Differences in the chemical structure of the lignins from sugarcane bagasse and straw

JOSÉ C. DEL RÍO^{*,1}, ALESSANDRO G. LINO^{1,2}, JORGE L. COLODETTE³, CLAUDIO F. LIMA², ANA GUTIÉRREZ¹, ÁNGEL T. MARTÍNEZ⁴, FACHUANG LU⁵, JOHN RALPH⁵ AND JORGE RENCORET¹

¹ Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, PO Box 1052, E-41080 Seville, Spain

² Department of Chemistry, Federal University of Viçosa, Viçosa, MG 36570-000, Brazil

³ Department of Forestry Engineering, Federal University of Viçosa, Viçosa, MG 36570-000, Brazil

⁴ Centro de Investigaciones Biológicas (CIB), CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

⁵ Departments of Biochemistry and Biological Systems Engineering, the Wisconsin Bioenergy Institute, and the DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wisconsin 53726, USA

* Corresponding author: (Tel: +34 954624711; Fax : +34 954624002; E-mail: delrio@irnase.csic.es)

1 ABSTRACT

2 Two major residues are produced by the sugarcane industry, the fibrous fraction following juice
3 extraction (bagasse), and the harvest residue (straw). The structures of the lignins from these
4 residues were studied by pyrolysis coupled to gas chromatography-mass spectrometry (Py-GC/MS),
5 nuclear magnetic resonance (NMR), and derivatization followed by reductive cleavage (DFRC).
6 Whereas the lignin from bagasse has a syringyl-rich *p*-hydroxyphenyl:guaiacyl:syringyl (H:G:S)
7 molar composition of 2:38:60, the lignin from straw is guaiacyl-rich (H:G:S of 4:68:28). The
8 compositional differences were also reflected in the relative abundances of the different interunit
9 linkages. Bagasse lignin was primarily β -O-4' alkyl-aryl ether units (representing 83% of NMR-
10 measurable units), followed by minor amounts of β -5' (phenylcoumarans, 6%) and other condensed
11 units. The lignin from straw has lower levels of β -ethers (75%) but higher relative levels of
12 phenylcoumarans (β -5', 15%) and dibenzodioxocins (5-5/4-O- β , 3%), consistent with a lignin
13 enriched in G-units. Both lignins are extensively acylated at the γ -hydroxyl of the lignin side-chain
14 (42% and 36% acylation in bagasse and straw), predominantly with *p*-coumarates (preferentially on
15 S-units) but also with acetates (preferentially on G-units) to a minor extent. Tetrahydrofuran
16 structures diagnostically arising from β - β -coupling (dehydrodimerization) of sinapyl *p*-coumarate
17 or its cross-coupling with sinapyl alcohol were found in both lignins, indicating that sinapyl *p*-
18 coumarate acts as a monomer participating in lignification. The flavone triclin was also found in the
19 lignins from sugarcane, as also occurs in other grasses.

20

21 **Keywords:** *Saccharum spp.*; NMR; DFRC; triclin; lignin acylation; *p*-coumarates

1. Introduction

Sugarcane (*Saccharum* spp. hybrids) is a perennial monocotyledonous plant belonging to the Gramineae (Poaceae) family that has received considerable recent attention due to its potential for biofuel production. The plant originates from Asia but it is well adapted to most tropical and subtropical climates where it has become one of the most important bioenergy crops. Its C4 photosynthesis makes it one of the most efficient species in carbon conversion and one of the most productive amongst all cultivated crops. Sugarcane is the main feedstock for sugar production, accounting for nearly two thirds of the world's production and is also the main raw material for the production of ethanol [1,2]. World sugarcane production was estimated to be around 1.91 Gt in 2013, about 0.77 Gt produced in Brazil alone, with an average productivity of 75 t ha⁻¹ [3].

During the harvest of sugarcane, the leaves and tops are left in field, whereas the stalks are transported to the mill where they are crushed to extract the sugar juice for the production of sugar and ethanol. Two major residues are produced by the sugarcane industry, the fibrous fraction following juice extraction (named bagasse), and the harvest residue (straw). These wastes are produced in large quantities, about 280 million metric tons of bagasse and straw per year [4], and they are expected to increase in the near future as this crop expands and new industrial plants are brought online.

The use of lignocellulosic biomass as a feedstock for the production of biofuels and other bio-based products, in the context of the so-called lignocellulosic biorefinery [5], has recently expanded, due in part to concerns about declining fossil fuel reserves, global climate change produced by greenhouse gas emissions from fossil fuels, and the need to secure more localized energy supplies. In this sense, there is an increasing trend towards the more efficient and rational use of agro-industrial wastes, including sugarcane bagasse and straw, as a means to maximize the use of agricultural resources. Sugarcane bagasse and straw are lignocellulosic materials basically composed of cellulose, hemicelluloses, and lignin, with lower amounts of extractives and ash, and are therefore attractive feedstocks for the production of new products, including the production of

second-generation bioethanol due to their high carbohydrate contents and their essential independence from competition with food/feed demands [4-9]. Currently, in most of the sugar industries, only the sugar juice is fermented to ethanol, whereas the bagasse and straw are mostly used for generating heat and electricity. The production of second-generation bioethanol from these 'waste' materials could be advantageous for the sugarcane industries as they can capitalize on the infrastructure already in place for the production of first-generation bioethanol. Converting bagasse and straw to ethanol requires an enzymatic saccharification of the polysaccharides to fermentable reducing sugars and their subsequent fermentation to ethanol or other liquid fuels. However, the architecture of the cell-wall, and particularly the presence of the recalcitrant lignin polymer, reduce the accessibility of the enzymes to the cellulose, and decrease the efficiency of the hydrolysis. Costly and harsh pretreatment processes are therefore needed to circumvent cell-wall recalcitrance by removing, or cleaving and redistributing, the lignin polymer [4-13]. Pretreatment is the most important limiting step in the production of ethanol from lignocellulosic materials. The efficiency of the pretreatment process is highly dependent on both the lignin content and structure, and therefore a knowledge of the structure of the lignin polymers in sugarcane bagasse and straw is important to develop appropriate pretreatment methods for delignification, as well as for potentially higher-value utilization of the lignin. However, bagasse and straw residues differ in their physical nature and chemical composition and the required delignification pretreatments will therefore be different for each residue; understanding the composition and structure of the lignins from sugarcane bagasse and straw is paramount.

Lignin is a complex aromatic heteropolymer produced by the combinatorial oxidative coupling of three main monolignols, *p*-coumaryl, coniferyl and sinapyl alcohols, differing in their degrees of methoxylation. When incorporated into the lignin polymer, these monolignols produce differently linked *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin units, respectively, generating a variety of structures within the polymer, including alkyl- β -aryl ethers, phenylcoumarans, resinols, spirodienones and dibenzodioxocins, among others [14]. The lignin composition varies between

74 plants from different taxa and even between different tissues from the same plant. Generally, lignin
75 from hardwoods is composed of S and G units in different ratios, whereas lignin from softwoods is
76 composed essentially of G units with minor amounts of H units, and lignins from grasses contain
77 the three units, with H-units still comparatively minor. In addition, the *p*-hydroxycinnamates (*p*-
78 coumarates and ferulates) widely occur in the grass polymers (hemicellulosic polysaccharides and
79 lignin), with *p*-coumarates acylating the γ -OH of the lignin side-chains, and predominantly on S-
80 units, whereas ferulates acylate arabinosyl residues of arabinoxylan chains and participate in both
81 polysaccharide-polysaccharide and lignin-polysaccharide cross-coupling reactions [15].

82 Despite the large effort devoted to the use of sugarcane bagasse and straw as feedstocks for the
83 production of second-generation bioethanol and other bioproducts [2,4,6-13], the structure of their
84 lignins has not been studied in detail to date, although some studies on the structure of bagasse
85 lignin have been published in recent years [16-18], and a systematic study of lignin deposition
86 during sugarcane stem development has also been recently performed [19]. Therefore, in this paper,
87 we report a more extensive and detailed structural characterization of the lignins in sugarcane
88 bagasse and straw. For this purpose, we used powerful analytical methodologies, including the
89 thermal degradative method, analytical pyrolysis coupled to gas chromatography and mass
90 spectrometry (Py-GC/MS), spectroscopic methods such as heteronuclear bidimensional solution-
91 state nuclear magnetic resonance (2D-NMR) spectroscopy, and the diagnostic chemical degradative
92 method, derivatization followed by reductive cleavage (DFRC), to delineate the compositional and
93 structural differences between these two lignins. Such data will help to maximize the exploitation of
94 these important agro-industrial wastes as feedstocks for the production of second-generation
95 bioethanol and other biobased products.

96

97 **2. Materials and methods**

98

99 *2.1. Samples*

100

101 The samples of sugarcane bagasse and straw were supplied by a mid-sized ethanol mill located
102 in the neighborhood of the São Pedro dos Ferros city, in the same-named County, in Minas Gerais
103 State, Brazil. Sugarcane plants were collected at the approximate age of 6 months from high-
104 performance sugarcane (*Saccharum spp.* hybrids) plantations (UFV/RIDESA RB867515 variety;
105 geographical coordinates are 20° 10' 0" South, 42° 31' 0" West, altitude of 373 m). This sugarcane
106 cultivar is named following international standard nomenclature, using the alphanumeric
107 designation in which the two first letters indicate the country (RB – Republic of Brazil), the two
108 first numbers (86) indicate the year of hybridization and the last four numbers (7515) indicate the
109 code initially assigned to the clone. The cultivar RB867515 was developed under the scientific
110 cooperation program between UFV (Universidade Federal de Viçosa, Brazil) and RIDESA (Rede
111 Interuniversitária para o Desenvolvimento do Setor Sucroenergético, Brazil), also called
112 UFV/RIDESA [1], and represents 69% of the sugarcane currently planted in Brazil. The sugarcane
113 plants were manually harvested and cleaned in the field where about one-third of their weight was
114 removed in the form of tops, limbs, and foliage; this material is generically called straw throughout
115 this study. The collected straw was immediately transported to the mill in trucks and crushed in a
116 hammer mill in order to homogenize the material well. The remaining two-thirds of the sugarcane
117 plants (stalks), comprised of the sugar itself (one-third = juice = sucrose) and bagasse (one third =
118 crushed stalks) were collected by suitable machines and immediately transported by truck to the
119 ethanol mill. The processing of sugarcane stalks involved two steps. The first was the crushing in a
120 semi-industrial chipper, with the fragments being processed by milling to extract the juice. The
121 resulting residue (bagasse) was used for subsequent characterization studies. Sugarcane bagasse and
122 straw samples were transported to the laboratory where they were air-dried and milled using an IKA
123 knife mill (Janke & Kunkel, Staufen, Germany) and then stored in wide-mouth glass jars until
124 required for analysis.

