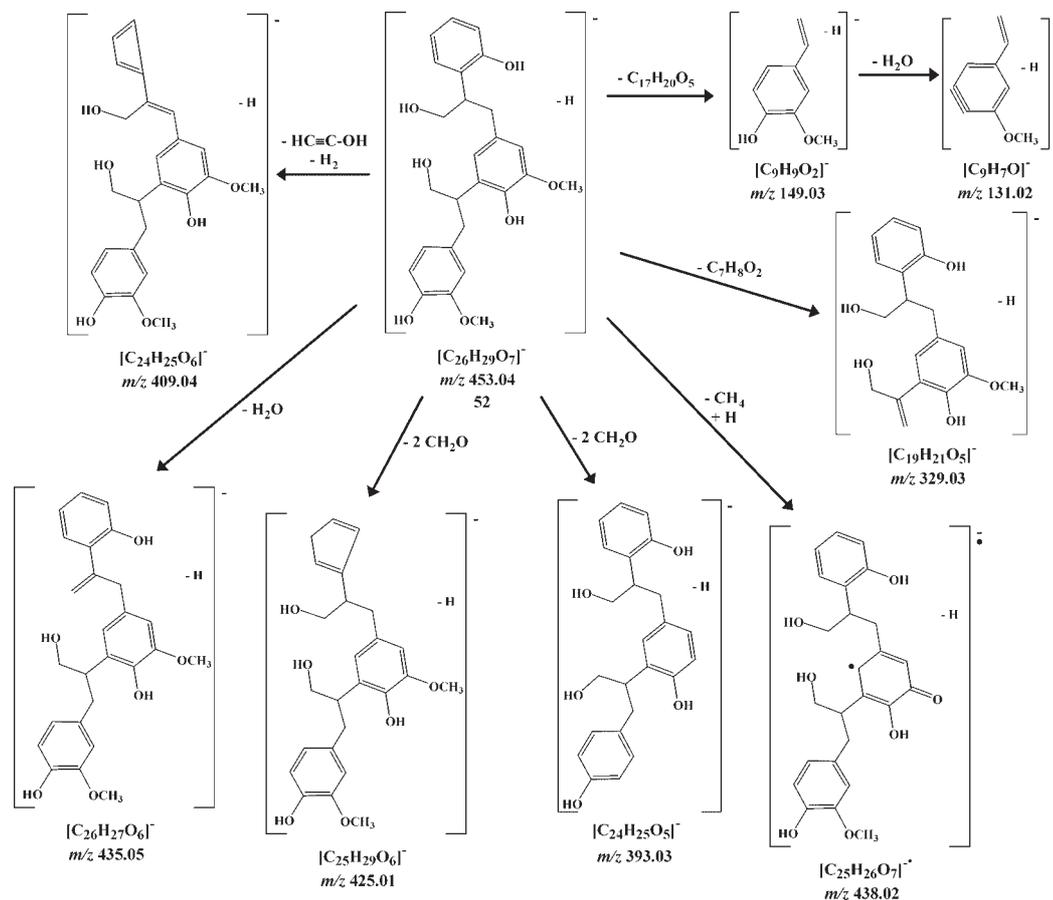
Scheme 8. Tentative breakdown processes from the product ion scan of deprotonated **50** at m/z 384.92.

deprotonated molecule **47**. The remaining product ions formed are indicated in Scheme 8. This last APPI-QqToF-CID-MS/MS study was included to indicate the presence of unique species: the open-chain acyclic dimers and oligomers obtained during the extraction of the wheat straw lignin by the AVIDEL procedure. The presence of these open-chain oligomers once more shows the microheterogeneity of the wheat straw lignin.

Finally, the product ion scans of the deprotonated open-chain trimer $[C_{26}H_{29}O_7]^-$ **52** at m/z 453.04 (Fig. 9) and the deprotonated cyclic trimer molecule $[C_{30}H_{29}O_{10}]^-$ **17** at m/z 549.17 (Fig. 10) were obtained. The deprotonated molecule $[C_{26}H_{29}O_7]^-$ **52** at m/z 453.04 lacks the C-4 carboxyl group on the first coniferyl unit and it contains, on the contiguous coniferyl residue, a C-4' free hydroxyl group. The open-chain trimer $[C_{26}H_{29}O_7]^-$ **52** is probably located in the middle of an 'opened up' repeating unit, which has been released by acid hydrolysis during the extraction procedure.

