Supplementary Information for

Profiling of the Formation of Lignin-derived Monomers and Dimers from *Eucalyptus* Alkali Lignin

Chengke Zhao,[ab] Zhenhua Hu,[a] Lanlan Shi,[a] Chen Wang,[a] Fengxia Yue,*[a] Suxiang Li,[a] Han Zhang,[a] and Fachuang Lu*[a]

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

[b] Flexible Printed Electronic Technology Center, School of Materials Science and Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

*Corresponding authors:

Fengxia Yue (yuefx@scut.edu.cn). State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou 510640, China.

Fachuang Lu (fefclv@scut.edu.cn). State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou 510640, China.

1 Experimental section

1.1 Chemicals and materials

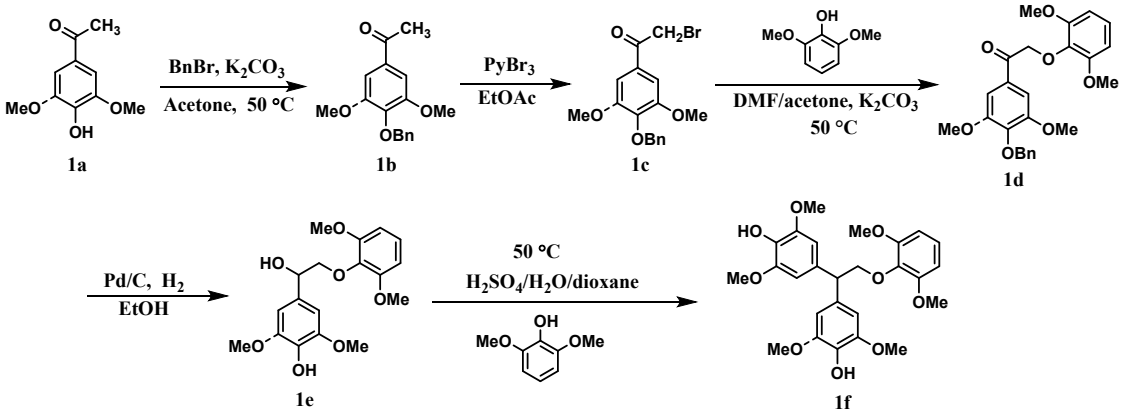
Chemicals including sodium hydroxide (97%), sodium acetate (99%), ammonium chloride (99.5%), anhydrous magnesium sulfate (98%), absolute ethanol (99.5%), dichloromethane (DCM, 99.5%), ethyl acetate (EtOAc, 99%), n-hexane (97%), acetic acid (99%), acetone (99.9%), dioxane (>99%), acetic anhydride (99%), sulfuric acid (98%), pyridine (99%), anhydrous potassium carbonate (99%), sodium borohydride (98%), sodium dihydrogen phosphate (99%), diisobutyl aluminum hydride ($(i\text{-}Bu)_2HAl$, 1.5 mol/L in toluene), acetylchloride (99%), tetrahydrofuran (≥99.9%), horseradish peroxidase (>300 units/mg), N,N-Dimethylformamide (DMF, 99.5%), benzyl bromide (>98%), N,O-bis(trimethylsilyl)-trifluoroacetamide (BSTFA, 98%), pyridinium bromide perbromide ($PyBr_3$, 90%), acetovanillone (>98%), acetosyringone (98%), 2,6-dimethoxyphenol (98%), sinapic acid (98%), Pd/C (10% Pd), hydrogen peroxide solution (30 wt.% in H_2O), Celluclast 1.5 L (a cellulase with enzymatic activity value of 700 EGU/g, Novozymes) are commercially available. Arylglycerol compound (4-hydroxy-3,5-dimethoxyphenyl glycerol) was synthetized by Yue et al.[1]

Eucalyptus and spruce wood chips (about 0.5×2×3 cm^3) were supplied by the local pulping mills. The cellulolytic enzyme lignin (CEL) of *Eucalyptus* was prepared previously.[2] The structural feature of the CEL was characterized by 2D HSQC NMR, as shown in Figure S1.

