Supporting Information for:

Structure-Reactivity Relationships Governing Hydrothermal Liquefaction of Lignin from Co-solvent Enhanced Lignocellulosic Fractionation (CELF)

Heather O. LeClerc,[1] Ronish M. Shrestha,[1] Feng Cheng,[1] Alex R. Maag,[1] Geoffrey A. Tompsett,[1] Brent Scheidemantle,[2] Zhaoxi Zheng,[3] Klaus Schmidt-Rohr,[3] Amy M. McKenna,[4,5] Sydney Niles,[4] Jialiang Zhang,[6] Marcus Foston,[6] Charles M. Cai,[2] Andrew R. Teixeira,[1] Michael T. Timko[1*]

1. *Department of Chemical Engineering, Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609, USA*
2. *Bourns College of Engineering-Center for Environmental Research and Technology (CE-CERT), University of California, Riverside, CA 92507, USA*
3. *Department of Chemistry, Brandeis University, 415 South Street, Waltham, MA 02453, USA*
4. *National High Magnetic Field Laboratory, 1800 Paul Dirac Dr., Tallahassee, FL 32310, USA*
5. *Department of Soil & Crop Sciences. Colorado State University, Fort Collins, Colorado 80523-1170, USA*
6. *Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130-4899, USA*

*Corresponding author, to whom correspondence should be addressed.

Lignin S/G and H/(S+G) ratios from solid-state NMR

The ratios of integrated peak intensities in dipolar dephased solid-state ^{13}C NMR spectra provide a general and relatively robust approach for determining the syringol (S) fraction in native and extracted lignins. This convenient spectral editing by gated decoupling[1, 2], also known as dipolar dephasing, provides selective spectra of C not bonded to H and of CH_3 groups (due to their fast three-fold rotational jump motion that partially averages the C-H dipolar couplings).

In the analysis of the ^{13}C NMR spectrum after dipolar dephasing, the peak area between 157 ppm and 142 ppm, centered on 149.5 ppm and therefore denoted as I_{149}, is set to 2.00, since two aromatic C-O carbons of most syringol or guaiacol (G) units, whether polymerized or hydrolyzed, resonate in this range (while the third, central C-O of most syringol units resonates below 142 ppm). Note that the C-O of *para*-hydroxy units is fairly well resolved from this spectral region, resonating above 157 ppm. Thus, the total S plus G amount is

$$S+G = c\ I_{149} \times 1.05^2/0.88/2 = c\ 1.25\ I_{149}/2 \qquad (1)$$

with a proportionality factor c that will later cancel. The factor of 0.88 takes into account the slight dipolar dephasing of C-O by long-range dipolar couplings; it can be directly verified by comparison with the aromatic C-O intensity in the full (undephased) ^{13}C spectrum of hard- or softwood. The dipolar dephasing removes background near 147 ppm from =CH of ferulate and similar alkenes, which are commonly found in grasses (see **Figure SI-6**). The factor of 1.05^2 in eq.(1) adds in the signal of spinning sidebands to the centerband integral and compensates for 5% lower cross-polarization efficiency relative to OCH_3.

While a syringol unit contains two OCH_3 groups, a guaiacol unit has only one OCH_3. Thus, the methoxy amount is

$$M = 2\ S + G \qquad (2)$$

This means that the methoxy intensity I_{55} at 56.6 ± 4 ppm is higher in syringol-rich samples. The overlap of OCH_3 with alicyclic CH in guaiacol lignin is removed by dipolar dephasing. However, like all CH_3 groups in a rigid environment, OCH_3 in lignin dephases to 57% of its full intensity during our standard recoupled dipolar dephasing of (30 + 8 + 30) μs.[1] The dephasing factor can be directly verified in clean hardwood by comparison with the full spectrum, taking into account a few percent of alicyclic CH background. The methoxy amount relates to the peak area between 50 and 60 ppm, I_{55}, by

$$M = c\ I_{55}/0.57/1.05 \qquad (3)$$

with the same proportionality factor c as above. Combining eqs.(1) and (3), the average OCH_3 number per S or G aromatic ring is

$$M/(S+G) = c\ I_{55}/0.57/(c\ 1.25\ I_{149}/2) = 1.4\ (2/I_{149})\ \ I_{55} \qquad (4)$$

According to eqs.(2) and (4),

$$[S + (S+G)]/(S+G) = M/(S+G) = 1.4\ (2/I_{149})\ \ I_{55} \qquad (5)$$

and the fraction of S is simply

$$S/(S+G) = 1.4\ I_{55}\ (2/I_{149}) - 1 \quad (6)$$

If the integral I_{149} in the dipolar-dephased spectrum was set to 2.00, which is easy to do in standard NMR spectral integration routines, this simplifies to

$$S/(S+G) = 1.4\ I_{55} - 1. \quad (7)$$

with the prefactor due to the dipolar-dephasing and spinning-sideband factors. This simplifying normalization of the integrals will be assumed in the following.

Using

$$(S+G)/S = 1 + 1/(S/G) \quad (8)$$

solved for

$$S/G = ([S/(S+G)]^{-1} - 1)^{-1} = ([1.4\ I_{55} - 1]^{-1} - 1)^{-1} \quad (9)$$

the OCH_3 peak integral in the dipolar dephased spectrum yields S/G. For instance, if $I_{55} = 1.03$, then $1.4\ I_{55} = 1.5$ (this is the average OCH_3 number per S or G aromatic ring according to eq.(M/(S+G) = … = $1.4\ (2/I_{149})\ \ I_{55}$)) and S/G = 1. If $I_{55} = 0.68$, then $1.4\ I_{55} = 1$ OCH_3 per ring and S/G = 0.00. For native lignins in biomass, S/G ≥ 0 in theory and in our experimental data, see **Table SI-1**.

Background signals near 55 ppm need to be considered carefully in order to achieve an accurate estimate of S/G. Most samples show a nearly uniform low background (even after dipolar dephasing), which integrated from 39 to 51 ppm amounts to 2 to 7% of the OCH_3 signal and is subtracted out. Furthermore, a foot or shoulder near 62 ppm is observed in most samples, see **Figures 1 and SI-1, 2**, accounting for an additional 10 to 20% compared to the signal between 52 and 60 ppm. Its fraction is largest in hardwood CELF lignins. Careful analysis has indicated that it has different origins in biomass and aggressively extracted lignins. Note that spectral coincidence does not prove that two chemical species are identical; for instance, hydrolysis shifts the aromatic C-O resonances of syringol to overlap with the main C-O peak of guaiacol near 148 ppm.