125 The extractives content (acetone-extractives and water-soluble material) was measured by
126 extracting ~50 g of sugarcane bagasse or straw with acetone in a Soxhlet for 16 h and then with hot
127 water at 100 °C for 3 h. The lignin content (estimated as Klason lignin) was measured
128 gravimetrically as the insoluble residue from the sulphuric acid hydrolysis of the pre-extracted
129 material according to the TAPPI method T222 om-8 [20]. The Klason lignin content was then
130 corrected for ash, and for protein as determined from the mass fraction of atomic nitrogen obtained
131 by CHN elemental analysis in a LECO CHNS-932 Elemental Analyzer (LECO Corp., St. Joseph,
132 Mich.) and using a 6.25 factor [21]. After filtration of the insoluble lignin, the content of acid-
133 soluble lignin was measured by UV-spectroscopic determination at 205 nm using $11,000 \text{ m}^2 \text{ kg}^{-1}$ as
134 the extinction coefficient, according to Tappi method UM 250 [20]. The content of α -cellulose and
135 hemicelluloses was determined after isolation of the holocellulose from pre-extracted samples by
136 delignification for 4 h using the acid chlorite method [22]. The content of α -cellulose was
137 determined after removing the hemicelluloses by alkaline extraction [22]. Finally, the ash content
138 was estimated as the residue after 6 h of heating at 575 °C under air in a muffle furnace. Three
139 replicates were used for each sample.

140

141 2.2. Ball milling and isolation of 'milled-wood' lignins (MWLs)

142

143 The MWLs were obtained according to the classical Björkman procedure [23]. Around 40 g of
144 extractive-free material were finely ball-milled in a Retsch PM100 planetary ball mill (Restch,
145 Haan, Germany) at 6.67 Hz (36 h) using an agate jar and 20 spheroid agate balls of 2 cm diameter
146 (this ball-milled material was used directly for whole-cell-wall-NMR to determine compositional
147 and structural features of the entire lignin fraction). The ball-milled material was then extracted
148 with dioxane-water, 96:4 (v/v) (20 cm^3 of solvent/g of milled fiber), and the isolated MWL was
149 subsequently purified as described elsewhere [24]. The final yields ranged from 10-15% of the
150 original Klason lignin content.

151

152 *2.3. Pyrolysis coupled to gas chromatography and mass spectrometry (Py-GC/MS)*

153

154 Pyrolysis of the whole-cell-walls and of their isolated MWLs (ca. 0.1 mg) was performed at
155 500 °C in an EGA/PY-3030D micro-furnace pyrolyzer (Frontier Laboratories Ltd., Fukushima,
156 Japan) connected to a GC 7820A (Agilent Technologies, Inc., Santa Clara, CA) and an Agilent
157 5975 mass-selective detector (EI at 70 eV). The column used was a 60 m x 0.25 mm i.d., 0.25 µm
158 film thickness, DB-1701 (J&W Scientific, Folsom, CA). The oven temperature was programmed
159 from 45 °C (4 min) to 280 °C (10 min) at 4 K min⁻¹. The carrier gas was Helium (2 cm³ min⁻¹). For
160 the pyrolysis in the presence of tetramethylammonium hydroxide (Py/TMAH), around 0.1 mg of
161 sample was mixed with 0.5 mm³ of TMAH (25%, w/w, in methanol), the pyrolysis was carried out
162 at 500 °C, and the released products analyzed as described above. Identification of the released
163 compounds was made by comparison of their mass spectra with those of the Wiley and NIST
164 libraries and with those reported in the literature [25] and, when possible, by comparison with the
165 retention times and mass spectra of authentic standards. Molar peak areas were calculated for each
166 lignin degradation product released, the summed areas were normalized, and the data for two
167 replicates were averaged and expressed as percentages.

168

169 *2.4. Nuclear magnetic resonance (NMR) spectroscopy*

170

171 2D-NMR spectra (heteronuclear single-quantum coherence, HSQC, and heteronuclear
172 multiple-bond correlation, HMBC) experiments were recorded on an AVANCE III 500 MHz
173 instrument (Bruker, Karlsruhe, Germany) at the NMR facilities of the General Research Services of
174 the University of Seville. For the NMR of the whole-cell-walls of sugarcane bagasse and straw,
175 around 50 mg of finely divided (ball-milled) extractive-free samples were swollen in 0.75 cm³
176 DMSO-*d*₆ according to the method previously described [26,27]. In the case of the isolated MWLs,

177 around 40 mg of underivatized or acetylated sample were dissolved in 0.75 cm³ of DMSO-*d*₆ or
178 0.75 cm³ of CDCl₃, respectively. The central solvent peaks were used as internal references (δ_C/δ_H
179 39.5/2.49 for DMSO and 70.0/7.26 for CDCl₃/CHCl₃). The detailed NMR experimental conditions
180 have been described elsewhere [28]. HSQC cross-signals were assigned by comparison with
181 literature data [24,26-31]. A semi-quantitative analysis of the volume integrals of the HSQC cross-
182 correlation signals was performed as previously described [28].

183

184 2.5. Derivatization followed by reductive cleavage (DFRC)

185

186 The DFRC degradation was performed according to the method previously developed by Lu
187 and Ralph [32-34], and the detailed explanation of the experimental protocol can be found
188 elsewhere [24,28]. Briefly, around 5 mg of lignin samples were stirred for two hours at 50 °C with
189 acetyl bromide in acetic acid, 8:92 (v/v) and then treated with powdered Zn (50 mg) for 40 min at
190 room temperature. The lignin degradation products were then acetylated for 1 h in 1.1 cm³ of
191 dichloromethane containing 0.2 cm³ of acetic anhydride and 0.2 cm³ of pyridine. The acetylated
192 lignin degradation products were analyzed by GC/MS using mass spectra and relative retention
193 times to authenticate the DFRC monomers and their *p*-coumarate conjugates as described [32-34].
194 In order to assess the presence of naturally acetylated lignin units, the DFRC method modified to
195 use propionylating reagents (so-called DFRC') was used, as previously described [35,36].

196 The lignin degradation products obtained after DFRC and DFRC' were analyzed using a
197 GCMS-QP2010plus instrument (Shimadzu Co., Kyoto, Japan) using a capillary column (DB-5MS
198 30 m × 0.25 mm I.D., 0.25 µm film thickness). The oven was heated from 140 °C (1 min) to 250 °C
199 at 3 K min⁻¹, then ramped at 10 K min⁻¹ to 280 °C (1 min) and finally ramped at 20 K min⁻¹ to 300
200 °C, and held for 18 min at the final temperature. The injector was set at 250 °C and the transfer line
201 was kept at 310 °C. The carrier gas was Helium at a rate of 1 cm³ min⁻¹. Quantitation of the released
202 (acetylated and propionylated) monomers was performed using tetracosane as internal standard and

by assuming response factors similar to those of the acetylated monomers reported previously [32]. The molar yields of the released lignin degradation products were determined on the basis of the molecular weights of the respective acetylated or propionylated compounds.

3. Results and discussion

3.1. Composition of the main constituents of sugarcane bagasse and straw

The abundances of the main constituents (water-soluble material, acetone extractives, Klason lignin, acid-soluble lignin, holocellulose, α -cellulose, and ash) of the sugarcane bagasse and straw samples selected for this study are shown in **Table 1**. The composition is in the range of values previously published for similar samples, except for the much higher content of extractives reported in previous papers [6,37]. However, those papers used 95% ethanol as solvent, which extracts more polar compounds; in addition, some papers also included the water-soluble material in the total extractives content, thus resulting in higher values. The bagasse and straw samples have a high content of carbohydrates (75.8% and 72.9%) and substantial lignin contents (20% and 18.9%). However, it is worth noting that this composition may vary depending of the cultivar, crop location, year of harvesting, and age of the crop plants, as well as depending of other environmental and cultivation conditions. In this work, the composition of the lignins in both sugarcane residues was analyzed *in situ* by Py-GC/MS and 2D-NMR. To obtain cleaner lignin spectra, MWLs were isolated from the sugarcane residues using a classical lignin isolation protocol [23] and analyzed by Py-GC/MS, 2D-NMR and DFRC.

3.2. Lignin composition of sugarcane bagasse and straw as determined by Py-GC/MS

228 The whole-cell-walls of sugarcane bagasse and straw and their respective MWLs were first
229 analyzed by Py-GC/MS (**Figure 1**). The identities of the released phenolic compounds and their
230 relative molar abundances are listed in **Table 2**. Pyrolysis released compounds derived from
231 carbohydrates as well as phenolic compounds derived from lignin and from *p*-hydroxycinnamates.
232 Among the phenolics released, the pyrograms showed compounds derived from the H, G and S
233 lignin units such as guaiacol **2**, 4-methylphenol **4**, 4-methylguaiacol **5**, syringol **11**, 4-
234 methylsyringol **15**, 4-vinylsyringol **23** and *trans*-4-propenylsyringol **29**. High amounts of 4-
235 vinylphenol **8** and 4-vinylguaiacol **9** were also released but, as usually occurs in the pyrolysis of
236 grasses, they mostly arise from *p*-coumarates and ferulates after decarboxylation [24,28,38,39].
237 Pyrolysis of the isolated MWLs released a similar distribution of phenolic compounds except for
238 the much lower relative abundances of 4-vinylguaiacol **8**; this compound is largely derived from
239 ferulates that are attached to polysaccharides and is therefore essentially absent from the MWLs.
240 The most prominent compound released was 4-vinylphenol **9**, derived essentially from *p*-coumarate
241 esters on MWL.

242 *p*-Hydroxycinnamates esterify to polysaccharides and/or lignin and are particularly abundant in
243 grasses [15,40-43]. As their efficiently released pyrolysis products can be conflated with those from
244 lignins, care must be exercised when estimating the lignin composition by pyrolysis. *p*-Coumarate-
245 and ferulate-derived 4-vinylphenol and 4-vinylguaiacol were therefore not be used for the
246 evaluation of the lignin H:G:S composition. The analysis revealed strong differences in the
247 composition of both lignins, with the lignin of bagasse being S-rich (S/G molar ratio of 1.3–1.6) and
248 the lignin in straw being G-rich (S/G molar ratio of 0.5) (**Table 2**). This is fairly readily explained
249 by the average maturity of the tissues in each as mature material is always richer in S-units; bagasse
250 is mainly derived from the mature stems, whereas the straw includes the leaves and other immature
251 parts of the plant.