1.2 Synthesis of model compounds

1.2.1 α-Condensed β-aryl ether compound

The synthetic pathway is shown in Scheme S1. The detailed experimental procedures for transforming **1a** to **1e** were similar to those in the literature.[2]



Scheme S1 Synthetic pathway for α-condensed β-aryl ether compound **1f**

Compound **1f**: **1e** (0.4 g, 1.25 mmol) and 2,6-dimethoxyphenol (2.89 g, 1.875 mmol) were dissolved in 6 mL dioxane/water (5:1, v/v) in a 25-mL flask. The flask was immersed in ice-water bath and sulfuric acid (98%, 1 mL) was added into the continuously stirring solution. The mixture was heated to 50 °C and kept for 3 h, after which the reaction was monitored by TLC (EtOAc/n-Hexane, 1:2, v/v) to ensure the starting material was used-up. The mixture was transferred into an 80-mL separating funnel charged with 30-mL saturated NH_4Cl. The organics were extracted by EtOAc (4×20 mL) and the combined organic phase was concentrated under reduced pressure after dried over anhydrous $MgSO_4$. The crude oil was purified by a flash column chromatography on 50-g SiO_2 with mobile phase EtOAc/n-hexane (1:2, v/v) to afford the aim product **1f** (38% in yield). ^{1}H NMR (600 MHz, DMSO-d6) δ 8.72 (s, 1H), 8.11 (s, 1H), 6.97 (dd, *J* = 11.1, 5.6 Hz, 1H), 6.94 (d, *J* = 1.9 Hz, 1H), 6.74 (dd, *J* = 8.2, 1.9 Hz, 1H), 6.69 (d, *J* = 8.1 Hz, 1H), 6.63 (d, *J* = 8.4 Hz, 2H), 6.61 (s, 2H), 4.30 (m, *J* = 26.6, 9.3, 7.5 Hz, 2H), 4.21 (t, *J* = 7.3 Hz, 1H), 3.75 (s, 3H), 3.73 (s, 6H), 3.71 (s, 6H); ^{13}C NMR (600 MHz, DMSO-d6) δ 153.73 (s), 148.23 (s), 147.72 (s), 145.34 (s), 137.20 (s), 134.46 (s), 134.11 (s), 133.31 (s), 124.08 (s), 120.77 (s), 115.64 (s), 112.93 (s), 106.18 (d, *J* = 4.1 Hz), 75.57 (s), 60.23 (s), 56.50 (s), 56.36 (s), 56.15 (s), 50.99 (s).