In fresh biomass, the shoulder near 62 ppm arises mainly from mobile OCH_2 and OCH_3 (e.g. methyl ester) groups of carbohydrates. The ~10% intensity of the 60-64 ppm foot in gently extracted lignins, see **Fig. 1a and 1d**, is an upper limit to the OCH_3 fraction in this spectral range in fresh biomass. We have documented values of 5% and 8% in hard- and softwood, respectively, through 2D NMR of ^{13}C-enriched woods, see **Figure SI-4**. In extracted lignins, the shoulder is stronger, with an unresolved center near 60 ppm, and behaves exactly like the main OCH_3 peak. There are no other major components such as mobile OCH_2 or nonaromatic OCH_3 in these samples, so the shoulder must be attributed to OCH_3 of lignin as well. This is confirmed by ^{1}H-^{13}C HetCor NMR, see **Figure SI-5**. It ends up compensating for intensity loss of the main OCH_3 peak and raises the S/G of the extracted lignins closer to the values in their precursor biomass.

Since softwood contains an insignificant amount of syringol lignin, and the processing will not result in OCH_3 substitution of guaiacol rings, S/G = 0 for all softwood lignins. For softwood Kraft lignin, slightly negative values of S/G are obtained from this formula; this indicates loss of $(O)CH_3$, for instance as methanol, during lignin extraction. Due to the potential for OCH_3 loss, these S/G values are considered as lower limits in **Table 1**.

The aromatic C-O resonance of *p*-hydroxyphenyl (H) moieties is resolved fairly well near 160 ppm, for both the ether and the OH form. Denoting the integral from 157 to 163 ppm as I_{160}, the H fraction is simply

$$H/(S+G) = I_{160} \quad (10)$$

with our standard normalization of $I_{149} = 2.00$ in the dipolar-dephased spectrum.

With the spectrum of softwood confirmed as a pure-G component, experiment-based deconvolution of ^{13}C NMR spectra becomes possible. **Figure SI-7** documents for native maple wood and extracted CELF lignin how the difference spectra show the characteristics of pure syringol lignin, for instance with a resolved peak centered on 106 ppm. The fractions of the aromatic C-O integrals in this deconvolution are the G and S fractions.

Degree of condensation from spectrally edited solid-state NMR

The dipolar dephased spectra of biomass and lignins exhibit a major aromatic peak near 135 ppm. It contains the signal from two carbons in syringol and from one carbon in a simple guaiacol ring without C5 substitution. This is then the same as the number of OCH_3 carbons associated with these rings, so the excess of this signal (I_{135}) over the dephasing-corrected methoxyl signal ($1.4 \times I_{55}$) is due to condensation, or in other words, additional bonds of aromatic C with other carbons, beyond the standard bond at C1 of the lignol ring.

The contribution of C1 of *p*-hydroxy rings to the peak near 127 ppm must also be considered; it is quantified by I_{160}. Note that due to **Equation 11** this peak provides an upper limit to S/G. The fractional degree of condensation, i.e. the fraction of carbon-substituted aromatic C other than C1 out of all S, G, or H rings, is with our standard normalization of $I_{149} = 2.00$ in the dipolar-dephased spectrum. The 4% correction of I_{160} in this equation takes into account overlap from the large C-O peak at 149.5 ppm.

$$\boldsymbol{I_{135}} = (S+(S+G)+H+DC)/(S+G) \geq S/(S+G) + 1 \quad (11)$$

$$\mathbf{DC} = (I_{135} - 1.4\, I_{55} - (I_{160}\text{-}0.04)/(1+ (I_{160}\text{-}0.04)) \quad (12)$$

Table SI-1. Lignin elemental composition as determined by elemental analysis and peak areas from ^{13}C NMR spectra after dipolar dephasing, with the area between 157 and 142 ppm set to 2.00 because 2 carbons of S or G units contribute here. The *p*-hydroxyphenyl fraction, H/(S+G), is obtained by subtracting the background value of 0.04 from the integral between 163 and 157 ppm.

Sample	C (wt%)	H (wt%)	N (wt%)	O (wt%)[a]	163-157 ppm	H/(S+G)	142-99 ppm[b]	64-52 ppm[c]	OCH_3/(S+G)
C-Pine Lignin	67.77	6.68	0.20	25.35	0.09	0.05	1.95	0.72	1.01
C-Bagasse Lignin	73.60	4.85	1.14	20.41	0.2	0.15	1.98	0.89	1.25
C-Corn Stover Lignin	63.43	5.83	1.18	29.56	0.34	0.30	2.47	0.81	1.13
C-Poplar Lignin	68.32	6.47	0.33	24.88	0.09	0.05	1.91	0.99	1.39
C-Maple Lignin	68.00	5.57	0.41	26.02	0.05	0.01	2.01	1.00	1.40
K-Spruce Lignin	62.75	5.91	0.22	(31.12)	0.04	0.00	1.87	0.63	0.92
Pine wood	-	-	-	-	0.05	0.01	1.45	0.71	0.99
Loblolly Pine Lignin	-	-	-	-	0.06	0.02	1.43	0.72	1.01
Sugarcane Bagasse	-	-	-	-	0.24	0.20	1.63	0.90	1.26
Enzymatic Poplar Lignin	-	-	-	-	0.07	0.03	1.7	1.12	1.57
Maple wood	-	-	-	-	0.04	0.00	1.6	1.13	1.59

[a] Oxygen determined by difference
[b] Excludes residual carbohydrate signal
[c] 60-52 ppm for lignin in wood to exclude carbohydrate signals; ×1.08 for softwood; ×1.05 for hardwood