252 The occurrence of *p*-hydroxycinnamates in sugarcane bagasse and straw was confirmed by
253 pyrolysis in the presence of tetramethylammonium hydroxide (TMAH), a reagent that prevents

254 decarboxylation during pyrolysis and liberates diagnostic methyl esters from the esterified
255 components [38,44]. **Figure 2** shows the Py/TMAH chromatograms of the whole cell-walls of
256 bagasse and straw, and of their respective MWLs. The compounds released and their relative molar
257 abundances are listed in **Table 3**. In general terms, Py/TMAH induces cleavage of alkyl-aryl ether
258 bonds in lignin and releases the methylated derivatives of hydroxybenzaldehydes (**4**, **11**, **17**),
259 hydroxyacetophenones (**14**, **20**) and hydroxybenzoic acids (**7**, **13**, **21**), among others [38,44,45]. As
260 is consistent with the comments above regarding the derivation of 4-vinylphenol and 4-
261 vinylguaiacol from the hydroxycinnamates, Py/TMAH of the whole cell-walls released high
262 amounts (43.8% and 38.4% of the total molar peak areas in bagasse and straw) of the fully
263 methylated derivative of *p*-coumaric acid, the *trans*-3-(4-methoxyphenyl)-propenoic acid methyl
264 ester **19**, as well as significant amounts (8.5% and 13%) of the fully methylated derivative of ferulic
265 acid, the *trans*-3-(3,4-dimethoxyphenyl)-propenoic acid methyl ester **26**. Py/TMAH of the isolated
266 MWLs released higher relative amounts of *p*-coumarate derivatives (68.7% and 54.8% from
267 bagasse and straw), and considerably lower amounts of the ferulate derivatives (2.3% and 4.5%). *p*-
268 Coumarates and ferulates were therefore present in important amounts in the whole cell-walls, with
269 a *p*-coumarate/ferulate molar ratio of 5.2 and 3.0, respectively (**Table 3**). However, this ratio
270 sharply increases to 31.1 and 12.8 in the MWLs, indicating that in both fractions, ferulates are
271 mostly attached to the carbohydrates in the cell-wall whereas *p*-coumarates are primarily attached to
272 the lignin polymer, as also occurs in other grasses [15]. Previous studies have shown that *p*-
273 coumarates in plants, including grasses, primarily acylate the γ -OH of the lignin side-chain, and
274 predominantly on S-units, whereas ferulates acylate cell-wall polysaccharides and participate in
275 both polysaccharide-polysaccharide and lignin-polysaccharide cross-coupling reactions [15,31,40-
276 43,46].

277

278 *3.3 Lignin structural units and inter-unit linkages in sugarcane bagasse and straw as elucidated by*

279 *2D-NMR*

280

281 Additional information regarding the structural characteristics of the lignins is available from
282 2D-NMR analysis of the whole-cell-walls and their isolated MWLs. The HSQC spectra of the
283 whole cell-walls were acquired by swelling finely ball-milled samples in DMSO-*d*₆ and forming a
284 gel in the NMR tube according to the method previously published [26,27], **Figures 3** and **4**.
285 Carbohydrate signals from xylans (X₂, X₃, X₄, X₅), including acetylated xylans (X'₂ and X'₃), were
286 predominant in the aliphatic side-chain region of the whole-cell-wall spectra, partially overlapping
287 with lignin signals. The spectra of the isolated MWLs, however, exhibited predominantly lignin
288 signals that matched those present in the HSQC spectra of their respective whole-cell-walls but that
289 are better resolved. The lignin and carbohydrate cross-signals assigned in the HSQC spectra are
290 listed in **Table 4** and the main lignin substructures are drawn in **Figure 5**.

291 In addition to the commonly observed lignin structural units (**Figure 5**), the occurrence of
292 intense signals in the range δ_C/δ_H 62.7/3.83–4.30, assigned to the C_γ/H_γ correlations of γ-acylated
293 units (**A'**), revealed that the lignins from sugarcane bagasse and straw are extensively acylated at the
294 γ-position of the lignin side-chain. The extent of lignin acylation was estimated in the HSQC
295 spectra of the isolated MWLs (where the signals from carbohydrates do not interfere) by integration
296 of the C_γ/H_γ signals from the normal (γ-OH) and the γ-acylated lignin units and varied from 42% in
297 the bagasse lignin to 36% in the straw lignin (**Table 5**), significantly higher than the 10.6%
298 previously reported [18]. The aliphatic side-chain region of the spectra also showed other prominent
299 signals corresponding to the C_α/H_α and C_β/H_β correlations of β-O-4' substructures (**A**, **A'**).
300 Interestingly, the HSQC spectrum of the MWL from the straw showed the presence of a signal at
301 δ_C/δ_H 80.8/4.58 (that was absent in the bagasse) characteristic for the C_β/H_β correlations of γ-
302 acylated β-O-4' substructures **A'** linked to G-units [24,28], indicating a significant degree of γ-
303 acylation of G-lignin units in the straw lignin.

304 Signals from other lignin substructures were also observed in the HSQC spectra, being more
305 evident in the isolated MWLs, and included signals for phenylcoumarans **B**, resinols **C**,

306 dibenzodioxocins **D** and spirodienones **F**. Signals for α,β -diaryl ethers **E** could also be detected in
307 the HSQC of the MWL from bagasse, although with low intensities, as revealed by the C_α/H_α
308 correlation at δ_C/δ_H 79.5/5.50, but were absent in the MWL from straw. In addition, signals for a
309 β - β' coupled tetrahydrofuran substructure **C₃**, arising from β - β' coupling of two γ -acylated sinapyl
310 alcohols, were observed in the HSQC spectrum of the MWL from bagasse, and in lower intensities
311 from straw, with their characteristic C_α/H_α and C_β/H_β correlations being at δ_C/δ_H 82.8/5.00 and
312 49.8/2.59. This tetrahydrofuran structure has also been observed in the lignins from other plants,
313 including other grasses such as elephant grass and corn [24,47]. Other signals in this region of the
314 spectra corresponded to the C_γ/H_γ correlations of cinnamyl alcohol end-groups **I** and γ -acylated
315 cinnamyl alcohol end-groups **I'**, as well as to the C_β/H_β correlations of α -keto- β -*O*-4' alkyl-aryl
316 ethers (**A_{ox}**).

317 In the aromatic/unsaturated region of the HSQC spectra are the main correlation signals
318 corresponding to the aromatic rings of the different lignin units (**H**, **G**, **S**), including C α -oxidized S-
319 lignin units **S'**, as well as to the aromatic rings and the unsaturated side-chains of the *p*-
320 hydroxycinnamates (**PCA**, **FA**). The signals for H-lignin units were only observed in low
321 intensities, confirming that the high amounts of 'H-units' released upon Py-GC/MS were mostly
322 due to *p*-coumarates. In addition, in this region of the HSQC spectra appeared the two distinctive
323 signals at δ_C/δ_H 94.0/6.56 and 98.7/6.22 corresponding to the C_8/H_8 and C_6/H_6 correlations of triclin
324 **T** [28]. Tricin is a flavone that is now widely found, and is apparently incorporated, into the lignin
325 structure in grasses [28,39,48-50] and that also occurs in the lignin of other monocotyledons, such
326 as coconut coir [51]. The HSQC spectra also show the C_3/H_3 correlation of triclin at δ_C/δ_H 104.5/7.03
327 as well as the correlations for $C_{2'}/H_{2'}$ and $C_{6'}/H_{6'}$ at δ_C/δ_H 103.9/7.30. Other signals in
328 aromatic/unsaturated region of the HSQC spectra are from cinnamyl alcohol end-groups **I** and
329 cinnamaldehyde end-groups **J**. In addition, a signal for the $C_{2,6}/H_{2,6}$ correlation of sinapaldehyde
330 end-groups **J_S**, was observed at δ_C/δ_H 106.8/7.06; however, no aromatic signals for coniferaldehyde
331 end-groups were detected in the spectra. Signals for γ -acylated cinnamyl alcohol end-groups **I'** were

also detected in this region of the spectra, with the characteristic C_α/H_α and C_β/H_β correlations observed at δ_C/δ_H 132.9/6.60 and 123.2/6.38. These signals were particularly evident in the HSQC spectrum of the MWL from bagasse, and derive from its higher extent of γ -acylation, but were present in lower amounts in the HSQC of the MWL from straw, with a lower degree of γ -acylation.

Some authors [18] have also reported the occurrence of *p*-hydroxybenzoates in the lignin from sugarcane bagasse based on the presence of a signal in the HSQC spectra at δ_C/δ_H 132.4/7.63, which is also observed in the HSQC of our samples. However, a close comparison with the HSQC spectra of other lignins from our own collection with proven occurrence of *p*-hydroxybenzoates indicates that this assignment was not correct as the $C_{2,6}/H_{2,6}$ correlation signal for *p*-hydroxybenzoates that should appear around δ_C/δ_H 131.2/7.68 [51] does not match with that signal. The absence of other known correlations in the HMBC spectra also discards this signal as belonging to *p*-hydroxybenzoates. *p*-Hydroxybenzoates have been widely found acylating the γ -OH of the lignin in palms and *Populus* species (willow, aspen, poplar) [47,51-56] but have never been found in grasses, and therefore, in addition to our not observing methylated *p*-hydroxybenzoate in the Py/TMAH analysis, we can dismiss the occurrence of *p*-hydroxybenzoates in the lignins from sugarcane.

