Biomacromolecules

Article

pubs.acs.org/Biomac

Identification of 4−O−5-Units in Softwood Lignins via Definitive Lignin Models and NMR

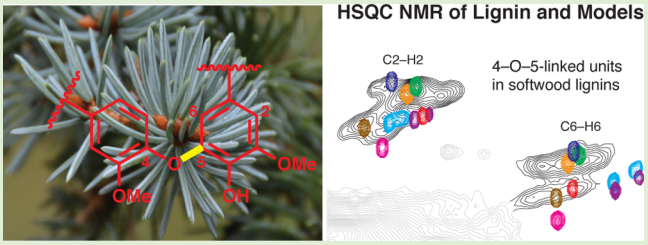
Fengxia Yue,[†,‡] Fachuang Lu,[*,†,‡] Sally Ralph,[§] and John Ralph[‡]

[†]State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou, 510640, China

[‡]Department of Biochemistry and The DOE Great Lakes Bioenergy Research Center, The Wisconsin Energy Institute, University of Wisconsin, Madison, Wisconsin 53726, United States

[§]USDA Forest Products Laboratory, Madison, Wisconsin 53706, United States

ABSTRACT: Lignins are complex and heterogeneous natural polymers in which the major units are characterized by certain prominent interunit linkages. Previous attempts to identify and quantify 4−O−5-linked units in softwood lignins by NMR were not successful. In this work, various lignin model compounds, including the tetramers formed by the 4−O−5-coupling of β−O−4-, β−β-, and β−5-model dimers, were synthesized. Such compounds are better able to model the corresponding structures in lignins than those used previously. 4−O−5-Linked structures could be clearly observed and identified in real softwood lignin samples by comparison of their 2D HSQC NMR spectra with those from the model compounds. When comparing NMR data of phenol-acetylated versus phenol-etherified model compounds with those of acetylated lignins, it was apparent that most of the 4−O−5-linked structures in softwood lignins are present as free-phenolic end units.



INTRODUCTION

Lignins are complex natural polymers that are formed via free-radical coupling of hydroxycinnamyl alcohols in a combinatorial manner, although various factors can dictate the ultimate structural features of lignins.[1,2] Lignins play essential roles in the development of plant cell walls and affect their use as renewable biomaterials.[3] As highly abundant natural aromatic polymers, lignins have drawn much research and industrial attention for decades, including for their potential use as adhesives, additives in biodegradable composite materials, and as stabilizing agents in ceramics and aqueous alumina suspensions for use in advanced materials.[4−6] However, the structure of lignin macromolecules is neither absolutely definable nor determinable because of their complexity and heterogeneity. There is no evidence for regularly repeating units of any size nor for any kind of control over the structure in either a sequence-type manner or, certainly, in terms of stereochemistry; lignin remains best described as a combinatorial racemic polymer for which astronomical numbers of isomers are possible.[1,2] However, lignins are well characterized by the relative frequencies of the various unit types and their characteristic interunit bonding patterns. Lignin structural studies therefore continue to play an important role both in understanding the nature of this enigmatic class of polymers and for optimizing their value for potential applications.[2,7,8]

Softwood lignins contain mostly guaiacyl (G) units derived from the coupling of coniferyl alcohol radicals, primarily in an endwise fashion, with the growing polymer. Low levels of H-units from *p*-coumaryl alcohol that will be ignored from here on are, however, quite prevalent in compression wood zones.[9,10]

Lignification, the polymerization process that generates lignins, starts with dimerization of phenolic coniferyl alcohol radicals, producing β−O−4, β−5, and β−β dehydrodimers, henceforth termed just dimers (Figure 1). In dicots and monocots, the other major monolignol, sinapyl alcohol, an analog of coniferyl alcohol with an additional methoxyl at C5, forms only β−β dimers; in principle, β−O−4 dimers can also be produced from sinapyl alcohol, although they have not been authenticated in biomimetic coupling reactions using peroxidase-H_2O_2.[2,11] As lignification progresses, 5−5- and 4−O−5-type linkages can be formed between two lignin oligomeric phenolic end-units; such interunit bonding does not result from monomer−monomer or monomer−oligomer coupling.[2] Such structures are important because, in principle, they lead to chain branching (but see the discussion below). Another coupling mode, β−1, has been shown to arise from coupling of a monolignol with a preformed β-aryl ether dimeric end-unit. The β−1 structure is relatively minor and does not lead to branching because the phenol originally present on the β-ether unit is trapped as a ketone in the resulting spirodienone unit and is therefore not able to form a phenolic radical for chain elongation at that site.[2,12,13]

As noted, the structure of lignins is mainly characterized by its prominent interunit linkages. The most prevalent are β−O−4-ethers, and the cleavage of these linkages by a degradative method such as thioacidolysis or DFRC results in low molecular weight products containing monomers, dimers,

Received: February 19, 2016
Revised: April 8, 2016
Published: April 14, 2016

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

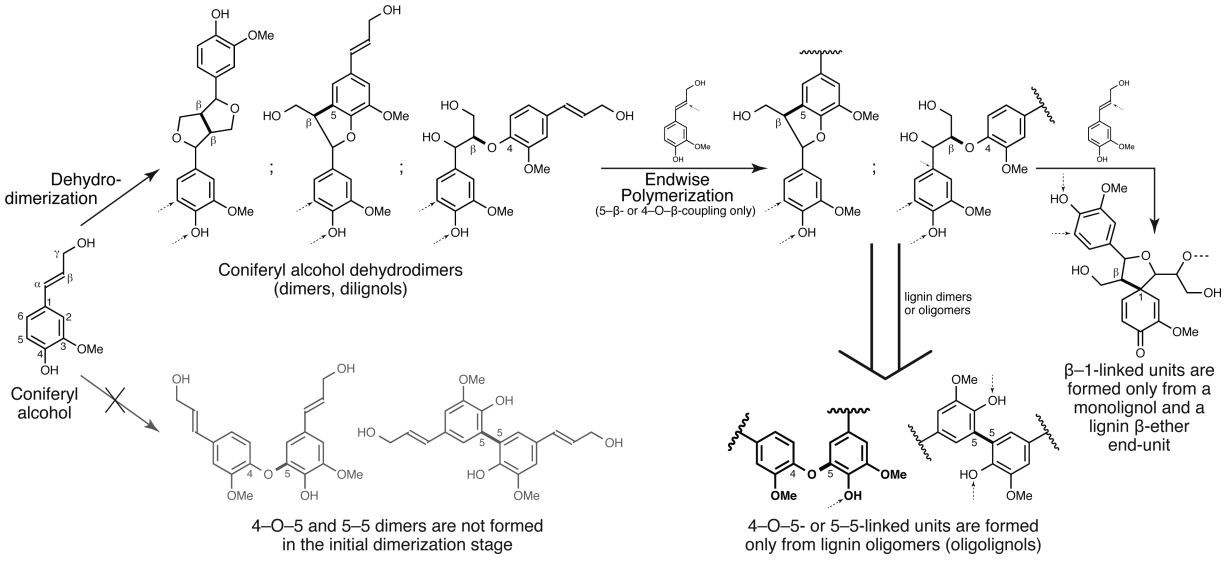


Figure 1. Dehydrodimerization of coniferyl alcohol followed by polymerization to produce (softwood) lignins. Dimerization produces one of three dimers, with at least one of the coniferyl alcohol radicals coupling at its favored *β*-position; that is, the often shown 4−O−5- or 5−5-coupled dimers (lower path, gray) do not arise from coniferyl alcohol in dimerization reactions. Further endwise polymerization occurs by coupling with coniferyl alcohol, invariably at its *β*-position, with the dimer (or higher oligomers) at the 5− or 4−O-position. In a special case, coupling of coniferyl alcohol, at its *β*-position, with a *β*-ether end unit at its 1-position gives rise to *β*−1 units, the spirodienones. Dimers and higher oligomers can 4−O−5- or 5−5-couple to produce the 4−O−5-units (bolded) that are the primary focus of this paper or 5−5-linked units (that go on to form dibenzodioxocins (see Figure 6). Dashed arrows represent sites where further radical coupling can occur during lignification.

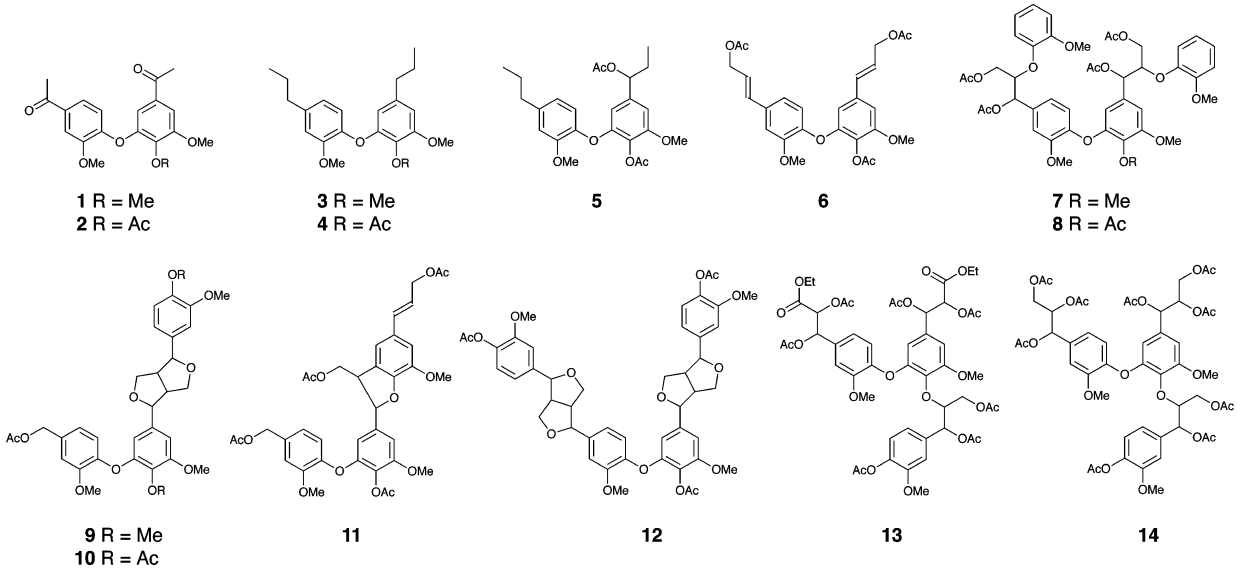


Figure 2. Structures of model compounds **1**−**6** (used previously) and **7**−**14** prepared and used in this study.

trimers, and residual oligomers.[14−16] Analysis of such degraded products, normally via GC or GC-MS, provides an estimation of the so-called "condensed units" characterized by other linkages such as 5−5, *β*−5, *β*−1, *β*−*β*, and 4−O−5.[17,18] It has been estimated that *β*−*β* and 4−O−5 linkages contribute 6 and 5%, respectively, to the total dimers released by thioacidolysis or permanganate oxidation from spruce lignin,[19,20] but the levels in lignin are in fact very difficult to extrapolate back from the relative levels of released dimers.

Characterization of lignins by NMR, especially heteronuclear 2D NMR, has become a routine practice as more and more advanced NMR methods and instruments have become available.[21] As importantly, various lignin model compounds having structures that more accurately model lignin's substructures have been increasingly synthesized to tremendously enhance the power of NMR analysis and have led to the discovery of novel structures in lignins of various origins.[21] Among the 2D NMR experiments, the short-range $^{1}H-^{13}C$ correlation experiment, HSQC (heteronuclear single-quantum coherence) is one of the most used methods for structural identification and relative quantitation because of its sensitivity and the high dispersion of its correlation signals. Coupled with the analysis of appropriate model compounds, it has been possible to identify various lignin units, characterized by their characteristic interunit linkages. In a 2D HSQC spectrum of lignin, the NMR correlations from lignin's aliphatics are well dispersed and diagnostic so that *β*−O−4, *β*−5, *β*−*β*, and *β*−1 units are easily observed and can, in principle, be quantitatively

