**Purification, structural characterization and modification of organosolv wheat straw lignin**

L. Mbotchak, C. Le Morvan, K. L. Duong, B. Rousseau, M. Tessier, A. Fradet*

**Supporting Information**

**$^1$H NMR spectra.**

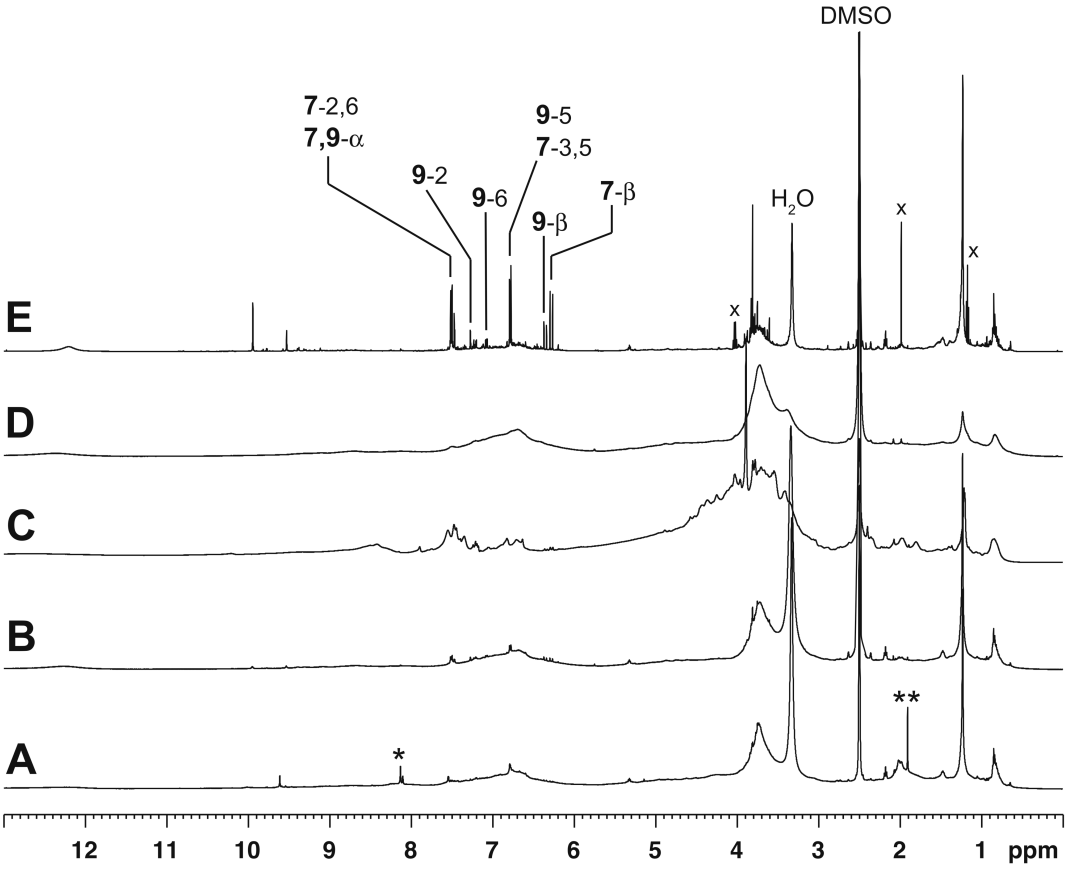


**Figure S1**. $^1$H NMR spectra of: A. Biolignin$^{TM}$; B. hydrolyzed Biolignin$^{TM}$ (Lh); C. purified Biolignin$^{TM}$ (Lp); D. hydrolysis residue (Rh); and E. ethyl acetate extraction residue (Rp). x: residual ethyl acetate ; *: formic acid ; **: acetic acid. Structure numbering is given in main text.