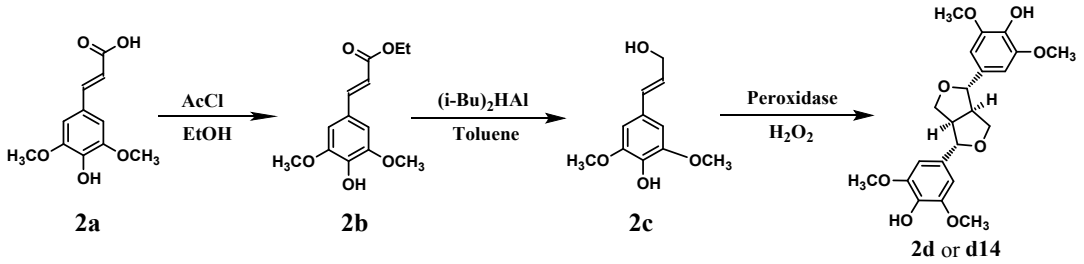
1.2.2 Syringaresinol

Compound **2b**: Sinapic acid (**2a**, 10 g, 54 mmol) was added in a 250-mL flask charged with absolute ethanol (100 mL) and acetylchloride (6.4 mL, 81 mmol). A mass of sinapic acid was not dissolved at this time. The mixture was kept stirring until the solution became clear (about 12 h later). The mixture was tested by TLC to ensure the finish of the reaction. Then, the solution was concentrated under reduced pressure and the generated hydrochloric acid (HCl) was removed by continuous evaporation with continuous addition of ethanol, and the product ethyl sinapate **2b** was finally obtained with the yield of 98%. ^{1}H NMR (600 MHz, DMSO-d6) δ 8.95 (s, 1H), 7.55 (d, J = 15.9 Hz, 1H), 7.03 (s, 2H), 6.53 (d, J = 15.9 Hz, 1H), 4.21-4.12 (m, 2H), 3.80 (s, 6H), 1.28-1.21 (m, 3H).

Compound **2c**: Ethyl sinapate (**2b**, 5 g, 19.8 mmol) was dissolved in toluene (80 mL) and cooled by ice-water bath. Then, 40 mL 1.5 mol/L diisobutyl aluminum hydride ($(i\text{-}Bu)_2HAl$) was slowly injected into the solution. Five hours later, the mixture was tested by TLC to ensure the reaction finished. The solvent was evaporated after the reaction was quenched by adding ethanol. Then, 60 mL deionized water was added, and the organics were extracted by EtOAc (4×50 mL). The combined solution was washed

with saturated NH_4Cl, dried over anhydrous $MgSO_4$ and concentrated by rotary evaporator to afford oily crude product. The crude oil was purified upon crystallization in the mixture of EtOAc and n-hexane. Finally, the product sinapyl alcohol (**2c**) was obtained as a white crystal with the yield of 84%. ^{1}H NMR (600 MHz, acetone-d6) δ 7.03 (s, 1H), 6.71 (s, 2H), 6.48 (dt, J = 15.9, 1.5 Hz, 1H), 6.24 (dt, J = 15.9, 5.5 Hz, 1H), 4.20 (td, J = 5.6, 1.5 Hz, 2H), 3.82 (s, 6H), 3.88 (t, J =5.65 Hz, 1H).

Compound **2d**: The coupling reaction of sinapyl alcohol was performed according to method described in the literatures[3, 4]. Compound **2c** (2.8 g, 13.3 mmol) was dissolved in 60-mL acetone. Then, it was dispersed into the solution of sodium dihydrogen phosphate (240 mL, pH = 4.5). The mixture was maintained agitating. After adding horseradish peroxidase (40 mg), 30% hydrogen peroxide (1.66 g) was slowly added over 20 min. After 2.5 h later, the mixture was saturated with NH_4Cl, and extracted with dichloromethane (3×200 mL). The organic phase was collected and concentrated after dried over anhydrous $MgSO_4$. The oily mixture was loaded onto 80 g SO_2 column and eluted with 40% EtOAc solution in n-hexane. ^{1}H NMR (600 MHz, $CDCl_3$) δ 6.58 (s, 4H), 5.53 (s, 2H), 4.73 (d, J = 3.9 Hz, 2H), 4.34–4.23 (m, 2H), 3.95–3.85 (m, 2H), 3.89 (s, 12H), 3.16–3.03(m, 2H). ^{13}C NMR (600 MHz, $CDCl_3$) δ 147.1 (s), 134.2(s), 132.0(s), 102.6(s), 86.0(s), 71.8(s), 56.3(s), 54.3(s).



