Identification of a diagnostic structural motif reveals a new reaction intermediate and condensation pathway in kraft lignin formation

Christopher S. Lancefield,[a] Hans L. J. Wienk,[b] Rolf Boelens,[b] Bert M. Weckhuysen[a] and Pieter C. A. Bruijnincx*[ac]

[a] Inorganic Chemistry and Catalysis, Debye Institute for Nanomaterials Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

[b] NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[c] Organic Chemistry and Catalysis, Debye Institute for Nanomaterials Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

**Corresponding author: p.c.a.bruijnincx@uu.nl*

Table of Contents

1. Experimental Details

1.1 Equipment and Methods

Routine 1D and 2D NMR spectra were recorded on an Agilent 400MR spectrometer (with OneNMR probe) or a Varian 400 MHz VNMRS system (with an Auto Switchable (ASW) probe, with tuning optimized for both ^{1}H and ^{19}F) using the standard NMR experiments, as present in the VNMRJ 4.2 software. The ^{1}H and ^{13}C NMR spectra were referenced internally using the residual solvent resonances and reported in ppm relative to TMS (0 ppm). Multiplicities are described using the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet and m = multiplet and the *J* couplings are reported in Hz. NMR spectra were processed using MestReNova 10.

High-field NMR spectra were acquired on a Bruker Avance II 600 MHz spectrometer equipped with a 5 mm CPTCI ^{1}H-^{13}C/^{15}N-^{2}H cryogenic probe with z-gradients at 25 °C. ^{1}H-^{13}C Heteronuclear single quantum coherence spectroscopy (HSQC) spectra were recorded using the Bruker pulse sequence 'hsqcetgpsp.3' using the following parameters: acquired from 13 to -1 ppm in F2 (^{1}H) with 2048 data points, 160 to 0 ppm in F1 (^{13}C) with 128 increments with a 1 s interscan delay (D1); cnst2 was set to 145 Hz. Processing 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. Volume integration of HSQC signals used Bruker's TopSpin 3.5 typically following manual phase correction and automatic baseline correction. The unfractionated Indulin AT lignin (Fig. 2) and some of the acetylated lignins were analyzed with increased resolution in F1 (256 or 384 increments) allowing for the identification of minor units such as vanillin and vanillic acid. Heteronuclear Multiple Bond Correlation (HMBC) spectra were recorded using the Bruker pulse sequence 'hmbcgplpndqf' using the following parameters: acquired from 11 to -1 ppm in F2 (^{1}H) with 2048 data points, 230 to 0 ppm in F1 (^{13}C) with 256 increments with a 1.5 s interscan delay (D1). Processing used one level of linear prediction (32 coefficients) in F1.

Quantitative $HSQC_0$ experiments were carried out according to previous reports using the non-constant time version of the pulse sequences available at http://pine.nmrfam.wisc.edu/download_pulseprogs.html.[1–3] Spectra were recorded using the following parameters: acquired from 13 to -1 ppm in F2 (^{1}H) with 2048 data points and either 50 to 90 ppm with 32 increments or 47-137 ppm with 72 increments in F1 (^{13}C) and a 7.5 s interscan delay; cnst2 was set to 145 Hz. Processing used Gaussian apodization (GB = 0.1, LB = 0.3 Hz) in F2 and squared cosine-bell and one level of linear prediction (8 or 18 coefficients respectively) in F1. Processing of integration data and linear regression was performed in Microsoft Excel 2016.

For NMR analyses lignins were dissolved at a concentration of approximately 170 mg/1 mL in the NMR solvent. For unacetylated lignins and synthetic Kraft reaction mixtures DMSO-d_6 was used as solvent and the central DMSO solvent peak was used as internal reference (δC 39.6, δH 2.49 ppm). For acetylated lignin samples $CDCl_3$ was used as a solvent and the minor $CHCl_3$ solvent peak was used as internal reference (δC 77.2, δH 7.26 ppm). The relative quantity of side chains and other units are expressed as a number per 100 aromatic units (100Ar) (based on comparison to the G+½S aromatic integral as previously reported).[4] Integrals from symmetrical units or those corresponding to non-diastereotopic CH_2's (i.e. β-β, β-1 stilbenes, SR, DHCA, CA) were halved to calculate the number of linkages/units per 100Ar.

For the ^{31}P-NMR measurements, the lignin samples were analyzed in triplicate using a standard phosphitylation procedure.[5] About 30 mg of dry lignin (accurately weighed) was dissolved in pyridine/$CDCl_3$ (300 µl, 1.6/1.0, v/v). Stock solutions of the internal standard (cholesterol, 20.9 mg/mL) and relaxation reagent (chromium (III) acetylacetonate, 10.5 mg/mL) were prepared separately using the same pyridine/$CDCl_3$ solvent mixture, with 150 µl and 75 µL respectively added to the lignin mixture. Subsequently, 75 µL of derivatization reagent (2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane) was added, after which the mixture was transferred into a 5-mm-OD NMR tube for analysis. ^{31}P NMR spectra were obtained on a Varian 400 MHz NMR spectrometer using a standard phosphorus pulse program with a relaxation delay of 5 s and 256 accumulated scans. Chemical shifts were referenced using the sharp signal arising from the product from residual water and 2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane at 132.2 ppm.

Gel Permeation Chromatography (GPC) measurements were performed on a Polymer Labs GPC 50 system, equipped with a series of three PLGel Mixed-E columns and a guard column and using THF spiked with 0.1 vol% acetic acid as mobile phase. Detection was done with an external Knauer UV detector at 280 nm and molecular weight determinations were based on calibration with polystyrene standards (M_n = 162, 570, 1060, 1400, 2240, 3690, 4760, 7130, 12800 and 19690). Samples were acetylated (pyridine/acetic anhydride overnight then concentrated *in vacuo*) before analysis.

1.2 Chemicals and Materials

All chemicals were used as received from commercial suppliers. Sodium sulfide nonahydrate (98%), acetovanillone (≥98%), guaiacol (≥99%), ruthenium (III) chloride (45-55% Ru), formaldehyde solution (37 wt% with 10-15% methanol), 5% Pd/C, isoeugenol (98% mixture of *cis* and *trans*), sodium thiosulfate, triethylamine (≥99%), acetic anhydride (≥98%), ethyl acetate-1-^{13}C (99% ^{13}C), ethyl acetate-2-^{13}C (99% ^{13}C), 2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane (95%), chromium(III)

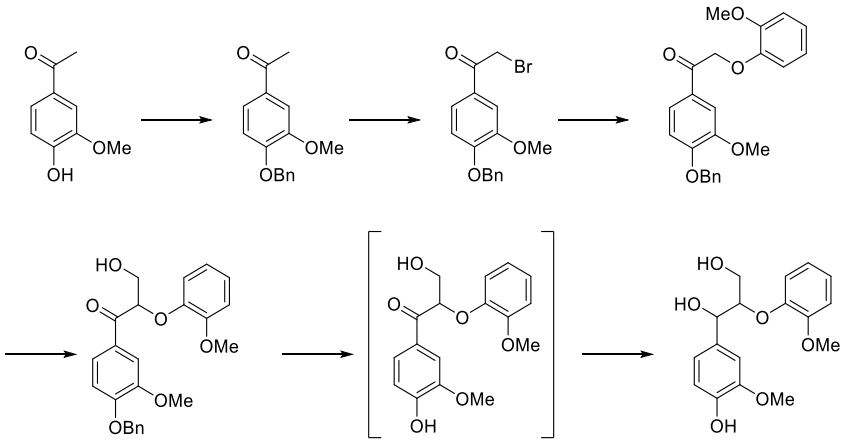
acetylacetonate (97%), dimethylformamide (99.8%) and Hoveyda-Grubbs Catalyst™ 2nd Generation were all purchased from Sigma-Aldrich. Benzyl chloride (99%), pyridine (99+%) and bromine (99+%) were purchased from Acros Organics. Acetyleugenol was purchased from TCI. Potassium carbonate and sodium hydroxide were purchased from VWR. Solvents (hexanes, ethanol, methanol, ethyl acetate, toluene, dichloromethane) were purchased from InterCheM (Netherlands). The Indulin AT lignin used throughout this study (Batch SA031) was purchased from MeadWestVaco. Other lignins were kindly donated by UPM (BioChoice), Raiz, Portugal (Eucalyptus), ECN (Indulin AT – labelled 'in kind'). Kraft black liquors were kindly donated by Celbi, Portugal (heavy black liquor) and Smurfit Kappa (normal black liquor) from which Kraft lignins were isolated by acidification with aqueous HCl.

TLC analysis was performed on glass backed silica gel 60 plates (Merck) and visualized under a UV light (254 nm) or by staining with a cerium molybdate stain (Hanessian's stain). Silica gel column chromatography was performed with silica gel 60A (40-63μ) from Fluorochem, UK. Aminopropyl silica was prepared by refluxing silica gel 60A in a 10% solution of (3-aminopropyl)triethoxysilane in toluene for 3 h. The silica gel was filtered, washed thoroughly with toluene and methanol and then dried under vacuum at 80 °C. Aminopropyl TLC plates were prepared by dipping standard TLC plates in a hexane/(3-aminopropyl)triethoxysilane and drying the plate at ~100 °C for 10 min. NMR solvents were purchased from Cambridge Isotope Labs *via* BUCHEM (Netherlands).

Spruce CEL was isolated based on previously reported methods from chips of spruce wood obtained locally in the Netherlands.[6] Briefly, 7.0 g of spruce wood chips were ball milled in a 125 mL agate jar containing agate balls (6 x 20 mm, 20 x 10 mm) for 48 hours using a Fritsch 'Pulverisette 7' planetary ball mill at speed setting 7. The ball milled wood was then digested with Celluclast 1.5L (Sigma) (1 mL/g) in acetate buffer (pH 5.2, 50 mM, 2 wt% loading) at 50 °C for 48 h with stirring. The mixture was filtered using a nylon membrane (45 μm) and the collected solid washed with water and dried. The collected solid (2.1g, enzyme lignin) was then ball milled in 50 mL agate jars (1 g per jar) containing agate balls (1 x 20 mm, 1 x 15 mm, 6 x 10 mm) for 18 hours at speed setting 7. The ball milled enzyme lignin was dissolved in dioxane/water (1:1, 50 mL) and centrifuged to remove small amounts of insoluble dark material. The soluble fraction was precipitated by pouring into acidified (AcOH, pH ~3) water (1 L) and the solid was collected by centrifugation and dried *in vacuo* over molecular sieves. The resulting powder (CEL) was completely soluble in DMSO-d_6 and used as such for characterization.

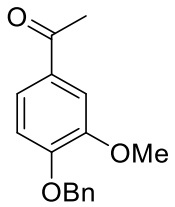
2. Synthesis of Model Compounds

Optimized, chromatography free guaiacylglycerol-β-guaiacyl ether (3) Synthesis



Scheme S1 Steps in the optimized synthesis of guaiacylglycerol-β-guaiacyl Ether (**3**)

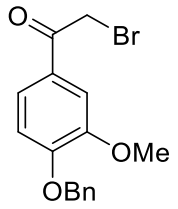
1-(4-(Benzyloxy)-3-methoxyphenyl)ethan-1-one (S1)



To a solution of acetovanillone (25.0 g, 150 mmol, 1.0 eq.) in DMF (50 mL) was added benzyl chloride (20.0 g, 158 mmol, 1.05 eq.) and K_2CO_3 (41.5 g, 301 mmol, 2 eq.). The mixture was stirred at 70 °C for 2 h until complete conversion as indicated by TLC. The reaction mixture was cooled to room temperature, diluted with EtOH (100 mL) and poured slowly into rapidly stirring water (1 L). The white precipitate was vigorously stirred for 10 minutes, collected by filtration and washed with water (500 mL). The solid was dried overnight *in vacuo* over 4 Å molecule sieves to give the title compound as a white solid (38.4 g, 100%).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.55 (d, *J* = 2.2 Hz, 1H), 7.50 (dd, *J* = 8.2, 2.2 Hz, 1H), 7.48 – 7.29 (m, 5H), 6.89 (d, *J* = 8.2 Hz, 1H), 5.23 (s, 2H), 3.94 (s, 3H), 2.55 (s, 3H) ppm.
Spectral data are in accordance with literature.[7]

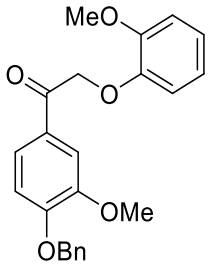
1-(4-(Benzyloxy)-3-methoxyphenyl)-2-bromoethan-1-one (S2)



To a solution of 1-(4-(benzyloxy)-3-methoxyphenyl)ethan-1-one (**S1**) (35 g, 137 mmol, 1 eq.) in a mixture of EtOH (700 mL) and dichloromethane (70 mL) was added dropwise a solution of Br_2 (24.0 g, 150 mmol, 1.1 eq.) in cyclohexane (80 mL). During the addition, the reaction mixture was sparged with a slow stream of nitrogen delivered *via* a glass pipette. After the bromine had been completely consumed (judged by the reaction color) the reaction was cooled on ice causing a white precipitate to form. This was collected by filtration and washed with a small amount of cold EtOH to yield the title compound as a voluminous white solid (40.7 g, 89%).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.56 (d, *J* = 2.1 Hz, 1H), 7.53 (dd, *J* = 8.4, 2.1 Hz, 1H), 7.47 – 7.30 (m, 5H), 6.91 (d, *J* = 8.4 Hz, 1H), 5.25 (s, 2H), 4.39 (s, 2H), 3.95 (s, 3H) ppm.
Spectral data are in accordance with literature.[8]

1-(4-(Benzyloxy)-3-methoxyphenyl)-2-(2-methoxyphenoxy)ethan-1-one (S3)



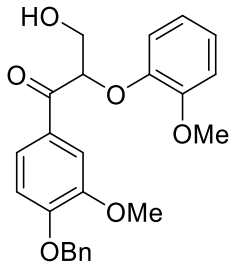
To a solution of 1-(4-(benzyloxy)-3-methoxyphenyl)-2-bromoethan-1-one (**S2**) (40.0 g, 119.33 mmol, 1 eq.) in acetone (200 mL) was added guaiacol (14.8 g, 119 mmol, 1 eq.) and K_2CO_3 (24.7 g, 1.5 eq.). The mixture was then heated to reflux for 2 h, allowed to cool to room temperature, filtered and concentrated *in vacuo*. The resulting oil was dissolved in MeOH (200 mL) and vigorously scratched with a spatula to induce crystallisation. The resulting solid was collected by filtration and washed with MeOH to yield the title compound as a white powder (41.0 g, 91%). TLC (Hexanes:EtOAc, 6:3 v/v): Rf = 0.38

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.67 – 7.55 (m, 2H), 7.48 – 7.28 (m, 5H), 7.00 – 6.77 (m, 5H), 5.27 (s, 2H), 5.23 (s, 2H), 3.94 (s, 3H), 3.87 (s, 3H) ppm.

^{13}C NMR (101 MHz, Chloroform-*d*) δ 193.3, 153.0, 149.8, 149.8, 147.7, 136.3, 128.8, 128.3, 128.2, 127.3, 122.6, 122.4, 120.9, 114.8, 112.3, 112.3, 111.0, 72.1, 70.9, 56.2, 56.0 ppm.

HR-MS (ESI) $C_{23}H_{23}O_5$ $[M+H]^+$ m/z required 379.1540; found 379.1548.

1-(4-(Benzyloxy)-3-methoxyphenyl)-3-hydroxy-2-(2-methoxyphenoxy)propan-1-one (S4)



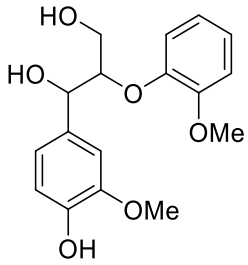
To a solution of 1-(4-(benzyloxy)-3-methoxyphenyl)-2-(2-methoxyphenoxy)ethan-1-one (**S3**) (30.0 g, 117 mmol, 1 eq.) in 1,4-dioxane (300 mL) formaldehyde (3.9 g, 129 mmol, 1.1 eq., 9.56 mL of a 37 wt% solution in water) and K_2CO_3 (16.2 g, 117 mmol, 1 eq.) were added. The mixture was then sealed and vigorously stirred overnight at room temperature, then filtered and concentrated *in vacuo*. The resulting oil was dissolved in MeOH (300 mL) and vigorously scratched with a spatula to induce crystallisation. The resulting white precipitate was collected by filtration and washed with methanol to yield the title compound as a white solid (26.7 g, 56%). TLC (Hexanes:EtOAc, 3:7 v/v): Rf = 0.66.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.66 (dd, *J* = 8.4, 1.8 Hz, 1H), 7.63 (d, *J* = 1.8 Hz, 1H), 7.48 – 7.28 (m, 5H), 7.04 – 6.96 (m, 1H), 6.93 – 6.85 (m, 3H), 6.85 – 6.77 (m, 1H), 5.38 (t, *J* = 5.3 Hz, 1H), 5.22 (s, 2H), 4.05 (d, *J* = 5.3 Hz, 2H), 3.92 (s, 3H), 3.84 (s, 3H) ppm.

^{13}C NMR (101 MHz, Chloroform-*d*) δ 195.1, 153.2, 150.5, 149.8, 147.1, 136.2, 128.9, 128.3, 128.3, 127.3, 123.7, 123.5, 121.3, 118.3, 112.4, 112.3, 111.5, 84.5, 71.0, 63.9, 56.2, 55.9.

HR-MS (ESI) $C_{24}H_{25}O_6$ $[M+H]^+$ m/z required 409.1646; found 409.1649.

Guaiacylglycerol-β-guaiacyl ether (3) / 3-Hydroxy-1-(4-hydroxy-3-methoxyphenyl)-2-(2-methoxyphenoxy)propan-1-one

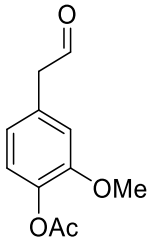


In a 250 mL flask a solution of 1-(4-(benzyloxy)-3-methoxyphenyl)-3-hydroxy-2-(2-methoxyphenoxy)propan-1-one (**S4**) (3.0 g, 7.34 mmol, 1 eq.) in EtOAc (30 mL) and EtOH (45 mL) was degassed for 5 min under vacuum. 5% Pd/C (150 mg) was added under nitrogen and the flask evacuated again. A hydrogen balloon was attached and after the mixture vigorously stirred overnight. The reaction mixture was then filtered through a 0.45 μm nylon filter and concentrated *in vacuo* to yield the title compound as a mixture of diastereomers (~2:1, major:minor, colourless oil, 2.2 g, 94%).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.19 – 6.72 (m, 7H), 5.05 – 4.88 (m, 1H), 4.19 – 4.12 (m, 0.64H), 4.04 – 3.99 (m, 0.36H), 3.94 – 3.85 (m, 6.64H), 3.70 – 3.59 (m, 1H), 3.48 (dd, *J* = 12.5, 3.8 Hz, 0.36H) ppm.

Spectral data are in accordance with literature.[9]

Acetyl Homovanillin (8)

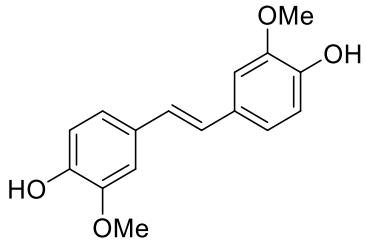


To a solution of acetyl eugenol (6.0 g, 14.5 mmol, 1 eq.) in a mixture of acetonitrile/ethyl acetate/water (3:3:1, 60 mL) containing H_2SO_4 (285 mg, 0.2 eq.) anhydrous $RuCl_3$ (30 mg, 0.01 eq.) was added followed by the portion-wise addition of $NaIO_4$ (7.5 g, 31.4 mmol, 2.4 eq.). The mixture was then allowed to stir for 3 h. The reaction was then quenched by addition of an aqueous solution of $Na_2S_2O_3$, diluted with water (100 mL) and extracted with EtOAc (2x100 mL). The combined EtOAc extracts were washed with aqueous $NaHCO_3$ and brine, dried over $MgSO_4$ and concentrated *in vacuo* below ~35 °C. The crude product was obtained as a pale-yellow oil (2.9 g) and was used without further

purification. The product is quite unstable, but can be stored for some time at 4 °C. The time required for completion of the reaction varies somewhat –stopping the reaction early results in a mixture of **7** and an intermediate diol.

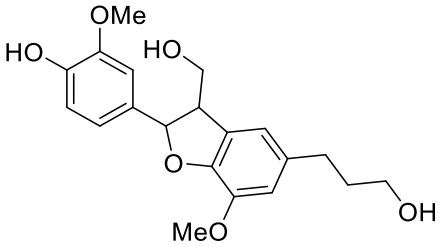
^{1}H NMR (400 MHz, Chloroform-*d*) δ 9.75 (t, *J* = 2.4 Hz, 1H), 7.03 (d, *J* = 8.2 Hz, 1H), 6.85 – 6.77 (m, 2H), 3.83 (s, 3H), 3.67 (d, *J* = 2.4 Hz, 2H), 2.32 (s, 3H) ppm.
^{13}C NMR (101 MHz, Chloroform-*d*) δ 199.2, 169.2, 151.5, 139.2, 130.8, 123.4, 122.0, 113.7, 56.0, 50.5, 20.8 ppm.
HR-MS (ESI) $C_{11}H_{13}O_4$ $[M+H]^+$ m/z required 209.0808; found 209.0812.

(*E*)-4,4'-(Ethene-1,2-diyl)bis(2-methoxyphenol) (β-1 stilbene, S5)



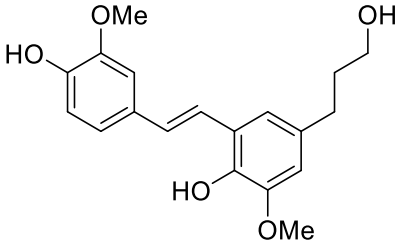
To neat isoeugenol (500 mg, 3.0 mmol) at 90 °C was added Hoveyda-Grubbs Catalyst™ 2nd Generation (~2 mg, 0.003 mmol, ~1 mol%) whilst stirring. Vigorous gas evolution followed and within ~20 s the reaction mixture had set solid. The reaction was then cooled to room temperature and the solid was broken up with a spatula and washed with diethyl ether. The solid was collected by filtration to give the product as a grey-pink solid (398 mg, 96%). The product was further purified by recrystallisation from methanol.
^{1}H NMR (400 MHz, DMSO-d_6) δ 9.03 (s, 2H), 7.13 (d, *J* = 1.9 Hz, 2H), 6.96 – 6.91 (m, 4H), 6.74 (d, *J* = 8.0 Hz, 2H), 3.82 (s, 6H) ppm.
^{13}C NMR (101 MHz, DMSO-d_6) δ 147.8, 146.1, 129.2, 125.7, 119.5, 115.6, 109.5, 55.6 ppm.
Spectral data are in accordance with literature.[10]

Dihydrodehydrodiconiferyl alcohol (S6)



Prepared from methyl ferulate *via* horseradish peroxide dimerization, hydrogenation and $LiAlH_4$ reduction according to literature methods. [11–13]

(E)-2-(4-Hydroxy-3-methoxystyryl)-4-(3-hydroxypropyl)-6-methoxyphenol (β-5 stilbene, S7)

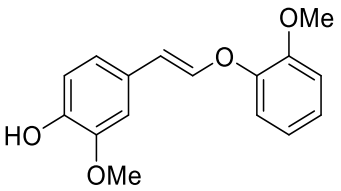


Prepared from dihydrodehydrodiconiferyl alcohol (**S6**) under Kraft conditions according to literature methods.[14] The compound appeared to be unstable upon prolonged standing in DMSO-d_6 at room temperature.

^{1}H NMR (600 MHz, DMSO-d_6) δ 9.07 (s, 1H), 8.54 (s, 1H), 7.19 (d, *J* = 16.4 Hz, 1H), 7.10 (s, 1H), 7.05 (d, *J* = 16.4 Hz, 1H), 6.96 (d, *J* = 1.2 Hz, 2H), 6.94 (d, *J* = 7.9 Hz, 1H), 6.75 (d, *J* = 8.1 Hz, 1H), 6.67 (s, 1H), 3.81 (s, 3H), 3.79 (s, 3H), 3.42 (t, *J* = 6.4 Hz, 2H), 2.56 – 2.50 (m, 2H), 1.80 – 1.66 (m, 2H) ppm.

^{13}C NMR (101 MHz, DMSO-d_6) δ 147.8, 147.7, 146.4, 141.6, 132.6, 129.3, 128.1, 124.0, 120.5, 119.6, 117.2, 115.6, 110.6, 109.7, 60.2, 55.8, 55.5, 34.6, 31.5 ppm.

(*E*)-2-methoxy-4-(2-(2-methoxyphenoxy)vinyl)phenol (S8)

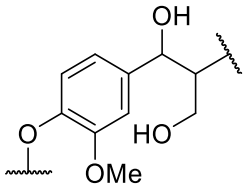


4-(1-Hydroxy-2-(2-methoxyphenoxy)ethyl)-2-methoxyphenol[15] (200 mg, 0.69 mmol) was dissolved 1 M NaOH (4 mL) and heated at 170 °C for 1 h. The solution was then cooled to ~4 °C resulting in the formation of a white precipitate. This was collected by filtration and washed with a few drops of ice cold water (NB: product is water soluble). The solid was partitioned between DCM (5 mL) and 1 M HCl

(5 mL) and vigorously shaken. The organic layer was collected, dried ($MgSO_4$) and concentrated *in vacuo* to give the title compound as a light yellow oil containing 97+% of the *E* isomer (51 mg, 28%). On standing in solution the *Z* isomer slowly forms.

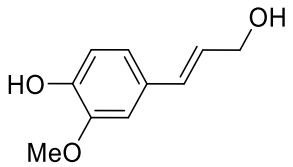
^{1}H NMR (600 MHz, DMSO-d_6) δ 7.50 (d, *J* = 2.1 Hz, 1H), 7.18 (dd, *J* = 8.0, 1.8 Hz, 1H), 7.11 (dd, *J* = 7.9, 1.9 Hz, 1H), 7.07 (td, *J* = 8.0, 2.0 Hz, 1H), 7.01 (dd, *J* = 8.3, 1.9 Hz, 1H), 6.94 (td, *J* = 7.6, 2.0 Hz, 1H), 6.74 (d, *J* = 8.1 Hz, 1H), 6.70 (d, *J* = 6.8 Hz, 1H), 5.55 (d, *J* = 6.8 Hz, 1H), 3.83 (s, 3H), 3.78 (s, 3H) ppm.
^{13}C NMR (101 MHz, DMSO-d_6) δ 149.2, 147.2, 145.8, 145.4, 139.8, 126.4, 123.6, 121.7, 120.8, 115.8, 115.3, 112.8, 112.5, 109.5, 55.8, 55.3 ppm.
Spectral data are in accordance with literature.[16]

β-O-4 model polymer (4)



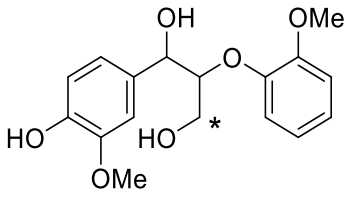
Phenolic β-O-4 model polymer was prepared according to literature methods.[17] Analysis of an acetylated sample (Ac_2O/Pyridine, overnight) of polymer by GPC indicated Mn = 1485, Mw = 2776. Quantitative ^{1}H NMR end-group analysis indicated an average degree of polymerisation of approximately 10 (i.e. 9 β-O-4 linkages per polymer chain) corresponding to an average chain molecular weight of ~1920 Da.

Coniferyl alcohol (2)

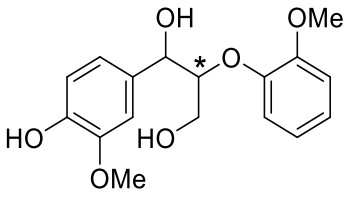


Prepared from methyl ferulate *via* DIBAL reduction in toluene according to literature methods.[18]

^{13}C Labelled β-O-4 model



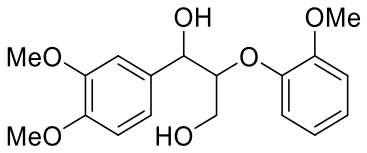
3-γ*



3-β*

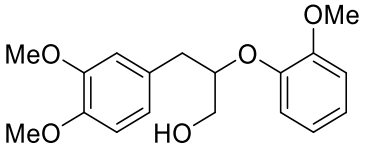
Prepared according to literature methods used for the synthesis of non-labelled compounds.[9,19] Substitution of C1 (98% ^{13}C) and C2 (99% ^{13}C) labelled ethyl bromoacetate (Sigma-Aldrich) in these syntheses gave the labelled model compounds **3-γ*** and **3-β*** respectively.

1-(3,4-Dimethoxyphenyl)-2-(2-methoxyphenoxy)propane-1,3-diol (S10)



Prepared according to literature methods.[20,21]

3-(3,4-Dimethoxyphenyl)-2-(2-methoxyphenoxy)propan-1-ol (S11)



1-(3,4-Dimethoxyphenyl)-2-(2-methoxyphenoxy)propane-1,3-diol (S10) (300 mg, 0.90 mmol, 1 eq.) was dissolved in a 4:1 mixture of acetic acid and acetyl bromide (5 mL) and stirred, protected from light, at room temperature for 16 hours. The reaction mixture was then concentrated *in vacuo* and residual acetic acid removed by azeotropic distillation with hexanes (5 x 20 mL). The remaining oil was dissolved in THF (6 mL) and a solution of $LiAlH_4$ in THF (2.4 M, 5 eq.) was added dropwise at -78 °C. The mixture was allowed to warm to room temperature and stirred for 2 hours. The reaction was quenched by the slow addition of EtOAc (5 mL) at 0 °C, acidified with 1 M HCl and partitioned between water (20 mL) and EtOAc (20 mL). The organic layer was washed with brine, dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography eluting with 20-40% EtOAc/hexanes to yield the product as a colourless oil (168 mg, 59%).

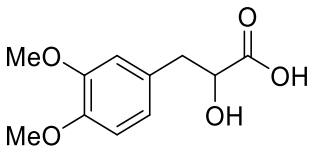
^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.00 (ddd, *J* = 8.2, 7.3, 1.7 Hz, 1H), 6.90 (dd, *J* = 8.2, 1.6 Hz, 1H), 6.85 (ddd, *J* = 7.7, 7.7, 1.6 Hz, 1H), 6.82 – 6.75 (m, 4H), 4.25 (dddd, *J* = 6.8, 6.8, 5.3, 3.1 Hz, 1H), 3.87

(s, 3H), 3.86 (s, 3H), 3.85 (s, 3H), 3.73 – 3.58 (m, 2H), 3.08 (dd, *J* = 13.9, 6.8 Hz, 1H), 2.92 (dd, *J* = 13.9, 6.8 Hz, 1H).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 151.3, 149.0, 147.8, 147.6, 130.5, 123.6, 121.6, 121.5, 120.2, 112.9, 112.2, 111.4, 85.2, 63.6, 56.0, 56.0, 55.9, 37.4.