The fragmentation breakdowns of these two aforementioned precursor ions **52** and **17** are indicated in Schemes 9 and 10, and reflect the commonality in the fragmentation of these anions.

Our experimental results may also indicate that the native wheat straw lignin is composed of different species of oligomers, formed from different types of monomer coniferyl residues. We have become aware of the diverse phenylcoumaran repeating units that can be composed of mixed oligomeric species. These mixed species were formed by attachment of the initial constituent 4-carboxylconiferyl residue, that is linked by the usual C-8-C-5' covalent bond, and the C-7-O-4' ether bond, to the next unit, which in that case is the *p*-hydroxycinnamyl alcohol residue (H), lacking the methoxyl groups, as shown for the oligomers **10**, **11**, **12**, **21**, **22**, **23**, **24**, **28**, **29**, **30**, **33**, **34** and **35**–**39**. These results are similar to those of Forss and Fremer, who have shown that the Finnish spruce wood 'glycolignin' polymer is not a single



Scheme 9. Tentative breakdown processes from the product ion scan of the deprotonated open-chain molecule $[C_{26}H_{29}O_7]^-$ **52** at m/z 453.18.

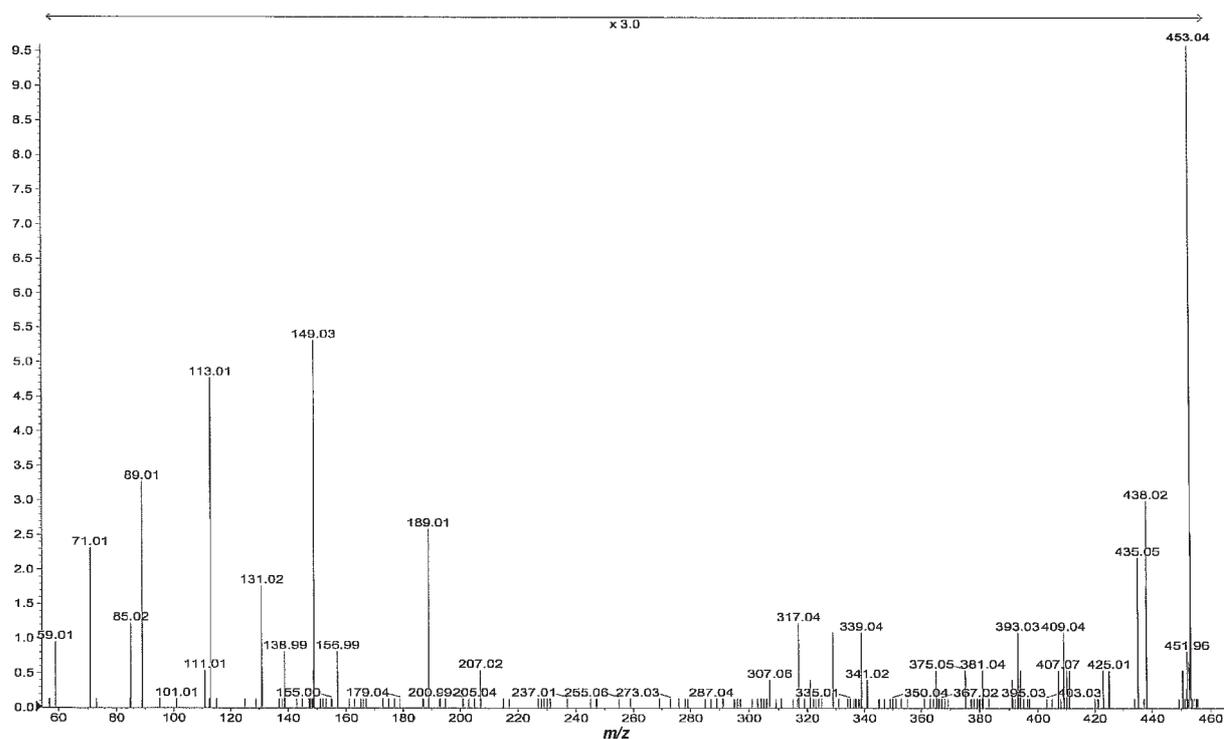


Figure 9. Product ion scan of the protonated open chain molecule $[C_{26}H_{29}O_7]^+$ **51** at m/z 453.04.