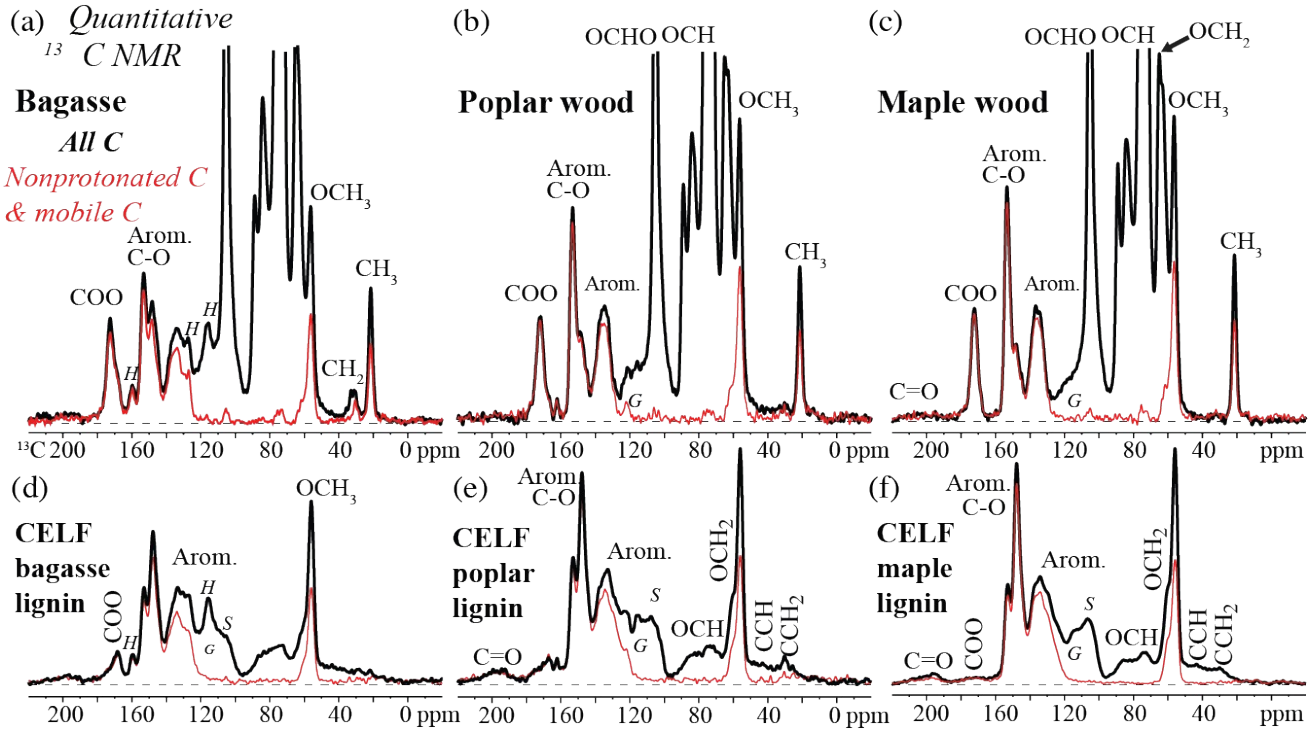

Figure SI-1. Quantitative solid-state ^{13}C NMR spectra of native lignin and carbohydrates in three types of wood or agricultural biomass (top) and the corresponding extracted CELF lignin (bottom).

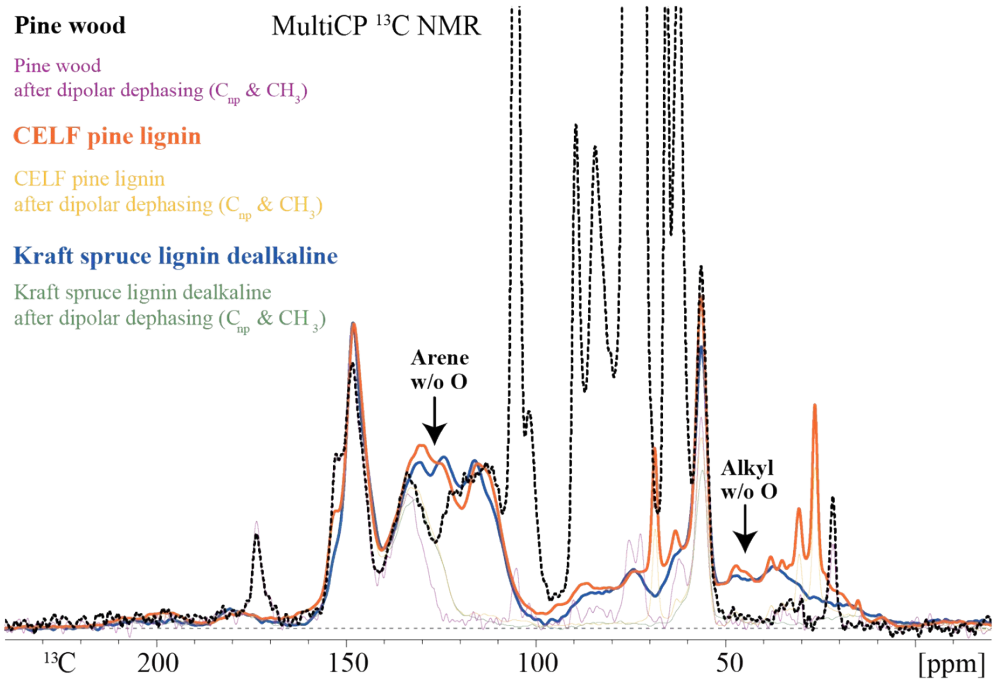

Figure SI-2. Quantitative ^{13}C solid-state NMR of CELF pine lignin and Kraft spruce lignin compared to lignin in unreacted pine wood.

