Electronic Supplementary Material (ESI) for Chemical Science.

Supplementary Material for:

Establishing Lignin Structure-Upgradeability Relationships Using Quantitative ^{1}H-^{13}C Heteronuclear Single Quantum Coherence Nuclear Magnetic Resonance (HSQC-NMR) Spectroscopy

Masoud Talebi Amiri, Stefania Bertella, Ydna M. Questell-Santiago, and Jeremy S. Luterbacher*

Laboratory of Sustainable and Catalytic Processing, Institute of Chemical Sciences and Engineering. École polytechnique fédérale de Lausanne (EPFL). CH-1015 Lausanne (Switzerland).

*E-mail: jeremy.luterbacher@epfl.ch

S1 Chemicals and Materials

All commercial chemicals were analytical reagents, and were used without further purification. 5% Ru on carbon, decane (>99%), sodium bicarbonate, propionaldehyde (≥98%), a-Cyano-4-hydroxycinnamic acid (>99.0%), trifluoroacetic acid (>99.0%), and dimethyl sulfoxide-d6 (99.9 atom % D) were purchased from Sigma Aldrich. Methanol (>99%) and 1,4-dioxane (99%) were purchased from ABCR GmbH. Fuming hydrochloric acid (37 %) and tetrahydrofuran (>99.0%) were purchased from VWR. The formaldehyde solution (37%) was purchased from Roth AG, 1-(4-hydroxy-3,5-dimethoxyphenyl) ethanone (97%) and 1-(4-hydroxy-3-methoxyphenyl) ethenone (98%) were purchased from ACROS. Kraft lignin (UPM BioPivaTM 100) was purchased from UPM Biochemicals.

Birch wood was procured from Dr. Michael Studer of the Bern University of Applied Sciences. The birch tree (*Betula pendula*, ca. 40 years old) was harvested in May of 2018 in Solothurn, Switzerland. The tree was debarked and the stem (trunk) was converted into wood chips then air-dried at 40 °C for 24 hours. These wood chips were then collected and transported to EPFL where they were sieved and sorted to remove residual bark and leaves. The wood chips were then milled using a 6 mm screen and then machine sieved with a 0.45 mm mesh to remove fines.

This beech wood was procured from Dr. Michael Studer of the Bern University of Applied Sciences. The beech wood (*Fagus sylvatica*) was harvested from Bern, Switzerland in April of 2018 and air-dried at 40 °C for 24 hours. These wood chips were then collected and transported to EPFL where they were sieved and sorted to remove residual bark and

leaves. The wood chips were then milled using a 6 mm screen and then machine sieved with a 0.45 mm mesh to remove fines.

S2 Experimental Methods

S2.1 Polymeric model compounds

The synthesis of the monomers (Figure S1), polymeric model compounds (Figure S2) and acetylation of model compounds (Figure S3) was performed using a slightly modified procedures first described by Kishimoto et al.[1,2] The full procedures and spectroscopic data are reported below.

S2.1.1 Monomer synthesis

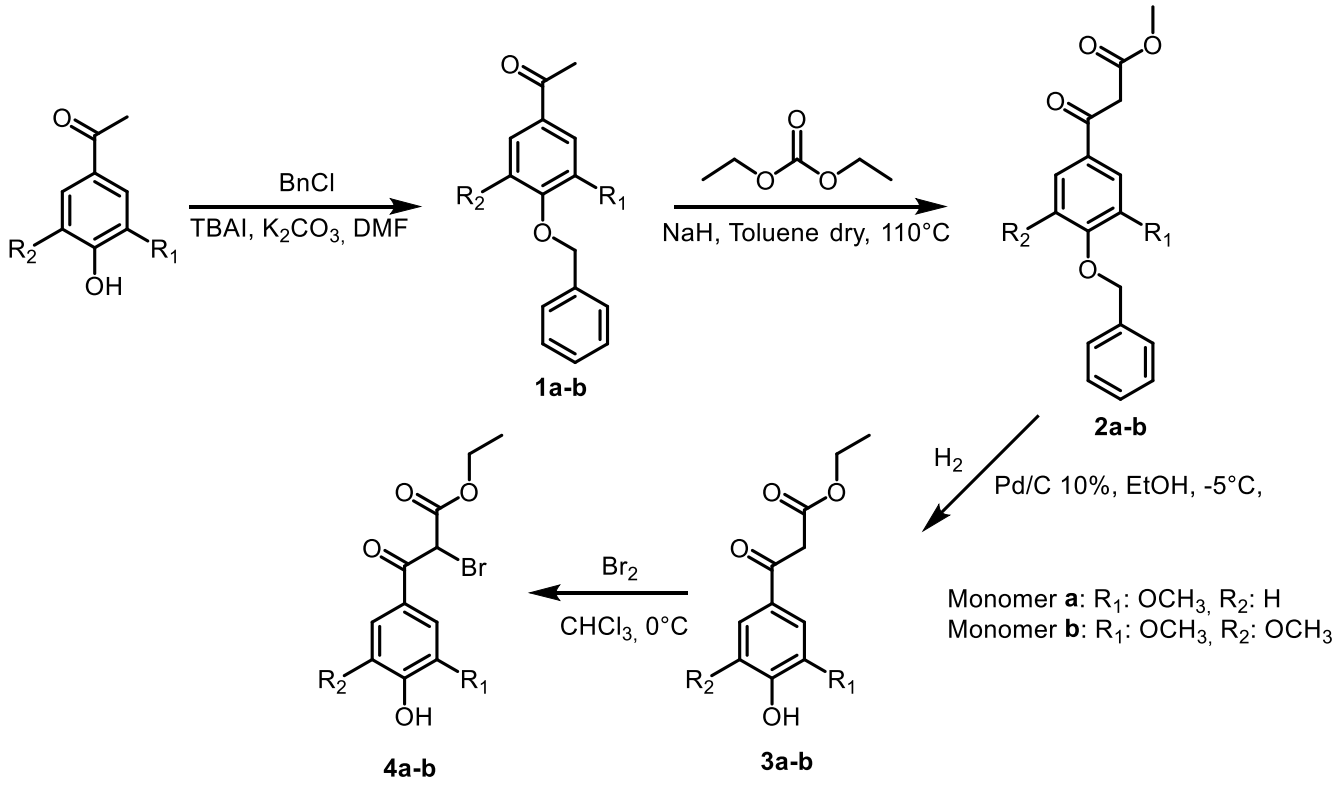


Figure S1 Synthesis of monomers **4a-b**

Synthesis of 1-(4-(benzyloxy)-3-methoxyphenyl)ethanone (1a)

1-(4-hydroxy-3-methoxyphenyl) ethanone (16.6 g, 0.1 mol) was dissolved in 150 ml of dimethylformamide (DMF). To this solution, K_2CO_3 (21 g, 0.15 mol), benzyl chloride (14 ml, 0.12 mol) and tetra-*n*-butyl ammonium Iodide (TBAI) (3.7 g, 0.01 mol) were added. The reaction mixture was stirred at 25 °C for 15 hours, after which 200 ml of ethyl acetate were added. The organic phase was washed with a saturated solution of NaCl in water (3x150 ml). The organic phase was then dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure to obtain a solid that was recrystallized from 20 ml of a solution of ethanol/hexane 1:4 v/v. The product **1a** was obtained as white crystals (22.5g, 87%). ^{1}H-NMR ($CDCl_3$): δ2.56 (s, 3H, Cβ-H), 3.97 (s, 3H, OCH_3), 5.26 (s, 2H, $CH_2C_6H_5$), 6.91 (d, 1H, J = 8.3, C5-H), 7.32–7.57 (m, 7H, Ar-H); ^{13}C-NMR ($CDCl_3$): δ26.2 (Cβ), 56.1 (OCH_3), 70.7 ($CH_2C_6H_5$), 110.5, 112.0, 123.1, 127.2, 128.1, 128.7, 130.8, 136.3, 149.5,152.3 (Ar-H), 196.8 (Cα).

Synthesis of 3-(4-(benzyloxy)-3-methoxyphenyl)-3-oxopropanoate (2a)

NaH 60% w/w dispersed in mineral oil (6.11 g, 0.15 mol) was suspended in 60 ml of dry toluene under N_2 atmosphere. A solution of compound **1a** (12.8 g, 0.05 mol) in 60 ml of dry toluene was added dropwise over a period of 2.5 hours. The reaction was stirred at 110 °C for 30 min and then cooled to 25 °C. Acetic acid (approx. 15 ml) was slowly added while cooling the reaction with an ice bath until neutralization. 70 ml of ethyl acetate were then added to the solution and extracted with a saturated solution of NaCl in water (3x150

ml). The organic phase was dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure. The residue was triturated with 20 ml of hexane to remove mineral oil and recrystallized with a solution of ethanol/hexane 15:1 v/v. The purified product **2a** was recovered as crystals (12.77 g, 78%). ^{1}H-NMR ($CDCl_3$): δ 1.25 (t, 3H, J = 7.1, $OCH_2C\underline{H}_3$), 3.92 (s, 2H, Cβ-$\underline{H}$), 3.94 (s, 3H, $OC\underline{H}_3$), 4.20 (q, 2H, J = 7.1, $OC\underline{H}_2CH_3$), 5.23 (s, 2H, $C\underline{H}_2C_6H_5$), 6.90 (d, 1H, J = 8.4, C5-$\underline{H}$), 7.30–7.55 (m, 7H, Ar-$\underline{H}$); ^{13}C-NMR ($CDCl_3$): δ 14.1 ($OCH_2\underline{C}H_3$), 45.7 ($\underline{C}$β), 56.0 ($O\underline{C}H_3$), 61.4 ($O\underline{C}H_2CH_3$), 70.8 ($\underline{C}H_2C_6H_5$), 110.7, 112.1, 123.3, 127.2, 128.2, 128.7, 129.5, 136.1, 149.7, 152.9 (Ar-H),167.7 ($\underline{C}$γ), 191.0 ($\underline{C}$α).

