

Supporting Information

Linkage Abundance and Molecular Weight Characteristics of Technical Lignins by Attenuated Total Reflection-FTIR Spectroscopy Combined with Multivariate Analysis

Christopher S. Lancefield,^[a] Sandra Constant,^[a] Peter de Peinder,^[b] and Pieter C. A. Bruijninx^{*[a, c]}

cssc_201802809_sm_miscellaneous_information.pdf

Table of Contents

General methods and information	S2
Table S1	S4
Table S2	S6
Table S3	S7
Table S4	S8
Table S5	S9
Table S6	S9
Table S7	S10
Figure S1	S11
Figure S2	S12
Figure S3	S13
References	S14

General Methods and Information

Commercial Lignins

Indulin AT kraft lignin (softwood, Meadwestvaco, US) and soda lignin Protobind™ 1000 (mixed wheat straw/Sarkanda grass, GreenValue S.A., Switzerland) were obtained from commercial sources. BioChoice and RAIZ Eucalyptus kraft lignins were kind donations from UPM Biochemicals (Finland) and the Forest and Paper Research Institute (RAIZ, Portugal), respectively. Alcell™ organosolv lignin (mixed hardwoods; maple, birch and poplar) was obtained from Repap Technology, Canada. One organosolv lignin was obtained from Sigma-Aldrich (product now discontinued) via Royal Dutch Shell (The Netherlands).

Lignin Preparation Methods

A: Purification of Lignin

Kraft lignin was dissolved in acetone/water (8:2, 20 mL/g) and then added dropwise to rapidly stirring water containing 0.02 M HCl (40 mL/g). The precipitated lignin was collected by filtration, washed thoroughly with water and allowed to air dry to a constant weight. The dried filter cake was manually ground in a pestle and mortar to give a fine powder, which was then rapidly stirred with Et₂O (20 mL/g) for 1 h to remove extractives. The lignin was finally filtered and allowed to air dry to a constant weight.

B: Solvent Fractionation of Lignin (1)

Kraft lignin was extracted sequentially with mixtures of EtOAc/Et₂O and/or EtOAc/MeOH and/or Acetone/MeOH (20 mL/g) of increasing polarity, as specified in Table S1, by stirring for 1 h. The soluble fractions were collected by centrifugation or filtration and concentrated *in vacuo*. The insoluble material from each step was returned for the next extraction. The resulting powders were then wetted with water to form a paste and allowed to air dry at 50 °C overnight. This was repeated 3 times to eliminate traces of residual organic solvent from the lignin.

C: Lignin Fractionation (2)

Fractionation of technical lignins were performed according to previously reported methods.¹

D: Ethanol Based Organosolv Lignin Extraction

This acid-catalysed, ethanol-based organosolv process was developed by the Energy Research Centre of the Netherlands (ECN, now ECN part of TNO). The organosolv procedures and lignin isolation methods have been reported in detail elsewhere.²⁻⁴ In short, the following parameters were used: 190 °C, 60 min, H₂SO₄ acidified, 60 wt% aqueous ethanol in a 20 L batch autoclave reactor (Büchi Glas Uster AG, Switzerland). The specific biomass related process conditions were: Wheat Straw - <10 mm, 11 L/kg DM, 20 mM H₂SO₄; Spruce - <6 mm, 5 L/kg DM, 10 mM H₂SO₄.

E: Dioxasolv Lignin

The dioxasolv procedure and lignin isolation methods have been reported in detail elsewhere.^{5,6} Softwood and Brewer's Spent Grain lignins (Entry 46 and 48, Table S1 respectively) were extracted using the following parameters: 1 h reflux, 1,4-dioxane/water (9:1), 0.2 M HCl, 10 L/kg DM; Softwood

lignin (Entry 47, Table S1) was extracted using the following parameters: 1 h reflux, 1,4-dioxane/water (8:2), 0.1 M HCl, 10 L/kg DM.

F: 140 °C Ethanosolv Lignin

This organosolv procedure was the same as that reported for method D but with the following parameters: 140 °C, 120 min, 60% w/w aqueous ethanol, 60 mM sulfuric acid and 10 L/kg DM.

G: 140 °C Acetone Lignin

This organosolv procedure and lignin isolation methods have been reported in detail elsewhere.⁷ The extraction used the following parameters: 140 °C, 60 min, 50% w/w aqueous acetone, 60 mM sulfuric acid, 10 L/kg DM.

NMR Methods

High-field NMR spectra were acquired on a Bruker Avance II 600 MHz spectrometer equipped with a 5 mm CPTCI ¹H-¹³C/¹⁵N-²H cryogenic probe with z-gradients at 25 °C. ¹H-¹³C Heteronuclear single quantum coherence spectroscopy (HSQC) spectra were recorded using either:

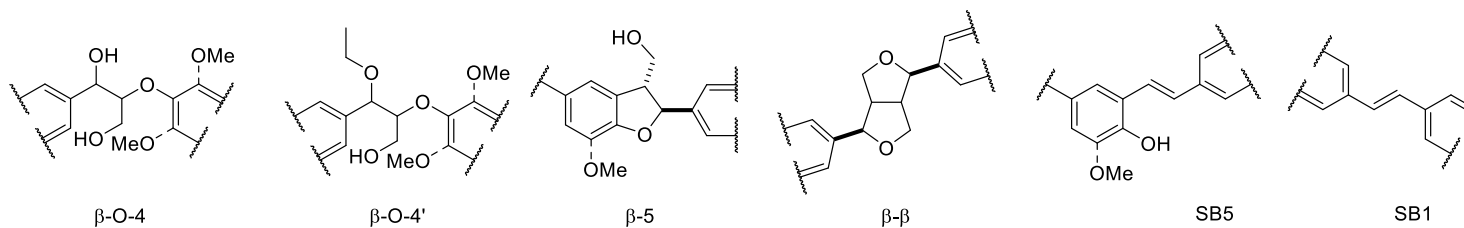
A: Bruker pulse program 'hsqcetgppsp.3' using the following parameters: acquired from 13 to -1 ppm in F2 (¹H) with 2048 data points, 160 to 0 ppm in F1 (¹³C) with 256 increments with a 1 s interscan delay (D1); cnst2 was set to 145 Hz.

