Article

NMR assignment for diaryl ether structures (4–O–5 structures) in pine wood lignin

Yanding Li, Takuya Akiyama, Tomoya Yokoyama, and Yuji Matsumoto

Biomacromolecules, **Just Accepted Manuscript** • DOI: 10.1021/acs.biomac.6b00257 • Publication Date (Web): 14 Apr 2016

NMR assignment for diaryl ether structures (4–O–5 structures) in pine wood lignin

Yanding Li,[†] Takuya Akiyama,[,†,‡] Tomoya Yokoyama,[†] Yuji Matsumoto[†]*

[†]Wood Chemistry Laboratory, Department of Biomaterial Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan

[‡]Japan Science and Technology Agency (JST), PRESTO, Kawaguchi, Saitama 332-0012, Japan

Keywords: diaryl ether, *gymnosperm*, lignins, 4–O–5 linkage, plant cell wall, softwood

Abstract

A 4–O–5-tetramer lignin model compound carrying β–O–4 linkages on each of the side-chain moieties was synthesized, as well as 4–O–5-coupled dehydrodiconiferyl alcohol. By comparison with their NMR data, two cross-signals in the HSQC spectrum of pine milled wood lignin recorded in DMSO-*d6* were assigned to H2/C2 and H6/C6 correlations on the aromatic rings of 4–O–5-linked units. Although the H2/C2 correlation peak appeared in the same region as syringyl units, nitrobenzene oxidation of the pine lignin did not yield any syringyl type product, but did release a 4–O–5-type product.

Introduction

Lignin is a natural aromatic polymer that is abundant in the secondary cell walls of woody plants (around 20-35 wt% of wood).[1] Lignins are derived from three monolignols, *p*-coumaryl, coniferyl, and sinapyl alcohols, that give rise to *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units during their polymerization process by radical coupling reactions followed by rearomatization steps. Whereas softwood species contain lignins primarily composed of G units together with much lower amount of H units, hardwood lignins additionally contain S units in different amounts in various wood species.[1] In the radical coupling step, G units can be coupled at the 5-position on the aromatic ring to form β–5-, 5–5- and 4–O–5-linked units, none of which can be formed solely from syringyl units. Consequently, the variation in syringyl/guaiacyl compositions observed in different wood lignins is reflected in the composition of different interunit linkage types in many cases.[2-5]

The 5-linked G unit is a key to understanding the polymer structure of lignins. Lignins are considered to be not linear but branched polymers as biphenyl (5–5-linked) structures and diaryl ether (4–O–5-linked) structures can be found in chemical degradation products of lignin.[6-11] They are recognized as candidate linkages for branch-points in a lignin polymer.[12] As well as the other major units in lignin such as those containing β–O–4, β–5, β–β linkages, all of which have been evidenced by both chemical degradation and NMR methods, the existence of 5–5 and 4–O–5 units also have been investigated by both analytical approaches.[13-16]

Among the major structural units in lignin, β–O–4, β–5, β–β linkages are readily differentiated by using short-range $^{1}H–^{13}C$ correlation NMR spectroscopy (heteronuclear single-quantum correlation (HSQC) experiment), in which the correlation peaks from their side-chain parts are well separated from each other, especially correlation peaks at β- and α-positions. A large

fraction of 5–5 units also can be identified from correlation peaks in the side-chain region (aliphatic region) of the HSQC spectrum as dibenzodioxocin structures that have an eight membered ring involving 5–5 linkage.[14-15] Although the atoms involved in the 5–5 linkage itself do not show a correlation peak in the HSQC spectrum due to the lack of protons attached directly to the carbons at the 5-positions, dibenzodioxocins exhibit two particular correlation peaks in the side-chain region that originate from the α- and β-positions in the dibenzodioxocin ring.

Whereas the structural evidence has been obtained for the biphenyl structure, limited information is available for the diaryl ether structure (4–O–5 linkage). As well as the 5–5 linkages, 4–O–5 linkages also potentially can be branch-points in lignin, and are therefore crucial structures to understand the shape of the polymer. Although 4–O–5-linked products have been obtained from lignin by chemical degradation methods, e.g., permanganate oxidation,[9] thioacidolysis,[10] and DFRC methods,[11] 4–O–5 units have not been fully assigned in the 2D HSQC spectra of lignins.

The 4–O–5 units are different from the other unit in that their correlation peaks are not expected to appear in the aliphatic region (side-chain region) of lignin HSQC spectra. However, this structure potentially can be identified in the aromatic region if crosspeaks from aromatic rings involved in 4–O–5 linkages show correlation peaks distinguishable from the others. This methodology has been proposed in a detailed 2D NMR study of pine lignin by the comparison of the chemical shifts of 4–O–5-model compounds carrying simple side-chain structures.[16] In the study, the author carefully avoided the assignment of candidate signals on the HSQC spectrum considering the overlapping signals from possible syringyl units in the lignin and the imperfect matching of chemical shifts between the lignin and the model compounds.

In the present study, a 4–O–5-dimer model compound and a tetrameric 4–O–5-linked dimeric model carrying β–O–4 linkages on each of the side-chain moieties were synthesized for the structural elucidation of softwood lignin. We investigated whether any cross-signal in the HSQC spectra of the softwood lignin can be assigned for 4–O–5 units based on the comparison with the NMR data of the model compounds. The result of the assignment is discussed together with nitrobenzene oxidation products of the lignin and the model compounds.

Experimental Section

General

Dioxane was distilled over Na metal. All other chemicals were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan), Tokyo Chemical Industry, Co., Ltd. (Tokyo, Japan) or Sigma-Aldrich (Tokyo, Japan), and used without further purification. The NMR spectra of synthesized compounds were taken on a JEOL JNM-A500 500 MHz spectrometer. The standard JEOL programs of one- and two-dimensional (proton, carbon, DEPT-135, ^{1}H–^{1}H COSY, ^{1}H–^{13}C HSQC, and ^{1}H–^{13}C HMBC) NMR experiments were used for the structural assignments of newly synthesized compounds. Electrospray ionization (ESI) mass spectra were recorded on a time-of-flight mass spectrometer (JEOL, AccuTOF JMS-T100LC). The NMR spectra of lignin samples were acquired by a Bruker Avance 600 MHz spectrometer fitted with a 5 mm TCI gradient cryoprobe. The central peak of the residual solvent was used as an internal reference (δ_H 7.26, δ_C 77.0 ppm for $CDCl_3$, δ_H 2.04, δ_C 29.8 ppm for acetone-*d*$_6$, δ_H 2.49, δ_C 39.5 ppm for DMSO-*d*$_6$). The traditional numbering system for lignins[17] was followed rather than the systematic IUPAC numbering scheme.

Isolation of milled wood lignins

Milled wood lignin (MWL) was isolated according to the Björkman method,[18] with minor modifications in the following procedure. Wood meal (80 mesh pass) from Japanese red pine (*Pinus densiflora* Siebold & Zucc.) and birch (*Betula maximowiczii* Regel) were prepared from the sapwood part of their wood stem disks by using a Wiley mill, and pre-extracted with ethanol/benzene (1:2, v/v) using a Soxhlet apparatus.

The pre-extracted pine wood meal (20 g for each run) was placed in a zirconium dioxide jar (500 mL) with ZrO_2 balls (10 mm diameter, 80 pieces), and ground further in a ball mill (planetary mill; Pulverisette 5, Fritsch Japan Co., Ltd.; 300 rpm) in an environment at 4 °C for 11 h: the milling was interrupted after every 15 min for 15 min to avoid excessive heat development. The ball-milled wood (BM) was extracted twice with with dioxane/water (96:4, v/v, 4 mL per 1 g BM) overnight at r.t. The extracts were combined, and freeze-dried to give crude MWL (11.5 g, 3.8 wt% yield from BM). The crude MWL dissolved in a small amount of dioxane/water (96:4, v/v, 2 mL per 1 g of crude MWL) was poured into 100 times the volume of water, and then filtered through a hydrophilic PTFE membrane filter (0.2 μm pore size; Advantec Toyo Kaisha, Ltd., Japan). The residue was dissolved again in a small amount of dioxane/water (96:4, v/v), then poured into a 10 times volume of diethyl ether to precipitate purified pine MWL (7.9 g, 74.3 wt% yield from the crude MWL). A part of the pine MWL (400 mg) was acetylated overnight at r.t. with Ac_2O (2 mL) and pyridine (6 mL).

Crude birch MWL was isolated from the pre-extracted wood meals, and purified in the same manner. The purified birch MWL was acetylated with Ac_2O and pyridine.

Alkaline Nitrobenzene oxidation

Alkaline nitrobenzene oxidation (NBO)[19] of pine MWL (10 mg) or model compound (5 μmol) was carried out using nitrobenzene (0.4 mL) and 2M NaOH (7 mL) at 170 ºC for 2 h as described previously.[8] The NBO-products were analyzed as trimethylsilyl derivatives by gas chromatography-mass spectroscopy (GC-MS, Shimadzu GC2010/PARVUM2, IC-1 column) equipped with a fused-silica capillary column (IC-1, 0.25 mm i.d. × 30 m, 0.4 μm thickness, GL Science Inc) to identify the products by the comparison with the peak retention time and mass spectra of the authentic compounds (vanillin, vanillic acid, syringaldehyde, syringic acid, compound **2**, and compound **3**). The identified products were quantified by gas chromatography with flame ionization Detector (GC-FID, Shimadzu GC-2014) using the same column and the following calibration curves: $Y_V = 1.0957\ X_V + 0.0298$ ($R^2 = 1.00$) ($0.55 < X_V < 5.65$), $Y_{VA} = 0.08395\ X_{VA} + 0.0035$ ($R^2 = 0.99$) ($0.0045 < X_{VA} < 0.608$), $Y_{4–O–5\text{-}VVA} = 0.71483\ X_{4–O–5\text{-}VVA} + 0.00864$ ($R^2 = 0.99$) ($0.007 < X_{4–O–5\text{-}VVA} < 0.13$). Where "Y" is the molar ratio of each authentic compound to EV, and "X" is the GC area ratio of each authentic compound to EV. Initial column temperature: 150 ºC (held for 25 min), raised at 3 ºC/min to 190 ºC, raised at 10 ºC/min to 230 ºC (held for 20 min), raised at 3 ºC/min to 280 ºC (held for 15 min). Total running time 94 min. Retention times: 8.2 min, vanillin (V); 10.6 min, 3-ethoxy-4-hydroxybenzaldehyde (EV, internal standard); 15.9 min, syringaldehyde (S); 23.0 min, vanillic acid (VA); 33.9 min, syringic acid (SA); 56.7 min, compound **2** (4–O–5-diV); 68.5 min, compound **3** (4–O–5-VVA).