HR-MS (ESI) $C_{18}H_{22}O_5Na$ $[M+Na]^+$ m/z required 341.1359; found 341.1365.

Methyl 3-(3,4-dimethoxyphenyl)-2-hydroxypropanoate (S12)



Prepared from 3-(3,4-dimethoxyphenyl)oxirane-2-carboxylate[22] according to literature methods.[23]

3. Synthetic Kraft Reactions

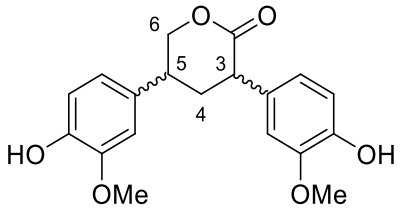
In a typical experiment, a Teflon-lined stainless steel autoclave (12 mL volume) was charged with a pre-mixed solution of white liquor and model compound. The autoclave was sealed and placed in a preheated oil bath at 170 °C. The reactions were then kept in the oil bath for 2 h before being removed and allowed to cool to room temperature. No stirring was used during these experiments. The autoclave content was acidified with 1 M HCl and either directly extracted with EtOAc or first centrifuged to remove the precipitate (2 minutes, 4000 rpm) and then extracted. Extracts were washed with brine, dried ($MgSO_4$) and concentrated *in vacuo*. Precipitates were washed with water and then dried *in vacuo* over molecular sieves.

Caution should be taken during work up as substantial quantities of H_2S are released – an efficient fume hood should be used at all times and other suitable precautions should be taken to avoid release of this gas into the general lab environment.

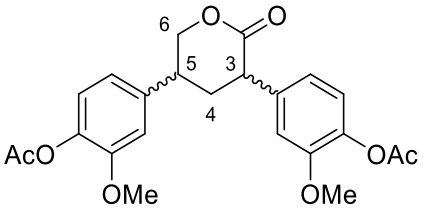
Synthetic Kraft reactions with the model compounds were run under dilute (10 mg/mL) and concentrated conditions (100 mg/mL). All experiments referred to in the main text were run using a synthetic Kraft liquor prepared by dissolving NaOH (0.48 g/100 mL) and $Na_2S.9H_2O$ (6 g/100 mL) in deionized water to give a solution containing 0.12 M NaOH and 0.25 M Na_2S.

An experiment using a 'strong white liquor' mixture prepared from NaOH (4.8 g/100 mL) and $Na_2S.9H_2O$ (6 g/100 mL), which mimics the starting cooking liquor used during the Kraft process, resulted in the formation of unrealistically large amounts of enol ethers (Figure S35). Given this result and the fact that a large proportion of the active alkali in the cooking liquor is known to be rapidly consumed early in the Kraft process by neutralization reactions with carbohydrates,[24] this strong white liquor was not considered a suitable starting point for the investigation of Kraft chemistry.

4. Isolation and Synthesis of Kraft-Derived Lactone



5a/b



S13a/b

The crude EtOAc soluble products obtained from the Kraft reaction of guaiacylglycerol-β-guaiacyl ether (**3**) (2 x 90 mg) under dilute conditions were purified by silica gel column chromatography (~10 g silica) eluting with 30-100% EtOAc/hexanes. Similar fractions (as judged by TLC) were combined and concentrated *in vacuo*. The fraction containing the lactone (22 mg) was re-columned (~4 g silica) eluting with 40-80% EtOAc/hexanes. All compounds eluted similarly so the fractions containing UV active components were split into four sequential groups and analyzed by NMR.

The fraction containing mostly lactones **5a/b** was acetylated (Pyridine/Ac_2O) and purified by pipette column eluting with 0-2% MeOH/DCM to give ~1 mg of compound for analysis. TLC (Hexanes:EtOAc, 3:7 v/v): Rf = 0.46. Detailed 2D NMR analysis combined with high resolution mass spectrometry allowed for the identification of these compounds as lactones **S13a/b** based on the assignments below.

HR-MS (ESI) $C_{23}H_{24}O_8Na$ $[M+Na]^+$ m/z required 451.1363; found 451.1369.

Major Diastereomer: S13a (*syn*)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.03 (d, *J* = 7.8 Hz, 1H), 7.02 (d, *J* = 8.6 Hz, 1H), 6.88 – 6.79 (m, 4H), 4.61 (ddd, *J* = 11.3, 5.2, 2.3 Hz, 1H, H6'), 4.41 (dd, *J* = 11.3, 10.7 Hz, 1H, H6''), 3.93 – 3.69 (m, 7H, 2xOMe, H3), 3.45 – 3.35 (m, 1H, H5), 2.59 – 2.49 (m, 1H, H4'), 2.38 – 2.24 (m, 7H, 2xOAc, H4'') ppm. Note: J^4 'W' coupling is observed between H6' and H4'.

^{13}C NMR (Chloroform-*d*) δ 170.9, 169.1 (2xC), 151.1 (2xC), 138.9 (2xC), 137.5 (2xC), 123.2 (2xC), 120.3, 119.3, 112.5, 111.2, 74.3 (C6), 55.9 (2xC), 48.1 (C3), 40.2 (C5), 36.5 (C4), 20.4 (2xC) ppm.

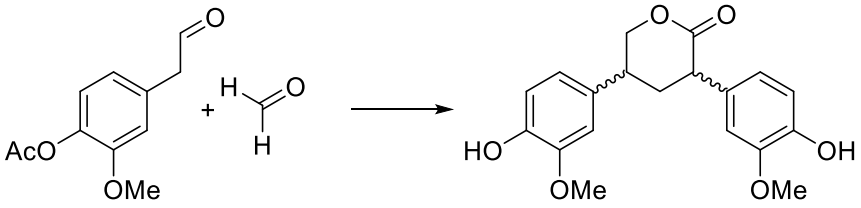
Minor Diastereomer: S13b (*anti*)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.03 – 6.97 (m, 2H), 6.86 – 6.79 (m, 4H), 4.53 (dd, J = 11.4, 5.3 Hz, 1H, H6'), 4.46 (dd, J = 11.4, 9.6 Hz, 1H, H6''), 4.00 – 3.96 (m, 1H, H3), 3.93 – 3.69 (m, 6H), 3.43 – 3.32 (m, 1H, H5), 2.56 – 2.45 (m, 2H, 2xH4), 2.36 – 2.23 (m, 6H) ppm.

^{13}C NMR (Chloroform-*d, partial from HSQC*) δ 73.1 (C6), 44.9 (C3), 37.3 (C5), 35.0 (C4) ppm.

Based on quantitative ^{1}H NMR analysis of the crude reaction mixture, using 1,3,5-trimethoxybenzene as an internal standard, the yield of **5a/b** was estimated to be 3%.

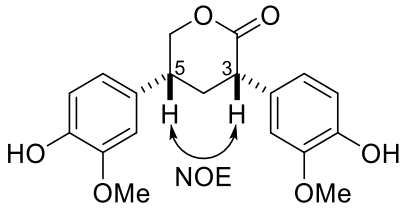
Having identified the new lactone species additional samples of **5** required for more detailed analysis were obtained from the reaction of acetylhomovanillin and formaldehyde as follows:



A solution of acetylhomovanillin **8** (2.8 g, 13.4 mmol, 1 eq.) and formaldehyde (1.09 g of 37 wt% solution, 13.4 mmol, 1 eq.) in the Kraft liquor (280 mL) was heated at 95 °C for 1 h with vigorous stirring. The solution was then allowed to cool to room temperature and acidified by the slow addition of 6 M HCl with stirring. The mixture was filtered to remove a brown precipitate and the filtrate was extracted with EtOAc. The organic extract was washed with brine, dried ($MgSO_4$) and concentrated *in vacuo* to give a yellow/orange foam. Purification was achieved by silica gel column chromatography eluting with 30-100% EtOAc/hexanes. Fractions containing **5** were concentrated *in vacuo* to give a light yellow oil containing mostly the lactones (103 mg) with some small impurities. The lactones were further purified by column chromatography using aminopropyl funtionalized silica gel eluting with 60-100% EtOAc/hexanes containing 1% triethylamine. From TLC analysis (using aminopropyl functionalized silica TLC plates) a small number of fractions (eluting first) containing only 1 spot were identified. These were pooled and concentrated *in vacuo* giving a mixture of **5a** and **5b** (17 mg), which was used for further analysis. Most material was still obtained as a mixture with contaminating impurities. Additionally, 2D TLC analysis indicated some instability of **5** on aminopropyl functionalized silica resulting in compound losses during purification. TLC (Hexanes:EtOAc, 3:7 v/v): Rf = 0.43.

HR-MS (ESI) $C_{19}H_{21}O_6$ $[M+H]^+$ m/z required 345.1333; found 345.1334 (mixture of **5a/b**).

Major Diastereomer 5a (*syn*)

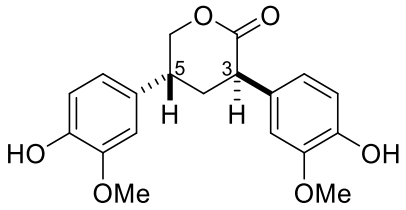


^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.91 (d, *J* = 5.5 Hz, 1H), 6.89 (d, *J* = 5.8 Hz, 1H), 6.81 – 6.71 (m, 4H), 5.60 (br. s, 2H), 4.57 (ddd, *J* = 11.3, 5.2, 2.5 Hz, 1H, H6'), 4.37 (dd, *J* = 11.3, 10.6 Hz, 1H, H6''), 3.90 (s, 6H), 3.87 – 3.78 (m, 1H, H3), 3.42 – 3.26 (m, 1H, H5), 2.54 – 2.42 (m, 1H, H4'), 2.35 – 2.26 (m, 1H, H4'') ppm.

^{13}C NMR (101 MHz, Chloroform-*d*) δ 172.0, 146.9, 146.7, 145.2, 145.2, 131.3, 130.9, 121.2, 119.9, 114.9, 114.8, 111.0, 109.7, 74.9 (C6), 56.1 (2xC), 48.2 (C3), 40.2 (C5), 37.0 (C6) ppm.

Assignment of **5a** as *syn* (and hence **S11a**) was made using NOESY analysis which showed an NOE between H5 and H3.

Minor Diastereomer 5b (*anti*)



^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.93 – 6.86 (m, 2H), 6.80 – 6.70 (m, 4H), 5.60 (s, 2H), 4.53 – 4.49 (m, 1H, H6'), 4.43 (dd, *J* = 11.4, 9.7 Hz, 1H, H6''), 3.97 – 3.90 (m, 1H, H3), 3.90 (s, 6H), 3.42 – 3.33 (m, 1H, H5), 2.53 – 2.42 (m, 2H, 2 x H4) ppm.

^{13}C NMR (101 MHz Chloroform-*d*) δ 172.7, 146.9, 146.8, 145.2, 145.1, 131.8, 131.3, 121.0, 119.9, 115.0, 114.6, 111.0, 110.0, 73.3, 56.1 (2xC), 44.8, 37.2, 35.3 ppm.

5. Lignin NMR Assignments

Table S1 ^{1}H and ^{13}C assignments of useful diagnostic signals for different structural units in kraft lignin HSQC spectra in DMSO-d_6. The cross peaks used in this study for quantification are indicated in bold. Due to broad peaks in lignin HSQC spectra the assignments are given to the center of the cross peak but are still approximate (±0.1 ppm for ^{1}H and ±1 ppm for ^{13}C).

Unit	Name	Diagnostic Peaks for Assignment ($^{13}C/^{1}H$ ppm)	Comments/Notes
A	β-O-4	**α: 70.9/4.77**, β: 84.1/4.30, γ: 59.9, 3.60/3.26	
B	β-5	**α: 86.8/5.49**, β: 53.2/3.47, γ: 62.9/3.73/3.62	
C	β-β	**α: 85.1/4.63**, β: 53.6/3.07, γ: 70.9/4.16/3.76	
D	dibenzodioxocin	α: 83.4/4.82, **β: 85.4/3.87**	
X	cinnamyl alcohol	α: 128.6/6.49, β: 128.4/6.26, **γ: 61.4/4.10**	
SR	secoisolariciresinol	α: 33.8/2.52, **β: 42.3/1.87**, γ: -	
DHCA	dihydrocinnamyl alcohol	α: 31.1/2.51, **β: 34.4/1.70**, γ: 60.0/3.42	
C'	epiresinol	**α: 87.0/4.34**, α': 81.2/4.77, β: 70.2/4.06/3.75, β' 68.8/3.77/3.12, γ: 53.7/2.84, γ': 49.2/3.70	
Z-EE	*Z*-enol ether	**α: 109.1/5.56**, β: 139.7/6.69	
E-EE	*E*-enol ether	**α: 112.0/6.14**, β: 142.8/7.29	
SB1	*trans*-stilbene (β-1)	**α: 125.6/6.97**	
SB2	*trans*-stilbene (β-5)	α: 128.2/7.07, **β: 120.1/7.22**	α cross peak overlaps with $H_{2/6}$. Sometimes assigned as SB1 or just assigned as 'stilbenes'.[25–27]
V	vanillin	**G_2: 110.6/7.40**, G_6: 125.8/7.41	
AV	acetovanillone	**G_2: 111.0/7.45**, G_6: 123.2/7.51	
VA	vanillic acid	**G_2: 112.6/7.45**, G_6: 123.3/7.46	
HA	aryl β-hydroxy propanoic acid	**α: 39.7/2.92**/2.70, β: 71.0/4.10	
Ar	Reduced β-O-4	α: 36.5/2.83, **β: 80.6/4.32**	
J	cinnamaldehyde (etherified)	G_6: 123.2/7.20, **α: 153.4/7.61**, β: 126.1/6.76	Not detected in this study. Previous assignment of 126.3/7.3 (α/β) in Kraft lignin[25] now consider incorrect based on the chemical shifts reported here and elsewhere.[28–30]
-	G-CH(OH)-COOH (mandelic acid)	α: 72.0/4.88	Not detected in this study. Previous assignment of 74.3/4.4[25–27] now considered incorrect based on the new data presented here determined from authentic commercially available compound.
-	homovanillic acid	α: 40.0/3.43	Detected by HMBC but not HSQC. Previous assignment of 39.5/2.4/2.7-2.9[25] now considered incorrect based on the new data presented here determined from authentic commercially available compound.
AG	arylglycerol (acetylated in $CDCl_3$)	α: 73.3/5.92, **β: 71.9/5.42**	Quantification of acetylated unit preferred; unacetylated AG appears at *ca*. α: 73.7/4.43, β: 75.3/3.58 but is significantly overlapped with other cross peaks

Table S2 ^{1}H and ^{13}C assignments and kraft lignin unit abundances reported by Crestini and Argyropoulos *et al.*[25], reproduced here for comparison. Assignments listed in bold could not be verified based on the data presented in Table S1

Unit	Name	Abundance (In house lignin)	Peaks Used for Assignment ($^{13}C/^{1}H$ ppm)	Comments/Notes
A	β-O-4	3.2	71.0/4.8 (Cα-H)	Based on previous work from these authors this value is not adjusted for it being a symmetrical unit[2]
B	β-5	0.8	86.5,5.5 (Cα-H)	
C	β-β	2.4	84.8/4.6 (Cα-H)	
X	cinnamyl alcohol	0.8	61.4/4.10 (Cγ-H)	
SR	secoisolariciresinol	3.2	42.3/1.9 (Cβ–H)	
DHCA	dihydrocinnamyl alcohol	3.4	34.4/1.7 (Cα–H)	
EE	enol ether	1.3	112.1/6.2 (Cα–H) 109.0/5.6 (Cα–H)	
SB	**stilbenes**	**4.8**	**128.2/7.2 (Cα–H, Cβ–H), 128.0/7.1 (Cα–H, Cβ–H)**	**Assigned as β-1 type stilbenes. Chemical shifts correspond to β-5 stilbenes and overlap with other signals (e.g. H units)**
J	**cinnamaldehyde (etherified)**	**0.2**	**126.3/7.3 (Cα/β–H)**	**Appears to be an incorrect assignment based on data in Table S1**
-	**G-CH(OH)-COOH (mandelic acid)**	**0.7**	**74.3/4.4 (Cα–H)**	**Appears to be an incorrect assignment based on data in Table S1**
-	**homovanillic acid**	**0.6**	**39.5/2.4 (Cα–H) 39.5/2.7–2.9 (Cα–H)**	**Appears to be an incorrect assignment based on data in Table S1**
-	**aryl ethyl ketones**	**0.6**	**31.3/2.5 (Cβ–H)**	**Ambiguous assignment - given chemical shifts correspond to DHCA**
-	**aryl hydroxyethyl ketone**	**0.6**	**21.0/1.4 (Cγ–H)**	**Ambiguous assignment - given chemical shifts also correspond to extractives**
V	benzaldehdyes	<0.1	126.5/6.9 (C6–H)	
	Total Ar units (excluding contentious assignments)	18.4		per 100 Ar

Table S3 BioChoice kraft lignin unit abundances reported by Jameel *et al.*[26]

Unit	Name	Abundance (BioChoice)	Comments/Notes
A	β-O-4	2.1	
B	β-5	0.8	
C	β-β	2.2	
X	cinnamyl alcohol	0.9	
SR	secoisolariciresinol	3.1	
DH CA	**dibenzodioxocin**	**0.4**	**Appears to be an incorrect assignment based on data in Table S1. Given chemical shift of 81.1/4.77 ppm the actual chemical shift for this unit is 83.4/4.82.**
EE	enol ether	1.9	
SB	**stilbenes**	**4.5**	**No structural assignment given (SB1 or SB5). Identified cross peaks overlaps with other signals (e.g. H units)**
-	**G-CH(OH)-COOH (Mandelic acid)**	**0.6**	**Appears to be an incorrect assignment based on data in Table S1.**
	Total Ar units (excluding ambiguous/incorrect assignments)	16.3	per 100 Ar

Table S4 BioChoice and Indulin kraft lignin unit abundances reported by Jameel *et al.*[27]

Unit	Name	Abundance (BioChoice)	Abundance (Indulin)	Comments/Notes
A	β-O-4	2.1	8.2	
B	β-5	1.6	1.1	
C	β-β	2.5	1.2	
X	cinnamyl alcohol	0.7	0.3	
SR	secoisolariciresinol	3.5	3.1	
DH CA	**dibenzodioxocin**	**0.8**	**0.6**	**Appears to be an incorrect assignment although chemical shifts are not given. Visual inspection suggests same assignment as in Table S3**
EE	enol ether	2.2	1.5	
SB	**stilbenes**	**7.3**	**6.7**	**No structural assignment given. Identified cross peaks overlaps with other signals (e.g. H units)**
-	**G-CH(OH)-COOH (Mandelic acid)**	**0.7**	**0.4**	**Appears to be an incorrect assignment although chemical shifts are not given. Visual inspection suggests same assignment as in Table S3**
	Total Ar units (excluding ambiguous/incorrect assignments)	18.6	19.7	per 100 Ar

Table S5 Indulin kraft lignin unit abundances reported by Huijgen, Gosselink and Bruijnincx *et al.*[31]

Unit	Name	Abundance (Indulin)	Comments/Notes
A	β-O-4	6.1	
B	β-5	0.3	
C	β-β	1.0	
SB	stilbenes	2.3	β-1 stilbenes only
	Total Ar units (excluding ambiguous/incorrect assignments)	13	per 100 Ar

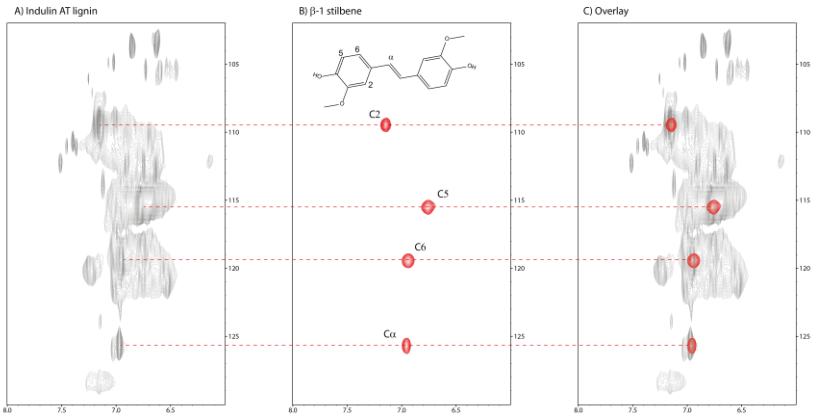


Figure S1 Assignment of β-1 stilbenes in the HSQC spectrum of Kraft lignin. This unit has previously been assigned using the commercially available 1-methoxy-4-[2-(4-methoxyphenyl)ethenyl]benzene.[31] Here this assignment is enforced through the synthesis of (*E*)-4,4'-(ethene-1,2-diyl)bis(2-methoxyphenol) (**S5**). Note the compound numbering in the figures does not reflect the IUPAC name but rather lignin convention.

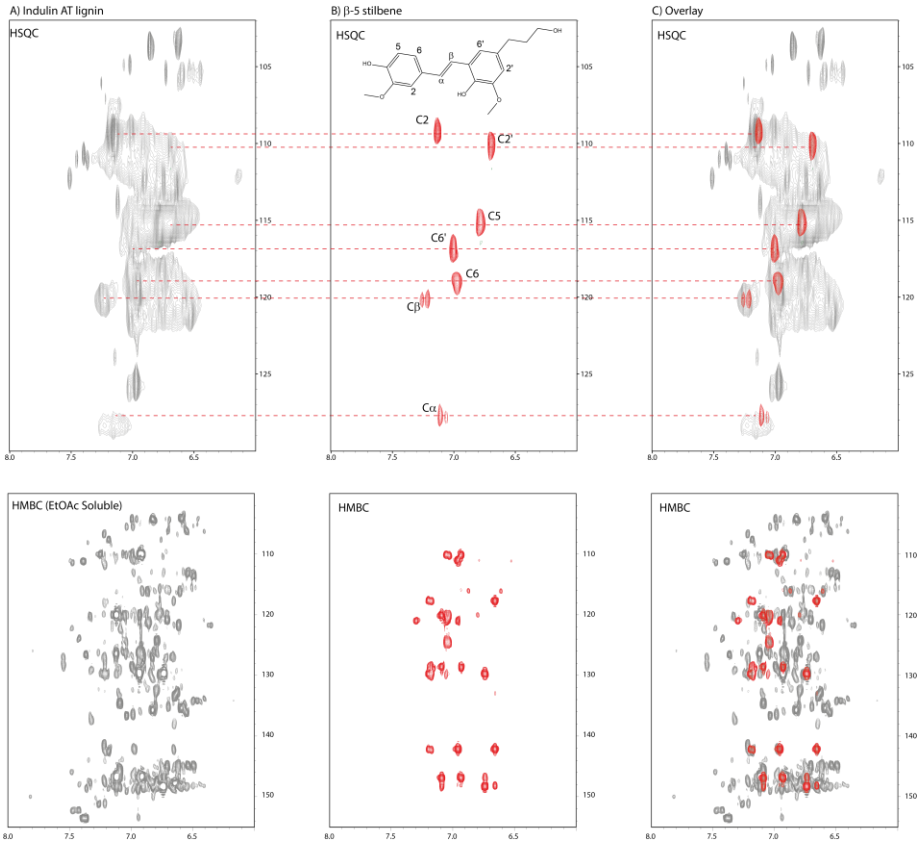


Figure S2 Assignment of β-5 stilbene in Kraft lignin with HSQC (Top) and HMBC (Bottom) using model compound **S7**. In some previous reports the cross peak at ~7.1/128.4 ppm has been assigned as a general indicator of stilbenes or specifically β-1 stilbenes-type stilbenes.[25–27] On its own, this cross peak is, in fact, not diagnostic for stilbenes at all as it overlaps with the $H_{2/6}$ cross peak arising from the small amounts of H units which are typically found in softwood lignins. The presence of H units in Kraft lignin is quite clear from the ^{31}P NMR analysis (Figure S28). Panel C shows the overlap of the seven signals of the β-5 stilbene with the Kraft lignin in the HSQC spectra, together with extensive overlap in the HMBC spectra, which allows for the unambiguous assignment of this unit in Kraft lignin.

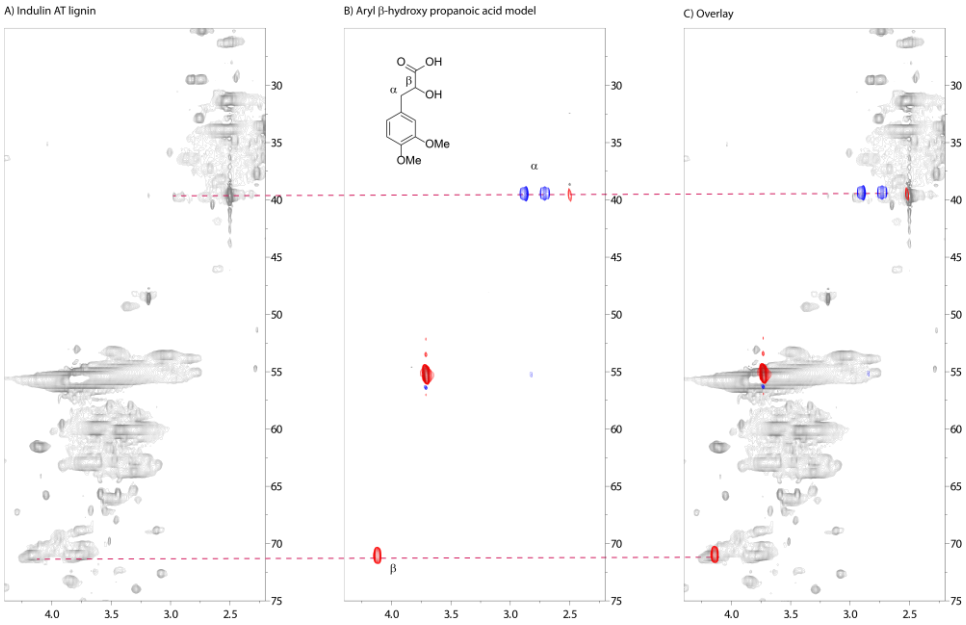


Figure S3 Assignment of aryl β-hydroxy propanoic acids in Kraft lignin, indicating lignin-carbohydrate condensations.[32]

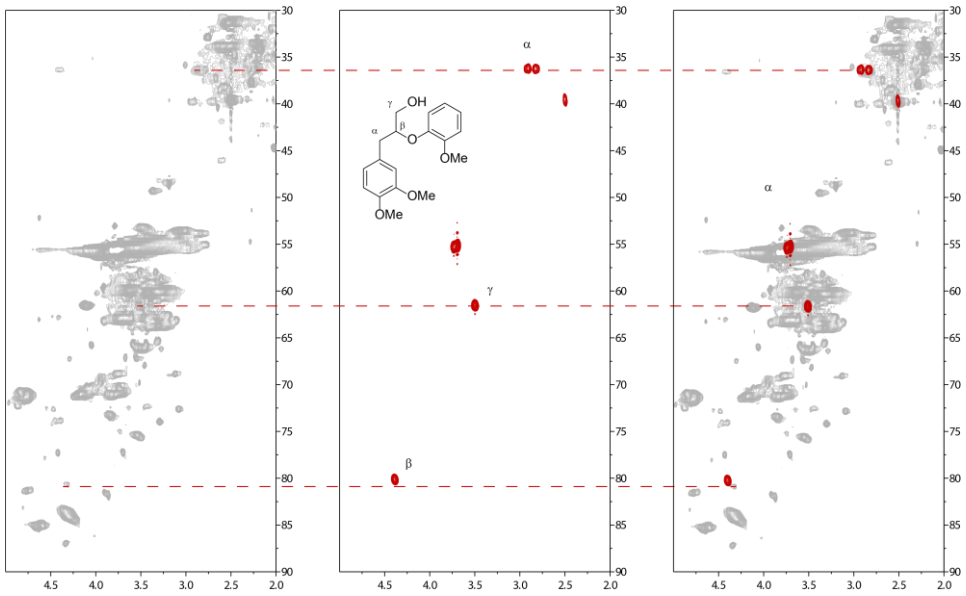


Figure S4 Assignment of reduced β-O-4 units in Kraft lignin.[32]

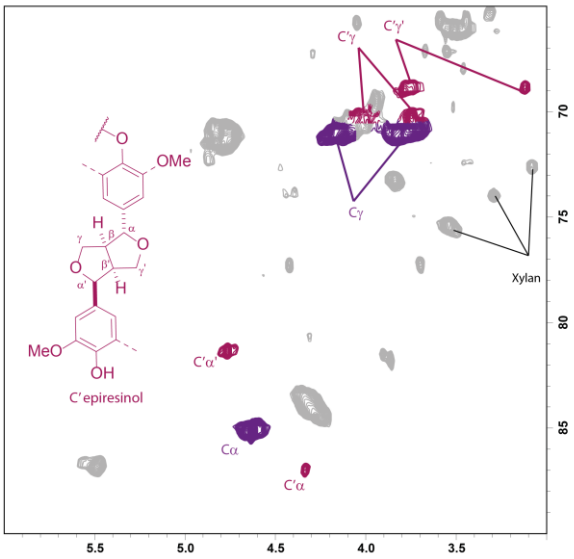


Figure S5 Assignment of epiresinols in Kraft lignin.[33,34] Epiresinols have previously been identified in hardwood Kraft lignin[35] but not in a softwood lignin.[34] This is almost certainly the result of their relatively low abundance in Kraft lignin making their observation difficult without the use of a high field cryoprobe equipped NMR spectrometer (as used here) or long experiment times). The C'α' cross peak at 4.75/81.3 ppm appears to overlap with other peaks. Based on integration of the C'α and C'α' cross peaks approximately 50% of the intensity of the peak at 4.75/81.3 ppm of the epiresinol comes from overlapped peaks. Some reports have assigned this cross peak to the α cross peak of dibenzodioxocins,[26] however the chemical shift indicates this to be a misassignment. Others have assigned this peak to arise from lignin-carbohydrate complexes (β-O-4 linkages with a sugar-benzyl ether).[25] We cannot rule this out, but given that the proposed reaction of carbohydrates with lignin derived quinone methides does not appear to proceed in aqueous environments, *in vitro* at least,[36] and the additional presence of large amounts of hydrogen sulfide nucleophiles in the Kraft reaction mixtures, formation and retention of such structures in appreciable quantities seems unlikely. Indeed, our model studies have shown that cross peaks in this region appear after Kraft treatment of β-O-4 polymers *in the absence* of carbohydrates. Additionally, analysis of fractionated lignin samples revealed that this peak is present even when there are no detectable carbohydrates (as judged by the absence of signals in the anomeric region of the HSQC). Traces of a third resinol diastereoisomer could also tentatively be assigned based on a cross peak at 4.80/83.1 ppm, which was only visible at lower contour levels (not shown).