B: Q-CAHSQC pulse program⁸ using the following parameters: acquired from 13 to -1 ppm in F2 (¹H) with 2048 data points, 160 to 0 ppm in F1 (¹³C) with 128 increments with a 6 s interscan delay (D1); cnst2 was set to 145 Hz.

In both cases the data was processed used Gaussian apodization (GB = 0.1, LB = 0.3 Hz) in F2 and squared cosine-bell and one level of linear prediction (32 coefficients) in F1.

NOTE: The detailed structural characterization of some of the included lignins have been reported by us previously. Where this is the case the references are given in Table S1. For this study the original HSQC NMR spectra were reprocessed for internal consistency and all GPC measurements were re-run to ensure an internally consistent data set.

Table S1 GPC, 2D HSQC analysis and isolation methods of the calibration lignins. Entries marked with * were also used for the second, kraft-only analysis. Values >1 reported to 2 significant figures, <1 reported to 1 significant figure.



Entry	Lignin	M _n	M _w	Log(M _n)	Log(M _w)	β -O-4	β -O-4' (α -OEt)	β -O-4+ β -O-4'	β -5	β - β	SB5	SB1	Prep. Method	NMR Method	Ref.
1*	Purified Indulin AT EtOAc (BNU) ^[a]	752	1093	2.9	3.0	1.1	0.0	1.1	0.5	1.1	9.3	3.5	A+B	A	
2*	Purified Indulin AT 5% EtOAc/MeOH (BNU)	1086	1691	3.0	3.2	3.1	0.0	3.1	1.3	1.6	10	2.3	A+B	A	
3*	Purified Indulin AT 10% EtOAc/MeOH (BNU)	1570	2579	3.2	3.4	5.2	0.0	5.2	2.5	2.0	9.3	1.3	A+B	A	
4*	Purified Indulin AT 20% EtOAc/MeOH (BNU)	2069	4098	3.3	3.6	8.3	0.0	8.3	2.9	2.3	5.4	0.9	A+B	A	
5*	Purified Indulin AT 30% EtOAc/MeOH (BNU)	2423	5999	3.4	3.8	10.1	0.0	10	3.8	2.5	5.9	0.6	A+B	A	
6*	Purified Indulin AT 40% EtOAc/MeOH (BNU)	3367	10577	3.5	4.0	11.3	0.0	11	4.0	2.8	4.5	0.2	A+B	A	
7*	Purified Indulin AT 100 %MeOH (BNU)	2508	9124	3.4	4.0	10.9	0.0	11	3.9	2.1	4.4	0.3	A+B	A	
8*	Purified Indulin AT 1:1 Acetone/MeOH (BNU)	3192	16625	3.5	4.2	12.7	0.0	13	4.0	2.9	3.8	0.3	A+B	A	
9	Crude Indulin AT (BNU)	1085	4669	3.0	3.7	8.2	0.0	10	2.0	1.6	9.0	2.2	Commercial	A	
10	Purified Indulin AT (BNU)	1238	5703	3.1	3.8	6.5	0.0	8.6	2.1	1.5	7.1	1.9	A	A	
11*	Purified Indulin AT SA31 EtOAc	829	1222	2.9	3.1	1.1	0.0	1.1	0.6	1.4	10.1	3.4	A+B	A	
12*	Purified Indulin AT SA31 5% EtOAc/MeOH	1169	2016	3.1	3.3	3.4	0.0	3.4	1.4	1.7	8.2	1.2	A+B	A	
13*	Purified Indulin AT SA31 10% EtOAc/MeOH	1743	2972	3.2	3.5	6.1	0.0	6.1	2.2	2.0	7.1	0.7	A+B	A	
14*	Purified Indulin AT SA31 20% EtOAc/MeOH	2345	4951	3.4	3.7	8.8	0.0	8.8	3.0	2.2	6.5	0.5	A+B	A	
15*	Purified Indulin AT SA31 30% EtOAc/MeOH	2831	7429	3.5	3.9	11.2	0.0	11	3.6	2.6	5.4	0.4	A+B	A	
16*	Purified Indulin AT SA31 40% EtOAc/MeOH	4049	14771	3.6	4.2	12.0	0.0	12	4.3	2.8	3.4	0.0	A+B	A	
17*	Purified Indulin AT SA31 1:1 Acetone/MeOH	5207	30649	3.7	4.5	13.0	0.0	13	4.2	2.7	2.8	0.0	A+B	A	
18	Indulin AT SA31 Purified	1264	6149	3.1	3.8	6.7	2.1	1.2	3.8	1.9	7.7	3.3	A	A	
19*	Crude Indulin AT SA31 EtOAc	637	934	2.8	3.0	1.5	0.0	1.5	0.5	1.2	8.2	7.9	B	A	9
20*	Crude Indulin AT SA31 5% EtOAc/MeOH	657	665	2.8	2.8	2.3	0.0	2.3	0.8	1.3	6.9	7.2	B	A	9