Scheme S2 Synthetic pathway for syringaresinol (**2d** or **d14**)

1.3 Reaction of model compounds

The reactions of lignin model compounds under alkali condition were performed in a 20-mL hydrothermal reactor with Teflon inner container. Model compound (20 mg) was dissolved in 8 mL 1 mol/L NaOH solution and reacted under 170 °C for 1 h in an oil bath. Then, the reactor was cooled by flowing water. The solution was transferred into a separating funnel, acidified with hydrochloric acid (2 mol/L), and extracted by dichloromethane (4×10 mL). The combined organics were washed with saturated NH_4Cl, dried over anhydrous $MgSO_4$, and then concentrated. The oil products were acetylated with pyridine (1 mL) and acetic anhydride (1 mL) for 4 h. The acetylated products were dissolved in 1-mL dichloromethane and the solution was transferred into a GC vial for GC-MS (Shimadzu GCMS-

TQ8040 triple quadrupole GC/MS/MS, equipped with a SH-Rxi-5Sil MS column) analysis.

GC-MS parameters: inlet temperature 300 °C, total flow rate of Helium carrier gas 21.1 mL/min, ion source temperature 250 °C, interface temperature 300 ºC, ion range 45-800 (m/z). Oven program: 80 °C for 5 min, ramp 8 °C/min to 250 °C, 5 °C/min to 300 °C, hold 30 min, total analysis time 66.25 min.

The reaction of Cα-condensed β-aryl ether compound (**1f**) was performed with a quantity of 90 mg and 1 mol/L NaOH solution of 15 mL. The finally concentrated reaction product was transferred onto a 1 mm^2 High-Performance Thin Layer Chromatograph (HPTLC, 20×20 cm^2, Merk), and spread with EtOAc/n-hexane solution (1:2, v/v) to separate the diarylethene products.

1.4 Pulping, lignins recovery and separation

The alkali and kraft pulping processes of wood feedstocks were conducted in a 1-L digester under the conditions seen in Table S1. The spent pulping liquid (100 mL) was collected and adjusted pH to 2.5 with 4 mol/L HCl after being on standing for 12 h. Then the turbid liquid was centrifuged and the precipitate was washed with deionized water (2×100 mL, pH <3) to remove salts and saccharides, then freeze-dried to afford crude lignin. The yields of different types of crude lignin was in range of 78-90% (based on the removed Klason lignin from feedstock).

Table S1 Conditions for pulping process.

Feedstock	NaOH dosage	Sulfidity	Liquor ratio	Temperature and time
Spruce	26%	0	4.5	170 ℃ for 2 h
	26%	30%	4.5	
Eucalyptus	22%	0	4.5	162 ℃ for 40 min, and raise to 170 ℃
	22%	30%	4.5	in 20 min and hold 50 min

The crude lignin (2.5 g) was dispersed in dichloromethane (DCM, 50 mL) in a glass container. The suspension was agitated for 30 min under room temperature. Afterwards, the suspension was filtrated by a glass filter (G4, pore size 5-15 μm). The procedure was repeated for 3 times. Finally, the combined filter liquor was condensed to afford the DCM-soluble fraction, and the solid fraction was dried under reduced pressure to afford the DCM-insoluble fraction.

1.5 2D HSQC NMR analysis

2D HSQC (heteronuclear single quantum coherence) NMR spectra of lignin model compounds or DCM-soluble/insoluble fractions of alkali lignin were recorded on a Bruker AVANCE III HD 600 MHz

spectrometer. Lignin model compound (~10 mg) was dissolved into DMSO-d6 (0.5 mL) in a NMR tube. The spectrum was recorded following the condition: F2 (^{1}H) (12-0) ppm with 2048 data points and 1 s recycle delay, and F1 (^{13}C) with 256 increments of 8 scans (the total time was about 37 min). DCM-soluble or DCM-insoluble fraction (~80 mg) was dissolved into DMSO-d6 (~80 mg), and the 2D HSQC NMR spectrum was recorded following the condition: F2 (^{1}H) (12-0) ppm with 2048 data points and 1 s recycle delay, F1 (^{13}C) with 256 increments of 64 scans (the total time was about 6 hours). The ^{1}H-^{13}C correlation signals of acquired spectra were calibrated with the peak of DMSO (δ (39.5, 2.49) ppm). Volume integration of contour signals in the spectra was performed with use of Bruker's Topspin 4.0.3 software for assessing the content of lignin structural units.