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

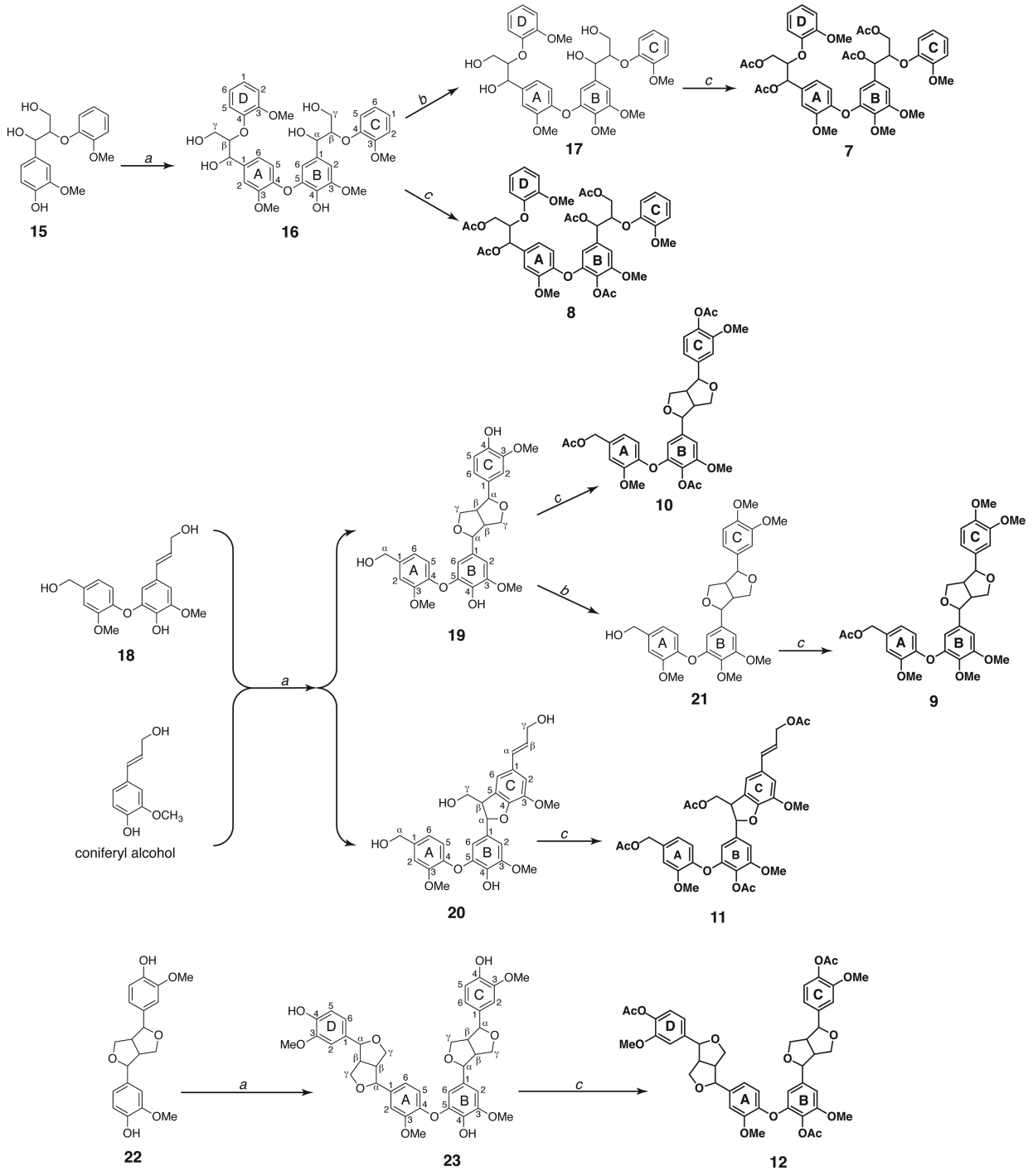


Figure 3. Synthetic routes to model compounds **7**–**12**: (a) peroxidase, H_2O_2; (b) iodomethane, K_2CO_3, acetone; (c) pyridine, acetic anhydride.

estimated if an appropriate pulse sequence and the right conditions are used. However, 4–O–5 and 5–5 units are more difficult to identify directly by the 2D HSQC experiment because carbons involved directly in such linkages have no attached protons and can not, therefore, be detected in such a one-bond $^{1}H-^{13}C$ correlation experiment. However, most of the 5–5-linked units are present as dibenzodioxocins, and these are readily recognized by the diagnostic α- and β-correlations in the aliphatic region of the spectrum. The 4–O–5-linked structures do not have similarly diagnostic aliphatic resonances. Fortunately, however, the units with 4–O–5-linkages are structurally similar to syringyl units with respect to their NMR characteristics suggesting that their C2–H2 and C6–H6 correlations would be different and easily recognized in the aromatic region in a 2D HSQC spectrum of a softwood lignin that contains no syringyl units.

Our previous attempts to identify the 4–O–5 linked units by 2D HSQC NMR were not successful either because the model compounds chosen as references did not sufficiently closely model the structures in lignins and/or because the actual abundance of the 4–O–5-linked units in such lignin samples was below the level that NMR could detect in a timely manner.[22] Model compounds for these units have been limited to simple models **1**–**6** (Figure 2) having 4–O–5 linkages but

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909–1920

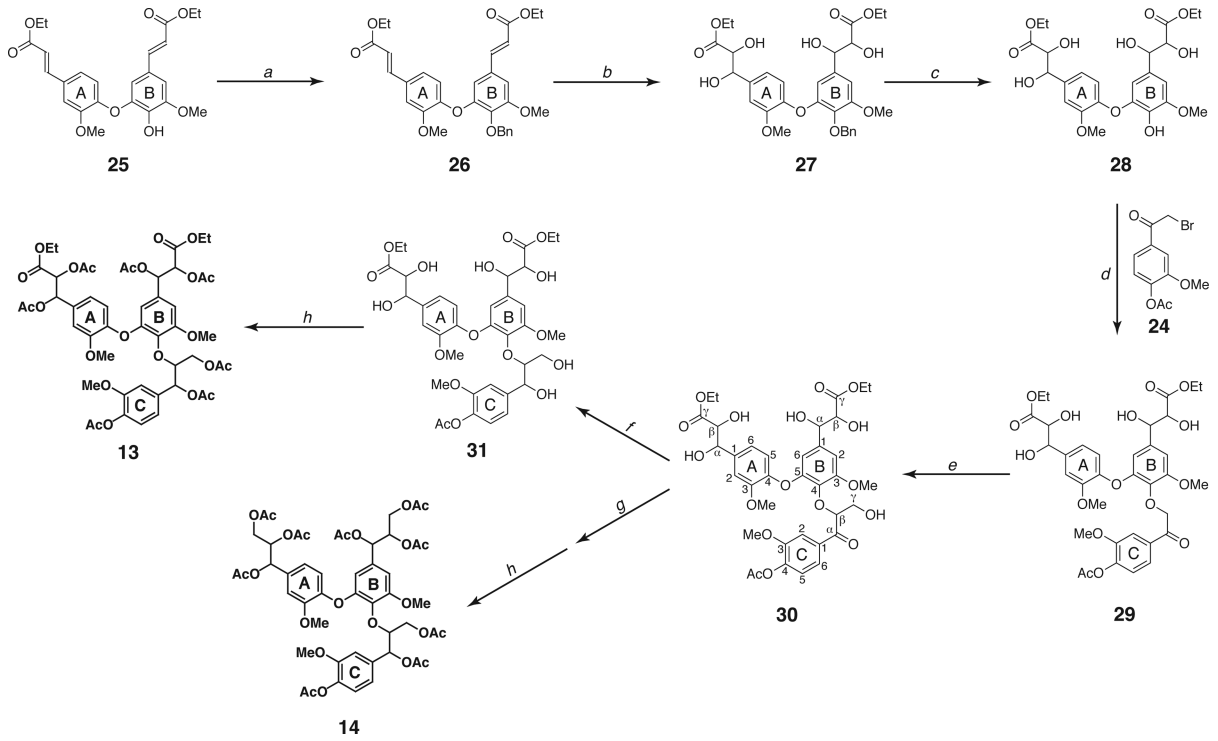


Figure 4. Synthetic routes to model compounds **13** and **14**: (a) benzyl bromide, K_2CO_3, DMF; (b) AD-mix-α, methanesulfonamide, *tert*-butanol/H_2O; (c) H_2, Pd/C, ethanol; (d) K_2CO_3, DMF; (e) K_2CO_3, formaldehyde, 1,4-dioxane; (f) borane-*tert*-butylamine complex, CH_2Cl_2; (g) $NaBH_4$, EtOH; (h) pyridine, acetic anhydride.

with inappropriately substituted side chains. In our recent work to understand 5-linked pinoresinol structures in softwood lignins, a 4−O−5-linked pinoresinol lignin model **10** (after acetylation) was synthesized.[23] Our continued interest in synthesizing improved lignin model compounds for a better understanding of lignin structures led to the syntheses described here of representative 4−O−5-linked model trimers and tetramers **7**−**14** (Figure 2) from dimers with β−O−4-, β−β, and β−5-linkages, models that are superior to those used previously. The NMR data from these models are then used to authenticate correlation peaks belonging to these elusive 4−O−5-units in the spectra of lignins isolated from softwoods.

EXPERIMENTAL SECTION

Materials. All chemicals and solvents used in this study were purchased from Aldrich (Milwaukee, WI, U.S.A.) and used as supplied.

Flash chromatography was performed with Biotage snap silica cartridges on an Isolera One (Biotage, Charlottesville, VA). All synthesized compounds were characterized by the usual array of NMR and GC-MS methods. NMR spectra were acquired on a Bruker Biospin (Billerica, MA, U.S.A.) AVANCE 500 (500 MHz) spectrometer fitted with a cryogenically cooled 5 mm TCI gradient probe with inverse geometry (proton coils closest to the sample) and spectral processing used Bruker's Topspin 3.1 (Mac) software. Standard Bruker implementations of one- and two-dimensional (gradient-selected COSY, HSQC, and HMBC) NMR experiments were used for routine structural assignments of newly synthesized compounds. The conditions used for all samples were 5−10 mg in 0.5 mL of NMR solvent (acetone-d_6) with the central solvent peaks (δ_H/δ_C 2.04/29.80) used as internal reference.

Synthesis of Model Compounds. The origins of compounds **1**−**6** have been described previously.[22] As shown in Figures 3 and 4, model compounds **7**−**14** were synthesized via multistep routes.

Synthesis of Model Compounds 7 and 8. Compound **16** was synthesized conveniently but in low yield from peroxidase-catalyzed homocoupling of compound **15** as follows. Compound **15** (805 mg, 2.51 mmol) was dissolved in acetone (40 mL) to which phosphate buffer (220 mL, pH 5.0) was added. Then H_2O_2−urea complex (130 mg, 1.38 mmol) dissolved in buffer (5 mL) was added into the acetone−buffer system and followed by addition of horseradish peroxidase (EC 3.2.1.4, 5 mg in 5 mL buffer). The solution turned brown slowly over about 5 min and then dark brown (by about 15 min) after the addition of the peroxidase. The solution became red in 30 min and the color remained until the end of the reaction. The mixture was kept stirring at room temperature and monitored by TLC (CH_2Cl_2/MeOH, 20:1, v/v). The reaction was continued for 70 min, after which the starting material **15** had been consumed and the reaction mixture was then acidified with HCl (1 M, 3 mL) and extracted with ethyl acetate (100 mL × 3). The combined organic phase was washed with saturated aqueous NaCl, dried over anhydrous $MgSO_4$, and filtered through sintered glass, and the solvent from the eluant was evaporated under reduced pressure at 40 °C. All of the products (755 mg) were loaded onto 1 mm normal-phase silica-gel plates (about 50 mg/plate) and developed with CH_2Cl_2/MeOH (20:1, v/v) as eluting solvent. Compound **16** (yellow oil) was a minor product separated from the mixture with an isolated yield of 2.9%. Each isolated product was characterized by NMR. Compound **16**, NMR, δ_H: 3.74/3.80 (3H/3H, s/s, C-OMe/D-OMe), 3.78 (3H, s, A-OMe), 3.84 (3H, s, B-OMe), 3.71 (2H, m, Aγ), 3.82 (2H, m, Bγ), 4.21 (1H, m, Bβ), 4.31 (1H, m, Aβ), 4.80/4.82 (1H, d, J = 5.10 Hz, Bα), 4.93/4.95 (1H, d, J = 5.10 Hz, Aα), 6.59/6.60 (1H, d, J = 2.19 Hz, B6), 6.66/6.67 (1H, d, J = 8.30 Hz, A5), 6.80/6.84 (2H, C6/D6), 6.91 (2H, B2/A6), 6.92/6.95 (2H, m, C1/D1), 6.95 (2H, m, C2/D2), 6.96 (2H, m, C5/D5), 7.21/7.22 (1H, d, J = 1.58 Hz, A2). δ_C: 56.14/56.20 (C-OMe/D-OMe), 56.15 (A-OMe), 56.50 (B-OMe), 61.69/61.70 (Aγ), 61.81 (Bγ), 73.57 (Aα), 73.64 (Bα), 86.44/86.46 (Aβ), 86.47/86.50 (Bβ), 106.92/106.94 (B2), 111.23/111.27 (B6), 112.65/112.67 (A2), 113.35/113.39/113.40 (C2/D2), 118.39/118.46 (A5), 119.45/119.49/119.58 (C5/D5), 120.10 (A6), 121.77/121.82/121.85 (C6/D6), 123.22/123.24/123.35 (C1/D1), 133.60 (B1), 137.80/137.82 (B5), 138.40/138.45 (A1), 144.78/144.82 (B4), 146.37/146.38 (A4), 148.91/148.92/148.94 (C4/D4), 149.19 (B3), 150.89/150.92 (A3),

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

151.80/151.82/151.88 (C3/D3). HR-MS (ESI) Calcd for $C_{34}H_{38}NaO_{12}$ [(M + Na)$^+$], 661.2256; found, 661.2249.