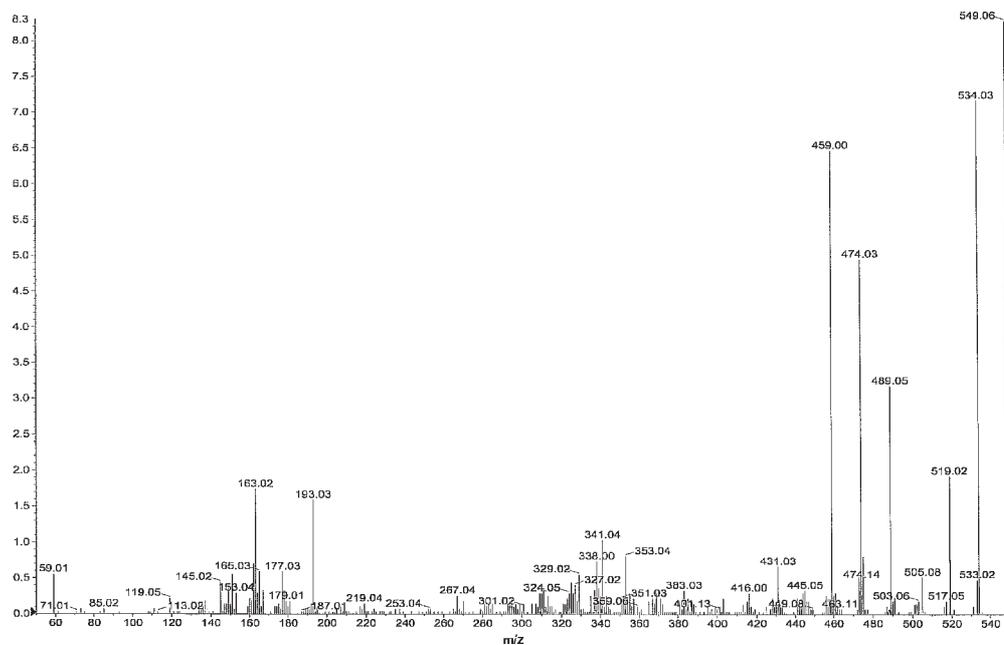
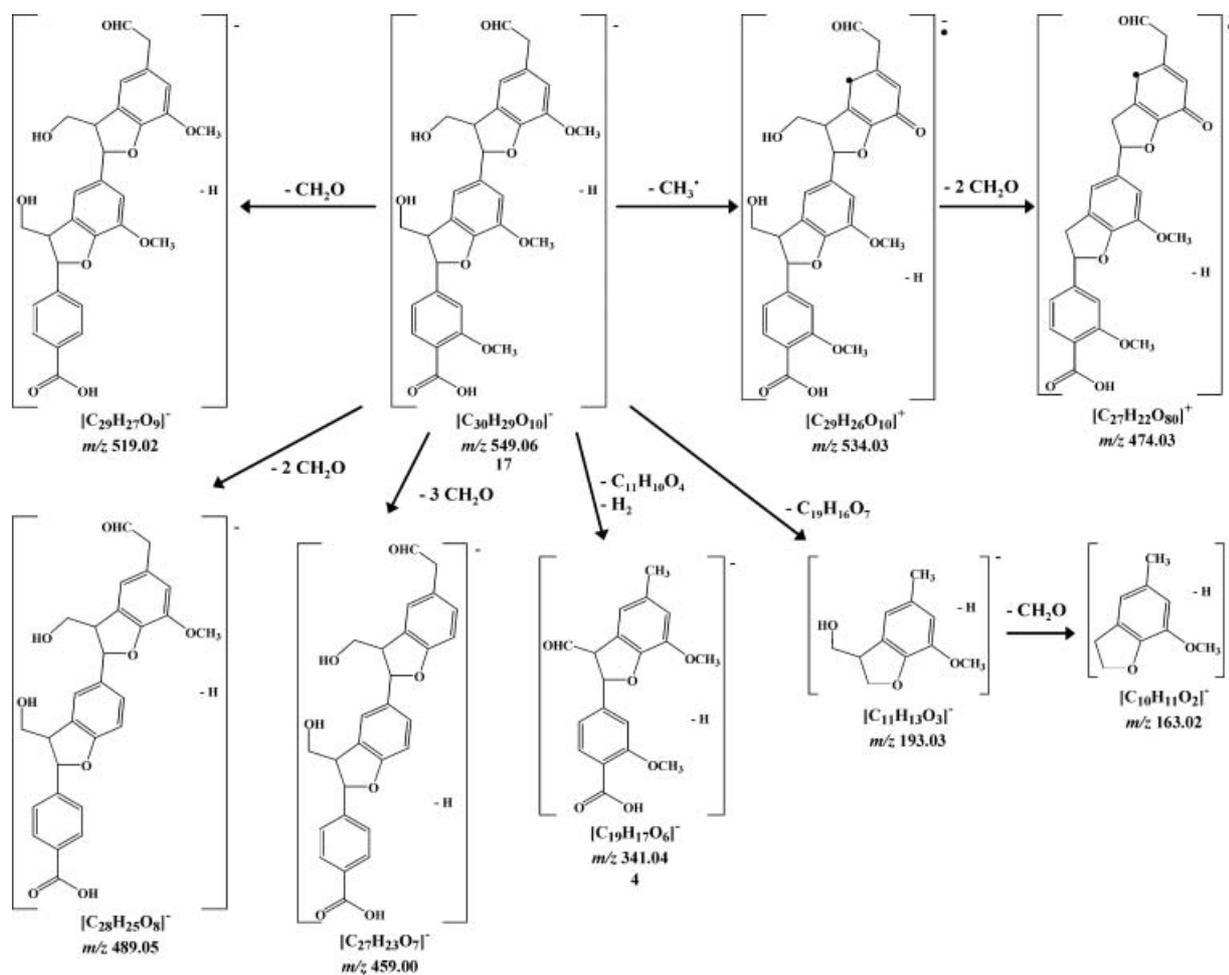