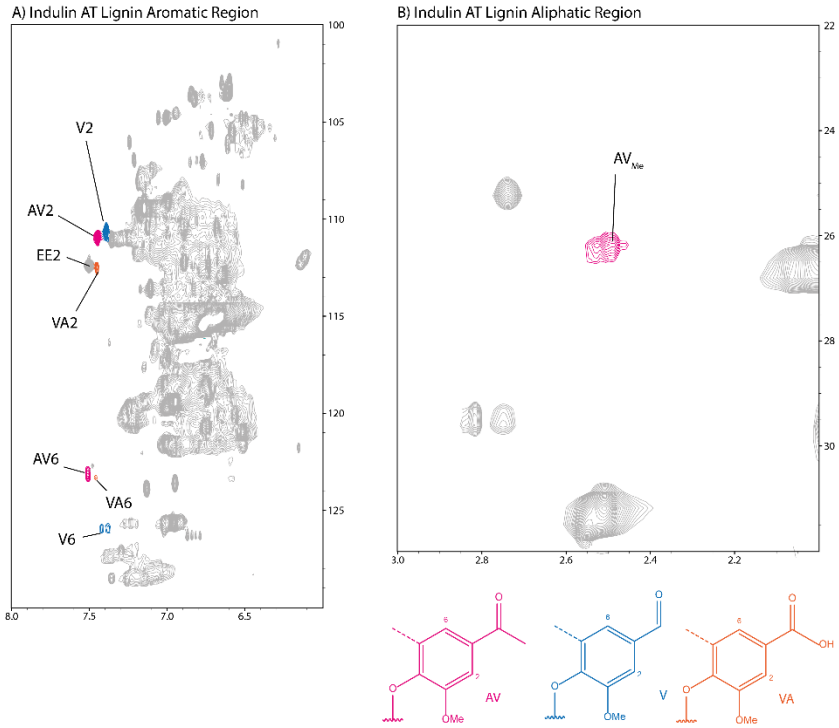


Figure S6 Assignment of vanillin, vanillic acid and acetovanillone groups in Kraft lignin. Only very small amounts of these units could be detected in this Kraft lignin.

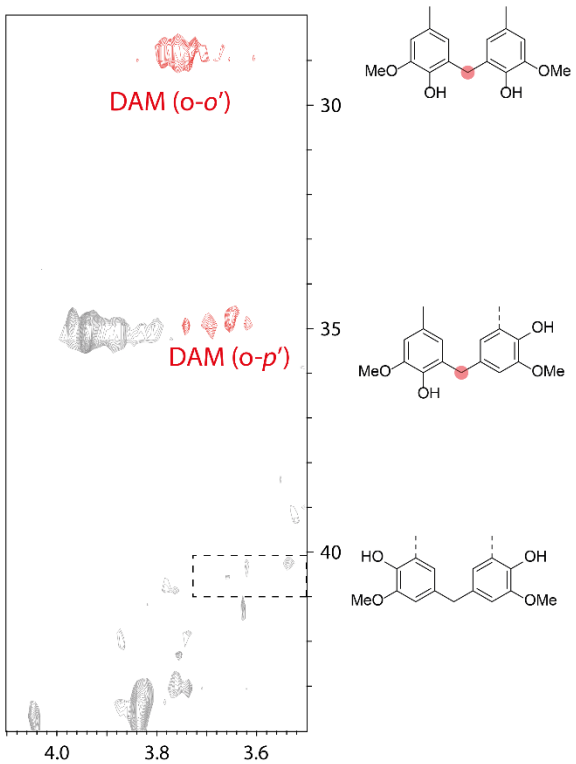


Figure S7 Assignment of diarylmethanes in Kraft lignin based on previous reports.[37,38] The dashed box in the bottom right corner of the figure indicates the region in which cross peaks for *p-p'* diarylmethane would appear.

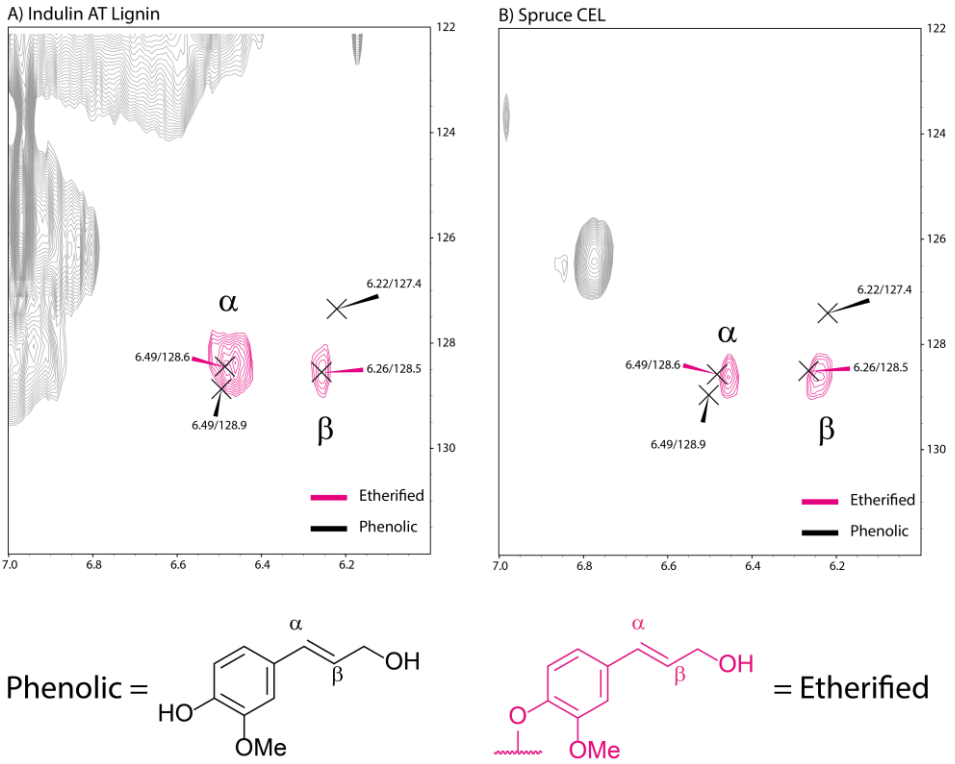


Figure S8 Assignment of etherified cinnamyl alcohol units in A) Indulin Kraft and B) spruce CEL. This shows that the major type of cinnamyl alcohols present in Kraft lignin are of the etherified type, suggesting they do not arise from the Kraft reaction of phenolic β-O-4 units, but rather are retained from the native lignin *and/or* form from etherified β-O-4 units. Assignment is based on available literature data for the appropriate model compounds.[28]

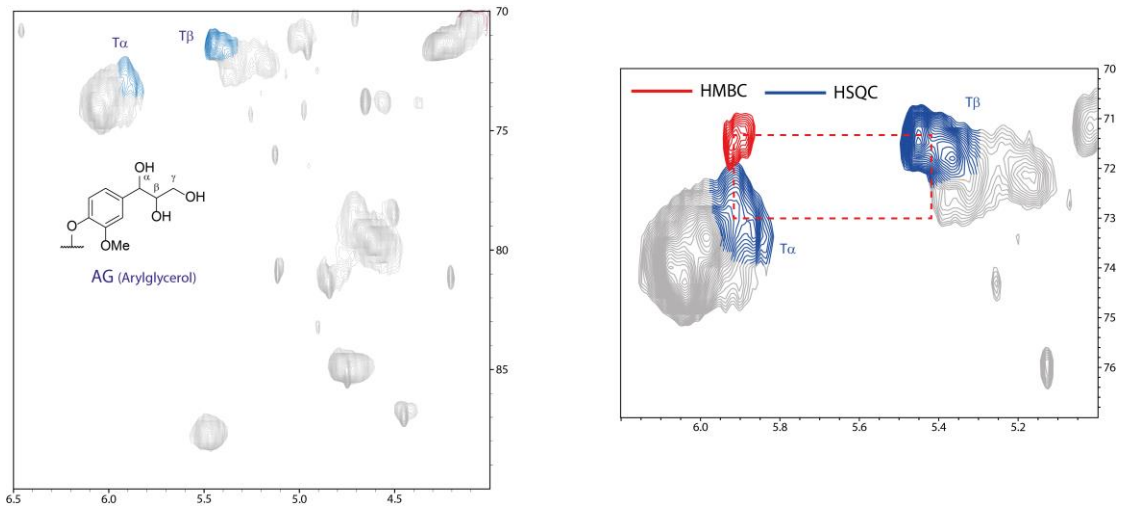


Figure S9 Assignment of arylglycerols in acetylated Kraft lignin based on literature examples.[39] Arylglycerols are also present in native lignins (2-3 per 100Ar in softwood lignin) so those in Kraft lignin are probably a mixture of retained native groups *and* newly generate units.

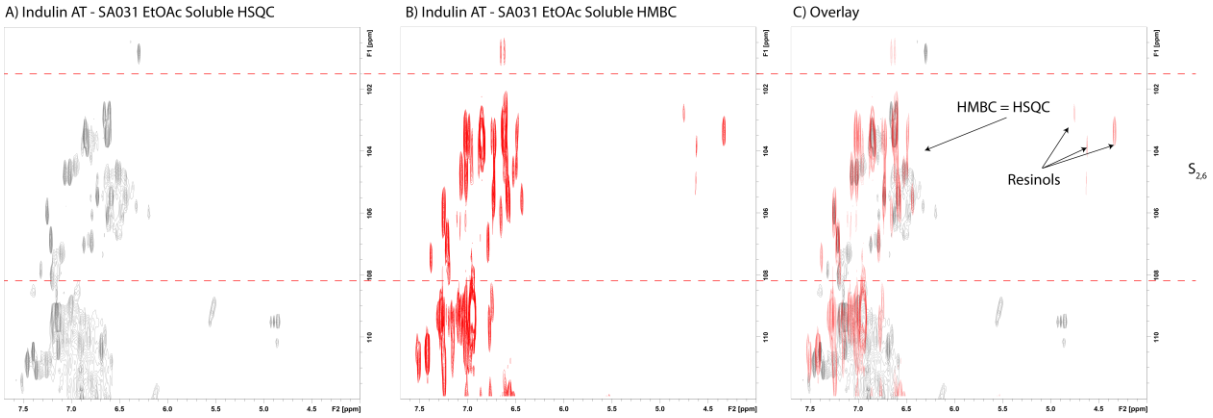


Figure S10 Assignment of syringyl units ($S_{2,6}$) in kraft lignin based on HMBC and HSQC analysis. The HMBC spectra shows the same cross peaks as the HSQC spectra for the $S_{2,6}$ units indicating symmetrical units as well as correlations to resinol units showing these are really associated with lignin. The occurrence of syringyl units in Indulin AT lignin is unusual and seems specific to this batch (SA031) presumably arising due contamination of the softwood feedstocks used for kraft pulping with some hardwoods in this case.

6. Lignin Fractionation

6.1 Fractionation Methods

Fractionation of crude Kraft lignin

Kraft lignin powder (Indulin AT, Batch No. SA031, 10 g) was stirred sequentially with 200 mL of EtOAc, 5% MeOH/EtOAc, 10% MeOH/EtOAc, 20% MeOH, 30% MeOH and 100% MeOH for 2 h each. After each extraction the solids were separated by filtration and retained for the next extraction. The soluble fraction was concentrated *in vacuo*. Yields: EtOAc – 9.7 wt%, 5.0% MeOH/EtOAc – 6.5 wt%, 10% MeOH/EtOAc – 8.9 wt%, 20% MeOH – 14 wt%, 30% MeOH – 9.0 wt%, 100% MeOH – 12 wt% and an insoluble fraction – 39 wt%.

Fractionation of acetylated Kraft lignin

Acetylation – Kraft lignin was acetylated using a 1:1 mixture of Ac_2O/pyridine (10 mL/g) at room temperature overnight. The acetylated lignin was recovered by azeotrope distillation of the Ac_2O/pyridine first with toluene (5 times), then with EtOH (5 times).

Acetylated Kraft lignin powder (6 g) was stirred sequentially with 120 mL of 10% acetone/Et_2O, 20% acetone/Et_2O, 30% acetone/Et_2O and 40% acetone/Et_2O for 1 h each. After each extraction the solids were separated by filtration and retained for the next extraction. The soluble fraction was concentrated *in vacuo*. Yields: 10% acetone/Et_2O – 27 wt%, 20% acetone/Et_2O – 15 wt%, 30% acetone/Et_2O – 16 wt%, 40% acetone/Et_2O – 15 wt% and insoluble – 28 wt%.

6.2 ^{13}C, HSQC NMR and GPC Lignin Fraction Characterization Data

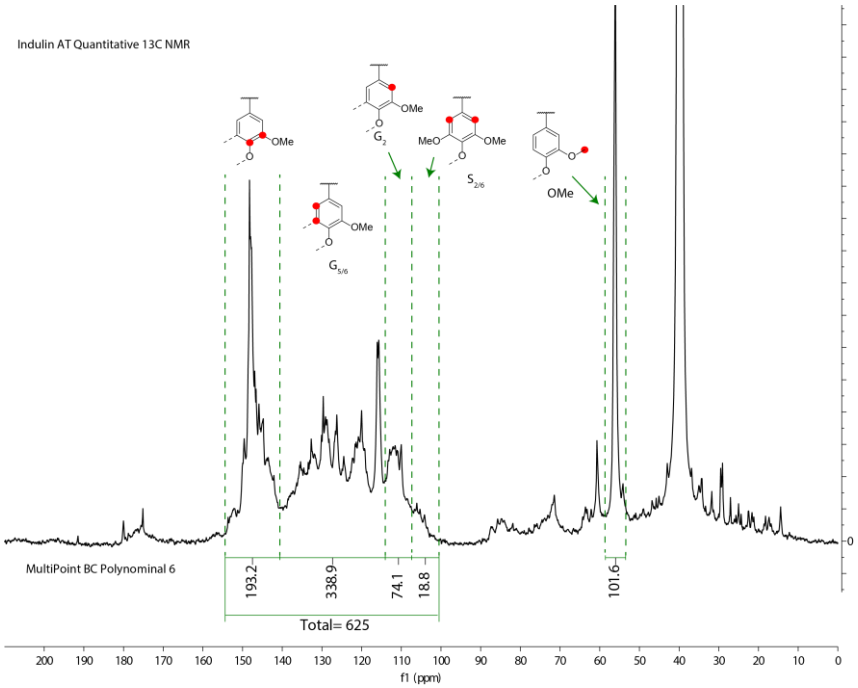


Figure S11 Quantitative ^{13}C NMR analysis of Indulin AT lignin in DMSO-d_6. 360 mg/mL, 0.01 M chromium (III) acetylacetonate, D1 = 2 s, acquisition time = 1.3107, SW = -15-234 ppm, NS = ~18000, apodization = 10 Hz, multi-point baseline correction applied in MestReNova.

Discussion and Analysis of Quantitative ^{13}C NMR Analysis of Crude Indulin AT Lignin

First the ^{13}C NMR spectra required careful baseline correction. In this particular case an interactive multi-point base line correction was applied in MestreNova using a 6th order polynomial. Small differences in baseline correction can have rather large effects on the results and the most suitable method may vary depending on different factors.

Following baseline correction the total integral for the aromatic region (100-155ppm) was set to 625. For this aromatic units were counted as 600 with an additional 25 carbons added for olefinic carbons from cinnamyl alcohols, stilbenes, enol ethers, fatty acids and other extractives. This value was estimated from HSQC analysis which is not ideal but does show that the contributions from olefinic carbons is not particularly large.

Integrals for the G_2 (107-113.8 ppm) and $S_{2/6}$ (100-107 ppm) peaks were then taken to match the spectral width used in the HSQC analysis, giving integrals of 74.1 and 18.8 respectively. These were then summed and corrected for enol ether contributions (-3.5 based on HSQC analysis) giving a total integral = 89.4. Based on a G:S ratio of 88:12 the theoretical value of this integral should be = 112.

Therefore, based on unfractionated lignin analysis, the error in using the “G2+S2/6” region is ~20%, giving an over estimation for linkage abundance in HSQC analysis.

Furthermore, analysis of the C-O and OMe regions suggests that neither radical coupling to form diaryl ethers (observed: 193, expected: 212) or demethylation (observed:102, expected: 112) are major reaction pathways, at least in the formation of Indulin AT kraft lignin.

As similar analysis of fractionated lignin samples (Figure S12) suggested that this error may be reasonably constant across the molecular weight range. For the low molecular weight 10% EtOAc/MeOH fraction the error was calculated to be ~24% and for the high molecular weight insoluble fraction the error was ~18%.

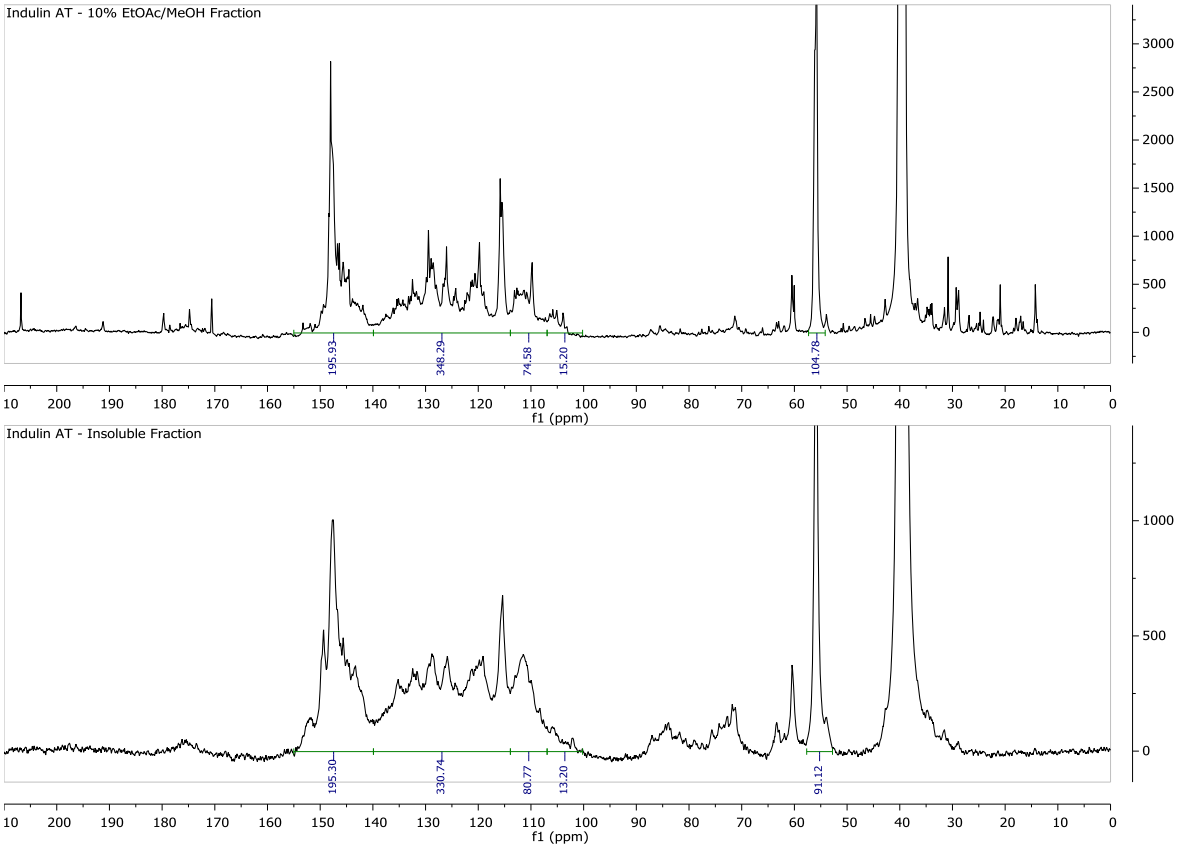


Figure S12 Quantitative ^{13}C NMR analysis of Indulin AT lignin fractions in DMSO-d_6. 360 mg/mL, 0.01 M chromium (III) acetylacetonate, D1 = 2 s, acquisition time = 1.3107, SW = -15-234 ppm, NS = ~18000, apodization = 10 Hz, 7th order polynomial baseline correction applied in MestReNova. Top: 10% EtOAc/MeOH. Bottom: MeOH Insoluble.

Table S6 GPC and 2D HSQC data, including molar masses, quantification of lignin units present in and total number of aromatic units that are accounted for in a spruce CEL, Indulin AT Kraft lignin and its fractions. Numbers in ( ) are the values calculated from HSQC0 experiments. Numbers in [ ] are standard deviations calculated from HSQC experiments run in triplicate.

Lignin	Spruce CEL	Indulin AT [a,b,c]	EtOAc	5% EtOAc/MeOH	10% EtOAc/MeOH	20% EtOAc/MeOH	30% EtOAc/MeOH	100% MeOH	Insoluble	Mw Adjusted HSQC[d]
Yield (wt%)	NA	NA	9.7	6.5	8.9	14	9	12	39	-
Mn[a]	3316	1176	631	657	793	1355	1853	2000	3396	-
Stand. Dev.[a]	ND	ND	5	0	4	49	105	126	617	-
Polydispersity (Mw/Mn)[a]	3.7	4.3	1.5	1.2	1.5	1.6	1.7	2.2	3.4	-
β–O-4	38.5	7.1 [0.25]	1.5 (2.1)	2.3 (2.9)	3.5 (4.0)	6.9 (7.7)	9.8 (11.2)	12.4 (12.6)	15.2 (15.0)	9.8
β–5	13.0	1.5 [0.00]	0.5 (0.5)	0.8 (0.7)	1.2 (0.9)	2.0 (1.9)	2.6 (3.0)	3.3 (3.0)	4.2 (3.9)	2.8
β–β (resinol)	3.0	1.6 [0.01]	1.2 (1.0)	1.3 (1.1)	1.5 (1.2)	1.8 (1.5)	2.1 (2.1)	2.3 (1.8)	2.9 (2.4)	2.1
5-5 (dibenzodioxocins)	2.4	0.2[e]	0	0	0	0	0	0	0	0
β–1 (spirodienone)	2.9	0	0	0	0	0	0	0	0	0
β–β– (secoisolariciresinol)	1.0	2.3 [0.07]	1.6 (1.4)	1.6 (1.3)	1.9 (1.2)	2.2 (1.9)	2.7 (2.2)	2.8 (2.1)	2.3 (1.8)	2.1
Dihydroconiferyl Alcohol	5.2	4.7 [0.04]	4.3 (4.1)	4.0 (4.0)	4.4 (3.5)	3.9 (3.9)	3.7 (3.9)	3.9 (4.4)	3.5 (3.5)	4.6
Coniferyl Alcohol	5.8	1.8 [0.02]	0.8 (0.9)	0.9 (0.9)	1.1 (1.0)	1.3 (1.1)	1.6 (1.6)	1.8 (1.6)	2.4 (2.1)	1.7
Coniferyl Aldehyde	3.7	0	0	0	0	0	0	0	0	0
epi-resinol	0	0.5 [0.02]	0.7 (0.5)	0.8 (0.6)	0.8 (1.2)	0.9	0.7	0.5	0.6	0.6
***Z*-enol ether**	0	1.1 [0.02]	0.8 (0.9)	1.0 (1.0)	1.0 (1.0)	1.1 (1.1)	1.0 (1.1)	1.1 (1.3)	1.2 (1.0)	1.0
***E*-enol ether**	0	2.8 [0.06]	1.4 (1.4)	1.9 (2.3)	2.0 (2.1)	2.6 (3.3)	2.9 (3.6)	2.7 (3.7)	3.1 (3.6)	2.1
***trans* β–1 stilbene**	0	3.0 [0.09]	7.9 (6.9)	7.2 (7.2)	4.9 (4.2)	2.0 (1.7)	0.8 (1.3)	0.5	0.4	1.5
***trans* β–5 stilbene**	0	6.8 [0.06]	8.2 (10.9)	6.9 (11.0)	6.7 (8.7)	5.2 (5.7)	3.8 (5.45)	3.9	2.9	3.8
Total Aromatic Groups (%)	80.0	40.8	40.3	39.6	38.1	36.8	38.0	41.3	44.9	38.4

[a] Calculated from 3 repeats [b]± standard error. ND – not determined. [c] These are separate experiments (TD (F1) = 128) to that used to prepared Table 1 and Fig. 2 in the manuscript (TD (F1) = 384). [d] calculated as the sum of (fraction yield x individual unit abundance) across all fractions. [e] Determined from high resolution experiment (TD (F1) = 384). Not detected using low resolution experiments.

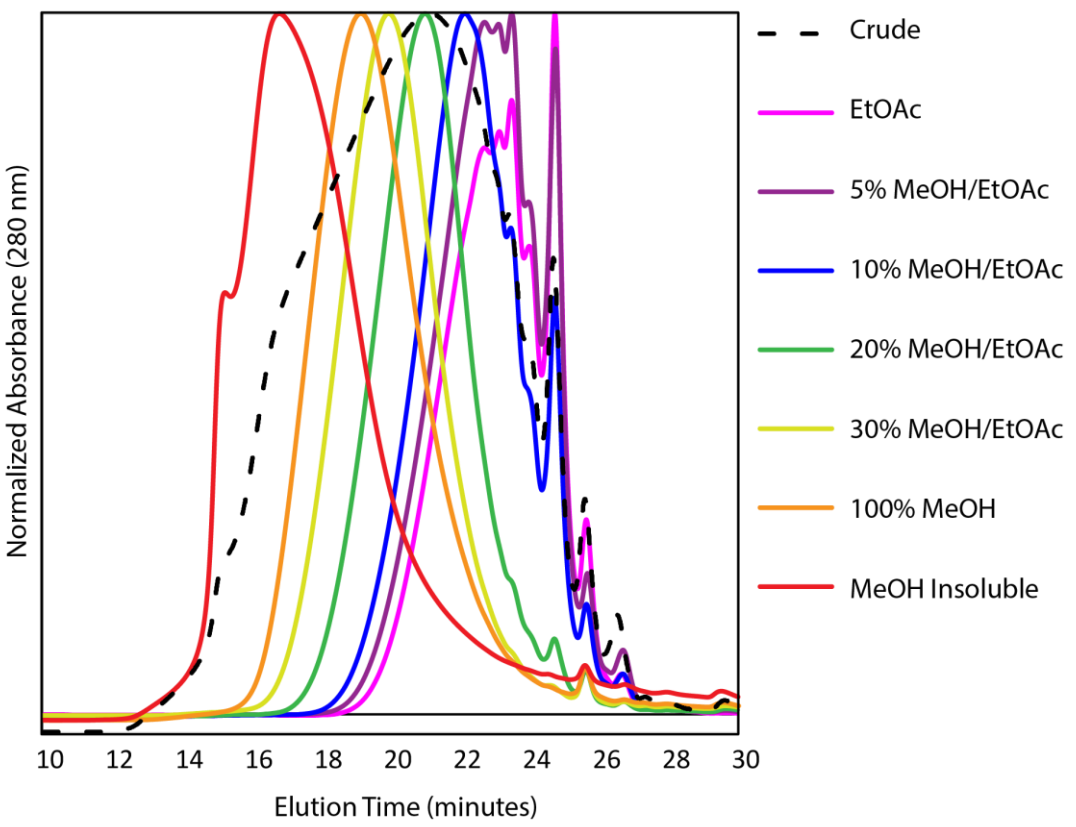


Figure S13 GPC elution profiles of the fractionated Indulin AT kraft lignin samples.

Table S7 GPC and 2D HSQC quantification of side chains present in acetylated Indulin AT Kraft lignin fractions.

Lignin	10% Acetone/Et_2O	20% Acetone/Et_2O	30% Acetone/Et_2O	40% Acetone/Et_2O	Insoluble	Mw Adjusted HSQC
Yield (wt%)	27.1	14.7	15.9	14.6	27.7	-
Mn[a]	666	1177	2074	3092	4756	-
Stand. Dev.[a]	8	6	44	83	904	-
β–O-4	5.0	7.1	10.2	13.1	15.9	10.3
β–5	0.73	1.3	2.5	3.3	4.0	3.8
β–β (resinol)	0.88	1.3	2.1	2.6	2.7	2.4
β–β– (secoisolariciresinol)	1.7	1.9	2.2	2.5	0.67	1.6
Coniferyl Alcohol	1.0	1.1	1.5	1.9	2.4	1.6
Dihydroconiferyl Alcohol	4.1	4.1	3.3	3.6	2.0	3.3
epi-resinol	0.53	0.61	0.89	1.13	0.88	0.8
***E*-enol ether**	2.1	2.4	3.2	4.0	4.0	3.1
***Z*-enol ether**	0.95	0.89	1.0	1.2	1.4	1.1
Arylglycerol	2.3	2.5	3.6	4.8	5.4	3.8

[a] Calculated from 3 repeats of the GPC measurement.

Table S8 Experimental conditions used for $HSQC_0$ experiments. D1 = 7.5 s in all experiments which was at least 5 x the longest T1 value. Fractions 10% EtOAc/MeOH and 20% EtOAc/MeOH were used to evaluate the difference between 2- and 3-point $HSQC_0$ experiments (Table S9)

Lignin Fraction	NS	SW F1	o2p	TD (F1)	approx. Expt time/hours 2-increment (3-increment)
Crude	40	90	92	72	12.2
EtOAc	40	90	92	72	12.2
5% EtOAc/MeOH	44	90	92	72	13.4
10% EtOAc/MeOH	120	90	92	72	36.5 (55)
20% EtOAc/MeOH	80	40	70	32	10.8 (16.3)
30% EtOAc/MeOH	80	40	70	32	10.8
100 %MeOH	80	40	70	32	10.8
Insoluble	80	40	70	32	10.8

NS = Number of scans. SW F1 = sweep width in F1 dimension. o2p = spectrum center in F1. TD (F1) = number of increments collected in F1.

Table S9 Comparison of quantification values obtained from 2-point and 3-point $HSQC_0$ experiments. The generally close agreement between the 2 data sets, together with the challenges in obtaining spectra with sufficient S/N for the 3rd $HSQC_0$ increment led us to focus on the use of the 2-point experiment. ND = not determined due to insufficient signal in the 3rd spectrum.