21*	Crude SA31 10% EtOAc/MeOH	799	1226	2.9	3.1	3.5	0.0	3.5	1.2	1.5	6.7	4.9	B	A	9
22*	Crude Indulin AT SA31 20% EtOAc/MeOH	1395	2135	3.1	3.3	6.9	0.0	6.9	2.0	1.8	5.2	2.0	B	A	9
23*	Crude SA31 30% EtOAc/MeOH	1939	3222	3.3	3.5	9.8	0.0	9.8	2.6	2.1	3.8	0.8	B	A	9
24*	Crude Indulin AT SA31 100 %MeOH	1960	4337	3.3	3.6	12	0.0	12	3.3	2.3	3.9	0.5	B	A	9
25*	Purified BioChoice EtOAc	981	1403	3.0	3.1	1.0	0.0	1.0	0.5	1.1	11	2.1	A+B	A	
26*	Purified BioChoice Kraft 5% EtOAc/MeOH	1455	2436	3.2	3.4	2.6	0.0	2.6	1.4	1.4	7.5	1.0	A+B	A	
27*	Purified BioChoice Kraft 10% EtOAc/MeOH	2173	3826	3.3	3.6	4.6	0.0	4.6	1.9	1.8	6.4	0.6	A+B	A	
28*	Purified BioChoice Kraft 20% EtOAc/MeOH	2914	5796	3.5	3.8	5.6	0.0	5.6	2.2	1.7	5.6	0.4	A+B	A	
29*	Purified BioChoice Kraft 30% EtOAc/MeOH	3658	10193	3.6	4.0	7.1	0.0	7.1	2.6	2.0	4.8	0.3	A+B	A	
30*	Purified BioChoice Kraft 40% EtOAc/MeOH	4066	12425	3.6	4.1	7.4	0.0	7.4	2.9	2.2	4.8	0.3	A+B	A	
31*	Purified BioChoice Kraft 1:1 Acetone/MeOH	4140	21959	3.6	4.3	9.3	0.0	9.3	3.2	2.5	4.0	0.3	A+B	A	
32	BioChoice Kraft	1384	5481	3.1	3.7	5.4	0.0	7.3	1.9	1.5	6.1	2.1	Commercial	A	
33	Purified BioChoice Kraft	1435	6216	3.2	3.8	4.1	0.0	1.2	4.1	1.3	1.4	8.2	A	A	
34	P1000	1005	3733	3.0	3.6	5.2	0.0	5.2	1.0	1.1	0.0	0.0	Commercial	B	4
35	P1000 f1	685	1312	2.8	3.1	1.2	0.0	1.2	0.3	1.1	0.0	0.0	C	B	
36	P1000 f2	1360	2508	3.1	3.4	5.8	0.0	5.8	1.0	1.1	0.0	0.0	C	B	
37	P1000 f3	1773	6747	3.2	3.8	12	0.0	12	1.6	1.1	0.0	0.0	C	B	
38	P1000 f4	1957	10728	3.3	4.0	12	0.0	12	1.8	1.5	0.0	0.0	C	B	
39	ECN High Temp Ethanol/Water Wheat Straw f1	788	1396	2.9	3.1	3.9	3.4	7.3	5.5	0.6	0.0	1.3	D+C	B	
40	ECN High Temp Ethanol/Water Wheat Straw f2	992	2194	3.0	3.3	6.3	4.1	10	6.6	0.8	0.0	0.5	D+C	B	
41	ECN High Temp Ethanol/Water Wheat Straw f3	1343	3253	3.1	3.5	11	4.5	15	5.6	0.7	0.0	0.0	D+C	B	
42	RAIZ Eucalyptus Kraft 50% EtOAc/Et ₂ O	581	891	2.8	2.9	0.6	0.0	0.6	0.0	4.3	1.7	2.7	B	A	
43	RAIZ Eucalyptus Kraft EtOAc	920	1543	3.0	3.2	1.7	0.0	1.7	0.0	2.7	1.3	1.2	B	A	
44	RAIZ Eucalyptus Kraft 10% EtOAc/MeOH	1185	2386	3.1	3.4	5.5	0.0	5.5	0.3	3.6	1.7	1.4	B	A	
45	RAIZ Eucalyptus Kraft MeOH	1486	5503	3.2	3.7	9.8	0.0	9.8	0.5	4.0	1.4	1.0	B	A	
46	Dioxasolv Softwood	1741	5257	3.2	3.7	26	0.0	26	5.5	3.9	0.0	0.0	E	B	10
47	Dioxasolv Softwood	2301	4462	3.4	3.6	31	0.0	31	10.0	2.0	0.0	0.0	E	A	11
48	Dioxasolv Brewer's Spent Grains	1937	5680	3.3	3.8	34	0.0	34	7.6	2.7	0.0	0.0	E	B	10
49	Alcell	985	3991	3.0	3.6	7.2	2.3	9.5	1.7	2.8	0.0	0.0	Commercial	B	4
50	Sigma-Aldrich Organosolv	1158	6999	3.1	3.8	9.6	0.0	9.6	0.8	0.4	0.0	0.0	Commercial	B	
51	ECN High Temp Ethanol/Water Spruce	985	2225	3.0	3.3	1.9	2.3	4.1	5.5	1.2	0.0	0.0	D	B	4