HSQC experiments of lignin samples

Acetylated or non-acetylated milled wood lignin prepared from pine wood or birch wood (50 mg) was dissolved in 0.6 mL of $CDCl_3$, acetone-d_6, or DMSO-d_6. In the HSQC experiments, 1846 data points were acquired from 10.5 to -0.5 ppm in F2 (^{1}H), with an acquisition time of 140

ms, and from 200 to 0 ppm in F1 (^{13}C) with 512 increments, 16-48 scans, and a 1.0 s interscan delay, with a total acquisition time of 2 h 39 min to 7 h 55 min depending on the number of scans used. Processing the final matrix to 2 k by 1 k data points was performed by means of a 90-degree-shifted squared sine-bell in both F2 and F1. Correlation peaks appearing in the aliphatic region of the HSQC spectrum of acetylated MWL ($CDCl_3$) were assigned based on the NMR data of acetylated lignin model compounds ($CDCl_3$) reported in the NMR database of lignin model compounds[20] and previous studies.[14, 21-23] The # refers to the library number of the NMR database[20]: in the case of β-O-4 (#3, #74, #29, #97, #98, #214, Hauteville et al.,[21] and Sipilä et al.[22]), β-β (#109 and #123), β-5 (#2005 and Li et al.[23]), and dibenzodioxocin (#278 and Karhunen et al.[14]).

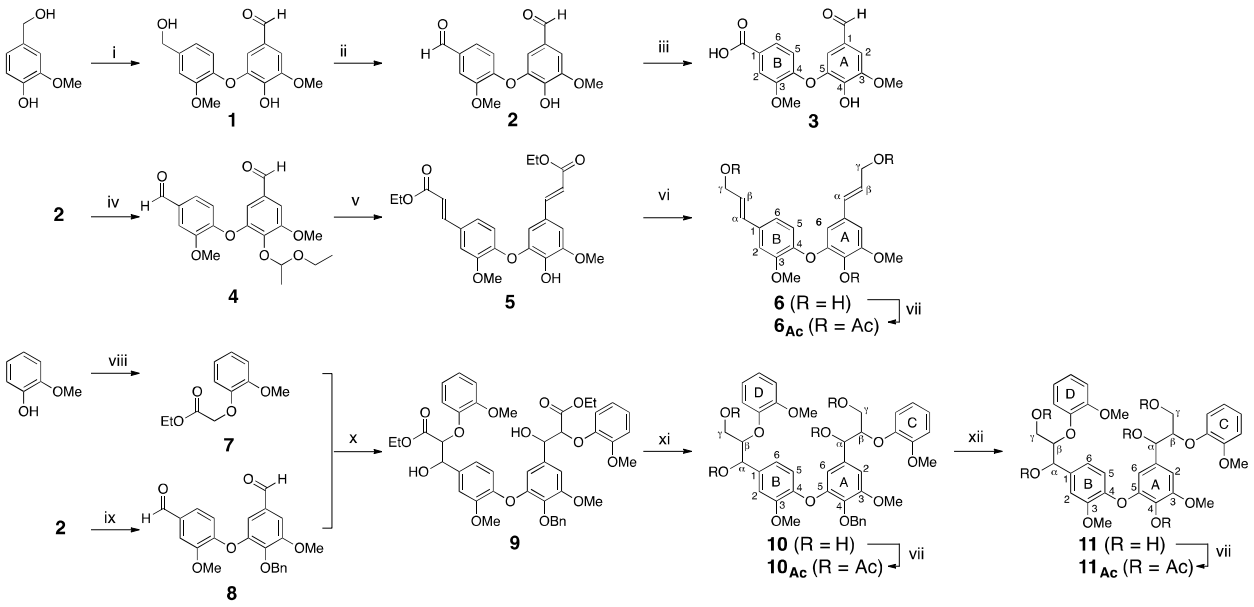


Figure 1. Synthetic route of the 4–O–5 lignin model compounds. i) Ag_2O; ii) DDQ; iii) Ag_2O, NaOH; iv) ethyl vinyl ether, PPTS; v) triethyl phosphonoacetate, NaH; vi) DIBAL-H; vii) Ac_2O, Py; viii) ethyl chloroacetate, K_2CO_3; ix) BnBr, K_2CO_3; x) LDA; xi) $NaBH_4$; xii) H_2, Pd/C.

Synthesized model compounds

As shown in Figure 1, 4–O–5-coupled dehydrodiferulic acid **5** was chemically synthesized starting from vanillyl alcohol *via* 4–O–5-coupled dehydrodivanillin **2** as described previously.[24] The 4–O–5-dimer model **6** was prepared by the reduction of compound **5** with DIBAL-H.[25] Compound **2** was oxidized to 4–O–5-coupled dehydrovanillin-vanillic acid **3** using silver oxide (I).[26] The 4–O–5-tetramer models **10** and **11** were synthesized by Nakatsubo's method[27] starting from compound **2** and ethyl 2-methoxyphenoxyacetate **7** as described below. Compounds **6**, **10**, and **11** were acetylated at r.t. with Ac_2O and pyridine.

4–O–5-coupled dehydrovanillin-vanillic acid (3)

To a stirred solution of compound **2** (3.02 g, 10 mmol) in 2.4 M NaOH (50 mL), Ag_2O (2.78 g, 12 mmol)[26] was added slowly. The suspension was stirred vigorously at room temperature for 30 min. The reaction progress was monitored by TLC (hexane-EtOAc, 1:2, with one drop AcOH). The reaction mixture was filtrated off through a Buchner funnel. The filtrate was acidified with 2 M HCl and stirred for 10 min. The resulting suspension was filtered to collect the insoluble crude product. The crude product was purified by crystallization from hexane-acetone (3:1) to afford compound **3** as white powder (0.31 g, 30%). ^{1}H-NMR (DMSO-d_6, 500 MHz) δ_H: 3.86 (3H, s, B3-OMe), 3.90 (3H, s, A3-OMe), 6.76 (1H, d, J = 8.0 Hz, B5), 7.11 (1H, s, A6), 7.35 (1H, s, A2), 7.50 (1H, d, J = 8.0 Hz, B6), 7.60 (1H, s, B2), and 9.73 (1H, s, A-CHO). ^{13}C-NMR (DMSO-d_6, 125 MHz) δ_C: 55.78 (B3-OMe), 56.24 (A3-OMe), 107.89 (A2), 113.36 (B2), 115.62

(A6), 116.65 (B5), 122.87 (B6), 126.03 (B1), 127.53 (A1), 143.18 (A5), 144.37 (B4), 149.28, 149.42 and 149.52 (A4, A3 and B3), 166.87 (B-COOH), and 190.93 (A-CHO).

4–O–5-coupled dehydrodiconiferyl alcohol (6)

A 200 mL dried flask equipped with a stirrer bar was charged with compound **5** (1.33 g, 3 mmol), and sealed with a rubber cap. Anhydrous toluene (50 mL) was injected through the septum. To the stirred solution over an ice bath at 0 ºC, 1.5 M diisobutylaluminium hydride (DIBAL-H)[25] in toluene (18 mmol) was added slowly by using a syringe over 30 min. The reaction mixture was stirred for additional 30 min. The reaction progress was monitored by TLC (hexane:EtOAc, 1:3) The resulting suspension was diluted with cold EtOAc (50 mL) and crushed ice (100 mL), and then acidified with 4 M HCl until the aqueous layer was pH<2. The organic layer was collected, and filtered through a silica gel column to remove insoluble materials. The aqueous layer was extracted with EtOAc (3 × 50 mL). These organic layers and filtrate above were combined, washed with brine several times until the pH of the aqueous layer was over 5, dried over Na_2SO_4, concentrated under the vacuum. The residue was purified by silica gel chromatography with hexane-EtOAc (1:3) to afford compound **6** as pale yellow oil (0.96 g, 90%). Mass (EI, 70eV) calculated for $C_{29}H_{46}O_6Si_3$ 574 (TMS derivative), found 574 *m*/*z*. ^{1}H-NMR (acetone-d_6, 500 MHz) δ_H: 3.73 (0.8H, broad-t, J = 5.5 Hz, Aγ-OH), 3.85 (4H, s, B3-OMe and Bγ-OH), 3.89 (3H, s, A3-OMe), 4.14 (2H, d, J = 5.0 Hz, Aγ), 4.22 (2H, d, J = 5.5 Hz, Bγ), 6.15 (1H, dt, J = 16.0, 5.5 Hz, Aβ), 6.34 (1H, dt, J = 16.0 Hz, 5.5 Hz, Bβ), 6.41 (1H, d, J = 16.0 Hz, Aα), 6.51 (1H, s, A6), 6.57 (1H, d, J = 16.0 Hz, Bα), 6.78 (1H, d, J = 8.0 Hz, B5), 6.87 (1H, s, A2), 6.93 (1H, d, J = 8.0 Hz, B6), 7.18 (1H, s, B2), and 7.67 (0.6H, br-s, A4-OH). ^{13}C-NMR (acetone-d_6, 125 MHz) δ_C: 56.24 (B3-OMe), 56.55 (A3-OMe), 63.20 (Aγ and Bγ), 105.95 (A2),

110.67 (A6), 111.41 (B2), 119.49 (B5), 119.97 (B6), 128.95 (Aβ), 129.32 (A1), 129.61 (Bα), 129.95 (Aα), 130.25 (Bβ), 134.32 (B1), 138.26 (A4), 145.45 (A5), 146.52 (B4), 149.75 (A3), and 151.58 (B3). After adding 1 drop of D_2O, the peak at δ_H: 3.73 and 7.67 disappeared. The integration of δ_H: 3.85 decreased from 4H to 3H.