Synthesis of ethyl 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanoate (3a)

Compound **2a** (8.93 g, 27.2 mmol) was dissolved in 190 ml of ethanol. The solution was cooled to 0 °C under N_2 atmosphere. When the temperature was reached, 10% Pd/C (0.92 g) was added to the solution and the reaction was carried out under H_2 atmosphere for 3 hours. The reaction mixture was then filtered and evaporated under reduced pressure. The crude product was purified over silica gel chromatographic column with ethyl acetate/hexane 2:3 v/v. Product **3a** was recovered as colourless syrup (5.81 g, 89.6%). 1H-NMR ($CDCl_3$): δ 1.27 (t, 3H, J =7.1,$OCH_2C\underline{H}_3$), 3.95 (s, 2H, Cβ-$\underline{H}$), 3.95 (s, 3H, $OC\underline{H}_3$), 4.21 (q, 2H, J = 7.1, $OC\underline{H}_2CH_3$), 6.95 (d, 1H, J = 8.2, C5-$\underline{H}$), 7.50 (dd, 1H, J_1 = 8.2, J_2 = 1.7, C6-$\underline{H}$), 7.55 (d, 1H, J = 1.7, C2-$\underline{H}$); ^{13}C-NMR ($CDCl_3$): δ 14.1 ($OCH_2\underline{C}H_3$), 45.7 ($\underline{C}$β), 56.0 ($O\underline{C}H_3$), 61.5 ($O\underline{C}H_2CH_3$), 110.1, 114.0, 124.1, 129.0, 146.8, 151.0 (Ar-H), 167.8 ($\underline{C}$γ), 191.0 ($\underline{C}$α).

Synthesis of ethyl 2-bromo-3-(4-hydroxy-3-methoxyphenyl)-3-oxopropanoate (4a)

Compound **3a** (6.85 g, 28.8 mmol) was dissolved in 35 ml of chloroform and cooled to -5 °C. A solution of Br_2 (1.55 ml, 30.2 mmol) in 35 ml of chloroform was added over a period of 3 hours. The reaction mixture was then stirred for 15 minutes and then 50 ml of ethyl acetate were added to the solution. The solution was then washed with a saturated solution of NaCl in water (3x100 ml). The organic phase was dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure. The crude product was purified over silica gel chromatographic column with dichloromethane/methanol 99:1 v/v. Compound **4a** was obtained as white crystals (6.70 g, 74.4%). ^{1}H-NMR ($CDCl_3$): δ 1.27 (t, 3H, J = 7.1, $OCH_2C\underline{H}_3$), 3.97 (s, 3H, $OC\underline{H}_3$), 4.29 (q, 2H, J = 7.1, $OC\underline{H}_2CH_3$), 5.66 (s, 1H, Cβ-<u>H</u>), 6.98 (d, 1H, J = 8.5, C5-<u>H</u>), 7.57–7.60 (m, 2H, C2-<u>H</u>, C6-<u>H</u>); ^{13}C-NMR ($CDCl_3$): δ 14.0 ($OCH_2\underline{C}H_3$), 46.2 (<u>C</u>β), 56.2 ($O\underline{C}H_3$), 63.3 ($O\underline{C}H_2CH_3$), 110.0, 114.0, 124.7, 126.1, 146.9, 151.5 (<u>Ar</u>-H), 165.4 (<u>C</u>γ), 186.7 (<u>C</u>α).

Synthesis of 1-(4-(benzyloxy)-3,5-dimethoxyphenyl)ethanone (1b)

1-(4-hydroxy-3,5-dimethoxyphenyl) ethanone (16.2 g, 82 mmol) was dissolved in 120 ml of dimethylformamide (DMF). To this solution, K_2CO_3 (17.1 g, 124 mmol), benzyl chloride (11.4 ml, 98.9 mmol) and tetra-*n*-butyl ammonium Iodide (TBAI) (3 g, 8.2 mmol) were added. The reaction mixture was stirred at 50 °C for 15 hours, then cooled to 25 °C. 180 ml of ethyl acetate were then added. The organic phase was washed with a saturated solution of NaCl in water (3x150 ml). The organic phase was dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure to obtain a solid that was recrystallized from 15 ml of a solution of ethanol/hexane 1:4 v/v. The product **1b** was obtained as white crystals (19.22 g, 81.4%). ^{1}H-NMR ($CDCl_3$): δ 2.58 (s, 3H, Cβ-<u>H</u>), 3.88 (s, 6H, $OC\underline{H}_3$), 5.10 (s, 2H, $C\underline{H}_2C_6H_5$), 7.20 (s, 2H, C2-<u>H</u>, C6-<u>H</u>), 7.26–7.47 (m, 5H, Ar-

H); ^{13}C-NMR ($CDCl_3$): δ 26.7 (Cβ), 56.6 (OCH_3), 75.3 ($CH_2C_6H_5$), 106.2, 128.3, 128.5, 128.7, 132.9, 137.6, 141.8, 153.7 (Ar-H), 197.2 (Cα).

Synthesis of 3-(4-(benzyloxy)-3,5-dimethoxyphenyl)-3-oxopropanoate (2b)

NaH 60% w/w dispersed in mineral oil (6.14 g, 0.15 mmol) was suspended in 70 ml of dry toluene under N_2 atmosphere. A solution of compound **1a** (14.2 g, 49.9 mmol) in 70 ml of dry toluene was added dropwise over a period of 2.5 hours. The reaction was stirred at 110 °C for 30 min and then cooled to 25 °C. Acetic acid (approx. 20 ml) was slowly added while cooling the reaction with an ice bath until neutralization. 100 ml of ethyl acetate were then added to the solution and extracted with a saturated solution of NaCl in water (3x200 ml). The organic phase was dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure. The residue was dissolved in 50 ml of methanol and extracted with hexane (3x40 ml) to remove mineral oil. The methanol was then evaporated at reduced pressure and the product **2b** was obtained pure as a yellow syrup (16.53 g, 92.5%). ^{1}H-NMR ($CDCl_3$): δ 1.26 (t, 3H, J = 7.1, OCH_2CH_3), 3.87 (s, 6H, OCH_3), 3.96 (s, 2H, Cβ-H), 4.21 (q, 2H, J = 7.1, OCH_2CH_3), 5.12 (s, 2H, $CH_2C_6H_5$), 7.21 (s, 2H, C2-H, C6-H), 7.28–7.48 (m, 5H,$CH_2C_6H_5$); ^{13}C-NMR ($CDCl_3$): δ14.1 (OCH_2CH_3), 46.0 (Cβ), 56.2 (OCH_3), 61.5 (OCH_2CH_3), 75.0 ($CH_2C_6H_5$), 106.1, 128.1, 128.2, 128.4, 131.3, 137.3, 142.0, 153.5 (Ar-H), 167.5 (Cγ), 191.3 (Cα).

Synthesis of ethyl 3-(4-(hydroxy)-3,5-dimethoxyphenyl)-3-oxopropanoate (3b)

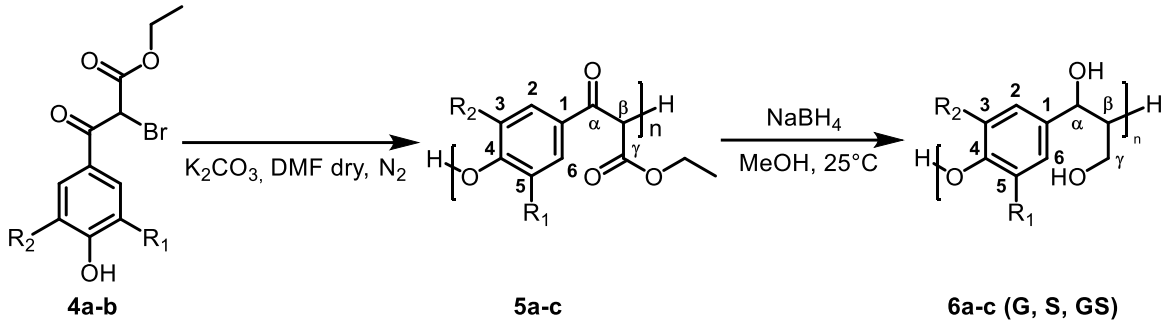
Compound **2b** (7.21 g, 20.1 mmol) was dissolved in 140 ml of ethanol. The solution was cooled to -5 °C under N_2 atmosphere. When the temperature was reached, 10% Pd/C

(0.73 g) was added to the solution and the reaction was carried under H_2 atmosphere for 2 hours. The reaction mixture was then filtered and evaporated under reduced pressure. The crude product was purified over silica gel chromatographic column with ethyl acetate/hexane 1:2 v/v. Product **3b** was recovered as yellow syrup (5.13 g, 95%). ^{1}H-NMR ($CDCl_3$): δ 1.25 (t, 3H, J = 7.1, $OCH_2C\underline{H}_3$), 3.93 (s, 2H, Cβ-H), 3.95 (s, 6H, $OC\underline{H}_3$), 4.20 (q, 2H, J = 7.1, $OC\underline{H}_2CH_3$), 7.24 (s, 2H, C2-H, C6-H); ^{13}C-NMR ($CDCl_3$): δ 14.2 ($OCH_2\underline{C}H_3$), 46.0 (Cβ), 56.5 ($O\underline{C}H_3$), 61.5 ($O\underline{C}H_2CH_3$), 106.1, 127.6, 140.5, 146.9 (Ar-H), 167.8 (Cγ), 190.9 (Cα).