52	ECN Low Temp Ethanol/Water Wheat Straw	1149	3350	3.1	3.5	9.2	12.0	21.3	5.4	1.2	0.0	0.0	F	B
53	ECN Low Temp Ethanol/Water Pine	1239	4152	3.1	3.6	6.9	15.2	22.1	7.4	1.4	0.0	0.0	F	B
54	ECN Low Temp Acetone/Water Wheat Straw	1340	5232	3.1	3.7	13.7	0.0	13.7	4.5	1.4	0.0	0.0	G	B

7

^[a] Batch Number Unknown (BNU) - a sample of Indulin AT kraft lignin for which we do not have a batch number, however, it is very similar to a lignin with batch number RF23 (data not shown).

Table S2 ³¹P NMR analysis of a selection of the calibration lignins in mmol/g (of dry lignin). Values >1 reported to 2 significant figures, <1 reported to 1 significant figure.

Entry	Lignin	Aliphatic OH	5-substituted ArOH	Guaiacyl ArOH	<i>p</i> -Hydroxy ArOH	Total ArOH	Total COOH	Ref.
1	P1000 f1	0.7	2.0	1.0	0.5	3.5	0.9	
2	P1000 f2	1.1	1.8	0.7	0.3	2.7	0.7	
3	P1000 f3	1.2	1.4	0.5	0.2	2.1	0.6	
4	P1000 f4	1.3	1.3	0.5	0.3	2.1	0.6	
5	High Temp Ethanol/Water Wheat Straw f1	1.0	1.1	1.0	0.4	2.6	0.2	
6	High Temp Ethanol/Water Wheat Straw f2	1.3	1.2	0.9	0.4	2.5	0.1	
7	High Temp Ethanol/Water Wheat Straw f3	1.4	1.1	0.7	0.3	2.1	0.2	
8	RAIZ Eucalyptus EtOAc	0.5	2.3	0.6	0.0	2.8	0.4	
9	RAIZ Eucalyptus 10% EtOAc/MeOH	0.6	1.3	0.3	0.0	1.6	0.2	
10	RAIZ Eucalyptus MeOH	0.7	0.7	0.3	0.0	1.0	0.3	
11	Crude Indulin AT SA31 EtOAc	1.0	1.5	2.2	0.2	3.8	0.7	⁹
12	Crude Indulin AT SA31 5% EtOAc/MeOH	1.3	1.7	2.4	0.2	4.4	0.6	⁹
13	Crude SA31 10% EtOAc/MeOH	1.4	1.7	2.2	0.2	4.1	0.5	⁹
14	Crude Indulin AT SA31 20% EtOAc/MeOH	1.5	1.6	1.7	0.2	3.5	0.4	⁹
15	Crude SA31 30% EtOAc/MeOH	1.9	1.5	1.6	0.2	3.3	0.4	⁹
16	Crude Indulin AT SA31 100 %MeOH	2.0	1.4	1.3	0.2	2.9	0.3	⁹
17	Alcell	1.0	1.7	0.6	0.1	2.4	0.2	⁴
18	Sigma-Aldrich Organosolv	1.4	0.9	0.7	0.6	2.3	1.4	
19	ECN Low Temp Ethanol/Water Wheat Straw	2.3	0.9	0.6	0.3	1.9	0.3	
20	ECN Low Temp Ethanol/Water Pine	2.6	0.7	1.0	0.1	1.8	0.1	
21	ECN Low Temp Acetone/Water Wheat Straw	1.8	0.8	0.6	0.3	1.7	0.5	

Table S3 Optimisation of PLS modelling pre-processing parameters

Entry	Property	Spectral Range (cm ⁻¹)	Pre-processing	Num. LVs	RMSEC	RMSECV	Bias	CV Bias	R ² Cal	R ² CV
1	M _n	4000-600		5	466	564	6.8E-13	12	0.80	0.71
2		4000-600	1st derivative	6	343	480	0	47	0.89	0.79
3		4000-600	2nd derivative	6	263	507	-6.8E-13	13	0.94	0.76
4		2000-750		5	471	571	0	19	0.79	0.70
5		2000-750	1st derivative	7	261	401	-4.5E-13	37	0.94	0.85
6	Log(M _n)	4000-600		9	0.046	0.075	4.4E-16	0.00044	0.96	0.90
7		4000-600	1st derivative	8	0.034	0.061	4.4E-16	0.0049	0.98	0.93
8		4000-600	2nd derivative	6	0.035	0.069	-8.9E-16	0.0051	0.98	0.91
9		2000-750		6	0.062	0.089	0	0.0086	0.93	0.85
10		2000-750	1st derivative	7	0.034	0.059	-1.8E-15	0.0081	0.98	0.94
11	M _w	4000-600		5	3529	4176	1.8E-12	208	0.57	0.42
12		4000-600	1st derivative	5	2851	4013	-2.7E-12	126	0.72	0.49
13		4000-600	2nd derivative	6	2096	4444	9.1E-13	-72	0.85	0.38
14		2000-750		4	3670	4128	9.1E-13	202	0.54	0.43
15		2000-750	1st derivative	10	1499	3015	-9.1E-13	243	0.92	0.70
16	Log(M _w)	4000-600		5	0.14	0.17	-8.9E-16	0.0088	0.86	0.78
17		4000-600	1st derivative	7	0.080	0.13	-4.4E-16	0.0087	0.95	0.88
18		4000-600	2nd derivative	6	0.077	0.17	1.3E-15	0.0012	0.96	0.79
19		2000-750		5	0.14	0.17	-8.9E-16	0.0088	0.86	0.78
20		2000-750	1st derivative	6	0.091	0.13	-4.4E-16	0.0029	0.94	0.88
21	β-O-4 ^a	4000-600		4	2.2	2.8	0	-0.19	0.90	0.84
22		4000-600	1st derivative	4	1.8	2.9	0	-0.030	0.94	0.84
23		4000-600	2nd derivative	6	1.2	3.5	1.8E-15	-0.048	0.97	0.76
24		2000-750		4	2.2	2.7	5.3E-15	-0.12	0.91	0.86
25		2000-750	1st derivative	4	1.8	2.8	1.8E-15	-0.019	0.94	0.85
26	β-5	4000-600		5	0.80	1.1	0	0.028	0.86	0.74
27		4000-600	1st derivative	5	0.67	1.1	0	0.050	0.90	0.74

