



HAL
open science

Characterization of liquid-liquid extraction fractions from lignocellulosic biomass by high performance liquid chromatography hyphenated to tandem high-resolution mass spectrometry

Carole Reymond, Alexis Dubuis, Agnès Le Masle, Cyril Colas, Ludovic Chahen, Emilie Destandau, Nadège Charon

► To cite this version:

Carole Reymond, Alexis Dubuis, Agnès Le Masle, Cyril Colas, Ludovic Chahen, et al.. Characterization of liquid-liquid extraction fractions from lignocellulosic biomass by high performance liquid chromatography hyphenated to tandem high-resolution mass spectrometry. *Journal of Chromatography A*, 2020, 1610, pp.460569. 10.1016/j.chroma.2019.460569 . hal-02551353

HAL Id: hal-02551353

<https://ifp.hal.science/hal-02551353v1>

Submitted on 22 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Characterization of liquid-liquid extraction fractions from lignocellulosic
2 biomass by high performance liquid chromatography hyphenated to tandem
3 high-resolution mass spectrometry (HPLC/MSⁿ)

4 Carole Reymond^a, Alexis Dubuis^a, Agnès Le Masle^{a,*}, Cyril Colas^{b,c}, Ludovic Chahen^a, Emilie
5 Destandau^b, Nadège Charon^a

6 ^a IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France

7 ^b Institut de Chimie Organique et Analytique, Université d'Orléans, CNRS UMR 7311, Rue de Chartres, 45067
8 Orléans, France

9 ^c Centre de Biophysique Moléculaire, CNRS UPR 4301, Université d'Orléans, rue Charles Sadron, 45071 Orléans,
10 France

11 *Corresponding author. E-mail address: agnes.le-masle@ifpen.fr (A. Le Masle)

12
13 **Abstract**

14 The conversion of lignocellulosic biomass is a major challenge in the field of renewable energies and
15 bio-based chemicals. The diversity of biomasses and processes leads to complex products having a
16 wide range of polarities and molecular weights. Nowadays, the molecular description of these
17 oxygenated matrices is still largely incomplete and new analytical strategies are required to have a
18 better understanding of biomass products properties. The present study proposes a reliable protocol
19 based on successive liquid-liquid extractions prior to high performance liquid chromatography
20 hyphenated to high-resolution tandem mass spectrometry (HPLC/MSⁿ) using a linear ion trap-Fourier
21 transform ion cyclotron resonance mass spectrometer (LTQ/FT-ICR). The protocol allowed to
22 fractionate an industrial sample coming from the sulfuric acid-based pretreatment of a wheat straw
23 into four key chemical families: carbohydrates, organic acids, phenols and neutral compounds. Each
24 fraction was separately analyzed, which limited matrix effects during mass spectrometry ionization
25 step. Electrospray and atmospheric pressure chemical ionization sources were used in both positive
26 and negative modes in order to ionize and detect a maximum of compounds. Thanks to HPLC/MSⁿ,
27 structures of heavy lignin-carbohydrate complexes (LCC) were elucidated (up to 600 g/mol) as well as
28 carbohydrate oligomers having acid functionalities. Mono, di, tri and tetra-aromatic compounds
29 coming from lignin were also detected. The results reported in this paper demonstrate the
30 complexity of pretreated biomass samples and propose an analytical approach from sample
31 simplification to data treatment in order to describe the biomass composition.

32
33 *Keywords: Lignocellulosic biomass, liquid-liquid extraction, high performance liquid chromatography,*
34 *high-resolution mass spectrometry, tandem mass spectrometry*

36 1. Introduction

37 Global increase in energy consumption combined with a depletion of fossil fuel and global warming
38 lead to a growing concern for renewable energy sources[1–3]. Lignocellulosic biomass also called
39 biomass of second generation is a promising and renewable resource to produce biofuels and
40 valuable bio-based chemicals [4]. Coming from various feedstocks such as forest residues,
41 agricultural wastes or dedicated crops, lignocellulosic biomass is the most abundant material of plant
42 cell walls. This composite material consists of linear or branched polymeric carbohydrates (cellulose
43 and hemicelluloses) and aromatic macromolecules of lignin [5]. Several transformation routes are
44 studied to convert this complex raw material into biofuels and bio-products [6–8]. As regards
45 biochemical transformation route, biomass is first pretreated to increase the accessibility of cellulose
46 to enzymes that are used to hydrolyze this polysaccharide to produce glucose. During this step, the
47 hemicelluloses were dissolved and hydrolyzed, which leads to complex aqueous matrices made up of
48 several hundreds of oxygenated compounds that are distributed over a wide range of polarities
49 and/or molecular weights, potentially being heat sensitive and having several chemical groups for a
50 single molecule. For all these reasons but also because of the low concentrations of some analytes,
51 the complete characterization of pretreated biomass sample is still a challenge [9–12]. A better
52 understanding of relationships between the products composition and reactivity is expected to
53 support development of innovative and efficient processes.

54 Mono-dimensional gas chromatography (GC) or comprehensive two-dimensional GC×GC coupled
55 with flame ionization detection and/or mass spectrometry (MS) are often used in literature to
56 characterize samples originating from biomass. However, these techniques presents limitation for
57 compounds with a molecular mass above 200 g/mol which usually need derivatization step [13–15].
58 High performance liquid chromatography (HPLC) techniques are relevant complementary approaches
59 to GC, especially when dealing with thermal sensitive, polar and/or high molecular weight
60 compounds. To go deeper in the characterization and structural elucidation, high-resolution mass
61 spectrometry (HRMS) and multi-stage tandem mass spectrometry (MS^n) are required [16]. Most of
62 publications using HRMS with or without HPLC mainly focus on lignin derivative compounds [17–21].
63 Recently, Jarrell *et al.* investigated lignin compounds produced from an organosolv switchgrass by
64 HPLC/ MS^n [22]. However, when considering very complex samples, HPLC may lack resolution leading
65 to peaks co-elutions and thus enhancement of matrix effects in MS ionization. To simplify matrix
66 composition and to limit co-elutions, a fractionation step prior to analysis should be performed.
67 Among these analytical pretreatments, liquid-liquid extraction (LLE) is an attractive technique easy to
68 implement to separate compounds of different polarities [23–25]. Moreover, it allows to recover the
69 whole sample, which is critical to fully characterize biomass products. Methyltertiobutylether (MTBE)
70 was already used to extract phenols, organic acids and neutral degradation products from biomass
71 hydrolysates [23]. MTBE extraction allows to isolate highly hydrophilic compounds such as
72 carbohydrates from other oxygenated molecules [12,26]. An interesting methodology based on
73 successive LLE and pH modifications has been proposed by Kanaujia *et al.* to go further the
74 fractionation of another kind of biomass product (fast pyrolysis oils) [24]. In this way, multi-step LLE
75 appears to be a promising approach to fractionate biomass products into selective chemical groups
76 [25].

77 In this study, a new protocol was set up to fractionate an aqueous biomass product coming from the
78 sulfuric acid-based pretreatment of a wheat straw in four fractions according to the main chemical

79 families: carbohydrates, organic acids, phenols and neutral compounds. Then, HPLC/MSⁿ analyses
80 using two ionization sources *i.e.* electrospray (ESI) and atmospheric pressure chemical ionization
81 (APCI) were performed to obtain structural information. The aim of this study is to characterize the
82 whole sample from carbohydrates to aromatic compounds coming from the hydrolysis of
83 hemicellulose and the solubilisation of lignin.

84 2. Experimental

85 2.1. Chemicals

86 The reagents used for the LLE protocol (methyl *tert*-butyl ether (MTBE) and methanol) were HPLC
87 grade purchased from Sigma-Aldrich (Saint-Quentin-Fallavier, France). Both sulfuric acid (H₂SO₄, 96%)
88 and sodium hydroxide (NaOH, 0.5 mol/L) were obtained from Carlo-Erba reagent (Val de Rueil,
89 France). Hydrochloric acid (HCl, 0.5 mol/L) was purchased from VWR (Fontenay sous Bois, France).
90 Methanol and formic acid used for LC/MS analysis were MS grade purchased from VWR (Fontenay
91 sous bois, France). Deionized water was produced by a Milli-Q water purifier (Millipore SAS,
92 Molsheim, France).

93 2.2. Sample

94 The biomass sample investigated in this work was provided by IFP Energies nouvelles (Solaize,
95 France). It corresponded to the water soluble fraction obtained by a sulfuric acid-based pretreatment
96 of a wheat straw. Xylose, glucose, 5-HMF (5-(hydroxymethyl)furfural) and furfural were quantified at
97 80.4 g/L, 8.5 g/L, 0.311 g/L and 0.081 g/L respectively using reference methods [27,28]. A pH of 2.1
98 was measured using a combined pH electrode (Fisher Scientific, Illkirch, France). The entire sample
99 (including particles) was introduced in the LLE protocol. Sample and fractions were filtered before
100 HPLC/MS analysis using Macherey-Nagel PTFE membrane filters (pore size: 0.20 μm, diameter: 13
101 mm, Düren, Germany).