Synthesis of ethyl 2-bromo-3-(4-hydroxy-3-dimethoxyphenyl)-3-oxopropanoate (4b)

Compound **3b** (4.62 g, 17.2 mmol) was dissolved in 20 ml of chloroform and cooled to 0 °C. A solution of Br_2 (0.95 ml, 18.5 mmol) in 20 ml of chloroform was added over a period of 2 hours. The reaction mixture was then stirred for 15 minutes and then 30 ml of ethyl acetate were added to the solution. The solution was then washed with a saturated solution of NaCl in water (3x80 ml). The organic phase was dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure. Compound **4b** was obtained as very viscous dark red syrup and used for polymerization without further purification (5.90 g, 98.6%). ^{1}H-NMR ($CDCl_3$): δ 1.27 (t, 3H, J = 7.1, $OCH_2C\underline{H}_3$), 3.96 (s, 6H, $OC\underline{H}_3$), 4.29 (q, 2H, J =7.1, $OC\underline{H}_2CH_3$), 5.62 (s, 1H, Cβ-H), 7.30 (s, 2H, C2-H, C6-H); ^{13}C-NMR (CDCl3): δ 13.9 ($OCH_2\underline{C}H_3$), 46.5 (Cβ), 56.5 ($O\underline{C}H_3$), 63.3 ($O\underline{C}H_2CH_3$), 106.7, 124.6, 140.8, 146.9 (Ar-H), 165.4 (Cγ), 186.5 (Cα).

S2.1.2 Polymer Synthesis



Monomer **a**, Polymer **a**: R_1: OCH_3, R_2: H
Monomer **b**, Polymer **b**: R_1: OCH_3, R_2: OCH_3
Polymer **c**: R_1: OCH_3, R_2: H (30 mol%), OCH_3 (70 mol%)

Figure S2 Synthesis of polymers G, S, GS (**6a-c**)

Synthesis of polymer 5a

Monomer **4a** (4.86 g, 15.3 mmol) and K_2CO_3 (3.18 g, 23 mmol) were inserted into a 100 ml 3-neck round bottom flask. To these compounds, 25 ml of dry dimethylformamide were added, and the reaction was stirred for 24 hours at 25 °C. The mixture was then poured in 400 ml of water at 0 °C, and the pH adjusted to 2.5 with a solution of HCl 2 M in H_2O. The resulting precipitate was filtered, washed with H_2O and dried at 45 °C in a vacuum oven for 3 days. The polymer **5a** was obtained as a pale-yellow solid (3.45 g, 95%). ^{1}H-NMR (DMSO-d_6): δ 1.10 (t, 3H, J = 6.8, $OCH_2C\underline{H}_3$), 3.78 (s, 6H, $OC\underline{H}_3$), 4.18 (q, 2H, J =6.8, $OC\underline{H}_2CH_3$), 6.66 (s, 1H, Cβ-$\underline{H}$), 7.12 (d, 1H, J =8.6, C5-$\underline{H}$), 7.61 (s, 1H, C2-$\underline{H}$), 7.71

(d, 1H, J =8.6, C6-H); ^{13}C-NMR (DMSO-d_6): δ 14.2 (OCH_2CH_3), 56.5 (OCH_3), 61.8 (OCH_2CH_3), 78.8 (Cβ), 112.5 (C2), 114.2 (C5), 123.9(C6), 128.6 (C4), 149.2 (C3), 151,8 (C1), 166.2 (Cγ), 189.9 (Cα).

Synthesis of polymer 6a (G)

In a 100 ml 2-necks round bottom flask, polymer **5a** (3.13 g, 13.3 mmol) was suspended in 50 ml of CH_3OH. $NaBH_4$ (4.06 g, 106 mmol) was added in very small portions and then the temperature of the reaction was set to 50 °C. The reaction was stirred for 24 hours, after which the mixture was neutralized with acetic acid. The reaction product was then poured in 500 ml of a solution of HCl 0.5 M in H_2O and centrifuged at 3700 rpm for 15 min. The pellet was collected and dried in a vacuum at 45 °C for 1 hour. The solid was the dissolved in 20 ml of dioxane and added dropwise to 400 ml of diethyl ether. The precipitate formed was filtered and dried. Polymer **6a** was obtained as a pale-yellow solid (1.18 g, 45.5%). ^{1}H-NMR (DMSO-d6): 3.47-3.63 (m, 2H, Cγ-H), 3.70-3.78(bs, 3H, OCH_3), 4.28 (s, 1H, Cβ-H), 4.75 (s, 1H, Cα-H), 6.68-7.02 (m, 3H, Ar-H); ^{13}C-NMR (DMSO-d6): δ 55.5 (OCH_3), 59.8 (Cγ), 71.4 (Cα), 83.6 (Cβ), 111.9 (C2), 115.0 (C5), 119.3 (C6), 135.1 (C1), 146.9 (C4), 149.1 (C3).

Synthesis of polymer 5b

Monomer **4b** (5.24 g, 15.1 mmol) and K_2CO_3 (3.12 g, 22.6 mmol) were inserted into a 100 ml 3-necks round bottom flask. To these compounds, 30 ml of dry dimethylformamide were added, and the reaction was stirred for 24 hours at 25 °C. The mixture was then poured in 400 ml of water at 0 °C, and the pH adjusted to 2 with a solution of HCl 2 M in H_2O. The precipitate formed was filtered, washed with H_2O and dried at 45 °C in a vacuum

oven for 2 days. The polymer **5b** was obtained as a pale-yellow solid (3.32 g, 83%). ^{1}H-NMR (DMSO-d_6): δ 1.26 (t, 3H, J = 6.8, $OCH_2C\underline{H}_3$), 3.81 (s, 6H, $OC\underline{H}_3$), 4.30 (q, 2H, J =6.8, $OC\underline{H}_2CH_3$), 5.77 (s, 1H, Cβ-H), 7.50 (s, 2H, C3-H, C5-H); ^{13}C-NMR (DMSO-d_6): δ 14.0 ($OCH_2\underline{C}H_3$), 56.1 ($O\underline{C}H_3$), 62.1 ($O\underline{C}H_2CH_3$), 84.6 (Cβ), 107.2 (C2, C6), 129.9 (C4), 140.1 (C1), 151.5 (C3, C5), 166.8 (Cγ), 198.6 (Cα).

Synthesis of polymer 6b (S)

Polymer **5b** (3.31 g, 12.4 mmol) was suspended in 55 ml of CH_3OH in a 250 ml 2-necks round bottom flask. $NaBH_4$ (3.76 g, 99.4 mmol) was added in very small portions and then the temperature of the reaction was set to 50 °C. The reaction was stirred for 24 hours, after which the mixture was neutralized with acetic acid. The reaction was then poured in 400 ml of a solution of HCl 0.5 M in H_2O and centrifuged at 3700 rpm for 15 min. The pellet was collected and dried in a vacuum at 45 °C for 1 hour. The solid was then dissolved in 15 ml of dioxane and added dropwise to 400 ml of diethyl ether. The precipitate was filtered and dried. Polymer **6b** was obtained as a yellow solid (1.36 g, 48.5%). ^{1}H-NMR (DMSO-d6): 3.23 (bs, 1H, Cγ-H_{threo}), 3.39 (bs, 1H, Cγ-$H_{erithro}$), 3.75 (s, 6H, $OC\underline{H}_3$), 4.02 (s, 1H, Cβ-H), 4.89 (s, 1H, Cα-H), 6.71 (s, 1H, C2-H/C6-H), 6.80 (s, 1H, C2-H/C6-H); ^{13}C-NMR (DMSO-d6): δ 56.3 ($O\underline{C}H_3$), 60.6 (Cγ), 72.6 (Cα), 86.5 (Cβ), 104.2 (C2/C6), 134.1 (C4), 137.2 (C1), 152.4 (C3), 152.5 (C5).

Synthesis of polymer 5c

Monomer **4a** (1.80 g, 5.7 mmol), monomer **4b** (4.61 g, 13.3 mmol) and K_2CO_3 (3.93 g, 28.5 mmol) were inserted into a 250 ml 2-necks round bottom flask. To these, 35 ml of dry dimethylformamide were added, and the reaction was stirred for 24 hours at 25 °C.