1.6 GC-MS analysis of DCM-soluble fraction

The DCM-soluble fraction (40 mg) from crude eucalyptus alkali lignin was acetylated with 1-mL pyridine and l-mL acetic anhydride for 5 h, after which the acetic anhydrides were removed under reduced pressure by continuously adding ethanol. The acetylated product was transferred into a 1.5-mL GC vial with 1 mL dichloromethane, then analyzed by GC-MS (Shimadzu GCMS-TQ8040 triple quadrupole GC/MS/MS, equipped with a SH-Rxi-5Sil MS column). The parameters and condition for analysis were same with that for analysis of model compound products (described in section **1.3**).

1.7 Molecular weight

The DCM-insoluble fraction or DCM-soluble fraction (5 mg) was acetylated with pyridine (250 μL) and acetic anhydride (250 μL). The acetylated product was dissolved in 2.5-mL tetrahydrofuran, of which 1 mL was transferred into a 1.5-mL sample vial after interception with a 0.2-μm filter. The molecular weight of samples was analyzed by using a High-Performance Liquid Chromatograph (Agilent infinity 1260 HPLC) equipped with a PLgel Mixed-C column and a Refractive Index Detector. Tetrahydrofuran was used as the mobile phase, temperature of column oven 30 ºC, flow velocity 1 mL/min. The reference polystyrenes with the weight-average molecular weight in range of 208-49000 g/mol were applied for calibration of the molecular weight of the samples.