Compound **6** in DMSO-*d*$_6$: ^{1}H-NMR (DMSO-*d*$_6$, 500 MHz) δ_H: 3.81 (3H, s, B3-OMe), 3.82 (3H, s, A3-OMe), 4.02 (2H, broad-s, Aγ), 4.10 (2H, br-s, Bγ), 4.73 (0.8H, br-s, Aγ-OH), 4.83 (0.8H, br-s, Bγ-OH), 6.11 (1H, dt, *J* = 16.0, 4.5 Hz, Aβ), 6.31 (1H, dt, *J* = 16.0, 4.5 Hz, Bβ), 6.34 (1H, d, *J* = 16.0 Hz, Aα), 6.41 (1H, s, A6), 6.49 (1H, d, *J* = 16.0 Hz, Bα), 6.63 (1H, d, *J* = 8.5 Hz, B5), 6.84 (1H, s, A2), 6.89 (1H, d, *J* = 8.5 Hz, B6), 7.16 (1H, s, B2), and 8.80 (0.8H, br-s, A4-OH). ^{13}C-NMR (DMSO-*d*$_6$, 125 MHz) δ_C: 55.68 (B3-OMe), 56.01 (A3-OMe), 61.52 and 61.54 (Aγ and Bγ), 105.27 (A2), 109.78 (A6), 110.39 (B2), 117.82 (B5), 118.97 (B6), 127.75 (A1), 128.20 (Bα), 128.46 (Aβ), 128.56 (Aα), 129.81 (Bβ), 132.65 (B1), 137.06 (A4), 144.23 (A5), 145.16 (B4), 149.08 (A3), and 149.94 (B3). After adding 1 drop of D_2O, the peak at δ_H: 4.73, 4.83 and 8.80 disappeared. The peak shape at δ_H: 4.02 and 4.10 became doublet with a coupling constant *J* = 4.5 Hz.

Acetates of compound **6** (**6Ac**) in $CDCl_3$: ^{1}H-NMR ($CDCl_3$, 500 MHz) δ_H: 2.05 (3H, s, Aγ-OCOMe), 2.08 (3H, s, Bγ-OCOMe), 2.22 (3H, s, A4-OCOMe), 3.82 (3H, s, B3-OMe), 3.83 (3H, s, A3-OMe), 4.63 (2H, dd, *J* = 6.5, 1.0 Hz, Aγ), 4.70 (2H, dd, *J* = 6.5, 1.0 Hz, Bγ), 6.09 (1H, dt, *J* = 16.0, 6.5 Hz, Aβ), 6.21 (1H, dt, *J* = 16.0, 6.5 Hz, Bβ), 6.43 (1H, d, *J* = 2.0 Hz, A6), 6.46 (1H, d, *J* = 16.0 Hz, Aα), 6.60 (1H, d, *J* = 16.0 Hz, Bα), 6.70 (1H, d, *J* = 2.0 Hz, A2), 6.87 (1H, d, *J* = 7.5 Hz, B5), 6.90 (1H, dd, *J* = 7.5, 1.5 Hz, B6), and 6.99 (1H, d, *J* = 2.0 Hz, B2). ^{13}C-NMR ($CDCl_3$, 125 MHz) δ_C: 20.34 (Aγ-OCO*Me*), 20.92 (Bγ-OCO*Me*), 20.96 (A4-OCO*Me*), 56.06 (OMe), 56.17 (OMe), 64.65 (Aγ), 64.96 (Bγ), 104.72 (A2), 108.99 (A6), 110.59 (B2), 119.86

(B6), 120.63 (B5), 122.84 (Bβ), 123.88 (Aβ), 130.24 (A4), 133.19 (B1), 133.31 (Aα), 133.67 (Bα), 134.61 (A1), 145.10 (B4), 150.28 (A5), 151.03 (B3), 152.62 (A3), 168.37 (A4-O*CO*Me), 170.73 (Aγ-O*CO*Me), and 170.82 (Bγ-O*CO*Me).

Acetates of compound **6** (**6Ac**) in acetone-d_6: ^{1}H-NMR (acetone-d_6, 500 MHz) δ_H: 1.99 (3H, s, Aγ-OCOMe), 2.03 (3H, s, Bγ-OCOMe), 2.18 (3H, s, A4-OCOMe), 3.83 (3H, s, B3-OMe), 3.87 (3H, s, A3-OMe), 4.63 (2H, dd, J = 6.5, 1.0 Hz, Aγ), 4.70 (2H, dd, J = 6.5, 1.0 Hz, Bγ), 6.25 (1H, dt, J = 16.0, 6.5 Hz, Aβ), 6.35 (1H, dt, J = 16.0, 6.5 Hz, Bβ), 6.49 (1H, d, J = 2.0 Hz, A6), 6.56 (1H, d, J = 16.0 Hz, Aα), 6.69 (1H, d, J = 16.0 Hz, Bα), 6.89 (1H, d, J = 9.0 Hz, B5), 6.96 (1H, d, J = 1.5 Hz, A2), 7.02 (1H, dd, J = 8.5, 2.0 Hz, B6), and 7.26 (1H, d, J = 2.5 Hz, B2). ^{13}C-NMR (acetone-d_6, 125 MHz) δ_C: 20.18 (Aγ-OCO*Me*), 20.72 (Bγ-OCO*Me*), 20.77 (A4-OCO*Me*), 56.31 (B3-OMe) and 56.59(A3-OMe), 64.93 (Aγ), 65.21 (Bγ), 105.43 (A2), 109.12 (A6), 111.83 (B2), 120.55 (B6), 121.63 (B5), 124.44 (Bβ), 125.43 (Aβ), 130.95 (A4), 133.33 (Aα), 133.76 (Bα), 134.74 (B1), 135.72 (A1), 145.67 (B4), 151.51 (A5), 152.35 (B3), 153.90 (A3), 168.38 (A4-O*CO*Me), 170.66 (Aγ-O*CO*Me), and 170.73 (Bγ-O*CO*Me).

Benzyl ether of 4–O–5-coupled dehydrodivanillic acid (8)

To a stirred suspension of compound **2** (3.02 g, 10 mmol), K_2CO_3 (2.78 g, 20 mmol) in DMF (20 mL), benzyl bromide (1.3 mL) was added dropwise. The reaction mixture was kept stirring at room temperature for 1.5 h. The progress of the reaction was monitored by TLC (Hexane-EtOAc, 2:1). The reaction mixture was diluted with water (50 mL), and extracted with CH_2Cl_2 (3 × 50 mL). The combined organic layer was washed with water (20 mL), brine (20 mL), dried over Na_2SO_4, and concentrated under vacuum. The residue was purified by silica gel chromatography with hexane-EtOAc (4:1) to afford compound 8 as white powder (3.60 g, 92%).

^{1}H-NMR (acetone-d_6, 500MHz) δH: 3.93 (3H, s, OMe), 4.01 (3H, s, OMe), 5.16 (2H, s, PhCH_2), 6.97 (1H, d, J = 9.0 Hz, B5), 7.17 (1H, s), 7.24-7.36 (5H, PhCH$_2$), 7.45 (1H, s), 7.49 (1H, d, J = 8.0 Hz, B6), 7.58 (1H, s, B2), 9.87 (1H, s, CHO), and 9.93 (1H, s, CHO).

Compound 9

An oven-dried 200 mL three-necked flask equipped with a stirrer bar was sealed with rubber caps, equipped with a nitrogen-gas-filled balloon. The flask was evacuated and filled with nitrogen gas. To the lithium diisopropylamide[27] (2 M LDA in THF-heptane-ethylbenzene from Aldrich Co., 12 mmol) was added through the septum by using a syringe, and diluted with anhydrous THF (6 mL). To the stirred suspension under a cold bath at -78 °C, a solution of ethyl 2-methoxyphenoxyacetate **7** (2.31 g, 11 mmol)[27] in anhydrous THF (10 mL) was added drop wise over a period of 1 h, and kept stirring for another 30min at -78 °C. To the suspension, a solution of compound **8** (1.96 g, 5 mmol) in anhydrous THF (20 mL) was added through a syringe during 1h. The reaction mixture was kept stirring for another 2 h at this temperature. The progress of the reaction was monitored by TLC (CH_2Cl_2-EtOAc, 4:1). Water (60 mL) was added dropwise to the reaction mixture, and organic layer was collected. The aqueous layer was extracted with EtOAc (3 × 50 mL). All organic layers were combined, washed with 4 M HCl (20 mL), water (60 mL), brine (60 mL), dried over Na_2SO_4, and then concentrated under vacuum. The residue was purified by silica gel chromatography with hexane-EtOAc (3:2) at multiple times to afford two fractions composed of the mixture of diastereomers of compound **9** (major fraction 2.8 g, minor fraction 0.9 g, total 91% yield). Each fraction contained two diastereomers of the theoretically possible 8 diastereomers for compound **9** (two diastereomers for each *erythro-erythro*, *threo-erythro*, *threo-erythro*, and *threo-threo* isomers). The two diastereomers

in each fraction exhibited almost identical peaks in the ^{1}H-NMR spectrum, but some carbon chemical shifts of the isomers' peaks were slightly different in the ^{13}C-NMR spectrum.

Compound **9** (major isomers): ^{1}H-NMR (acetone-d_6, 500 MHz) δ_H: 1.13~1.20 (6H, m, Aγ- and Bγ-COOCH$_2$*CH$_3$*), 3.70/3.71 (3H, s, OMe of two isomers), 3.76 (3H, s, OMe), 3.80 (3H, s, B3-OMe), 3.89 (3H, s, A3-OMe), 4.02~4.17 (4H, m, Aγ- and Bγ-COO*CH$_2$*CH$_3$), 4.61 (1H, dd, *J* = 6.5, 1.5 Hz, Aβ), 4.71 (1H, dd, *J* = 7.0, 1.0 Hz, Bβ), 4.87 (0.8H, d, *J* = 5.0 Hz, Aα-OH), 4.90 (0.8H, d, *J* = 4.5 Hz, Bα-OH), 4.97 (1H, dd, *J* = 6.5, 5.0 Hz, Aα), 5.03 (2H, s, Ph*CH$_2$*), 5.10 (1H, dd, *J* = 7.0, 4.5 Hz, Bα), 6.61/6.62 (1H, d, *J* = 1.5 Hz, A6), 6.77 (1H, m, B5), 6.73~6.81 and 6.91~6.94 (4H+4H, m, C- and D-rings), 7.01 (1H, s, B6), 7.04 (1H, br-s, A2), 7.03~7.29 (3H, m, *Ph*CH$_2$), 7.34/7.35 (1H, d, *J* = 1.5 Hz, B2), and 7.43~7.44 (2H, m, *Ph*CH$_2$). ^{13}C-NMR (acetone-d_6, 125 MHz) δ_C: 14.46 (COOCH$_2$*CH$_3$*) and 14.50 (COOCH$_2$*CH$_3$*), 56.24 (OMe), 56.28 (OMe), 56.37 (B3-OMe), 56.47 (A3-OMe), 61.28 (Aγ- and Bγ-COO*CH$_2$*CH$_3$), 74.60 (Aα and Bα), 75.30 (Ph*CH$_2$*), 83.65 and 83.68 (Aβ and Bβ), 107.47 (A2), 110.68/110.72 (A6 of two isomers), 113.28/113.31 (B2), 113.82 and 113.89 (C2 and D2), 117.54, 117.58, 117.60 and 117.63 (C- or D-ring), 119.44/119.47 (B5), 120.44/120.45 (B6), 121.52, 121.61, 123.61 and 123.65 (C- or D-ring), 128.32, 128.82 and 128.91 (*Ph*CH$_2$), 137.91 (A1), 138.29/138.31 (B1), 139.04 (A4), 139.23 (*Ph*CH$_2$), 146.25/146.27 (B4), 148.26 and 148.40 (C4 and D4), 151.26 and 151.29 (C3 and D3), 151.32 (B3), 151.34 (A5), 154.67 (A3), 170.14 (Aγ-*CO*OEt), and 170.26 (Bγ-*CO*OEt). MS (ESI) calculated for $C_{45}H_{48}O_{14}Na$ $[M+Na]^+$ 835.29, found 835.29.