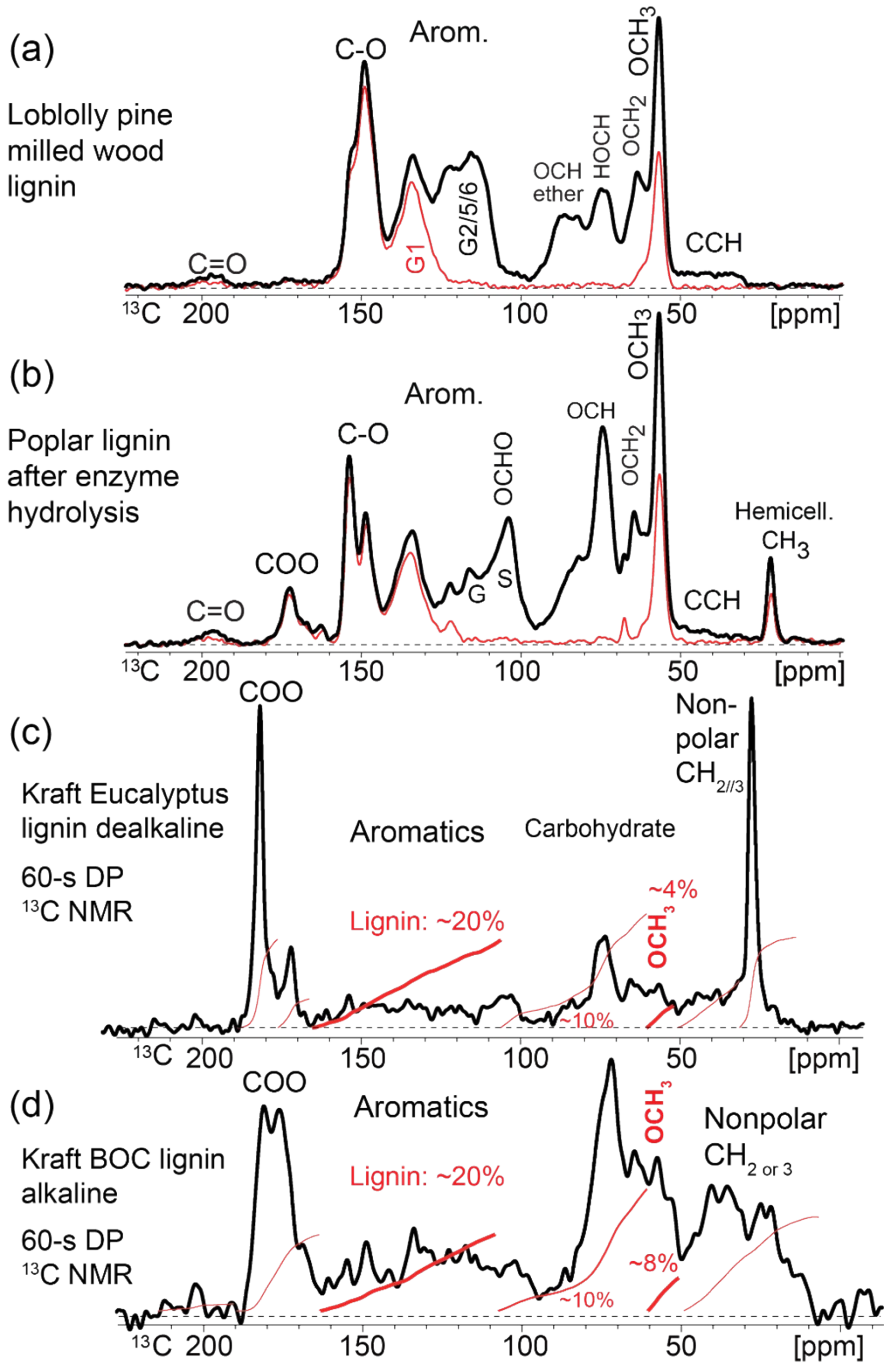

Figure SI-3. Nearly quantitative solid-state ^{13}C NMR analysis of **(a, b)** high-quality lignins and **(c, d)** low-quality hardwood Kraft "lignins". MultiCP ^{13}C NMR of **(a)** loblolly pine milled wood lignin and **(b)** poplar lignin from enzymatic hydrolysis. The spectra in **(a)** and **(b)** show the expected 6:3 aromatic:O-alkyl intensity ratio. Fully relaxed direct polarization ^{13}C NMR of **(c)** "Eucalyptus Kraft lignin", dominated by signals of an O=C-CH_n structural fragment and residual carbohydrate, and **(d)** "BOC Kraft lignin alkaline", showing major COO, nonpolar alkyl, and carbohydrate bands. DP

spectra were measured to ensure full detection of this component, while multiCP ^{13}C NMR spectra showed similar lineshapes. In both **(c)** and **(d)**, aromatics contribute <20% of all signals while characteristic lignin bands (compare **(a)** and **(b)**) account for <25% of all C.

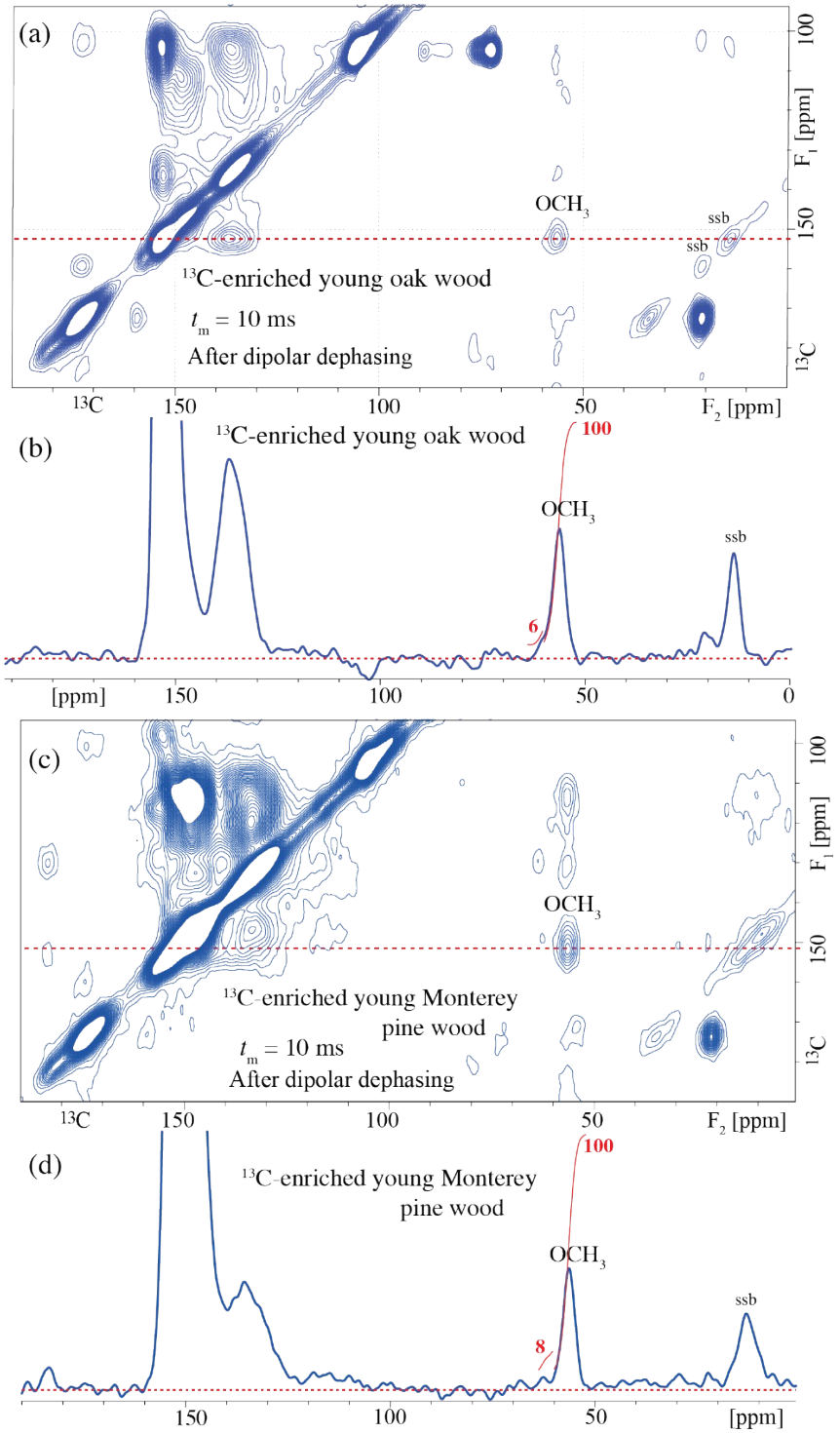

Figure SI-4. Assessment of the lignin OCH_3 ^{13}C NMR line shape in **(a, b)** ^{13}C-enriched young oak hardwood (Isolife, NL) and **(c, d)** ^{13}C-enriched young Monterey pine softwood (Isolife, NL),

by solid-state ^{13}C-^{13}C exchange NMR after dipolar dephasing. **(a/c)** Two-dimensional exchange spectrum with 10 ms mixing time. **(b, d)** Horizontal cross section through the spectrum in a) and c), respectively, along the red dashed line, at the aromatic C-O peak maximum. The peak integrals between 52 and 60 ppm and 60 and 64 ppm are indicated in red.