28		4000-600	2nd derivative	5	0.61	1.2	0	-0.022	0.92	0.68
29		2000-750		5	0.73	1.1	-4.4E-16	0.071	0.88	0.73
30		2000-750	1st derivative	4	0.74	1.1	0	0.035	0.88	0.75
31		4000-600		4	0.46	0.55	0	-0.018	0.70	0.58
32		4000-600	1st derivative	6	0.30	0.56	2.2E-16	-0.059	0.87	0.59
33	β - β	4000-600	2nd derivative	6	0.26	0.61	2.2E-16	-0.0074	0.90	0.51
34		2000-750		4	0.45	0.54	6.7E-16	-0.019	0.72	0.59
35		2000-750	1st derivative	5	0.33	0.53	2.2E-16	-0.048	0.85	0.61

^a Total quantity of β -O-4 motifs defined as β -O-4 + β -O-4' (i.e. α -OH + α -OEt) - α -OEt results from exchange of the (benzylic) α -OH in the β -O-4 unit under acidic organosolv pulping conditions in the presence of EtOH.

Table S4. Results of partial least-squares (PLS) regression between hydroxyl group content as determined by ³¹P NMR GPC and the ATR-FTIR spectra. A spectral range of 750-2000 cm⁻¹ was used together with 1st derivative pre-processing.

Entry	Unit	Num. LVs	Range	RMSEC	RMSECV	Bias	CV Bias	R ² Cal	R ² CV
1	Aliphatic OH	4	2.1	0.12	0.22	2.22E-16	0.0038	0.96	0.84
2	5-Substituted ArOH (G _{cond.} + S)	4	1.5	0.19	0.41	2.22E-16	0.059	0.79	0.17
3	Guaiacyl OH	5	2.2	0.059	0.17	2.22E-16	0.024	0.99	0.93
4	Total Phenols (inc. H)	3	3.4	0.28	0.49	0	0.056	0.89	0.70
5	Total COOH	6	1.3	0.048	0.13	0	-0.0026	0.97	0.80

Table S5. GPC, 2D HSQC analysis and isolation methods of the set of lignins used for validation of the PLS model.

Validation Lignin	Mn	Mw	Log ₁₀ (M _n)	Log ₁₀ (M _w)	β-O-4	β-5	β-β	SB5	SB1	Prep. Method
Purified Indulin AT EtOAc (BNU)	730	1243	2.9	3.1	1.5	0.7	1.1	9.4	4.6	A+B
Purified Indulin AT 5% EtOAc/MeOH (BNU)	966	1817	3.0	3.3	3.3	1.2	1.5	9.7	2.0	A+B
Purified Indulin AT 10% EtOAc/MeOH (BNU)	1594	2940	3.2	3.5	5.5	2.2	1.9	7.9	1.1	A+B
Purified Indulin AT 20% EtOAc/MeOH (BNU)	1879	4185	3.3	3.6	8.5	3.0	2.1	4.5	0.9	A+B
Purified Indulin AT 30% EtOAc/MeOH (BNU)	2127	5683	3.3	3.8	9.6	3.5	2.3	5.1	0.4	A+B
Purified Indulin AT 40% EtOAc/MeOH (BNU)	2388	8527	3.4	3.9	12.8	4.2	2.8	4.1	0.5	A+B
Purified Indulin AT 1:1 Acetone/MeOH (BNU)	2275	14403	3.4	4.2	12.1	3.6	2.4	3.2	0.3	A+B

Table S6. Results measured by GPC and HSQC NMR and predicted by the PLS model for the validation set of kraft lignins. Predicted values are an average of two ATR-FTIR measurements. Meas. = measured, Pred. = predicted.

Property		EtOAc/MeOH						Acetone/MeOH
		0%	5%	10%	20%	30%	40%	50%
Log(M _n)	Meas.	2.85	3.03	3.16	3.31	3.39	3.43	3.38 ^[a]
	Pred.	2.84	3.05	3.20	3.34	3.44	3.52	3.57
Log(M _w)	Meas.	3.05	3.27	3.43	3.64	3.91	4.03	4.16 ^[a]
	Pred.	2.98	3.21	3.40	3.63	3.85	4.04	4.37

^[a] sample not fully soluble in THF after acetylation and values most likely underestimated with respect to the whole sample.