*Biolignin$^{TM}$:*

- 6-8 ppm: broad signal of aromatics with well-resolved peaks indicating the presence of low molar mass aromatics,
- 3-6 ppm: broad signal of lignin side chains and polysaccharides,
- 3.72 ppm: methoxy groups,
- 0.83 and 1.23 ppm: non-lignin aliphatic methyl and methylene groups,
- 1.48 and 2.17 ppm: methylenes in β and α position of carboxyl groups, respectively,
- 5.32 ppm: unsaturated double bonds of fatty acids,

- 2.0 ppm: methyl group of acetates.
- 1.91 and 8.13 ppm: residual acetic and formic acids from the delignification process.

*Hydrolyzed Biolignin$^{TM}$, Lh:*
- The peaks at 1.91, 2.0 and 8.13 ppm (acetates and residual acetic and formic acid) have disappeared.

*Water-soluble hydrolysis residue, Rh:*
- 1.5-2.5 and 3.0-4.5 ppm: complex signals of polysaccharides,
- 6-9 ppm: low molar mass aromatic compounds.

The alkaline hydrolysis leads to the cleavage of ester bonds, including those involved in carbohydrate-lignin complexes, and allows the elimination of residual polysaccharides, together with some aromatic compounds.

*Purified Biolignin$^{TM}$, Lp:*
- aromatic region: no residual sharp signals,
- aliphatic region (0.8-2.2 ppm): large decrease in peak intensity, especially those at 0.83 and 1.23 ppm.

*Organo-soluble residue, Rp:*
- saturated (0.8-2 ppm) and unsaturated (double bond at 5.32 ppm) fatty acids.
- *p*-coumaric acid (**7**): resonances at 6.28 (**7**-β), 6.80 (**7**-3,5) and 7.49-7.50 ppm (**7**-α and **7**-2,6).
- ferulic acid (**9**): 6.35 (**9**-β), 7.08 (**9**-6) and 7.27 ppm (**9**-2).

The $^{1}$H NMR study shows that the consecutive hydrolysis and solvent extraction treatment of Biolignin$^{TM}$ produce a final purified Biolignin$^{TM}$ sample, Lp, with highly reduced contaminant content.

**HSQC 2D NMR spectra of hydrolysis (Rh) and solvent extraction (Rp) residues**

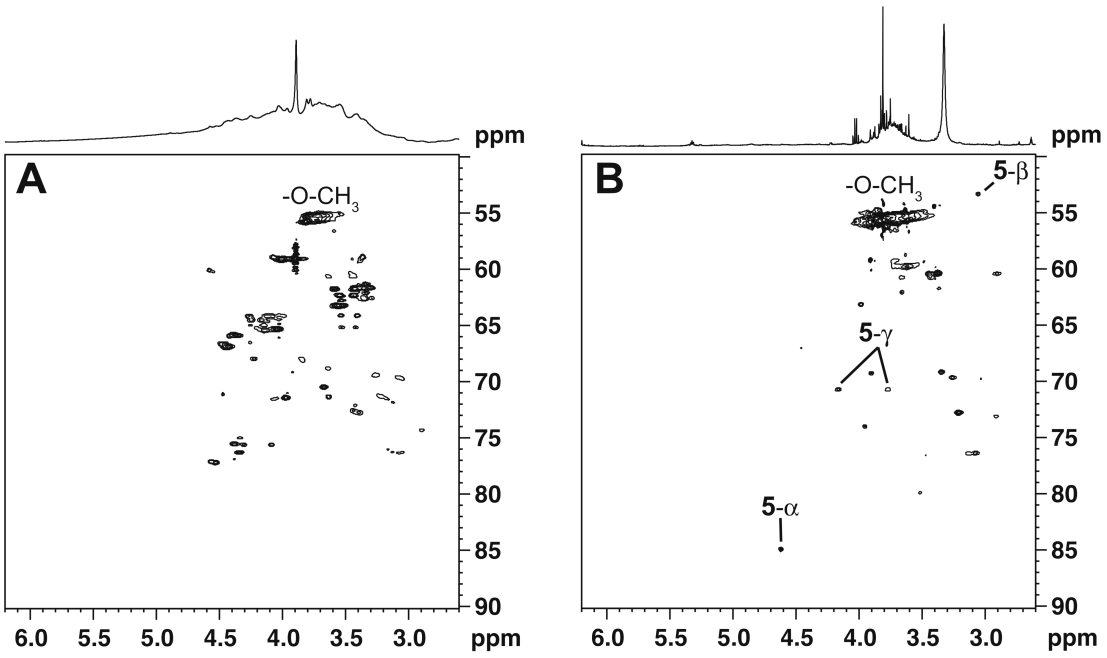


**Figure S2.** 2D $^{13}C$-$^{1}H$ HSQC NMR spectra of: A. Rh; and B. Rp. Expansion of the lignin side chain domain. Structure numbering is given in main text.