102 2.3. Liquid-liquid fractionation

103 A fractionation protocol corresponding to successive LLE was set up to generate four fractions from
104 the aqueous biomass sample: carbohydrates (AQ1), organic acids (AQ3), phenols (ORG3) and neutral
105 products (ORG2) (Figure 1). To do so, 4.5 mL of the sample were introduced and weighed in a 30 mL
106 separating funnel. Then 1.5 mL of MTBE were added and weighed in the separating funnel. After 30 s
107 of vigorous shaking, the two phases were left 5 min for equilibration. The organic and aqueous
108 phases were separated in two flasks. The LLE extraction was performed two more times on the
109 aqueous phase using each time 1.5 mL of MTBE (4.5 mL used in total). The resulting organic (ORG1)
110 phases were combined and weighed to perform a mass balance. AQ1 was kept for analysis. ORG1
111 was fully reintroduced in the separating funnel and 4 mL of 0.5 mol/L sodium hydroxide solution
112 were added. This volume may be adapted according to the sample acidity to reach pH 12 in the
113 aqueous fraction (AQ2a). Organic (ORG2) and aqueous (AQ2a) phases were shaken and collected
114 after equilibration. ORG2 was concentrated by evaporation under a nitrogen flow and weighed for
115 mass balance. The dry extract was dissolved in 250 μL of methanol and transferred into a vial for
116 analysis. AQ2a was acidified with 0.5 mol/L hydrochloric acid to reach pH 7 (AQ2b) under pH
117 combined electrode control. AQ2b (9 mL) was then extracted in triplicate with MTBE (3x3 mL). AQ3
118 was collected for analysis. ORG3 was evaporated under nitrogen and the dry extract was dissolved in

119 250 μL of methanol. ORG3 was transferred into a vial for analysis. The entire LLE protocol was
120 conducted at room temperature ($20 \pm 2^\circ\text{C}$).

121 2.4. *Liquid chromatography/high-resolution mass spectrometry*

122 All analyses were performed using an Agilent 1290 UHPLC system consisting of a binary pump, an
123 autosampler, a temperature-controlled column compartment and a photodiode array (PDA) UV
124 detector. The extra-column variance was measured to be equal to $9 \mu\text{L}^2$. The chromatographic
125 system was hyphenated to a Fourier-transform ion cyclotron resonance mass spectrometer (LTQ-FT-
126 ICR Thermo Scientific; 7 T magnet). Ionization was carried out with the ESI and APCI sources, both
127 working in positive or negative modes.

128 For HPLC/MS analysis, 1 μL of each sample was injected. Separation was achieved on a Kinetex C18
129 column ($100 \times 3 \text{ mm}$, particle size $2.6 \mu\text{m}$, Phenomenex, France) at 30°C . The mobile phase solvents
130 used were a 0.01% (v/v) formic acid in water (A) and a 0.01% (v/v) formic acid in methanol (B) at 600
131 $\mu\text{L}/\text{min}$. A linear gradient was used as follows: 0.0 - 4.0 min, 1% B; 4.0 - 30.0 min, from 1 to 99% B;
132 30.0 - 35.0 min, 99% B; 35.0 - 36.0 min, from 99 to 1% B; 36.0-40.0 min, 1% B. UV signal was recorded
133 from 210 to 400 nm.

134 Ionization efficiency was optimized with a design of experiments approach according to a previous
135 work [29]. For ESI source, a splitter was set up at the outlet of the HPLC column and at the entrance
136 of the MS device to reduce the solvent flow (split ratio 1:2.2). To avoid excessive pressure in the UV
137 cell, HPLC flow was split before PDA. Flow rates of sheath, auxiliary and sweep gases as well as
138 vaporizer and transfer capillary temperatures, transfer capillary voltage, spray voltage and corona
139 discharge values are summarized in Table 1. In order to have at least 10 points for each
140 chromatographic peak, a resolving power of 12500 at m/z 400 was used for the FT-ICR mass
141 spectrometer. The mass accuracy of the determinations was 2 ppm.

142 MS^n experiments were performed using the data dependent acquisition functionality. The most
143 intense ion at each time was selected, subjected to isolation and then to fragmentation in the LTQ
144 (linear ion trap in front of the FT-ICR mass spectrometer). An isolation window of 2 m/z was used and
145 fragmentation was done with collision energy of 35%.

146 3. Results and discussion

147 3.1. *A selective LLE fractionation*

148 3.1.1. *Validation of the LLE protocol*

149 A fractionation protocol based on successive LLE was developed to improve HPLC/MS
150 characterization of aqueous biomass samples. Four fractions, selective of targeted chemical families
151 are produced from the whole biomass sample to limit co-elutions and to help chromatograms
152 understanding. More precisely, high water-soluble compounds (*i.e.* carbohydrates) are expected to
153 be recovered mainly in aqueous phase 1 (AQ1), non ionizable compounds at pH 12 mainly in organic
154 phase 2 (ORG2), phenols ($\text{pK}_a > 9$) mainly in organic phase 3 (ORG3) and carboxylic acids ($\text{pK}_a < 7$)
155 mainly in aqueous phase 3 (AQ3) as illustrated in Figure 1.

156 The first key step of the LLE protocol is to get on one hand an aqueous phase containing
157 carbohydrates and, on the other hand, an organic phase containing all other oxygenated compounds.

158 MTBE was reported in previous studies as an efficient solvent to extract organic species from
159 aqueous biomass samples [12,23]. To confirm these observations, partition experiments were
160 realized in the MTBE/water solvent system buffered at pH 2.8 using 35 model molecules
161 representative of the main oxygenated chemical families (Figure S1 and Table S1). These partition
162 measurements demonstrated a good extraction of phenols, organic acids, aldehydes, ketones, esters
163 and alcohols in ORG1 while very hydrophilic compounds such as carbohydrates and their derivatives
164 were recovered in AQ1. The second step of the protocol consisted in addition of diluted sodium
165 hydroxide solution in ORG1 to transfer ionizable compounds in aqueous phase 2 (AQ2) at pH 12,
166 namely carboxylic acids and phenols. At this step, all non ionizable compounds at pH 12 ($pK_a > 14$)
167 remained in ORG2. Then, an acidification is performed on AQ2 to reach pH 7, intermediate between
168 phenols pK_a (9-10) and carboxylic acids pK_a (4-5). Therefore, phenols are neutral at pH 7 and
169 carboxylic acids under their ionic form, which allows a final MTBE extraction to separate re-
170 protonated phenols in ORG3 from carboxylates in AQ3. Mass balances using model molecules were
171 done on the global scheme to assess a controlled fractionation selectivity (data available in
172 Supplementary information Figure S2). However, some neutral compounds (furfural, 5-HMF, ethanol,
173 acetic acid, etc.) were partitioned between organic and aqueous phases and they were consequently
174 detected in different LLE fractions.

175 *3.1.2. HPLC-UV chromatograms of the biomass sample and LLE fractions*

176 The LLE protocol was applied to an aqueous product coming from the sulfuric acid-based
177 pretreatment of a wheat straw. A first extraction was performed to evaluate the mass repartition of
178 main compounds contained in the sample. The resulting mass balance done on dry extracts of the
179 four fractions revealed a significant proportion of carbohydrates and presence of minority products
180 (AQ1: 89% w/w, ORG2: 2% w/w, ORG3: 4% w/w, AQ3: 3% w/w, losses: 2% w/w). The LLE protocol
181 was repeated on the whole sample to produce the fractions used for analysis. Organic fractions were
182 concentrated as described in the experimental section to improve the analysis of minor products.
183 HPLC-UV chromatograms obtained for the whole sample and LLE fractions are presented in Figure 2.

184 HPLC-UV chromatogram of the entire sample presents about sixty peaks with very different
185 intensities, revealing many co-elutions, and is therefore difficult to process without prior
186 fractionation. All HPLC-UV chromatograms are overlaid to suggest effective fractionation of
187 compounds in the LLE fractions. It is interesting to notice that some carbohydrate-type molecules in
188 AQ1 carry chromophore groups responsible for their UV response. Hyphenation with high-resolution
189 mass spectrometry is required to achieve the structural characterization of these polyfunctional
190 compounds.

191 *3.2. Characterization of LLE fractions by HPLC/MSⁿ*

192 Despite some fractions represented a small part of the entire sample, their characterization is crucial.
193 Indeed, during the biomass pretreatment which is the first step of the biochemical transformation,
194 some inhibitors of the hydrolysis enzymatic and alcoholic fermentation are released. These inhibitors
195 can have an impact on the yield of the reaction even at low concentration. Some of them are already
196 known such as phenolic compounds or furans, however many of them are still unknown. In order to
197 characterize all classes of components in the initial biomass sample, LLE fractions were analyzed by
198 HPLC/MSⁿ using reverse phase chromatography with ESI and APCI ionizations in positive and negative
199 modes (ESI-/+ and APCI-/+). Deprotonated $[M-H]^-$ molecules and protonated $[M+H]^+$ molecules are

200 mostly formed in negative and positive modes respectively. High-resolution mass spectrometry
201 allowed to measure m/z ratio with an accuracy of 2 ppm. Molecular formulae were calculated
202 considering carbon, hydrogen and oxygen between 1 and 100 atoms for each. For ESI+ and APCI+,
203 nitrogen was also considered for molecular formulae calculations. In total for the four LLE fractions,
204 480 molecular formulae were measured in ESI-, 444 in ESI+, 109 in APCI- and 122 in APCI+. Although
205 ESI- was the most suitable ionization mode for this sample by detecting the highest number of
206 compounds, other detections were complementary to ESI- and thus required for a complete
207 characterization of the sample. A Venn diagram is presented in supplementary information to
208 illustrate this complementarity (Figure S3). To represent all detected components, van Krevelen
209 diagrams were built, representing H/C *versus* O/C ratios in order to reveal areas corresponding to
210 chemical classes (Figure 3).