The mixture was then poured in 400 ml of water at 0 °C, and the pH adjusted to 1.5-2 with a solution of HCl 2 M in H_2O. The resulting precipitate was filtered, washed with H_2O and dried at 40 °C in a vacuum oven for 24 hours. Polymer **5c** was obtained as a pale-yellow solid (4.09 g, 84%). ^{1}H-NMR (DMSO-d6): δ 1.14 (t, 3H, J = 6.8, $OCH_2C\underline{H}_3$), 3.63-3.81 (m, $OC\underline{H}_3$), 4.17 (m, 2H, $OC\underline{H}_2CH_3$), 6.26 (bs, 1H, Cβ-<u>H</u>), 7.3-7.62 (m, Ar-<u>H</u>); ^{13}C-NMR (DMSO-d6): δ 14.2 ($OCH_2\underline{C}H_3$), 56.2 ($O\underline{C}H_{3b}$), 56.5 ($O\underline{C}H_{3a}$), 62.0 ($O\underline{C}H_2CH_3$), 78.7 (<u>C</u>β$_a$), 84.7 (<u>C</u>β$_b$), 107.2 (<u>C</u>2$_b$, <u>C</u>6$_b$), 112.3 (<u>C</u>2$_a$), 114.4 (<u>C</u>5$_a$), 124.1(<u>C</u>6$_a$), 128.6 (<u>C</u>4$_a$), 130.0 (<u>C</u>4$_b$), 139.2 (<u>C</u>1$_b$), 149.3 (<u>C</u>3$_a$), 151.6 (<u>C</u>3$_b$, <u>C</u>5$_b$), 151.9 (<u>C</u>1$_a$), 166.6 (<u>C</u>γ), 189.9 (<u>C</u>α$_a$), 198.5 (<u>C</u>α$_b$).

Synthesis of polymer 6c (GS)

Polymer **5c** (4.05 g, 15.8 mmol) was suspended in 70 ml of CH_3OH in a 250 ml 2-necks round bottom flask. $NaBH_4$ (4.78 g, 126 mmol) was added in very small portions and then the temperature of the reaction was set to 50 °C. The reaction was stirred for 24 hours, after which the mixture was neutralized with acetic acid. The reaction mixture was then poured in 400 ml of a solution of HCl 0.5 M in H_2O and centrifuged at 3700 rpm for 15 min. The pellet was collected and dried in a vacuum at 45 °C for 1 hour. The solid was then dissolved in 15 ml of dioxane and added dropwise to 400 ml of diethyl ether. The resulting precipitate was filtered and dried. Polymer **6c** was obtained as a yellow solid (1.86 g, 54.4%). 3.23 (bs, Cγ-<u>H</u>$_{\text{threo b}}$), 3.39 (bs, Cγ-<u>H</u>$_{\text{erithro b}}$), 3.55-3.64 (bs, Cγ-<u>H</u>$_a$), 3.74 (bs, $OC\underline{H}_3$), 4.01 (s, Cβ-<u>H</u>$_b$), 4.28 (s, Cβ-<u>H</u>$_a$), 4.75 (s, Cα-<u>H</u>$_a$), 4.89 (s, Cα-<u>H</u>$_b$), 6.71-7.03 (m, Ar-<u>H</u>); ^{13}C-NMR (DMSO-*d*6): δ 55.5 ($O\underline{C}H_{3b}$), 56.0 ($O\underline{C}H_{3a}$), 64.9 (<u>C</u>γ$_a$), 66.3 (<u>C</u>γ$_b$), 71.4 (<u>C</u>α$_a$), 72.2 (<u>C</u>α$_b$), 86.1 (<u>C</u>β$_b$), 86.8 (<u>C</u>β$_a$), 104.2 (<u>C</u>2$_b$/<u>C</u>6$_b$), 111.4 (<u>C</u>2$_a$), 115.4 (<u>C</u>5$_a$),

119.1 (C6$_a$), 134.3 (C4$_b$), 134.7 (C1$_a$), 137.8 (C1$_b$), 146.9 (C4$_a$), 149.2 (C3$_a$), 152.2 (C3$_b$), 152.1 (C5$_b$).

Acetylation of polymers

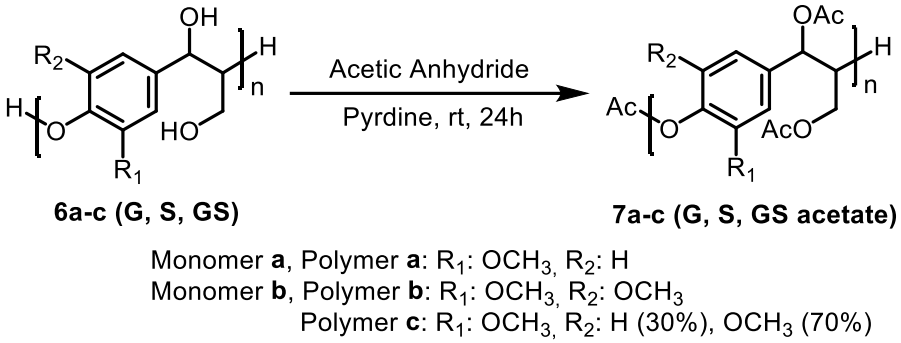


Figure S3 Acetylation of polymers G, S, GS (**7a-c**)

100 mg of polymeric model compound was inserted in a round bottom flask with 4 ml of acetic anhydride and 4 ml of pyridine. The reaction was stirred for 24 hours at 25 °C. The mixture was diluted with 10 ml of ethyl acetate and extracted with a saturated solution of NaCl in water (3x10 ml). The organic phase was dried over anhydrous Na_2SO_4, filtered and evaporated under reduced pressure. The residue was dissolved in 2 ml of dichloromethane, precipitated in 5 ml of hexane and centrifuged at 3700 rpm for 15 min. The pellet was collected and dried in a vacuum at 40 °C for 48 hours to obtain the acetate polymers **7a-c** in quantitative yields.

7a (acetate): ^{1}H-NMR ($CDCl_3$): δ 1.59-2.09 (m, α-OCOC$\underline{H}_3$, γ-OCOC$\underline{H}_3$), 2.86-4.05 (m, Ar-OCOC$\underline{H}_3$, OC$\underline{H}_3$, Cγ-$\underline{H}$), 4.22-4.70 (m, Cβ-$\underline{H}$), 6.28-6.91 (m, Cα-$\underline{H}$), 7.64-8.45 (m, Ar-

H); ^{13}C-NMR ($CDCl_3$): δ 21.0 (OCOCH$_3$), 56.5 (OCH$_3$), 67.3 (Cγ), 70.1 (Cα), 79.5 (Cβ), 123.0 (C2), 125.4 (C5), 127.1 (C6), 131.3 (C1), 143.8 (C4), 149.7 (C3), 160.5 (Ar-OCOCH$_3$), 166.2 (α-OCOCH$_3$), 194.28 (γ-OCOCH$_3$).

7b (acetate): ^{1}H-NMR (DMSO-*d*6): δ 1.82-2.11 (m, α-OCOCH$_3$, γ-OCOCH$_3$, Ar-OCOCH$_3$), 3.75 (s, OCH$_3$), 4.03-4.27 (m, Cγ-H), 4.53-4.63 (m, Cβ-H), 5.94 (s, Cα-H), 6.66 (s, Ar-H); ^{13}C-NMR ($CDCl_3$): δ 20.4, 21.0, 22.5 (OCOCH$_3$), 56.3 (OCH$_3$), 66.8 (Cγ), 74.8 (Cα), 80.5 (Cβ), 104.1, 104.4 (C6, C2), 132.9 (C1), 136.2 (C4), 152.7, 152.9 (C3, C5), 169.8 (Ar-OCOCH$_3$), 170.3 (α-OCOCH$_3$), 170.4, (γ-OCOCH$_3$).

7c (acetate): ^{1}H-NMR (DMSO-*d*6): δ 1.79-2.15 (m, α-OCOCH$_3$, γ-OCOCH$_3$, Ar-OCOCH$_3$), 3.75 (s, OCH$_3$, Ar-OCOCH$_3$), 4.00-4.27 (m, Cγ$_a$-H, Cγ$_b$-H), 4.53-4.89 (m, Cβ$_a$-H, Cβ$_b$-H), 5.92 (bs, s, Cα$_b$-H), 6.66 (s, Ar$_b$-H), 6.71-7.14 (m, Cα$_a$-H, Ar$_a$-H); ^{13}C-NMR (DMSO-*d*6): 19.1, 20.3, 21.03, 22.5, 22.9 (OCOCH$_3$), 56.1 (OCH$_{3b}$), 56.3 (OCH$_{3a}$), 66.8 (Cγ$_a$), 66.9 (Cγ$_b$), 70.6 (Cα$_a$), 74.0 (Cα$_b$), 78.6 (Cβ$_a$), 80.5 (Cβ$_b$), 104.0, 104.8 (C6$_b$, C2$_b$), 123.2 (C2$_a$), 124.9 (C5$_a$), 127.1 (C6$_a$), 130.2 (C1$_a$), 131.1 (C1$_b$), 137.4 (C4$_b$), 152.7, 152.9 (C3$_b$, C5$_b$), 169.8, 170.3, 170.4, 194.3 (Ar-OCOCH$_3$).