Compound **9'** (minor isomers): ^{1}H-NMR (acetone-d_6, 500 MHz) δ_H: 1.07 (3H, t, *J* = 7.0 Hz, Aγ-COOCH$_2$*CH$_3$*), 1.18 (3H, t, *J* = 6.5 Hz, Bγ-COOCH$_2$*CH$_3$*), 3.74 (3H, s, OMe), 3.76 (3H, s, OMe), 3.81 (3H, s, B3-OMe), 3.87 (3H, s, A3-OMe), 4.01 (2H, m, Aγ-COO*CH$_2$*CH$_3$), 4.14 (2H, q, *J* = 7.0 Hz, Bγ-COO*CH$_2$*CH$_3$), 4.64 (1H, m, Aβ), 4.71 (1H, m, Bβ), 4.80 (0.3H, d, *J* = 5.0 Hz,

Aα-OH), 4.92 (0.8H, d, *J* = 5.5 Hz, Bα-OH), 4.99 (1H, d, *J* = 5.5 Hz, Aα), 5.02 (2H, s, Ph*CH*$_2$), 5.08 (1H, d, *J* = 7.0 Hz, Bα), 6.57 (1H, d, *J* = 2 Hz, A6), 6.75 (1H, d, *J* = 8.0 Hz, B5), 6.78~6.83 and 6.90~6.95 (4H+4H, m, C- and D-rings), 6.98 (1H, d, *J* = 2 Hz, A2), 7.01 (1H, dd, *J* = 8.0, 1.5 Hz, B6), 7.21~7.29 (3H, m, *Ph*CH$_2$), 7.34 (1H, d, *J* = 1.5 Hz,, B2), and 7.41~7.42 (2H, m, *Ph*CH$_2$). ^{13}C-NMR (acetone-d_6, 125 MHz) δ_C: 14.38 (Aγ-COOCH$_2$*CH$_3$*), 14.50 (Bγ-COOCH$_2$*CH$_3$*), 56.02 (OMe), 56.31 (OMe), 56.34 (B3-OMe), 56.49 (A3-OMe), 61.27 (Aγ-COO*CH$_2$*CH$_3$), 61.31 (Bγ-COO*CH$_2$*CH$_3$), 74.40/74.56 (Aα of two isomers), 75.12/75.26 (Bα), 75.27 (Ph*CH$_2$*), 83.64 (Aβ), 83.68 (Bβ), 107.29 (A2), 110.62 (A6), 113.22 (B2), 113.85 (C2 and D2), 117.48, 117.51 and 117.71 (C- or D-ring), 119.33 (B5), 120.41 (B6), 121.58, 123.57 and 123.62 (C- or D-ring), 128.34, 128.82 and 128.92 (*Ph*CH$_2$), 136.74 (A1), 138.27 (B1), 139.09 (A4), 139.15 (*Ph*CH$_2$), 146.21 (B4), 148.37 (D4), 148.65 (C4), 151.12 and 151.25 (C3 and D3), 151.25 (B3), 151.37 (A5), 154.75 (A3), 169.69 (Aγ-*CO*OEt), and 170.25 (Bγ-*CO*OEt). MS (ESI) calculated for $C_{45}H_{48}O_{14}Na$ $[M+Na]^+$ 835.29, found 835.32.

Benzylated 4–O–5-tetramer model (10)

To a stirred solution of major isomers of compound **9** (2.5 g, 3.01 mmol) in THF (30 mL) under an iced bath at 0 ºC, excess of $NaBH_4$ (1.13 g, 30 mmol, 10 eq.)[28] was added slowly over 5 min. After 55 min, the ice-water bath was removed and the reaction suspension was stirred at room temperature for 1 day. The completion of the reaction was monitored by TLC (EtOAc, 100%). Water (30 mL) and AcOH (10 mL) was added under an ice-water bath to quench the excess of $NaBH_4$. The reaction mixture was extracted with EtOAc (3 × 50 mL). The combined organic layers were washed with saturated $NaHCO_3$ (20 mL), water (30 mL), brine (30 mL), and concentrated under vacuum. The residue was separated on silica-gel chromatography with

hexane-EtOAc (1:9) to obtain a fraction containing compound **10**. The product was further purified by preparative TLC (1 mm thickness, $CHCl_3$-EtOAc, 1:2, 5 time developments) to afford compound **10** (0.47 g, 21%) as colorless syrup. ^{1}H-NMR (DMSO-d_6, 500 MHz) δ_H: 3.57~3.66 (4H, m, Aγ and Bγ), 3.62 (3H, s, OMe), 3.68 (3H, s, OMe), 3.70 (3H, s, B3-OMe), 3.78(3H, s, A3-OMe), 4.21 (1H, m, Aβ), 4.34 (1H, m, Bβ), 4.61~4.67 (3H, m, Aγ-OH, Bγ-OH and Aα), 4.80 (1H, m, Bα), 4.88 (2H, s, Ph*CH*$_2$), 5.47 (1H, d, J = 5.0 Hz, Aα-OH), 5.51 (1H, dd, J = 5.0, 3.0 Hz, Bα-OH), 6.41 (1H, d, J = 1.0 Hz, A6), 6.57 (1H, d, J = 8.2 Hz, B5), 6.77~6.91 (7H, m, C- and D-rings), 6.83 (1H, br-s, A2), 6.87 (1H, m, B6), 6.99 (1H, m, C- or D-ring), 7.18 (1H, br-s, B2), and 7.25~7.34 (5H, m, *Ph*CH$_2$). ^{13}C-NMR (DMSO-d_6, 125 MHz) δ_C: 55.48 (OMe), 55.52 (OMe), 55.60 (OMe), 55.88 (OMe), 60.04 (Bγ), 60.11 (Aγ), 71.52 (Aα and Bα), 74.21 (Ph*CH*$_2$), 83.51 and 83.56 (Aβ and Bβ), 106.52 (A2), 109.30 (A6), 112.23/112.25 (B2 of two isomers), 112.58 (C2), 112.68 (D2), 115.94, 115.99 and 116.04 (C- or D-ring), 118.02/118.04 (B5), 119.49 (B6), 120.58, 120.69, 121.08, 121.10, and 121.16 (C- and D-rings), 127.71, 127.99 and 128.08 (*Ph*CH$_2$), 136.80/136.82 (A4), 137.69 (*Ph*CH$_2$), 138.46 (A1), 138.49/138.52 (B1), 144.04 (B4), 147.92 (C4), 148.02 (D4), 149.57 (A5), 149.64/149.66 (B3), 149.69 and 149.79 (C3 and D3), and 152.98 (A3). MS (ESI) calculated for $C_{41}H_{44}O_{12}Na$ $[M+Na]^+$ 751.27, found 751.30.

Acetates of compound **10** (**10Ac**) in $CDCl_3$: ^{1}H-NMR ($CDCl_3$, 500 MHz) δ_H: 1.98 (3H, s, OCO*Me*), 2.00 (3H, s, OCO*Me*), 2.01/2.02 (3H, s, Aα-OCO*Me* of two isomers), 2.06/2.07 (3H, s, Bα-OCO*Me*), 3.70/3.71 (3H, s, OMe), 3.75 (3H, s, OMe), 3.76 (3H, s, B3-OMe), 3.82 (3H, s, A3-OMe), 4.15~4.21 (2H, m, Aγ$_1$ and Bγ$_1$), 4.37~4.45 (2H, m, Aγ$_2$ and Bγ$_2$), 4.56 (1H, m, Aβ), 4.65 (1H, m, Bβ), 5.00 (2H, s, Ph*CH*$_2$), 5.93 (1H, d, J = 6.0 Hz, Aα), 6.03 (1H, d, J = 5.0 Hz, Bα), 6.58 (1H, d, J = 2.0 Hz, A6), 6.62 (1H, m, B5), 6.72~6.85 (6H, m, C- and D-rings), 6.80

(1H, br-s, A2), 6.83 (1H, m, B6), 6.92~6.97 (2H, m, C- and/or D ring), 7.04 (1H, d, J = 1.5 Hz, B2), and 7.17~7.30 (5H, m, *Ph*CH_2). ^{13}C-NMR ($CDCl_3$, 125 MHz) δ_C: 20.74/20.76 (OCO*Me*), 20.90 (OCO*Me*), 20.93 (OCO*Me*), 21.02 (OCO*Me*), 55.64 (OMe), 55.71 (OMe), 55.90 (OMe), 56.14 (OMe), 62.62 and 62.69 (Aγ and Bγ), 73.65/73.70 (Aα of two isomers), 73.86 (Bα), 74.97 (Ph*CH_2*), 80.15 and 80.19/80.21 (Aβ and Bβ), 107.33/107.44 (A2), 111.20/111.32 (A6), 112.01 (B2), 112.45 (C2), 112.48 (D2), 117.55/117.69 (B5), 119.28 and 119.34 (C ring or D ring), 119.82 (B6), 120.85, 120.87, 121.92, 123.46 and 123.51 (C- and D-ring), 127.65, 128.03, 128.06 and 128.14 (*Ph*CH_2), 131.84/131.90 (B1), 132.46/132.51 (A1), 137.51 and 137.54 (*Ph*CH_2), 139.05/139.10 (A4), 146.21/146.30 (B4), 147.14 (C4 and D4), 149.61/149.68 (A5), 149.76/149.83 (B3), 150.92 and 151.03/151.05 (C3 and D3), 153.92 (A3), 169.47/169.51 and 169.55 (Aα- and Bα-O*CO*Me), 170.71/170.73 and 170.74 (Aγ- and Bγ-O*CO*Me).