Compound **17** was prepared from compound **16** via methylation. Compound **16** (5 mg, 8 μmol) was dissolved in acetone (5 mL) to which K_2CO_3 powder (11 mg, 80 μmol) was added, at which point the solution became light yellow. Then iodomethane (0.1 mL, Aldrich, 99%, stabilized with copper) was added into the reaction solution and the vessel was sealed. The resultant mixture was kept stirring at room temperature, overnight. The reaction was checked by TLC (CH_2Cl_2/MeOH, 20:1, v/v) to ensure that no starting material remained. After the reaction was completed, the inorganics were filtered off and the filtrate was evaporated under reduced pressure at 40 °C to remove acetone (but not to dryness). After acidification with HCl (1 M, 1 mL), the mixture was extracted with ethyl acetate (100 mL × 3). The combined organic phase was washed with saturated aqueous NaCl and dried over anhydrous $MgSO_4$, filtered, and the solvent evaporated under reduced pressure at 40 °C. Compound **17** was obtained in almost quantitative yield and no further purification was needed. Compound **17**, NMR, δ_H: 3.71 (3H, s, B4-OMe), 3.74/3.81 (6H, C-OMe/D-OMe), 3.76 (3H, s, A-OMe), 3.83 (3H, s, B3-OMe), 3.7−3.80 (4H, Aγ/Bγ), 4.20 (1H, m, Bβ), 4.31 (1H, m, Aβ), 4.79−4.84 (1H, Bα), 4.94/4.96 (1H, d, *J* = 4.85 Hz, Aα), 6.52/6.53 (1H, d, *J* = 1.88 Hz, B6), 6.70/6.71 (1H, d, *J* = 8.20 Hz, A5), 6.78−6.97 (8H, C1/C2/C5/C6/D1/D2/D5/D6), 6.90 (1H, B2), 6.94 (1H, A6), 7.24 (1H, d, *J* = 1.58 Hz, A2). δ_C: 56.14/56.21 (A-OMe/C-OMe/D-OMe), 56.31 (B3-OMe), 60.68 (B4-OMe), 61.74 (γ), 73.58 (Aα/Bα), 86.29/86.35 (Bβ), 86.50/86.57 (Aβ), 106.95/106.98 (B2), 110.26/110.31 (B6), 112.79 (A2), 113.36/113.41/113.43 (C2/D2), 119.46 (A5), 119.53/119.58/119.62 (C5/D5), 120.16 (A6), 121.75/121.83/121.87 (C6/D6), 123.28/123.31/123.37 (C1/D1), 138.65 (B1), 139.08/139.14 (A1), 139.69/139.71 (B4), 145.80/145.82 (A4), 148.86/148.94 (C4/D4), 151.24/151.28 (B5), 151.30/151.33 (A3), 151.82/151.83/151.91 (C3/D3), 154.42 (B3).

Compounds **7** and **8** (Figure 3) were obtained quantitatively via acetylation of **17** and **16** with acetic anhydride−pyridine (1 mL, 1:1, v/v), followed by the usual workup. Compound **7**, NMR, δ_H: 1.91 (3H, s, γ-OAc), 1.95 (3H, s, γ-OAc), 1.97/1.98 (3H, s, Bα-OAc), 2.07 (3H, s, Aα-OAc), 3.72/3.82, (6H/3H, s/s, B4-OMe/C3-OMe/D3-OMe), 3.81 (3H, s, A3-OMe), 3.87 (3H, s, B3-OMe), 4.15 (1H, dd, *J* = 11.95, 1.67 Hz, Bγ1), 4.21 (1H, dd, *J* = 11.95, 3.92 Hz, Aγ1), 4.32 (1H, dd, *J* = 11.95, 1.91 Hz, Bγ2), 4.37 (1H, dd, *J* = 11.95, 1.91 Hz, Aγ2), 4.72 (1H, m, Bβ), 4.85 (1H, m, Aβ), 5.90 (1H, d, *J* = 5.15 Hz, Bα), 6.05 (1H, d, *J* = 5.15 Hz, Aα), 6.51/6.52 (1H, d, *J* = 2.40 Hz, B6), 6.76/6.77 (1H, d, *J* = 8.24 Hz, A5), 6.80−6.99 (8H, C1/C2/C5/C6/D1/D2/D5/D6), 6.92 (1H, B2), 6.99 (1H, A6), 7.27 (1H, d, *J* = 1.82 Hz, A2). δ_C: 20.61/20.64 (γ-OAcs), 20.77/20.79 (Bα-OAc), 20.88 (Aα-OAc), 56.03/56.26 (C-OMe/D-OMe), 56.14 (A-OMe), 56.44 (B3-OMe), 60.71 (B4-OMe), 63.10/63.12 (Aγ/Bγ), 74.54 (Bα), 74.56 (Aα), 80.24 (Aβ), 80.27 (Bβ), 107.57 (B2), 110.79/110.81 (B6), 113.21 (A2), 113.58/113.65/113.67 (C2/D2), 119.58 (A5), 119.70/119.72 (C5/D5), 120.75/120.77 (A6), 121.53/121.58/121.59 (C6/D6), 123.94/124.00 (C1/D1), 133.43 (B1), 133.95 (A1), 140.44 (B4), 146.43/146.45 (A4), 148.24/148.27 (C4/D4), 151.09/151.10 (B5), 151.42 (A3), 151.84/151.94/151.97 (C3/D3), 154.73 (B3), 169.78/169.80 (Bα-OAc) 169.92 (Aα-OAc), 170.21/170.72/170.76 (γ-OAc). Compound **8**, NMR, δ_H: 1.90 (3H, s, Bγ-OAc), 1.96 (3H, s, Aγ-OAc), 1.98 (3H, s, Bα-OAc), 2.07 (3H, s, Aα-OAc), 2.14 (3H, s, Ar-OAc), 3.71 (3H, s, D-OMe), 3.78 (3H, s, A-OMe), 3.82 (3H, s, C-OMe), 3.85 (3H, s, B-OMe), 4.15 (1H, dd, *J* = 11.95, 1.94 Hz, Bγ1), 4.20 (1H, dd, *J* = 11.95, 3.98 Hz, Aγ1), 4.33 (1H, dd, *J* = 11.95, 2.35 Hz, Bγ2), 4.37 (1H, dd, *J* = 11.95, 2.50 Hz, Aγ2), 4.71 (1H, m, Bβ), 4.85 (1H, m, Aβ), 5.94 (1H, d, *J* = 5.16 Hz, Bα), 6.05 (1H, d, *J* = 5.16 Hz, Aα), 6.51/6.52 (1H, d, *J* = 2.20 Hz, B6), 6.82/6.85 (C6/D6), 6.82/6.83 (1H, d, *J* = 8.24 Hz, A5), 6.91−7.01 (C1/D1/C2/D2/C5/D5), 6.92 (1H, d, *J* = 1.55 Hz, B2), 7.00 (1H, A6), 7.27/7.28 (1H, d, *J* = 1.50 Hz, A2). δ_C: 20.13 (Ar-OAc), 20.59 (Bγ-OAc), 20.64 (Aγ-OAc), 20.74/20.76 (Bα-OAc), 20.87 (Aα-OAc), 56.03 (C-OMe), 56.14 (D-OMe), 56.26 (A-OMe), 56.57 (B-OMe), 62.94 (Bγ), 63.05/63.08 (Aγ), 74.47 (Bα), 74.56 (Aα), 80.22 (Aβ), 80.25 (Bβ), 106.59 (B2), 109.69 (B6), 113.28/113.30 (A2), 113.61/113.64/113.66 (C2/D2), 119.54/119.67/119.95 (C5/D5), 120.79 (A5), 120.83 (A6), 121.58 (C6/D6), 123.97/124.02/124.08 (C1/D1), 130.78 (B4), 134.84 (A1), 136.27 (B1), 145.49/145.50 (A4), 148.16/148.19/148.23 (C4/D4), 150.82 (B5), 151.77 (A3), 151.90/151.92/151.96 (C3/D3), 153.56 (B3), 168.32 (Ar-OAc), 169.77/169.79 (Bα-OAc) 169.93 (Aα-OAc), 170.71/170.72/170.75 (γ-OAc).

Synthesis of Model Compounds 9−11. Compounds **19** and **20** (Figure 3) were each synthesized via the peroxidase-catalyzed free-radical coupling reaction of coniferyl alcohol with a synthesized 5−O-linked coniferyl alcohol compound **18**, as published.[23] Note that compound **18** can only be obtained synthetically as radical coupling with coniferyl alcohol always occurs via its coupling at its β-position.[1,2,11] Compound **21** was produced by methylation of **19** under similar conditions to those described above for compound **17**. Compound **21**, NMR, δ_H: 3.01 (2H, m, Bβ + Cβ), 3.71 (6H, s, B4-OMe/C4-OMe), 3.75 (Bγ), 3.77 (2H, m, Bγ1 + Cγ1), 3.78 (3H, s, C3-OMe), 3.79 (3H, s, A3-OMe), 3.86 (3H, s, B3-OMe), 4.12 (1H, m, Bγ2), 4.17 (1H, m, Cγ2), 4.61 (2H, d, *J* = 5.68 Hz, Aα), 4.62 (1H, d, *J* = 4.30 Hz, Cα), 4.63 (1H, d, *J* = 4.30 Hz, Bα), 6.41 (1H, d, *J* = 1.68 Hz, B6), 6.79 (1H, d *J* = 1.81 Hz, B2), 6.83 (1H, A5), 6.87 (2H, C5/C6), 6.88 (1H, A6), 6.95 (1H, br-s, C2), 7.13 (1H, d, *J* = 1.70 Hz, A2); δ_C: 55.01 (Bβ), 55.32 (Cβ), 55.96 (C3-OMe), 56.11 (A3-OMe), 56.37 (B3-OMe), 59.95 (B4-OMe/C4-OMe), 66.35 (Aα), 72.20 (Bγ), 72.20 (Cγ), 86.10 (Bα), 86.37 (Cα), 105.54 (B2), 108.53 (B6), 110.76 (C2), 112.31 (C5), 112.38 (A2), 118.97 (C6), 119.57 (A6), 120.36 (A5), 135.10 (C1), 138.36 (B1), 139.48 (B4), 139.96 (A1), 145.08 (A4), 149.68 (C4), 150.30 (C3), 151.76 (A3), 151.94 (B5), 154.85 (B3).

Compounds **9−11** were obtained via acetylation of **21**, **19**, and **20** with acetic anhydride−pyridine (1 mL, 1:1. v/v). Compound **9**, NMR, δ_H: 2.04 (3H, s, Aα-OAc), 3.72 (6H, s, B4-OMe/C4-OMe), 3.76/3.77 (2H, m, Bγ1 + Cγ1), 3.78 (3H, s, C3-OMe), 3.82 (3H, s, A3-OMe), 3.87 (3H, s, B3-OMe), 4.13/4.18 (1H/1H, dd/dd, *J* = 8.48, 6.83/8.48, 6.83 Hz, Bγ2 + Cγ2), 4.64 (1H, d, *J* = 4.60 Hz, Cα), 4.65 (1H, d, *J* = 4.24 Hz, Bα), 5.05 (2H, s, Aα), 6.48 (1H, d, *J* = 1.68 Hz, B6), 6.82 (1H, d, *J* = 6.51, A5), 6.83 (1H, s, B2), 6.87 (2H, C5/C6), 6.92 (1H, dd, *J* = 8.30, 1.96 Hz, A6), 6.95 (1H, s, C2), 7.15 (1H, d, *J* = 1.79 Hz, A2); δ_C: 20.81 (Aα-OAc), 55.04 (Bβ), 55.32 (Cβ), 55.96 (C3-OMe), 56.25 (A3-OMe), 56.39 (B3-OMe), 59.98 (B4-OMe/C4-OMe), 66.24 (Aα), 72.24 (Bγ), 72.27 (Cγ), 86.06 (Bα), 86.37 (Cα), 106.07 (B2), 109.31 (B6), 110.78 (C2), 112.39 (C5), 114.03 (A2), 118.98 (C6), 119.72 (A5), 121.60 (A6), 133.39 (A1), 135.10 (C1), 138.58 (B1), 139.87 (B4), 146.66 (A4), 149.66 (C4), 150.31 (C3), 151.25 (B5), 151.54 (A3), 154.92 (B3), 170.85 (Aα-OAc). Compound **10**, NMR, δ_H: 2.04 (3H, s, Aα-OAc), 2.16/2.21 (3H/3H, s/s, B4-OAc/C4-OAc), 3.06 (2H, m, Bβ + Cβ), 3.80 (3H, s, A-OMe), 3.80 (3H, s, C-OMe), 3.84 (3H, s, B-OMe), 3.84 (2H, m, Bγ1 + Cγ1), 4.17 (1H, m, Bγ2), 4.22 (1H, m, Cγ2), 4.69 (1H, d, *J* = 4.0 Hz, Bα), 4.72 (1H, d, *J* = 4.1 Hz, Cα), 5.06 (2H, s, Aα), 6.48 (1H, d, *J* = 1.8 Hz, B6), 6.86 (1H, d, *J* = 1.6 Hz, B2), 6.89 (1H, d, *J* = 8.2 Hz, A5), 6.92 (1H, dd, *J* = 8.1, 1.8 Hz, C6), 6.94 (1H, dd, *J* = 8.2, 1.8 Hz, A6), 6.99 (1H, d, *J* = 8.1 Hz, C5), 7.09 (1H, d, *J* = 1.8 Hz, C2), 7.15 (1H, d, *J* = 1.8 Hz, A2); δ_C: 20.17/20.43 (B4-OAc/C4-OAc), 20.80 (Aα-OAc), 55.19 (Bβ), 55.34 (Cβ), 56.11 (C-OMe), 56.26 (B-OMe), 56.52 (A-OMe), 66.15 (Aα), 72.44 (γs), 85.93 (Bα), 86.12 (Cα), 105.01 (B2), 108.08 (B6), 110.95 (C2), 114.02 (A2), 118.54 (C6), 121.10 (A5), 121.58 (A6), 123.41 (C5), 130.41 (B4), 134.37 (A1), 140.02 (C4), 141.38 (C1), 141.68 (B1), 145.67 (A4), 151.06 (B5), 151.98 (A3), 152.20 (C3), 153.70 (B3), 168.49/169.03 (B4-OAc/C4-OAc), 170.85 (Aα-OAc). HR-MS (ESI) Calcd for $C_{34}H_{40}NO_{12}$ [(M + NH_4)$^+$], 654.2546; found, 654.2517. Compound **11**, NMR, δ_H: 1.94 (3H, s, Bγ-OAc), 2.01 (3H, s, Cγ-OAc), 2.06 (3H, s, Aα-OAc), 2.18 (3H, s, B4-OAc), 3.69 (1H, m, Bβ), 3.74 (3H, s, A-OMe), 3.82 (3H, s, C-OMe), 3.83 (3H, s, B-OMe), 4.24 (1H, m, Bγ1), 4.40 (1H, m, Bγ2), 4.65 (2H, dd, *J* = 6.5, 1.2 Hz, Cγ), 5.06 (2H, s, Aα), 5.50 (1H, d, *J* = 6.1 Hz, Bα), 6.22 (1H, dt, *J* = 15.9, 6.5 Hz, Cβ), 6.47 (1H, d, *J* = 1.6 Hz, B6), 6.62 (1H, dt, *J* = 15.9, 1.2 Hz, Cα), 6.90 (1H, d, *J* = 1.8 Hz, B2), 6.91 (1H, d, *J* = 8.2 Hz, A5), 6.93 (1H, dd, *J* = 8.2, 1.6 Hz, A6), 7.0 (1H, br-s, C6), 7.01 (1H, br-s, C2), 7.12 (1H, d, *J* = 1.6 Hz, A2); δ_C: 20.16 (B4-OAc), 20.65 (Cγ-OAc), 20.78/20.81 (Aα-OAc/Bγ-OAc), 51.45 (Bβ), 56.14