The structural characteristics of the lignins from sugarcane bagasse and straw, such as the relative abundances of the different interunit linkages and cinnamyl end-groups, the percentage of γ -acylation, the molar abundances of the lignin units (H, G, and S), *p*-coumarates, ferulates, and tricinn, estimated from volume integration of the signals in the HSQC spectra, are indicated in **Table 5**. The data confirmed the completely different composition of the two lignins noted above from the Py-GC/MS data. The H:G:S molar composition of the lignin from bagasse (2:38:60) was quite different to that from the straw (4:68:28). The S/G molar ratios obtained by 2D-NMR closely matched those determined by Py-GC/MS, indicating a predominance of S units in the lignin from sugarcane bagasse (S/G molar ratio of 1.6) and a predominance of G units in the lignin from sugarcane straw (S/G molar ratio of 0.4). This data also confirms that the content of H-units in the

lignins from both sugarcane residues is quite low (<5%, **Table 5**), and that the high abundance of *p*-hydroxyphenyl compounds observed upon Py-GC/MS was due to pyrolytic decarboxylation of *p*-coumarates. 2D-NMR data indicated that *p*-coumarates are present in high abundance in the whole-cell-walls, as well as in their isolated MWLs, whereas ferulates are present in much lower abundance in the isolated MWLs than in the respective whole-cell-walls, as already observed by Py-GC/MS and Py/TMAH. This is because, as already indicated above, ferulates are primarily attached to the polysaccharides whereas *p*-coumarates are predominantly attached to the lignin. This information, together with the high extent of acylation of the γ -OH of the lignin side-chain observed in these lignins (42% and 36% of lignin acylation in bagasse and straw), seems to indicate that *p*-coumarates are the groups acylating the γ -OH of the lignin side-chain, as has been observed in the lignins of other plants [15,24,28,31,34,40,43,46]. The flavone triclin was also present in both lignins, being more abundant in the straw. Triclin has been suggested to act as a lignin monomer in grasses and its incorporation into the lignin polymer through 4-*O*- β coupling reactions has been indicated [28], and such coupling has been validated in biomimetic coupling reactions [50]. Triclin cannot couple with another triclin molecule, so it can only undergo cross-coupling reactions with monolignols and therefore must be present at the start of a lignin chain; triclin therefore seems to act as a nucleation or initiation site for lignification [28,50].

The differences in composition between the lignins from sugarcane bagasse and straw are also reflected in the differences in the relative abundances of the various interunit linkages between them. Thus, the bagasse lignin is mostly made up of β -*O*-4' alkyl-aryl ether linkages (accounting for 83% of all the interunit linkage types measured), followed by minor amounts of β -5' phenylcoumarans (6%) and other condensed units. The straw lignin, on the other hand, has lower levels of β -*O*-4' alkyl-aryl ether linkages (75% of all the interunit linkages) and has higher amounts of β -5' phenylcoumarans (15%) and dibenzodioxocins (3%), as is consistent with a lignin enriched in G-units. Minor amounts of α,β -diaryl ethers (2%) were also found in the bagasse lignin, but were absent from the straw. Other condensed substructures, such as β - β' resinols and β -1' spirodienones,

are present in minor amounts in both lignins. The unusually low proportions of β - β' resinol substructures observed in these lignins, accounting for only 2% of all interunit linkages in the bagasse lignin and 3% in the straw lignin, seem to be related to the high extent of acylation of the γ -OH in these lignins, as also noted in other highly acylated lignins [24,31]. This is because the formation of normal β - β' resinol structures requires a free γ -OH in order to rearomatize the intermediate quinone methide (following the radical dehydrodimerization step). If the γ -OH of a monolignol is acylated, normal resinol structures cannot be formed, and instead, unconventional β - β' tetrahydrofuran structures are formed from β - β' homocoupling of two γ -acylated monolignols or from β - β' cross-coupling of a normal (γ -OH) monolignol with a γ -acylated monolignol [47,55-57]. Interestingly, significant amounts of a β - β' tetrahydrofuran structure (**C**₃) formed by β - β' homocoupling of two γ -acylated monolignols were found in these lignins, being particularly abundant in the lignin from sugarcane bagasse (4% of all interunit linkages), as is consistent with its higher level of γ -acylation (42% of all side-chains).

397

3.4. Details of lignin acylation by *p*-coumarate (and more minor levels of acetate) from 2D-NMR and DFRC

400

p-Coumarates are widely present in grasses, and in other plants, acylating the γ -OH of the lignin side-chain [15,24,28, 31,34,40,43,46]. In order to demonstrate that the *p*-coumarates present in the lignins from sugarcane bagasse and straw are of the same type, we performed 2D-NMR HMBC experiments that correlate protons with carbons separated by 2- or 3-bonds. **Figure 6** shows the MWL HMBC sections for the carbonyl carbon correlations. Strong signals are observed for the *p*-coumarate ester carbonyl correlations at δ_C 166.0 with their α - and β -protons at δ_H 7.41 and 6.27, and with protons in the range δ_H 4.0-4.8 ppm in both HMBC spectra confirming that *p*-coumarate groups acylate the γ -OH of the lignin side-chains, as usually occurs in grasses. The HMBC spectra, however, did not show similar correlations for acetate groups (at δ_C 169.8), despite previous

410 findings that lignin γ -acetates are widely found, including in grasses [24,28,31,35,36,58].
411 Apparently, the extent of γ -acetylation, if any, is below the detection limit of the HMBC experiment
412 here.

413 Additional information regarding the nature of the acylation of the γ -OH was obtained from
414 DFRC, a degradation method that cleaves (α - and) β -ether bonds but leaves γ -esters intact [32-34].
415 **Figure 7** shows the chromatograms of the DFRC degradation products obtained from the MWLs
416 from sugarcane bagasse and straw. The DFRC of the isolated MWLs produced the *cis*- and *trans*-
417 isomers of the acetylated *p*-hydroxyphenyl (*t*H), guaiacyl (*c*G and *t*G), and syringyl (*c*S and *t*S)
418 lignin monomers arising from normal, non- γ -*p*-coumaroylated β -ether lignin units. In addition, the
419 chromatograms revealed the presence of important peaks corresponding to the *cis*- and *trans*-
420 isomers of S-lignin units acylated with *p*-coumarates. Both saturated (*cS_{pC}* and *tS_{pC}*) and
421 unsaturated (*cS_{pC'}* and *tS_{pC'}*) DFRC-derived products were particularly abundant in the MWL from
422 sugarcane bagasse, together with minor amounts of the guaiacyl analogs (*cG_{pC}* and *tG_{pC}*). The
423 release of these compounds confirmed that *p*-coumarate groups are acylating the γ -OH of these
424 lignins, and predominantly on S units, as usually occurs in grasses [15,24,28,31,34,40,43,46].
425 Because the original DFRC degradation method masked the presence of natural acetates, a
426 modification in which acetylating reagents are substituted for their propionylating counterparts
427 (DFRC') is used to reveal native lignin acetylation [35,36]. By this method it was possible to detect
428 the presence of minor amounts of γ -acetylated G- and S-lignin units in the MWL isolated from both
429 sugarcane bagasse and straw, demonstrating that native acetylation of the γ -OH also occurred in
430 these lignins, although to a low extent.

431 **Table 6** presents the results from the DFRC and DFRC' degradations of the MWLs isolated
432 from sugarcane bagasse and straw, including the molar yields of the released monomers (H, G, G_{ac},
433 G_{pC}, S, S_{ac}, S_{pC}), as well as the percentages of naturally acetylated guaiacyl (%G_{ac}) and syringyl
434 (%S_{ac}) and *p*-coumaroylated guaiacyl (%G_{pC}) and syringyl (%S_{pC}) lignin units (the values for S_{pC}
435 include the sum of both the saturated and the unsaturated *p*-coumarate analogs). The analyses

confirm the high level of γ -acylation in the lignin of sugarcane bagasse, that are mostly due to *p*-coumarates that are predominantly attached to S-units (23.3% of the S-units and only 1.5% of G-units are *p*-coumaroylated). Lower amounts of γ -*p*-coumaroylated guaiacyl (0.8%) and syringyl (10.9%) conjugates were released in the DFRC from the MWL from sugarcane straw, despite the substantial amounts of *p*-coumarates present in this lignin (21% with respect to the total lignin aromatic units, as estimated by 2D-NMR, **Table 5**) and the fact that they are acylating the γ -OH, as seen by HMBC. Similar results were also observed in the lignin from wheat straw [28], where only traces of γ -*p*-coumaroylated syringyl conjugates were released, despite the occurrence of significant amounts of *p*-coumarates being attached to the γ -OH. This finding seems to indicate that, as also occurred in the lignin from wheat straw, *p*-coumarate groups mostly acylate the γ -OH in condensed lignin substructures that are not amenable to DFRC analysis, a slightly troubling observation that requires further study. This is however supported by the below-noted observation of the products of sinapyl *p*-coumarate homodimerization units **C₃**. The assumption is also supported by the presence of an intense signal for γ -acylated β -O-4' substructures linked to G-units in the HSQC spectrum of the MWL from sugarcane straw (**Figure 4B**) that indicates a significant γ -acylation of G-units in this lignin, and which is absent in the spectrum of the MWL from sugarcane bagasse (**Figure 3B**). The DFRC' data also revealed a minor extent of γ -acylation of these lignins with acetate groups, but which takes place predominantly on G units (3.7% and 3.3% of G-lignin units were acetylated in the lignins from bagasse and straw, respectively) rather than on S units (only 0.8% and 1.8%). Low levels of lignin acetylation, with a preference for G-units, have also been found in other grasses, such as bamboo, wheat straw, or the pith of elephant grass [24,28,36]. In contrast, in the lignin of most plants, γ -acetylation occurs predominantly on S-units [31,36,58]. This fact seems to indicate that as yet unknown acetyl transferases with a higher affinity towards coniferyl alcohol than towards sinapyl alcohol, contrary to what occurs in most plants, are probably involved in monolignol acetylation in grasses.

462 3.5. Diversity of β - β (resinol and tetrahydrofuran) structures in the lignins from sugarcane bagasse
463 and straw

464

465 The HSQC spectra of the lignins from sugarcane bagasse and straw clearly show the presence
466 of significant amounts of a tetrahydrofuran structure (**C₃**) arising from β - β coupling of two γ -
467 acylated monolignols. As *p*-coumarates are the main group acylating the lignin, and predominantly
468 on S-units, it is evident that the tetrahydrofuran structure **C₃** identified in the HSQC spectra has
469 been produced from the β - β' -coupling of two sinapyl *p*-coumarate monomers, as shown in **Figure**
470 **8**. These units have been found in the lignins from other grasses, such as maize and elephant grass
471 [24,47], and analogous structures bearing acetates or *p*-hydroxybenzoates have also been identified
472 in the lignins from other plants [47,55,56,59].