211 AQ3, ORG2 and ORG3 covered really specific areas on the van Krevelen diagram, which demonstrates
212 the specificity of LLE protocol. AQ3 and ORG3 were mainly centered on an area located at H/C and
213 O/C ratios of 1 and 0.4, this area is known as phenolic zone [21]. ORG2 might be related to lipids with
214 rather high H/C ratios and low O/C ratios [21]. AQ1 was distributed into two different zones,
215 including an area typically attributed to carbohydrates (H/C and O/C ratios being higher than 1.5 and
216 0.8 respectively). Discussion of each fraction composition is detailed next. Most abundant ions
217 belonging to each of the four fractions were subjected to isolation and CID (collision-induced
218 dissociation) experiments. High-resolution mass spectra were measured to obtain elemental
219 compositions of fragments.

220 3.2.1. Aqueous fraction 1

221 According to LLE protocol, a large part (89% w/w) of the initial sample is focused in AQ1.
222 Components present in this fraction have a high solubility in water and might be related to
223 carbohydrate compounds. According to UV chromatogram in Figure 2, two groups of peaks can be
224 distinguished: a first group corresponds to the less retained compounds on the non-polar column,
225 having retention times inferior to 10 min (less than 12% of methanol in the HPLC mobile phase),
226 while a second group consists on the strongly compounds on the HPLC column, eluting after 10 min.
227 So it was assumed that AQ1 contains carbohydrates owing unsaturated chemical functions, which
228 explains observations about retention behavior on a non-polar HPLC stationary phase and possibility
229 to see peaks from UV detection. HPLC/MS analysis of AQ1 using several ionization conditions (ESI+/-
230 and APCI +/-) delivered numerous compounds that were represented on a van Krevelen diagram
231 (Figure 3). The size of the circle represents the mass intensity of the measured ions. First of all, by
232 using different colors for retained (black dots) and un-retained compounds (green and orange dots)
233 on chromatographic column, a new chemical family of compounds which might correspond to lignin-
234 carbohydrate complexes (LCC) can be clearly separated from other carbohydrates. This class of
235 molecules which combine lignin derivatives and carbohydrate units, was already related in the
236 literature [30,31]. Due to the phenol moiety, LCC have lower H/C and O/C ratios than 'conventional'
237 carbohydrates like glucose or xylose. This distinction can also be highlighted when taking into
238 account the double bond equivalent (DBE) of compounds (Figure S4). For the same molecular mass, a
239 retained compound exhibits a higher DBE than a non-retained compound due to its aromatic ring(s)
240 attached to the carbohydrate part. This additional type of information could help usefully to assume
241 structural elucidation of compounds analyzed in AQ1.

242 Considering non lignin-carbohydrate compounds whose retention times are inferior to 10 min,
243 several groups can be distinguished from Figure 4:

244 (1) some compounds detected in both negative and positive ionization modes are located on a line
245 where H/C ratios are two times higher than the O/C ones ($H/C = 2 \times O/C$).

246 (2) Under this line $H/C = 2 \times O/C$, compounds were ionized only in negative mode (green dots), which
247 might indicate the presence of carboxylic acid function. One may suggest detection of uronic acids
248 in this area of the van Krevelen diagram.

249 (3) Above the line $H/C = 2 \times O/C$, nitrogen compounds are mainly detected (orange dots).

250 The line $H/C = 2 \times O/C$ corresponds to dehydration reactions with loss of a H_2O molecule. Compounds
251 located on this line may be reaction intermediates, such the ones coming from degradation of xylose
252 to produced furfural or from glucose to produce 5-hydroxymethylfurfural (5-HMF). Oligomers
253 constituted of several $C_xH_{2x}O_x$ carbohydrate units (xylose, glucose, mannose, etc.) are located on this
254 dehydration line due to the loss of water to condense two sugar units. From a mathematical point of
255 view, a polymer with infinity of xylose (or glucose) units has H/C and O/C ratios which tend to 1.60
256 and 0.80 respectively (1.67 and 0.83 for glucose series). These two points are indicated by points (1)
257 and (2) on Figure 4. Moreover, these two specific points correspond to the first dehydration of xylose
258 (or glucose) meaning that compounds on the dehydration line under these points are carbohydrate
259 degradation products.

260 Fragmentations in negative and positive-ion modes ESI were performed from MS^2 to MS^7
261 experiments. Usual carbohydrates neutral losses such as H_2O (18 Da), $C_2H_4O_2$ (60 Da), $C_5H_8O_4$ (132
262 Da), $C_5H_{10}O_5$ (150 Da) or $C_6H_{10}O_5$ (162 Da) are shown in Figure S5. These fragments confirm the
263 presence of carbohydrates in AQ1 which are specifically extracted by the LLE protocol. Loss of xylose
264 can give $C_5H_8O_4$ or $C_5H_{10}O_5$ fragments in negative-ion mode according to how xylose is linked to the
265 rest of the molecule. $C_5H_{10}O_5$ fragment might also come from specific fragmentation of hexose ($^{0,1}X^-$
266 according to Domon and Costello nomenclature [32]). Xylose losses were more abundant than
267 glucose ones ($C_6H_{10}O_5$), which is consistent with the higher concentration of xylose (80.4 g/L) than
268 glucose (8.5 g/L) in the initial sample. Loss of $C_3H_4O_2$ (72 Da) corresponds to the remaining part of
269 xylose after a loss of $C_2H_4O_2$. Moreover, for retained components (retention times between 10 and
270 20 min), many losses of CO_2 and $\bullet CH_3$ radical happened mainly after MS^3 fragmentation stage. CO_2
271 loss is well known neutral fragment from carboxylic acid and $\bullet CH_3$ radical from methoxy group on
272 aromatic ring. These observations were consistent with the hypothesis of the existence of lignin-
273 carbohydrate complexes.

274 To illustrate the fragmentation process with some examples of proposed structures, fragmentations
275 of the most intense peaks on base peak chromatogram (BPC) are detailed. Peaks having various
276 retention times were chosen and located on the van Krevelen diagram dedicated to AQ1 (Figure 5).
277 Domon and Costello nomenclature was used to depict the fragmentation spectrum.

278 For retained compounds (eluted after 10 min), the most intense peak using negative-ion mode ESI
279 (Figure 5) but also the other detection modes was peak (b) eluted at 13.5 min which can be assessed
280 to feruloyl xyloside (Figure 6). In negative-ion mode ESI, MS^2 and MS^3 consisted in the loss of the
281 xylose part, leading to Y_0^- fragment, then fragmentations of the lignin part started with $\bullet CH_3$ and CO_2

282 losses. Positive-ion mode ESI provided complementary information with a different fragmentation
283 process: Z_0^+ fragment was formed, then methoxy group fragmentation consisted in a CH_4O loss
284 instead of a $\bullet CH_3$ radical loss. Last fragmentation step was dedicated to a loss of CO. Moreover, UV
285 spectrum of compound b was the same as ferulic acid, which confirmed its identification. Compound
286 b' had the same molecular formula and UV spectrum as compound b but its fragmentation was
287 different. This isomer might correspond to a ferulic acid linked to a xylose unit on a different position
288 (for example on the hydroxyl group of ferulic acid instead of carboxylic acid group). Other
289 fragmentations confirmed the presence of LCC for retained compounds. Fragmentations of peaks a
290 and c allowed to identify coumaroyl xyloside and feruloyl dixyloside respectively (Figure 6). The
291 carbohydrate part was removed first and only afterwards the fragmentation of the phenolic part
292 took place. The positions of the glycosidic bond and the bond between the phenolic and the
293 carbohydrate parts could not be strictly determined. However, by considering all the fragments, the
294 phenolic group might be link to the xylose on its position 4 for compounds (a), (ab) and (c). For the
295 compound (c), three positions of the glycosidic bond were possible: 1-4, 1-3 and 1-2 (Figure S6). For
296 some components and especially for molecules with a phenol moiety that is chemically complex (*ie*
297 polyaromatic), no detailed structure could be proposed. As regards peak d, the carbohydrate part
298 was a hexose unit while the phenolic part has a $C_{20}H_{19}O_7$ formula. MS^3 to MS^7 fragmentations gave
299 precious information on the lignin part such as $\bullet CH_3$ losses, which may indicate methoxy group on
300 aromatic ring. Also, phenolic part has a DBE of 11 for only 20 atoms of carbon which should indicate
301 two aromatic rings.

302 Concerning carbohydrates eluted at the beginning of the run (between 0 and 10 min), many
303 compounds had similar fragments. Indeed, many of them are composed of an acidic sugar (such as
304 methylglucuronic acid or glucuronic acid) attached to 1, 2 or 3 pentose and/or hexose units. MS^2
305 spectra of peaks α , β and γ are shown in Figure 7. Positions of the chemical groups were not taken
306 into account here. Methylglucuronic acid group was confirmed by detection of a fragment at m/z
307 207.0510 which lost CH_4O corresponding to a methoxy group. Peaks δ , ϵ and ζ were identified as
308 glucuronic acid with one, two and three pentose units respectively.