Acetates of compound **10** (**10Ac**) in acetone-d_6: ^{1}H-NMR (acetone-d_6, 500 MHz) δ_H: 1.912/1.914 (3H, s, OCOMe of two isomers), 1.947/1.949 (3H, s, OCOMe), 1.98/1.99 (3H, s, Aα-OCOMe), 2.07 (3H, s, Bα-OCOMe), 3.72/3.73 (3H, s, OMe), 3.79 (3H, s, B3-OMe), 3.81 (3H, s, OMe), 3.89 (3H, s, A3-OMe), 4.15~4.22 (2H, m, $A\gamma_1$ and $B\gamma_1$), 4.31~4.44 (2H, m, $A\gamma_2$ and $B\gamma_2$), 4.74 (1H, m, Aβ), 4.85 (1H, m, Bβ), 5.00 (2H, s, Ph*CH_2*), 5.93 (1H, d, J= 5.5 Hz, Aα), 6.03 (1H, d, J = 5.0 Hz, Bα), 6.56 (1H, d, J= 2.0 Hz, A6), 6.76 (1H, dd, J= 8.0, 1.5 Hz, B5), 6.80~6.87 (2H, m, C- and/or D-ring), 6.91~7.03 (6H, m, C- and D-ring), 6.96 (1H, br-s, A2), 6.98 (1H, m, B6), 7.22~7.29 (3H, m, *Ph*CH_2), 7.26 (1H, m, B2), and 7.37~7.39 (2H, m, *Ph*CH_2). ^{13}C-NMR (acetone-d_6, 125 MHz) δ_C: 20.63 (OCO*Me*), 20.64 (OCO*Me*), 20.79/20.81 (OCO*Me*), 20.90 (OCO*Me*), 56.11 (OMe), 56.18 (OMe), 56.30 (OMe), 56.56 (A3-OMe), 63.12/63.14 and 63.17 (Aγ and Bγ), 74.57 (Aα), 74.67 (Bα), 75.27 (Ph*CH_2*), 80.29 and 80.32 (Aβ and Bβ), 107.67 (A2), 110.89 (A6), 113.28 (B2), 113.68 and 113.74 (C2 and D2), 119.55 (B5), 119.67,

119.74 and 119.79 (C- and/or D-ring), 120.79/120.82 (B6 of two isomers), 121.58, 121.62, 123.97, 123.99 and 124.03 (C- and D-rings), 128.39 and 128.84 (*Ph*CH_2), 133.65 (A1), 133.93 (B1), 139.00 (*Ph*CH_2), 139.34 (A4), 146.51/146.53 (B4), 148.30 and 148.31 (C4 and D4), 151.23 (A5), 151.42 (B3), 151.88 and 151.99/152.02 (C3 and D3), 154.99 (A3), 169.80 and 169.82 (Aα-O*CO*Me), 169.92 (Bα-O*CO*Me), 170.72 (O*CO*Me), and 170.75 (O*CO*Me).

4–O–5-tetramer model (11)

A stirred suspension of compound **10** (350 mg, 0.45 mmol) and Pd/C (35 mg) in THF (20 mL) was fitted with a hydrogen-gas-filled balloon. The reaction mixture was kept stirring overnight at room temperature. The completion of the reaction was monitored by TLC (EtOAc, 100%). After the reaction, Pd/C was filtered off through a glass filter with an aid of celite, and washed with acetone (20 mL). Filtrate was concentrated under vacuum to afford compound **11** as white powder (308 mg, 99%). ^{1}H-NMR (DMSO-d_6, 500 MHz) δ_H: 3.58~3.65 (4H, m, Aγ and Bγ), 3.62 (3H, s, OMe), 3.69 (3H, s, OMe), 3.73 (3H, s, B3-OMe), 3.76 (3H, s, A3-OMe), 4.20 (1H, dt, *J* = 5.2, 4.7 Hz, Aβ), 4.30 (1H, m, Bβ), 4.56 (1H, br-t, *J* = 5.0 Hz, Aγ-OH), 4.61 (1H, dd, *J* = 5.0, 4.5 Hz, Aα), 4.64 (1H, br-t, *J* = 5.0 Hz, Bγ-OH), 4.77 (1H, dd, *J* = 5.0, 4.5 Hz, Bα), 5.35 (1H, d, *J* = 4.5 Hz, Aα-OH), 5.46 (1H, dd, *J* = 4.5, 2.5 Hz, Bα-OH), 6.42 (1H, br-s, A6), 6.45 (1H, dd, *J* = 8.0, 1.0 Hz, B5), 6.76 (1H, d, *J* = 7.5 Hz, C- or D-ring), 6.79~6.83 (2H, m, C- and/or D-ring), 6.81 (1H, br-s, A2), 6.82 (1H, m, B6), 6.84~6.88 (3H, m, C- and/or D ring), 6.91 (1H, dd, *J* = 8.0, 2.0 Hz, C- or D-ring), 6.98 (1H, ddd, *J* = 8.0, 4.0, 2.0 Hz, C- or D-ring), 7.14 (1H, br-s, B2), and 8.51 (0.8H, s, A4-OH). ^{13}C-NMR (DMSO-d_6, 125 MHz) δ_C: 55.48 (OMe), 55.63 (OMe), 55.83 (OMe), 55.87 (OMe), 60.05 (Bγ), 60.21 (Aγ), 71.55 (Aα and Bα), 83.58 (Aβ), 83.63 (Bβ), 106.53 (A2), 110.50/110.52 (A6 of two isomers), 112.07/112.09 (B2), 112.61 and 112.71 (C2

and D2), 115.80, 115.83 and 116.00 (C- and/or D-ring), 116.63/116.67 (B5), 119.32 (B6), 120.59, 120.69, 120.98 and 120.12 (C- and/or D-ring), 132.80 (A1), 136.43/136.45 (A4), 137.38/137.45 (B1), 143.54/143.56 (A5), 144.94 (B4), 148.03 (C4 and D4), 148.27 (A3), 149.12/149.15 (B3), 149.68 and 149.80 (C3 and D3). MS (ESI) calculated for $C_{34}H_{38}O_{12}Na$ $[M+Na]^+$ 661.23, found 661.26.

Acetates of compound **11** (**11Ac**) in $CDCl_3$: ^{1}H-NMR ($CDCl_3$, 500 MHz) δ_H: 1.96/1.97 (3H, s, OCOMe of two isomers), 1.98/2.00 (3H, s, Aα-OCOMe), 2.01 (3H, s, OCOMe), 2.06/2.07 (3H, s, Bα-OCOMe), 2.15/2.16 (3H, s, A4-OCOMe), 3.68/3.69 (3H, s, OMe), 3.77~3.78 (6H, m, 2 × OMe), 3.80/3.81 (3H, s, A3-OMe), 4.15~4.20 (2H, m, $A\gamma_1$ and Bγ1), 4.36~4.44 (2H, m, $A\gamma_2$ and $B\gamma_2$), 4.53~4.55 (1H, m, Aβ), 4.66 (1H, dt, J = 5.2, 4.7 Hz, Bβ), 5.94 (1H, d, J = 5.0 Hz, Aα), 6.02 (1H, d, J = 5.0 Hz, Bα), 6.50/6.51 (1H, d, J = 1.8 Hz, A6), 6.72~6.74 (1H, m, C- or D-ring), 6.75 (1H, m, B5), 6.78~6.87 (5H, m, C- and D-rings), 6.81 (1H, m, A2), 6.86 (1H, m, B6), 6.93~6.99 (2H, m, C- and/or D-ring), and 7.05/7.06 (1H, d, J = 1.5 Hz, B2). ^{13}C-NMR ($CDCl_3$, 125 MHz) δ_C: 20.26 (A4-OCO*Me*), 20.70/20.71(OCO*Me* of two isomers), 20.75 (OCO*Me*), 20.86/20.90 (Aα-OCO*Me*), 21.01 (Bα-OCO*Me*), 55.65 (OMe), 55.74 (OMe), 56.04 (B3-OMe), 56.24 (A3-OMe), 62.55 and 62.58 (Aγ and Bγ), 73.59/73.63 (Aα), 73.83 (Bα), 80.06, 80.15/80.18 (Aβ and Bβ), 106.37/106.45 (A2), 109.91/110.03 (A6), 112.31/112.34 (B2), 112.51 and 112.54 (C2 and D2), 119.21, 119.31, 119.46 and 119.54 (C- and/or D-ring), 119.59 (B5), 119.89/119.93 (B6), 120.93, 120.95, 123.51, 123.56 and 123.61 (C- and D-rings), 130.49/130.54 (A4), 132.96/132.98 (B1), 135.08/135.12 (A1), 145.26/145.33 (B4), 147.09 and 147.12 (C4 and D4), 149.52/149.57 (A5), 150.43/150.47 (B3), 151.01 and 151.03/151.06 (C3 and D3), 152.50 (A3), 168.15 (A4-O*CO*Me), 169.34 (Aα-O*CO*Me), 169.56 (Bα-O*CO*Me), and 170.72 (Aγ-O*CO*Me and Bγ-O*CO*Me).

Acetates of compound **11** (**11Ac**) in acetone-d_6: ^{1}H-NMR (acetone-d_6, 500 MHz) δ_H: 1.89/1.90 (3H, s, Aγ-OCOMe of two isomers), 1.95/1.96 (3H, s, Bγ-OCOMe), 1.98/1.99 (3H, s, Aα-OCOMe), 2.07 (3H, s, Bα-OCOMe), 2.14 (3H, s, A4-OCOMe), 3.71/3.72 (3H, s, OMe), 3.78/3.79 (3H, s, B3-OMe), 3.82 (3H, s, OMe), 3.85 (3H, s, A3-OMe), 4.16 (1H, dd, J = 12.0, 1.5 Hz, Aγ), 4.21 (1H, dd, J = 12.0, 3.0 Hz, Bγ), 4.33 (1H, dd, J = 12.0, 1.5 Hz, Aγ), 4.39 (1H, dd, J = 12.0, 3.0 Hz, Bγ), 4.72 (1H, m, Aβ), 4.85 (1H, m, Bβ), 5.93 (1H, d, J= 5.0 Hz, Aα), 6.06 (1H, d, J = 5.0 Hz, Bα), 6.52/6.53 (1H, d, J= 2.0, A6), 6.80~6.88 (2H, m, C- and/or D-ring), 6.84 (1H, dd, J = 8.3, 2.0 Hz, B5), 6.90~6.92 (1H, m, C- or D-ring), 6.94~7.04 (5H, m, C- and D-rings), 6.97 (1H, s, A2), 6.98~7.02 (2H, m, C- and/or D-ring), 7.01 (1H, m, B6), and 7.27 (1H, br-s, B2). ^{13}C-NMR (acetone-d_6, 125 MHz) δ_C: 20.14 (A4-OCO*Me*), 20.60 (OCO*Me*), 20.64 (OCO*Me*), 20.75/20.77 (Aα-OCO*Me* of two isomers), 20.88 (Bα-OCO*Me*), 56.12 (OMe), 56.20 (OMe), 56.33 (OMe), 56.62 (OMe), 62.98 (Aγ), 63.10/63.12 (Bγ), 74.53 (Aα), 74.61 (Bα), 80.26 and 80.29 (Aβ and Bβ), 106.69 (A2), 109.82 (A6), 113.41/113.42 (B2), 113.71 and 113.75/113.77 (C2 and D2), 119.59, 119.71 and 119.98 (C- and/or D-ring), 120.84 and 120.89 (B5 and B6), 123.99, 124.04 and 124.10 (C- and/or D-ring), 130.87 (A4), 134.86 (B1), 136.29 (A1), 145.60 (B4), 148.21 and 148.29 (C4 and D4), 150.84 (A5), 151.81 (B3), 151.94 and 151.99 (C3 and D3), 153.60 (A3), 168.34 (A4-O*CO*Me), 169.79/169.82 (Aα-O*CO*Me), 169.95 (Bα-O*CO*Me), 170.74 (O*CO*Me), and 170.77 (O*CO*Me).