Table S7. Results of partial least-squares (PLS) regression between molecular weight and inter-unit abundances determined by GPC and 2D HSQC NMR for the 28 kraft lignin samples and their ATR-FTIR spectra using 1st derivative pre-processing, with and without baseline correction. Values reported to 2 significant figures.

Entry	Unit	Range	Num. LVs	With Baseline Correction		Without Baseline Correction	
				RMSEC	RMSECV	RMSEC	RMSECV
2	Log(M _n)	0.91	5	0.025	0.046	0.025	0.047
4	Log(M _w)	1.7	6	0.039	0.097	0.040	0.098
5	β-O-4	12	6	0.25	0.46	0.25	0.46
6	β-5	3.8	6	0.095	0.18	0.095	0.17
7	β-β	1.8	3	0.13	0.19	0.13	0.19
8	SB5	8.0	3	0.79	1.0	0.79	1.0
9	SB1	7.9	5	0.26	0.48	0.25	0.48

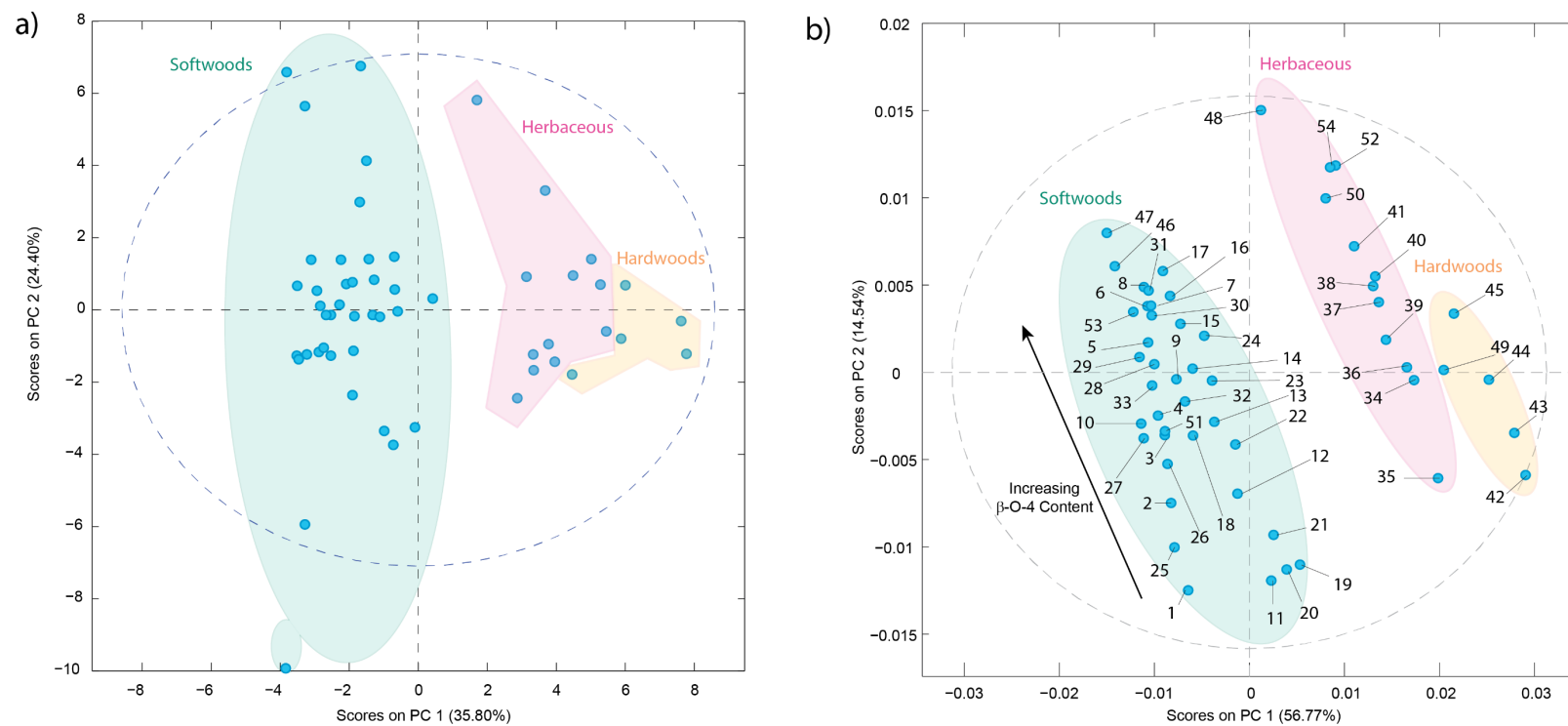


Figure S1. Comparison of PCA analysis a) without 1st derivative pre-processing and b) with 1st derivative pre-processing. The labels on b) correspond to the lignin entries in Table S1. The shaded areas are for illustrative purposes only.