473 Besides tetrahydrofuran structure **C₃**, the β - β coupling and cross-coupling of sinapyl alcohol
474 and γ -*p*-coumaroylated sinapyl alcohol can form non-symmetrical tetrahydrofurans **C₁** and **C₂**, as
475 detailed in **Figure 8**. A wide variety of tetrahydrofuran structures arising from β - β coupling and
476 cross-coupling of normal γ -OH and γ -acylated (with acetates, *p*-coumarates and *p*-
477 hydroxybenzoates) monolignols, have been synthesized in past years and their different correlation
478 signals assigned in the HSQC spectra [47,55,56,59]. In those papers, the different synthetic
479 structures were acetylated to improve the spectral properties. Therefore, in order to have detailed
480 information of the different β - β resinol and tetrahydrofuran structures present in the lignins here,
481 the MWLs were acetylated and the HSQC correlation signals were compared with those previously
482 published [47,55,56,59]. **Figure 9** shows the 2D-NMR HSQC spectra of the (acetylated) MWLs
483 from sugarcane bagasse and straw, with the C_α/H_α and C_β/H_β and C_γ/H_γ correlations of the different
484 resinol and tetrahydrofuran structures arising from the β - β coupling and cross-coupling of sinapyl
485 alcohol and sinapyl *p*-coumarate. The signals observed closely match those previously reported for
486 synthesized model compounds and related lignins and provide compelling evidence for the
487 occurrence of the various tetrahydrofuran structures arising from the β - β homo-coupling of sinapyl

488 *p*-coumarate (**C**₃) as well as its cross-coupling with sinapyl alcohol (**C**₁ and **C**₂), together with the
489 presence of traditional β–β resinol structures (**C**). The occurrence of the tetrahydrofuran structures
490 **C**₁–**C**₃ in these lignins conclusively demonstrates that the sinapyl *p*-coumarate conjugate acts as a
491 monomer participating in coupling and cross-coupling reactions during lignification of sugarcane,
492 and therefore implicates the presence of corresponding *p*-coumaroyl-CoA:monolignol transferases
493 (PMTs) in this plant. In fact, the PMTs involved in the *p*-coumaroylation of monolignols have been
494 discovered in corn [43], and in the model grass *Brachypodium distachyon* [60] in which a candidate
495 gene has been recently identified and its action validated [61].

496

497 **4. Conclusions**

498

499 This paper describes the differences in composition and structure of the lignins from sugarcane
500 bagasse and straw, the two major residues produced from the sugarcane industry. Whereas the
501 lignin from sugarcane bagasse is S-rich (H:G:S molar ratio of 2:38:60), the lignin from sugarcane
502 straw is G-rich (4:68:28). The compositional differences are also reflected in differences in the
503 relative abundances of the various interunit linkages in both lignins. Thus, the lignin from bagasse
504 is mostly made up by alkyl-aryl ether units (β–O–4', accounting for 83% of all the measured
505 interunit linkage types), followed by minor amounts of phenylcoumarans (β–5', 6%) and other
506 condensed units. The lignin from straw, on the other hand, has lower amounts of alkyl-aryl ether
507 units (75% of all interunit linkages) and has higher amounts of condensed structures such as
508 phenylcoumarans (15%) and dibenzodioxocins (5–5'/4–O–β', 3%), as is consistent with a lignin
509 enriched in G-units. The degree of lignin acylation by *p*-coumarate and acetate also differs
510 strikingly, being higher in the bagasse. The structural differences observed in the lignins of the two
511 sugarcane residues, bagasse from the stems after expelling the sugar, and straw from the residual
512 plant (including leaves), indicates that the pretreatment methods needed to reduce the recalcitrance
513 of the cell wall will have to be optimized for each residue. In particular, the structures of the lignins

514 in both residues suggest that the bagasse material will pretreat more easily (because of its higher
515 syringyl content and lower condensation degree) than the straw (with a higher content of condensed
516 structures). In addition, the two residues contain, as well as the polysaccharides that may be
517 saccharified and fermented to liquid biofuels and lignins that might be valorized, significant levels
518 of *p*-coumarate, ferulate, and triclin that might be valuable to extract as commodity chemicals.
519 Optimizing the utilization of all of the plant biomass that is sustainably available will improve
520 biorefinery operations and help to offer fuels and chemicals with a reduced carbon footprint
521 compared to fossil-derived products.

522

523 **Acknowledgements**

524

525 This study has been funded by the Spanish projects AGL2011-25379, AGL2014-53730-R and
526 CTQ2014-60764-JIN (co-financed by FEDER funds), the CSIC project 2014-40E-097 and the EU-
527 projects LIGNODECO (KBBE-2009-3-244362) and INDOX (KBBE-2013-7-613549). John Ralph
528 was funded through the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science
529 DE-FC02-07ER64494). Jorge Rencoret thanks the CSIC for a JAE-DOC contract of the program
530 “Junta para la Ampliación de Estudios” co-financed by Fondo Social Europeo (FSE). A.G. Lino
531 thanks CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for financial
532 support. C.F. Lima and J.L. Colodette are grateful to National Council for Scientific and
533 Technological Development (CNPq) for research fellowships. We also thank Dr. Manuel Angulo
534 for performing the NMR analyses that were acquired on a Bruker Avance III 500 MHz instrument
535 from the NMR facilities of the General Research Services of the University of Seville (SGI-
536 CITIUS).

537

538 **References**

- 539 [1] Santos F, Borém A, Caldas C. Sugarcane: Bioenergy, Sugar and Ethanol – Technology and
540 Prospects. Viçosa, Brazil: Universidade Federal de Viçosa; 2012.
- 541 [2] de Souza AP, Grandis A, Leite DCC, Buckeridge MS. Sugarcane as a bioenergy source:
542 history, performance and perspectives for second generation bioethanol. *Bioenerg Res* 2014
543 Aug; 7: 24–35.
- 544 [3] FAOSTAT [database on the Internet]. Rome - Italy: Food and Agriculture Organization of the
545 United Nations; 2015 - [cited 2015 Jun 10] Production, Crops, Sugar Cane; Available from
546 <http://faostat3.fao.org> Files updated Annually.
- 547 [4] Ortiz PS, de Oliveira S. Exergy analysis of pretreatment processes of bioethanol production
548 based on sugarcane bagasse. *Energy* 2014 Nov; 76: 130–8.
- 549 [5] Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The path
550 forward for biofuels and biomaterials. *Science* 2006; 311(5760): 484–9.
- 551 [6] Andrade MF, Colodette JL. Dissolving pulp production from sugar cane bagasse. *Ind Crops*
552 *Prod* 2014 Jan; 52: 58–64.
- 553 [7] Cardona CA, Quintero JA, Paz IC. Production of bioethanol from sugarcane bagasse: status
554 and perspectives. *Bioresour Technol* 2010; 101(13): 4754–66.
- 555 [8] Rabelo SC, da Costa AC, Vaz Rossel CE. Industrial waste recovery. In: Santos F, Borém A,
556 Caldas C, editors. Sugarcane: Bioenergy, Sugar and Ethanol – Technology and Prospects,
557 Viçosa, Brazil: Universidade Federal de Viçosa; 2012, p. 449–70.
- 558 [9] Oliveira LRM, Nascimento VM, Gonçalves AR, Rocha GJM. Combined process system for
559 the production of bioethanol from sugarcane straw. *Ind Crops Prod* 2014 Jul; 58: 1–7.

- 560 [10] Rocha GJM, Gonçalves AR, Oliveira BR, Olivares EG, Rossell CEV. Steam explosion
561 pretreatment reproduction and alkaline delignification reactions performed on a pilot scale
562 with sugarcane bagasse for bioethanol production. *Ind Crops Prod* 2012; 35(1): 274–79.
- 563 [11] Rocha GJM, Nascimento VM, da Silva VFN, Corso DLS, Gonçalves AR. Contributing to the
564 environmental sustainability of the second generation ethanol production: delignification of
565 sugarcane bagasse with sodium hydroxide recycling. *Ind Crops Prod* 2014 Aug; 59: 63–8.
- 566 [12] Mou H, Heikkilä E, Fardim P. Topochemistry of environmentally friendly pretreatments to
567 enhance enzymatic hydrolysis of sugar cane bagasse to fermentable sugars. *J Agric Food*
568 *Chem* 2014; 62(16): 3619–25.
- 569 [13] Batalha LAR, Han Q, Jameel H, Chang H-M, Colodette JL, Gomes FJB. Production of
570 fermentable sugars from sugarcane bagasse by enzymatic hydrolysis after autohydrolysis and
571 mechanical refining. *Bioresour Technol* 2015 Mar; 180: 97–105.
- 572 [14] Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, et al. Lignins: Natural polymers
573 from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochem Rev* 2004; 3(1–2): 29–
574 60.
- 575 [15] Ralph J. Hydroxycinnamates in lignification. *Phytochem Rev* 2010; 9(1): 65–83.
- 576 [16] Sun J-X, Sun X-F, Sun R-C, Fowler P, Baird MS. Inhomogeneities in the chemical structure
577 of sugarcane bagasse lignin. *J Agric Food Chem* 2003; 51(23): 6719–25.
- 578 [17] Zhang A-P, Liu C-F, Sun R-C, Xie J. Extraction, purification, and characterization of lignin
579 fractions from sugarcane bagasse. *BioResources* 2013; 8(2): 1604–14.
- 580 [18] Zeng J, Tong Z, Wang L, Zhu JY, Ingram L. Isolation and structural characterization of
581 sugarcane bagasse lignin after dilute phosphoric acid plus steam explosion pretreatment and
582 its effect on cellulose hydrolysis. *Bioresour Technol* 2014 Feb; 154: 274–81.

- 583 [19] Bottcher A, Cesarino I, dos Santos AB, Vicentini R, Mayer JLS, Vanholme R, et al.
584 Lignification in sugarcane: biochemical characterization, gene discovery, and expression
585 analysis in two genotypes contrasting for lignin content. *Plant Physiol* 2013; 163(4): 1539–57.
- 586 [20] Tappi Standard Test Methods 2004-2005. Tappi Test Methods. Atlanta, GA: Tappi Press;
587 2004.
- 588 [21] Darwill A, McNeil M, Albersheim P, Delmer D. The primary cell-walls of flowering plants.
589 In: Tolbert N, editor. *The Biochemistry of Plants*, New York: Academic Press; 1980, p. 91–
590 162.
- 591 [22] Browning BL. *Methods of wood chemistry*. Vol. II; New York: Wiley-Interscience
592 Publishers; 1967.
- 593 [23] Björkman A. Studies on finely divided wood. Part I. Extraction of lignin with neutral
594 solvents. *Sven Papperstidn* 1956; 59(13): 477–85.
- 595 [24] del Río JC, Prinsen P, Rencoret J, Nieto L, Jiménez-Barbero J, Ralph J, et al. Structural
596 characterization of the lignin in the cortex and pith of elephant grass (*Pennisetum purpureum*)
597 stems. *J. Agric. Food Chem.* 2012; 60(14): 3619–34.
- 598 [25] Ralph J, Hatfield RD. Pyrolysis-GC/MS characterization of forage materials. *J Agric Food*
599 *Chem* 1991; 39(8): 1426–37.
- 600 [26] Kim H, Ralph J, Akiyama T. Solution-state 2D NMR of ball-milled plant cell-wall gels in
601 DMSO-*d*₆. *Bioenergy Res* 2008; 1(1): 56–66.
- 602 [27] Rencoret J, Marques G, Gutiérrez A, Nieto L, Santos JI, Jiménez-Barbero J, et al. HSQC-
603 NMR analysis of lignin in woody (*Eucalyptus globulus* and *Picea abies*) and non-woody
604 (*Agave sisalana*) ball-milled plant materials at the gel state. *Holzforschung* 2009; 63(6): 691–
605 98.