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

(A-OMe), 56.45 (C-OMe), 56.55 (B-OMe), 65.44 (Cγ), 65.97 (Bγ), 66.11 (Aα), 88.01 (Bα), 104.50 (B2), 107.58 (B6), 112.33 (C2), 113.93 (A2), 116.38 (C6), 121.57 (A6), 121.64 (A5), 122.32 (Cβ), 128.52 (C5), 130.43 (B4), 131.68 (C1), 134.54 (Cα), 134.74 (A1), 140.70 (B1), 145.10 (A4), 145.30 (C3), 149.08 (C4), 151.37 (B5), 152.14 (A3), 153.83 (B3), 168.42 (B4-OAc), 170.73 (Cγ-OAc), 170.83/170.86 (Aα-OAc/Bγ-OAc). HR-MS (ESI) Calcd for $C_{36}H_{42}NO_{13}$ [$(M + NH_4)^+$], 696.2651; found, 696.2668.

Synthesis of Model Compound 12. Peroxidase-catalyzed homocoupling of pinoresinol, compound **22** (225 mg, 0.63 mmol), was conducted under similar conditions to the homocoupling reaction of compound **15** described above. Following the normal workup, the product mixture (180 mg) was loaded onto 1 mm thick normal-phase silica-gel plates (about 50 mg/plate) and developed with CH_2Cl_2/MeOH, (20:1, v/v) as eluting solvent. Each isolated product was characterized by NMR. Compound **23** was a minor product separated with an isolated yield of 3.6%. Compound **23**, NMR, δ_H: 3.03 (2H, m, Bβ + Cβ), 3.10 (2H, m, Aβ + Dβ), 3.75/3.82/4.12/4.16/4.21 (8H, Aγ/Bγ/Cγ/Dγ), 3.81/3.83 (3H/3H, s/s, C-OMe/D-OMe), 3.84 (3H, s, A-OMe), 3.86 (3H, s, B-OMe), 4.61/4.62 (1H/1H, d/d, J = 4.80/4.80 Hz, Cα/Dα), 4.66 (1H, s, J = 4.41, Bα), 4.73 (1H, s, J = 4.41, Aα), 6.52 (1H, d, J = 1.68, B6), 6.76 (1H, d, J = 8.44, A5), 6.77/6.78 (1H/1H, s/s, C5/D5), 6.82 (1H, d, J = 1.20, B2), 6.80/6.83 (1H/1H, dd/dd, J = 8.35, 1.80/8.35, 1.80 Hz, C6/D6), 6.88 (1H, dd, J = 8.31, 1.85 Hz, A6), 6.95 (1H, d, J = 1.82, C2), 6.98 (1H, d, J = 1.82, D2), 7.12 (1H, d, J = 1.82, A2); δ_C: 55.06 (Bβ), 55.14/55.20 (Cβ/Dβ), 55.28 (Aβ), 56.14/56.15 (C-OMe/D-OMe), 56.23 (A-OMe), 56.55 (B-OMe), 72.11 (Bγ), 72.16 (Aγ), 72.19/72.30 (Cγ/Dγ), 86.24 (Bα), 86.32 (Aα), 86.53/86.56 (Cα/Dα), 106.05 (B2), 110.41/110.42 (B6), 110.46/110.50 (C2/D2), 111.65 (A2), 115.46/115.48 (C5/D5), 118.73 (A5), 119.01 (A6), 119.53/119.57 (C6/D6), 133.41 (B1), 133.99 (C1/D1), 138.00 (B4), 138.38 (A1), 144.93 (B5), 146.53/146.54 (A4), 146.79/146.82 (C4/D4), 148.25/148.27 (C3/D3), 149.63 (B3), 151.26 (A3). HR-MS (ESI) Calcd for $C_{40}H_{42}NaO_{12}$ [$(M + Na)^+$], 737.2569; found, 737.2559.

Model compound **12** was obtained via acetylation of compound **23** with acetic anhydride−pyridine (1 mL, 1:1, v/v). Compound **12**, NMR, δ_H: 2.16 (3H, s, B-OAc), 2.21/2.22 (6H, s/s, C4-OAc/D4-OAc), 3.05 (2H, m, Bβ + Cβ), 3.14 (2H, m, Aβ + Dβ), 3.80 (6H, s, C-OMe/D-OMe), 3.81 (3H, s, A-OMe), 3.84 (3H, s, B-OMe), 3.83/3.89/4.16/4.22/4.26/4.28 (8H, Aγ/Bγ/Cγ/Dγ), 4.68 (1H, s, J = 4.08, Bα), 4.72 (1H, s, J = 4.08, Aα), 4.78/4.79 (1H/1H, d/d, J = 4.45/4.45 Hz, Cα/Dα), 6.46 (1H, d, J = 1.38, B6), 6.85 (1H, d, J = 1.60, B2), 6.88 (1H, d, J = 8.22, A5), 6.92/6.96 (C6/D6), 6.94 (1H, dd, J = 4.72, 1.70 Hz, A6), 6.98/7.00 (C5/D5), 7.09/7.13 (1H/1H, d/d, J = 1.60/1.68 Hz, C2/D2), 7.16 (1H, d, J = 1.78, A2); δ_C: 20.19 (B-OAc), 20.45 (C-OAc/D-OAc), 55.19/55.20 (Bβ/Cβ), 55.30/55.33 (Aβ/Dβ), 56.12/56.13 (C-OMe/D-OMe), 56.23 (A-OMe), 56.50 (B-OMe), 72.43/72.45/72.50 (Aγ/Bγ/Cγ/Dγ), 85.93 (Bα), 86.13 (Cα/Dα), 86.21 (Aα), 104.82 (B2), 107.88/107.91 (B6), 110.94/110.98 (C2/D2), 111.78 (A2), 118.53/118.56 (C6/D6), 119.17 (A6), 121.23 (A5), 123.42/123.43 (C5/D5), 130.09 (B4), 139.98 (C4/D4), 140.02 (A1), 141.29 (B1), 141.68/141.78 (C1/D1), 144.95/144.97 (A4), 151.26 (B5), 152.11 (A3), 152.20/152.21 (C3/D3), 153.68 (B3), 168.50 (B4-OAc), 169.02 (C4/D4-OAc).

Synthesis of Model Compounds 13 and 14. Compounds **24** and **25** (Figure 4) were synthesized as reported.[24,25] Model compounds **13** and **14** were synthesized via the multistep routes shown in Figure 4.

Compound **26** was synthesized by benzylation of compound **25** with benzyl bromide to protect the phenolic hydroxyl groups, which was accomplished in the traditional way. Diferulate **25** (5.40 g, 12.20 mmol) was dissolved in 20 mL of *N*,*N*-dimethylformamide (DMF), to which anhydrous K_2CO_3 powder (2.53 g, 18.3 mmol) and 1.6 mL of benzyl bromide (2.29 g, 13.42 mmol) were added. The resultant mixture was stirred overnight, after which time TLC (hexanes/EtOAc, 3:1, v/v) showed no compound **25** remaining. The DMF was removed by evaporation at 55 °C under reduced pressure. EtOAc and water (200 mL each) were added to the mixture, followed by the addition of 1 M HCl to quench the residual base. After all liquids were transferred to a separatory funnel, the two phases were separated. The water phase was re-extracted with EtOAc (100 mL). The combined EtOAc phase was washed with saturated NH_4Cl, dried over anhydrous $MgSO_4$, filtered, and the eluant evaporated to dryness. The crude product was purified by flash chromatography (Biotage, 100 g silica column) using hexanes/EtOAc (3:1, v/v) to obtain compound **26** as light yellow oil (5.5 g, 10.30 mmol, 84.4% isolated yield). Compound **26**, NMR, δ_H: 1.24 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.28 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 3.91 (3H, s, B-OMe), 3.96 (3H, s, A-OMe), 4.16 (2H, q, J = 7.1 Hz, B-OCH$_2$Me), 4.19 (2H, q, J = 7.1 Hz, A-OCH$_2$Me), 5.08 (2H, s, Bn-CH$_2$), 6.43 (1H, d, J = 15.9 Hz, Bβ), 6.50 (1H, d, J = 15.9 Hz, Aβ), 6.85 (1H, d, J = 8.3 Hz, A5), 6.86 (1H, d, J = 1.9 Hz, B6), 7.19 (1H, dd, J = 8.3, 1.9 Hz, A6), 7.24 (1H, d, J = 1.9 Hz, B2), 7.25 (1H, m, Bn-4), 7.26−7.30 (2H, Bn, 3/5), 7.37 (2H, dd, J = 7.3, 1.7 Hz, Bn, 2/6), 7.48 (1H, d, J = 1.9 Hz, A2), 7.54 (1H, d, J = 15.9 Hz, Bα), 7.63 (1H, d, J = 15.9 Hz, Aα); δ_C: 14.54 (B-OCH$_2$Me), 14.58 (A-OCH$_2$Me), 56.34 (A-OMe), 56.63 (B-OMe), 60.64 (B-OCH$_2$Me), 60.68 (A-OCH$_2$Me), 75.33 (Bn-CH$_2$), 108.31 (B2), 112.62 (A2), 113.09 (B6), 118.12 (Aβ), 118.85 (Bβ), 119.34 (A5), 122.82 (A6), 128.52 (Bn-C4), 128.84 (Bn-C3/5), 128.88 (Bn-C2/6), 131.24 (B1), 131.52 (A1), 138.62 (Bn-C1), 141.57 (B4), 144.45 (Bα), 144.73 (Aα) 148.79 (A4), 151.03 (B5), 151.65 (A3), 155.45 (B3), 166.89 (Bγ), 167.08 (Aγ).