309 The highest intensity peak detected on BPC in negative-ion mode ESI (peak η , m/z
310 513.1469/ $C_{19}H_{29}O_{16}$) had similar fragments than peak β and was a bit more retained on the HPLC
311 stationary phase. It might have similar structure with an extra acetyl group ($COCH_3$) linked to a
312 pentose.

313 For compounds eluted between 0 and 10 min, sugar acids were detected in negative-ion mode ESI.
314 Many oligomers combining sugar acids and pentose and/or hexose units were present in AQ1.
315 Moreover, on the van Krevelen diagram (Figure 5), same components with different numbers of
316 pentose units are located on a same line which converges toward point (1). Thanks to this linear
317 structuration and to the molecular formulae, it should be possible to make predictions on the
318 structures of unknown molecules without fragmentation experiments, provided one compound has
319 already been identified on the same line. For example, feruloyl xyloside and feruloyl dixyloside (peaks
320 b and c respectively) were identified by the use of MS^n analyses. The line going through these two
321 points and point (1) also goes through $C_{25}H_{34}O_{16}$ ($H/C = 1.36$; $O/C = 0.34$) and $C_{30}H_{42}O_{20}$ ($H/C = 1.40$;
322 $O/C = 0.66$) molecules. Thus, they may be assessed to feruloyl trixyloside and tetraxyloside.

323 It has to be noticed that for this complex fraction, despite the use of chromatographic separation
324 ahead the mass spectrometer, it remained some co-elutions especially at the beginning of the run for
325 the very polar compounds. These co-elutions may lead to matrix effect and ion suppression in the
326 atmospheric ion source. Thus, some compounds might to be not be detected. A dedicated
327 chromatographic separation and ionization method should be optimized for this class of molecules.

328 *3.2.2. Aqueous fraction 3*

329 AQ3 is dedicated to carboxylic acids according LLE protocol. This fraction represents 3% (w/w) of our
330 initial biomass sample. Despite the low mass percentage of this fraction, it was necessary to carry out
331 the fractionation protocol to isolate minor products and to identify species that were not detected in
332 presence of the major fraction (AQ1) due to matrix effects. On the van Krevelen diagram,
333 components in this fraction are mainly centered around H/C of 1.0 and O/C of 0.4, meaning that they
334 are composed of aromatic rings (Figure 8 (1)). Interaction with the apolar stationary phase (retention
335 times between 10 and 20 min) are in concordance with the hypothesis of aromatic acids. These
336 aromatic carboxylic acids can be separated between mono, di, tri and tetra-aromatic compounds by
337 plotting the DBE value as a function of molecular mass (Figure 8 (2)). Few aliphatic carboxylic acids
338 were also detected in this fraction. At the opposite of aromatic acids, they have a high H/C ratio and
339 a low DBE. Moreover, furans such as 5-HMF were observed. Indeed because of their low log P, they
340 were distributed in several LLE fractions. During fragmentation process in negative-ion mode ESI,
341 principal losses were CO₂ coming from fragmentation of carboxylic acid function and •CH₃ radical
342 from fragmentation of methoxy group attached to an aromatic ring (Figure 8 (3)). Among the 17
343 compounds submitted to fragmentation, 10 of them lost the carboxylic acid moiety as the major
344 fragment during the first CID event. When methoxy functions are present on an aromatic ring,
345 methyl radical losses are in competition with CO₂ loss. Radical methyl groups were removed one by
346 one during each fragmentation stage. Therefore experimental data enable us to identify coumaric,
347 syringic and ferulic acids as mono-aromatic compounds. For coumaric acid, exclusive loss of CO₂ was
348 observed, whereas for ferulic and syringic acids additional losses of one and two methyl radicals
349 were exhibited respectively. For few dimeric components, water was the most abundant loss in MS²
350 experiment although CO₂ loss exhibited a lower intensity. This loss was associated to CH₂O loss which
351 might indicate β-O-4 linkage [17]. However, fragmentation by CID shown limitations for untargeted
352 structural identifications of di, tri and tetra-aromatic compounds. To propose confident structures,
353 fragmentation should be done on target components in order to obtain the complete fragmentation
354 tree (Table S2).

355 In addition, this fraction was highly connected with LCC measured in AQ1. Indeed, during LCC
356 fragmentations, molecular formulae of the phenolic part free of the carbohydrate moiety were
357 measured and these formulae were also measured in AQ3 in many cases. For example, for the peak
358 d, phenolic part was C₂₀H₂₀O₇ (Figure 6); this fragment was also measured as deprotonated
359 compounds in AQ3. UV spectrum of peak d was also similar to the peak corresponding to C₂₀H₂₀O₇ in
360 AQ3 (not shown). This observation suggested that LCC are mainly composed of phenolic acid
361 associated with carbohydrates. Since carbohydrates do not absorb in UV, comparison of LCC and
362 phenolic acid UV spectrum should be a supplementary clue for identification in case of the complex
363 was not fragmented.

364 *3.2.3. Organic fraction 2*

365 ORG2 represents a minor part (2% w/w) of our initial sample and is related to compounds that are
366 neutral at pH >12. Mass between 90 and 600 m/z were detected, mainly in positive-ion mode ESI,
367 which suggests carbonyl groups. On the van Krevelen diagram, components are distributed between
368 0.5 and 2.0 for H/C and between 0.1 and 0.5 for O/C but an important part is centered around H/C of
369 1.5 and O/C of 0.2 (Figure S7 (1)). The two most intense peaks observed on UV chromatogram (Figure
370 3) were identified as 5-HMF ($t_R=5.2$ min) and furfural ($t_R=5.5$ min). On the contrary to AQ3, DBE
371 values were comprised between 2 and 8 for 95% of compounds (Figure S7 (2)). Low DBE associated
372 to relative high H/C ratio was a first clue to attest the presence of aliphatic chains. This hypothesis
373 was confirmed with fragmentation experiments and detection of C_4H_8 neutral loss which may come
374 from the fragmentation of aliphatic chains from lipids (Figure S7 (3)). Losses of water were abundant
375 and might come from hydroxyl groups. Moreover, CO losses occurring during the first fragmentation
376 stage (MS^2) may indicate presence of carbonyl functions such as ketone and/or aldehyde groups [33].
377 Another abundant loss was C_2H_2O (ketene) which is known to be characteristic of acetyl group.

378 3.2.4. Organic fraction 3

379 The last fraction (ORG3) represents 4% (w/w) of our initial sample and is expected to be composed of
380 phenols with a pKa varying between 7 and 12. Similarly to AQ3, the van Krevelen diagram of ORG3 is
381 centered around H/C of 1 and O/C of 0.4, corresponding to an aromatic area (Figure S8 (1)). Graph
382 representing DBE as a function of molecular mass also proves the phenolic nature of components in
383 this fraction (Figure S8 (2)). Fragmentations of monoaromatic compounds allowed structural
384 identification with a first methyl radical loss from methoxy group and then a combination of CO and
385 CO_2 losses. Marcum *et al.* already reported a study on fragmentation of small molecules related to
386 lignin which demonstrated that CO losses often occur after the loss of methyl radical during the
387 fragmentation of methoxy group [19]. Hydroxybenzaldehyde, vanillin, syringaldehyde,
388 acetosyringone and coniferaldehyde were identified as monomeric compounds in ORG3 (Table S3).
389 For di, tri and tetra-aromatic compounds, fragmentation patterns depend on the linkage between
390 units. For example, ion at m/z 273.0765 ($C_{15}H_{14}O_5$) was detected at two retention times (12.6 and
391 17.3 min) and had two different fragmentation patterns. The first one implies (CO+ H_2O) fragment as
392 the main MS^2 neutral loss, whereas the second one shows a methyl radical loss. In both cases, the di-
393 aromatic compounds did not fragment in two mono-aromatic units indicating carbon-carbon bond
394 between lignin units such as 5-5, β - β or β -5 links. At the opposite, tetra-aromatic ion at m/z 683.2144
395 ($C_{38}H_{36}O_{12}$) was fragmented in two di-aromatic fragments: $C_{18}H_{17}O_6$ and $C_{20}H_{19}O_6$. Then, $C_{18}H_{17}O_6$ ion
396 fragment lose 3 methyl radical fragments during MS^3 , MS^4 and MS^5 stages, which may correspond to
397 the fragmentation of three methoxy group but also suppose carbon-carbon bound in the di-aromatic
398 ion fragment.

399 In many case, ion fragment corresponding to guaiacol ($C_7H_8O_2$) was measured during the
400 fragmentation of tri and tetra-aromatic compounds (Table S3).