S2.1.3 Characterization of the polymeric model compound

Gel Permeation Chromatography

The average molecular weights of polymeric model compounds 6a-c were determined by gel permeation chromatography (GPC), after acetylation. Acetylation was carried out using a solution of acetic anhydride and pyridine for 24 hours at room temperature. GPC analyses were conducted in tetrahydrofuran (1 ml/min @40 °C) using an Agilent 1260

infinity equipped with a refractive index detector and 2x Agilent PL-Gel Mixed C+ guard column set. The calibration was performed with polystyrene standards from 277000 to 1250 Da.

Table S1 GPC analysis of acetylated synthetic polymers **7a-c**

Polymer	Mn	Mw	PD index
7a (G acetate)	6500	9500	1.5
7b (S acetate)	6500	11000	1.7
7c (GS acetate)	5600	9600	1.7

MALDI-TOF-MS of polymeric model compounds

MALDI-TOF-MS spectra of model compounds were recorded using a Bruker AutoFlex Speed instrument (Bremen, Germany). Samples were prepared according to the procedure described by Kosyakov et al.[3] Approximately 10 mg of lignin model compounds were dissolved in 1ml of tetrahydrofuran (THF). A solution of 10 mg of α-cyano-4-hydroxycinnamic acid (CHCA) in 1 ml of THF and 10 µL of trifluoroacetic acid (TFA) was prepared. The final samples were prepared according to the “s-m-s” technique. 1 µL of the solution containing the model compound was loaded on the MALDI plate, 1 µL of the solution containing the matrix was then deposited, and finally 1 µL of the model compound solution was deposited again. In these measurements the laser power was set higher than 50% and lower than 85%.

The recorded spectrum of synthetic polymer G (6a) is shown below and confirms that the polymer features a uniform structure with guaiacyl propane units connected by a β-O-4 linkage.

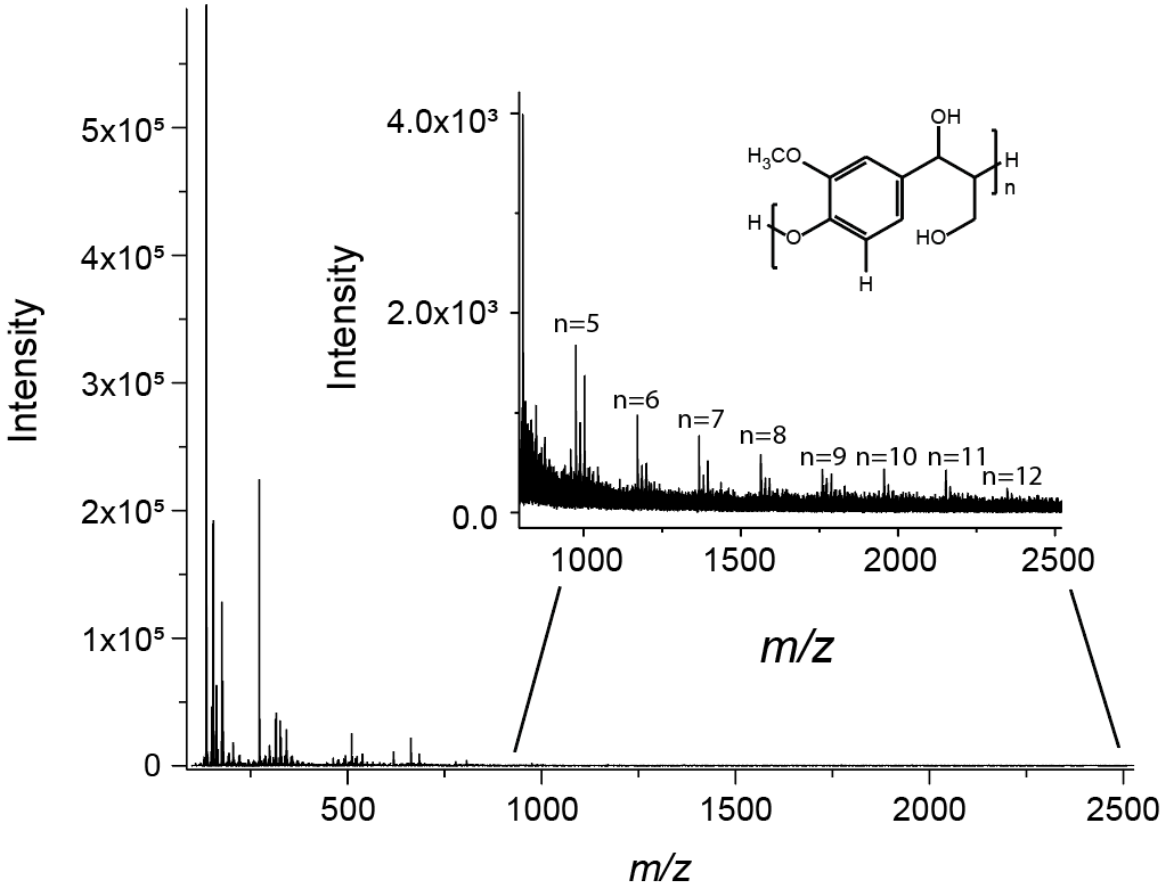


Figure S4 MALDI-TOF-MS spectrum of synthetic polymer G (**6a**)

S2.2 Isolated lignin samples

For lignin isolation, the biomass is either used as is, without removing the extractives, after removal of extractives, after which the biomass is further dried. In cases with extraction, the biomass is washed with an 80 wt% ethanol aqueous solution three times and dried in the vacuum oven at 60 °C and 50 mbar for at least 16 hours. Biomass

extraction is mentioned in the following section detailing each sample preparation and also listed in Table S3.

S2.2.1 Formaldehyde-stabilized lignin (Samples SF1-SF5)

The isolation of formaldehyde-stabilized lignin was performed according to two different approaches based on procedures modified from previously published methods[4,5] to provide a range of ether linkage content within the resulting lignin samples.

For the preparation of the samples SF1, SF2, and SF5, in a 60 ml glass reactor, 1 g of biomass was mixed with 9 ml of 1,4-dioxane, 1 ml of formaldehyde solution (37 wt%), and 420 µl of HCl solution (37 wt%). The reaction was conducted in an oil bath set at 95 ^{o}C and stirred with a stir bar set at 300 rpm for 3.5 hours. After the reaction, the slurry was filtered and washed with 13.5 ml of 1,4-dioxane to separate cellulose. The filtrate was neutralized by addition of a saturated $NaHCO_3$ solution (5 ml). The solvent was then evaporated in rotavap at 40 ^{o}C and 60 mbar pressure. The lignin was precipitated by adding 25 ml of Milli-Q water and stirring at room temperature for 1 hour in the case of SF1 and SF2, and 30 minutes for SF5. The precipitated lignin was recovered by filtration and dried overnight in a desiccator under vacuum.

For the preparation of samples SF3 and SF4, 4.5 g of extracted and dried biomass were mixed with 25 ml of 1,4-dioxane, 5.2 ml of formaldehyde solution (37 wt%), and 2.1 ml of HCl solution (37 wt%) in a 60 ml glass reactor. The reaction was conducted in an oil bath set at 95 ^{o}C and stirred with a stir bar at 300 rpm for 3.5 hours. The reaction solution was swirled by hand every 30 minutes to ensure the homogeneity. After the reaction, the slurry was filtered and washed with 1,4-dioxane and methanol to separate the cellulose. The

filtrate was neutralized by addition of a saturated $NaHCO_3$ solution (35 ml) and precipitated by solvent removal in a rotary evaporator set at 35 ºC and 60 mbar. The precipitated lignin was then separated by filtration and dried overnight in a desiccator under vacuum. The sources of biomass for samples SF3 and SF4 were wild birch and beech, respectively.[4]

S2.2.2 Propionaldehyde stabilized lignin (Samples SP6 and SP7)

Similarly, the isolation of propionaldehyde-stabilized lignin was performed based on procedures modified from previously published methods.[4,5] In a 100 ml round-bottom flask, 4.5 g of extracted and dried biomass was mixed with 25 ml of 1,4-dioxane, 4.8 ml of propionaldehyde, and 0.85 ml of HCl solution (37 wt%). A reflux condenser was used on top of the round-bottom flask to reflux the propionaldehyde. The reaction was conducted in an oil bath set at 85 ºC and stirred with a stir bar set at 300 rpm for 3 hours. After the reaction, the slurry was filtered and washed with 1,4-dioxane and methanol to remove the cellulose. The filtrate was neutralized by addition of $NaHCO_3$ (1.680 g) and stirred for 30 minutes. Afterwards, the excess amount of $NaHCO_3$ and NaCl was removed by filtration followed by washing with 1,4-Dioxane. The filtrate was then concentrated using a rotavap at 40 ºC and 25 mbar final pressure resulting in a dark brown oil. 10 ml of ethyl acetate were added to the oil and the lignin was precipitated by drop-wise addition of this solution to hexanes and separated by filtration. The precipitated lignin was then washed with diethyl ether for full sugar removal and dried using a rotavap at 40 ºC at 25 mbar. The sources of biomass for samples SP6 and SP7 were beech and birch, respectively.

To measure the number of monomer units in a chain of propionaldehyde-stabilized lignin we have performed a GPC analysis on this sample with the same conditions as section S2.1.3. Figure S5 presents the Molecular Weight Distribution plot that is obtained by GPC.