	A	B	C	X	SR	DHCA	C'	EE	SB1	SB5
10% EtOAC/MeOH 2-point $HSQC_0$	4.0	0.9	1.2	1.0	1.2	3.5	0.6	3.1	4.2	8.7
10% EtOAC/MeOH 3-point $HSQC_0$	3.9	1.0	1.2	1.1	1.0	3.1	0.6	3.0	4.5	8.1
Difference (2-point minus 3-point)	0.1	-0.1	0.0	-0.1	0.2	0.4	0.0	0.1	-0.3	0.6
20% EtOAC/MeOH 2-point $HSQC_0$	7.7	1.9	1.5	1.1	1.9	3.9	1.2	4.4	1.7	5.7
20% EtOAC/MeOH 3-point $HSQC_0$	7.8	1.9	1.8	1.3	1.8	4.1	1.5	4.3	ND	ND
Difference (2-point minus 3-point)	-0.1	0.0	-0.3	-0.2	0.1	-0.2	-0.3	0.1	-	-

6.3 Discussion of HSQC Analysis of Lignin Fractions

As outlined above Indulin AT Kraft lignin was fractionated based on changing solubility in solvent mixtures of different polarity, similar to previous reports[25,26,40,41] with the fractions being analyzed by GPC and HSQC.

From analysis of a spruce CEL (M_n = 3316) and the crude Indulin AT lignin (M_n = 1176) we had already observed that the molecular weight of the Kraft lignin was significantly reduced relative to the native lignin (which has already undergone some depolymerization during extraction). This led us to conclude that depolymerization reactions outweigh recondensation reactions during the Kraft pulping process. Accordingly, low molecular weight Kraft lignin fractions should be enriched in Kraft chemistry derived products and depleted in native lignin structures, with the reverse being true of the high molecular weight fractions. Indeed, we found that the abundances of lignin units varied greatly depending on the molecular weight and proving very insightful in terms of elucidating the chemistry occurring during the process, as discussed below.

β–O-4 and β–5 units: Native β–O-4 and β–5 units displayed a very clear positive correlation with molecular weight as shown in Figures S14 and S15, consistent with these units being degraded during the Kraft process as the lignin is depolymerized and new phenolic groups are released. Even in the most degraded (lowest Mw) fraction some of these units remain, though, possibly indicating that some of these are part of Kraft resistant linkages, for example those involved in 4-O-5 linkages or ones adjacent to enol ethers as both such structures would prevent the formation of the phenolic group required for degradation of the β–O-4 and β–5 linkages. Quantitative $HSQC_0$ gave very similar results to the standard HSQC experiment, as also illustrated in Figures S14 and S15.

Interestingly, when normalized according to abundance, these two units were found to have very similar molecular weight profiles (Figure S16), indicating that they have similar distributions in the native lignin polymer and similar sensitivities towards the pulping conditions.

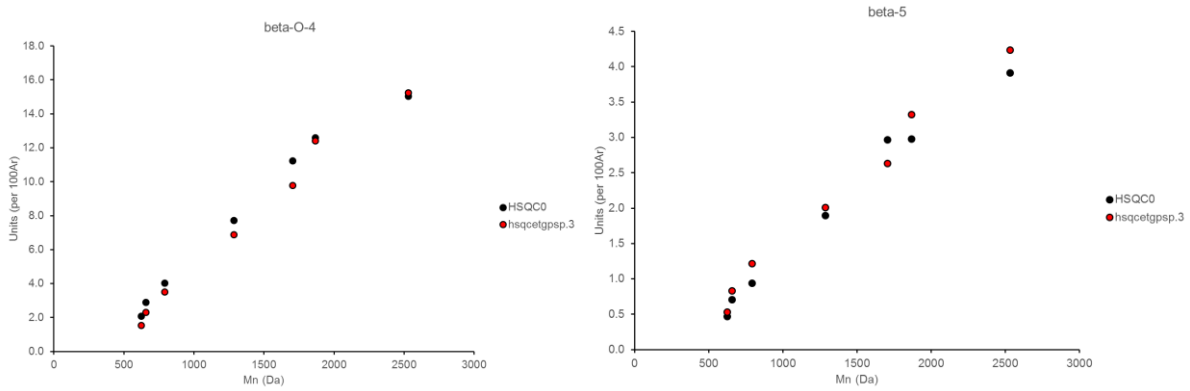


Figure S14 and Figure S15 Left: β-O-4 and Right: β-5 unit abundance as determined by standard HSQC (red) and quantitative HSQC0 (black) NMR methods against GPC determined molecular weight.

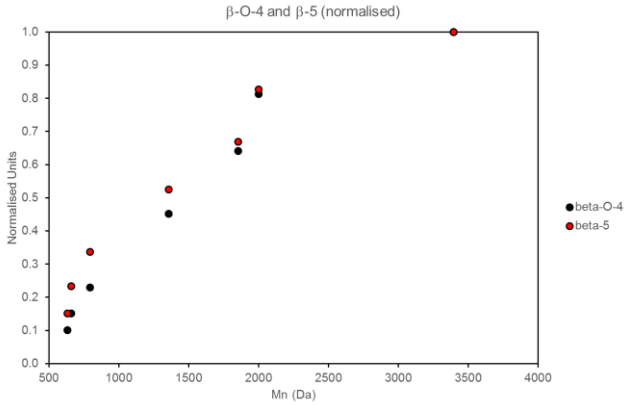


Figure S16 Normalised β-O-4 and β-5 lignin unit abundance against GPC determined molecular weight.

β–β units: This class of linkages consists of resinols, epiresinols and secoisolariciresinols and were found to have a more complex distribution (Figure S17-19). Natural resinol units (Figure S17) showed a similar trend to that observed for β–O-4 and β–5 units, i.e. they increase in abundance as molecular weight increased. This is again in line with the hypothesis that degradation of these units during the Kraft process also requires a free phenolic group. $HSQC_0$ experiments gave overall similar results to the standard HSQC method, with a slight overestimation of resinols in the standard HSQC experiment being noted.

Epimerized resinol units were found to initially increase in abundance (Figure S18), but then decrease in the higher molecular weight fractions. This pattern likely reflects at least two competing factors; i) epimerization of the resinol unit requires at least one free phenolic group to be present which should favor their increased abundance in the low molecular weight fractions and ii) resinol units are known to be poorly released from softwood lignins during degradative analyses indicating that they are, for

the most part (but not exclusively), incorporated into the lignin polymer through chemically resistant bonds,[42] and should therefore be associated with higher molecular weight fractions even after depolymerization. Furthermore, as resinols are incorporated early in lignin biosynthesis they should be quite remote from the (reactive) phenolic ends of the polymer, thus protecting them from degradation relative to other units. Due to their low abundance leading to poor signal to noise ratios, epiresinols were not included in this $HSQC_0$ analysis.

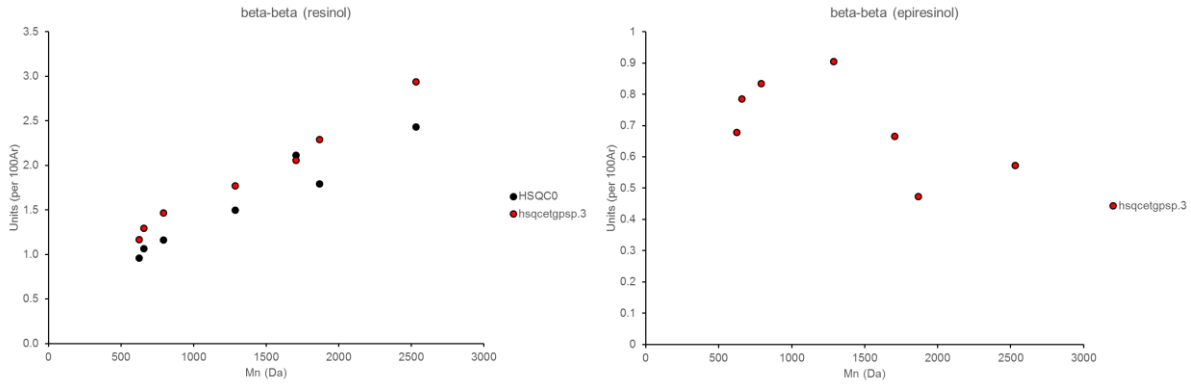


Figure S17 and Figure S18 Left: resinol and Right: epiresinol unit abundance as determined by standard HSQC (red) and quantitative $HSQC_0$ (black) against GPC determined molecular weight.

The abundance of secoisolariciresinols (SR), found to be the second major type of β–β coupled linkage in Kraft lignin, showed a similar trend (Figure S19) and is present in similar abundance to natural resinols, except in the highest molecular weight fraction where their abundance fell. This last point is somewhat surprising given the structural similarly of both units, however it is probably the result of the natural distribution of these units in the lignin. Indeed, the propensity of these units to be released by degradative analysis methods such as DFRC is markedly different[43] suggesting substantially different coupling pathways are preferred for their incorporation into the lignin polymer in each case. Quantitative $HSQC_0$ experiments gave a similar overall trend as the standard HSQC method but did indicate a quite significant overestimation of these units by the standard method. As errors caused by differences in *J* couplings are typically quite small, this observation may be the result of a relaxation effect, although we have not confirmed this.

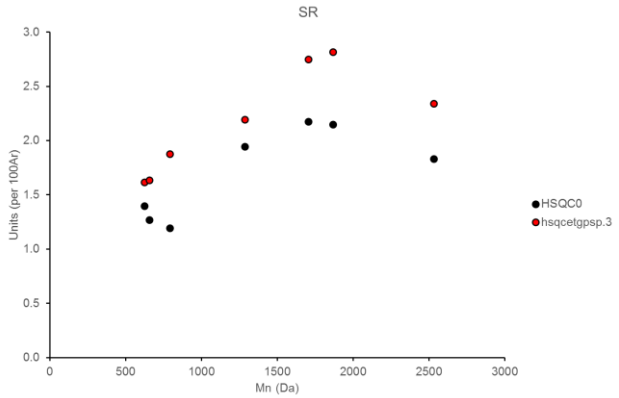


Figure S19 Secoisolariciresinol unit abundance as determined by standard HSQC (red) and quantitative HSQC0 (black) against GPC determined molecular weight.

Enol ethers: The presence of enol ethers was found to increase with molecular weight (Figure S20), consistent with their anticipated role in blocking further lignin depolymerization. Additionally, the *E:Z* ratio appeared to change slightly with molecular weight, with the *E* isomer always being the most abundant (Figure S21).

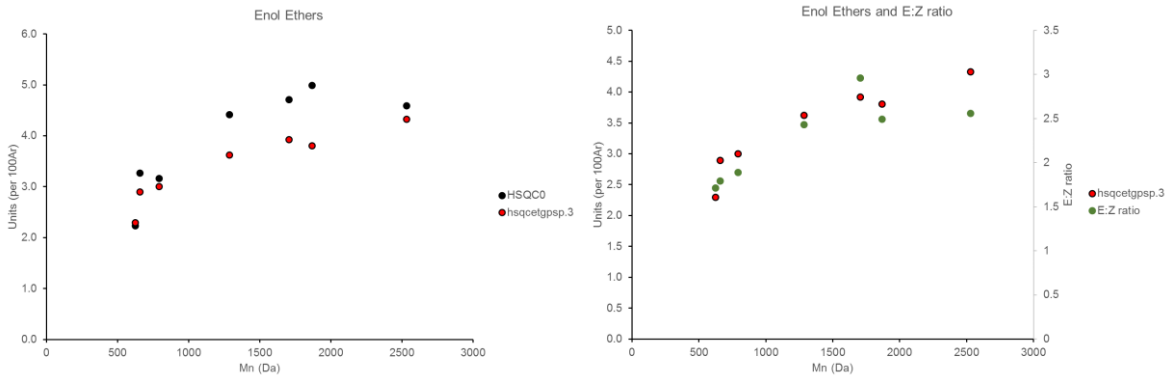


Figure S20 and Figure S21 Left: Enol ether unit abundance as determined by standard HSQC (red) and quantitative $HSQC_0$ (black) against GPC determined molecular weight. Right: Enol ether unit abundance and the *E:Z* ratio against GPC determined molecular weight.

β-5 and β-1 stilbenes: Both stilbenes were found to be most abundant in the lowest molecular weight lignin fractions (Figures S22 and S23), consistent with their formation during the Kraft process and their relatively high stability under the Kraft pulping conditions. Interestingly, the abundance of β-1-derived stilbenes fell much more rapidly than the β-5-derived stilbenes indicating that these stilbenes must be more efficiently formed and released from the lignin polymer than β-5 lignin units. A similar

observation has been made during thioacidolysis studies were β–1 derived stilbenes are over-represented in the dimeric fractions given their abundance in natural lignins.[42] One explanation put forward for this has been that because β–1 units can only be etherified at one phenolic end during lignification and the chances of this end being 4-O-etherified are high, the probability of releasing a dimeric unit are consequently higher than for other units which require two suitably cleavable linkages to be present.[42] Due to the poorer resolution and S/N provided by $HSQC_0$ experiments quantification of these stilbenes was only performed for the four lowest molecular weight fractions. This suggested that the abundances of the β-5-derived stilbenes was underestimated in the standard HSQC method whilst β-1-derived stilbenes were slightly overestimated. In both cases, however, the same trend in unit abundances is observed using both NMR methods.

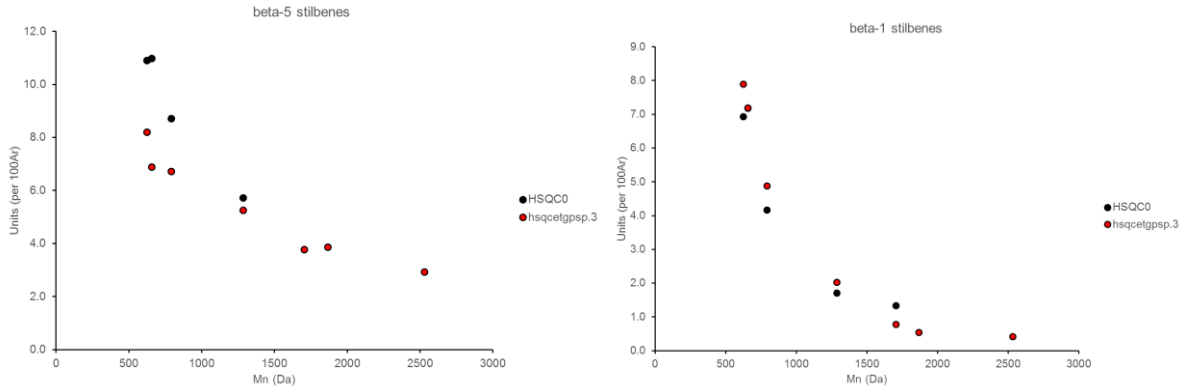


Figure S22 and **Figure S23** Left: β-5 stilbene and Right: β-1 stilbene unit abundance as determined by standard HSQC (red) and quantitative $HSQC_0$ (black) against GPC determined molecular weight.

Cinnamyl alcohol end groups: The abundance of these units were found to increase with molecular weight (Figure S24). Although coniferyl alcohol is a known intermediate in the Kraft process,[44,45] formed by degradation of phenolic β-O-4 units, most of the cinnamyl alcohol groups observed in Kraft lignin are instead likely to be surviving end groups from the native lignin or are formed from non-phenolic β-O-4 units during the Kraft process (See Figure S33) rather than being from degradation of phenolic β-O-4 units. Indeed, HSQC analysis indicates that these groups are etherified rather than free phenolic, suggesting that this is in fact the case (Figure S8). This is also consistent with the high reactivity of phenolic coniferyl alcohol groups compared to etherified cinnamyl alcohols under pulping conditions.[46] Quantitative $HSQC_0$ gave very similar results to the standard HSQC experiment in this case.

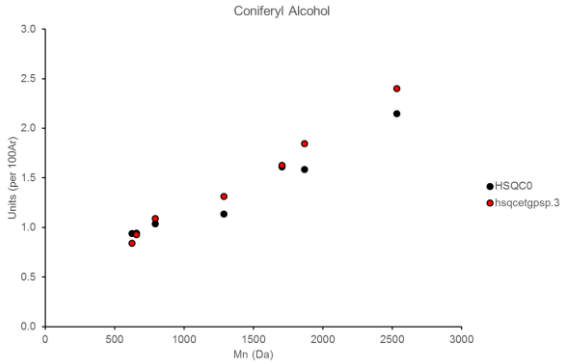


Figure S24 Cinnamyl alcohol unit abundance as determined by standard HSQC (red) and quantitative $HSQC_0$ (black) against GPC determined molecular weight.

Dihydrocinnamyl alcohol end groups: These end groups were found to have a relatively similar abundance in all molecular weight fractions (Figure S25). This is probably the result of their stability under Kraft conditions and we found no evidence in our model studies that such units form from either coniferyl alcohol or β-O-4 linkages (by reduction of intermediates). Quantitative $HSQC_0$ experiments showed that these units are overestimated in the standard HSQC experiment.

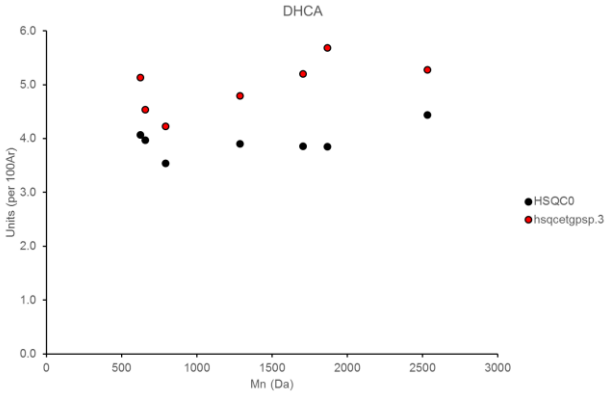


Figure S25 Dihydrocinnamyl alcohol unit abundance as determined by standard HSQC (red) and quantitative $HSQC_0$ (black) against GPC determined molecular weight.

Arylglycerol end groups: The abundance of these units could best be determined in the acetylated lignins. A clear positive correlation of the abundance of these units with molecular weight is seen (Figure S26, Table S7), consistent with them being associated with etherified units, similar to enol ethers. No $HSQC_0$ analysis was conducted on these samples.

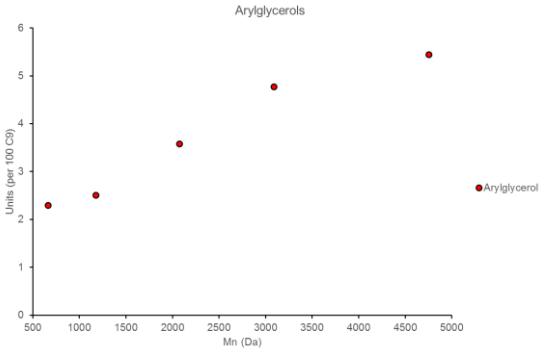


Figure S26 Arylglycerol unit abundance as determined by standard HSQC (red) against GPC determined molecular weight.

Table S10 Relative volume integral data obtained for Kraft lignin model compounds from HSQC analysis. Small errors are observed in most cases when using the standard HSQC method indicating that difference in *J*-couplings and relaxation profiles have minor effects. Some errors are seen for stilbenes and cinnamyl alcohols, which are significantly reduced using the $HSQC_0$ method. Bold indicates the peaks used for quantification in lignin. [a] obtained using the hsqcetgpsp.3 pulse sequence and a D1 of 1 s. [b] obtained using the $HSQC_0$ method (2 points) with a D1 of 5 s. Data were processed identically as for lignin analyses. [c] HSQC-COSY peaks were observed for and between the α/β cross peaks, but were not included in the integration.

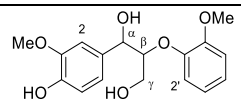
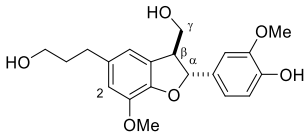
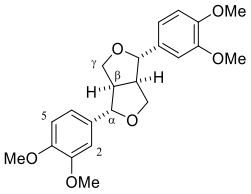
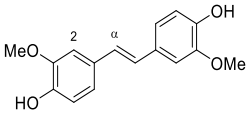
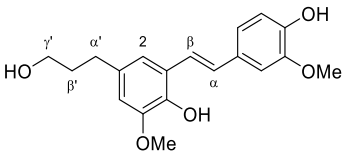
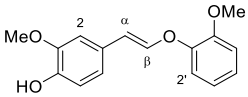
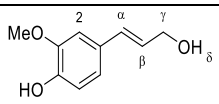
Model	Signal	HSQC Integral[a]	$HSQC_0$ Integral[b]
	G2+G2′ (2H)	200	-
	α	**103**	-
	G2	100	-
	α	**101**	-
	G2+G5 (2H)	200	-
	α	**107**	-
	β	98/99	-
	γ	88	-
	G2	100	100
	α	**79**	**105**
	G2	100	100
	β	**77[c]**	**89**
	α' (2H)	177	196
	β' (2H)	180	189
	γ' (2H)	195	202
	G2	100	-
	α	**98**	-
	β	78	-
	G2	100	100
	γ (2H)	**174**	**207**

Table S11 $HSQC_0$ derived signal attenuation factors for lignin units in different molecular weight fractions of Indulin AT lignin. ND – not determined.

		$HSQC_0$ Signal Attenuation Factors												
Lignin Fraction/Unit	**Mw**	**G2**	**S2/6**	**SB1**	**SB5**	**B**	**A**	**C**	**C'**	**DHCA**	**SR**	**EE**	**EE**	**X**
EtOAc	624	-1.10	-1.12	-1.55	-1.43	-1.04	-1.17	-0.96	-0.98	-1.40	-1.68	-1.02	-1.06	-1.13
5% EtOAc/MeOH	658	-1.12	-1.11	-1.68	-1.46	-1.12	-1.16	-1.03	-1.12	-1.46	-1.39	-1.05	-1.32	-1.13
10% EtOAc/MeOH	791	-1.12	-1.10	-1.50	-1.45	-1.07	-1.16	-0.99	-1.04	-1.38	-1.39	-1.02	-1.13	-1.06
20% EtOAc/MeOH	1286	-1.46	-1.44	-1.78	-1.67	-1.31	-1.33	-1.11	-1.37	-1.36	-1.55	-1.38	-1.50	-1.02
30% EtOAc/MeOH	1706	-1.45	-1.49	-2.01	-1.74	-1.35	-1.32	-1.25	-1.38	-1.32	-1.53	-1.22	-1.45	-1.17
100 %MeOH	1868	-1.57	-1.74	ND	ND	-1.41	-1.45	-1.28	-1.56	-1.42	-1.66	-1.48	-1.58	-1.28
Insoluble	2531	-1.66	-1.80	ND	ND	-1.52	-1.53	-1.47	ND	-1.51	-1.68	-1.71	-1.38	-1.39

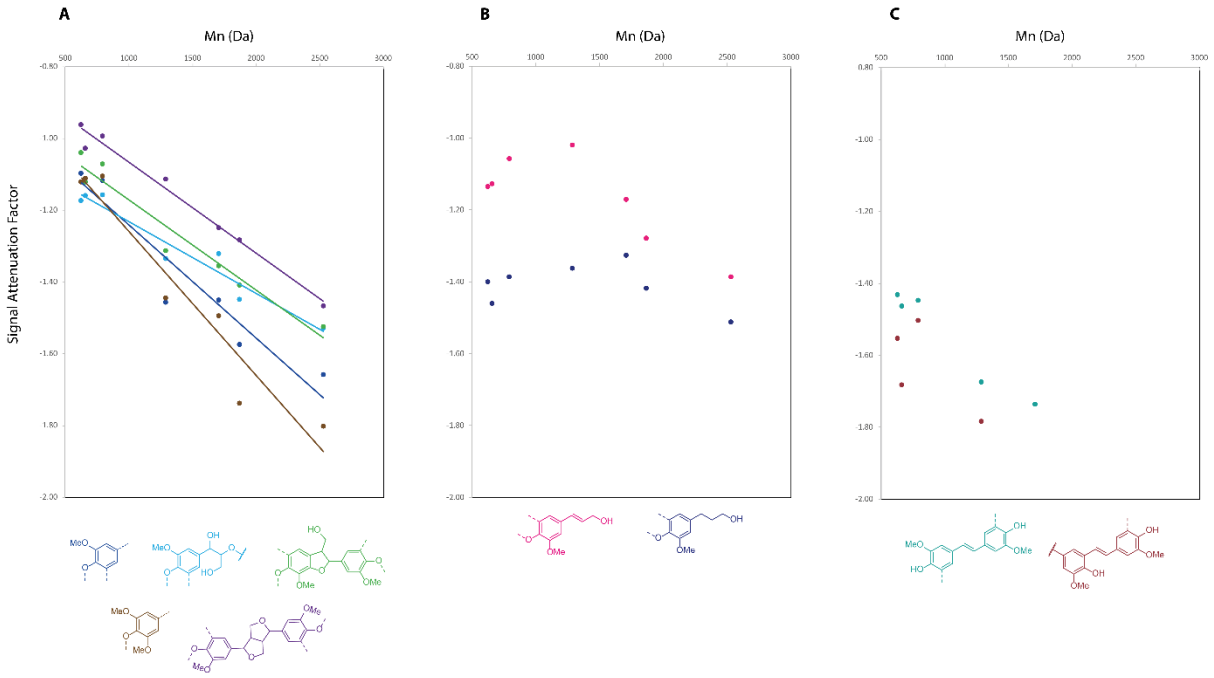


Figure S27 Plots showing the correlation of $HSQC_0$ derived signal attenuation factors with molecular weight for different lignin units. **A,** G, S, β-O-4, β-5 and β-β appear to show an almost linear correlation. **B**, Coniferyl alcohol and dihydroconiferyl alcohol groups have non-linear correlations. **C**, The signal attenuation factors for the stilbenes do appear to correlate with molecular weight, but have been plotted separately given the limited amount of data points.

7. ^{31}P NMR Fractionation Data

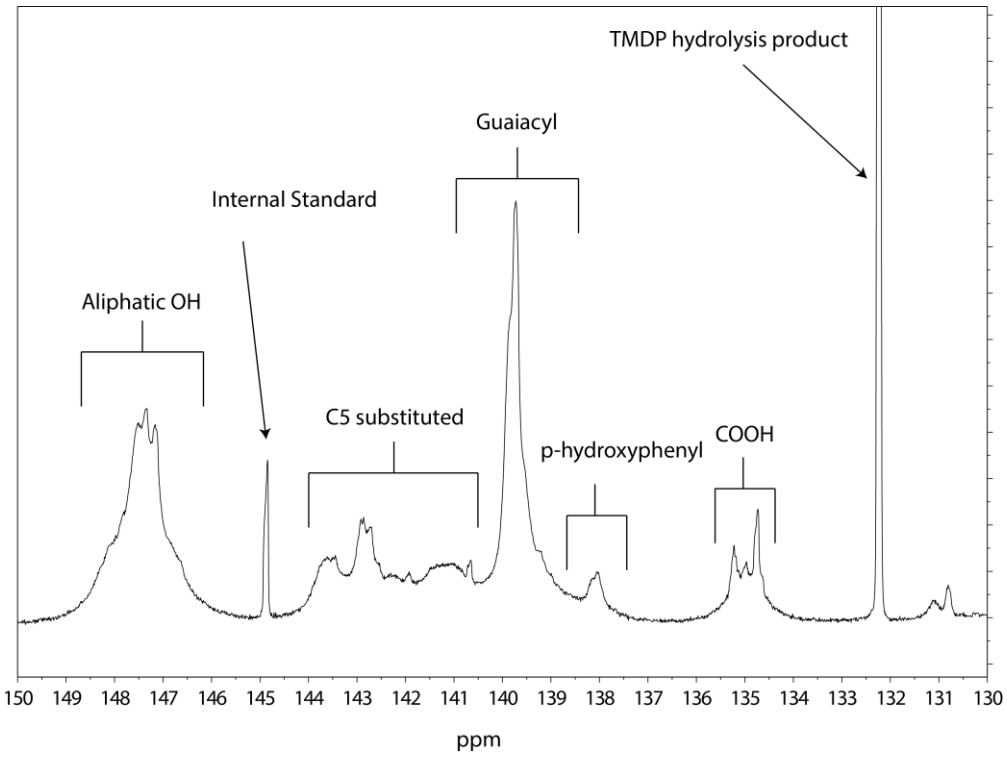


Figure S28 ^{31}P NMR analysis of unfractionated, phosphitylated Indulin AT lignin showing the presence of and regions used for the quantification of aliphatic OH groups, C5 substituted guaiacyl/syringyl, unsubstituted guaiacyl and *p*-hydroxyphenol groups, and carboxylic acids in fractionated samples.

Table S12 Hydroxyl group content of Kraft lignin fractions quantified by ^{31}P NMR using phosphitylated cholesterol as an internal standard and expressed as mmol per g lignin. All measurements were performed in triplicate.

		Units (mmol/g)					Standard Deviation					
Lignin	**Mn**	**Aliphatic**	**Condensed**	**Guaiacyl**	**H**	**Acids**	**Mn**	**Aliphatic**	**Condensed**	**Guaiacyl**	**H**	**Acids**
EtOAc	631	0.96	1.45	2.16	0.18	0.73	5.4	0.024	0.039	0.074	0.010	0.025
MeOH 5%	657	1.28	1.71	2.45	0.23	0.57	0.5	0.035	0.089	0.152	0.022	0.047
MeOH 10%	793	1.39	1.71	2.22	0.22	0.50	4.0	0.051	0.066	0.095	0.017	0.024
MeOH 20%	1355	1.51	1.57	1.71	0.19	0.40	49	0.017	0.021	0.039	0.012	0.008
MeOH 30%	1853	1.92	1.51	1.62	0.20	0.40	105	0.034	0.028	0.030	0.014	0.017
MeOH	2000	1.96	1.45	1.33	0.15	0.31	126	0.070	0.038	0.016	0.044	0.077
Insoluble	3396	2.53	1.05	1.07	0.09	0.13	617	0.099	0.040	0.044	0.006	0.007

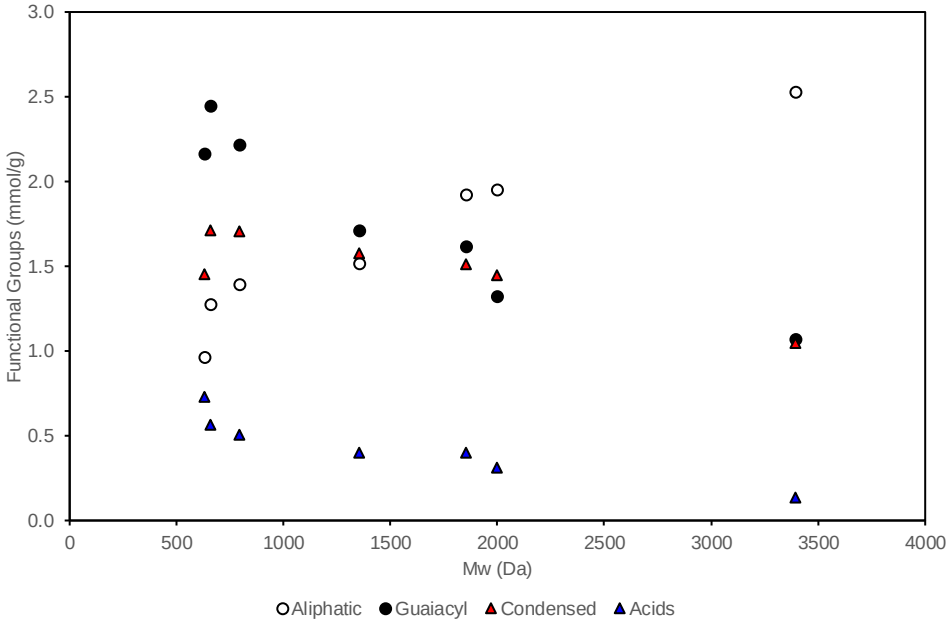


Figure S29 Hydroxyl, phenol and acid content of Indulin Kraft lignin as a function of molecular weight.