Results and Discussion

As model compounds for the 4–O–5 structure in lignin, three types of diaryl ethers were synthesized: 4–O–5-coupled dehydrodiconiferyl alcohol (**6**) that is a model for a coniferyl alcohol end-group of lignin, non-phenolic tetramer model (**10**) for 4–O–5 structures at the internal unit of lignin, and phenolic tetramer (**11**) for 4–O–5 structures at the phenolic end of lignin (Figure 1). As well as pine lignin sample (MWL), these model compounds were examined for a chemical degradation method and NMR spectroscopy, and their results were compared for the analysis of 4–O–5 structures in lignin.

Nitrobenzene oxidation products released from 4–O–5 structures

Among the 4–O–5 models, compound **6** and **11** in Figure 1 were subjected to nitrobenzene oxidation method (NBO) to investigate whether this analysis method is applicable to the 4–O–5 structures in lignin or not. Unlike β–O–4 and 5–5 structures that give simple benzaldehydes, vanillin and 5–5-coupled dehydrodivanillin, respectively,[8] both model compounds **6** and **11** did not yield 4–O–5-coupled dehydrodivanillin (**2**). Instead, 4–O–5-coupled dehydrovanillin-vanillic acid (**3**) was obtained as their degradation product accompanied by larger amounts of vanillin and vanillic acid (Table 1). Compound **2** and **3** were subjected to the NBO method to check their stability under the reaction condition.

Compound **3** remained fairy stable under the harsh oxidation condition; 74% of this compound was recovered (Table 1). On the other hand, compound **2** was not recovered under the same condition, and degraded into compound **3**, vanillin and vanillic acid. This result implied the possibility that a part of vanillin and vanillic acid obtained from 4–O–5 model compounds **6** and **11** were formed *via* compound **2**. The low yield of 4–O–5-type products from the model

compounds **6** and **11** limited the opportunity for the quantification of 4–O–5 structures in lignin by the NBO method. However, it was suggested that compound **3** released from the 4–O–5 units by the NBO method is useful for investigating the existence of 4–O–5 structures in a lignin sample.

The NBO of pine MWL released a small amount of compound **3**, suggesting that at least 0.56% of aromatic ring in the pine lignin was detected as 4–O–5 structure based on the calculation that molecular weight of a phenylpropanoid unit is 200 although the actual 4–O–5 content in the MWL must be higher than this value considering the low yield of compound **3** obtained from the model compounds **6** and **11** (Table 1). In addition, neither syringaldehyde nor syringic acid was found in the NBO products of the pine MWL at the detectable level by GC-FID analysis, indicating that the pine lignin sample does not contain syringyl units, and that there was no accidental contamination by hardwood meals during the grinding process to make wood meal from pine wood chips using a Wiley-mill.

Table 1. Yields of nitrobenzene oxidation products from 4–O–5 lignin model compounds and pine MWL.

Sample	NBO products from model compounds (mol %) and pine MWL (μmol g^{-1})					
	4–O–5-diV **2**	4–O–5-VVA **3**	V	VA	S	SA
4–O–5-diV **2**	n.d.	18.1 ± 1.7	56.2 ± 5.8	9.9 ± 1.5	---	---
4–O–5-VVA **3**	n.d.	74.0 ± 0.2	1.4 ± 0.1	1.0 ± 0.1	---	---
4–O–5-dimer **6**	n.d.	6.7 ± 2.8	36.4 ± 4.4	10.8 ± 3.1	---	---
4–O–5-tetramer **11**	n.d.	5.3 ± 1.1	17.0 ± 1.6	19.0 ± 2.7	---	---
Pine MWL	n.d.	0.56[a]	33.29 [a]	5.05 [a]	n.d.	n.d.

(13.95[b]) (1664.3 [b]) (252.3 [b])

[a] Yield was expressed in molar yield (%) based on the assumption that the molecular weight of phenyl propane (C_6-C_3) unit is 200. [b] Yield was expressed in μmol per one gram MWL. Chemical structures of compound **2**, **3**, **6** and **11** were shown in Figure 1. V: vanillin, VA: vanillic acid, S: syringaldehyde, SA: syringic acid.

The HSQC spectra of pine MWL

The HSQC spectrum of the acetylated pine MWL ($CDCl_3$ solvent) showed different correlation peaks that can be assigned for β–O–4, β–β (resinol), β–5 (phenylcoumaran) in its aliphatic region (Figure 2).[20-23] In addition, the correlation peaks for dibenzodioxocins (5–5/4–O–β) were readily found in the aliphatic region.[14-15, 29] It was confirmed that the pine MWL sample prepared in the present study had an inter-unit linkage distribution typical of guaiacyl lignin.[14] In the aromatic region of the HSQC spectrum, two correlation peaks at 6.70/104.2 and 6.75/105.8 ppm were assigned to the H2/C2 correlation of 4–O–5 structures, which were found at a low contour level close to the noise level of the spectrum (Figure 3B-2), although these correlation peaks were not visible at the higher contour level (Figure 3B-1) that is used for the detection of the major signals from lignins (Figure 2, and 3A). The ^{1}H and ^{13}C chemical shifts of the two correlations were in good agreement with those of the model compounds obtained as acetate forms in $CDCl_3$ solvent (6.70/104.72 ppm in $\mathbf{6_{Ac}}$, 6.80/107.40 in $\mathbf{10_{Ac}}$, and 6.81/106.37 in $\mathbf{11_{Ac}}$) as shown in Table 2. The H2/C2 correlations were within the aromatic region that is occupied by syringyl units when they are present (Figure 3A and 3B-2). However, it can be safely concluded that these signals do not originate from syringyl units but from 4–O–5-structures as the lack of syringyl units in the MWL was indicated by the results of nitrobenzene oxidation mentioned above. Although the correlation around 6.5/110 ppm in the MWL can possibly be assigned to H6/C6 correlation of 4–O–5 structures based on the NMR data of lignin models, the cross-signal is obscured by

correlations from other types of guaiacyl units. To obtain more reliable assignments for 4–O–5 structures, the NMR data of the model compounds were collected under different sample and solvent conditions.

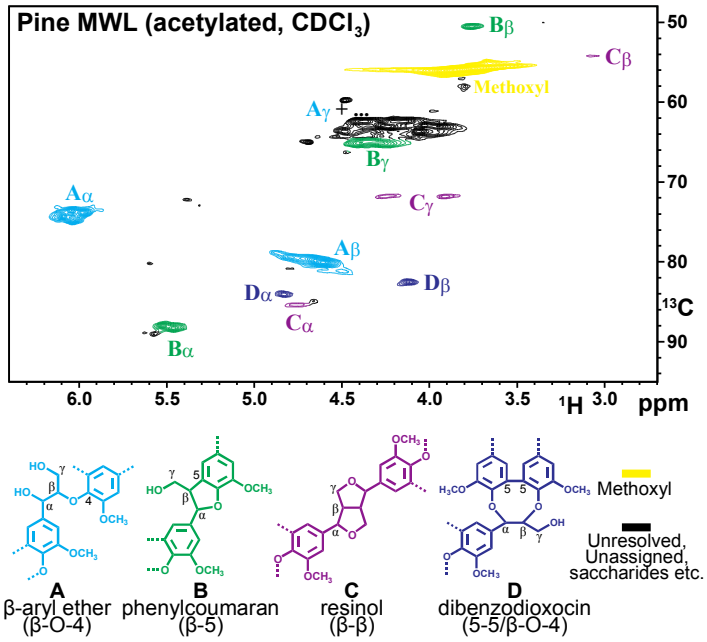


Figure 2. Partial short-range $^{1}H–^{13}C$ (HSQC) NMR spectrum (aliphatic regions) of milled wood lignin isolated from pine wood trunk (acetylated, $CDCl_3$). MWL: milled wood lignin.

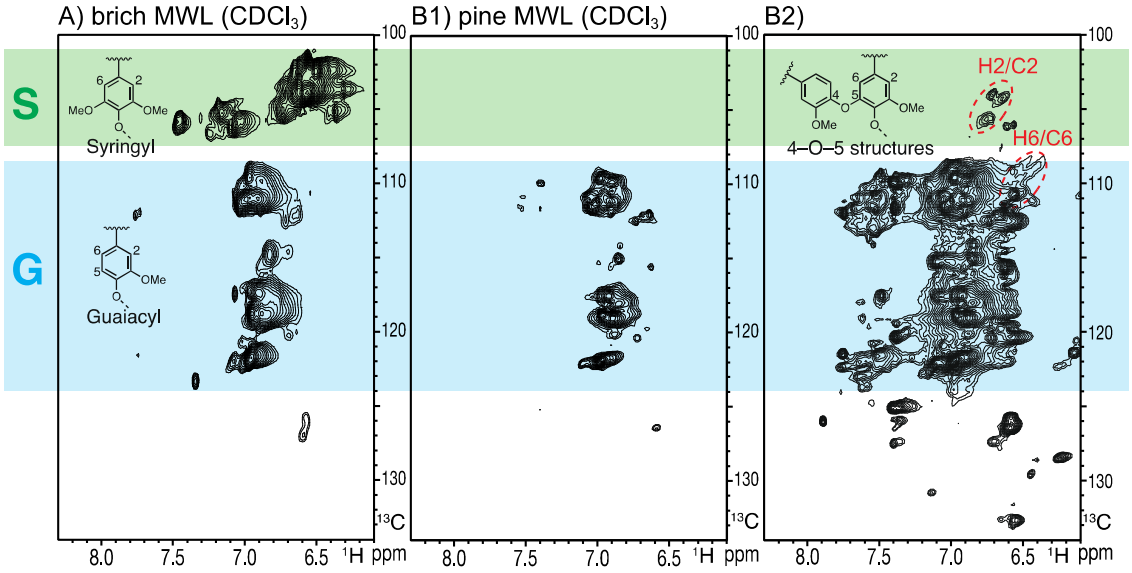


Figure 3. HSQC spectra (aromatic regions) of milled wood lignins isolated from pine wood and birch wood trunks (acetylated, $CDCl_3$). A) birch MWL; B1) pine MWL and B2) pine MWL at the lower contour level than for B1. MWL: milled wood lignin, S: syringyl region, G: guaiacyl region.