401 4. Conclusion

402 In this work, a multi-technique analytical approach combining liquid-liquid extractions, LC/HRMS and
403 multi-stage fragmentations, was developed to achieve an exhaustive characterization of biomass
404 sample coming from the industrial pretreatment of wheat straw. First, this approach consisted in
405 organizing the biomass sample in four fractions according to chemical family with a reliable liquid-
406 liquid extraction protocol. Then, HPLC/ MS^n experiments were performed using high-resolution LTQ-

407 FT-ICR/MS on the fractions. The combination of LLE protocol and LC separation highly limited the risk
408 of co-elutions and thus ion suppression phenomenon which may occur in atmospheric pressure
409 ionization source. In this way, the composition of the sample should be more representative than
410 using direct introduction or LC/HRMS analysis. Also, isomers could be distinguish which is not
411 possible by direct introduction mass spectrometry. High resolution mass spectrometry allowed to
412 measure elemental compositions with a list of chemical formulae for each fractions. In order to go
413 deeper in the understanding of the biomass composition, multi-stage fragmentations up to MS⁷ were
414 conducted. Typical fragments obtained were in accordance with fraction chemical families. By
415 combining LLE fractions properties, retention time, HPLC-UV spectrum, molecular formula as well as
416 structural information delivered by fragmentation experiments, structures were proposed for
417 compounds up to 600 g/mol. To the best of our knowledge, for the first time carbohydrates with
418 carboxylic acid function and heavy lignin-carbohydrate complexes were elucidated using HPLC/MSⁿ
419 method. The ability of this analytical approach to describe the main chemical families of a
420 lignocellulosic biomass sample is a promising tool and should be applied on several biomass samples
421 in order to progress in the comprehension of relationships between the products composition and
422 reactivity. Moreover, this approach may also be useful during the optimization of pretreatment
423 processes in order to observe the impact of the different process parameters (temperature, time
424 reaction, etc.) on the chemical composition of liquid samples.

425 5. References

- 426 [1] J. Mohtasham, Review Article-Renewable Energies, Energy Procedia 74 (2015) 1289–1297.
427 <https://doi.org/10.1016/j.egypro.2015.07.774>.
- 428 [2] I. Dincer, C. Acar, A review on clean energy solutions for better sustainability, Int. J. Energy Res.
429 39 (5) (2015) 585–606. <https://doi.org/10.1002/er.3329>.
- 430 [3] N. Abas, A. Kalair, N. Khan, Review of fossil fuels and future energy technologies, Futures 69
431 (2015) 31–49. <https://doi.org/10.1016/j.futures.2015.03.003>.
- 432 [4] F.H. Isikgor, C.R. Becer, Lignocellulosic biomass: A sustainable platform for the production of
433 bio-based chemicals and polymers, Polym. Chem. 6 (25) (2015) 4497–4559.
434 <https://doi.org/10.1039/C5PY00263J>.
- 435 [5] A. Brandt, J. Gräsvik, J.P. Hallett, T. Welton, Deconstruction of lignocellulosic biomass with ionic
436 liquids, Green Chem. 15 (3) (2013) 550. <https://doi.org/10.1039/c2gc36364j>.
- 437 [6] P.F.H. Harmsen, W.J.J. Huijgen, L.M. Bermudez Lopez, R.R.C. Bakker, Literature review of
438 physical and chemical pretreatment processes for lignocellulosic biomass, 2010 (accessed 3
439 November 2017).
- 440 [7] P. Gallezot, Conversion of biomass to selected chemical products, Chem. Soc. Rev. 41 (4) (2012)
441 1538–1558. <https://doi.org/10.1039/c1cs15147a>.
- 442 [8] P. Kumar, D.M. Barrett, M.J. Delwiche, P. Stroeve, Methods for Pretreatment of Lignocellulosic
443 Biomass for Efficient Hydrolysis and Biofuel Production, Ind. Eng. Chem. Res. 48 (8) (2009)
444 3713–3729. <https://doi.org/10.1021/ie801542g>.
- 445 [9] L.J. Jönsson, B. Alriksson, N.-O. Nilvebrant, Bioconversion of lignocellulose: inhibitors and
446 detoxification, Biotechnol. Biofuels 6 (1) (2013) 16. <https://doi.org/10.1186/1754-6834-6-16>.
- 447 [10] H.B. Klinker, A.B. Thomsen, B.K. Ahring, Inhibition of ethanol-producing yeast and bacteria by
448 degradation products produced during pre-treatment of biomass, Appl. Microbiol. Biotechnol.
449 66 (1) (2004) 10–26. <https://doi.org/10.1007/s00253-004-1642-2>.

- 450 [11] C. Luo, D.L. Brink, H.W. Blanch, Identification of potential fermentation inhibitors in conversion
451 of hybrid poplar hydrolyzate to ethanol, *Biomass and Bioenergy* 22 (2) (2002) 125–138.
452 [https://doi.org/10.1016/S0961-9534\(01\)00061-7](https://doi.org/10.1016/S0961-9534(01)00061-7).
- 453 [12] B. Du, L.N. Sharma, C. Becker, S.-F. Chen, R.A. Mowery, G.P. van Walsum, C.K. Chambliss, Effect
454 of varying feedstock-pretreatment chemistry combinations on the formation and accumulation
455 of potentially inhibitory degradation products in biomass hydrolysates, *Biotechnol. Bioeng.* 107
456 (3) (2010) 430–440. <https://doi.org/10.1002/bit.22829>.
- 457 [13] Y. Liu, Q. Shi, Y. Zhang, Y. He, K.H. Chung, S. Zhao, C. Xu, Characterization of Red Pine Pyrolysis
458 Bio-oil by Gas Chromatography–Mass Spectrometry and Negative-Ion Electrospray Ionization
459 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, *Energy Fuels* 26 (7) (2012)
460 4532–4539. <https://doi.org/10.1021/ef300501t>.
- 461 [14] J.H. Marsman, J. Wildschut, F. Mahfud, H.J. Heeres, Identification of components in fast
462 pyrolysis oil and upgraded products by comprehensive two-dimensional gas chromatography
463 and flame ionisation detection, *J. Chromatogr. A* 1150 (1-2) (2007) 21–27.
464 <https://doi.org/10.1016/j.chroma.2006.11.047>.
- 465 [15] T. Sfetsas, C. Michailof, A. Lappas, Q. Li, B. Kneale, Qualitative and quantitative analysis of
466 pyrolysis oil by gas chromatography with flame ionization detection and comprehensive two-
467 dimensional gas chromatography with time-of-flight mass spectrometry, *J. Chromatogr. A* 1218
468 (21) (2011) 3317–3325. <https://doi.org/10.1016/j.chroma.2010.10.034>.
- 469 [16] M. Staš, J. Chudoba, D. Kubička, J. Blažek, M. Pospíšil, Petroleomic Characterization of Pyrolysis
470 Bio-oils: A Review, *Energy Fuels* 31 (10) (2017) 10283–10299.
471 <https://doi.org/10.1021/acs.energyfuels.7b00826>.
- 472 [17] H. Sheng, W. Tang, J. Gao, J.S. Riedeman, G. Li, T.M. Jarrell, M.R. Hurt, L. Yang, P. Murria, X. Ma,
473 J.J. Nash, H.I. Kenttämä, (-)ESI/CAD MS_n Procedure for Sequencing Lignin Oligomers Based on a
474 Study of Synthetic Model Compounds with β -O-4 and 5-5 Linkages, *Anal. Chem.* 89 (24) (2017)
475 13089–13096. <https://doi.org/10.1021/acs.analchem.7b01911>.
- 476 [18] B.C. Owen, L.J. Hauptert, T.M. Jarrell, C.L. Marcum, T.H. Parsell, M.M. Abu-Omar, J.J. Bozell, S.K.
477 Black, H.I. Kenttämä, High-performance liquid chromatography/high-resolution multiple stage
478 tandem mass spectrometry using negative-ion-mode hydroxide-doped electrospray ionization
479 for the characterization of lignin degradation products, *Anal. Chem.* 84 (14) (2012) 6000–6007.
480 <https://doi.org/10.1021/ac300762y>.
- 481 [19] C.L. Marcum, T.M. Jarrell, H. Zhu, B.C. Owen, L.J. Hauptert, M. Easton, O. Hosseinaei, J. Bozell, J.J.
482 Nash, H.I. Kenttämä, A Fundamental Tandem Mass Spectrometry Study of the Collision-
483 Activated Dissociation of Small Deprotonated Molecules Related to Lignin, *ChemSusChem* 9 (24)
484 (2016) 3513–3526. <https://doi.org/10.1002/cssc.201600678>.
- 485 [20] E. Kiyota, P. Mazzafera, Sawaya, Alexandra C H F, Analysis of soluble lignin in sugarcane by
486 ultrahigh performance liquid chromatography-tandem mass spectrometry with a do-it-yourself
487 oligomer database, *Anal. Chem.* 84 (16) (2012) 7015–7020. <https://doi.org/10.1021/ac301112y>.
- 488 [21] J. Hertzog, V. Carré, Y. Le Brech, A. Dufour, F. Aubriet, Toward Controlled Ionization Conditions
489 for ESI-FT-ICR-MS Analysis of Bio-Oils from Lignocellulosic Material, *Energy Fuels* 30 (7) (2016)
490 5729–5739. <https://doi.org/10.1021/acs.energyfuels.6b00655>.
- 491 [22] T.M. Jarrell, C.L. Marcum, H. Sheng, B.C. Owen, C.J. O'Lenick, H. Maraun, J.J. Bozell, H.I.
492 Kenttämä, Characterization of organosolv switchgrass lignin by using high performance liquid
493 chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-