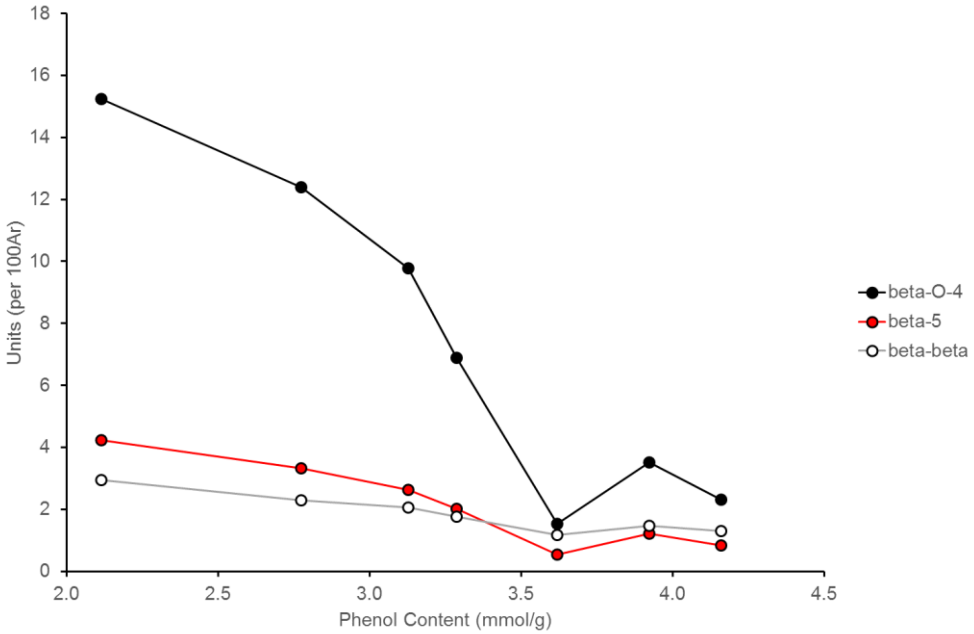


Figure S30 Native lignin unit abundance in Indulin Kraft lignin fractions against total phenol hydroxy group content.

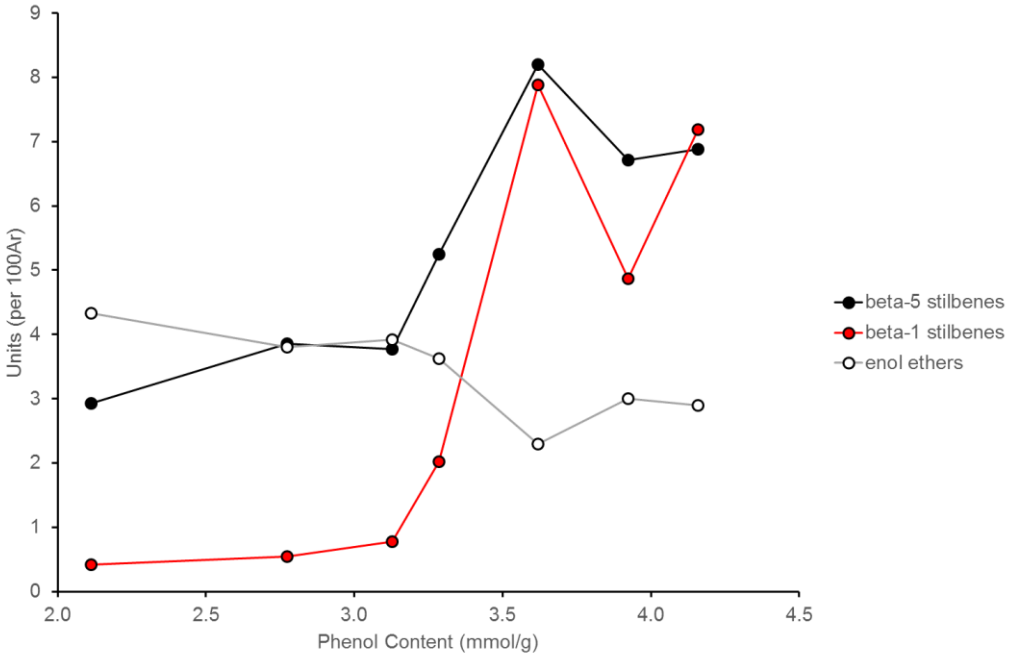


Figure S31 Kraft derived lignin unit abundance in Indulin Kraft lignin fractions against total phenol hydroxy group content.

8. Additional NMR Spectra

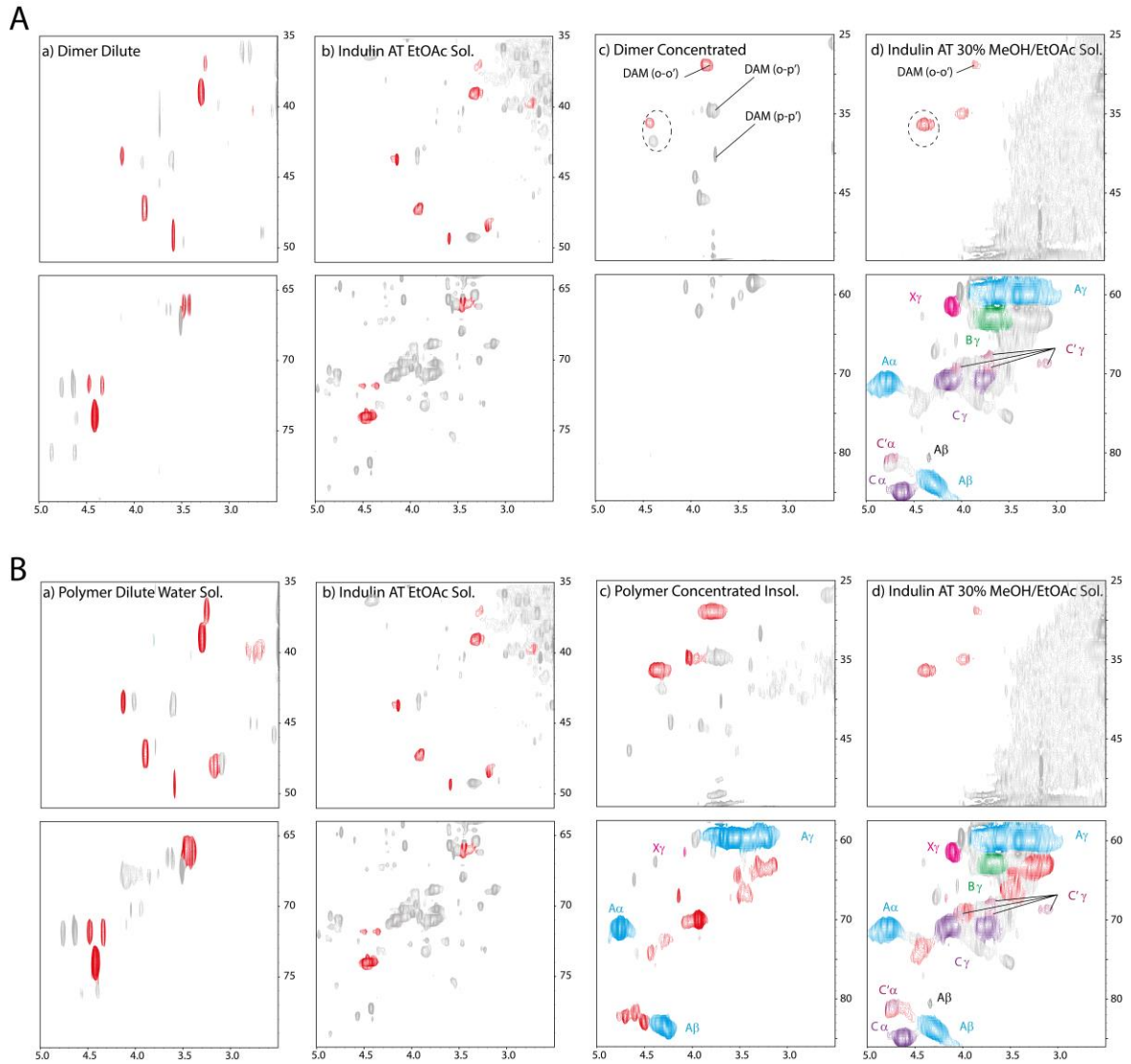


Figure S32 Comparison of A) the synthetic Kraft products obtained from dimer **3** under dilute (a) and concentrated (c) reaction conditions with the EtOAc soluble (b) and 30% MeOH/EtOAc (d) fraction of Indulin AT Kraft lignin; B) the synthetic Kraft products obtained from β-O-4 polymer under dilute (a) and concentrated (c) reaction conditions with the EtOAc soluble (b) and 30% MeOH/EtOAc (d) fraction of Indulin AT Kraft lignin. Color coding matches that used in the main text.

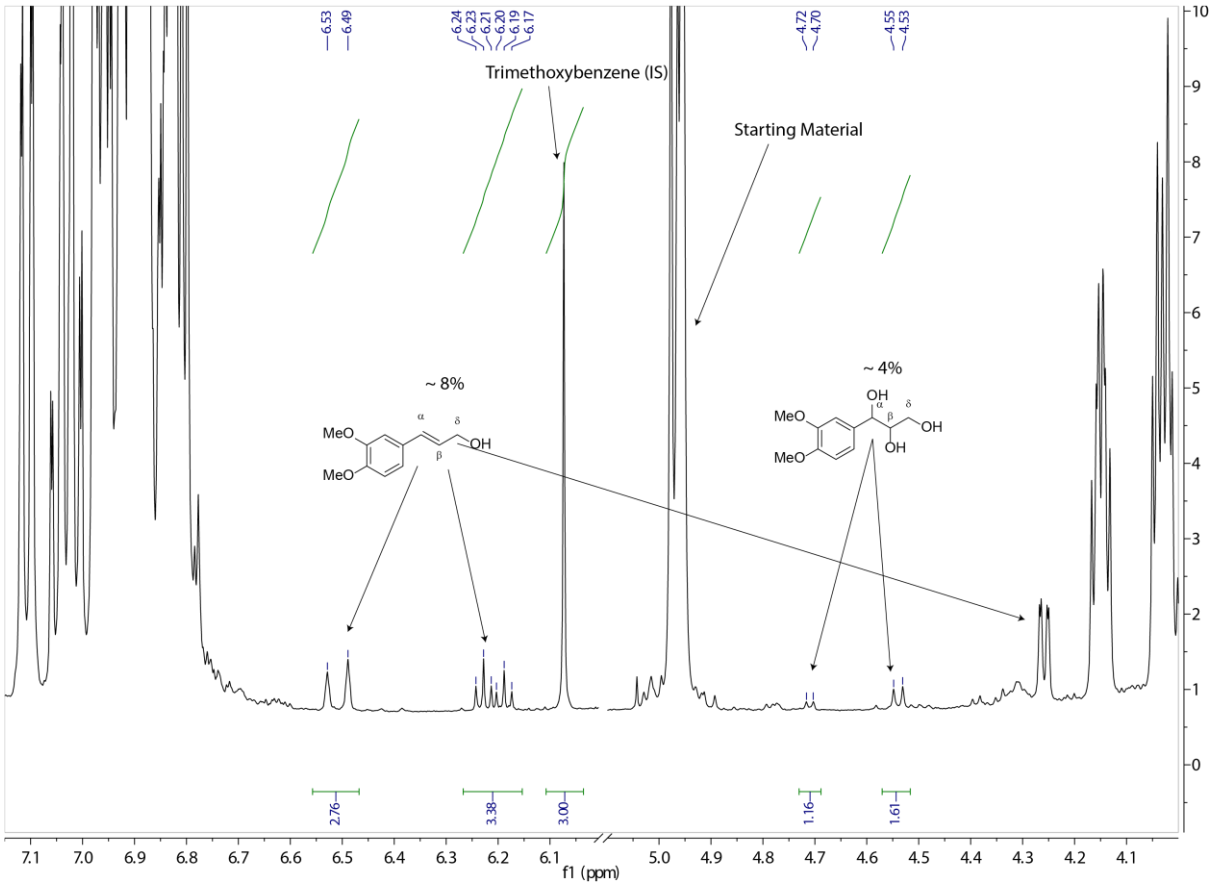


Figure S33 1D Proton NMR spectrum in $CDCl_3$ of the crude reaction mixture obtained from the Kraft reaction (0.12 M NaOH, 0.25 M Na_2S) of 1-(3,4-dimethoxyphenyl)-2-(2-methoxyphenoxy)propane-1,3-diol (**S10**) at a concentration of 10 mg/mL. Cinnamyl alcohol and arylglycerols have been assigned based on comparison to literature data.[47,48]

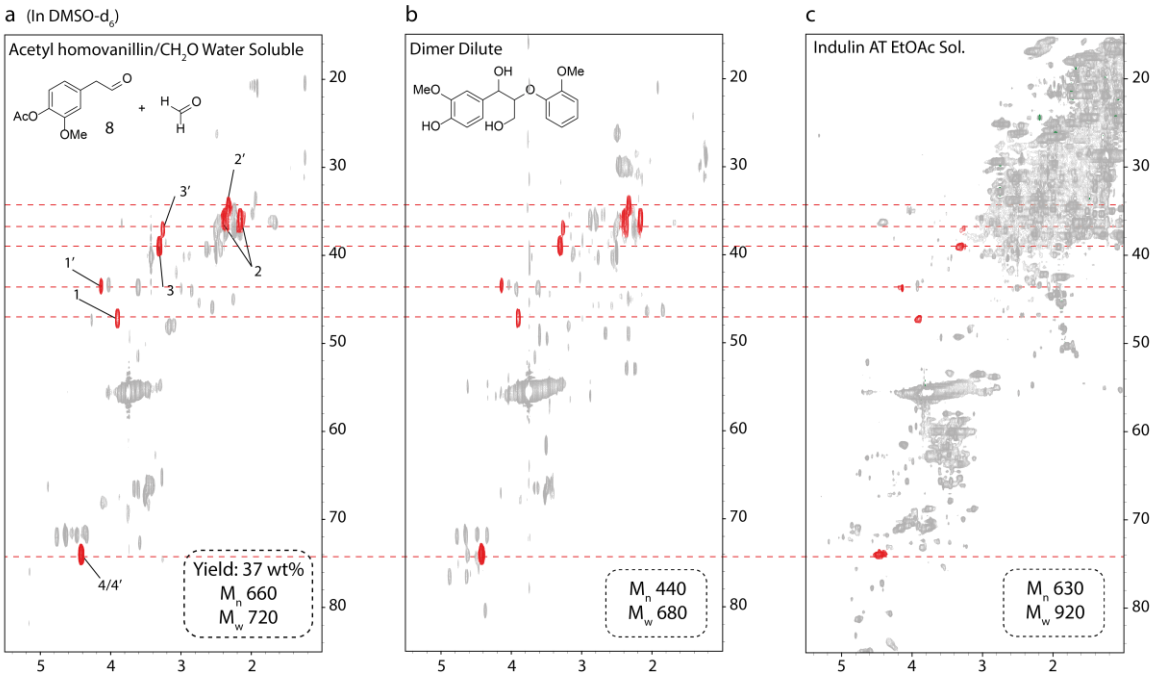


Figure S34 $^{1}H–^{13}C$ HSQC spectra, in DMSO-d_6 showing the identification of lactone **5** in both synthetic and real Kraft lignins: **a**, the acidic water-soluble Kraft reaction products obtained from acetylhomovanillin (**8**) and formaldehyde (**7**); **b**, the Kraft products from model compound **3**; **c**, the EtOAc soluble fraction of Indulin AT Kraft lignin.

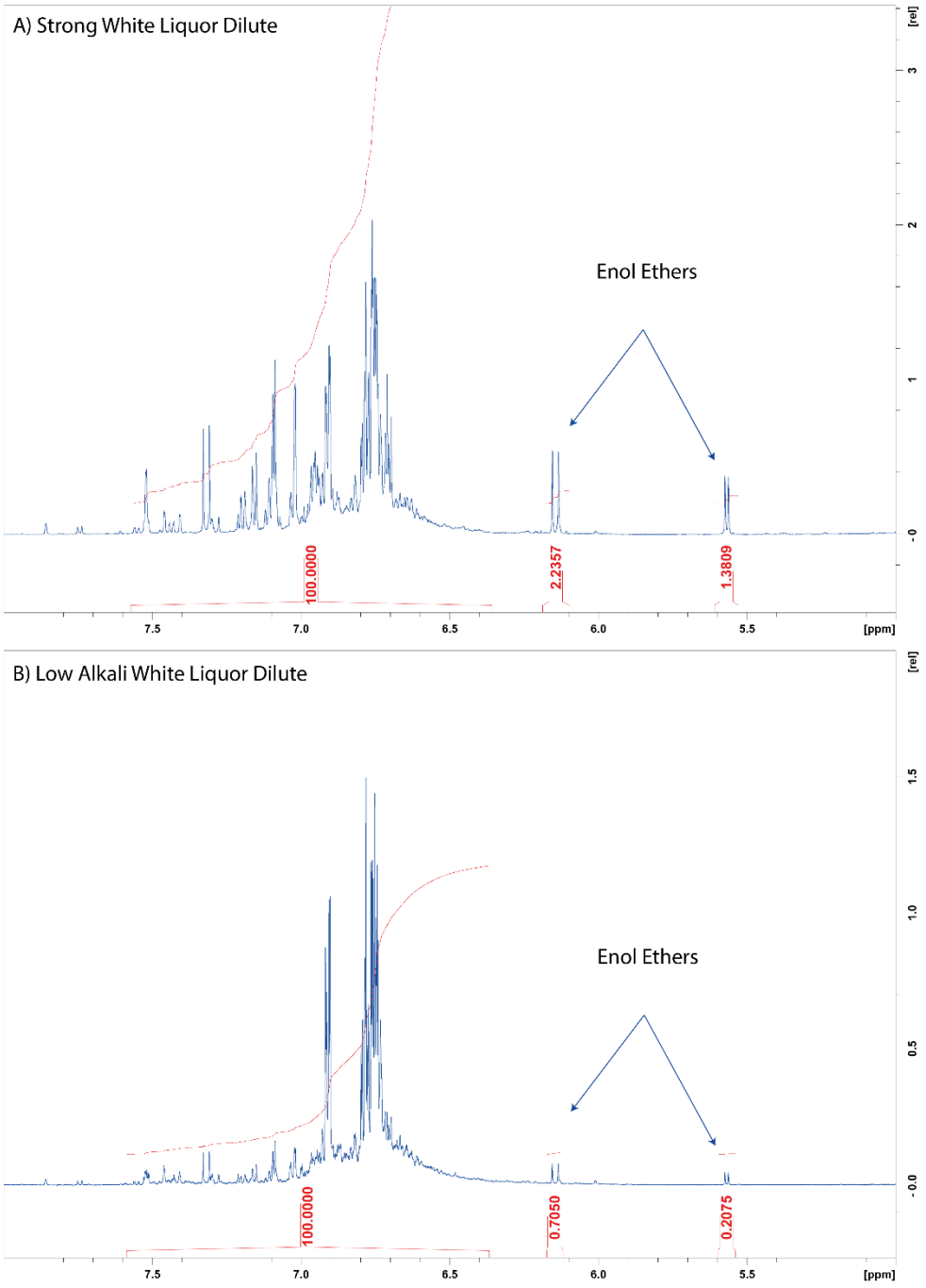


Figure S35 1D Proton NMR spectra in DMSO-d_6 of the crude reaction mixtures obtained from Kraft reaction of guaiacylglycerol-β-guaiacyl ether **3** at a concentration of 10 mg/mL using A) a strong white liquor (1.2 M NaOH, 0.25 M Na_2S) and B) a low alkali (0.12 M NaOH, 0.25 M Na_2S) white liquor. For comparison the integrals of both aromatic regions have been set to 100.

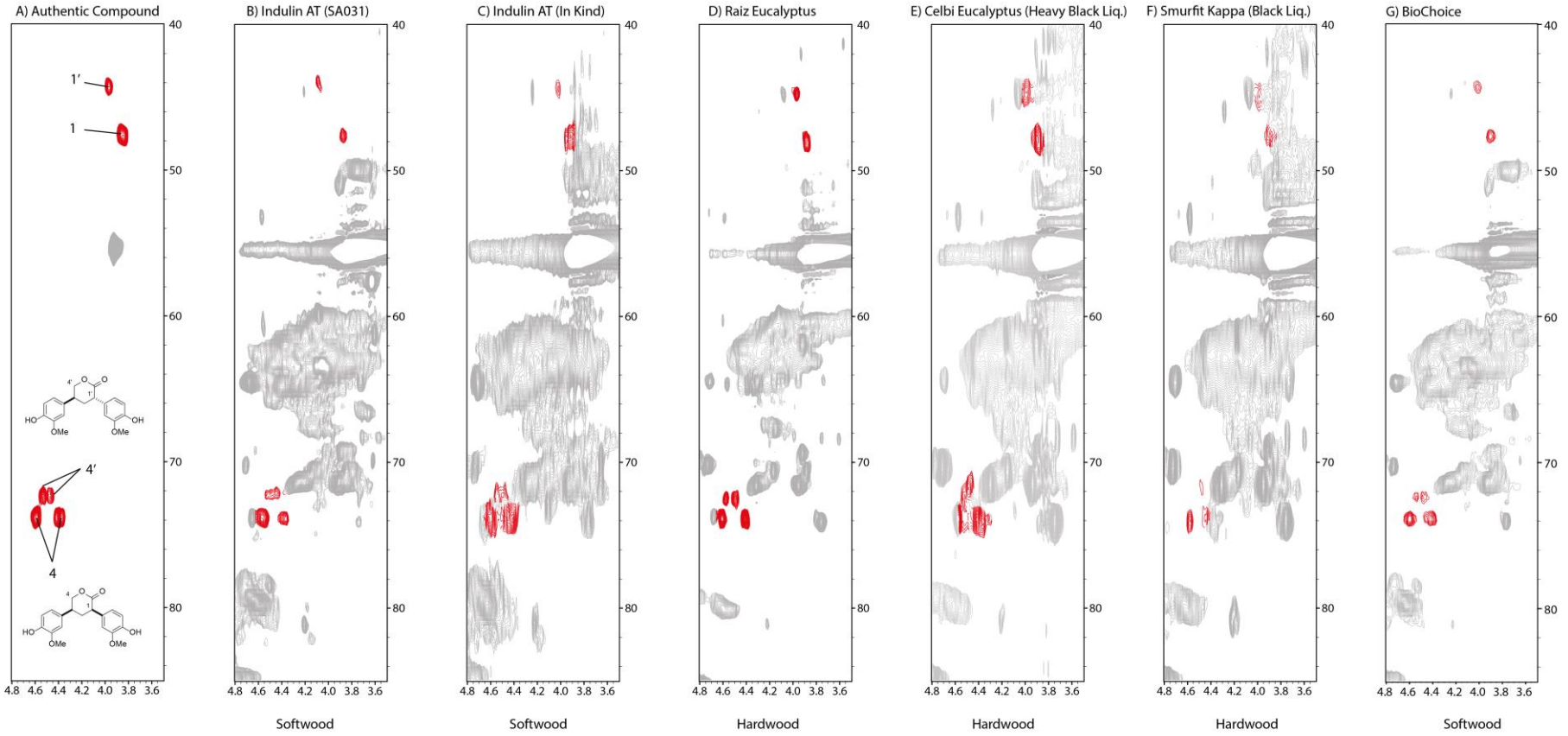


Figure S36 Assignment of Kraft derived lactones (Red) in various soft and hardwood acetylated Kraft lignins (in $CDCl_3$). A, C, E, and F were acquired with 128 increments in F1, B, D and G were observed with 256 or 384 increments in F1 (160 ppm SW)

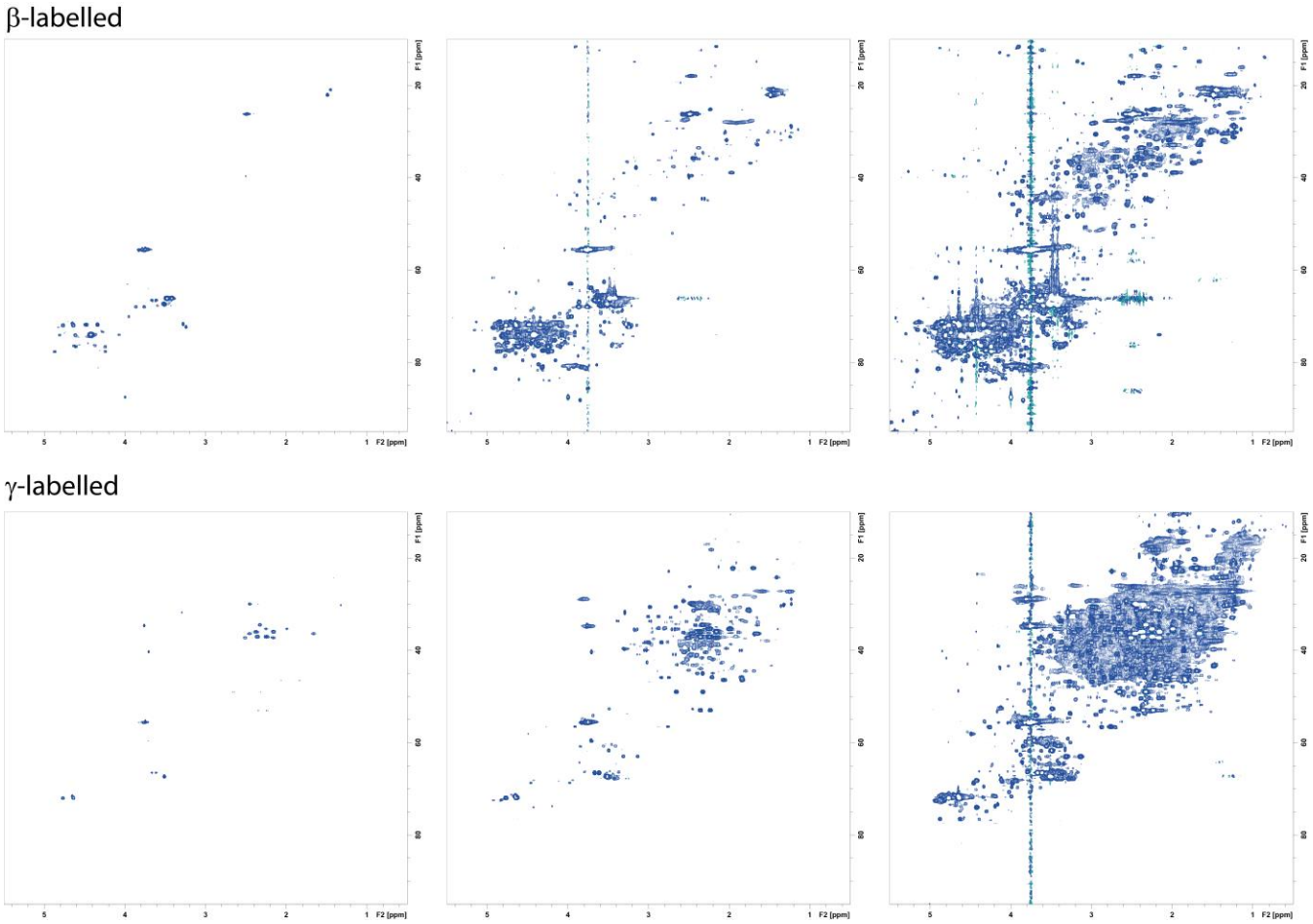


Figure S37 Sections of the 2D ^{1}H–^{13}C HSQC spectra of the synthetic Kraft lignins obtained from the dilute Kraft reactions of β (Top) and γ (Bottom) ^{13}C labelled guaiacylglycerol-β-guaiacyl ether models (**3**) plotted at different contour levels to show the abundances of major and minor new chemical species.