Compound **27** was obtained by Sharpless catalytic asymmetric dihydroxylation of compound **26** using AD-mix-α as the oxidant.[26] To a well-stirred *tert*-BuOH/H_2O (120 mL, 1:1, v/v) solution at 0 °C (ice−water bath) was added AD-mix-α (1.4 g/mmol of substrate based on each double bond), followed by the addition of methanesulfonamide (1.0 equiv, based on each double bond). The resultant mixture was continuously stirred for 5 min before compound **26** (3.1 g, 5.80 mmol) was added. The resulting solution was well-stirred at 0 °C for 2 h, and then the mixture was warmed to room temperature and stirred for 24 h. The reaction was monitored by TLC (CH_2Cl_2/MeOH, 20:1, v/v) to ensure that no starting material remained. The reaction was quenched by adding excess saturated $Na_2S_2O_3$ solution (50 mL). Then, the resultant solution was kept stirring for an additional 30 min, and the color of the resultant solution became dark green. The product was extracted with EtOAc and water (200 mL × 2), and the combined organic phase was washed with distilled water and saturated NH_4Cl solution, dried over anhydrous $MgSO_4$, filtered, and concentrated under reduced pressure. The residue was purified by flash chromatography (Biotage, 100 g silica column) using CH_2Cl_2/MeOH (20:1, v/v) to give compound **27** (2.0 g, 3.3 mmol, 57.4% isolated yield). Note that compound **27** (and subsequently compounds **28** and **29**) should be produced as largely a single isomer, the isomer in which each side chain is the αS/βR isomer, because of the fairly selective asymmetric dihydroxylation reaction using AD-mix-α.[26] Compound **27**, NMR, δ_H: 1.17 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.20 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 3.80 (3H, s, A-OMe), 3.87 (3H, s, B-OMe), 4.10 (2H, q, J = 7.1 Hz, B-OCH$_2$Me), 4.13 (2H, q, J = 7.1 Hz, A-OCH$_2$Me), 4.17/4.19 (1H, d, J = 3.6, Bβ), 4.25/4.27 (1H, d, J = 3.6, Aβ), 4.84/4.85 (1H, d, J = 3.7 Hz, Bα), 4.96/4.97 (1H, d, J = 3.7 Hz, Aα), 5.03 (2H, s, Bn-CH$_2$), 6.54 (1H, d, J = 1.8 Hz, B6), 6.81 (1H, d, J = 8.2 Hz, A5), 6.91 (1H, d, J = 1.8 Hz, B2), 6.95 (1H, dd, J = 8.2, 1.8 Hz, A6), 7.23 (1H, d, J = 1.8 Hz, A2), 7.25−7.32 (3H, Bn, 3/4/5), 7.44 (2H, dd, J = 7.3, 1.6 Hz, Bn, 2/6); δ_C: 14.43 (B-OCH$_2$Me), 14.47 (A-OCH$_2$Me), 56.20 (A-OMe), 56.41 (B-OMe), 61.41 (B-OCH$_2$Me), 61.43 (A-OCH$_2$Me), 75.18 (Aα), 75.20 (Bα), 75.26 (Bn-CH$_2$), 76.48 (Aβ/Bβ), 106.70 (B2), 110.03 (B6), 112.62 (A2), 119.54 (A5), 119.77 (A6), 128.34 (Bn-C4), 128.83 (Bn-C3/5), 128.88 (Bn-C2/6), 138.29 (B1), 138.63 (B4), 138.73 (A1), 139.22 (Bn-C1), 145.88 (A4), 151.35 (A3), 151.40 (B5), 154.65 (B3), 172.88 (Bγ), 173.05 (Aγ).

Compound **28** was obtained by debenzylation of compound **27** via standard catalytic hydrogenation with palladium on activated carbon (Pd/C, 5% Pd) in ethanol. Compound **27** (400 mg, 0.67 mmol) was dissolved in ethanol (10 mL) with stirring, followed by addition of 50 mg of Pd/C. The resultant mixture was stirred under a hydrogen-filled balloon for 3 h until TLC showed that no starting material **27** remained. The solid catalyst was filtered off using a polyamide membrane (0.2 μm), and the product was recovered by evaporation of the filtrate resulting in compound **28** as oil that became solid (330 mg,

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

0.65 mmol, 97.0% yield) when dried under high vacuum. Compound **28**, NMR, δ_H: 1.16 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.20 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 3.82 (3H, s, A-OMe), 3.85 (3H, s, B-OMe), 4.08 (2H, q, J = 7.1 Hz, B-OCH$_2$Me), 4.13 (2H, q, J = 7.1 Hz, A-OCH$_2$Me), 4.16 (1H, d, J = 3.6, Bβ), 4.25 (1H, d, J = 3.6, Aβ), 4.80 (1H, d, J = 3.6 Hz, Bα), 4.94 (1H, d, J = 3.6 Hz, Aα), 6.56 (1H, d, J = 1.8 Hz, B6), 6.75 (1H, d, J = 8.2 Hz, A5), 6.87 (1H, d, J = 1.8 Hz, B2), 6.90 (1H, dd, J = 8.2, 1.8 Hz, A6), 7.19 (1H, d, J = 1.8 Hz, A2); δ_C: 14.41 (B-OCH$_2$Me), 14.46 (A-OCH$_2$Me), 56.23 (A-OMe), 56.50 (B-OMe), 61.34 (B-OCH$_2$Me), 61.42 (A-OCH$_2$Me), 75.16 (Aα), 75.23 (Bα), 76.47 (Bβ), 76.62 (Aβ), 106.71 (B2), 111.11 (B6), 112.50 (A2), 118.48 (A5), 119.69 (A6), 133.00 (B1), 137.98 (B4), 138.04 (A1), 144.79 (B5), 146.50 (A4), 149.19 (B3), 150.91 (A3), 172.98 (Bγ), 173.10 (Aγ).

Compound **29** was prepared via a typical β-ether synthetic method from compound **28** and compound **24**, as described. Compounds **28** (320 mg, 0.63 mmol) and **24** (180 mg, 0.63 mmol) were dissolved in DMF (10 mL) with K_2CO_3 (100 mg, 0.72 mmol). The reaction was monitored by TLC (CH_2Cl_2/MeOH, 20:1, v/v). After the reaction was completed, K_2CO_3 was filtered off, and DMF was evaporated under reduced pressure at 55 °C. The product was extracted with EtOAc (50 mL × 2) and water (50 mL × 2), and the combined organic phase was washed with distilled water and saturated NH_4Cl solution, dried over anhydrous $MgSO_4$, filtered, and concentrated under reduced pressure at 40 °C. Compound **29** was obtained as a light yellow oil (400 mg, 0.56 mmol, 89% yield). Compound **29**, NMR, δ_H: 1.18 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.20 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 2.26 (3H, s, C4-OAc), 3.75 (3H, s, A-OMe), 3.83 (3H, s, B-OMe), 3.85 (3H, s, C-OMe), 4.10 (2H, q, J = 7.1 Hz, B-OCH$_2$Me), 4.14 (2H, q, J = 7.1 Hz, A-OCH$_2$Me), 4.16/4.17 (1H, d, J = 3.6, Bβ), 4.25/4.26 (1H, d, J = 3.6, Aβ), 4.84/4.85 (1H, d, J = 3.7 Hz, Bα), 4.96/4.97 (1H, d, J = 3.7 Hz, Aα), 5.22 (2H, s, Cβ), 6.50 (1H, d, J = 1.7 Hz, B6), 6.89 (1H, d, J = 1.7 Hz, B2), 6.91 (1H, d, J = 8.2 Hz, A5), 6.97 (1H, dd, J = 8.2, 1.7 Hz, A6), 7.17 (1H, d, J = 8.2 Hz, C5), 7.20 (1H, d, J = 1.7 Hz, A2), 7.72 (1H, dd, J = 8.2, 1.7 Hz, C6), 7.74 (1H, d, J = 1.7 Hz, C2); δ_C: 14.44 (B-OCH$_2$Me), 14.48 (A-OCH$_2$Me), 20.45 (C4-OAc), 56.14 (A-OMe), 56.36 (B-OMe), 56.41 (C-OMe), 61.44 (A/B-OCH$_2$Me), 75.09 (Aα), 75.13 (Bα), 76.06 (Cβ), 76.40 (Bβ), 76.42 (Aβ), 106.45 (B2), 109.35 (B6), 112.57 (A2), 112.94 (C2), 119.79 (A6), 120.40 (A5), 122.50 (C6), 123.76 (C5), 134.87 (C1), 137.61 (B4), 138.65 (B1), 139.28 (A1), 144.87 (C4), 145.10 (A4), 151.28 (B5), 151.52 (A3), 152.34 (C3), 154.02 (B3), 168.64 (C4-OAc), 172.87 (Bγ), 173.02 (Aγ), 194.38 (Cα).

For the synthesis of compound **30**, compound **29** (340 mg, 0.48 mmol) and formaldehyde solution (37% [w/w], 38.9 mg, 0.48 mmol) were dissolved in 1,4-dioxane (10 mL), to which K_2CO_3 (4.0 equiv) was added.[27] The reaction mixture was sealed, kept stirring at room temperature overnight, and monitored by TLC. After the reaction was completed, K_2CO_3 was filtered off, and the solvent was evaporated under reduced pressure. The product was extracted with EtOAc (30 mL × 2) and water (30 mL × 2), and the combined organic phase was washed with distilled water and saturated NH_4Cl solution, dried over anhydrous $MgSO_4$, filtered, and concentrated under reduced pressure. All of the products (260 mg) were loaded onto 1 mm thick normal-phase silica-gel plates (about 60 mg/plate) and developed with methanol−dichloromethane (1:20) as eluting solvent. Each isolated product was characterized by NMR. Compound **30** was separated with an isolated yield of 30.0% as a pair of isomers; although compound **29** is a largely stereochemically pure compound (see above for the synthesis of **27**), the introduction of a new racemic optical center will lead to a pair of diastereomers, A($\alpha S/\beta R$)-B($\alpha S/\beta R$)-C(R/S). Compound **30**, NMR, δ_H: 1.17 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.21 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 2.25 (3H, s, C4-OAc), 3.75 (3H, s, B-OMe), 3.77 (3H, s, A-OMe), 3.80 (3H, s, C-OMe), 3.95 (2H, d, J = 7.3 Hz, Cγ), 4.10 (2H, q, J = 7.1 Hz, B-OCH$_2$Me), 4.14 (2H, q, J = 7.1 Hz, A-OCH$_2$Me), 4.16/4.17 (1H, d, J = 3.6, Bβ), 4.25/4.26 (1H, d, J = 3.6, Aβ), 4.84/4.85 (1H, d, J = 3.7 Hz, Bα), 4.97/4.98 (1H, d, J = 3.7 Hz, Aα), 5.36 (1H, t, J = 3.9 Hz, Cβ), 6.53 (1H, d, J = 1.7 Hz, B6), 6.84 (1H, dd, J = 8.2, 2.9 Hz, A5), 6.90 (1H, d, J = 1.7 Hz, B2), 6.97 (1H, d, J = 8.2 Hz, A6), 7.14 (1H, d, J = 8.2 Hz, C5), 7.22 (1H, d, J = 1.7 Hz, A2), 7.71 (1H, dd, J = 8.2, 1.8 Hz, C6), 7.74 (1H, d, J = 1.8 Hz, C2); δ_C: 14.45 (B-OCH$_2$Me), 14.49 (A-OCH$_2$Me), 20.46 (C4-OAc), 56.14 (A-OMe), 56.30 (B/C-OMe), 61.45 (B-OCH$_2$Me), 61.47 (A-OCH$_2$Me), 63.63 (Cγ), 75.06 (Aα), 75.11 (Bα), 76.38 (Bβ), 76.40 (Aβ), 87.28/87.36 (Cβ), 106.41/106.48 (B2), 109.39/109.42 (B6), 112.41/112.42 (A2), 113.39/113.41 (C2), 119.83/119.85 (A6), 119.93/119.97 (A5), 122.92/122.94 (C6), 123.54 (C5), 135.77 (C1), 137.23 (B4), 138.66/138.68 (B1), 139.37/139.40 (A1), 144.61 (C4), 144.77 (A4), 150.82 (B5), 151.20/151.25 (A3), 152.14 (C3), 153.60 (B3), 168.65 (C4-OAc), 172.81/172.84 (Bγ), 173.00/173.04 (Aγ), 194.38/196.40 (Cα).