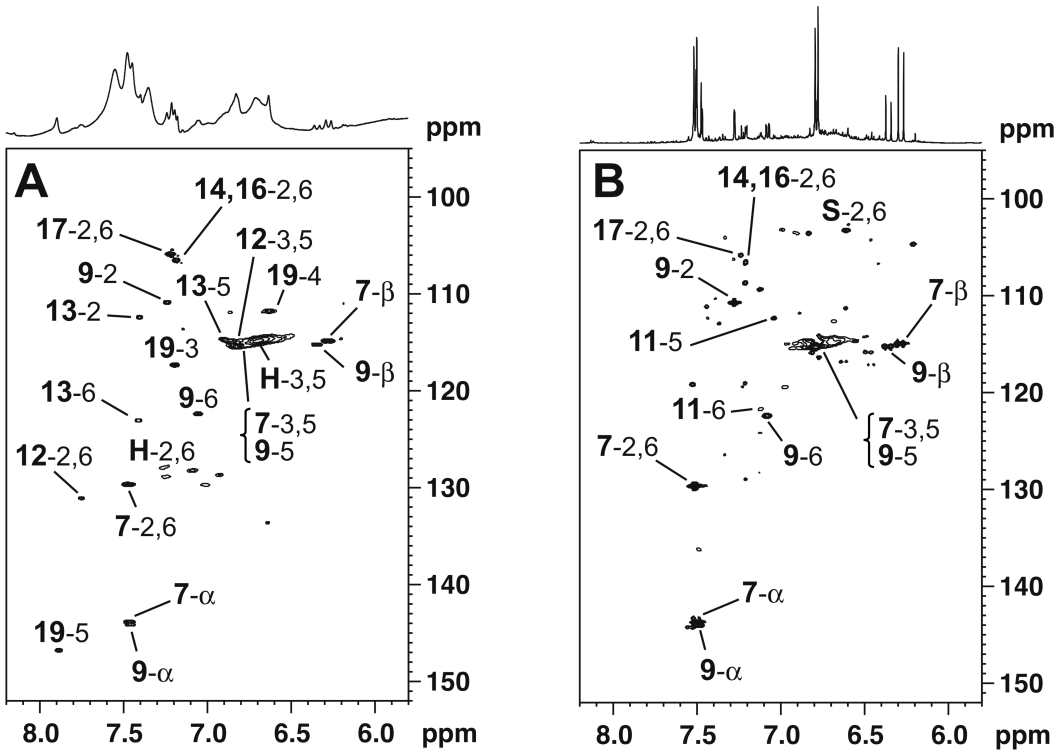


**Figure S3.** 2D $^{13}C$-$^{1}H$ HSQC NMR spectra of: A. Rh; and B. Rp. Expansion of the aromatic domain. Structure numbering is given in main text.

*Aliphatic domain (Figure S2):*
- Rh spectrum: carbohydrate resonances (not discussed in this work),
- Rp spectrum: none of the typical lignin signals are observed, except the correlations of β-β interunit linkages (**5**).