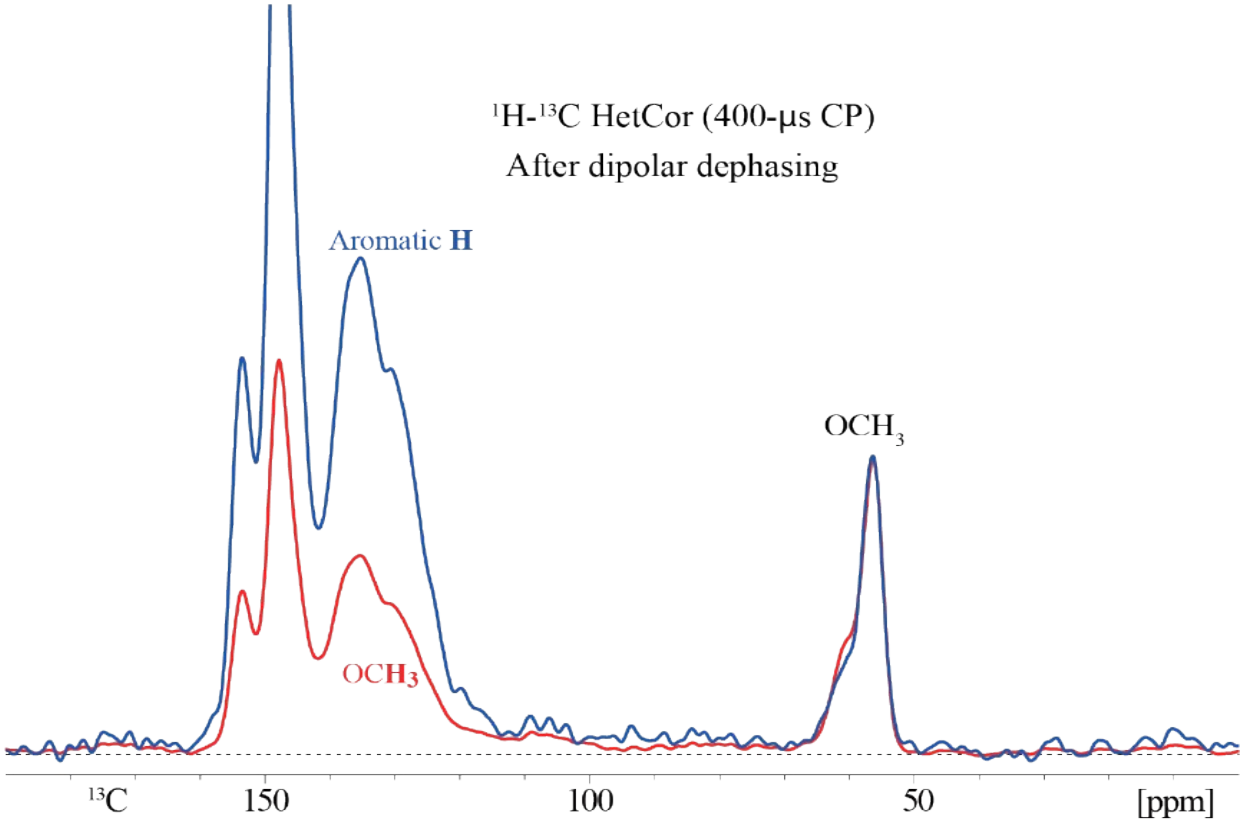

Figure SI-5. Assessment of the lignin OCH_3 ^{13}C NMR line shape (rightmost peak) in maple wood CELF lignin, by solid-state two-dimensional ^{1}H-^{13}C HetCor NMR after dipolar dephasing. Partial projections across the aromatic (blue) and OCH_3 (red) ^{1}H ranges are compared in terms of their OCH_3 line shape. The difference is insignificant, confirming that the shoulder at >60 ppm is also from lignin OCH_3 groups.