For the reduction of compound **30**, two different reduction methods were applied in order to obtain different model compounds **13** and **14**. Compound **31** was obtained by the reduction of compound **30** using borane−*tert*-butylamine complex (powder, 97%) as the reducing agent. Compound **30** (50 mg, 0.067 mmol) was dissolved in CH_2Cl_2 (5 mL), to which borane−*tert*-butylamine complex (11.6 mg, 0.13 mmol) was added. The reaction mixture was kept stirring at room temperature, overnight. After the reaction was completed (monitored by TLC), the excess reducing agent was quenched by slowly adding 1 M aqueous HCl solution until the mixture turned clear. Then, the reaction mixture was evaporated to remove CH_2Cl_2. The resultant product was dissolved in EtOAc (25 mL) and hydrolyzed by shaking with aqueous HCl (1 M, 20 mL). The aqueous phase was further extracted with EtOAc (20 mL). The two EtOAc fractions were combined and mixed with 1 M aqueous HCl (20 mL) in a separatory funnel. The two phases in the funnel were well-mixed by shaking vigorously for 10 min. This operation was repeated to make sure (by TLC) that all borate intermediates were converted to product **31**. The resulting organic phase was washed with 0.4 M $NaHCO_3$ (20 mL × 2) and saturated NH_4Cl (20 mL) solution. The separated EtOAc layer was dried over $MgSO_4$, filtered, and evaporated at 40 °C under reduced pressure to result in product **31** (35 mg, 70.1% yield) as a mixture of four isomers, two from the *syn* (*threo*) β-ether and two from the *anti* (*erythro*) form, not all of which are fully resolved in the NMR spectra. Compound **31**, NMR, δ_H: 1.17 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.21 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 2.20/2.21 (3H, s, C4-OAc), 3.35/3.77 (1H, Cγ1), 3.53/3.89 (1H, Cγ2), 3.71/3.74 (3H, s, C-OMe), 3.80 (3H, s, A-OMe), 3.87/3.89 (3H, s, B-OMe), 4.09 (2H, q, J = 7.1 Hz, B-OCH$_2$Me), 4.16 (2H, q, J = 7.1 Hz, A-OCH$_2$Me), 4.16/4.17 (1H, d, J = 3.6, Bβ), 4.27/4.28 (1H, d, J = 3.6, Aβ), 4.20/4.38 (1H, m, Cβ), 4.84/4.85 (1H, d, J = 3.7 Hz, Bα), 4.97/4.98 (1H, d, J = 3.7 Hz, Aα), 5.13 (1H, m, Cα), 6.51/6.53 (1H, d, J = 1.7 Hz, B6), 6.90/6.91 (1H, d, J = 1.7 Hz, B2), 6.93 (1H, A5), 6.94 (1H, C5), 7.03 (1H, C6), 6.99 (1H, A6), 7.14/7.21 (1H, C2), 7.26 (1H, d, J = 1.7 Hz, A2); δ_C: 14.44 (B-OCH$_2$Me), 14.49 (A-OCH$_2$Me), 20.47 (C4-OAc), 56.02/56.04 (C-OMe), 56.22/56.24 (A-OMe), 56.57 (B-OMe), 63.81/61.43 (Cγ), 61.45/61.48 (A/B-OCH$_2$Me), 73.33/73.72 (Cα), 75.07 (Aα), 75.12 (Bα), 76.33 (Bβ), 76.37/76.41 (Aβ), 87.66/87.68/89.19 (Cβ), 106.42/106.47 (B2), 109.19/109.21/109.26 (B6), 111.52/112.09 (C2), 112.52/112.53 (A2), 119.45/119.90 (C6), 119.92 (A6), 120.14/120.19/120.22 (A5), 122.86 (C5), 136.75/137.24 (B4), 138.39/138.60 (B1), 139.54/139.59 (A1), 139.61/139.68/139.93 (C4), 141.27/141.65 (C1), 144.49/144.52/144.54 (A4), 151.26/151.28 (B5), 151.35/151.37/151.39 (A3), 151.71/151.77 (C3), 153.88/154.12 (B3), 169.02 (C4-OAc), 172.84/172.86 (Bγ), 173.03 (Aγ).

Compound **13** was obtained quantitatively via acetylation of compound **31** with acetic anhydride−pyridine (1 mL, 1:1, v/v), followed by the usual workup, again as a mixture of four isomers. Compound **13**, NMR, δ_H: 1.16 (3H, t, J = 7.1 Hz, B-OCH$_2$Me), 1.21 (3H, t, J = 7.1 Hz, A-OCH$_2$Me), 1.77/1.78/1.85/1.87 (3H, s, Cγ-OAc), 1.87/1.90 (3H, s, Cα-OAc), 1.97 (3H, s, Aβ-OAc), 1.98 (3H, s, Bα-OAc), 2.06 (3H, s, Bβ-OAc), 2.10 (3H, s, Aα-OAc), 2.21 (3H, s, C4-OAc), 3.70/3.76 (3H, s, C-OMe), 3.80/3.84 (3H, s, A-OMe), 3.85 (3H, s, B-OMe), 3.89/4.20 (1H, Cγ1), 4.14 (4H, A/B-OCH$_2$Me), 4.19/4.38 (1H, Cγ2), 4.91/4.97 (1H, m, Cβ), 5.31 (1H, Bβ), 5.45 (1H, Aβ), 6.09 (1H, Cα), 6.13 (1H, d, J = 3.7 Hz, Bα), 6.32 (1H, d, J = 3.7 Hz, Aα), 6.32/6.38 (1H, d, J = 1.7 Hz, B6), 6.88/6.91 (1H, d, J = 1.7 Hz, B2), 6.94−6.99 (3H, A5/C5/C6), 7.05 (1H, A6), 7.14/7.17

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

(1H, C2), 7.28 (1H, d, J = 1.7 Hz, A2); δ_C: 14.33 (B-OCH$_2$Me), 14.39 (A-OCH$_2$Me), 20.31/34 (Bβ-OAc), 20.44/20.47 (C4-OAc), 20.51 (Aβ/Bα-OAc), 20.53 (Cγ-OAc), 20.66 (Aα-OAc), 20.84/20.85/20.90 (Cα-OAc), 56.09/56.16 (C-OMe), 56.29/56.33/56.35 (A-OMe), 56.50 (B-OMe), 62.13/62.14 (A/B-OCH$_2$Me), 63.30/63.96/64.01 (Cγ), 73.98/74.02 (Bα), 74.07 (Aα), 74.47/74.52/74.56 (Bβ), 74.78/74.81 (Aβ), 74.97/75.13/76.03/76.11 (Cα), 81.32/81.42/81.49/81.53 (Cβ), 106.23/106.29/106.43/106.51 (B2), 108.48/108.63/109.07/109.16 (B6), 112.06/112.11/112.45/112.47 (C2), 112.60/112.63 (A2), 119.99/120.11/120.12/120.14 (A5), 120.17/120.33 (A6), 120.85/120.96/121.30/121.54 (C6), 123.25/123.29/123.43 (C5), 132.54/132.57/132.65/132.74 (B1), 133.79/133.84/134.05/134.17 (A1), 136.74/136.71/136.97 (C1), 137.10/137.19/138.27 (B4), 140.49/140.66 (C4), 145.15/145.27/145.63/145.70 (A4), 151.24/151.32/151.63/151.65 (B5), 151.89/151.93 (A3), 152.00/152.02/152.04 (C3), 154.35/154.38 (B3), 167.29 (Bγ), 167.46 (Aγ), 168.87/168.92 (C4-OAc), 169.57/169.60 (Bα-OAc), 169.78/169.79 (Aα-OAc), 169.91/169.94/169.95 (Cα-OAc), 170.01/170.03/170.05 (Bβ-OAc), 170.22 (Aβ-OAc), 170.59/170.65 (Cγ-OAc). HR-MS (ESI) Calcd for $C_{48}H_{60}NO_{23}$ [(M + NH_4)$^+$], 1018.3551; found, 1018.3550.

For the synthesis of compound **14**, sodium borohydride was used as the reducing agent. Compound **30** (45 mg, 0.06 mmol) was reduced by stirring with sodium borohydride (27.4 mg, 0.72 mmol) in ethanol (5 mL, 95%) at room temperature. The reaction mixture was kept stirring overnight to ensure the completeness of the reduction. The reaction was quenched by slowly adding 1 M aqueous HCl solution to deplete the excess reducing agent until the mixture turned clear and H_2 evolution ceased. The mixture was stirred at room temperature for 30 min to ensure the cleavage of the borate intermediates. Then, the reaction mixture was evaporated to remove ethanol. The resultant product dissolved in water (30 mL) was slowly neutralized by adding NaOH (1 M) with stirring until a pH value of ∼5.0. The aqueous phase was freeze-dried under high vacuum (about 100 mT). The dried crude product was then acetylated with acetic anhydride-pyridine (10 mL, 1:1, v/v) at room temperature, overnight. Following the normal workup, the acetylated mixture was then extracted twice with EtOAc (30 mL × 2) and water (30 mL × 2), and the combined organic phase was washed with distilled water and saturated NH_4Cl solution, dried over anhydrous $MgSO_4$, filtered, and concentrated under reduced pressure. All of the products (60 mg) were loaded onto 1 mm thick normal-phase silica-gel plates (about 60 mg/plate) and developed with CH_2Cl_2/MeOH (20:1, v/v). Each isolated product was characterized by NMR. The final model compound **14** was separated with an isolated yield of 63.3%, again as a mixture of four isomers. Compound **14**, NMR, δ_H: 1.77/1.78/1.86/1.87/1.90/1.93/1.94/1.95/2.00/2.02/2.08/2.10/2.11 (24H, Aα/Aβ/Aγ/Bα/Bβ/Bγ/Cα/Cγ-OAc), 2.21 (3H, s, C4-OAc), 3.70/3.72/3.76/3.79/3.80 (6H, A/C-OMe), 3.83/3.84 (3H, s, B-OMe), 3.85/4.18 (2H, Bγ), 3.89/4.20 (2H, Cγ), 3.97/4.26 (2H, Aγ), 4.91/4.96 (1H, m, Cβ), 5.30 (1H, Bβ), 5.45 (1H, Aβ), 5.80 (1H, d, J = 3.7 Hz, Bα), 6.02 (1H, d, J = 3.7 Hz, Aα), 6.08 (1H, Cα), 6.32/6.38 (1H, d, J = 1.7 Hz, B6), 6.82/6.85 (1H, d, J = 1.7 Hz, B2), 6.94−6.99 (3H, A5/C5/C6), 7.02 (1H, A6), 7.14/7.17 (1H, C2), 7.22 (1H, d, J = 1.7 Hz, A2); δ_C: 20.44/20.47 (C4-OAc), 20.53/20.56/20.66/20.68/20.71/20.73/20.81/20.84/20.86/20.91 (Aα/Aβ/Aγ/Bα/Bβ/Bγ/Cα/Cγ-OAc), 56.09/56.11/56.17/56.24/56.27/56.28 (A/C-OMe), 56.46/56.48/56.49 (B-OMe), 62.73/62.76 (Bγ), 62.87 (Aγ), 64.00/64.06 (Cγ), 72.82/72.86 (Bβ), 73.04 (Aβ), 74.00/74.05 (Bα), 74.07/74.11 (Aα), 74.96/75.09/76.07/76.12 (Cα), 81.45/81.50 (Cβ), 106.05/106.09/106.26/106.36 (B2), 108.85/108.91/109.45/109.52 (B6), 112.12/112.13/112.45 (C2), 112.67/112.69/112.71 (A2), 119.98/120.06/120.09/120.12 (A5), 120.43/120.49/120.55 (A6), 120.95/121.18/121.44/121.57 (C6), 123.03/123.24/123.27/123.42 (C5), 133.11/133.16/133.25/133.33 (B1), 134.40/134.70/134.77 (A1), 136.79/136.99 (C1), 137.20/137.29/138.36/138.39 (B4), 140.47/140.48/140.65 (C4), 145.16/145.23/145.43/145.67/145.69 (A4), 151.22 (B5), 151.58/151.60/151.92/151.97/152.00 (A3/C3), 154.57/154.68 (B3), 168.87 (C4-OAc), 168.91/168.93/169.86/169.88/169.92/170.01/170.11/170.21/170.58/170.60/170.64/170.66 (Aα/Aβ/Aγ/Bα/Bβ/Bγ/Cα/Cγ-OAc). HR-MS (ESI) Calcd for $C_{48}H_{60}NO_{23}$ [(M + NH_4)$^+$], 1018.3551; found, 1018.3553.

Lignin HSQC Spectra. Cellulolytic enzyme lignin (CEL) preparations were isolated from ball-milled white spruce (*Picea glauca*) and loblolly pine (*Pinus taeda*) according to the standard procedure.[28] The yields of crude CELs were 55% (pine) and 58% (spruce) of the original Klason lignin. Acetylation was performed in a 1:1 mixture of acetic anhydride and pyridine as above. 2D HSQC NMR spectra were acquired on the spectrometer described above using 80 mg of acetylated CEL in 0.5 mL acetone-d_6 with the central solvent peaks (δ_H/δ_C 2.04/29.80) used as internal references.

RESULTS AND DISCUSSION

Much of the information on lignin structure and the distribution of the various interunit linkage types comes from degradative methods of analysis. Thioacidolysis is designed to selectively cleave the predominant β−O−4-linkages in lignin so that the released monomers and, following desulfurization by Raney nickel, dimers can be analyzed by GC or GC-MS.[29] The DFRC method similarly cleaves lignin and the monomers and dimers can be analyzed by the same methods.[16] The detection of the 4−O−5-linked dimer **4** from thioacidolysis/Raney-Ni, as well as the isolation of 4−O−5-linked DFRC dimer **6**, imply that some of the 4−O−5-linked units in lignin must be integrated into softwood lignins with (cleavable) β−O−4-linkages on their side chains and be either free-phenolic units or be 4−O-etherified to the β-position of the next unit. In the DFRC method, for example, each double bond in compound **6** (Figure 2) results only in the case that a β-ether is cleaved, so such a dimer can only arise from structures in which both moieties are β-ethers. Compounds **7** and **8** would therefore be more ideal lignin models for such structures in lignin polymers. Trimeric products consisting of 4−O−5/β−5 or 4−O−5/5−5 linked units have been previously isolated and identified.[30]

NMR methods, particularly the 2D ^{1}H−^{13}C correlation methods, such as HSQC, are particularly valuable for identifying structures in the lignin polymer, even without the need to isolate the lignin, that is, on the whole cell wall fraction of plant materials.[21,31−34] NMR has been responsible, for example, for the discovery that 5−5-linked units in lignins are largely in the form of 8-membered ring heterocycles, the dibenzodioxocins,[21,35] and for the revelation that β−1 units are actually present in lignins as spirodienones (explaining a long-held conundrum that only one of the phenolics in the classical β−1 structure was ever etherified in the polymer; it is not a phenol in the lignin, so can not undergo radical coupling).[12,13,21] It was, therefore, also a conundrum that 4−O−5-units were reportedly a significant fraction of the interunit linkages in softwood lignins, noted at up to 3.5−4% from various acidolytic methods,[36] but could not be observed in NMR spectra. For softwood lignins that contain no syringyl units, we reasoned that the 4−O−5-linked units should be unique in terms of the 5−O-substituent so that their C2−H2 and C6−H6 correlations in an HSQC NMR spectrum would be well dispersed from those of other guaiacyl units and nearer where syringyl units resonate, and therefore could be readily identified in the otherwise empty aromatic region of a softwood HSQC spectrum.[37] This contention, that 4−O−5-units in lignins should have diagnostic contours that are, in principle, readily identified, contrasts with prior literature comments that they are invisible to 2D ^{1}H−^{13}C correlation experiments because the carbons involved in the bonding are not protonated; that statement is true but it misses the point that adjacent C/H pairs are diagnostic for this structural unit. However, our previous attempt to identify 4−O−5-units in a loblolly pine lignin by comparing HSQC NMR with a series of

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

simple 4−O−5-models was not successful.[22] We concluded that any 4−O−5-units should have been readily detectable if present at a level of ∼3% with the HMQC/HSQC NMR methods of the time. Our failure to find C−H correlations consistent with those of a series of 4−O−5 model compounds was probably because either the abundance of 4−O−5 structures in that loblolly pine lignin preparation was too low (and was, therefore, overquantified by extrapolating degradative analysis data) or the model compounds used did not sufficiently closely model the real 4−O−5 structures in lignins, and we were simply not looking in the right regions of the spectra.