- 606 [28] del Río JC, Rencoret J, Prinsen P, Martínez ÁT, Ralph J, Gutiérrez A. Structural
607 characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and
608 reductive cleavage methods. J Agric Food Chem 2012; 60(23): 5922–35.
- 609 [29] Ralph SA, Ralph J, Landucci L. NMR database of lignin and cell wall model compounds
610 https://www.glbrc.org/databases_and_software/nmrdatabase/ (accessed: January 2009), 2009.
- 611 [30] Ralph J, Landucci LL. NMR of lignins. In: Heitner C, Dimmel DR, Schmidt JA, editors.
612 Lignin and lignans - Advances in chemistry, Boca Raton, FL: CRC Press; 2010: p. 137–234.
- 613 [31] del Río JC, Rencoret J, Marques G, Gutiérrez A, Ibarra D, Santos JI, et al. Highly acylated
614 (acetylated and/or *p*-coumaroylated) native lignins from diverse herbaceous plants. J Agric
615 Food Chem 2008; 56(20): 9525–34.
- 616 [32] Lu F, Ralph J. Derivatization followed by reductive cleavage (DFRC method), a new method
617 for lignin analysis: protocol for analysis of DFRC monomers. J Agric Food Chem 1997;
618 45(7): 2590–2.
- 619 [33] Lu F, Ralph J. The DFRC method for lignin analysis. Part 1. A new method for β -aryl ether
620 cleavage: lignin model studies. J Agric Food Chem 1997; 45(12): 4655–60.
- 621 [34] Lu F, Ralph J. Detection and determination of *p*-coumaroylated units in lignins. J Agr Food
622 Chem 1999; 47(5): 1988–92.
- 623 [35] Ralph J, Lu F. The DFRC method for lignin analysis. 6. A simple modification for identifying
624 natural acetates in lignin. J Agric Food Chem 1998; 46(11): 4616–9.
- 625 [36] del Río JC, Marques G, Rencoret J, Martínez AT, Gutiérrez A. Occurrence of naturally
626 acetylated lignin units. J Agric Food Chem 2007; 55(14): 5461–8.

- 627 [37] Rocha GJM, Nascimento VM, Gonçalves AR, Silva VFN, Martín C. Influence of mixed
628 sugarcane bagasse samples evaluated by elemental analysis and physical-chemical
629 composition. *Ind Crops Prod* 2015 Feb; 64: 52–8.
- 630 [38] del Río JC, Gutiérrez A, Rodríguez IM, Ibarra D, Martínez AT. Composition of non-woody
631 plant lignins and cinnamic acids by Py-GC/MS, Py/TMAH and FT-IR. *J Anal Appl Pyrol*
632 2007; 79(1–2): 39–46.
- 633 [39] Rencoret J, Prinsen P, Gutiérrez A, Martínez AT, del Río JC. Isolation and structural
634 characterization of the milled wood lignin, dioxane lignin, and cellulolytic lignin preparations
635 from brewer's spent grain. *J Agric Food Chem* 2015; 63(2): 603–13.
- 636 [40] Grabber JH, Quideau S, Ralph J. *p*-Coumaroylated syringyl units in maize lignin:
637 Implications for β -ether cleavage by thioacidolysis. *Phytochemistry* 1996; 43(6): 1189–94.
- 638 [41] Grabber JH, Ralph J, Hatfield RD. Cross-linking of maize walls by ferulate dimerization and
639 incorporation into lignin. *J Agr Food Chem* 2000; 48(12): 6106–13.
- 640 [42] Grabber JH, Lu F. Formation of syringyl-rich lignins in maize as influenced by feruloylated
641 xylans and *p*-coumaroylated monolignols. *Planta* 2007; 226(3): 741–51.
- 642 [43] Hatfield RD, Marita JM, Frost K, Grabber J, Ralph J, Lu F, et al. Grass lignin acylation: *p*-
643 coumaroyl transferase activity and cell wall characteristics of C3 and C4 grasses. *Planta* 2009;
644 229(6): 1253–67.
- 645 [44] del Río JC, Martín F, González-Vila FJ. Thermally assisted hydrolysis and alkylation as a
646 novel pyrolytic approach for the structural characterization of natural biopolymers and
647 geomacromolecules. *Trends Anal Chem* 1996; 15(2): 70–9.

- [45] Martín F, del Río JC, González-Vila FJ, Verdejo T. Thermally assisted hydrolysis and alkylation of lignins in the presence of tetra-alkylammonium hydroxides. *J Anal Appl Pyrol* 1995; 35(1): 1–13.
- [46] Ralph J, Hatfield RD, Quideau S, Helm RF, Grabber JH, Jung H-J. G. Pathway of *p*-coumaric acid incorporation into maize lignin as revealed by NMR. *J Am Chem Soc* 1994; 116(21): 9448–56.
- [47] Lu F, Ralph J. Novel β – β structures in lignins incorporating acylated monolignols. *Appita* 2005 May: 233–7.
- [48] Wen J-L, Su, S-L, Xue B-L, Sun R-C. Quantitative structural characterization of the lignins from the stem and pith of bamboo (*Phyllostachys pubescens*). *Holzforschung* 2013; 67(6): 613–27.
- [49] You T-T, Mao J-Z, Yuan T-Q, Wen J-L, Xu F. Structural elucidation of the lignins from stems and foliage of *Arundo donax* Linn. *J Agric Food Chem* 2013; 61(22): 5361–70.
- [50] Lan W, Lu F, Regner M, Zhu Y, Rencoret J, Ralph SA, et al. Tricin, a flavonoid monomer in monocot lignification. *Plant Phys* 2015; 167(4): 1284–95.
- [51] Rencoret J, Ralph J, Marques G, Gutiérrez A, Martínez AT, del Río JC. Structural characterization of lignin isolated from coconut (*Cocos nucifera*) coir fibers. *J Agric Food Chem* 2013; 61(10): 2434–45.
- [52] Landucci LL, Deka GC, Roy DN. A ^{13}C NMR study of milled wood lignins from hybrid *Salix* clones. *Holzforschung* 1992; 46(6): 505–11.
- [53] Sun RC, Fang JM, Tomkinson J. Fractional isolation and structural characterization of lignins from oil palm trunk and empty fruit bunch fibres. *J Wood Chem Technol* 1999; 19(1–2): 335–56.

- 671 [54] Kuroda K, Ozawa T, Ueno T. Characterization of sago palm (*Metroxylon sagu* Rottb.) lignin
672 by analytical pyrolysis. J Agric Food Chem 2001; 49(4): 1840–7.
- 673 [55] Lu F, Ralph J, Morreel K, Messens E, Boerjan W. Preparation and relevance of a cross-
674 coupling product between sinapyl alcohol and sinapyl *p*-hydroxybenzoate. Org Biomol Chem
675 2004; 2(20): 2888–90.
- 676 [56] Lu F, Karlen SD, Regner M, Kim H, Ralph SA, Sun R-C, et al. Naturally *p*-
677 hydroxybenzoylated lignins in palms. BioEnergy Res 2015; (in press; DOI 10.1007/s12155-
678 015-9583-4)
- 679 [57] Lu F, Ralph J. Preliminary evidence for sinapyl acetate as a lignin monomer in kenaf. Chem
680 Commun 2002 Jan; 1: 90–1.
- 681 [58] Ralph J. An unusual lignin from kenaf. J Nat Prod 1996; 59(4): 341–2.
- 682 [59] Lu F, Ralph J. Novel tetrahydrofuran structures derived from β – β -coupling reactions
683 involving sinapyl acetates in kenaf lignins. Org Biomol Chem 2008; 6(20): 3681–94.
- 684 [60] Withers S, Lu F, Kim H, Zhu Y, Ralph J, Wilkerson CG. Identification of a grass-specific
685 enzyme that acylates monolignols with *p*-coumarate. J Biol Chem 2012; 287(11): 8347–55.
- 686 [61] Petrik D, Karlen SD, Cass C, Padmakshan D, Lu F, Liu S, et al. *p*-Coumaroyl-
687 CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in
688 *Brachypodium distachyon*. Plant J 2014; 77(5): 713–26.

689 Figure Captions

690 **Figure 1.** Py-GC/MS chromatograms of the whole cell-walls of sugarcane (**A**) bagasse and (**B**)
691 straw, and of the MWLs from sugarcane (**C**) bagasse and (**D**) straw. The identities and relative
692 abundances of the released phenolic compounds are listed in **Table 2**. Levoglucosan arises from
693 carbohydrates.

694 **Figure 2.** Py-TMAH-GC/MS chromatograms of the whole cell-walls of sugarcane (**A**) bagasse and
695 (**B**) straw, and of the MWLs from sugarcane (**C**) bagasse and (**D**) straw. The identities and relative
696 abundances of the released compounds are listed in **Table 3**.

697 **Figure 3.** Side-chain (δ_C/δ_H 48-90/2.5-5.8, top) and aromatic (δ_C/δ_H 90-155/5.5-8.0, bottom) regions
698 from the 2D HSQC NMR spectra of the whole cell-walls from (**A**) sugarcane bagasse and (**B**) its
699 isolated MWL. See **Table 4** for signal assignments and **Figure 5** for the main lignin structures
700 identified.

701 **Figure 4.** Side-chain (δ_C/δ_H 48-90/2.50-5.80, top) and aromatic (δ_C/δ_H 90-155/5.50-8.00, bottom)
702 regions from the 2D HSQC NMR spectra of the whole cell-walls from (**A**) sugarcane straw and (**B**)
703 its isolated MWL. See **Table 4** for signal assignments and **Figure 5** for the main lignin structures
704 identified.