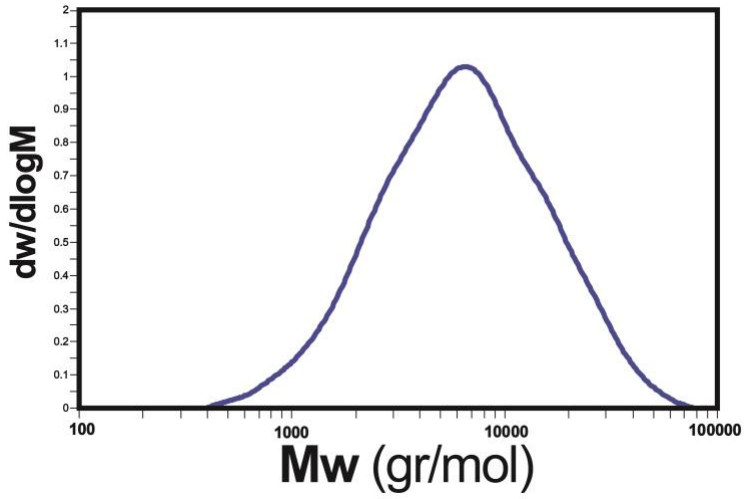


Figure S5 Molecular weight distribution of propionaldehyde-stabilized lignin (SP7)

From this analysis, we obtained a number average molecular weight (Mn) of 4560 gr/mol which indicates that an average oligomer contains approximately 15 monomer units.

S2.2.3 Mild dilute acid-catalyzed (MDAC) lignin extraction (Samples SA8 and SA9)

The procedure for isolating MDAC lignin was also modified from previous work.[6] 5.0 g of biomass and 120 mL of dioxane/H_2O (9:1, v/v) containing 0.2 mol/L HCl were added to a round bottom flask. The mixture was heated to reflux under vigorous stirring for 45 (SA8) min or 150 (SA9) min. Once cooled down to room temperature, the mixture was vacuum-filtered and the filter cake was washed with 50 mL dioxane/H_2O (9:1, v/v). The filtrate was collected and the pH valued was adjusted to 3-4 with a saturated $NaHCO_3$ solution. The

solution was then concentrated at 40 °C under reduced pressure to about 25 mL. The dark brown oil was added slowly to cold water to precipitate the lignin. The resulting powder was collected by centrifugation, washed with H_2O, and dried in a vacuum oven set at 50 °C for 24 h.

S2.2.4 Non-stabilized lignin (SU10)

In a 60 ml glass reactor, 4.5 g of extracted and dried biomass was mixed with 25 ml of 1,4-dioxane, 3.3 ml of MiliQ water, and 2.1 ml of HCl solution (37 wt%). The reaction was conducted in an oil bath set at 95 ^oC and stirred with a stir bar at 300 rpm for 3.5 hours. The reaction solution was swirled by hand every 30 minutes to ensure the homogeneity. After the reaction, the slurry was filtered and washed with 1,4-dioxane and methanol to separate the cellulose. The lignin was precipitated by solvent removal in a rotary evaporator at 35 ^oC and 60 mbar. The precipitated lignin was then separated by filtration and overnight in a desiccator under vacuum.

S2.2.5 Hydrogenolysis of lignin into lignin monomers

200 mg of isolated lignin was added to a 50-mL high-pressure Parr reactor along with 100 mg of catalyst (5 wt% Ru/C) and 20 ml of solvent. In the case of FA, PA, and non-stabilized lignin the solvent was 1,4-Dioxane and, in the case of MDAC lignin, the solvent was methanol.

The reactor was stirred with a magnetic stir bar and heated with high-temperature heating tape (Omega) connected to a variable power supply controlled by a PID temperature controller (Omega) with a K-type thermocouple that measured the reaction temperature

through a thermowell. Once closed, the reactor was purged three times and then pressurized with 40 bar of H_2. The reactor was heated to the desired temperature and then held at that temperature for the specified residence time. After reaction, the reactor was cooled with an external flow of compressed air to room temperature.

S3 Analytical Methods

S3.1 Calculation of lignin monomers yield from hydrogenolysis by GC analysis

After the hydrogenolysis reaction, the reactor was cooled down to room temperature. 200 µl of internal standard solution (2 g of decane in 50 ml 1,4-Dioxane) was added to the reaction mixture and stirred with a spatula for homogeneity. The mixture was then filtered through a syringe filter. A sample of the filtrate was used for analysis with a GC (Agilent 7890B series) equipped with an HP5-column and a flame ionization detector (FID). The injection temperature was 300 °C. The column temperature program was: 40 °C (3 min), 30 °C/min to 100 °C, 40 °C/min to 300 °C and 300 °C (5 min). The detection temperature was 300 °C.

The effectiveness and validity of this method has been shown in our previous work.[5,7] The monomer yield was calculated based on the area of the monomer and the area of decane in the GC chromatogram as previously reported.[5] The detailed calculation is as follows:

$$n_{\mathrm{decane}} = \frac{W_{\mathrm{decane\ in\ sample}}}{MW_{\mathrm{decane}}} \qquad \text{(Equation S1)}$$

$$n_{monomer} = \frac{A_{monomer\ in\ sample}}{A_{decane\ in\ sample}} \times n_{decane} \times \frac{ECN_{decane}}{ECN_{monomer}} \qquad \text{(Equation S2)}$$

In the equations,

$W_{decane\ in\ sample}$ (mg): the weight of decane used as an internal standard in each analyzed sample;

MW_{decane} (mg mmol^{-1}): the molecular weight of decane (142 mg mmol^{-1});

n_{decane} (mmol): the molar amount of decane in each analyzed sample;

$n_{monomer}$ (mmol): the molar amount of monomer in each analyzed sample;

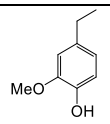
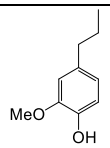
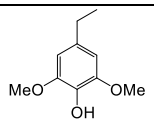
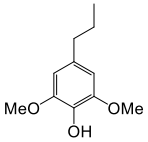
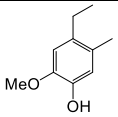
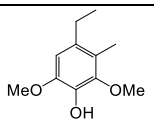
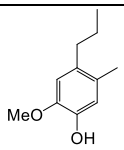
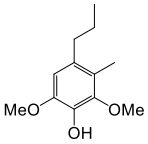
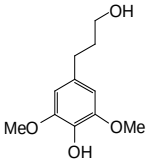
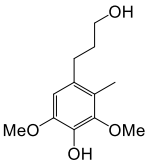
$A_{monomer\ in\ sample}$: the peak area of monomer in the GC-FID chromatogram;

$A_{decane\ in\ sample}$: the peak area of decane in the GC-FID chromatogram;

ECN_{decane}: the effective carbon number (10) of decane;

$ECN_{monomer}$: the effective carbon number of the lignin monomer molecule (Table S2);

Table S2 Effective carbon number (ECN) for lignin monomers

Lignin monomer structure	Effective carbon number calculated based on adjusted ECN rule ($ECN_{monomer}$)
	7
	8
	7
	8
	8
	8
	9
	9
	7.4
	8.4

S3.2 Lignin monomers yield prediction by 2D-HSQC$_0$ NMR

S3.2.1 Preparation of sample for 2D-HSQC$_0$ NMR

Approximately 50 mg sample of isolated lignin was added to NMR tube and 0.7 ml of DMSO-d was added to dissolve the lignin by using a sonication bath and vortex mixer. After making sure that the lignin was fully dissolved, a known amount of the Tetramethylsilane (TMS) was added as an internal standard to the NMR tube. The vortex mixer was used to dissolve the TMS in the solution. NMR spectra were acquired on a Bruker Avance III 500 MHz spectrometer. Table S3 presents the detailed information of the prepared samples.

Table S3 Detailed prepration data of the isolated lignin samples

Sample	Biomass source	Extraction and drying	Isolation method	Lignin sample (mg)	TMS (mg)	Solvent (ml)
SF1	Birch	No	FA	50.0	2.6	0.7
SF2	Birch	No	FA	58.9	2.2	0.7
SF3	Birch	Yes	FA	21.5	4.1	0.7
SF4	Beech	Yes	FA	51.1	2.5	0.7
SF5	Birch	No	FA	50.6	5.3	0.7
SP6	Beech	Yes	PA	47.8	6.0	0.7
SP7	Birch	Yes	PA	53.3	7.6	0.7
SA8	Birch	Yes	MDAC	53.7	3.2	0.7
SA9	Birch	Yes	MDAC	52.6	3.2	0.7
SU10	Birch	Yes	Non-stabilized	51.3	2.9	0.7

S3.2.2 Pulse sequence

The main focus in gradient-selective HSQC is to ignore the signal attenuation during the coherent transfer. The detailed explanation of the pulse sequence was reported previously.[8] In this work, the strengths of the pulsed field gradients applied along the z-

axis are: g_1: 80%, g_2: 20.1%, g_3: 60%, g_4: 15.075%, g_5: 40%, g_6: 10.05% of the maximum of 53 G/cm, each with a duration of 1 ms followed by a 200 µs gradient recovery period.[8]

S3.2.3 Calculation of the moles of chemical groups in the sample

Principally, this method involved performing consecutive 2D HSQC NMR data acquisitions on the same sample, which is shown as $HSQC_i$ (i = 1, 2, 3 …). We performed three data acquisitions for each sample in this study. Each acquisition runs right after the next, but with a different corresponding relaxation time. This produces peak intensities for each chemical group at different relaxation times. By extrapolating the absolute peak intensities and their corresponding sequence (i), it was possible to predict the peak intensities at time zero with the equation below:

$$\ln(V_{i,x}) = \ln(2V_{0,x}) + i * \ln(f_{A,x} * \frac{1}{2}) \qquad \text{(Equation S3)}$$