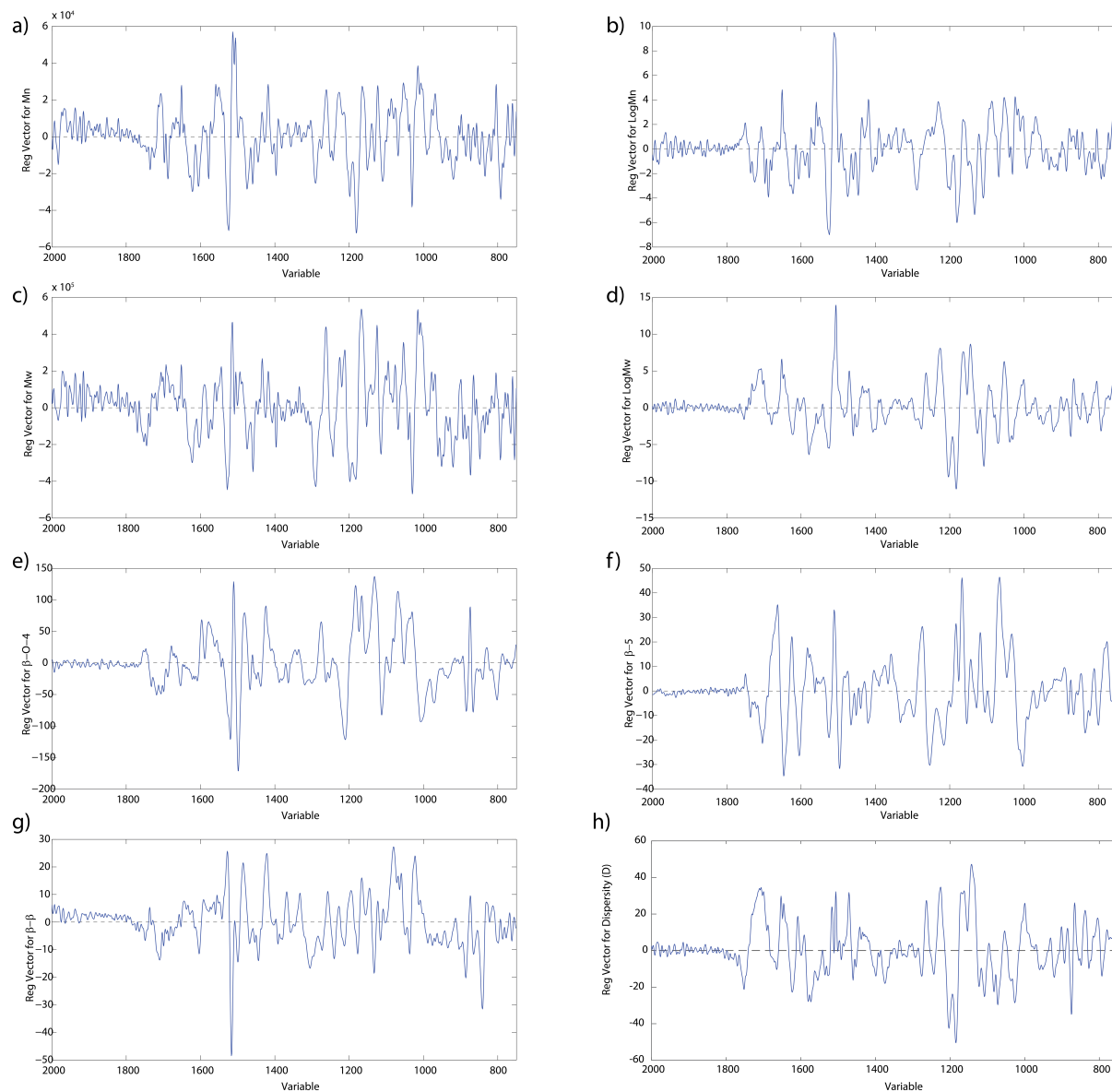


Figure S2. The regression vectors associated with PLS models reported in Table 1: a) M_n ; b) $\text{Log}(M_n)$; c) M_w ; d) $\text{Log}(M_w)$ e) β -O-4; f) β -5; g) β - β content and h) dispersity.