2 Supplementary data

2.1 Figures and Tables

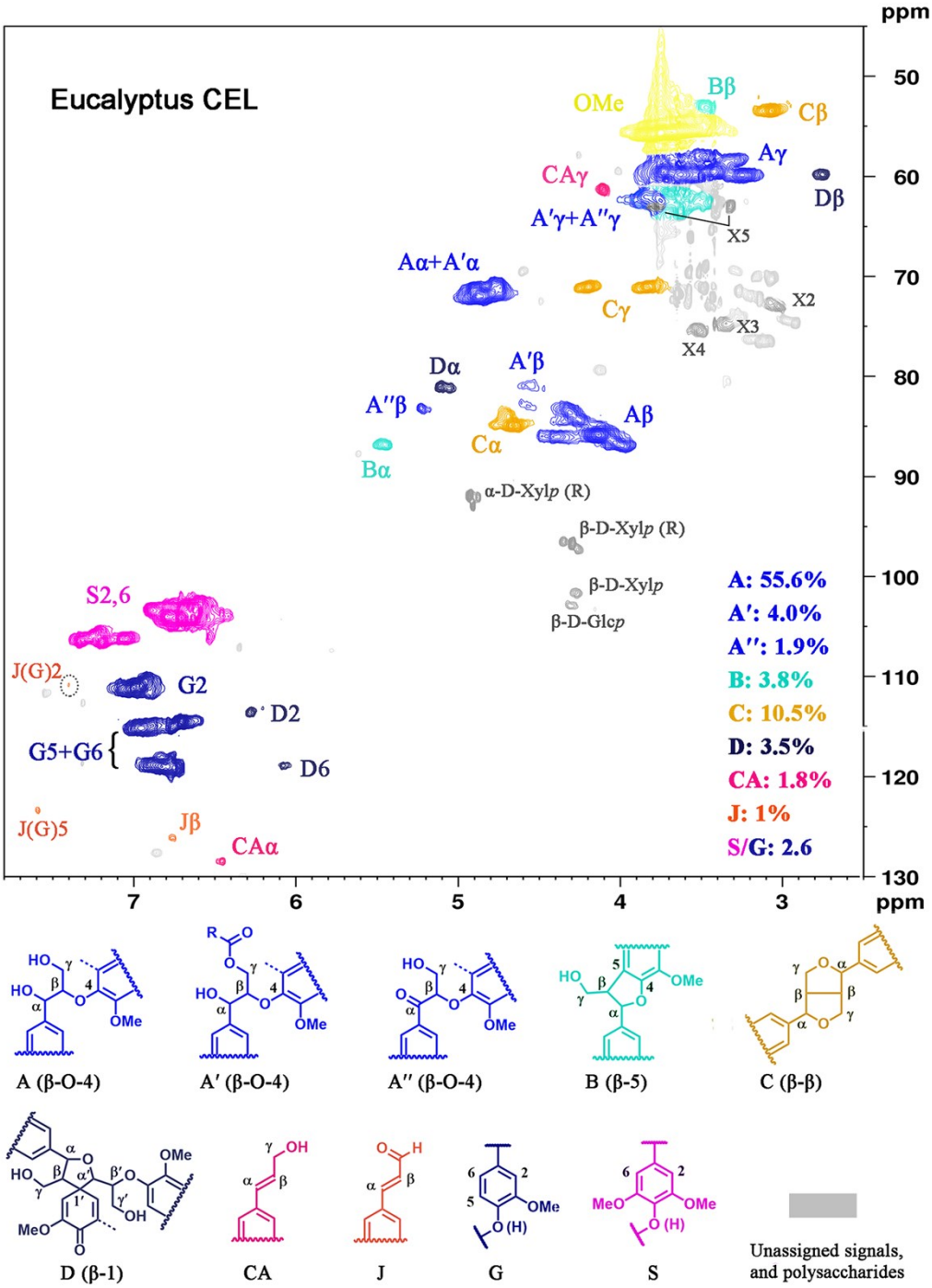


Figure S1 2D HSQC NMR of *Eucalyptus* cellulolytic enzyme lignin (CEL). Abbreviations of polysaccharides: X (xylan), β-D-Glc*p* (β-D-glucopyranoside), α-D-Xyl*p* (α-D-xylopyranoside), and β-D-Xyl*p* (β-D-xylopyranoside).

.