An NMR solvent influences the chemical shift values, a feature that can possibly be utilized for obtaining more reliable assignments of lignin spectra. Although the choice of NMR solvent is limited to DMSO-d_6 and some uncommon NMR solvents, e.g., DMSO-d_6, Pyridine-d_5, DMSO-d_6/Pyridine-d_5 (4:1), or dioxane-d_8/D_2O (9:1),[9,30-36] for an underivatized lignin sample due to its solubility limitations, acetylated lignins can be solubilized in common NMR solvents such as $CDCl_3$, acetone-d_6 or DMSO-d_6.[14,21,29,30,37,38] There is a large difference in the chemical shift values between underivatized and acetylated lignins.[39] For example, acetylation of lignin causes a large downfield shift in the α-proton from β–O–4 structures ($\Delta\delta_H$: □+1 ppm) that results in a well-dispersed ^{1}H NMR lignin spectrum, and makes assignments of the correlation signals in the aliphatic regions of HMBC and HSQC-TOCSY spectra of lignins more reliable. The NMR

spectra of acetylated lignin samples have typically been recorded using $CDCl_3$ as solvent,[14,21,23,29,37,38] and the same sample condition was employed in our study, as described above (Figure 2 and 3). However, the NMR spectra of underivatized lignins are required for investigating acylating group that are naturally present in some lignins, and are often recorded in DMSO-d_6[32,36,39] or DMSO-d_6/pyridine-d_5 (4:1).[33,34]

The NMR sample conditions for lignins have been converging on the following three conditions:[9,14,21,23,29-32,35-38] the acetatylated lignin using $CDCl_3$ or acetone-d_6, and the underivatized lignin using DMSO-d_6. For the assignment of lignin spectra, the NMR data from a large number of lignin model compounds have been collected under the three conditions[13,20-23,40-42] as the chemical shift values, especially in the aliphatic region, vary under the different conditions. Although the effects on the chemical shift values have not been investigated for the aromatic region in detail, such an effect could be utilized for the assignment of 4–O–5 structures. In order to evaluate the influence of sample and solvent conditions on the chemical shift values, the NMR data of the 4–O–5 models **6**, **10** and **11** were collected under the three conditions: acetate form in $CDCl_3$ or acetone-d_6, and non-acetate form in DMSO-d_6 (Table 2).

Table 2. Selected NMR chemical shifts for 4–O–5-model compounds at H2/C2 and H6/C6 positions on the 5-linked aromatic rings (A-rings) measured in three sample conditions.

Model	In DMSO-*d6* (ppm)		Model	In $CDCl_3$ (ppm)		In acetone-*d6* (ppm)	
	H2/C2	H6/C6	(acetates)	H2/C2	H6/C6	H2/C2	H6/C6
6	6.84/105.27	6.41/109.78	**6Ac**	6.70/104.72	6.43/108.99	6.96/105.43	6.49/109.12
10	6.83/106.52	6.41/109.30	**10Ac**	6.80/107.40	6.58/111.30	6.96/107.67	6.56/110.89
11	6.81/106.53	6.42/110.51	**11Ac**	6.81/106.37	6.50/109.91	6.97/106.69	6.53/109.82
$\Delta\delta_H/\Delta\delta_C$[a]	Δ0.03/Δ1.26	Δ0.01/Δ1.21	$\Delta\delta_H/\Delta\delta_C$[a]	Δ0.11/Δ2.68	Δ0.15/Δ2.31	Δ0.01/Δ2.24	Δ0.07/Δ1.77

[a]Differences in the chemical shifts among H2/C2 or H6/C6 signals on the A-rings of 4–O–5-dimer **6**, and 4–O–5-tetramers **10** and **11**. Their chemical structures were shown in Figure 1.

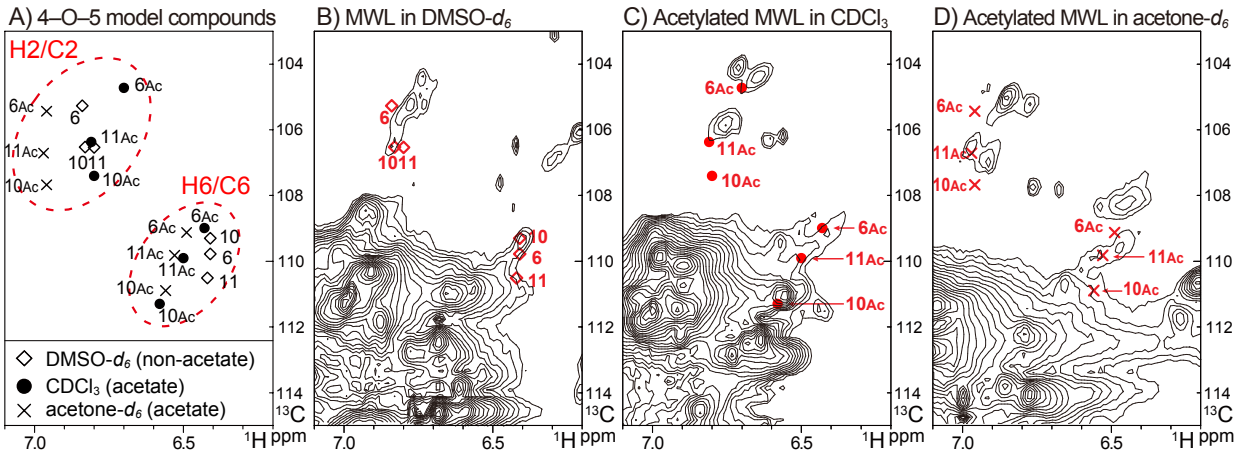


Figure 4. Selected NMR data plot of 4–O–5 lignin model compounds, and partial HSQC spectra (δ_H/δ_C 6.1-7.2/100-115) of pine MWL measured in different sample conditions. A) H2/C2 and H6/C6 positions on the 5-linked aromatic rings (A-ring) of a 4–O–5-dimer **6**, 4–O–5-tetramers **10** and **11**, and their acetates **6Ac**, **10Ac**, **11Ac**; B) MWL in DMSO-d_6 (H2/C2 correlation at 6.76/105.2 ppm and H6/C6 correlation at 6.40/109.4 ppm); C) acetylated MWL in $CDCl_3$ (H2/C2 correlations at 6.70/104.2 and 6.75/105.8 ppm); D) acetylated MWL in acetone-d_6 (H2/C2 correlations at 6.84/105.1 and 6.95/106.7 ppm). NMR data of the models were plotted over the lignin spectrum recorded in the same sample and solvent condition. See Figure 1 for the chemical structures of model compounds, Table 2 for chemical shift values of model compounds, and Figure 3 for the whole aromatic region of the lignin spectra.

Although the H2/C2 and H6/C6 positions had the similar δ_H values among three model compounds $\mathbf{6_{Ac}}$, $\mathbf{10_{Ac}}$, and $\mathbf{11_{Ac}}$ (Table 2), they showed a wide range of C2 and C6 carbon shifts in $CDCl_3$ solvent ($\Delta\delta_C$: 2.68 and 2.31 ppm for C2 and C6 carbons, respectively). The wide

variation in carbon chemical shifts can be utilized to distinguish the structural difference of 4–O–5 structures in a lignin sample, such as differences in their side-chain parts and/or the phenolic end type from the others. However, as already shown in Figure 2, the pine MWL used in this study showed a weak obscure signal for its H6/C6 correlation presumably due to the low content and/or the structural variation in the 4–O–5 structures. Although two kinds of the H2/C2 correlations were observed under the sample and solvent condition, further interpretation of these correlations could not be attained because of the lack of H6/C6 signal information and an insufficient number of model compounds.

A solvent change from $CDCl_3$ to acetone-d_6 caused substantial downfield shifts to proton at H2 position of the models $\mathbf{6_{Ac}}$, $\mathbf{10_{Ac}}$, and $\mathbf{11_{Ac}}$ (approximately +0.2 ppm), whereas the chemical shifts of H6 proton, and C2 and C6 carbons were similar between two solvents (Fig 4, Table 2). Consequently, similarly to the case of $CDCl_3$ solvent, both at 2 and 6-positions, the δ_H and δ_C values of the three models showed narrow and wide variations in acetone-d_6 solvent, respectively. Also in acetone-d_6 solvent, the chemical shifts of H2/C2 correlation peaks of the acetylated pine MWL were in good agreement with those of model compounds. The chemical shifts of H2 protons were approximately 0.2 ppm downfield from those observed in the acetylated lignin using $CDCl_3$ solvent, as reflected by the results of model compounds.

In contrast, in the case of DMSO-d_6 solution, the chemical shifts at the H2/C2 and H6/C6 positions were less influenced by the structural differences among the three non-acetate models **6**, **10** and **11** (Table 2). Their H2/C2 and H6/C6 positions appeared at 6.83/105.90 ppm (± 0.02/0.63) and 6.41/109.90 ppm (± 0.01/0.61), respectively. The HSQC spectrum of non-acetylated pine MWL showed not only H2/C2 correlation peak (6.76/105.2 ppm), but a clear single correlation peak assigned to the H6/C6 position on 4–O–5 structures (6.40/109.4 ppm).

This observation suggests that the detection of the H6/C6 correlation peak in the MWL was accomplished by suppressing the broadening of the peak attributed in part to the structural variation of 4–O–5 structures.