*Aromatic domain (Figure S3):*
- Low molar mass aromatic compounds, identified by comparison with the spectra of authentic samples: *p*-hydroxybenzoic acids (**12**, **13** and **14**), *p*-hydroxycinnamic acids (**7** and **9**) and carbonyl derivatives (**16** and **17**).
- Furoic acid (**19**) (Rh spectrum): correlations at 117.6/7.20, 112.0/6.64 and 147.0/7.90 ppm, (positions 3, 4 and 5 of furan ring, respectively), identified by comparison with the spectra of an authentic sample. This compound might be formed during the alkaline hydrolysis step by oxidation or by Cannizzaro or crossed Cannizzaro reactions involving furaldehyde and other aldehydes initially present in Biolignin$^{TM}$, such as vanillin. This is supported by the fact that vanillic acid, **13**, is not present in Biolignin$^{TM}$, but is detected in Rh after basic hydrolysis.
- S-2,6: β-β dimer deriving from sinapyl alcohol.

**HMBC 2D NMR spectrum of solvent extraction residue (Rp)**
- Syringaldehyde and syringic acid give a single cross-peak in HSQC (Figure S3), but are differentiated by their $^{13}C$–$^{1}H$ HMBC long-range correlations between carbonyls and protons in 2 and 6 positions at 191.05/7.21 and 167.2/7.20 ppm, respectively (Figure S4).
- The aromatic rings linked to the α positions of β-β interunit linkages, **5**, are exclusively of **S** type (long range $^{13}C$–$^{1}H$ **5**-α/**S**-2,6 correlation at 84.8/6.66 ppm).

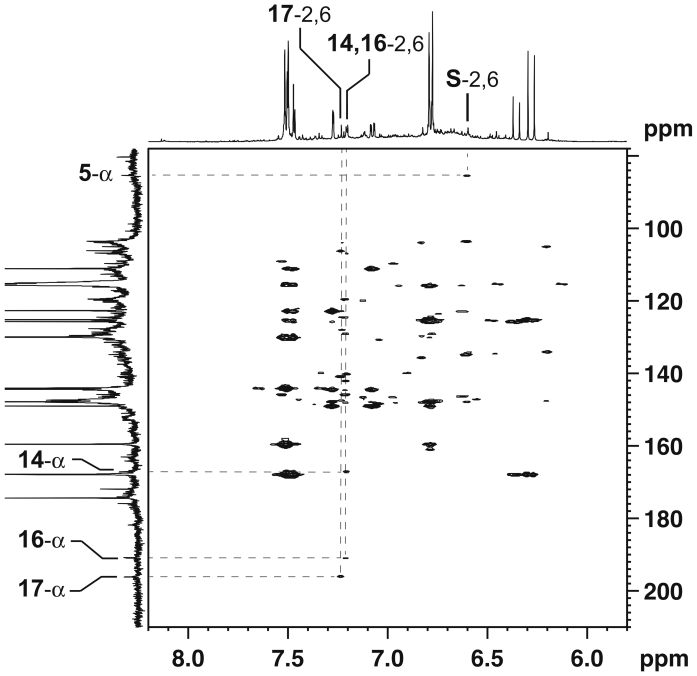


**Figure S4.** 2D $^{13}C$-$^{1}H$ HMBC NMR spectrum of Rp. Expansion of the 80-200/6-8 ppm domain. Structure numbering is given in main text.