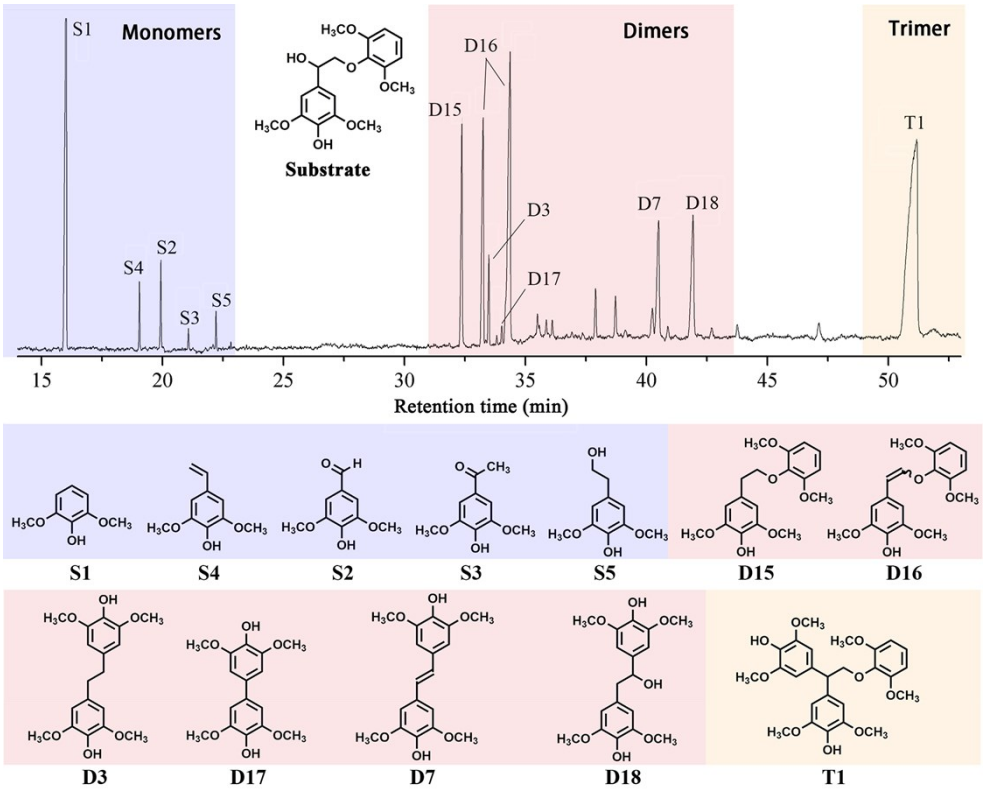


Figure S2 Preliminary analysis of the products from "S-S" β-aryl ether compound (syringyl glycol β-syringyl ether) by using GC-MS after alkaline treatment (1 mol/L NaOH, 170 ºC, 1 h). According to a preliminary analysis, "S-S" β-aryl ether compound reacted to produce several kinds of condensed products (**D3**, **D7**, **D17**, **D18** and **T1**).

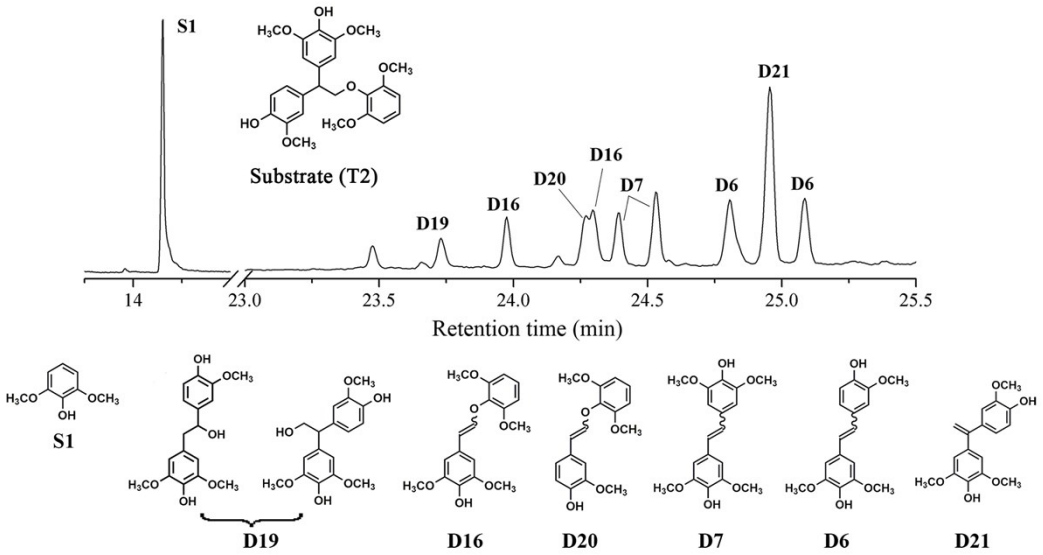


Figure S3 Preliminary analysis of the products from the Cα-condensed β-aryl ether compound (**T2**) by using GC-MS after alkaline treatment (1 mol/L NaOH, 170 ºC, 1 h). According to a preliminary analysis, apart from styryl ether products (**D16** and **D20**), diaryl ethanols (**D19**) and diarylethenes (**D6**, **D7** and **D21**) were presumably produced.