- 494 ion mode electrospray ionization, *Green Chem* 16 (5) (2014) 2713–2727.
495 <https://doi.org/10.1039/C3GC42355G>.
- 496 [23] S.-F. Chen, R.A. Mowery, V.A. Castleberry, G.P. van Walsum, C.K. Chambliss, High-performance
497 liquid chromatography method for simultaneous determination of aliphatic acid, aromatic acid
498 and neutral degradation products in biomass pretreatment hydrolysates, *J. Chromatogr. A* 1104
499 (1-2) (2006) 54–61. <https://doi.org/10.1016/j.chroma.2005.11.136>.
- 500 [24] P.K. Kanaujia, D.V. Naik, D. Tripathi, R. Singh, M.K. Poddar, L.S.K. Konathala, Y.K. Sharma,
501 Pyrolysis of *Jatropha Curcas* seed cake followed by optimization of liquid-liquid extraction
502 procedure for the obtained bio-oil, *Journal of Analytical and Applied Pyrolysis* 118 (2016) 202–
503 224. <https://doi.org/10.1016/j.jaap.2016.02.005>.
- 504 [25] Y. Wei, H. Lei, L. Wang, L. Zhu, X. Zhang, Y. Liu, S. Chen, B. Ahring, Liquid–Liquid Extraction of
505 Biomass Pyrolysis Bio-oil, *Energy Fuels* 28 (2) (2014) 1207–1212.
506 <https://doi.org/10.1021/ef402490s>.
- 507 [26] S.-F. Chen, High-Performance Liquid Chromatographic Methods for Quantitative Assessment of
508 Degradation Products and Extractives in Pretreated Lignocellulose, 2007.
- 509 [27] A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, and D. Templeton: NREL, Determination of
510 Sugars, Byproducts, and Degradation Products in Liquid Fraction Process Samples: Laboratory
511 Analytical Procedure (LAP); Issue Date: 12/08/2006.
- 512 [28] A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, and D. Crocker: NREL,
513 Determination of Structural Carbohydrates and Lignin in Biomass: Laboratory Analytical
514 Procedure (LAP) (Revised July 2011).
- 515 [29] C. Reymond, A. Le Masle, C. Colas, N. Charon, A rational strategy based on experimental designs
516 to optimize parameters of a liquid chromatography-mass spectrometry analysis of complex
517 matrices, *Talanta* 205 (2019) 120063. <https://doi.org/10.1016/j.talanta.2019.06.063>.
- 518 [30] T.-Q. Yuan, S.-N. Sun, F. Xu, R.-C. Sun, Characterization of lignin structures and lignin-
519 carbohydrate complex (LCC) linkages by quantitative ¹³C and 2D HSQC NMR spectroscopy, *J.*
520 *Agric. Food Chem.* 59 (19) (2011) 10604–10614. <https://doi.org/10.1021/jf2031549>.
- 521 [31] K.S. Boes, R.H. Narron, S. Park, N.R. Vinueza, Mass Spectrometry Exposes Undocumented Lignin-
522 Carbohydrate Complexes in Biorefinery Pretreatment Stream, *ACS Sustainable Chem. Eng.* 6 (8)
523 (2018) 10654–10659. <https://doi.org/10.1021/acssuschemeng.8b01986>.
- 524 [32] B. Domon, C.E. Costello, A systematic nomenclature for carbohydrate fragmentations in FAB-
525 MS/MS spectra of glycoconjugates, *Glycoconjugate J* 5 (4) (1988) 397–409.
526 <https://doi.org/10.1007/BF01049915>.
- 527 [33] L.M. Amundson, V.A. Gallardo, N.R. Vinueza, B.C. Owen, J.N. Reece, S.C. Habicht, M. Fu, R.C.
528 Shea, A.B. Mossman, H.I. Kenttämä, Identification and Counting of Oxygen Functionalities and
529 Alkyl Groups of Aromatic Analytes in Mixtures by Positive-Mode Atmospheric Pressure Chemical
530 Ionization Tandem Mass Spectrometry Coupled with High-Performance Liquid Chromatography,
531 *Energy Fuels* 26 (5) (2012) 2975–2989. <https://doi.org/10.1021/ef2019098>.

532

533 **Figure captions**

534 **Table 1:** Experimental conditions for ESI and APCI in positive and negative modes

535

536 **Figure 1:** A selective LLE fractionation for aqueous biomass samples

537 **Figure 2:** HPLC-UV chromatograms of the whole sample (black chromatogram) and LLE fractions
538 obtained (AQ1: green, AQ3: purple, ORG2: blue, ORG3: red) at 254 nm
539 **Figure 3:** van Krevelen diagram using all detection modes (ESI+/- and APCI+/-) for each LLE fraction
540 **Figure 4:** van Krevelen diagram for AQ1. Colors were used to differentiate compounds eluted at the
541 beginning of the run (with less than 12% of MeOH in the mobile phase in orange and green) and
542 those eluted later (black and grey). Positive and negative ion-mode detections were also
543 differentiate.
544 **Figure 5:** Base peak chromatogram of AQ1 in negative-ion mode ESI and associated van Krevelen
545 diagram for C_xH_yO_z compounds with red dots for the annotated peaks
546 **Figure 6:** Proposition of fragmentation scheme from MS¹ to MS⁷ of compound (b) in negative-ion
547 mode ESI and positive-ion mode ESI, compounds (a), (c) and (d) in negative-ion mode ESI for fraction
548 AQ1. Relative intensity of each fragment is given in bracket
549 **Figure 7:** MS² mass spectra of peak α (red), β (blue) and γ (green) present in fraction AQ1 in negative-
550 ion mode ESI
551 **Figure 8:** AQ3 van Krevelen diagram (1), DBE as a function of molecular mass (2) and losses in
552 negative-ion mode ESI for MS² (red dots) until MS⁷ (black outline circles) according to retention time
553 (3)