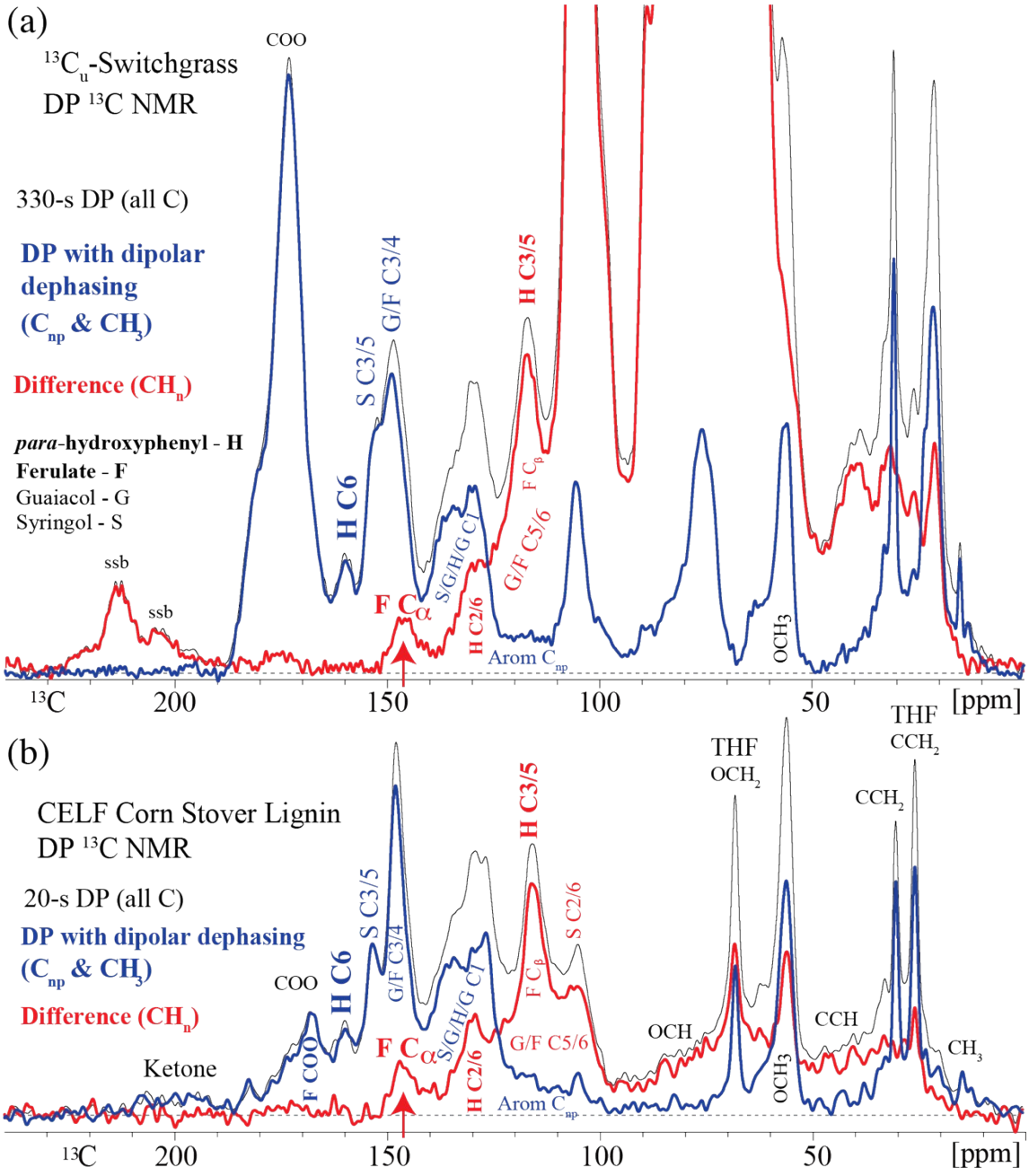

Figure SI-6. Quantitative solid-state ^{13}C NMR of grass lignins with spectral editing and peak assignments (S: syringol; G: guaiacol; F: ferulate; H: *p*-hydroxyphenyl; Cn: carbon n in a lignol ring). **(a)** ^{13}C-enriched switchgrass from Isolife (The Netherlands). **(b)** CELF lignin from corn stover. The spectrum after dipolar dephasing (blue trace) was scaled up by 1/0.88 and subtracted from the full spectrum (thin black trace) to yield the spectrum of C bonded to H (red trace). Alkene =CH signals of ferulate overlapping with aromatic C-O are highlighted by arrows. The spectral editing also highlights signals of H.

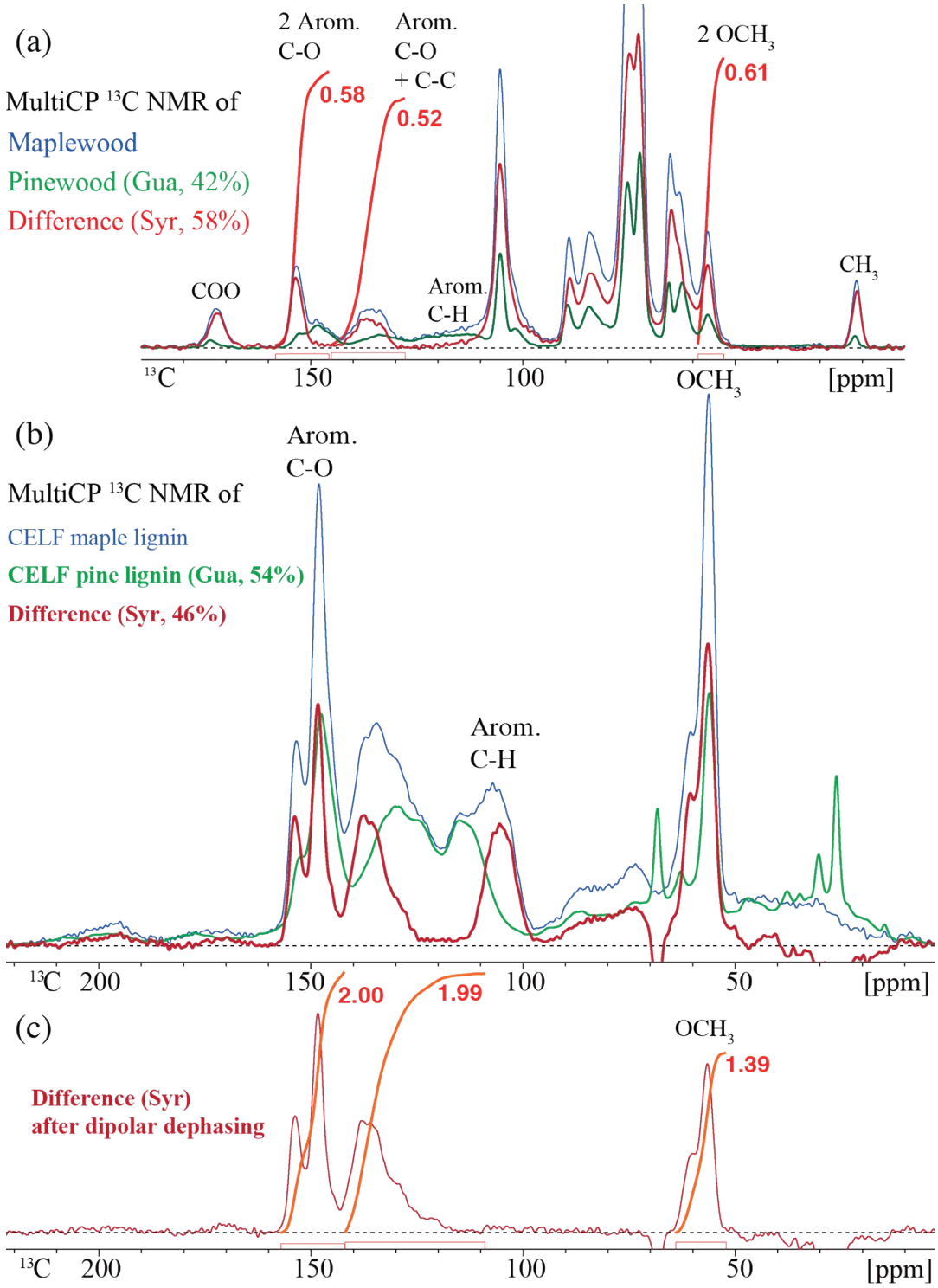

Figure SI-7. Determination of S/G by experiment-based deconvolution of the ^{13}C NMR spectrum of maple wood and its CELF lignin, using the corresponding spectrum of pine as a pure-G component. **(a)** Thin blue trace: spectrum of maple wood; green trace: spectrum of pine wood, scaled to match the maplewood intensity near 125 ppm; red trace: difference spectrum, with the simple characteristics of the symmetric syringol unit. Integrals relative to the total