Given that 4−O−5-coupled units are evident in softwoods from degradation methods but not from NMR (at the time), it seemed imperative to resolve this dilemma. With these two points in mind, this study pursued two directions. First, a new route was developed to prepare more sophisticated and realistic 4−O−5-linked lignin models. Recently, our interest in the 5-linked pinoresinol structures of softwood lignins led us to synthesize various novel 5-linked lignin model compounds including 4−O−5-linked models.[23] Following appropriate derivatization, 4−O−5-linked models **9**−**11** (Figure 2), suitable for characterizing acetylated lignins were produced. By using a similar biomimetic free-radical coupling approach, compounds **7**, **8**, and **12** were also synthesized from the corresponding dimers. As demonstrated previously, the functionalities or substituents on the side-chain of a 4−O−5-linked dimer or trimer can profoundly affect chemical shifts of C2 or H2 on the aromatic rings. As listed in Table 1, the newly synthesized model compounds **7**−**14** have similar NMR characteristics regarding the chemical shifts for C2−H2 and C6−H6 of the 4−O−5-linked unit, but are quite different from those in the simpler models **1**−**6** (Figure 2).

Table 1. NMR Data for C2−H2 and C6−H6 Correlations for 4−O−5-Linked Model Compounds 1−14 (Figure 2)

model	C2−H2	C6−H6
1	109.1/7.48	113.8/7.23
2	108.2/7.48	112.0/7.15
3	109.2/6.58	111.1/6.17
4	107.5/6.63	110.2/6.18
5	105.0/6.77	107.9/6.32
6	105.5/6.96	109.1/6.48
7	107.6/6.92	110.8/6.51, 6.52
8	106.6/6.97	109.7/6.51, 6.52
9	106.1/6.83	109.3/6.48
10	105.0/6.86	108.1/6.48
11	104.5/6.90	107.6/6.46
12	104.8/6.85	107.9/6.46
13	106.3/6.88	108.6/6.32
	106.5/6.91	109.1/6.38
14	106.05, 106.09/6.82	108.85, 108.91/6.32
	106.26, 106.36/6.85	109.45, 109.52/6.38

Second, two CELs from white spruce and loblolly pine wood were isolated and used for NMR analysis. We chose the CEL fraction over the Milled Wood Lignin (MWL) from loblolly pine that was used in our previous study in order to have a better chance of detecting 4−O−5 structures in lignin−CEL preparations have higher yields (and are, therefore, more representative) than the traditional MWLs that, for softwoods, are typically obtained in ∼15% yield and therefore do not necessarily represent the lignin well. In Figure 5, panels B and C show the short-range C−H correlations (HSQC) in the aromatic region of the acetylated CELs run in acetone-d_6. These two spectra are extremely similar. In the region of δ_H/δ_C = 6.4−7.1/104−111 ppm, two distinguishable C−H correlation clusters were clearly evident. When the chemical shift values of compounds **1**−**6** from Table 1 are superimposed on the spectrum (Figure 5B), however, none of these 4−O−5 model correlations appears to fit with the spectrum, confirming our previous contention that the 4−O−5 units in lignin have little in common with these models **1**−**6** previously studied. This is not surprising given that compounds **1**−**6** either contain benzylic keto-groups on both side chains (compounds **1** and **2**) or have completely reduced (alkyl) side chains (compounds **3**−**5**), and compound **6** is a 4−O−5-linked coniferyl alcohol dimer that simply can not be present in lignin−coniferyl alcohol coupling products invariably result from the coupling of at least one of them at its β-position; 5−5- and 4−O−5-coupled units in lignin therefore must only occur from the radical coupling of preformed dimers or oligomers, as has been pointed out in various reviews.[1,2,11] Obviously better models than compounds **1**−**6** were needed to determine to which structures the observed lignin correlations belong. When HSQC spectra of model compounds **7**−**14** are overlaid onto the spectrum of spruce lignin (Figure 5A), it is clear that the C2−H2 and C6−H6 correlations from compounds **8**, **10**, **11**, and **12** match well, suggesting that those C−H correlations (δ_H/δ_C, 6.4−6.6/107−111 ppm and δ_H/δ_C, 6.7−7.1/104−107 ppm) in the spruce lignin spectrum belong to C6−H6 and C2−H2 correlations of 4−O−5-linked structures, respectively. It is also clear that the C6−H6 or C2−H2 correlations for compounds **9**−**12** are close to each other despite the different substitutions on their side chains. However, the correlations for model **8** are different enough from those of models **9**−**12** that the 4−O−5-linked β−O−4 structures in lignin can be distinguished from the 4−O−5-linked β−5 or β−β structures by HSQC NMR. Although this represents a significant improvement (over the data shown for models **1**−**6** in Figure 5B), it remains apparent that we still do not have the required arsenal of models to fully assign all of the intricate features apparent in the correlation maps here for C2−H2 or for C6−H6 in Figure 5A; we do not wish to make unsupported assignments here, but contend that, at some point, this will be possible.

One of the most intriguing observations comes from comparing data from models in which the 5-linked moiety's phenol is 4−O-etherified or not. Comparing data from model **7** versus **8** shows that small differences in chemical shift occur when the phenolic−OH is etherified and these differences are significant enough to put the correlations from **7**, the etherified model, in clear areas of the spectrum. The same result was obtained when comparing data from model **9** versus **10**, at least for the C2−H2 correlation. Because of the important ramifications of this observation that evidence for etherified 4−O−5-units appears to be lacking, i.e., that such 4−O−5 units might not be true lignin branch points (see below), two more 4−O−5-linked models, **13** and **14** (Figure 2), in which the phenolic−OH is etherified via a 4−O−β-linkage rather than a simple methyl group as in models **7** and **9**, were synthesized to verify this deduction. As the results show (Figure 5), the C2−H2 correlations (chemical shift) of model **13** matched those of the lignin spectrum at δ_C/δ_H 105.8−106.5/6.86−6.92 ppm, whereas the C2−H2 correlations of model **14** deviated slightly from those in the lignin spectrum. However, the C6−H6

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

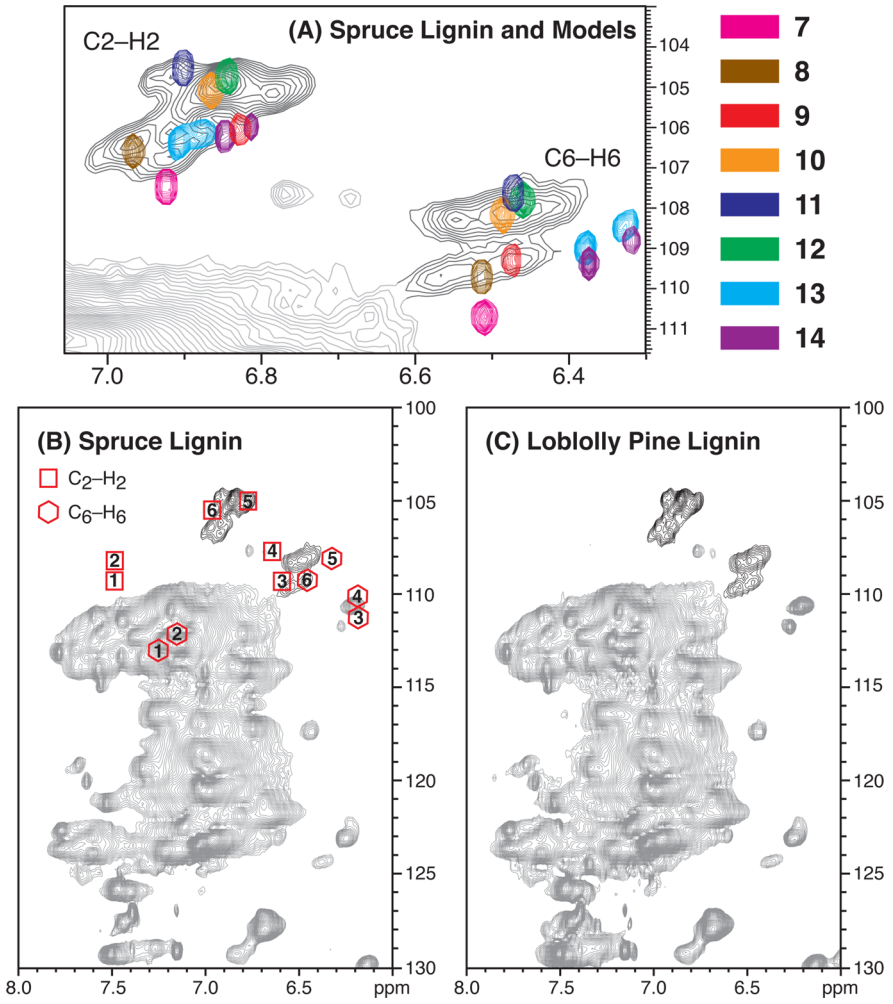


Figure 5. Partial HSQC NMR (aromatic region) of softwood lignins: (A) spruce lignin overlaid with model compound data showing the diagnostic C2−H2 and C6−H6 correlations; (B) spruce lignin; (C) loblolly pine lignin.

correlations from both **13** and **14**, located at δ_C/δ_H 108.2−109.7/6.32−6.38 ppm, did not match any correlation peaks in the lignin spectrum (Figure 5A). These models are not quite perfect, but the data also appear to suggest that essentially all the 4−O−5 units in spruce lignin, at least those that join β−O−4- or β−β-linked units, are present in their free-phenolic forms; that is, that once the 4−O−5 coupling reaction occurs (between two dimers or higher oligomers), no detectable further 4−O-coupling (of the 5−O-linked moiety) to the β-position of a monolignol (Figure 6) occurs during further lignification in planta. Such a contention is crucial to an understanding of whether 4−O−5-coupling of lignin oligomers results solely in the "U-type" conjoining of two chains or is a real "Y-type" branch-point (Figure 6B). The data here rather strongly support the contention, from elegant DFRC, followed by phosphitylation and ^{31}P NMR experiments,[38] that 4−O−5-linked structures that are formed during lignification are not, within the limits of detection here, etherified and are therefore not further polymerized via chain extension by the addition of new monomers. The significance is, if further supporting evidence remains consistent with these now independent sets of observations, that 4−O−5-units are not an indication of real branching in the lignin polymer! That same contention is being demonstrated for dibenzodioxocins, derived from 5−5-coupling of two oligomers followed by the addition of one more monomer (to create the dibenzodioxocin),[35] with the evidence so far being that the chain is not further extended (Figure 6A). The dogma that softwood lignins are "highly branched" is apparently not true, and lignins (even softwood lignins) are far more linear than often conceptualized.