Figure 10. Product ion scan of the trimer deprotonated open-chain molecule $[C_{30}H_{29}O_{10}]^-$ **17** at m/z 549.17.



Scheme 10. Tentative breakdown processes of the deprotonated trimer molecule $[C_{30}H_{29}O_{10}]^-$ **17** at m/z 549.17.

polymer but probably a series of heterogeneous related polymers.⁵¹

We would like to suggest that, according to our experimental results, the identified mixed oligomeric units may exist, either in the native extracted lignin or as products of degradation during the extraction process. To our knowledge the CIMV extraction process is one of the most chemically gentle and non-destructive processes that has been reported.³⁹ For that reason, we cautiously propose that the formation of the different acyclic oligomers may be the result of the unzipping (opening) of the reactive constituent cyclic oligomers of native lignin which almost certainly takes place during this smooth extraction process.

CONCLUSIONS

It has been proposed by Atalla that the intracellular regulation of the structure of lignins occurs through the provision of templates for their spatial organization, prior to the development of covalent linkages during radical coupling reactions.²¹ Therefore, a constant and linear structure of lignin can only be considered within a synthesis performed in the presence of a matrix which could be either cellulose or hemicellulose. This matrix may control the anarchical synthesis which would inevitably be produced in a system where contact between radicals can occur at random.^{6,14,17–21}

Our present investigation indicated that the native lignin polymer is composed of a mixture of different linear polycondensed coniferyl units. The main constituent repeating phenylcoumaran units are formed from two contiguous di-coniferyl residues, which are linked by the C-8–C-5' covalent bond and the ether C-7–O-4' linkage, hence forming the intermediate furan-like ring of the constituent repeating phenylcoumaran units.