Figure 1

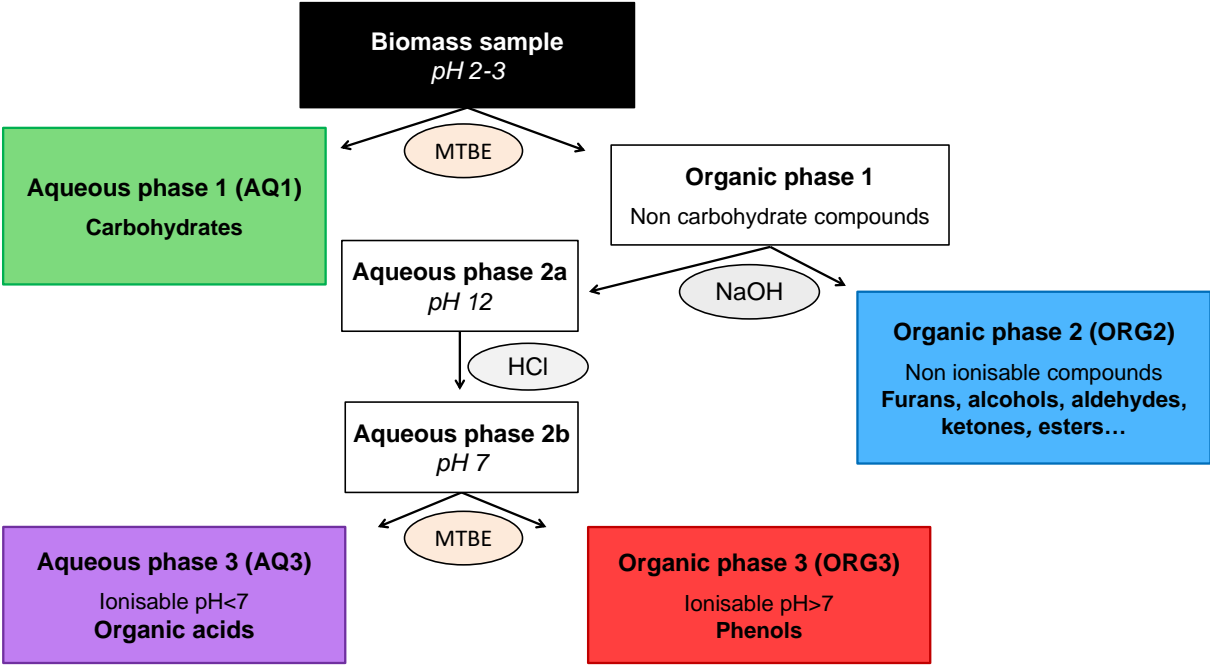


Figure 2

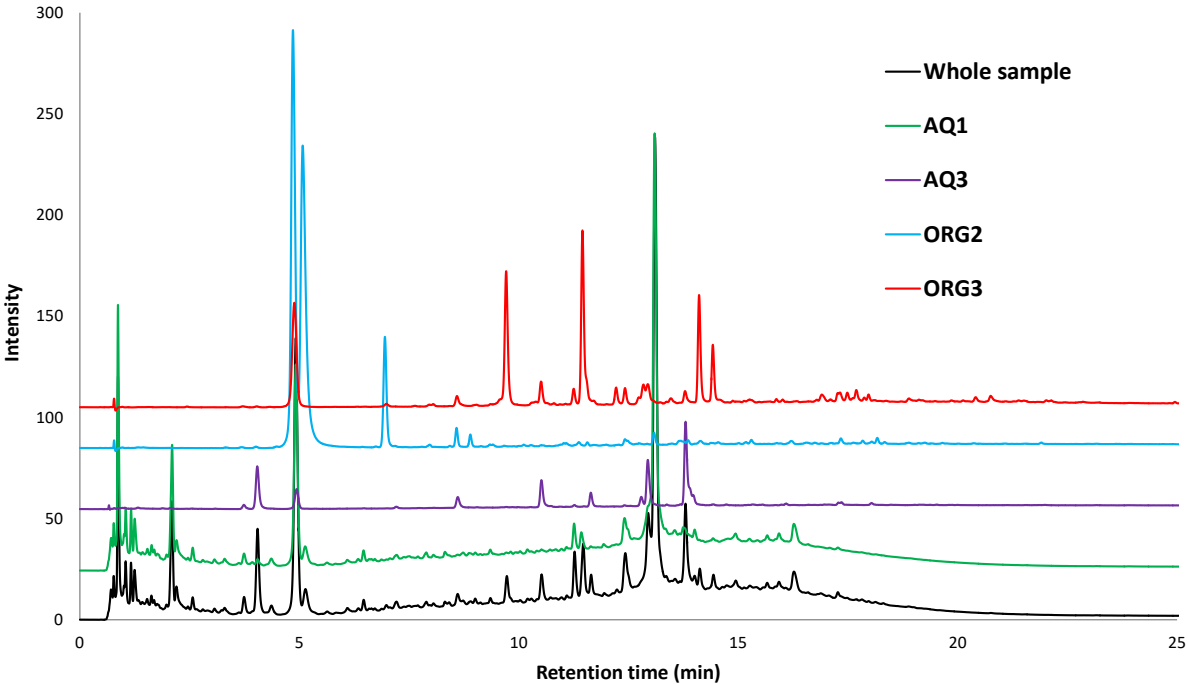


Figure 3

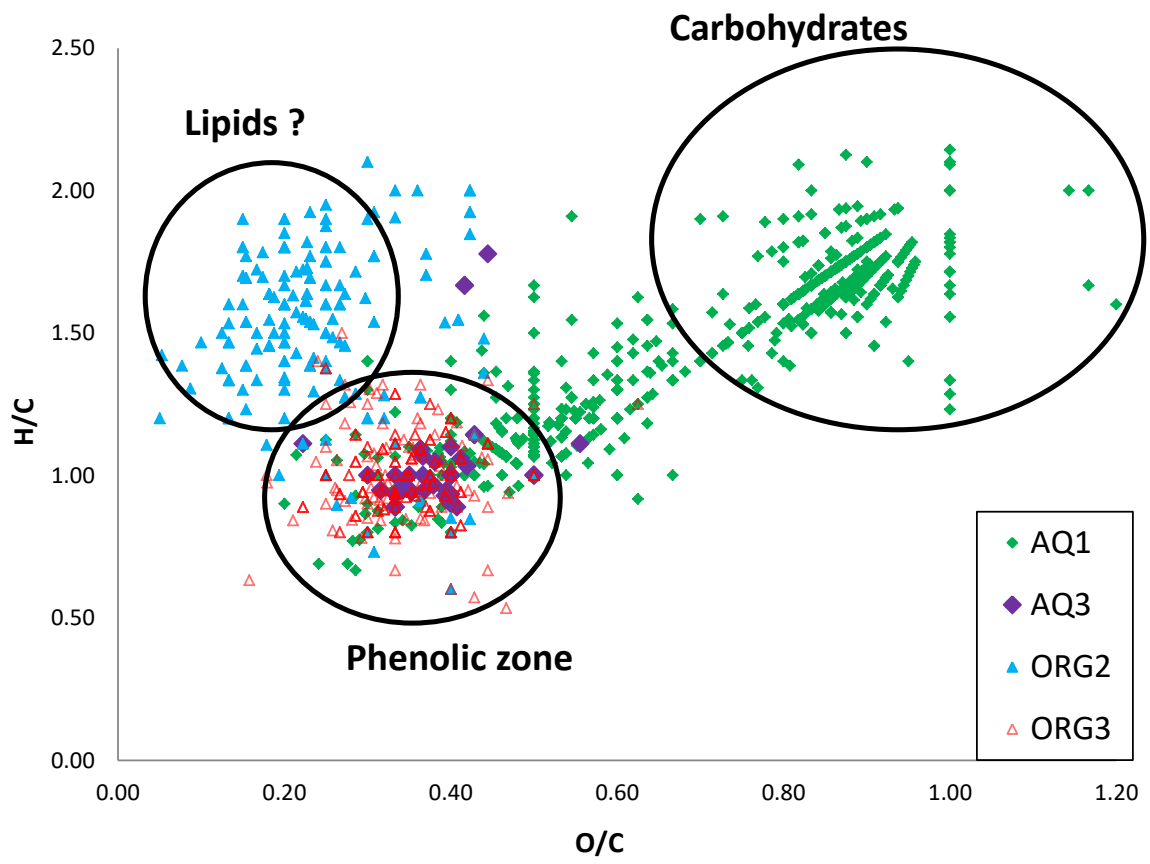


Figure 4