Finally, we address the apparent dilemma regarding 4−O−5-levels in softwoods. If such units were at the levels noted in the literature, they would be much easier to detect by NMR. Although quantitation via HSQC is rather poor, and particularly over-represents end groups with higher mobility, the quantification of aromatic residues that are primarily in the chain appears to be reasonable.[33] Using simple integration of the NMR contours in the softwood lignin HSQC spectra, we estimate that 4−O−5-linked structures represent ∼1% of the total guaiacyl units; that is, as anticipated from the initial study, the levels are indeed very low. We therefore suspect that extrapolations of levels made from the various degradative methods are uniformly too high for 4−O−5 units. This is analogous to the way that β−1 units are now presumed to be too high, 7% according to Adler's wonderful but now outdated review,[36] from their rather remarkable levels in the recognizable dimers from acidolytic methods; β−1 dimers are a significant fraction of the thioacidolysis dimers, for example;[19] they are now considered to be <3−4% from NMR methods.[13] What we can conclude is that 4−O−5 units have now finally been rather

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920

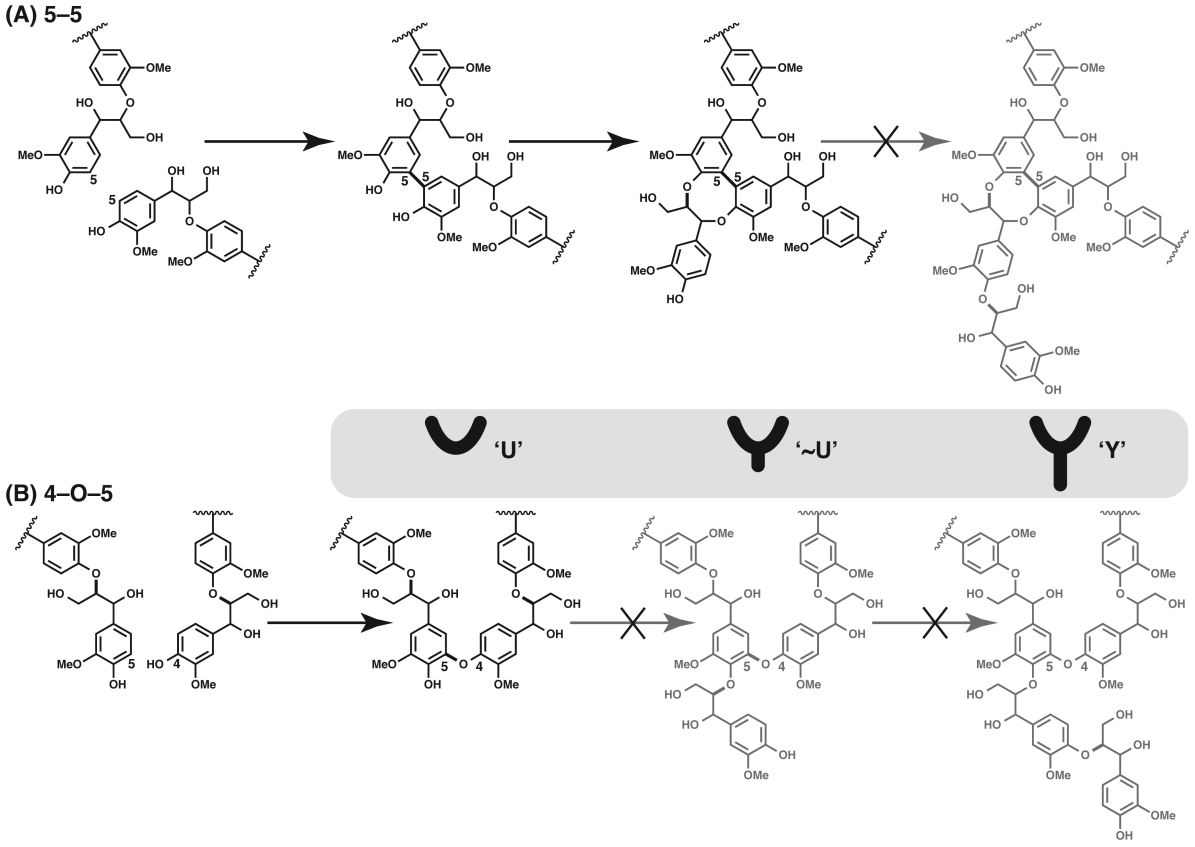


Figure 6. Do 5–5 (dibenzodioxocin) and 4–O–5 units in lignins allow for the formation of "Y-type" or only "U-type" branches (gray box) in lignin? The inability to detect etherification of the free-phenolic end-group in either case suggests that polymerization does not continue from these units and that they are just "U-type" branches. [The addition of coniferyl alcohol to the initially formed 5–5-coupled unit appears to form a short "Y-type" branch, but as it does not appear to polymerize further, it is still almost a "U-type" structure ("~U")]. In both cases, as the grayed structures do not appear to form (i.e., with polymerization terminating with the last black structure), these units, traditionally thought of as being evidence of branch points, represent little more than chain connectors. Lignin chains are therefore best described as essentially linear, even in softwoods. To date, there is no compelling evidence in softwood lignins for true "Y-type" branching.

compellingly demonstrated by NMR methods and so are indeed to be considered authentic components of the lignin polymer, but they have almost certainly been overestimated in the past.

CONCLUSIONS

In this study, the anticipated 4–O–5-units in softwood lignins have been quite compellingly identified by their diagnostic C–H correlations in the aromatic region of high-sensitivity HSQC NMR spectra. The levels of such units appear to be ~1% of the lignin, significantly lower than estimates obtained from extrapolating acidolytic degradative data. Authentication has required synthesized trimeric and tetrameric lignin model compounds that include the 4–O–5-linked moieties having various complete side chain functionalities that more closely resemble their counterpart structures in softwood lignins. Analysis of the various models has also demonstrated that the 4–O–5-linked β–O–4 dimeric units in the lignin polymer can be differentiated from the 4–O–5-linked β–5 or β–β structures by HSQC NMR. Most importantly for its implications to the understanding of the existence of true "Y-type" branching of lignins (Figure 6), etherified 4–O–5-linked units can not be detected, suggesting that most or all of the 4–O–5-linked structures in lignins may be present as end units. If this contention holds up, it means that 4–O–5-units are not real "Y-type" branch-points in the polymer. Along with recent similar disclosures that dibenzodioxocin units, formed by addition of a single monolignol to a 5–5-coupled unit, are not further etherified either, the two traditional structures (5–5 and, now, 4–O–5) long-thought to represent lignin polymer branch-points may not, in fact, be so; to date, evidence for their further coupling with additional monolignols during lignification has not been found. Consequently, the notion that even softwood lignins are "highly branched" is increasingly being mollified. Lignins may therefore be better described as being essentially linear, a profound departure from conventional wisdom that is gaining support.

AUTHOR INFORMATION

Corresponding Author

*E-mail: fachuanglu@wisc.edu. Tel.: +86 2087113953. Fax: (608) 890-2043.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank the China Scholarship Council, State Education Department, for supporting F.Y. as a visiting student in the Department of Biochemistry, the University of Wisconsin, and at the Great Lakes Bioenergy Research Center (GLBRC). This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909–1920

Science, DE-FC02-07ER64494) and in part by Stanford University's Global Climate and Energy Project.

REFERENCES

(1) Ralph, J.; Brunow, G.; Harris, P. J.; Dixon, R. A.; Schatz, P. F.; Boerjan, W. In *Recent Advances in Polyphenol Research*; Daayf, F., El Hadrami, A., Adam, L., Ballance, G. M., Eds.; Wiley-Blackwell Publishing: Oxford, U.K., 2008; Vol. *1*, pp 36−66.
(2) Ralph, J.; Lundquist, K.; Brunow, G.; Lu, F.; Kim, H.; Schatz, P. F.; Marita, J. M.; Hatfield, R. D.; Ralph, S. A.; Christensen, J. H.; Boerjan, W. *Phytochem. Rev.* **2004**, *3*, 29−60.
(3) Whetten, R.; Sederoff, R. *Plant Cell* **1995**, *7*, 1001−1013.
(4) Sarkanen, K. V.; Ludwig, C. H. *Lignins: Occurrence, Formation, Structure and Reactions*; Wiley-Interscience: New York, 1971.
(5) Lepifre, S.; Froment, M.; Cazaux, F.; Houot, S.; Lourdin, D.; Coqueret, X.; Lapierre, C.; Baumberger, S. *Biomacromolecules* **2004**, *5*, 1678−1686.
(6) Hambardzumyan, A.; Foulon, L.; Chabbert, B.; Aguie-Beghin, V. *Biomacromolecules* **2012**, *13*, 4081−8.
(7) Fu, C.; Mielenz, J. R.; Xiao, X.; Ge, Y.; Hamilton, C. Y.; Rodriguez, M., Jr.; Chen, F.; Foston, M.; Ragauskas, A.; Bouton, J.; Dixon, R. A.; Wang, Z.-Y. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3803−3808.
(8) Rinaldi, R.; Jastrzebshi, R.; Clough, M. T.; Ralph, J.; Kennema, M.; Bruijnincx, P. C. A.; Weckhuysen, B. M. *Angew. Chem., Int. Ed.* **2016**, n/a.
(9) Timmel, T. E. *Compression Wood in Gymnosperms*; Springer: Heidelberg, 1986.
(10) Nanayakkara, B.; Manley-Harris, M.; Suckling, I. D.; Donaldson, L. A. *Holzforschung* **2009**, *63*, 431−439.
(11) Boerjan, W.; Ralph, J.; Baucher, M. *Annu. Rev. Plant Biol.* **2003**, *54*, 519−546.
(12) Zhang, L.; Gellerstedt, G. *Chem. Commun.* **2001**, 2744−2745.
(13) Zhang, L.; Gellerstedt, G.; Ralph, J.; Lu, F. *J. Wood Chem. Technol.* **2006**, *26*, 65−79.
(14) Brunow, G. *Methods to Reveal the Structure of Lignin*; Wiley-VHC: Weinheim, Germany, 2001.
(15) Lapierre, C. In *Forage Cell Wall Structure and Digestibility*; Jung, H. G., Buxton, D. R., Hatfield, R. D., Ralph, J., Eds.; American Society of Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI, 1993; pp 133−166.
(16) Lu, F.; Ralph, J. *J. Agric. Food Chem.* **1997**, *45*, 2590−2592.
(17) Rolando, C.; Monties, B.; Lapierre, C. In *Methods in Lignin Chemistry*; Dence, C. W., Lin, S. Y., Eds.; Springer-Verlag: Berlin; Heidelberg, 1992; pp 334−349.
(18) Peng, J.; Lu, F.; Ralph, J. *J. Agric. Food Chem.* **1998**, *46*, 553−560.
(19) Lapierre, C.; Pollet, B.; Monties, B.; Rolando, C. *Holzforschung* **1991**, *45*, 61−68.
(20) Parkas, J.; Brunow, G.; Lundquist, K. *Bioresources* **2007**, *2*, 169−178.
(21) Ralph, J.; Landucci, L. L. In *Lignin and Lignans; Advances in Chemistry*; Heitner, C., Dimmel, D. R., Schmidt, J. A., Eds.; CRC Press (Taylor & Francis Group): Boca Raton, FL, 2010; pp 137−234.
(22) Ralph, S. A. *Thirteenth International Symposium on Wood, Fiber, and Pulping Chemistry*, Auckland, New Zealand, May 16−19, 2005; Appita: Australia, 2005; pp 315−319.
(23) Yue, F.; Lu, F.; Sun, R.; Ralph, J. *Chem. - Eur. J.* **2012**, *18*, 16402−16410.
(24) Lan, W.; Lu, F.; Regner, M.; Zhu, Y.; Rencoret, J.; Ralph, S. A.; Zakai, U. I.; Morreel, K.; Boerjan, W.; Ralph, J. *Plant Physiol.* **2015**, *167*, 1284−1295.
(25) Ralph, J.; Quideau, S.; Grabber, J. H.; Hatfield, R. D. *J. Chem. Soc., Perkin Trans. 1* **1994**, 3485−3498.
(26) Sharpless, K. B.; Amberg, W.; Bennani, Y. L.; Crispino, G. A.; Hartung, J.; Jeong, K.-S.; Kwong, H.-L.; Morikawa, K.; Wang, Z.-M.; Xu, D.; Zhang, X. L. *J. Org. Chem.* **1992**, *57*, 2768−2771.
(27) Helm, R. F.; Ralph, J. *J. Agric. Food Chem.* **1992**, *40*, 2167−2175.
(28) Chang, H.-M.; Cowling, E. B.; Brown, W.; Adler, E.; Miksche, G. *Holzforschung* **1975**, *29*, 153−9.
(29) Lapierre, C.; Monties, B.; Rolando, C. *Holzforschung* **1986**, *40*, 113−118.
(30) Onnerud, H. *Holzforschung* **2003**, *57*, 377−384.
(31) Lu, F.; Ralph, J. *J. Biobased Mater. Bioenergy* **2011**, *5*, 169−180.
(32) Lu, F.; Ralph, J. *Plant J.* **2003**, *35*, 535−544.
(33) Mansfield, S. D.; Kim, H.; Lu, F.; Ralph, J. *Nat. Protoc.* **2012**, *7*, 1579−1589.
(34) Kim, H.; Ralph, J. *Org. Biomol. Chem.* **2010**, *8*, 576−591.
(35) Karhunen, P.; Rummakko, P.; Sipila, J.; Brunow, G.; Kilpelainen, I. *Tetrahedron Lett.* **1995**, *36*, 169−170.
(36) Adler, E. *Wood Sci. Technol.* **1977**, *11*, 169−218.
(37) Smit, R.; Suckling, I. D.; Ede, R. M. Oral Presentations. *Proceedings of the ACM Workshop on Role Based Access Control*, Fairfax, VA, Nov 6−7, 1997, ACS: New York, 1997; Vol *1*, pp L4−1.
(38) Crestini, C.; Melone, F.; Sette, M.; Saladino, R. *Biomacromolecules* **2011**, *12*, 3928−3935.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on May 4, 2016, with errors in the TOC/abstract graphic and Figure 4. The corrected version was reposted on May 27, 2016.

DOI: 10.1021/acs.biomac.6b00256
Biomacromolecules 2016, 17, 1909−1920