aromatic C-O area (157 – 142 ppm) in the maple wood spectrum are also shown. The S fraction is 58%, the G fraction 42%, corresponding to an S/G ratio of 1.4. **(b)** Corresponding deconvolution of the ^{13}C NMR spectrum of maple wood CELF lignin. Thin blue trace: spectrum of maple wood CELF lignin; green trace: spectrum of pine wood CELF lignin, scaled to match the maplewood intensity near 120 ppm; red trace: difference spectrum, with the simple characteristics of the symmetric syringol unit; the pair of C-O peaks at 153 and 147 ppm is due to phenol ether linkages and C-OH end groups, respectively. The S fraction is 46%, the G fraction 54%, corresponding to an S/G ratio of 0.85. **(c)** Difference spectrum after dipolar dephasing, with integrals; the total aromatic C-O area (157 – 142 ppm), from two C-OCH_3 sites, has been set to 2.00.

Table SI-2. Elemental composition of biocrude and char phases from hydrothermal liquefaction of CELF lignin samples. All HTL experiments were conducted at 300 °C for 60 minutes. Oxygen is determined by difference.

		Biocrude				Char			
	S/G	**C (wt%)**	**H (wt%)**	**N (wt%)**	**O (wt%)**	**C (wt%)**	**H (wt%)**	**N (wt%)**	**O (wt%)**
C-Pine	0.10	69.55	7.33	0.10	23.02	74.82	5.55	0.13	19.5
C-Bagasse	0.52	74.34	8.29	0.47	16.90	67.38	4.64	1.03	26.95
C-Corn Stover	0.53	67.13	7.14	0.63	25.10	73.42	5.24	2.06	19.28
C-Poplar	0.79	70.65	6.95	0.41	21.99	70.72	4.46	0.84	23.98
C-Maple	0.68	65.76	7.15	0.26	26.83	69.6	4.55	0.66	25.19
K-Spruce	-	66.12	7.17	0.40	26.31	69.95	5.01	0.75	24.29

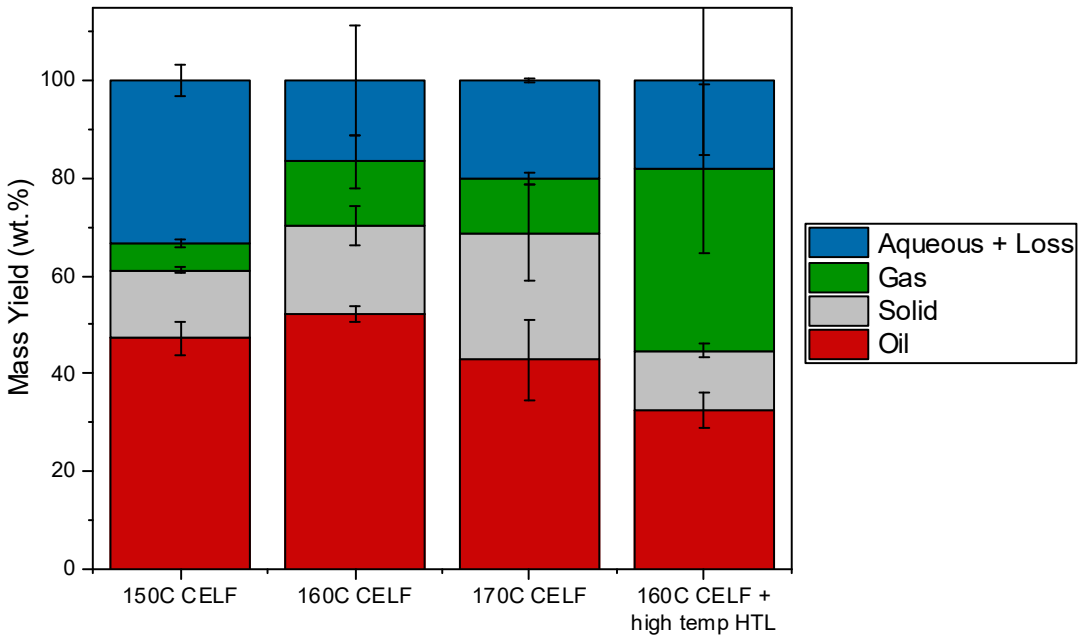

Figure SI-8. Hydrothermal liquefaction product distribution for the reaction of poplar lignin extracted via the CELF process at varying conditions. 150C was extracted at 150 °C, 25 min, 0.5% acid, and 2:1 THF:water ratio. 160C was extracted at 160 °C, 15 min, 0.5% acid, and 1:1 THF:water ratio. 170C was extracted at 170 °C, 15 min, 0.5% acid, and 2:1 THF:water ratio. High temperature HTL was performed at 350 °C for 60 minutes. CELF poplar in the remainder of this study is equivalent to 160C CELF.