705 **Figure 5.** Main structures present in the lignins of sugarcane bagasse and straw as identified in the
706 NMR spectra of **Figures 3** and **4**. **A**: β -O-4'-structures; **A'**: β -O-4'-structures with acylated (by
707 acetate or *p*-coumarate) γ -OH; **A_{ox}**: C $_{\alpha}$ -oxidized β -O-4' structures; **B**: phenylcoumaran structures
708 formed by β -5'-coupling; **C**: resinol structures formed by β - β' -coupling; **C₃**: tetrahydrofuran
709 structures formed by β - β' -coupling of monolignols acylated at the γ -OH; **D**: dibenzodioxocin
710 structures formed by 5-5'- followed by 4-O- β -coupling; **E**: α,β -diaryl ether structures; **F**:
711 spirodienone structures formed by β -1'-coupling; **I**: *p*-hydroxycinnamyl alcohol end-groups; **I'**: *p*-
712 hydroxycinnamyl alcohol end-groups acylated at the γ -OH; **J**: cinnamaldehyde end-groups; **T**:

713 tricin end-groups; **PCA**: *p*-coumarate moieties; **FA**: ferulate moieties; **H**: *p*-hydroxyphenyl units;
714 **G**: guaiacyl units; **S**: syringyl units; **S'**: oxidized syringyl units bearing a carbonyl group at C α . The
715 structures are colored to match the assigned contours in the NMR spectra in **Figures 3** and **4**.

716 **Figure 6.** Sections of the HMBC spectra (δ_C/δ_H 164–171/3.8–7.8) of the MWLs isolated from (**A**)
717 sugarcane bagasse and (**B**) straw showing the main correlations for the carbonyl carbons of *p*-
718 coumarates (PCA) acylating the γ -OH of the lignin side-chains. Appropriate sections of the HSQC
719 spectra showing the C γ /H γ correlations of the acylated lignin (δ_C 60–66) and the C α /H α and C β /H β
720 correlations of *p*-coumarates (δ_C 112–115 and 142–147, respectively), are also depicted. Note the
721 absence of signals for the carbonyl carbons from acetates acylating the γ -position of the lignin side-
722 chains (at δ_C 169.8).

723 **Figure 7.** Chromatograms of the DFRC degradation products from the MWLs isolated from (**A**)
724 sugarcane bagasse and (**B**) straw showing the presence of sinapyl (and minor coniferyl) units
725 acylated by *p*-coumarate moieties. *tH*, *cG*, *tG*, *cS* and *tS* are the normal *cis*- and *trans*-*p*-
726 hydroxyphenyl (H), coniferyl (G) and sinapyl (S) alcohol monomers (as their acetate derivatives).
727 *cG_{pC}*, *tG_{pC}*, *cS_{pC}* and *tS_{pC}* are the *cis*- and *trans*-coniferyl and sinapyl dihydro-*p*-coumarates (as their
728 acetate derivatives). *cS_{pC}* and *tS_{pC}* are the *cis*- and *trans*-sinapyl *p*-coumarates (as their acetate
729 derivatives). IS: internal standard (*n*-tetracosane).

730 **Figure 8.** Resinol **C** and tetrahydrofurans produced from β – β' -homo-dimerization of sinapyl *p*-
731 coumarate (**C**₃) and its cross-coupling with sinapyl alcohol (**C**₁, **C**₂). (Adapted from [47,59]).

732 **Figure 9.** 2D-NMR HSQC of the (acetylated) MWLs from (**A**) sugarcane bagasse and (**B**) straw
733 showing the C α /H α and C β /H β and C γ /H γ correlations of the different β – β resinol **C** and
734 tetrahydrofuran **C**₁–**C**₃ structures arising from the coupling and cross-coupling of sinapyl alcohol
735 and sinapyl *p*-coumarate, as depicted in **Figure 8**. Signals were assigned from previously published
736 literature [47,59].

Table 1 – Abundance (percentage) of the main constituents of the dry weight of sugarcane (*Saccharum spp.*) bagasse and straw.

	Sugarcane bagasse ^a	Sugarcane straw ^a
Water-soluble material	1.3 ± 0.2	2.1 ± 0.2
Acetone extractives	0.9 ± 0.1	1.4 ± 0.1
Klason lignin ^b	17.8 ± 0.6	17.0 ± 0.2
Acid-soluble lignin	2.2 ± 0.2	1.9 ± 0.2
Holocellulose (α-cellulose) ^c	75.8 ± 0.5 (40.1 ± 0.2)	72.9 ± 0.7 (37.9 ± 0.3)
Ash	2.0 ± 0.1	4.7 ± 0.5

^a Average of three replicates

^b Corrected for proteins and ash

^c The α-cellulose contents are shown in parentheses

Table 2 – Identities and relative molar abundances of the compounds released after Py-GC/MS of whole cell-walls (CW) of sugarcane (*Saccharum spp.*) bagasse and straw and their isolated MWLs.

		Sugarcane Bagasse		Sugarcane Straw	
Compounds		CW	MWL	CW	MWL
1	phenol	2.4	2.7	3.1	3.0
2	guaiacol	2.5	2.2	3.9	8.0
3	3-methylphenol	1.0	0.7	1.1	0.8
4	4-methylphenol	2.4	4.2	3.1	4.4
5	4-methylguaiacol	2.3	3.0	4.0	8.9
6	4-ethylphenol	1.4	3.7	2.1	2.9
7	4-ethylguaiacol	0.8	1.0	1.8	2.8
8	4-vinylguaiacol	10.7	3.3	17.0	8.7
9	4-vinylphenol	51.8	41.7	43.8	30.0
10	eugenol	0.4	0.3	0.8	0.8
11	syringol	3.0	4.2	2.1	4.3
12	<i>cis</i> -isoeugenol	0.3	0.3	0.5	0.7
13	<i>trans</i> 4-propenylphenol	0.5	1.0	0.5	0.8
14	<i>trans</i> -isoeugenol	1.7	1.5	3.1	3.0
15	4-methylsyringol	2.3	5.1	1.7	3.5
16	vanillin	2.3	1.9	2.6	2.5
17	propynylguaiacol	0.2	0.4	0.2	0.3
18	propynylguaiacol	0.3	0.5	0.3	0.6
19	4-ethylsyringol	0.5	1.2	0.3	0.9
20	vanillic acid methyl ester	0.0	0.3	0.0	0.4
21	acetovanillone	0.5	0.9	0.8	1.3
22	4-hydroxybenzaldehyde	0.0	1.3	0.0	0.3
23	4-vinylsyringol	2.9	4.0	1.9	2.5
24	guaiacylacetone	0.2	0.2	0.5	0.8
25	4-allylsyringol	0.8	1.1	0.6	0.7
26	<i>cis</i> -4-propenylsyringol	0.5	0.8	0.3	0.6
27	propynylsyringol	0.4	0.3	0.4	0.0
28	propynylsyringol	0.4	0.3	0.2	0.0
29	<i>trans</i> -4-propenylsyringol	4.8	3.2	2.0	2.1
30	syringaldehyde	0.6	3.2	0.0	1.3
31	syringic acid methyl ester	0.1	0.3	0.1	0.1
32	acetosyringone	0.6	1.8	0.4	1.4
33	syringylacetone	0.6	1.2	0.4	0.9
34	propiosyringone	0.2	0.4	0.1	0.3
35	syringyl vinyl ketone	0.2	0.3	0.2	0.1
36	<i>trans</i> -coniferaldehyde	0.2	1.3	0.3	0.2
37	<i>trans</i> -sinapaldehyde	0.3	0.3	0.0	0.0
S/G ratio ^a		1.3	1.6	0.5	0.5

^aAll G- and S-derived peaks were used for the estimation of the S/G molar ratio, except for 4-vinylguaiacol (peak 8, which also, and predominantly, arises from ferulates), and the analogous 4-vinylsyringol (peak 23).

Table 3. Identity and relative molar abundances of the compounds released after Py/TMAH of whole cell-walls of sugarcane (*Saccharum spp.*) bagasse and straw and their isolated MWLs.

Label	Compound	Bagasse CW	Bagasse MWL	Straw CW	Straw MWL
1	4-methoxybenzeneethylene	18.0	3.0	12.3	2.5
2	1,2-dimethoxybenzene	2.7	0.4	4.8	1.3
3	3,4-dimethoxytoluene	1.9	0.3	2.9	1.0
4	4-methoxybenzaldehyde	3.8	2.9	2.3	2.0
5	1,2,3-trimethoxybenzene	1.4	0.6	1.7	0.5
6	3,4-dimethoxyphenylethylene	5.6	0.7	7.6	1.9
7	4-methoxybenzoic acid methyl ester	0.3	0.5	0.6	1.7
8	3,4,5-trimethoxytoluene	1.6	0.4	1.3	0.4
9	1,2-dimethoxy-4-propenylbenzene	1.3	0.8	1.8	1.8
10	3,4,5-trimethoxybenzeneethylene	1.0	0.7	0.7	0.6
11	3,4-dimethoxybenzaldehyde	2.1	1.7	3.8	5.2
12	<i>cis</i> 3-(4-methoxyphenyl)-propenoic acid methyl ester	1.5	2.7	2.1	2.5
13	3,4-dimethoxybenzoic acid methyl ester	0.8	1.9	1.4	5.4
14	3,4-dimethoxyacetophenone	0.6	0.4	0.8	1.0
15	1-(3,4,5-trimethoxyphenyl)-1-propene	0.9	0.9	0.6	0.8
16	<i>cis</i> 1-(3,4-dimethoxyphenyl)-2-methoxyethylene	0.1	0.5	0.3	1.1
17	3,4,5-trimethoxybenzaldehyde	1.0	3.0	0.7	2.9
18	<i>trans</i> 1-(3,4-dimethoxyphenyl)-2-methoxyethylene	0.2	0.6	0.4	1.2
19	<i>trans</i> 3-(4-methoxyphenyl)-propenoic acid methyl ester	43.8	68.7	38.4	54.8
20	3,4,5-trimethoxyacetophenone	0.9	0.6	0.5	1.5
21	3,4,5-trimethoxybenzoic acid methyl ester	0.8	3.1	0.9	2.6
22	3,4,5-trimethoxyphenyl-2-propanone	0.4	0.4	0.2	0.6
23	<i>cis</i> 3-(3,4-dimethoxyphenyl)-propenoic acid methyl ester	0.3	0.0	0.7	0.0
24	<i>cis</i> 1-(3,4,5-trimethoxyphenyl)-2-methoxyethylene	0.2	1.2	0.1	1.5
25	<i>trans</i> 1-(3,4,5-trimethoxyphenyl)-2-methoxyethylene	0.2	1.7	0.1	0.9
26	<i>trans</i> 3-(3,4-dimethoxyphenyl)-propenoic acid methyl ester	8.5	2.3	13.0	4.5
	<i>p</i> -Coumarate/Ferulate ratio ^a	5.2	31.1	3.0	12.8

^a Relative molar abundance of *p*-coumarates (peaks 12 and 19) relative to ferulates (peaks 23 and 26)

Table 4. Assignments of $^1\text{H}/^{13}\text{C}$ correlation signals in the 2D HSQC spectra from the whole-cell-walls of sugarcane (*Saccharum spp.*) bagasse and straw and their isolated MWLs in DMSO- d_6 .