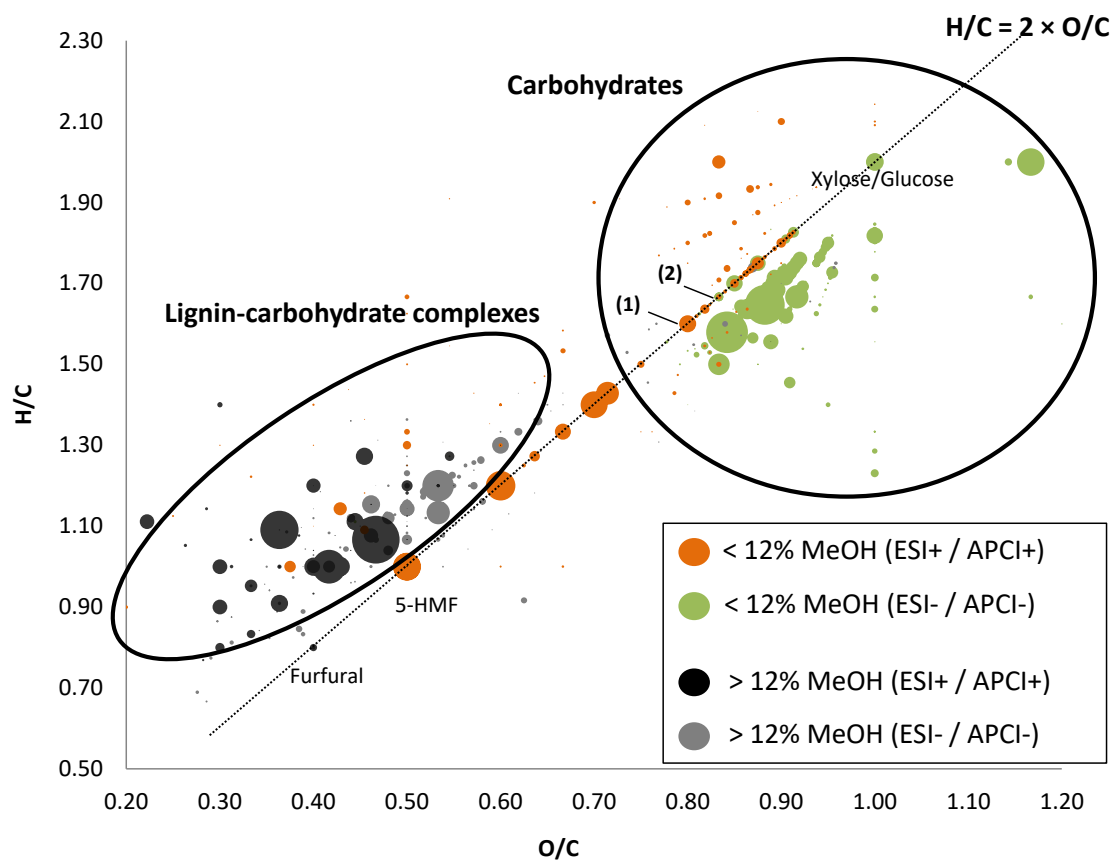


Figure 5

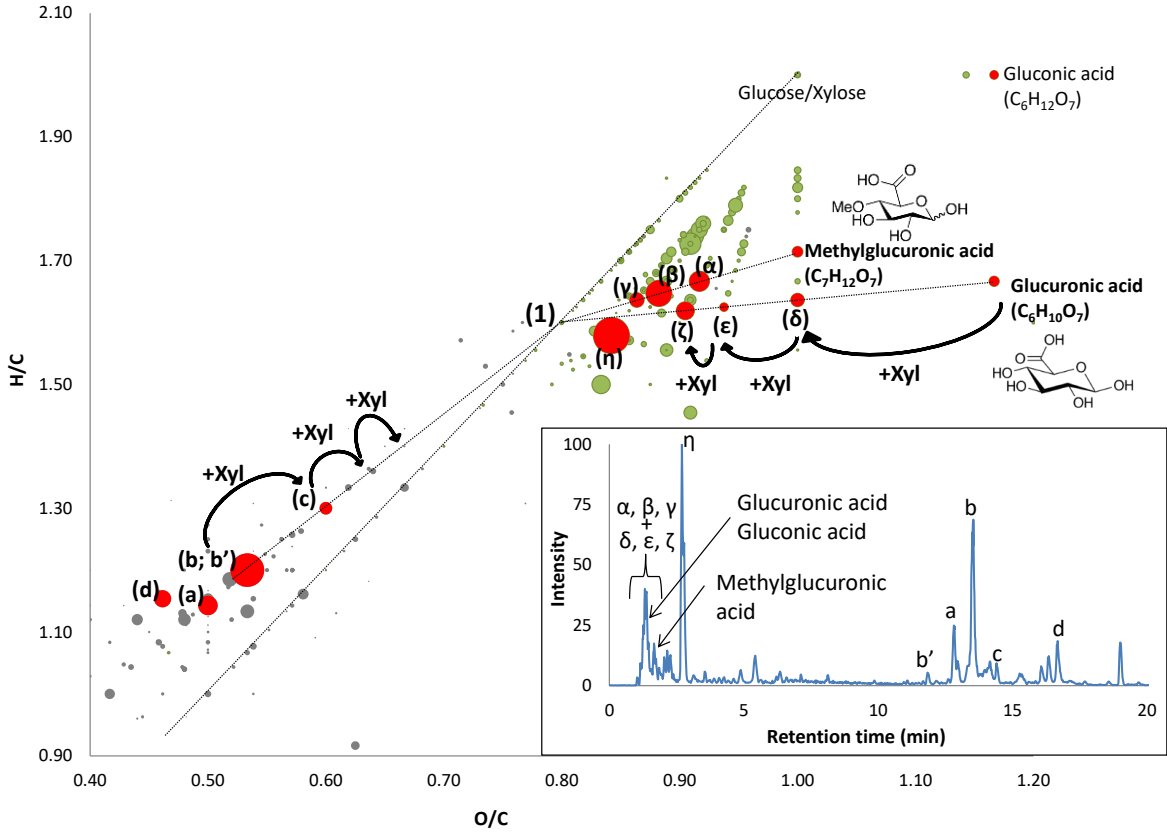


Figure 6

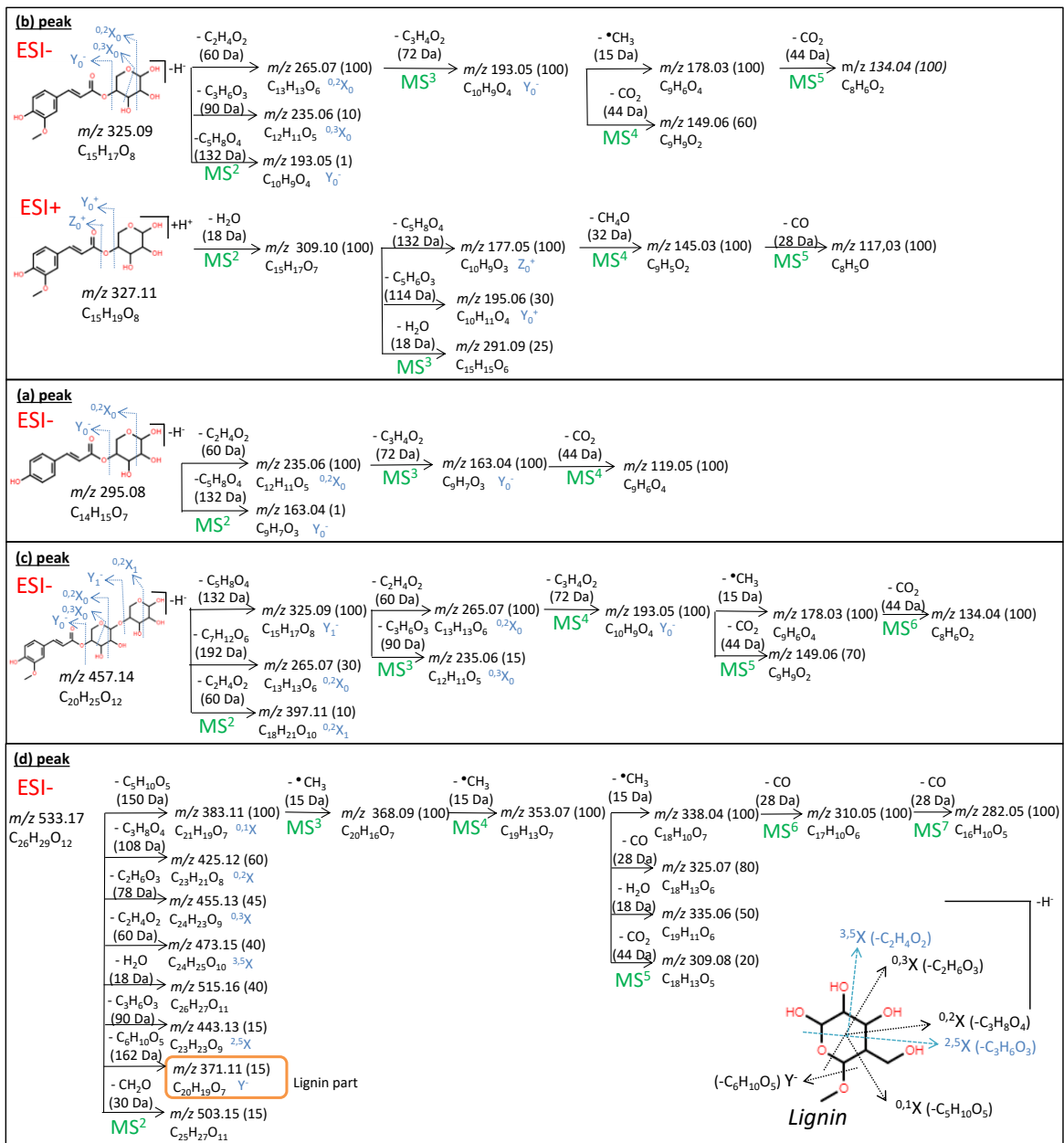


Figure 7

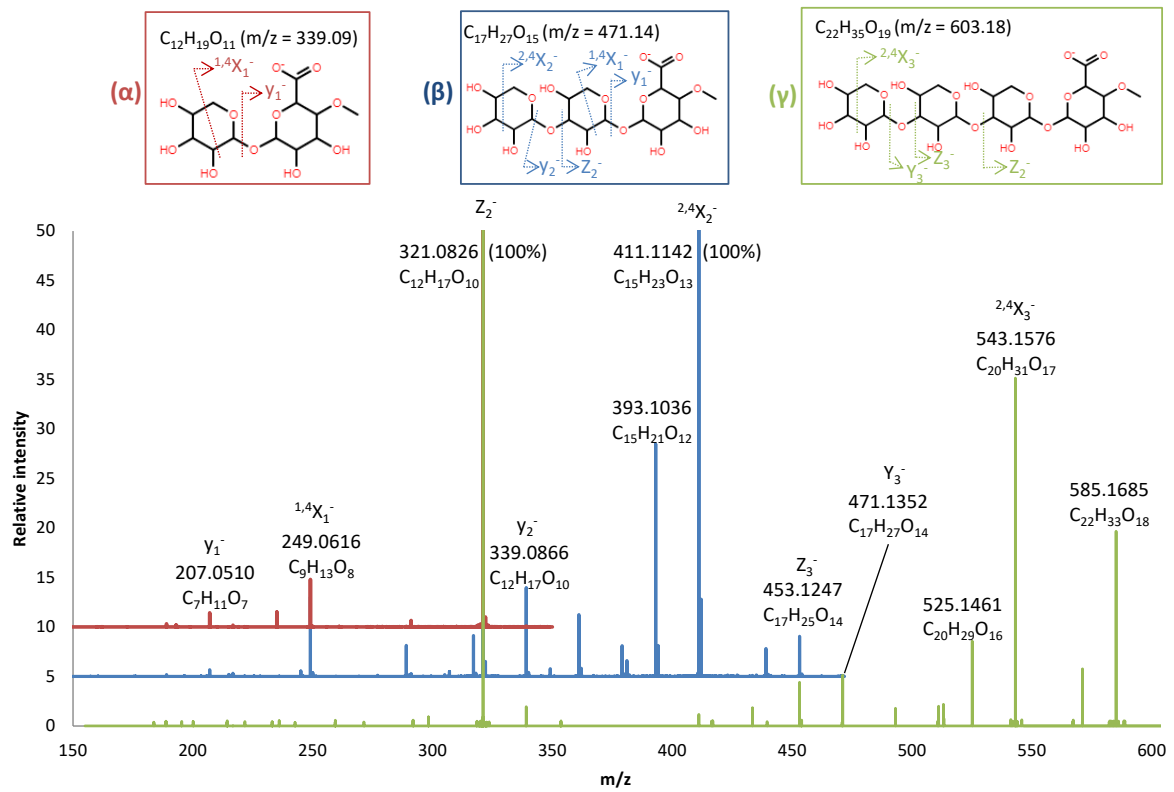


Figure 8