## HMBC 2D NMR spectrum of s-Lp-Et

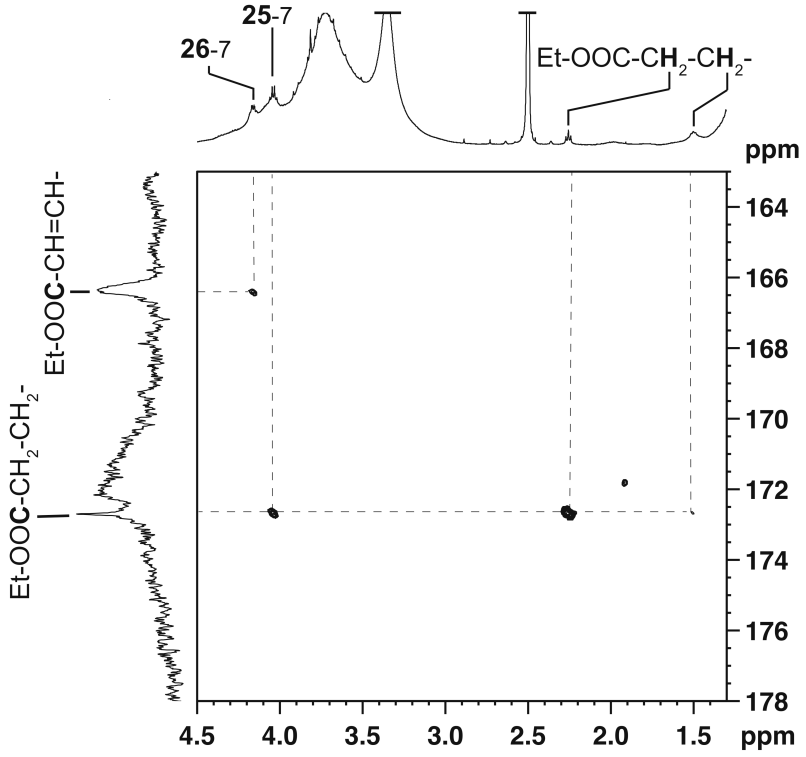


**Figure S5.** 2D $^{13}C$-$^{1}H$ HMBC NMR spectrum of s-Lp-Et. Expansion of the 164-178/1.5-4.5 ppm domain showing the long-distance correlations corresponding to ethyl esters. Structure numbering is given in main text.

## $^{13}$C NMR spectra of Lp, s-Lp-Et and Lp-EG

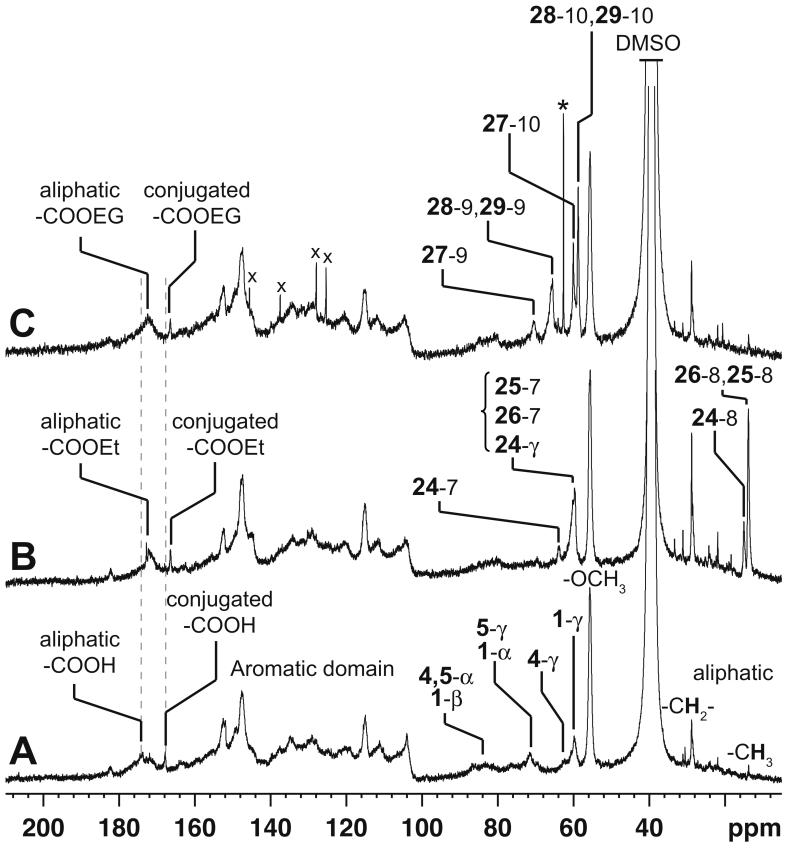


**Figure S6.** $^{13}$C NMR spectra of: A. purified Biolignin$^{TM}$, Lp; B. ethanol-modified lignin s-Lp-Et; and C. ethane-1,2-diol-modified lignin Lp-EG, showing the shift of COOH resonances to COOR resonances. x: Residual p-toluenesulfonic acid; *: residual ethane-1,2-diol; -COOEt: ethyl esters; -COOEG: hydroxyethyl esters. Structure numbering is given in main text.