Where i is the number of the corresponding sequence, $V_{i,x}$ is the absolute intensity (volume) of peak x in the corresponding spectrum i, $f_{A,x}$ is the amplitude attenuation factor specific for peak x, and $V_{0,x}$ is the intensity corresponding to chemical group x in the extrapolated signal. $V_{i,x}$ is obtained by integration of the peak volume in TopSpin 3.5. $V_{0,x}$ is directly proportional to sample concentrations due notably to the elimination of relaxation time effects. By adding a known number of moles of an internal standard to the sample, the number of moles of each chemical group can be calculated with Equation S4.

$$n_x = \frac{V_{0,x}}{C - H_x} * \frac{n_{IS} * C - H_{IS}}{V_{0,IS}} \quad \text{(Equation S4)}$$

Where n_x is the predicted moles of the chemical group in sample, $V_{0,x}$ is the intensity of the chemical group *x* in the sample, $C - H_x$ is the number of effective C—H bonds of the chemical group of interest, n_{IS} is the known number of moles of internal standard in the sample, $C - H_{IS}$ is the number of C—H bonds of internal standard (12 for TMS), and $V_{0,IS}$ is the intensity of the internal standard.

S3.2.4 Prediction of monomer yields and distribution

The gradient-selective HSQC$_0$ was used to quantify the absolute number of moles of protected, unprotected and C—C linkages. Knowing that only protected and unprotected ether bonds can be cleaved by hydrogenolysis, the total number of moles of the protected and unprotected lignin can be added up to estimate the maximum moles of monomers that can be produced by hydrogenolysis by assuming that each ether linkage will result in one monomer produced (Equation S5).

$$n_{\text{Total Monomers}} = n_{\text{Protected}} + n_{\text{Unprotected}} \quad \text{(Equation S5)}$$

As expected from lignin's structure, the predicted number of moles of α, β, and γ in their protected form and the moles of protection group should result in the same number as all of these atoms are part of the same chemical functionality. Therefore, the amount of protected lignin is calculated based on the average of these chemical groups:

$$n_{\text{Protected Lignin}} = \text{Average}(n_{\text{protected},\alpha}, n_{\text{protected},\beta}, n_{\text{protected},\gamma}, n_{\text{protection group}}) \quad \text{(Equation S6)}$$

The same logic applies for calculating the number of moles of the unprotected linkages. However, in the case where no unprotected form of α and/or β peaks were observed in the spectrum, the unprotected portion was assumed to be zero.

$$n_{\text{Unrotected Lignin}} = \text{Average}(n_{\text{unprotected},\alpha}, n_{\text{unprotected},\beta}, n_{\text{unprotected},\gamma}) \quad \text{(Equation S7)}$$

These measurements could also be used to predict the distribution of monomers with various functionalities by quantification of the syringyl and guaiacyl units (along with the amount of their hydroxymethylated form in the case of formaldehyde-stabilized lignin). However, the assigned peaks for syringyl and guaiacyl units (G2 and S2/6 in Figure 3) represent all of these units whether they are in the form of protected, unprotected or linked to lignin by C—C linkages. However, we assume that there is a uniform distribution of these groups across these different units, regardless of their linkages, which is not necessarily true, especially given that G-units may have a greater propensity to condense due to their open positions on the aromatic ring. Nevertheless, this leads to accurate predictions, as discussed in the main manuscript. As a result, the amount of syringyl and guaiacyl units are calculated indirectly. First, we calculate the ratio of the total guaiacyl to syringyl in the lignin named as G/S:

$$G/S = \frac{n_{\text{Total G units}}}{n_{\text{Total S units}}} \quad \text{(Equation S8)}$$

We then use this ratio to calculate the number of each type of monomer:

$$n_{\text{syringyl monomers}} = \frac{n_{\text{Total Monomers}}}{1 + G/S} \quad \text{(Equation S9)}$$

$$n_{\text{guaiacyl monomers}} = n_{\text{Total Monomers}} - n_{\text{monomers from syringyl}} \quad \text{(Equation S10)}$$

As a general rule for the integration of the peaks, the integration boundary is defined by auto integration of the software based on the lowest contour level that does not cause overlapping with neighboring peaks, especially in the case of aldehyde-stabilized lignin where traces of sugars can be found in the spectrum. Because the tale of the neighboring peaks can create large errors in the integration, the oblique integration mode can be used as a guide for setting the contour level to avoid the interference of the tale of neighboring peaks. For example in the case of aldehyde-stabilized lignin, traces of sugars in the isolated lignin interfered with one of the peaks from protected γ. Therefore, only one of the C—H bonds was used for integration of the peak corresponding to the Cγ as explicitly depicted on Figure 3 by the label “Signal used for integration”.

S3.2.5 The effect of t_1 noise and background noise

The t_1 noise is a ridge of noise around large peaks that is parallel to the F_1 axis in two-dimensional spectra. Figure S6 shows the t_1 noise of internal standard (TMS) along the F_1 axis.

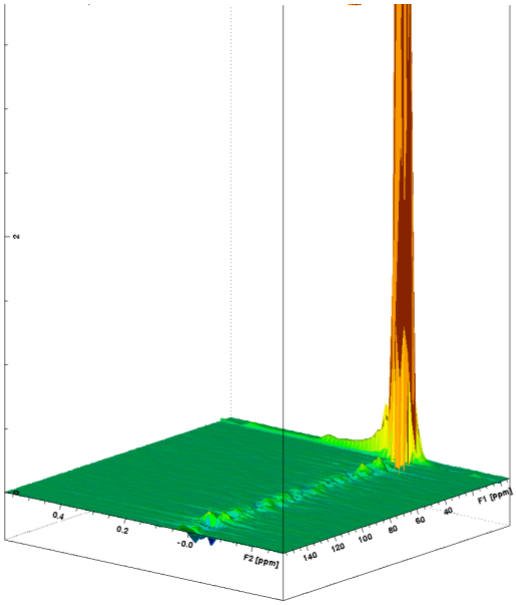


Figure S6 The t_1 noise caused by internal standard (TMS)

The other peaks in the spectra, which were large enough to create significant noise included the solvent (DMSO) and the methoxy group peaks from the lignin structure (Figure S7). Even though the intensity of these two peaks are lower than that of the TMS, they are visible in the spectrum.

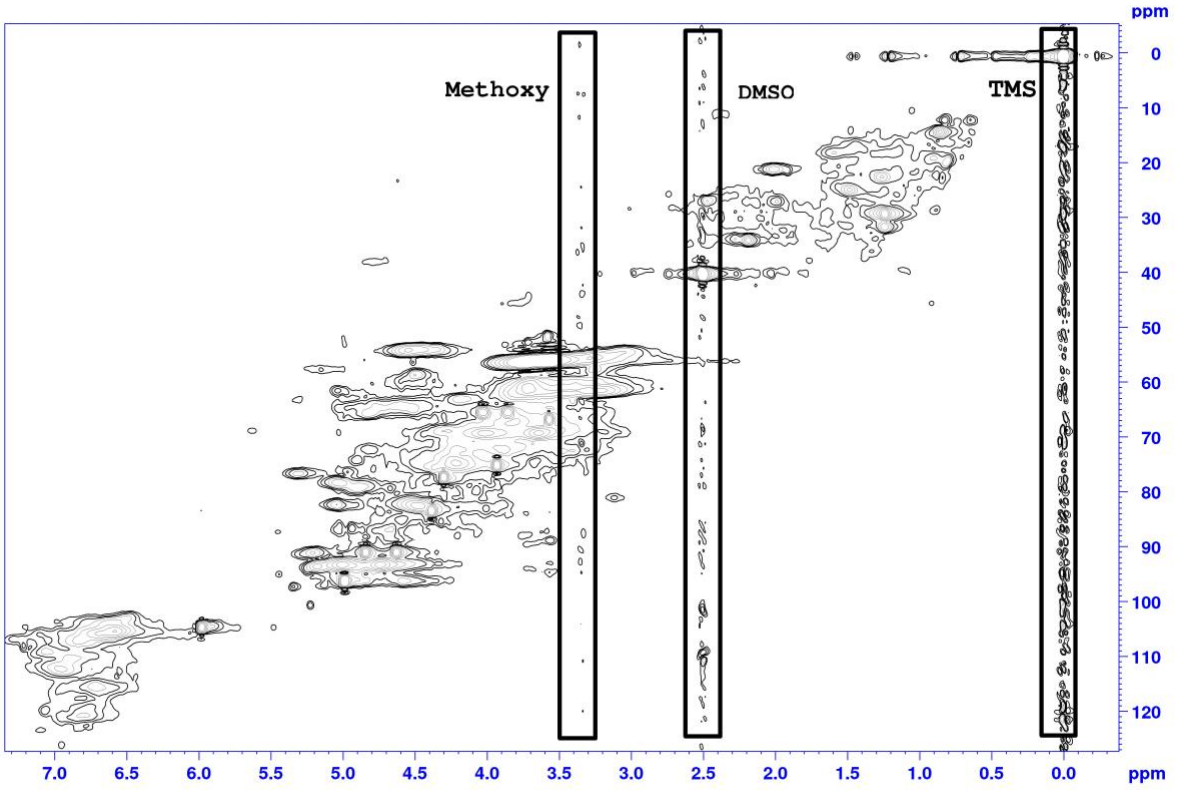


Figure S7 The main sources of t_1 noise in lignin samples

Importantly, the t_1 noise from the methoxy group interferes with one of the C—H bonds of γ carbon. This interference is a source of error in integration of the peak intensity of this chemical group, and for this reason (and as previously mentioned), the integration of the γ position is done only with the peak, for which no overlapping with the t_1 noise occurs (see Figure 3, peak marked with “Signal used for integration”).