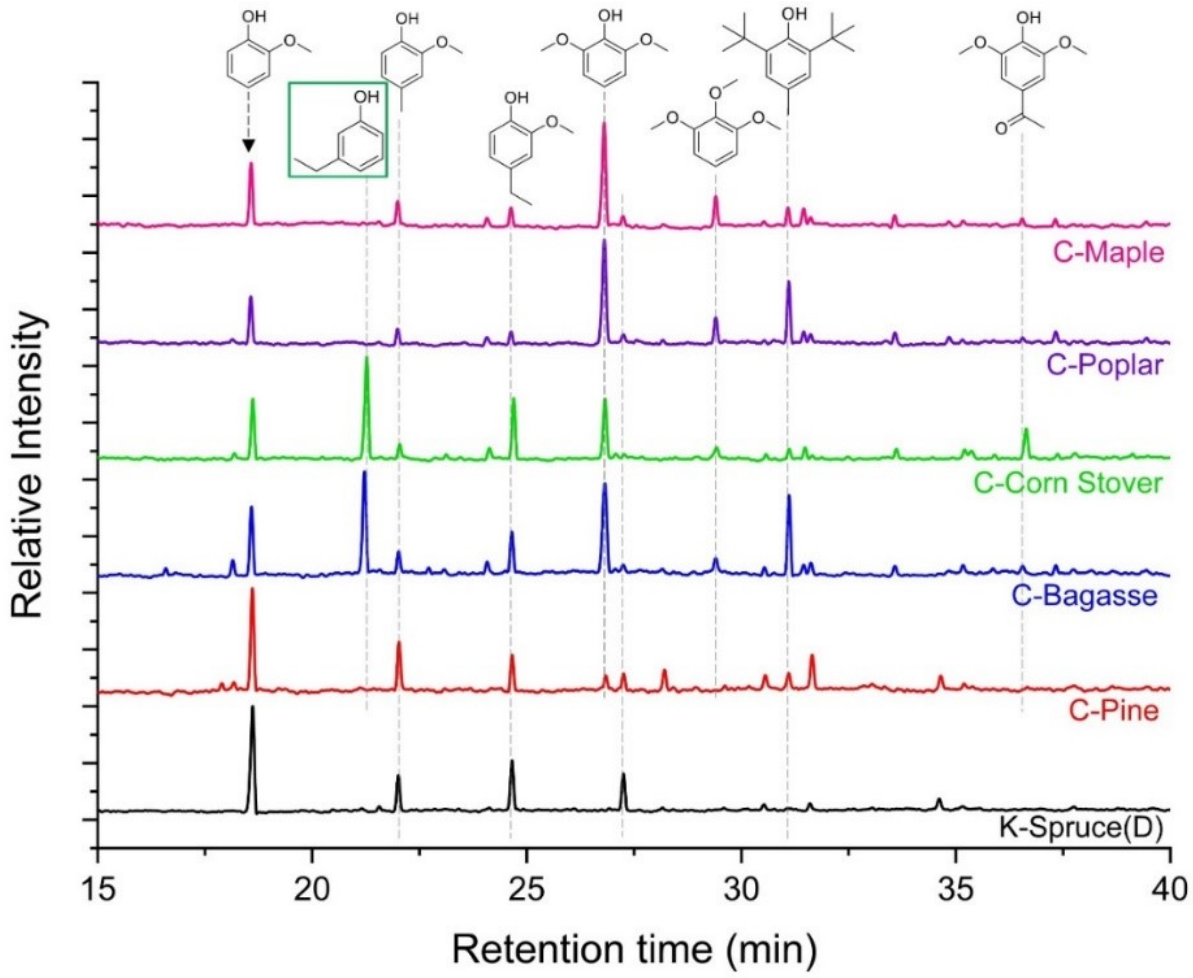

Figure SI-9. Raw GC-MS chromatograms for biocrudes from all CELF lignins examined, plus the comparison Kraft lignin. All experiments were completed at 300 °C for 1 h. Butylated hydroxytoluene, green square, is due to residual THF from CELF pretreatment.

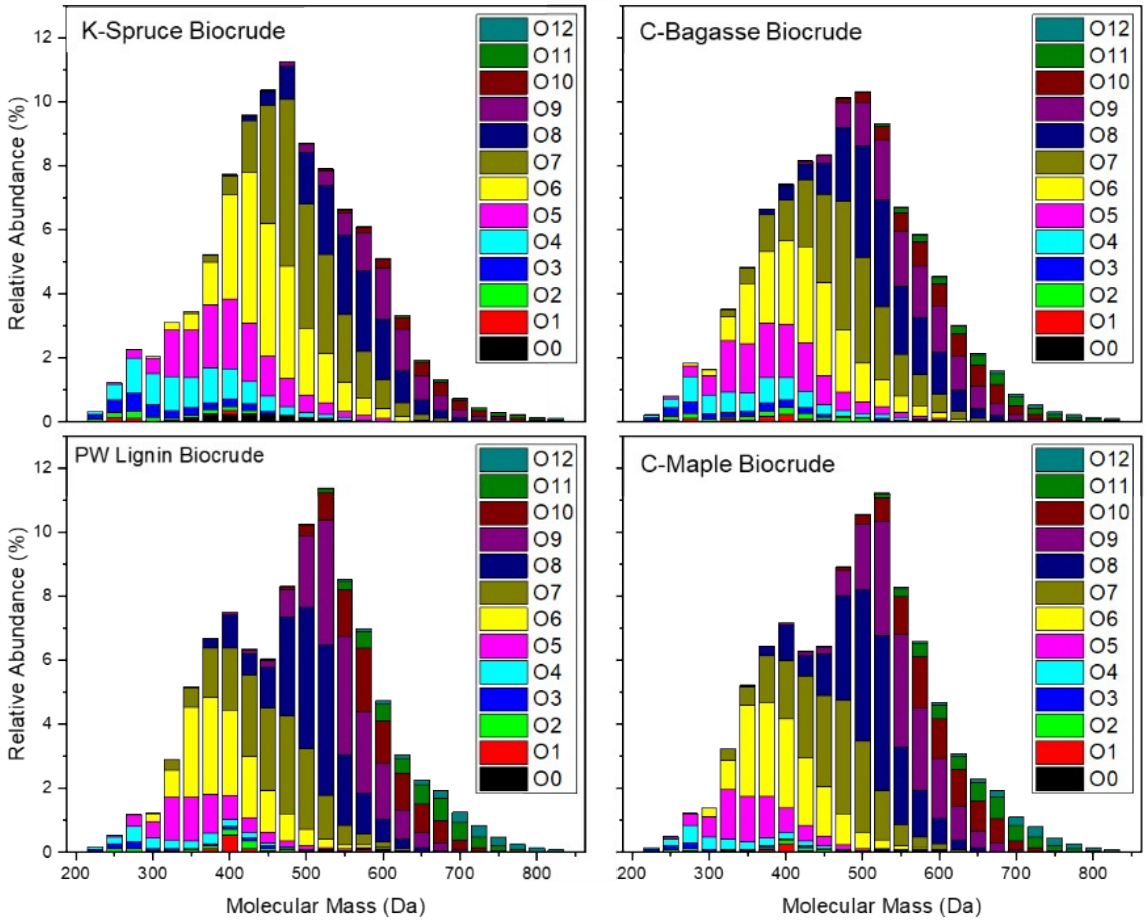

Figure SI-10. Oxygen heteroatom class abundance as a function of molecular weight derived from the analysis of biocrude samples with 21 T (+) APPI FT-ICR MS for **(a)** Kraft spruce biocrude **(b)** C-Bagasse biocrude, **(c)** C-Poplar biocrude, and **(d)** C-Maple biocrude.

References

1. Mao, J. D.; Schmidt-Rohr, K., Accurate Quantification of Aromaticity and Nonprotonated Aromatic Carbon Fraction in Natural Organic Matter by ^{13}C Solid-State Nuclear Magnetic Resonance. *Environ. Science Technol.* **2004,** *38*, 2680-2684.
2. Opella, S. J.; Frey, M. H., Selection of nonprotonated carbon resonances in solid-state nuclear magnetic resonance. *J. Am. Chem. Soc.* **1979,** *101*, 5854-5856.