Label	$\delta_{\text{C}}/\delta_{\text{H}}$ (ppm)	Assignment
Lignin cross-peak signals		
C _{3β}	49.8/2.59	C _β /H _β in γ-acylated β-β' tetrahydrofuran substructures (C ₃)
B _β	53.6/3.47	C _β /H _β in β-5' phenylcoumaran substructures (B)
C _β	53.6/3.07	C _β /H _β in β-β' resinol substructures (C)
-OCH ₃	55.6/3.73	C/H in methoxyls
A _γ	59.4 /3.40 and 3.72	C _γ /H _γ in normal (γ-hydroxylated) β-O-4' substructures (A)
F _β	59.7/2.76	C _β /H _β in spirodienone substructures (F)
I _γ	61.3/4.09	C _γ /H _γ in cinnamyl alcohol end-groups (I)
B _γ	62.6/3.68	C _γ /H _γ in β-5' phenylcoumaran substructures (B)
A' _γ	62.7/3.83-4.30	C _γ /H _γ in γ-acylated β-O-4' substructures (A')
I' _γ	64.0/4.79	C _γ /H _γ in γ-acylated cinnamyl alcohol end-groups (I')
C _γ	71.0/3.83 and 4.19	C _γ /H _γ in β-β' resinol substructures (C)
A _α /A' _α	71.8/4.87	C _α /H _α in β-O-4' substructures (A, A')
F _{β'}	79.5/4.11	C _β /H _{β'} in spirodienone substructures (F)
E _α	79.5/5.50	C _α /H _α in α,β-diaryl ether substructures (E)
A' _{β(G)}	80.8/4.58	C _β /H _β in γ-acylated β-O-4' substructures linked to a G unit (A')
F _α	81.2/5.10	C _α /H _α in spirodienone substructures (F)
C _{3α}	82.8/5.00	C _α /H _α in γ-acylated β-β' tetrahydrofuran structures (C ₃)
A _{oxβ}	82.8/5.23	C _β /H _β in α-oxidized β-O-4' substructures (A _{ox})
A' _{β(S)}	83.0/4.33	C _β /H _β in γ-acylated β-O-4' substructures linked to a S unit (A')
D _α	83.4/4.82	C _α /H _α in 5-5' dibenzodioxocin substructures (D)
A _{β(G)}	83.6/4.29	C _β /H _β in β-O-4' substructures linked to a G unit (A)
C _α	84.8/4.67	C _α /H _α in β-β' resinol substructures (C)
F _{α'}	84.9/4.88	C _α /H _{α'} in spirodienone substructures (F)
D _β	85.4/3.86	C _β /H _β in 5-5' dibenzodioxocin substructures (D)
A _{β(S)}	85.9/4.12	C _β /H _β in β-O-4' substructures linked to a S unit (A)
B _α	86.9/5.47	C _α /H _α in phenylcoumaran substructures (B)
T ₈	94.0/6.56	C ₈ /H ₈ in tricin units (T)
T ₆	98.7/6.22	C ₆ /H ₆ in tricin units (T)
S _{2,6}	103.8/6.69	C ₂ /H ₂ and C ₆ /H ₆ in etherified syringyl units (S)
T _{2,6'}	103.9/7.30	C ₂ /H ₂ and C ₆ /H ₆ in tricin units (T)
T ₃	104.5/7.03	C ₃ /H ₃ in tricin units (T)
S' _{2,6}	106.1/7.32 and 106.4 /7.19	C ₂ /H ₂ and C ₆ /H ₆ in α-oxidized syringyl units (S')
J _{2,6(S)}	106.8/7.06	C ₂ /H ₂ and C ₆ /H ₆ in cinnamaldehyde end-groups in S units (J)
G ₂	110.9/7.00	C ₂ /H ₂ in guaiacyl units (G)
FA ₂	111.0/7.32	C ₂ /H ₂ in ferulic acid units (FA)
PCA _β and FA _β	113.5/6.27	C _β /H _β in <i>p</i> -coumarate (PCA) and ferulate (FA)
H _{3,5}	114.5/6.62	C ₃ /H ₃ and C ₅ /H ₅ in <i>p</i> -hydroxyphenyl units (H)
G ₅ /G ₆	114.9/6.72 and 6.94 118.7/6.77	C ₅ /H ₅ and C ₆ /H ₆ in guaiacyl units (G)
PCA _{3,5}	115.5/6.77	C ₃ /H ₃ and C ₅ /H ₅ in <i>p</i> -coumarate (PCA)
I' _β	123.2/6.38	C _β /H _β in γ-acylated cinnamyl alcohol end-groups (I')
FA ₆	123.3/7.10	C ₆ /H ₆ in ferulate (FA)
J _β	126.3/6.76	C _β /H _β in cinnamaldehyde end-groups (J)
H _{2,6}	128.0/7.23	C ₂ /H ₂ and C ₆ /H ₆ in <i>p</i> -hydroxyphenyl units (H)
I _β	128.4/6.23	C _β /H _β in cinnamyl alcohol end-groups (I)
I _α	128.4/6.44	C _α /H _α in cinnamyl alcohol end-groups (I)
PCA _{2,6}	130.0/7.46	C ₂ /H ₂ and C ₆ /H ₆ in <i>p</i> -coumarate (PCA)
I' _α	132.9/6.60	C _α /H _α in γ-acylated cinnamyl alcohol end-groups (I')
PCA _α and FA _α	144.4/7.41	C _α /H _α in <i>p</i> -coumarate (PCA) and ferulate (FA)
J _α	153.4/7.61	C _α /H _α in cinnamaldehyde end-groups (J)
Polysaccharide cross-peak signals		
X ₅	63.2/3.26 and 3.95	C ₅ /H ₅ in β-D-xylopyranoside
X ₂	72.9/3.14	C ₂ /H ₂ in β-D-xylopyranoside
X' ₂	73.5/4.61	C ₂ /H ₂ in 2-O-acetyl-β-D-xylopyranoside
X ₃	74.1/3.32	C ₃ /H ₃ in β-D-xylopyranoside
X' ₃	74.9/4.91	C ₃ /H ₃ in 3-O-acetyl-β-D-xylopyranoside
X ₄	75.6/3.63	C ₄ /H ₄ in β-D-xylopyranoside

Graphical Abstract

Table 5. Structural characteristics (lignin inter-unit linkage types, end-groups, γ -acylation, aromatic units, and S/G Ratio, cinnamate content and *p*-coumarate/ferulate ratio, and tricin content) from integration of $^1\text{H}/^{13}\text{C}$ correlation signals in the HSQC spectra from the whole cell-walls of sugarcane (*Saccharum spp.*) bagasse and straw and their isolated MWLs.

	Sugarcane Bagasse CW	Sugarcane Bagasse MWL	Sugarcane Straw CW	Sugarcane Straw MWL
Lignin inter-unit linkages (%)				
β -O-4' Aryl ethers (A/A')	-	80	-	75
α -Oxidized β -O-4' aryl ethers (A_{ox})	-	3	-	0
Phenylcoumarans (B)	-	6	-	15
Resinols (C)	-	2	-	3
Tetrahydrofurans (C₃)	-	4	-	1
Dibenzodioxocins (D)	-	0	-	3
α,β -diaryl ethers (E)	-	2	-	0
Spirodienones (F)	-	3	-	3
Lignin end-groups ^a				
Cinnamyl alcohol end-groups (I)	-	3	-	7
γ -Acylated cinnamyl alcohol end-groups (I')	-	2	-	3
Cinnamaldehyde end-groups (J)	-	4	-	2
Lignin side-chain γ -acylation (%)	-	42	-	36
Lignin aromatic units ^b				
H (%)	3	2	5	4
G (%)	37	38	66	68
S (%)	60	60	29	28
S/G ratio	1.6	1.6	0.4	0.4
<i>p</i> -Hydroxycinnamates ^c				
<i>p</i> -Coumarate (%)	68	48	35	21
Ferulate (%)	26	4	18	5
<i>p</i> -Coumarate/Ferulate ratio	2.6	12.0	1.9	4.2
Tricin ^c	2	2	4	4

^aExpressed as a fraction of the total lignin inter-unit linkage types **A-F**

^bMolar percentages (H+G+S=100)

^c*p*-Coumarate, ferulate and tricin molar contents (integrals) as percentages of lignin content (H+G+S)

Table 6. Abundance ($\mu\text{mol/g}$ of lignin) of the monomers obtained from DFRC degradation of the MWLs from sugarcane (*Saccharum spp.*) bagasse and straw, and relative percentages of the different acylated (acetylated and *p*-coumaroylated) lignin monomers.

	H	G	G _{ac}	G _{pC}	S	S _{ac}	S _{pC} ^a	%G _{ac} ^b	%G _{pC} ^c	%S _{ac} ^d	%S _{pC} ^e
MWL Bagasse	2	128	5	2	98	1	30	3.7	1.5	0.8	23.3
MWL Straw	4	112	4	1	24	0.5	3	3.4	0.8	1.8	10.9

^a includes the sum of sinapyl *p*-coumarates (S_{pC}) and sinapyl dihydro-*p*-coumarates (S_{pC}).

^b %G_{ac} is the percentage of acetylated G units (G_{ac}) with respect to the total G units (G, G_{ac}, G_{pC}).

^c %G_{pC} is the percentage of *p*-coumaroylated G units (G_{pC}) with respect to the total G units (G, G_{ac}, G_{pC}).

^d %S_{ac} is the percentage of acetylated S units (S_{ac}) with respect to the total S units (S, S_{ac}, S_{pC}).

^e %S_{pC} is the percentage of total *p*-coumaroylated S units (including sinapyl *p*-coumarates, S_{pC}, and sinapyl dihydro-*p*-coumarates, S_{pC}), with respect to the total S units (S, S_{ac}, S_{pC}).

Figure 1

Figure 2

Figure 3

Figure 6

Figure 7

Figure 8

A **B** **Figure 9**