The effect of background noise could also be observed for lignin samples with very low contents of ether linkages such as Kraft lignin. We applied our method to Kraft lignin (UPM BioPivaTM 100 provided by UPM Biochemicals). Hydrogenolysis of Kraft lignin resulted

in a yield of 0.85 % of monomers. However, the accurate integration of peaks in the NMR spectrum was difficult and we observed wide variations (up to 400%) in the quantification of functionalities that should have led to similar results (such as the α, β or γ signals, which supposedly belong to the same linkage). These variations were likely due to the very small intensity of most peaks compared to the noise levels and the overlap with other unknown structures in Kraft lignin. We concluded that an accurate quantification of the different functionalities could not be made by NMR for Kraft lignin.

S3.2.6 Monomer yield prediction for formaldehyde-stabilized lignin

Considering the negligible amount of unprotected lignin, the predicted total amount of monomers was calculated directly from the amount of protected β—O—4 linkages:

$$n_{\text{Total Monomers}} = \text{Average}(n_{MA_\alpha}, n_{MA_\beta}, n_{MA_\gamma}, n_{MA_1}) \quad \text{(Equation S11)}$$

The peak that is used for calculation of MA_γ is marked with a hashed circle in Figure 3 labeled “Signal used for integration”. The prediction of the monomer distribution in the case of formaldehyde-stabilized was not straight forward due to the overlap of the peaks for syringyl units with its hydroxymethylated form (S2/6 and S2 in Figure 3d). The different number of effective C—H bonds for these peaks creates more complexity as there is 1 C—H bond for S2 and 2 for S2/6. However, the peak of the guaiacyl unit does not overlap with its hydroxymethlated structure and they both have the same effective C—H bonds number that equals 1. On the other hand, the total amount of hydroxymethylated units (syringyl and guaiacyl) can be predicted by integration of the HM peak HM (see Figure 3d). Therefore, we first calculated the amount of hydroxymethylated (HM) units:

$$n_{\text{HM units}} = \text{Eq. S3}\{V_{i,x} = V_{HM}\}, \text{Eq. S4}\{C - H_x = 1\} \quad \text{(Equation S12)}$$

And the number of hydroxymethylated guaiacyl units were calculated by integration of the G2 peak (marked in green in Figure 3d):

$$n_{\text{HM Guaiacyl}} = \text{Eq. S3}\{V_{i,x} = V_{G2(\text{green})}\}, \text{Eq. S4}\{C - H_x = 1\} \quad \text{(Equation S13)}$$

And finally, the amount of hydroxymethylated syringyl were calculated as shown below:

$$n_{\text{HM Syringyl}} = n_{\text{HM units}} - n_{\text{HM Guaiacyl}} \quad \text{(Equation S14)}$$

The total amount of guaiacyl units in the lignin structure were calculated as shown below:

$$n_{\text{Total Guaiacyl}} = \text{Eq. S3}\{V_{i,x} = V_{G2(\text{green})} + V_{G2(\text{blue})}\}, \text{Eq. S4}\{C - H_x = 1\} \quad \text{(Equation S15)}$$

To calculate the total amount of syringyl units, we first integrated the peak of syringyl units and its hydroxymethylated structure together:

$$n_{S2/6,S2} = \text{Eq. S3}\{V_{i,x} = V_{S2/6} \,\&\, V_{S2}\}, \text{Eq. S4}\{C - H_x = 1\} \quad \text{(Equation S16)}$$

Then the total number of syringyl units was calculated by subtracting the number of hydroxymethylated syringyl units:

$$n_{\text{Total Syringyl}} = \frac{(n_{S2/6,S2} - n_{\text{HM Syringyl}})}{2} + n_{\text{HM Syringyl}} \qquad \text{(Equation S17)}$$

The G/S ratio was then calculated as described below:

$$\frac{G}{S} = \frac{n_{\text{Total Guaiacyl}}}{n_{\text{Total Syringyl}}} \qquad \text{(Equation S18)}$$

The amount of syringyl and guaiacyl units were then calculated as previously described, based on the equations S9 and S10, respectively.

Table S4 Number of effective C—H bonds for integration of peaks for formaldehyde-stabilized lignin

Peak annotation*	Effective C—H bonds
TMS	12
HM	2
MA_{γ}	1
MA_{α}	1
MA_{β}	1
MA_{1}	2
S2/6	2
G2 (Green)	1
G2 (Blue)	1

* Based on Figure 3d

S3.2.7 Monomer yield prediction for propionaldehyde-stabilized lignin

In the case of propionaldehyde-stabilized lignin, there were three peaks that were identified as part of the protection group which are shown as PA_1, PA_2, and PA_3 (Figure 3e). However, PA_2 and PA_3 are not propotional to the quantity of protection group. Similar functionality in the structure of the lignin overlaps with these peaks, which can cause error

in the integration (this mainly the case for PA_2). Therefore, the calculation of the number of protected β—O—4 linkages ignores these two signals and uses:

$$n_{Protected} = Average(n_{PA_\alpha}, n_{PA_\beta}, n_{PA_\gamma}, n_{PA_1}) \quad \text{(Equation S19)}$$

The number of unprotected β—O—4 linkages was calculated with:

$$n_{Unprotected} = Average(n_{A_\alpha}, n_{A_\beta}, n_{A_\gamma}) \quad \text{(Equation S20)}$$

And the number of resinol linkages (C—C linked) was calculated with:

$$n_{Resinol} = Average(n_{C_\alpha}, n_{C_\beta}, n_{C_\gamma}) \quad \text{(Equation S21)}$$

The total number of monomers was then calculated using the equation S5. The total number of syringyl and guaiacyl units was calculated based on the integration of the peaks S2/6 and G2 in Figure 3e. The monomer distribution was then predicted based on equations S8 to S10.

Table S5 Number of effective C—H bonds for integration of peaks for propionaldehyde-stabilized lignin

Peak annotation*	Effective C—H bonds
TMS	12
C_γ	2
C_α	1
C_β	1
A_γ	1
A_α	1
A_β	1
PA_γ	1
PA_α	1
PA_β	1
PA_1	1
S2/6	2
G2	1

* Based on Figure 3e

S3.2.8 Monomer yield prediction for mild dilute-acid catalyzed and non-stabilized lignin

Taking into account that there is no protected linkages in these samples, the total number of monomers was calculated based on the number of unprotected β—O—4 linkages:

$$n_{\text{Total Monomers}} = n_{\text{unprotected}} = \text{Average}(n_{A_\alpha}, n_{A_\beta}, n_{A_\gamma}) \quad \text{(Equation S21)}$$

The amount of resinol linkages (C—C linked) and the monomer distribution were calculated as described in previous section for propionaldehyde-protected lignin.

Table S6 Number of effective C—H bonds for integration of peaks for MDAC and non-stabilized lignin

Peak annotation*	Effective C—H bonds
TMS	12
C_γ	2
C_α	1
C_β	1
A_γ	1
A_α	1
A_β	1
S2/6	2
G2	1

* Based on Figure 3f and 3g

References

[1] T. Kishimoto, Y. Uraki, M. Ubukata, *Org. Biomol. Chem.* **2008**, *6*, 2982–2987.
[2] T. Kishimoto, Y. Uraki, M. Ubukata, *J. Wood Chem. Technol.* **2008**, *28*, 97–105.
[3] D. S. Kosyakov, N. V. Ul'yanovskii, E. A. Sorokina, N. S. Gorbova, *J. Anal. Chem.* **2014**, *69*, 1344–1350.
[4] M. T. Amiri, G. R. Dick, Y. M. Questell-Santiago, J. S. Luterbacher, *Nat. Protoc.* **2019**, *14*, 921.
[5] L. Shuai, M. T. Amiri, Y. M. Questell-Santiago, F. Héroguel, Y. Li, H. Kim, R. Meilan, C. Chapple, J. Ralph, J. S. Luterbacher, *Science* **2016**, *354*, 329–333.
[6] A. Das, A. Rahimi, A. Ulbrich, M. Alherech, A. H. Motagamwala, A. Bhalla, L. da Costa Sousa, V. Balan, J. A. Dumesic, E. L. Hegg, et al., *ACS Sustain. Chem. Eng.* **2018**, *6*, 3367–3374.
[7] W. Lan, M. T. Amiri, C. M. Hunston, J. S. Luterbacher, *Angew. Chem.* **n.d.**, *130*, 1370–1374.
[8] K. Hu, J. J. Ellinger, R. A. Chylla, J. L. Markley, *Anal. Chem.* **2011**, *83*, 9352–9360.