We therefore propose that the main constituent of the heterogeneous lignin polymer, containing the repeating phenylcoumaran units, can be attached by various ester bonds formed between the opposite terminal carboxylic end groups of the lignin oligomers (compounds **14**, **15**, **22**, **30** and **35**) and the primary hydroxyl groups C-6' of hemicelluloses matrices. Such structures, by virtue of their crossed covalent linkage nature, will be 'tridimensional'.

Notwithstanding this new finding, we wish to repeat that the native lignin polymer, despite its microheterogeneity, is extremely chemically reactive when subjected to harsh chemical methods (i.e. extraction and all chemical degradations), and hence tends to either break down quite easily and/or react rapidly to produce new compounds, not originally present in the native lignin.^{24–26} This may explain the inconsistencies in the structures that have been proposed in the last fifty years, especially after extraction of the native lignin with extremely harsh chemical procedures.^{2,6–10}

We would like to express some words of caution to the practising NMR researchers who may use high-resolution CP/MAS ¹³C NMR as the sole means of determining the molecular structure of lignin, especially when we know that the latter is a micro-heterogeneous mixture of approximately more than 57 oligomeric compounds. It is crucial to point out that the ratio between the signals at 153.1 and 147.5 ppm can be confused with the ratio of etherified and free syringyl and

coniferyl units. We have found that this ratio typically indicates the relative amount of the β -C-5' containing the cyclic lignin units *vis a vis* the open-chain α -O-4' lignin. In the supplementary information, please note that the broad NMR resonance at 153.1 ppm can be easily confused with resonance signals of syringyl alcohol units which are definitely absent in our native lignin.^{44–47} However, we should acknowledge that the fact that we can obtain a comprehensive solid-state NMR spectrum was beyond any of our expectations and this confirms the robustness of this technique.^{41–43} Most prominently, we could recognize and account, for the first time, the presence of several misidentified carboxylic groups in the solid-state CP/MAS NMR spectrum. These carboxylic groups were attributed as being inconsequential groups from proteins or other unknown extracted compounds. The presence of several types of carboxyl groups has been established as an intrinsic part of the various structures of the lignin backbone, which were identified by APPI-MS and MS/MS.

In conclusion, we can say without any doubt that photoionization provides superior ionization of native lignin to ESI and/or APCI.^{36–38} In this manuscript, we presented the analysis of approximately 63 (57 new species and 6 other species that were measured using APPI-MS in both positive and negative ion modes) related oligomers obtained from native lignin by direct introduction into the APPI source, without LC separation, allowing extremely thermally labile and reactive oligomers to be studied. Positive and negative ion APPI-MS was performed with no off-line sample cleanup, or any kind of chromatographic purification. In addition, APPI-CID-QqToF-MS/MS analysis using low-energy CID permitted the separation of these novel 57 ionic species, in addition to the validation of the fragmentation pathways. MS/MS analyses specifically permitted us to infer the exact molecular structure of each individual novel constituent of native wheat straw lignin.

Acknowledgements

Michel Delmas acknowledges the financial support from CIMV, Paris France. Joseph Banoub acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada for a Discovery Grant.

REFERENCES

1. Page DH. *Wood Fiber* 1976; **7**: 246; Fengel D. *J. Polymer Sci.: Part C* 1971; **36**: 383.
2. Adler E. *Wood Sci. Technol.* 1977; **11**: 169.
3. Earl WL, VanderHart DL. *Macromolecules* 1981; **14**: 570.
4. Jones DW. In *Cellulose and Cellulose Derivatives*, vol. 5, part 4, High Polymers, Bikales MN, Segal L (eds). Wiley Interscience: New York, 1971; 117; Ellefsen O, Tonnesen BA. in *ibid.* 1971; 151; Tonnesen BA, Ellefsen O. in *ibid.* 1971; 265; Atalla RH. *Adv. Chem. Ser.* 1979; **181**: 55.
5. Karlsson O, Pettersson B, Westermark U. *J. Pulp. Pap. Sci.* 2001; **27**: 196.
6. (a) Freudenberg K. *Nature* 1959; **183**: 1152; (b) Freudenberg K, Hess H. *Liebigs Ann. Chem.* 1926; **448**: 126; (c) Freudenberg K, Richtzenhain H. *Ann.* 1942; **552**: 126.
7. (a) Freudenberg K, Neish AC. *Constitution and Biosynthesis of Lignin (Molecular Biology, Biochemistry, and Biophysics)*. Springer-Verlag: New York, 1969; 132; (b) Alder E. *Wood Sci. Technol.* 1977; **11**: 169.