Conclusion

Diaryl ether structures (4–O–5-linked units) were found in pine MWL through the use of the HSQC experiment. The underivatized lignin spectrum recorded in DMSO-d_6 revealed the H2/C2 and H6/C6 correlation signals from the 5-linked moiety of the 4–O–5-linked units' aromatic ring; correlation peaks were assigned based on the chemical shift values of 4–O–5-dimeric and tetrameric model compounds. In the case of acetylated lignin spectra recorded in $CDCl_3$ and acetone-d_6, only the H2/C2 correlation signals can be assigned as the candidate signals for H6/C6 correlations are obscured by correlation signals from the other types of guaiacyl units. The assignments were possible by exploiting the absence of syringyl units in the lignin sample, as evidenced by the complete absence of syringyl type products from nitrobenzene oxidation of the pine lignin. In addition, by this chemical degradation method, a 4–O–5-type product was released from both the pine lignin and from 4–O–5-model compounds, indicating that nitrobenzene oxidation is a useful method for the detection of 4–O–5 structures in a lignin sample.

Corresponding Author

Takuya Akiyama

Wood Chemistry Laboratory, Department of Biomaterial Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan

*E-mail: akiyama@woodchem.fp.a.u-tokyo.ac.jp

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. †These authors contributed equally.

Acknowledgments

This work was supported by the PRESTO program from the Japan Science and Technology Agency (JST). The authors would like to thank Dr. Y. Hori, Technology Advancement Center, the University of Tokyo for collection of the mass spectra. NMR experiments were carried out at the Yokohama City University in Japan, with support from the Project for Creation of Research Platforms and Sharing of Advanced Research Infrastructure by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT).

Abbreviations

BnBr, benzyl bromide; COSY, (^{1}H–^{1}H) correlation spectroscopy; DEPT-135, distortionless enhancement by polarization transfer; DIBAL-H, diisobutylaluminium hydride; HSQC, (^{1}H–^{13}C) heteronuclear single-quantum coherence (NMR spectroscopy); HMBC, (^{1}H–^{13}C) heteronuclear multiple-bond correlation (NMR spectroscopy); LDA, lithium diisopropylamide; PPTS, pyridinium *p*-toluenesulfonate; Py, pyridine.

References

(1) Akiyama, T.; Goto, H.; Nawawi, D. S.; Syafii, W.; Matsumoto, Y.; Meshitsuka, G. *Holzforschung* **2005**, 59, 276-281.

(2) Wagner, A.; Tobimatsu, Y.; Phillips, L.; Flint, H.; Geddes, B.; Lu, F.; Ralph, J. *Proc. Natl. Acad. Sci. U S A.* **2015**, 112, 6218-6223.

(3) Ralph, J.; Akiyama, T.; Kim, H.; Lu, F.; Schatz, P. F.; Marita, J. M.; Ralph, S. A.; Srinivasa Reddy, M. S.; Chen, F.; Dixon, R. A. *J. Biol. Chem.*, **2006**, 281, 8843-8853.

(4) Stewart, J. J.; Akiyama, T.; Chapple, C. C.; Ralph, J.; Mansfield, S. D. *Plant Physiol.* **2009**, 150, 621-635.

(5) Nawawi, D. S.; Syafii, W.; Akiyama, T.; Matsumoto, Y. *Holzforschung*, **2015**, DOI 10.1515/hf-2015-0107

(6) Leopold, B. *Acta Chem. Scand.* **1952**, 6, 38-48.

(7) Pew, J. C. *J. Org. Chem.* **1963**, 28, 1048-1054.

(8) Tamai, A.; Goto, H.; Akiyama, T.; Matsumoto, Y. *Holzforschung*, **2015**, 69, 951-958.

(9) Lange, W.; Faix, O. *Holzforschung*, **1999**, 53, 519-524.

(10) Lapierre, C.; Pollet, B.; Monties, B.; Rolando, C. *Holzforschung* **1991**, 45, 61-68.

(11) Peng, J.; Lu, F; Ralph, J. *J. Agri. Food Chem*. **1998**, 46, 553-560.

(12) Lundquist, K.; Li, S.; Parkås, J. BioResources, **2011**, 6, 920-926.

(13) Drumond, M.; Aoyama, M.; Chen, C. L.; Robert, D. *J. Wood Chem. Technol.* **1989**, 9, 421-441.

(14) Karhunen, P.; Rummakko, P.; Sipilä, J.; Brunow, G.; Kilpeläinen, I. *Tetrahedron Lett.* **1995**, 36, 169-170.

(15) Karhunen, P.; Rummakko, P.; Sipilä, J.; Brunow, G.; Kilpeläinen, I. *Tetrahedron Lett.* **1995**, 36, 4501-4504.

(16) Ralph, S.A. Proceeding of 13th International Symposium on Wood, Fiber, and Pulping Chemistry; Auckland, New Zealand, 2005, vol. 3, pp. 315-319.

(17) Sarkanen, K.V.; Hergert, H. L. In *Lignins: Occurrence, formation, structure, and reactions*; Sarkanen, K.V. Eds.; Wiley-Interscience: New York, 1971; pp. 43-94.

(18) Björkman, A. *Svensk Papperstidn.* **1956**, 59, 477-485.

(19) Chen, C. L. In *Methods in lignin chemistry*; Lin, S. Y., Dence, C. W., Eds.; Springer-Verlag: Berlin, 1992; pp. 301-321.

(20) Ralph, S. A.; Ralph, J.; Landucci, L. L. *NMR Database of Lignin and Cell Wall Model Compounds*, 2009, https://www.glbrc.org/databases_and_software/nmrdatabase/

(21) Hauteville, M.; Lundquist, K.; von Unge, S. *Acta Chem. Scand.* **1986**, B40, 31-35.

(22) Sipilä, J.; Syrjänen, K. *Holzforschung* **1995**, 49, 325-331.

(23) Li, S.; Iliefski, T.; Lundquist, K.; Wallis, A. F. A. *Phytochemistry* **1997**, 46, 929-934.

(24) Ralph, J.; Quideau, S.; Grabber, J. H.; Hatfield, R. D. *J. Chem. Soc. Perkin Transactions 1*, **1994**, (23), 3485-3498.

(25) Quideau, S.; Ralph, J. *J Agri. Food Chem.* **1992**, 40, 1108-1110.

(26) Pearl, I. A. *Org. Synth.* **1950**, 30, 101-106.

(27) Nakatsubo, F. In *Methods in Enzymology v161, Biomass Part B, Lignin, Pectin and Chitin*; Wood, W. A. Kellogg, S. T. Eds.; Academic Press: San Diego, CA, USA, 1988; pp. 57-65.

(28) Kishimoto, T.; Uraki, Y; Ubukata, M. *Org. Biomol Chem.* **2006**, 4, 1343-1347.

(29) Terashima, N.; Akiyama, T.; Ralph, S.; Evtuguin, D.; Neto, C.P.; Parkås, J.; Paulsson, M.; Westermark, U.; Ralph, J. *Holzforschung* **2009**, 63, 379-384.

(30) Fukagawa, N.; Meshitsuka, G.; Ishizu, A. *J. Wood Chem. Technol.* **1991**, 11, 373-396.

(31) Zhang, L.; Henriksson, G.; Gellerstedt, G. *Org. Biomol. Chem.*, **2003**, 1, 3621-3624.

(32) del Río, J. C.; Rencoret, J.; Gutiérrez, A.; Nieto, L.; Jiménez-Barbero, J.; Martínez, Á. T. *J. Agric. Food Chem.* **2011**, 59, 11088–11099.

(33) Kim, H.; Ralph J. *Org. Biomol. Chem.* **2010**, 8, 576-591.

(34) Kim, H.; Ralph J. *RSC Adv.* **2014**, 4, 7549-7560.

(35) Santos, R. B.; Capanema, E. A.; Balakshin, M. Y., Chang, H. M., Jameel, H. *J. Agric. Food Chem.* **2012**, 60, 4923−4930.

(36) Bardet, M.; Robert, D.; Lundquist, K.; Von Unge, S. *Magn. Reson. Chem.* **1998**, 36, 597-600.

(37) Capanema, E.A.; Balakshin, M.Y.; Kadla, J.F. *J. Agric. Food Chem.* **2004**, 52, 1850-1860.

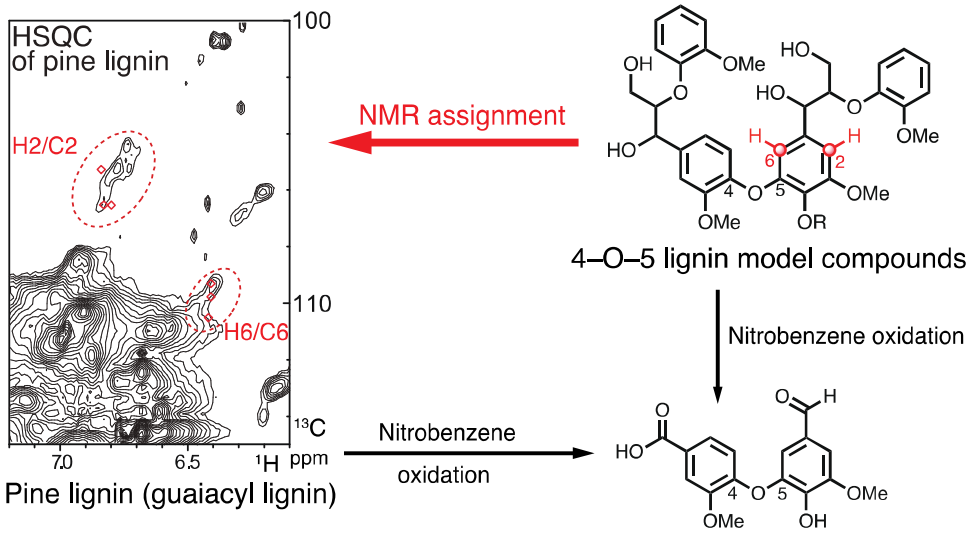
(38) Zhang L.; Gellerstedt G.; Ralph J.; Lu F. *J. Wood Chem. Technol.* **2006**, 26, 65-79.

(39) Martínez, Á. T.; Rencoret, J.; Marques G.; Gutiérrez, A.; Ibarra, D; Jiménez-Barbero, J.; del Río, J. C. *Phytochemistry* **2008**, 69, 2831-2843.

(40) Bardet, M.; Robert, D.; Lundquist, K.; Parkås, J.; Robert, D., von Unge, S. *Magn. Reson. Chem.* **2006**, 44, 976-979.

(41) Lundquist, K.; Li, S.; Parkås, J. *^{1}H NMR database of lignin model compounds in different solvents*, 2009, http://publications.lib.chalmers.se/publication/93716-1h-nmr-database-of-lignin-model-compounds-in-different-solvent

(42) Kishimoto, T.; Uraki, Y.; Ubukata, M. *Org. Biomol. Chem.* **2008**, 6, 2982-2987.



For Table of Contents/Abstract Graphic Only