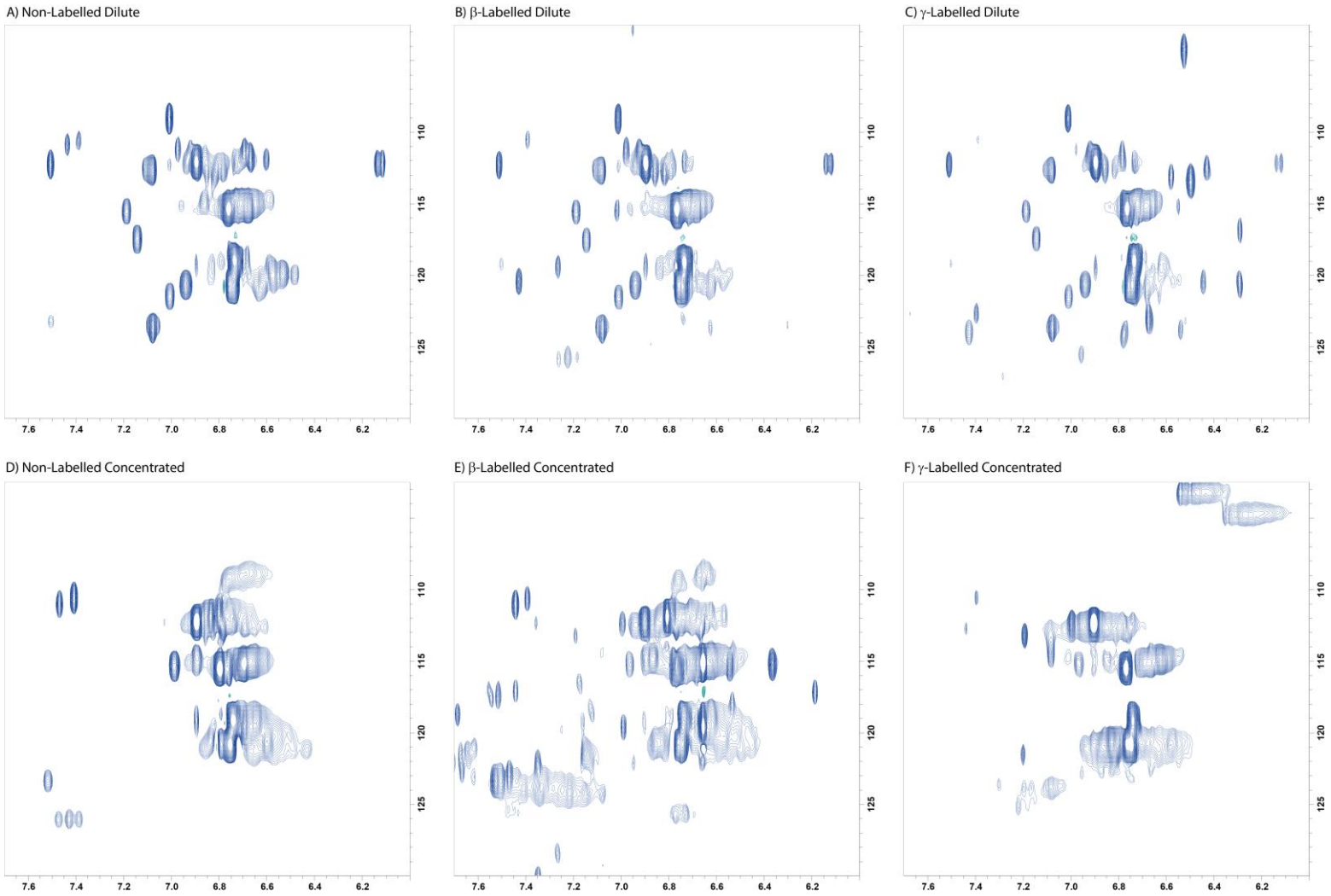


Figure S38 Aromatic regions of the 2D HSQC spectra obtained from the Kraft reaction of non-labelled (A & D) and ^{13}C labelled (B, D, E & F) compounds at 10 mg/mL (A-C) and 100 mg/mL (D-F) reaction concentrations.

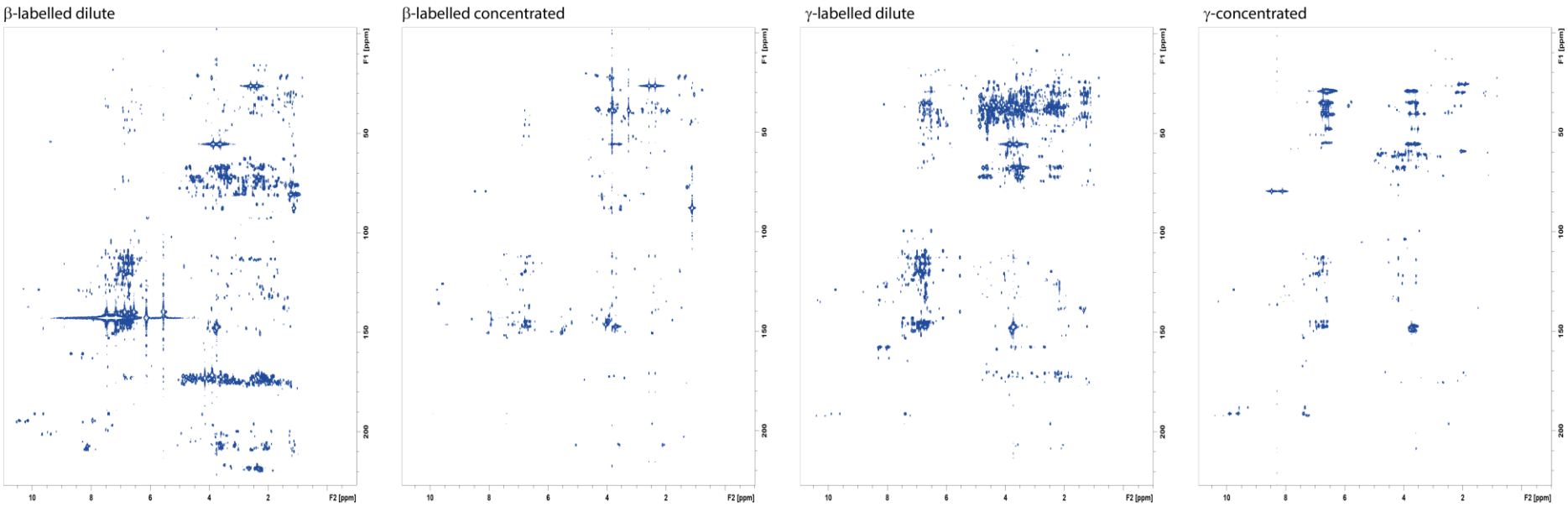


Figure S39 2D ^{1}H–^{13}C HMBC spectra of the synthetic Kraft lignins obtained from the Kraft reactions of β and γ ^{13}C labelled guaiacylglycerol-β-guaiacyl ether models (**3**) obtained from dilute and concentrated Kraft reactions.

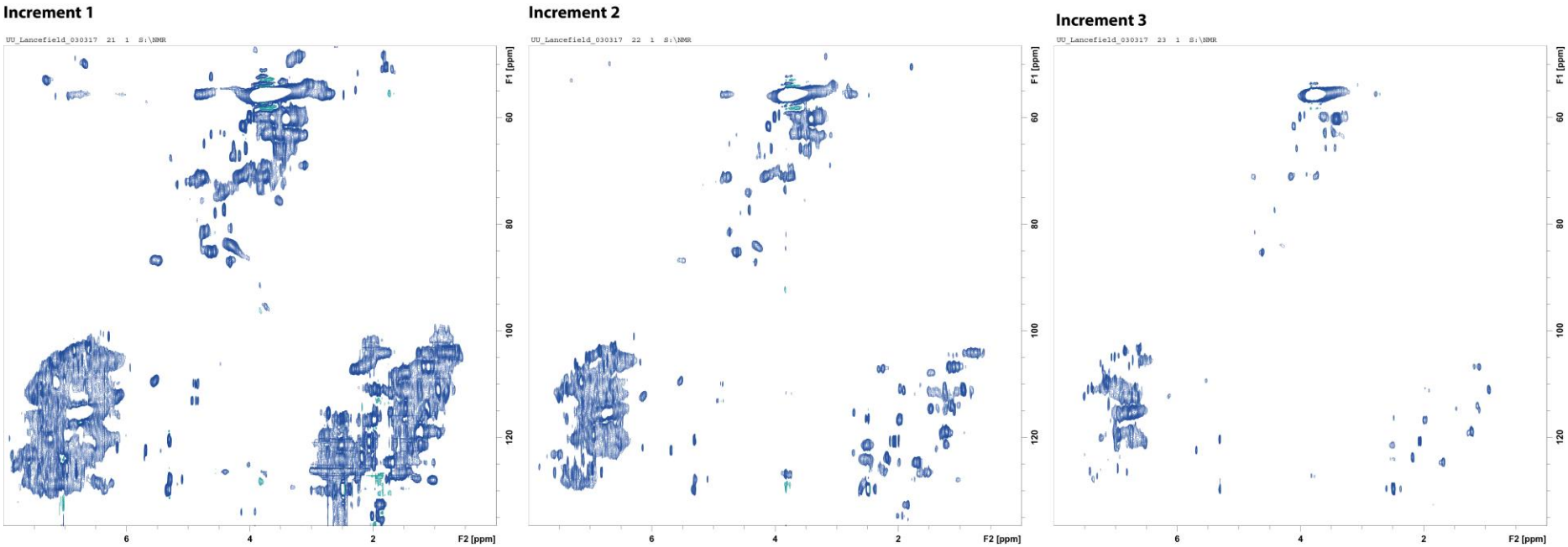


Figure S40 $HSQC_0$ spectra for 10% EtOAc/MeOH fraction showing the decrease in absolute intensity of cross peak across the three incremented experiments. Spectra are plotted at identical contour levels for comparison.

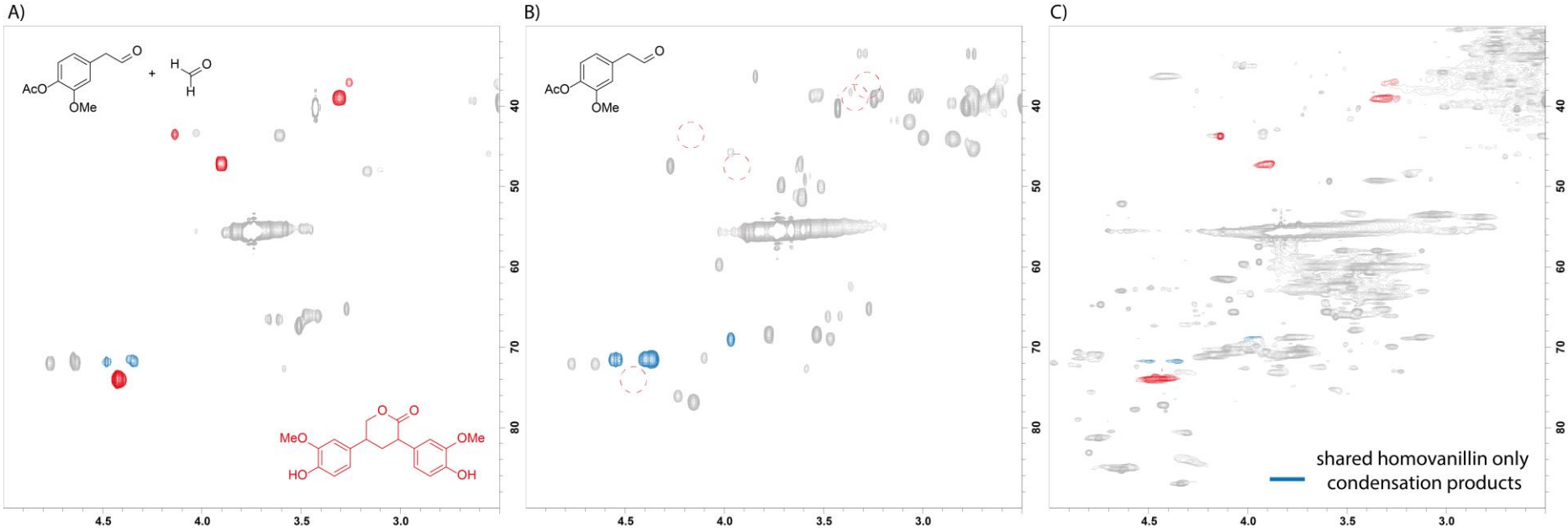


Figure S41 Sections of the 2D HSQC spectra comparing the Kraft reaction products from A) acetyl homovanillin and formaldehyde, B) acetyl homovanillin only and C) the ethyl acetate soluble Kraft lignin fraction. The contours coloured red highlight the presence of lactone **4** in formaldehyde containing reactions (A) and in lignin (C), as well as its virtual absence when no formaldehyde is present (B). Blue contours highlight condensation products in model reactions and lignin deriving only from homovanillin.

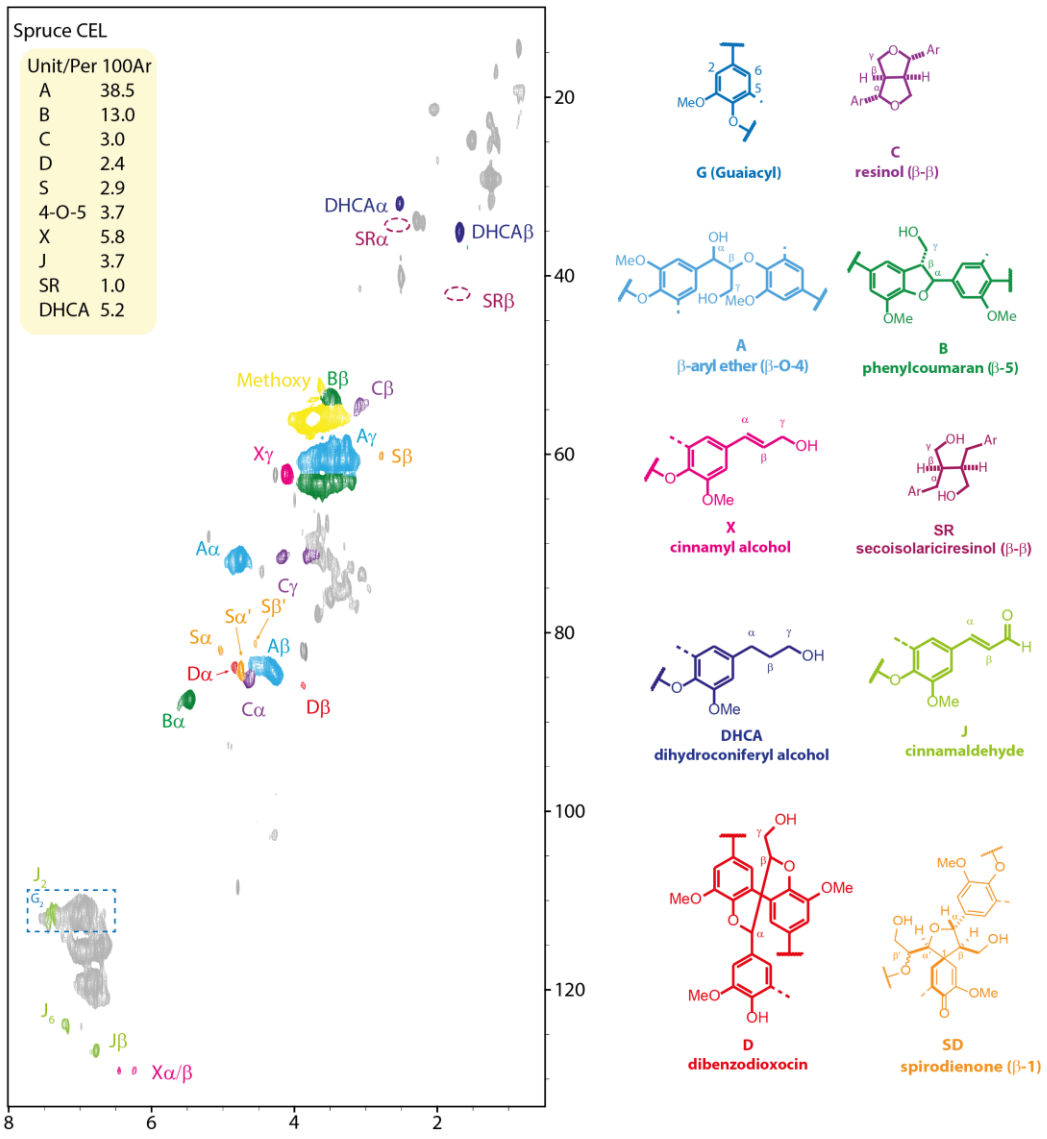


Figure S42 2D HSQC analysis of spruce CEL showing the major and minor lignin units. Data was acquired using the hsqcetgpsp.3 pulse sequence with D1 = 1 s.

9. References

1. Hu, K., Westler, W. M. & Markley, J. L. Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero 1H-13C HSQC (HSQC0). *J. Am. Chem. Soc* **133,** 1662–1665 (2011).

2. Sette, M., Lange, H. & Crestini, C. Quantitative HSQC Analyses of Lignin: A Practical Comparison. *Comput. Struct. Biotechnol. J.* **6,** e201303016 (2013).

3. Fardus-Reid, F., Warren, J. & Le Gresley, A. Validating heteronuclear 2D quantitative NMR. *Anal. Methods* **8,** 2013–2019 (2016).

4. Zhang, L. & Gellerstedt, G. Quantitative 2D HSQC NMR determination of polymer structures by selecting suitable internal standard references. *Magn. Reson. Chem.* **45,** 37–45 (2007).

5. Jiang, Z.-H., Argyropoulos, D. S. & Granata, A. Correlation analysis of31P NMR chemical shifts with substituent effects of phenols. *Magn. Reson. Chem.* **33,** 375–382 (1995).

6. Chen, T. Y. *et al.* Structural variations of lignin macromolecule from different growth years of Triploid of Populus tomentosa Carr. *Int. J. Biol. Macromol.* **101,** 747–757 (2017).

7. Huang, G.-J. *et al.* A concise synthesis of viscolin, and its anti-inflammatory effects through the suppression of iNOS, COX-2, ERK phosphorylation and proinflammatory cytokines expressions. *Eur. J. Med. Chem.* **48,** 371–378 (2012).

8. Fuglseth, E., Anthonsen, T. & Hoff, B. H. New chiral building blocks from acetovanillone using lipase A and B from Candida antarctica. *Tetrahedron: Asymmetry* **17,** 1290–1295 (2006).

9. Miles-Barrett, D. M. *et al.* Use of Bisulfite Processing To Generate High-β-O-4 Content Water-Soluble Lignosulfonates. *ACS Sustain. Chem. Eng.* **5,** 1831–1839 (2017).

10. Trita, A. S. *et al.* Synthesis of potential bisphenol A substitutes by isomerising metathesis of renewable raw materials. *Green Chem.* **19,** 3051–3060 (2017).

11. Lancefield, C. S. & Westwood, N. J. The synthesis and analysis of advanced lignin model polymers. *Green Chem.* **17,** 4980–4990 (2015).

12. Lemière, G. *et al.* 3',4-Di-O-methylcedrusin: synthesis, resolution and absolute configuration. *J. Chem. Soc. Perkin Trans. 1* 1775–1779 (1995).

13. Pieters, L. *et al.* Synthesis and biological evaluation of dihydrobenzofuran lignans and related compounds as potential antitumor agents that inhibit tubulin polymerization. *J. Med. Chem.* **42,** 5475–5481 (1999).

14. Adler, E., Marton, J. & Falkehag, I. The Behaviour of Lignin in Alkaline Pulping I. Model Experiments with Phenylcoumarans. *Acta Chem. Scand.* **18,** 1311–1312 (1964).

15. Gierer, J. & Norén, I. Über die reaktionen des lignins bei der sulfatkochung. *ACTA Chem. Scand.* **16,** 1713–1729 (1962).

16. Ralph, J., Helm, R. F. & Quideau, S. Lignin–feruloyl ester cross-links in grasses. Part 2. Model compound syntheses. *J. Chem. Soc. Perkin Trans. 1* 2971 (1992). doi:10.1039/p19920002971

17. Kishimoto, T., Uraki, Y. & Ubukata, M. Chemical synthesis of β-O-4 type artificial lignin. *Org. Biomol. Chem.* **4,** 1343 (2006).

18. Quideau, S. & Ralph, J. Facile Large-Scale Synthesis of Coniferyl, Sinapyl, and p-Coumaryl Alcohol. *Food Chem* **40,** 1108–1 (1992).

19. Lancefield, C. S., Ojo, O. S., Tran, F. & Westwood, N. J. 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.* **54,** 258–262 (2015).

20. Ralph, J. & Helm, R. F. Rapid proton NMR method for determination of threo:erythro ratios in lignin model compounds and examination of reduction stereochemistry. *J. Agric. Food Chem.* **39,** 705–709 (1991).

21. Gall, D. L. *et al.* Stereochemical Features of Glutathione-dependent Enzymes in the *Sphingobium* sp. Strain SYK-6 β-Aryl Etherase Pathway. *J. Biol. Chem.* **289,** 8656–8667 (2014).

22. Coote, S. J., Davies, S. G., Middlemiss, D. & Naylor, A. Tricarbonylchromium(0) promoted stereoselective cyclisations of the N-3,4-dimethoxyphenethyl derivatives of the 1-phenyl ethanolamines halostachine, ephedrine and pseudoephedrine to 1-phenyl-N-methyl-7,8-dimethoxy-1,2,4,5-tetrahydrobenzazepines. *Tetrahedron: Asymmetry* **1,** 33–56 (1990).

23. Kelley, C. J. *et al.* Polyphenolic acids of Lithospermum ruderale (Boraginaceae). I. Isolation and structure determination of lithospermic acid. *J. Org. Chem.* **40,** 1804–1815 (1975).

24. Brännvall, E., Bäckström, M., Mocchiutti, P., Inalbon, M. & Rånby, B. Improved impregnation efficiency and pulp yield of softwood kraft pulp by high effective alkali charge in the impregnation stage. *Holzforschung* **70,** 149–153 (2016).

25. Crestini, C., Lange, H., Sette, M. & Argyropoulos, D. S. On the structure of softwood kraft lignin. *Green Chem.* **19,** 4104–4121 (2017).

26. Jiang, X. *et al.* Fractionation and Characterization of Kraft Lignin by Sequential Precipitation with Various Organic Solvents. *ACS Sustain. Chem. Eng.* **5,** 835–842 (2017).

27. Hu, Z., Du, X., Liu, J., Chang, H. M. & Jameel, H. Structural Characterization of Pine Kraft Lignin: BioChoice Lignin vs Indulin AT. *J. Wood Chem. Technol.* **36,** 432–446 (2016).

28. Ralph, S. & Ralph, J. NMR Database of Lignin and Cell Wall Model Compounds.

29. Rencoret, J. *et al.* Isolation and structural characterization of the milled-wood lignin from Paulownia fortunei wood. *Ind. Crops Prod.* **30,** 137–143 (2009).

30. Yuan, T.-Q., Sun, S.-N., Xu, F. & Sun, R.-C. Characterization of Lignin Structures and Lignin–Carbohydrate Complex (LCC) Linkages by Quantitative ^{13}C and 2D HSQC NMR Spectroscopy. *J. Agric. Food Chem.* **59,** 10604–10614 (2011).

31. Constant, S. *et al.* New insights into the structure and composition of technical lignins: a comparative characterisation study. *Green Chem.* **18,** 2651–2665 (2016).

32. Gierer, J. & Wännström, S. Formation of Alkali-Stahle C-C-Bonds Between Lignin and Carbohydrate Fragments During Kraft Pulping. *Holzforschung* **38,** 181–184 (1984).

33. Rahman, M. M. A., Dewick, P. M., Jackson, D. E. & Lucas, J. A. Lignans of Forsythia intermedia. *Phytochemistry* **29,** 1971–1980 (1990).

34. Tran, F. *et al.* Selective modification of the β–β linkage in DDQ-treated Kraft lignin analysed by 2D NMR spectroscopy. *Green Chem.* **17,** 244–249 (2015).

35. Lancefield, C. S. *et al.* Investigation of the Chemocatalytic and Biocatalytic Valorization of a Range of Different Lignin Preparations: The Importance of β-O-4 Content. *ACS Sustain. Chem. Eng.* **4,** 6921–6930 (2016).

36. Ralph, J. & Landucci, L. L. *Lignin and Lignans: Advances in Chemistry*. (CRC Press, Taylor & Francis, Boca Raton, FL, USA, 2010).

37. Huang, X., Korányi, T. I., Boot, M. D. & Hensen, E. J. M. Ethanol as capping agent and formaldehyde scavenger for efficient depolymerization of lignin to aromatics. *Green Chem.* **17,** 4941–4950 (2015).

38. Yelle, D. J. & Ralph, J. Characterizing phenol-formaldehyde adhesive cure chemistry within the wood cell wall. *Int. J. Adhes. Adhes.* **70,** 26–36 (2016).

39. Ralph, J. *et al.* Effects of Coumarate 3-Hydroxylase Down-regulation on. *J. Biol. Chem.*

281, 8843–8853 (2006).

40. Cui, C., Sun, R. & Argyropoulos, D. S. Fractional Precipitation of Softwood Kraft Lignin: Isolation of Narrow Fractions Common to a Variety of Lignins. *ACS Sustain. Chem. Eng.* **2,** 959–968 (2014).

41. Jääskeläinen, A.-S., Liitiä, T., Mikkelson, A. & Tamminen, T. Aqueous organic solvent fractionation as means to improve lignin homogeneity and purity. *Ind. Crops Prod.* **103,** 51–58 (2017).

42. Yue, F., Lu, F., Regner, M., Sun, R. & Ralph, J. Lignin-Derived Thioacidolysis Dimers: Reevaluation, New Products, Authentication, and Quantification. *ChemSusChem* **10,** 830–835 (2017).

43. Ralph, J. *et al.* Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochem. Rev.* **3,** 29–60 (2004).

44. Kondo, R. & Sarkanen, K. V. Formation and Reaction of Coniferyl Alcohol During Alkaline Pulping. *J. Wood Chem. Technol.* **4,** 301–311 (1984).

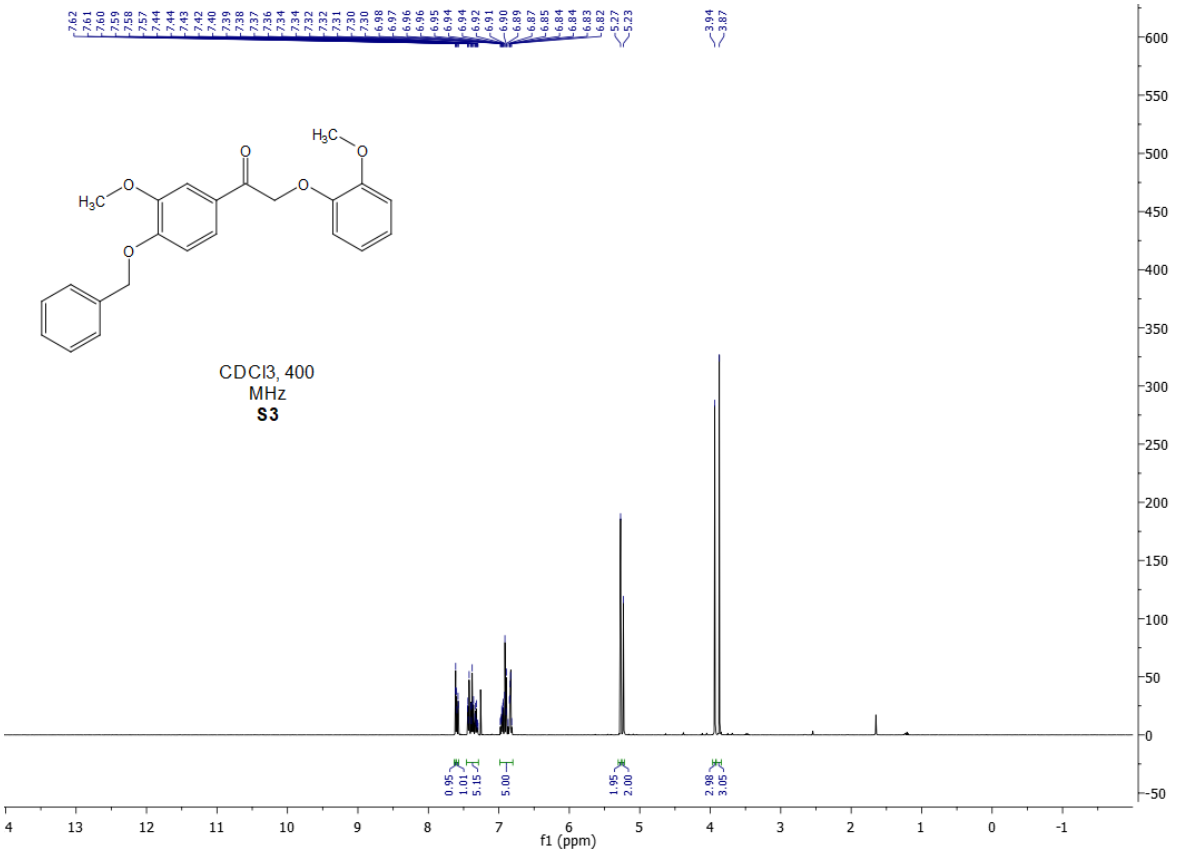
45. Berthold, F., Lindfors, E.-L. & Gellerstedt, G. Degradation of Guaiacylglycerol-β-Guaiacyl Ether in the Presence of NaHS or Polysulphide at Various Alkalinities. Part II. Liberation of Coniferyl Alcohol and Sulphur. *Holzforschung* **52,** 481–489 (1998).

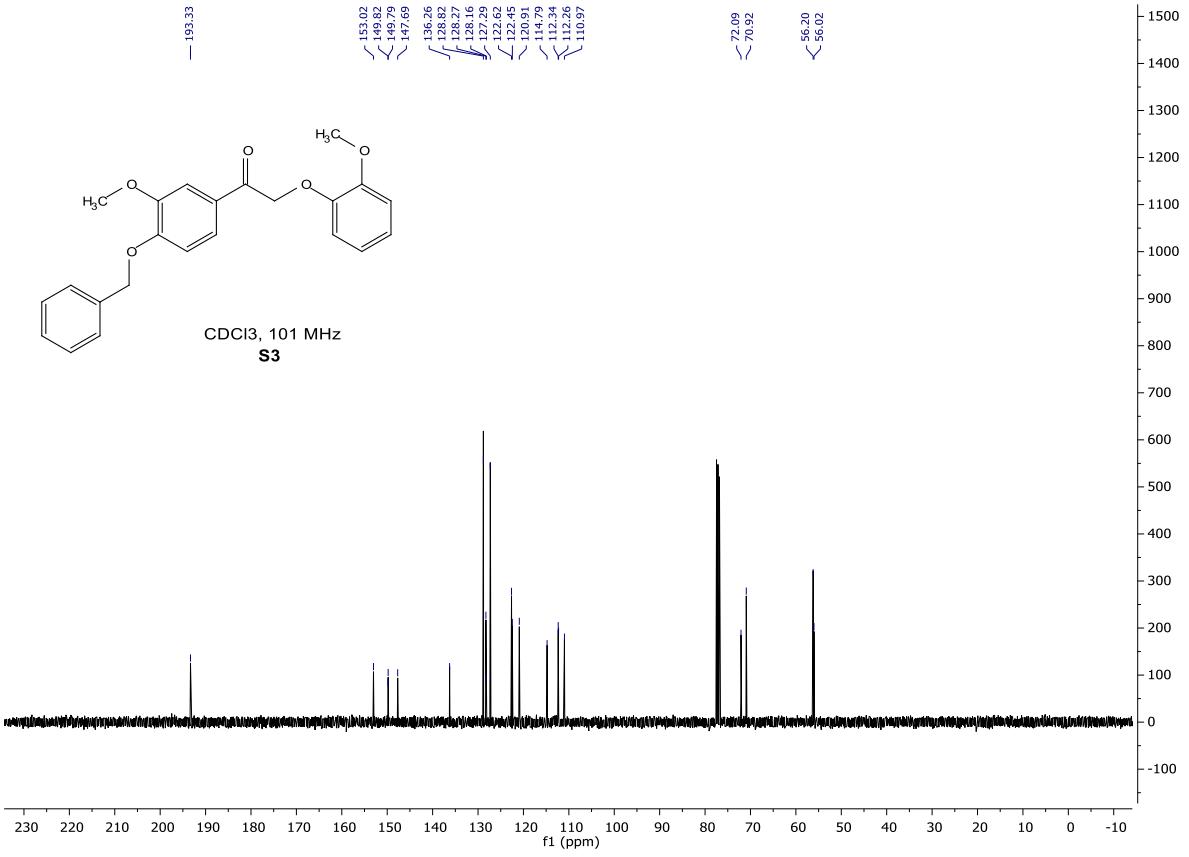
46. Grierer, J. & Lindeberg, O. Reactions of lignin during sulfate pulping. Part XV. The behaviour of intermediary coniferyl alcohol structures. *Acta Chem. Scand. B* **32,** 577–587 (1978).