8. Higuchi T. In *Biosynthesis of Lignin. Biosynthesis and Biodegradation of Wood Components*. Academic Press: New York, 1985; 141.
9. Salmén L, Olsson AM. *J. Pulp Pap. Sci.* 1998; **24**: 99.
10. Bouchard J, Overend RP, Chornet E, Calsteren M-RV. *J. Wood Chem. Technol.* 1992; **12**: 335 Fengel D. *J. Wood Chem. Technol.* 1992; **12**: 335.
11. Sarkanen KV, Ludwig CW. Definition and nomenclature. In *Lignins: Occurrence, Formation and Reactions*, Sarkanen KV, Ludwig CW (eds). Wiley Interscience: New York, 1971; 1.
12. Sarkanen KV. Lignin precursors and their polymerization. In *Lignins: Occurrence, Formation and Reactions*, Sarkanen KV, Ludwig CW (eds). Wiley Interscience: New York, 1971; 95.
13. Sarkanen KV, Islam A, Anderson CD. Ozonation. In *Methods in Lignin Chemistry*, Lin SY, Dence CW (eds). Springer-Verlag: Berlin Heidelberg, 1992; 387; and references cited therein.
14. Reale S, Di Tullio A, Spreti N, De Angelis F. *Mass Spectrom. Rev.* 2004; **23**: 87.
15. Eglinton TI, Goni MA, Boon JJ, Van der Hage ERE. *Holzforchung* 2000; **54**: 39.
16. Monties B. *Polym. Degrad. Stab.* 1998; **59**: 53.
17. Lewis NG, Davin LB. The biochemical control of monolignol coupling and structure during lignan and lignin biosynthesis. In *Lignin and Lignan Biosynthesis*, Lewis NG, Sarkanen S (eds). ACS Symposium Series, ACS: Washington, DC, 1998; **697**: 1.
18. Lewis NG, Davin LB, Sarkanen S. Lignin and lignan biosynthesis: distinctions and reconciliations. In *Lignin and Lignan Biosynthesis*, Lewis NG, Sarkanen S (eds). ACS Symposium Series, ACS: Washington, DC, 1998; **697**: 334.
19. Önerud H, Zhang L, Gellerstedt G, Henriksson G. *The Plant Cell* 2002; **14**: 1953; Ragauskas AJ, Wiulliams CK, Davidson BH, Britovesk G, Cairney J, Eckert CA, Frederick WJ Jr, Hallet JP, Leak DJ, Liotta CL, Mienlenz JR, Murphy R, Templer R, Tschaplinski T. *Science* 2006; **311**: 484.
20. Davin LB, Wang HB, Crowell AL, Bedgar DL, Martin DM, Sarkanen S, Lewis NG. *Science* 1997; **275**: 362.
21. Atalla RH. Cellulose and the hemicelluloses: patterns for the assembly of lignin. In *Lignin and Lignan Biosynthesis*, Lewis NG, Sarkanen S (eds). ACS Symposium Series, ACS: Washington, DC, 1998; **697**: 172.
22. Karlsson O, Westermark U. *J. Pulp Pap. Sci.* 1996; **22**: J397.
23. Hortling B, Ranua M, Sundquist J. *Nordic Pulp Pap. Res. J.* 1990; **5**: 33.
24. Obst JR, Kirk TK. Isolation of lignin. In *Methods in Enzymology: Biomass*, Wood WA, Kellog ST (eds). Academic Press: San Diego, 1998; 3.
25. Yamamoto E, Bokkelman GH, Lewis NG. In *Phenylpropanoids Metabolism in Cell Walls, in Plant Cell Wall Polymer, Biogenesis and Degradation*, Lewis NG, Paice MG (eds). American Chemical Society: Washington, DC, 1989; 68.
26. Conner AH. Size exclusion chromatography of cellulose and cellulose derivatives. In *Handbook of Size Exclusion Chromatography*, Wu C (ed). Marcel-Dekker: New York, 1995; 331.