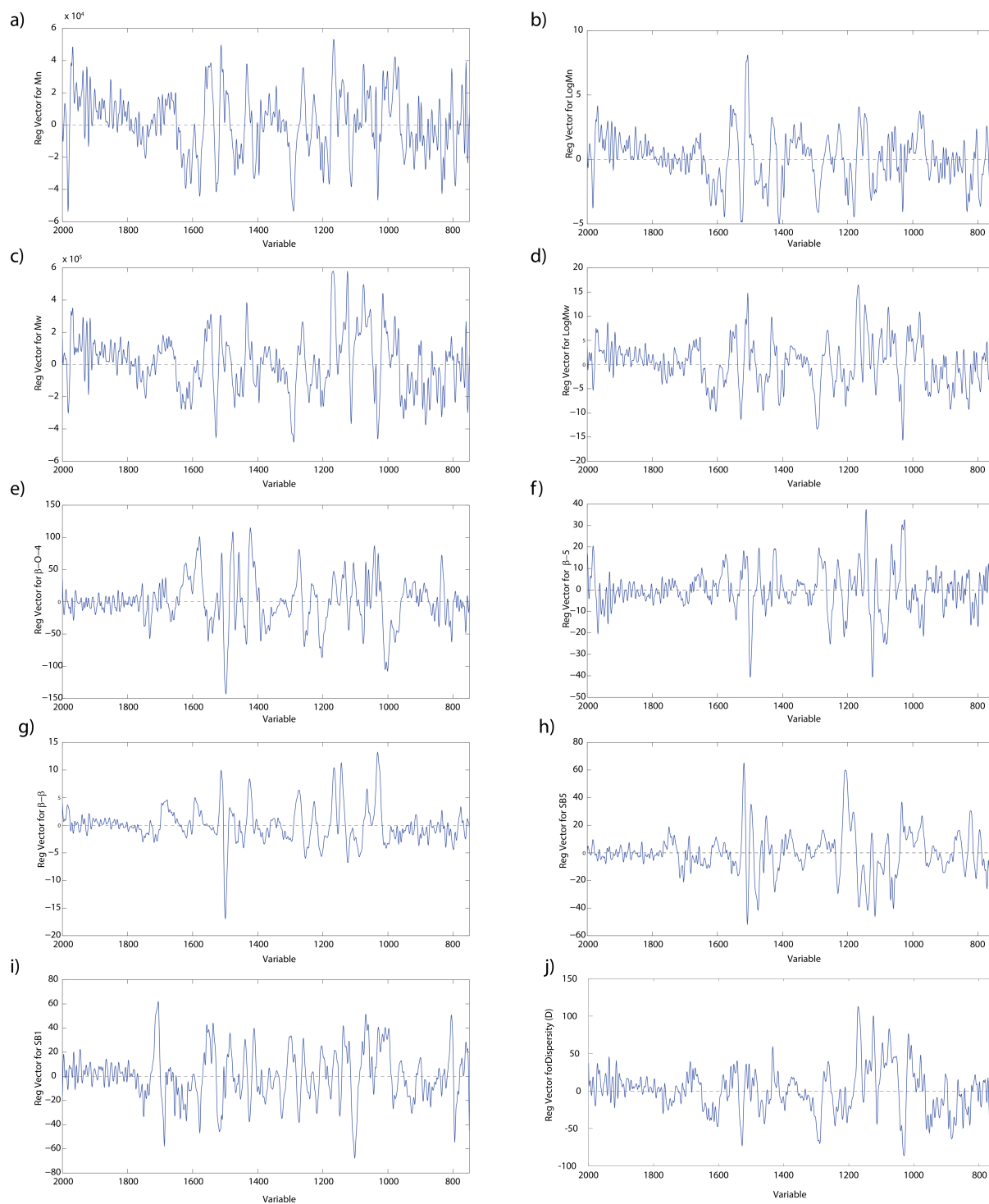


Figure S3. The regression vectors associated with PLS models reported in Table 2: a) M_n ; b) $\text{Log}(M_n)$; c) M_w ; d) $\text{Log}(M_w)$ e) $\beta\text{-O-4}$; f) $\beta\text{-5}$; g) $\beta\text{-}\beta$; h) SB5; i) SB1 content and j) dispersity.

References

- 1 R. Gosselink, J. van Der Putten, and D. van Es, (2015). Fractionation of technical lignin. WO/2015/178771.
- 2 J. Wildschut, A. T. Smit, J. H. Reith and W. J. J. Huijgen, Ethanol-based organosolv fractionation of wheat straw for the production of lignin and enzymatically digestible cellulose, *Bioresour. Technol.*, 2013, 135, 58–66.
- 3 P. de Wild, W. Huijgen and H. Heeres, Pyrolysis of wheat straw-derived organosolv lignin, *J. Anal. Appl. Pyrolysis*, 2012, 93, 95–103.
- 4 S. Constant, H. L. J. Wienk, A. E. Frissen, P. de Peinder, R. Boelens, D. S. van Es, R. J. H. Grisel, B. M. Weckhuysen, W. J. J. Huijgen, R. J. A. Gosselink and P. C. A. Bruijninx, New insights into the structure and composition of technical lignins: a comparative characterisation study, *Green Chem.*, 2016, 18, 2651–2665.
- 5 C. S. Lancefield, O. S. Ojo, F. Tran and N. J. Westwood, Isolation of Functionalized Phenolic Monomers through Selective Oxidation and C-O Bond Cleavage of the β -O-4 Linkages in Lignin, *Angew. Chemie Int. Ed.*, 2015, 54, 258–262.
- 6 D. M. Miles-Barrett, A. R. Neal, C. Hand, J. R. D. Montgomery, I. Panovic, O. S. Ojo, C. S. Lancefield, D. B. Cordes, A. M. Z. Slawin, T. Lebl and N. J. Westwood, The synthesis and analysis of lignin-bound Hibbert ketone structures in technical lignins, *Org. Biomol. Chem.*, 2016, 14, 10023–10030.
- 7 A. Smit and W. Huijgen, Effective fractionation of lignocellulose in herbaceous biomass and hardwood using a mild acetone organosolv process, *Green Chem.*, 2017, 19, 5505–5514.
- 8 H. Koskela, I. Kilpeläinen and S. Heikkinen, Some aspects of quantitative 2D NMR, *J. Magn. Reson.*, 2005, 174, 237–244.
- 9 C. S. Lancefield, H. L. J. Wienk, R. Boelens, B. M. Weckhuysen and P. C. A. Bruijninx, Identification of a diagnostic structural motif reveals a new reaction intermediate and condensation pathway in kraft lignin formation, *Chem. Sci.*, 2018, 9, 6348–6360.
- 10 R. Jastrzebski, S. Constant, C. S. Lancefield, N. J. Westwood, B. M. Weckhuysen and P. C. A. Bruijninx, Tandem catalytic depolymerization of lignin by water-tolerant Lewis acids and rhodium complexes, *ChemSusChem*, 2016, 9, 2074–2079.
- 11 C. S. Lancefield, L. W. Teunissen, B. M. Weckhuysen and P. C. A. Bruijninx, Iridium-catalysed primary alcohol oxidation and hydrogen shuttling for the depolymerisation of lignin, *Green Chem.*, 2018, 20, 3214–3221.