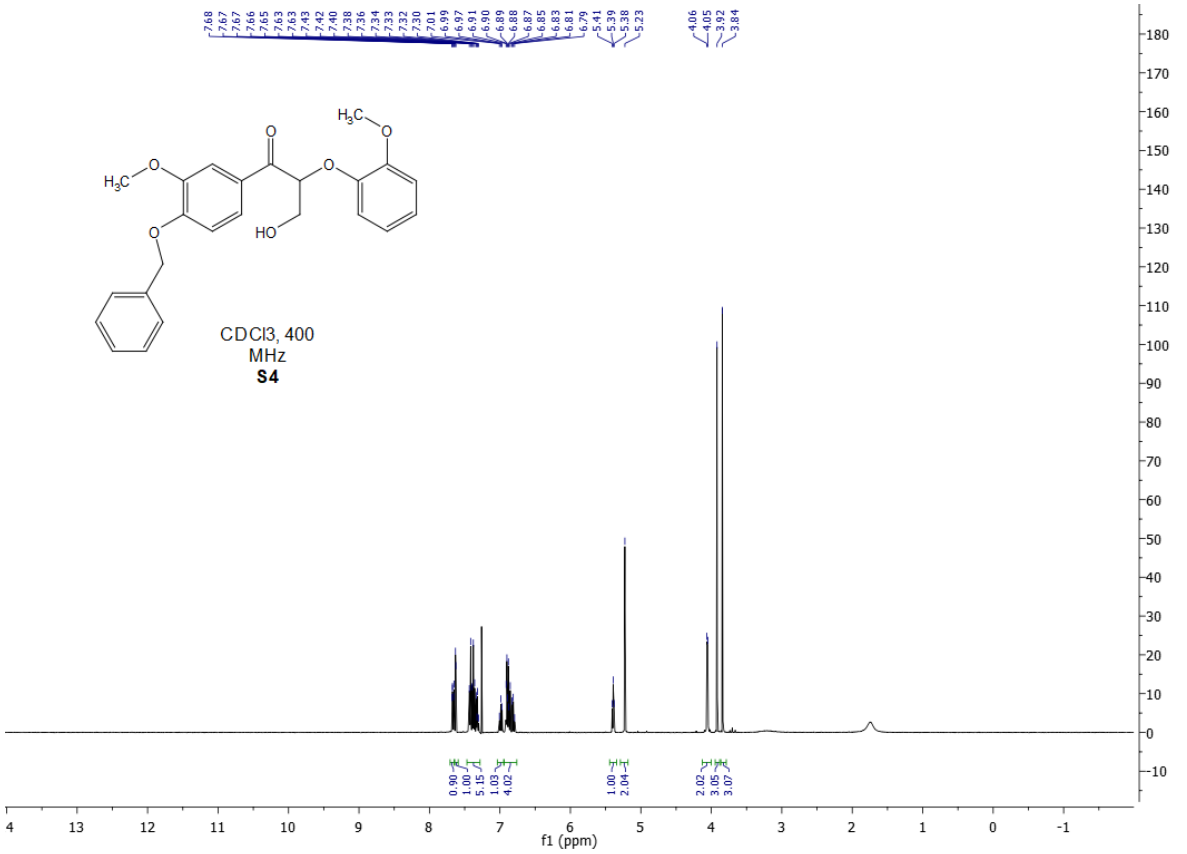
47. Miles-Barrett, D. M. *et al.* The synthesis and analysis of lignin-bound Hibbert ketone structures in technical lignins. *Org. Biomol. Chem.* **14,** 10023–10030 (2016).

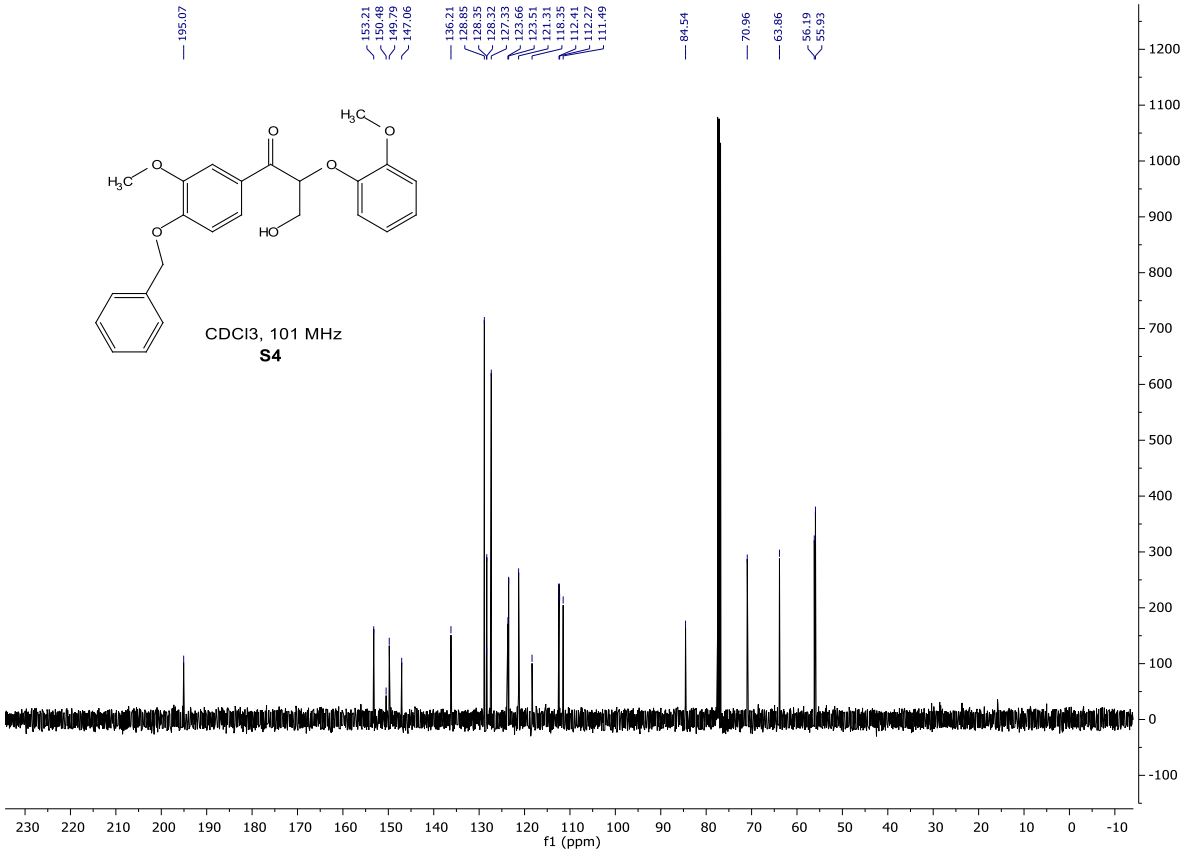
48. Meiß, R., Kumar, K. & Waldmann, H. Divergent Gold(I)-Catalyzed Skeletal Rearrangements of 1,7-Enynes. *Chem. - A Eur. J.* **21,** 13526–13530 (2015).

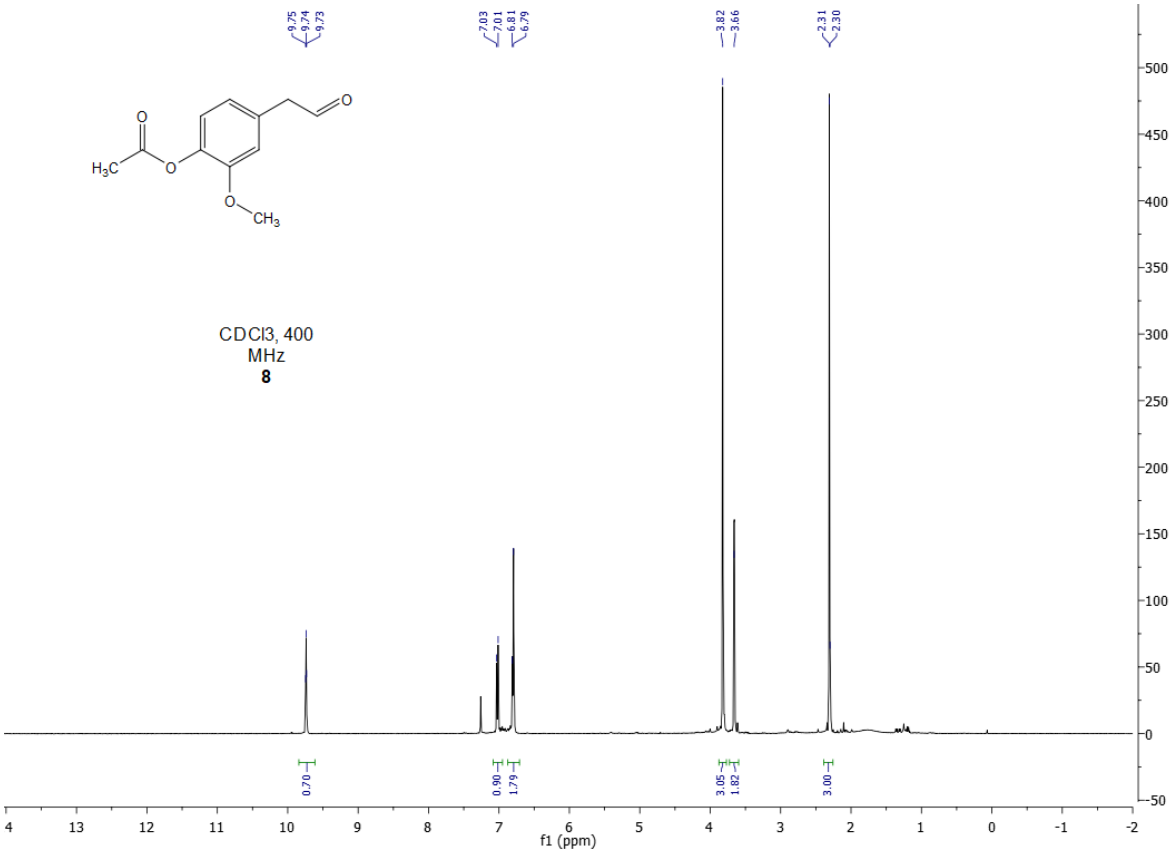
10. NMR Spectra of Novel Compounds



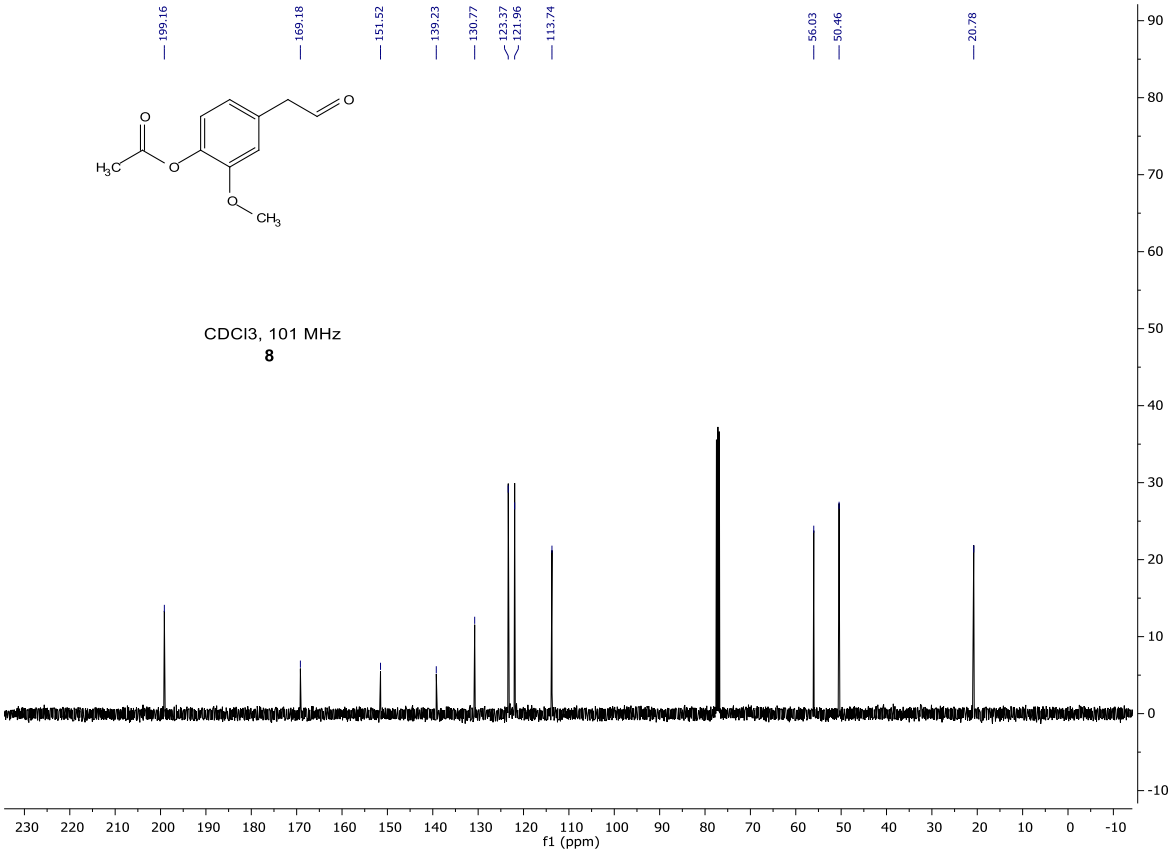



CDCl3, 400 MHz
S4
f1 (ppm)

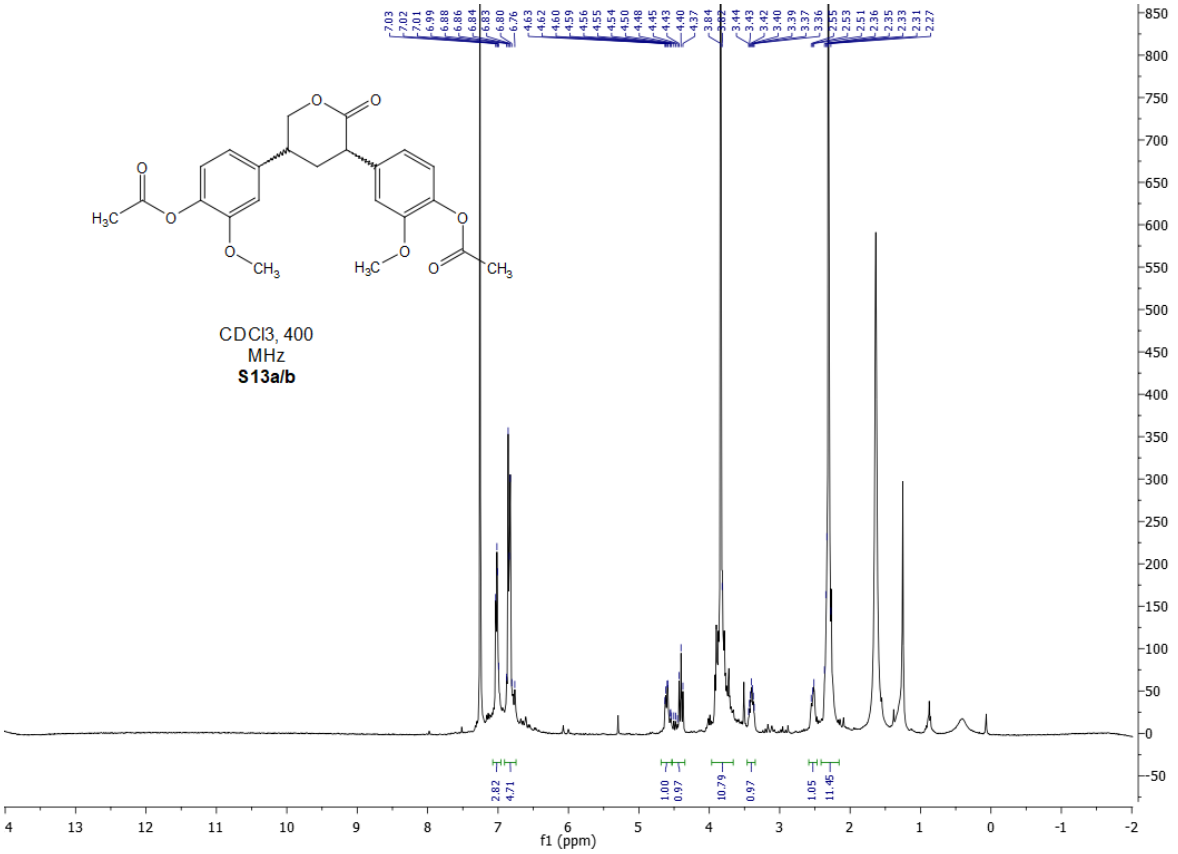

CDCl3, 101 MHz
S4
f1 (ppm)

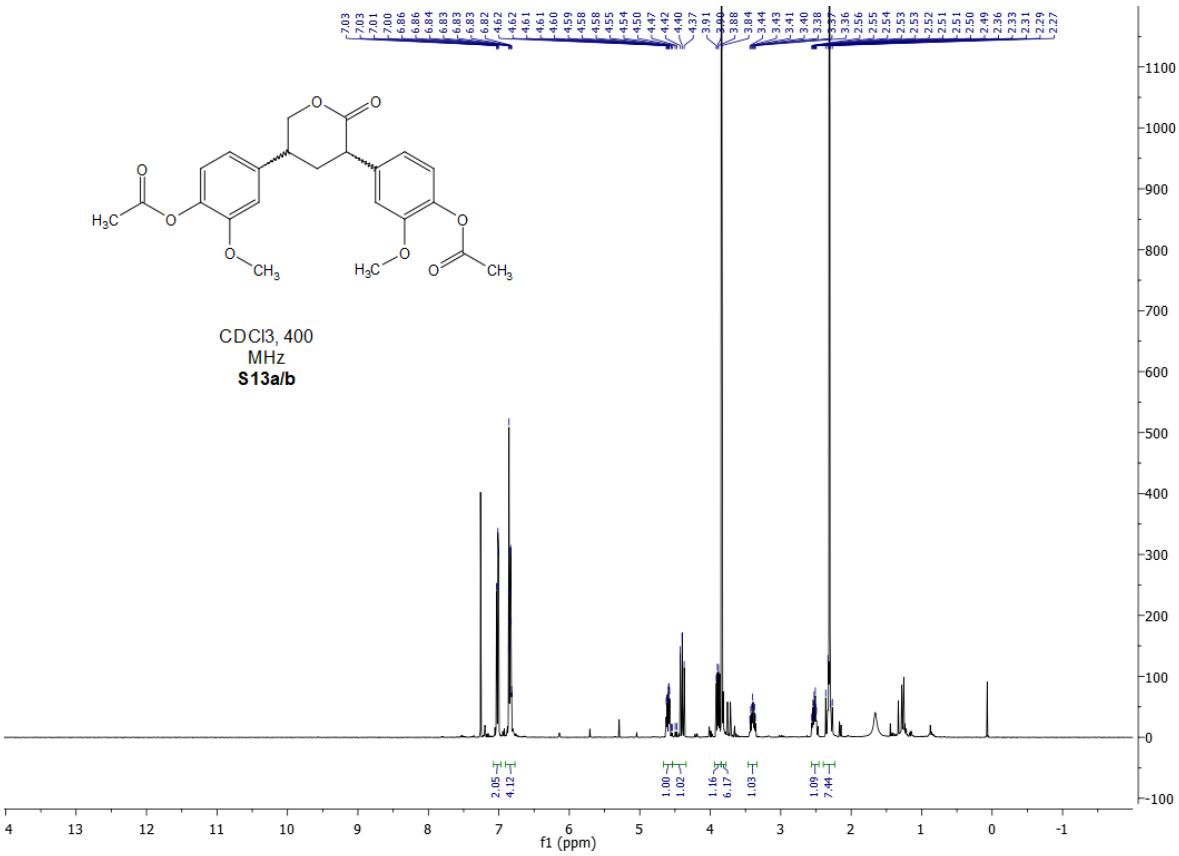


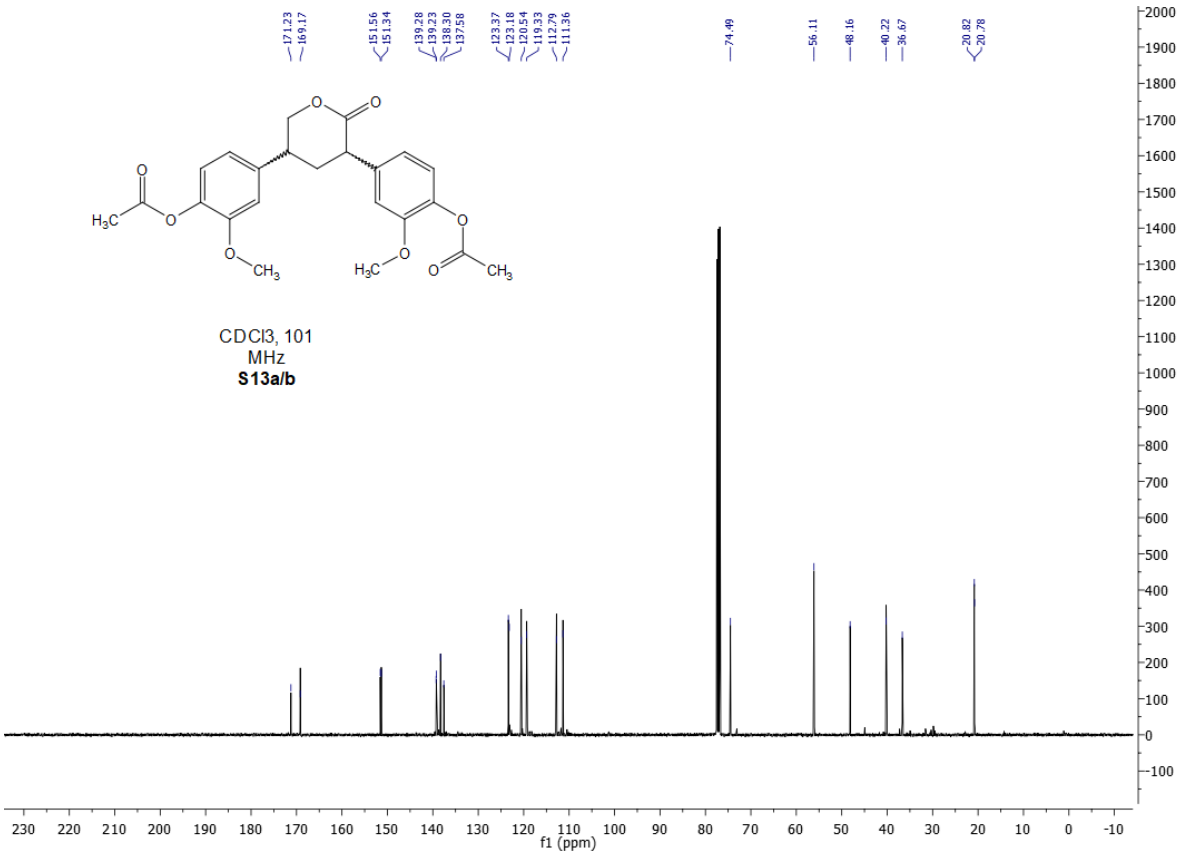
CDCl3, 400 MHz
8



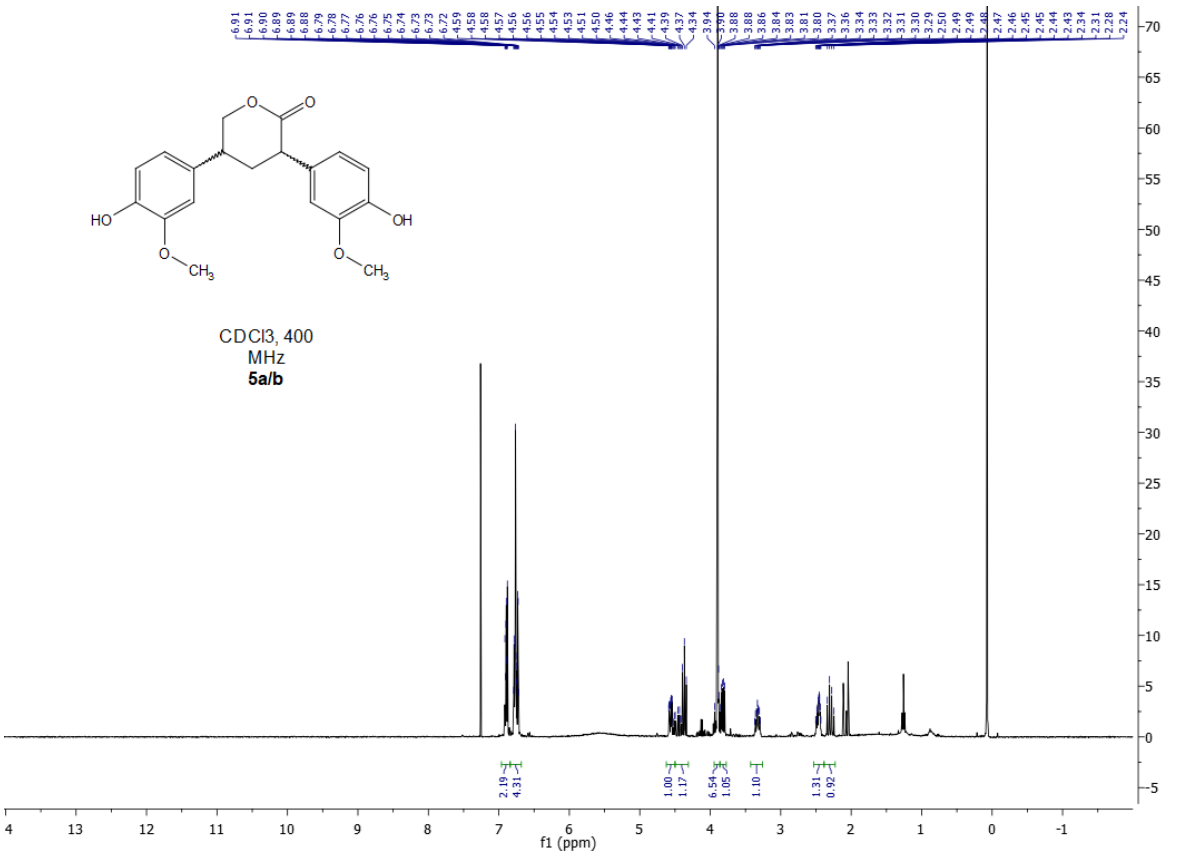
CDCl3, 101 MHz
8


CDCl3, 400
MHz
S13a/b
f1 (ppm)

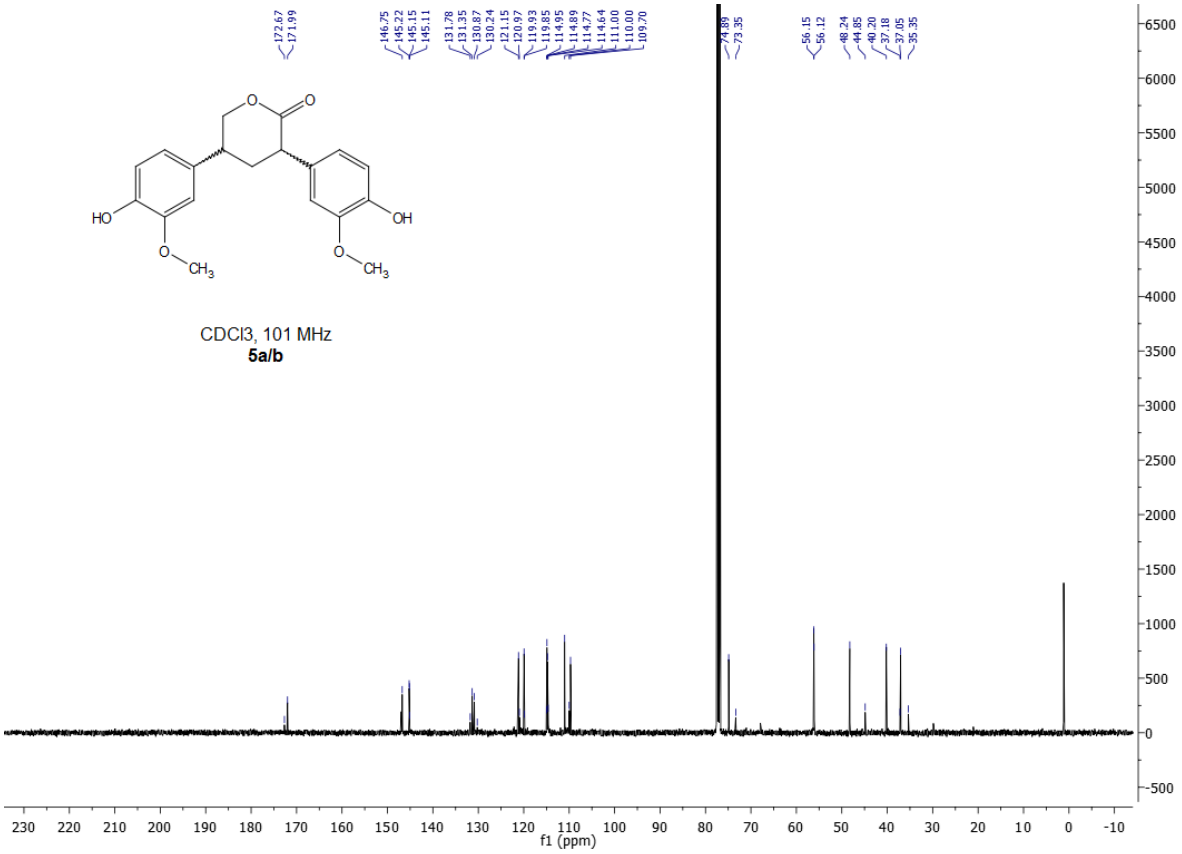

CDCl3, 400
MHz
S13a/b
f1 (ppm)