27. Ohra-aho T, Tenkanen M, Tamminen T. *J. Anal. Appl. Pyrolysis.* 2005; **74**: 123.
28. MacKay J, Dimmel DR, Boon JJ. *J. Wood Chem. Technol.* 2001; **21**: 19.
29. Fachuang I, Ralph J. *J. Agric. Food Chem.* 1997; **45**: 2590.
30. Metzger JO, Bicke C, Faix O, Tuszynski W, Angermann R, Karas M, Strupat K. *Agnew. Chem. Int. Ed. Engl.* 1992; **31**: 762.
31. De Angelis F, Nicoletti R, Spreti N, Veri F. *Agnew. Chem. Int. Ed. Engl.* 1999; **38**: 1283.
32. De Angelis F, Fregonese P, Veri F. *Rapid Commun. Mass Spectrom.* 1996; **10**: 1304.
33. Evtuguin DV, Domingues PM, Amado FML, Pascoal Neto C, Ferrer Correia AJ. *Holzforchung* 1999; **53**: 525.
34. Evtuguin DV, Pascoal Neto C, Silva AM, Domingues PM, Amado FML, Robert D, Faix O. *J. Agric. Food Chem.* 2001; **49**: 4252; Evtuguin DV, Amado FML. *Macromol. Biosci.* 2003; **3**: 339.
35. Banoub J, Delmas M. *J. Mass Spectrom.* 2003; **38**: 900.
36. Cai SS, Syage JA. *J. Chromatogr.* 2006; **1110**: 15; Syage JA, Evans MD, Hanold KA. *Am. Lab.* 2000; **32**: 24.
37. Syage JA, Hanold KA, Lynn TC, Horner JA, Thakur RA. *J. Chromatogr. A* 2004; **1050**: 137; Hanold KA, Fischer SM, Cormia PH, Miller CE, Syage JA. *Anal. Chem.* 2000; **76**: 2842.
38. Syage J. *J. Am. Soc. Mass Spectrom.* 2004; **15**: 1521; Robb DB, Covey TR, Bruins AP. *J. Anal. Chem.* 2000; **72**: 3653.
39. Delmas M, Avignon G. *Brevet Fr.* 1997; **97**: 13658.
40. Lam HQ, Le Bigot Y, Delmas M, Avignon G. *Ind. Crops Prod.* 2001; **14**: 139.
41. Schaefer J, Stejskal EO. *J. Am. Chem. Soc.* 1976; **98**: 1031.
42. Peerson OB, Wu XL, Kustanovich I, Smith SO. *J. Mag. Reson.* 1993; **104**: 334.
43. Morcombe CR, Zilm KW. *J. Mag. Reson.* 2003; **132**: 479.
44. Atalla RH, Vander Hart DL. *Science* 1984; **17**: 283.
45. Torri G, Sozzani P, Focher B. In *Morphology and Structure of Cellulose materials as Studied by MAS NMR Spectroscopy from Molecular Materials to Solids*, vol. 5, Morazzoni F (ed). Polo Ed. Chimico: Milan, 1993.
46. Lu FC, Ralph J. *Plant J.* 2003; **35**: 535.
47. NMRPREDICT version 3.2.2, Mestelab Research. Available: http://www.modgraph.co.uk/product_nmr.htm.
48. El-Anead A, Banoub J. *Rapid Commun. Mass Spectrom.* 2005; **19**: 1683.
49. The chemical structures of all the individual analyzed protonated and deprotonated molecules; cations and anions; and radical cations and anions obtained by APPI-MS and APPI-QqToF-MS/MS experiments, which have been presented in this manuscript, including the established fragmentation patterns, are available upon request on CD from the authors.
50. Sack TM, Lapp RL, Gross ML, Kimble BJ. *Int. J. Mass Spectrom. Ion Processes* 1984; **61**: 191.
51. Forss KG, Fremer K-E. In *The Nature and Reactions of Lignin – A New Paradigm*, Forss KG, Fremer K-E (eds). Oy Nord Print Ab: Helsinki, 2003; 556.