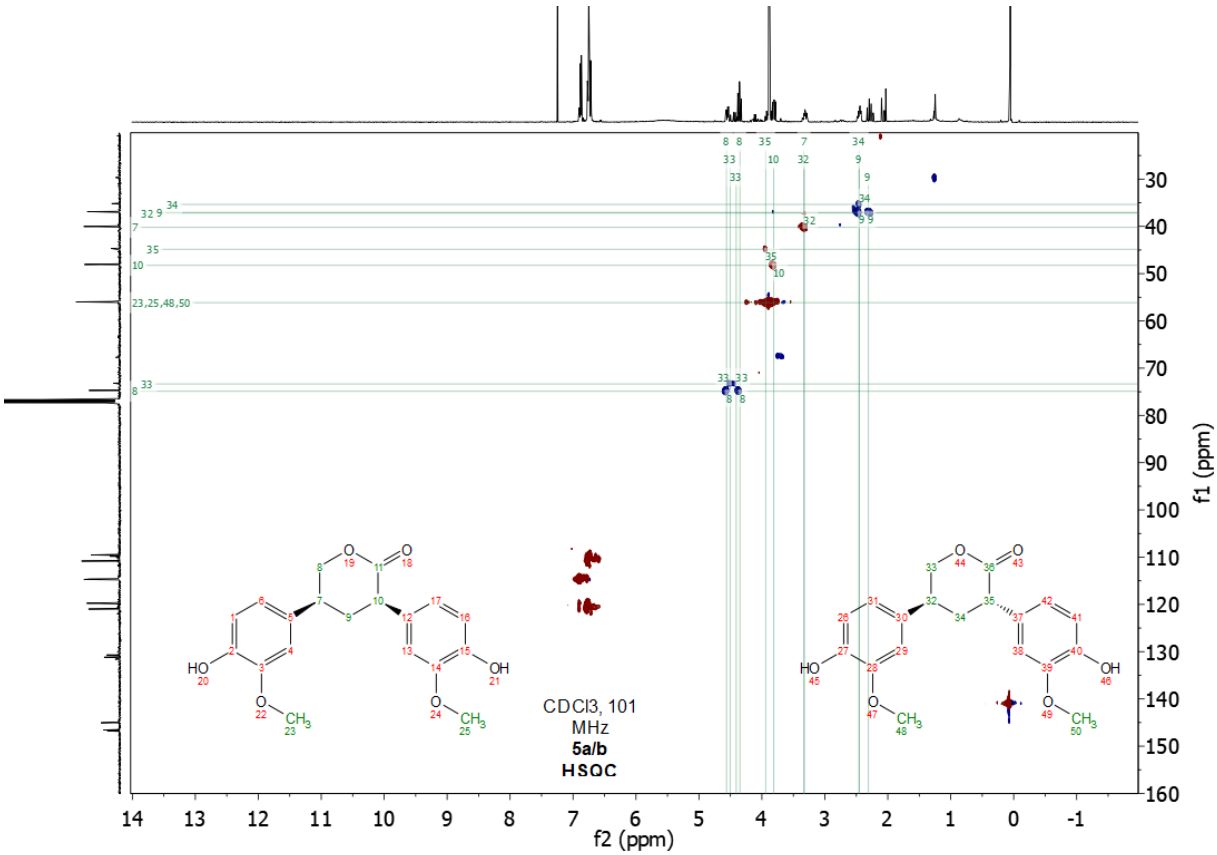


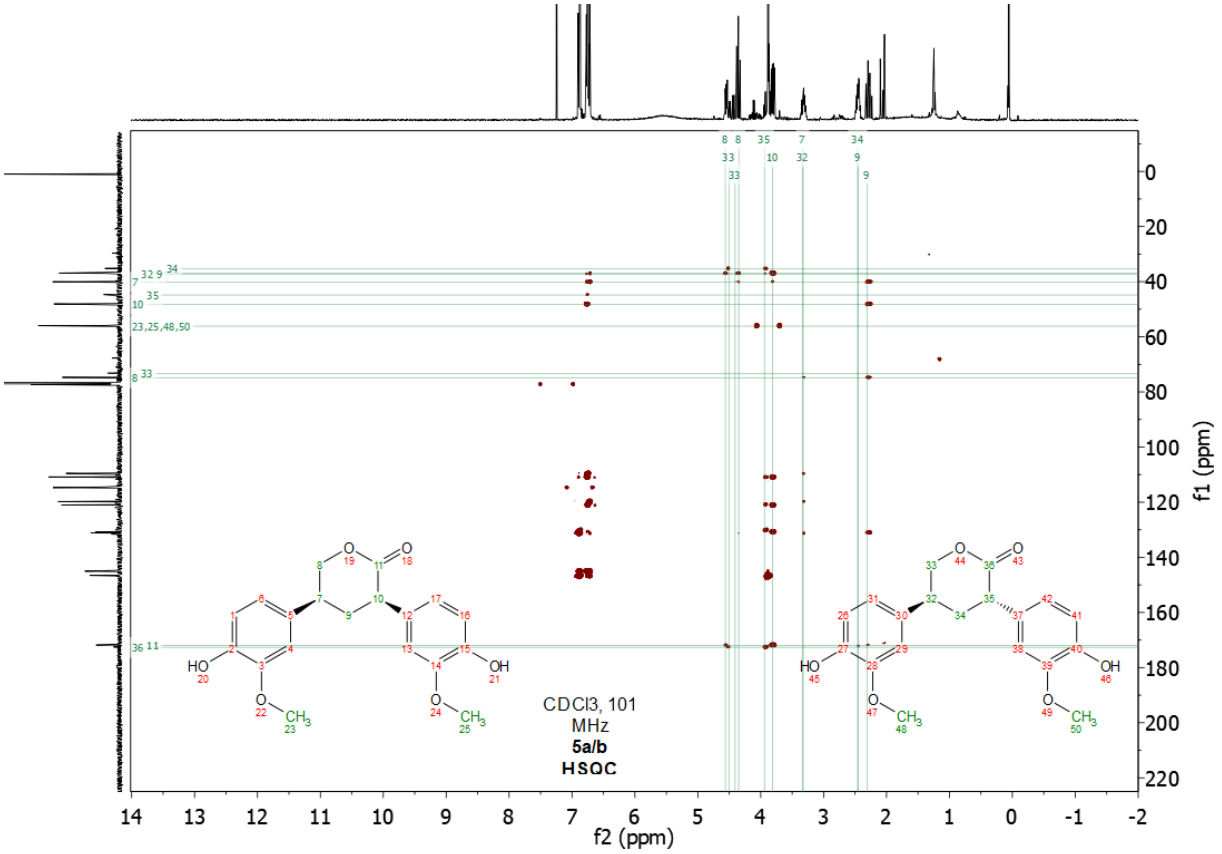
CDCl3, 101 MHz
S13a/b

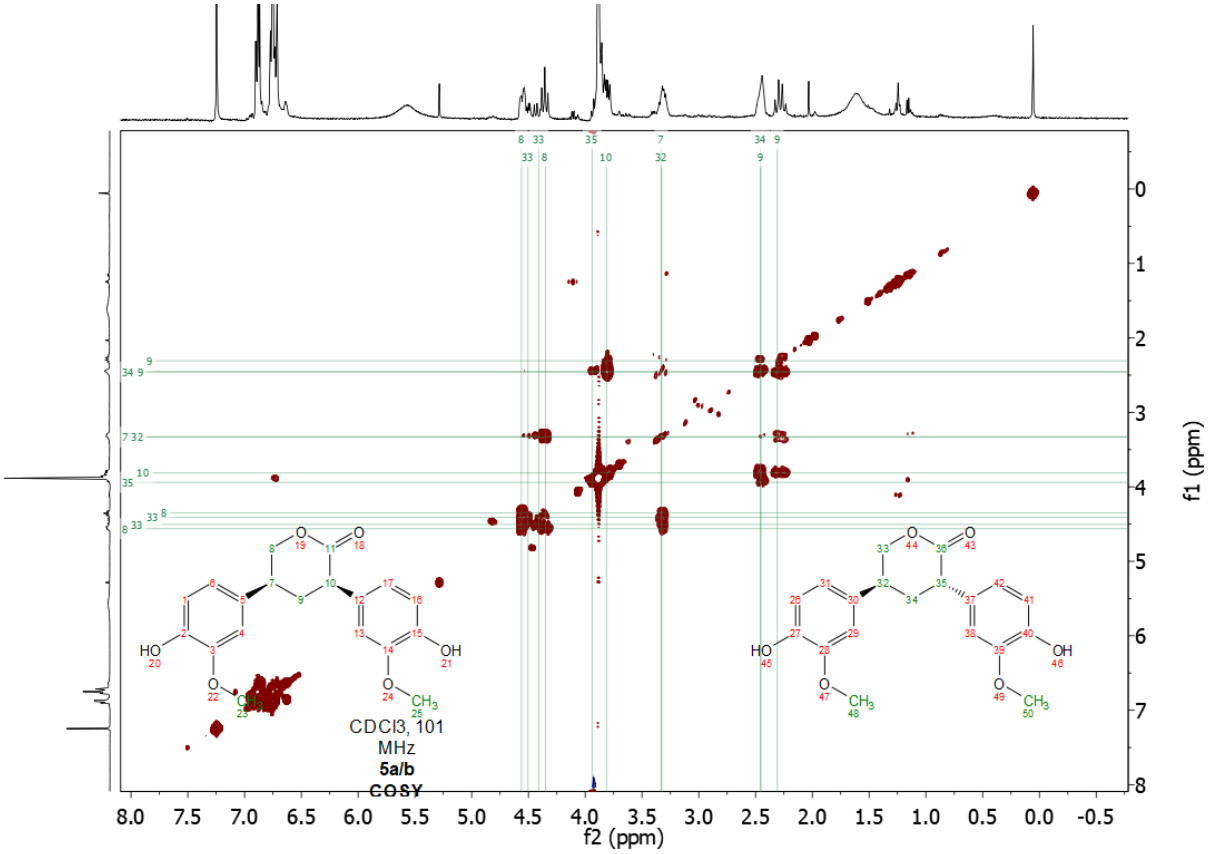


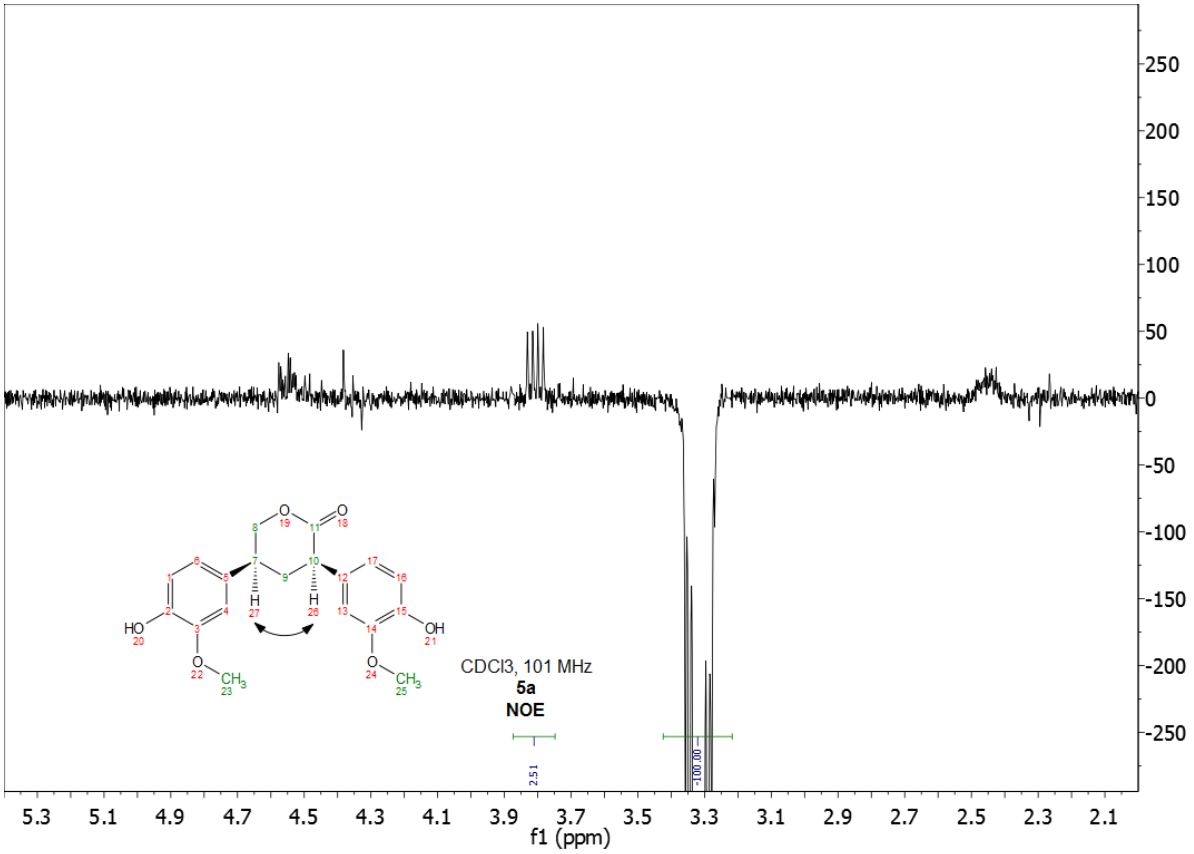
CDCl3, 400 MHz
5a/b

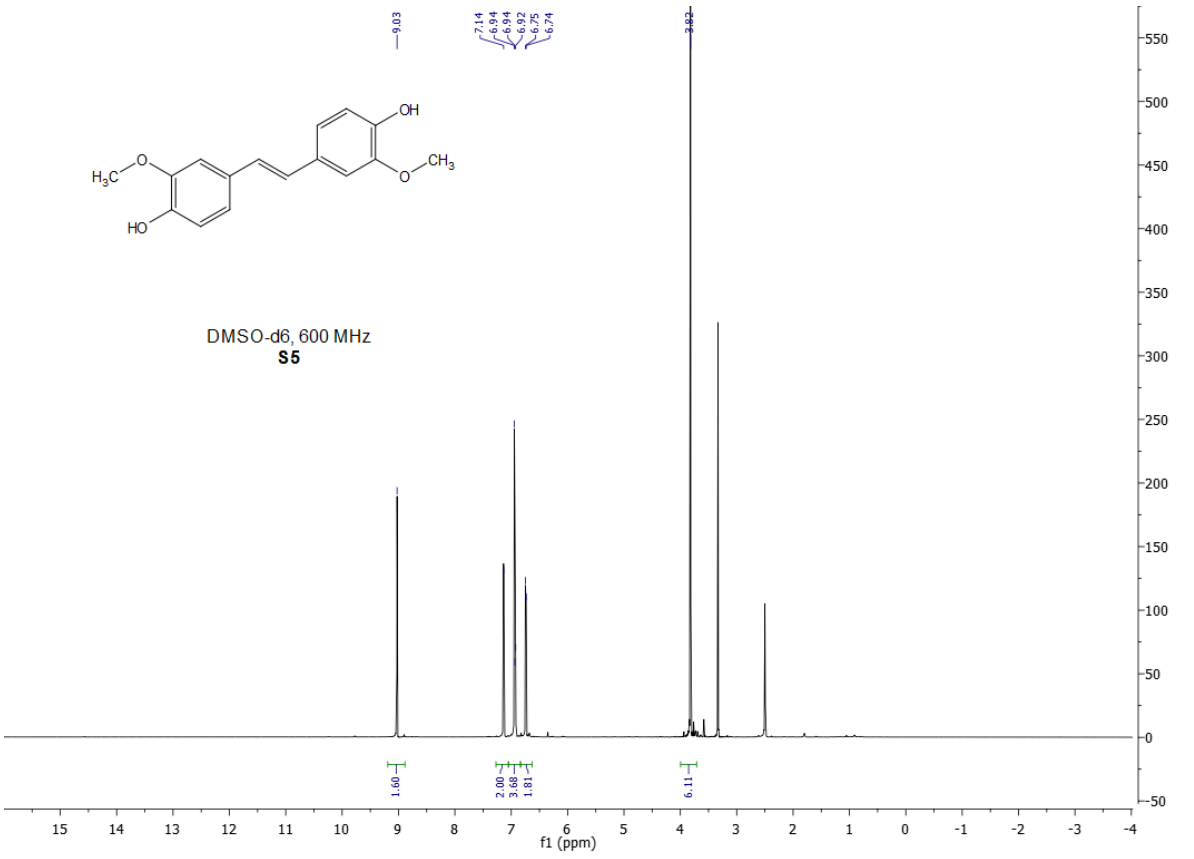

CDCl3, 101 MHz
5a/b
f1 (ppm)

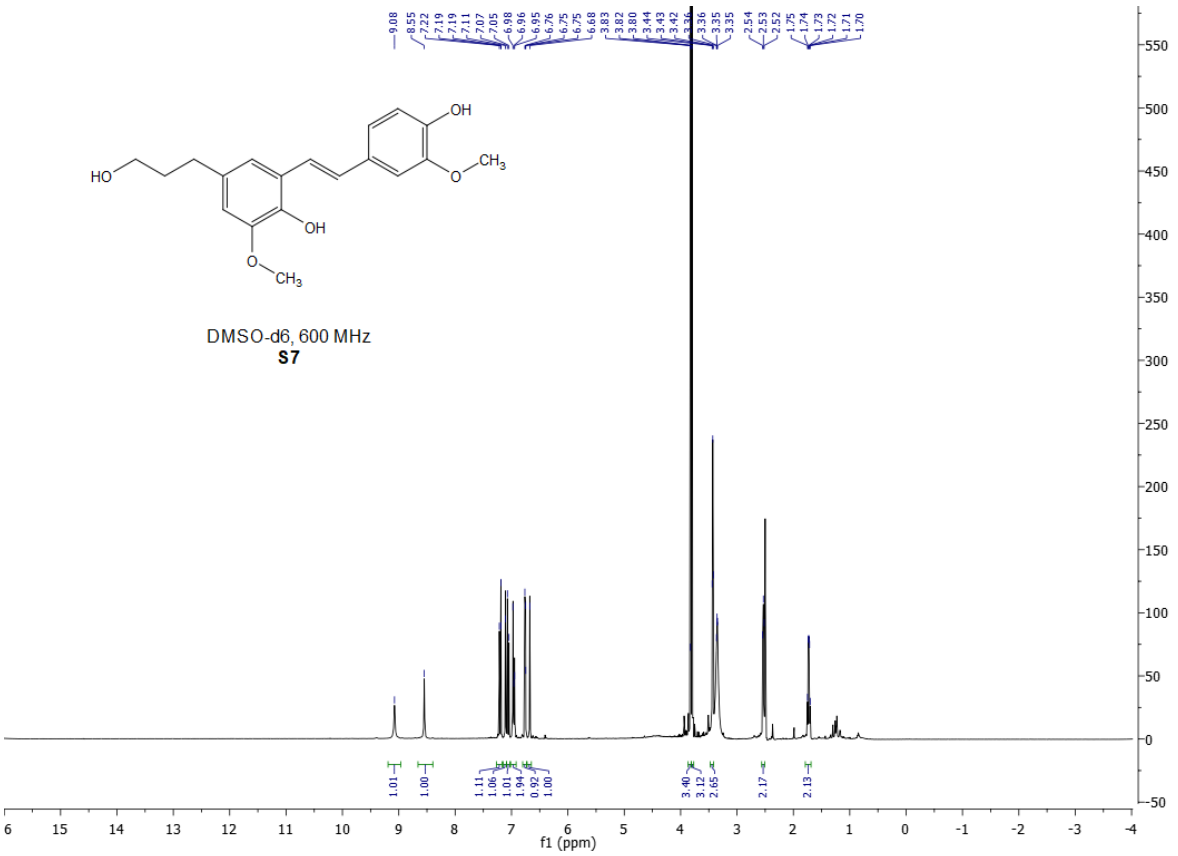

CDCl3, 101 MHz
5a/b
HSQC
f2 (ppm)
f1 (ppm)


CDCl3, 101 MHz
5a/b
HSQC
f2 (ppm)
f1 (ppm)

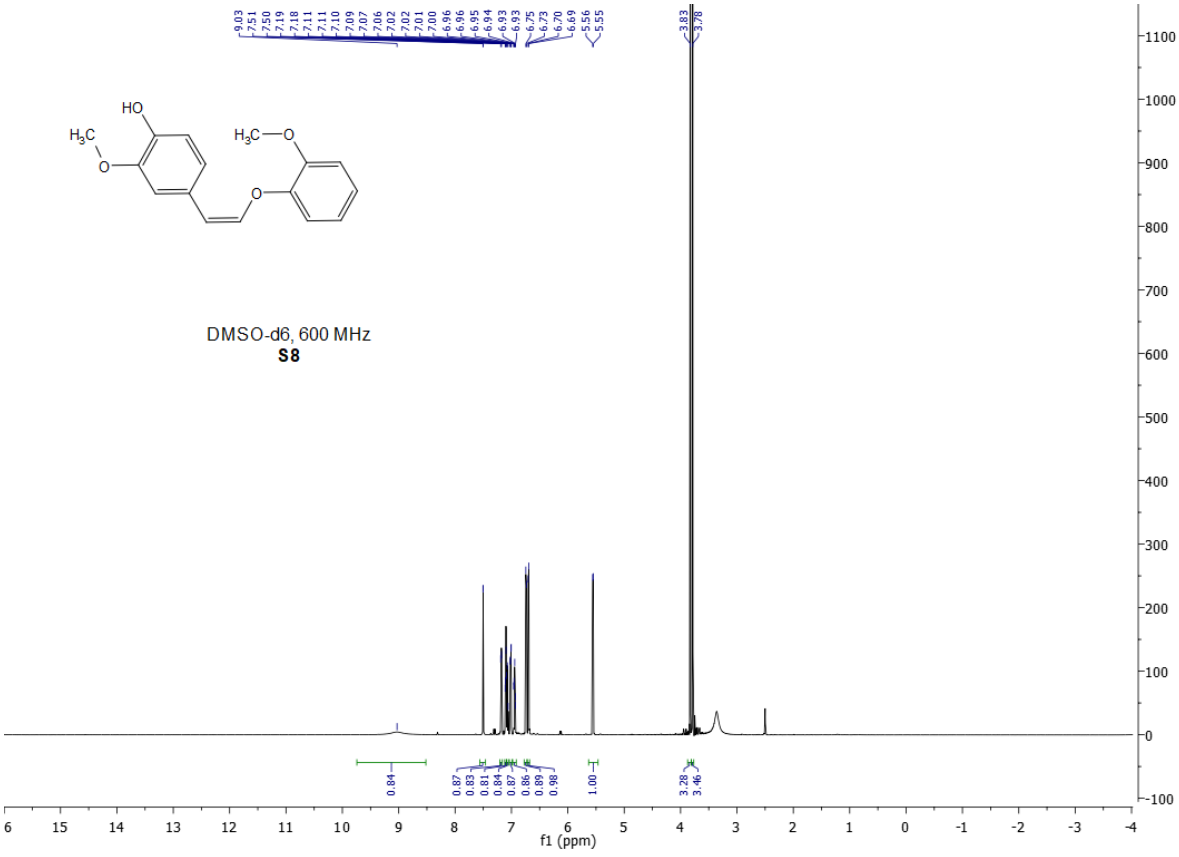

CDCl3, 101 MHz
5a/b
COSY
f2 (ppm)
f1 (ppm)


CDCl3, 101 MHz
5a
NOE
2.51
-100.00
f1 (ppm)

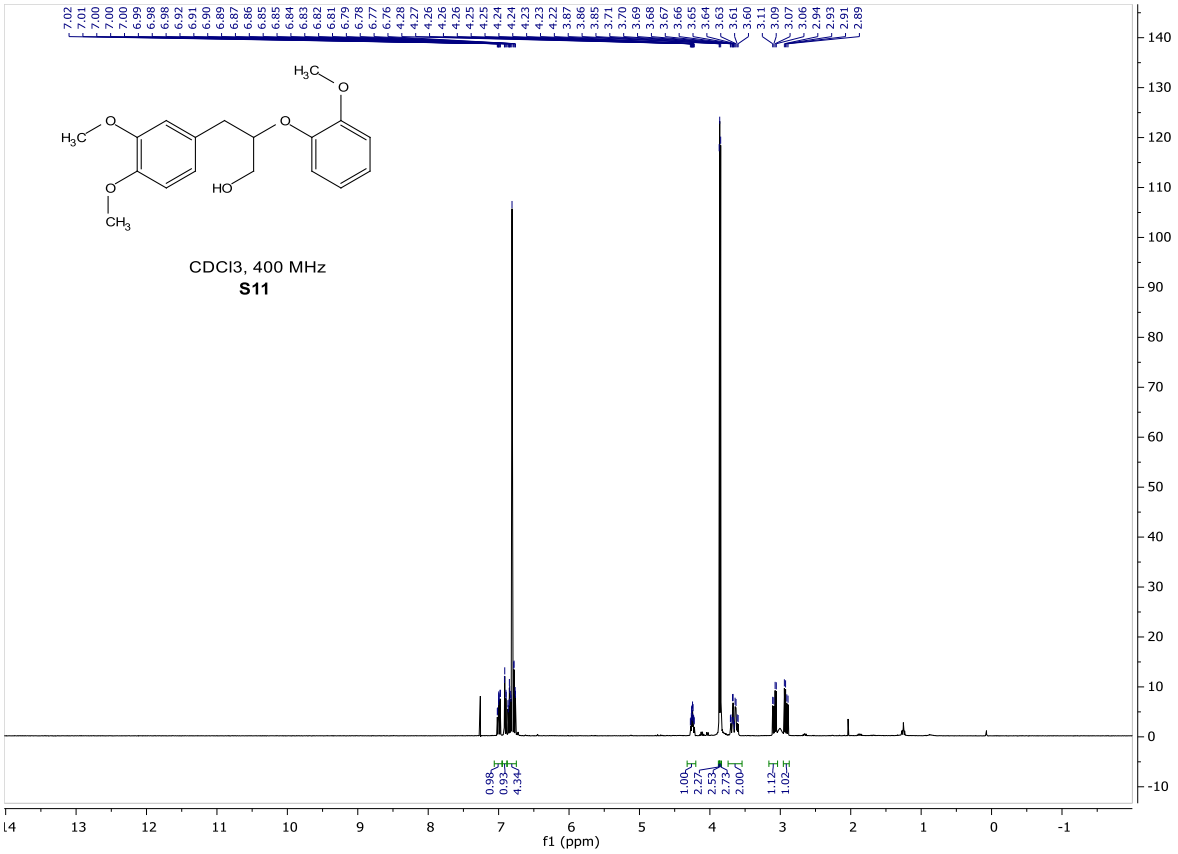

DMSO-d6, 600 MHz
S5
9.03
7.14
6.94
6.94
6.92
6.75
6.74
3.82
1.60
2.00
3.68
1.81
6.11
f1 (ppm)

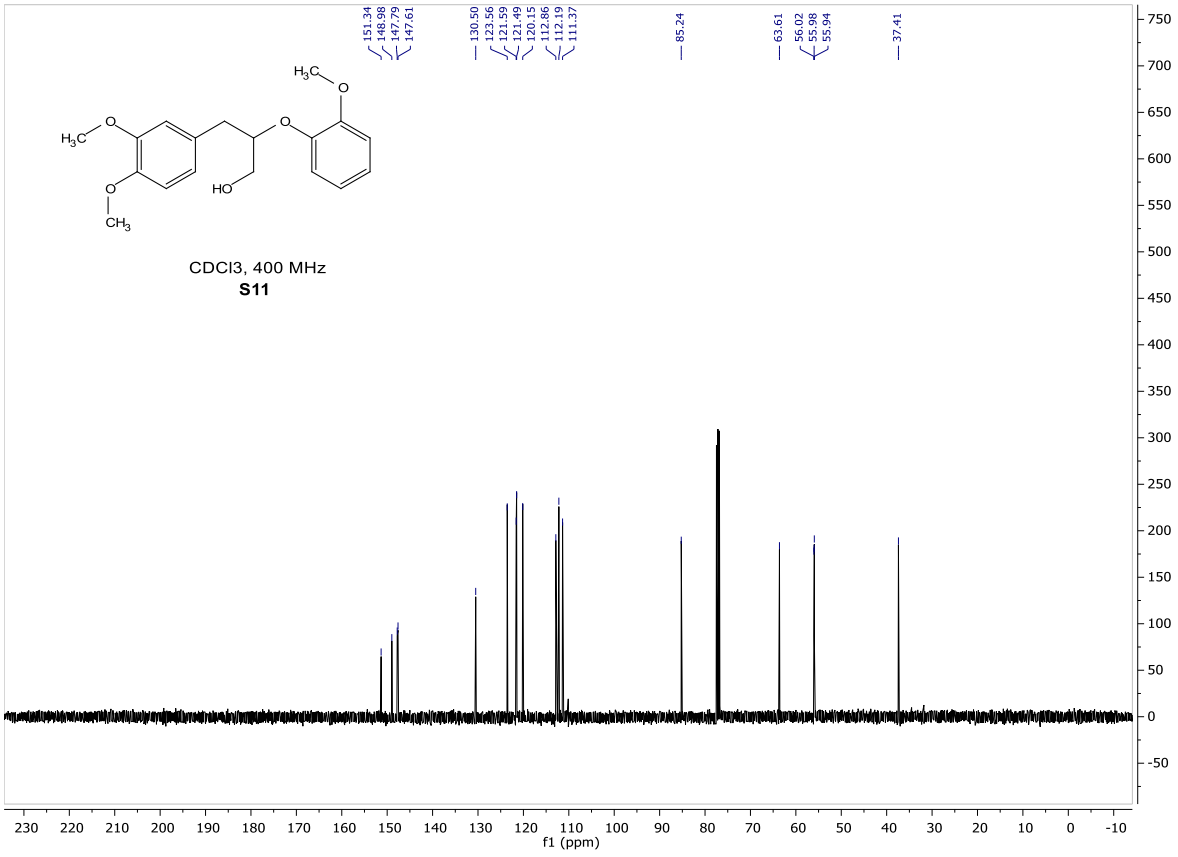


DMSO-d6, 600 MHz
S7



DMSO-d6, 600 MHz
S8


CDCl3, 400 MHz
S11
f1 (ppm)


CDCl3, 400 MHz
S11
f1 (ppm)