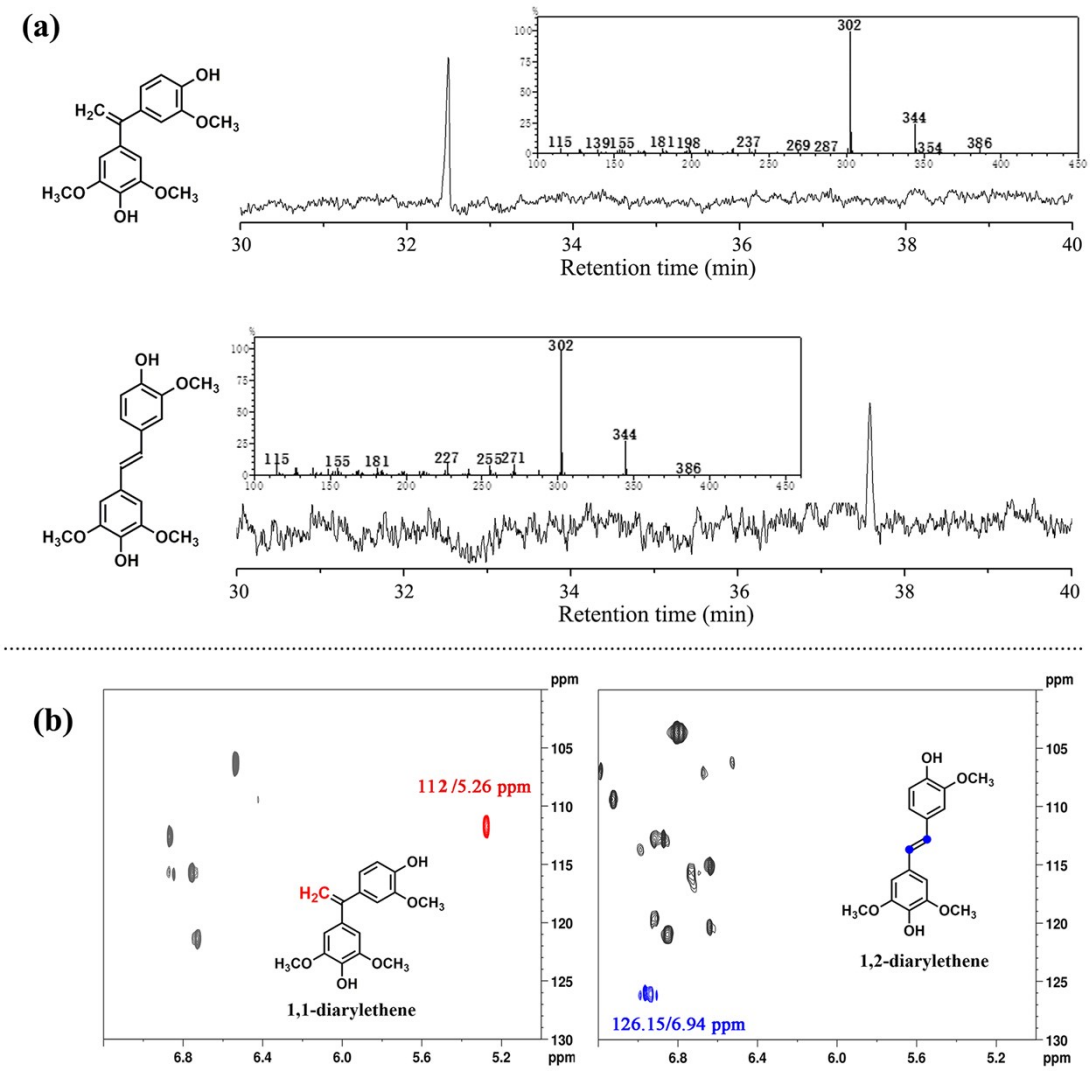


Figure S4 Evidence for the generation of two kinds of diarylethenes (1,1-diarylethene and 1,2-diarylethene) from Cα-condensed β-aryl ether compound (**T2**). After separation of the products of **T2**, 1,1-diarylethene and 1,2-diarylethene were authenticated by analytical techniques. (a) Shows that the compounds have different retention time whereas a similar MS information; (b) shows the representative signals of "=CH_2" and "=CH-" in 2D HSQC NMR in 1,1-diarylethene and 1,2-diarylethene.[5,6]

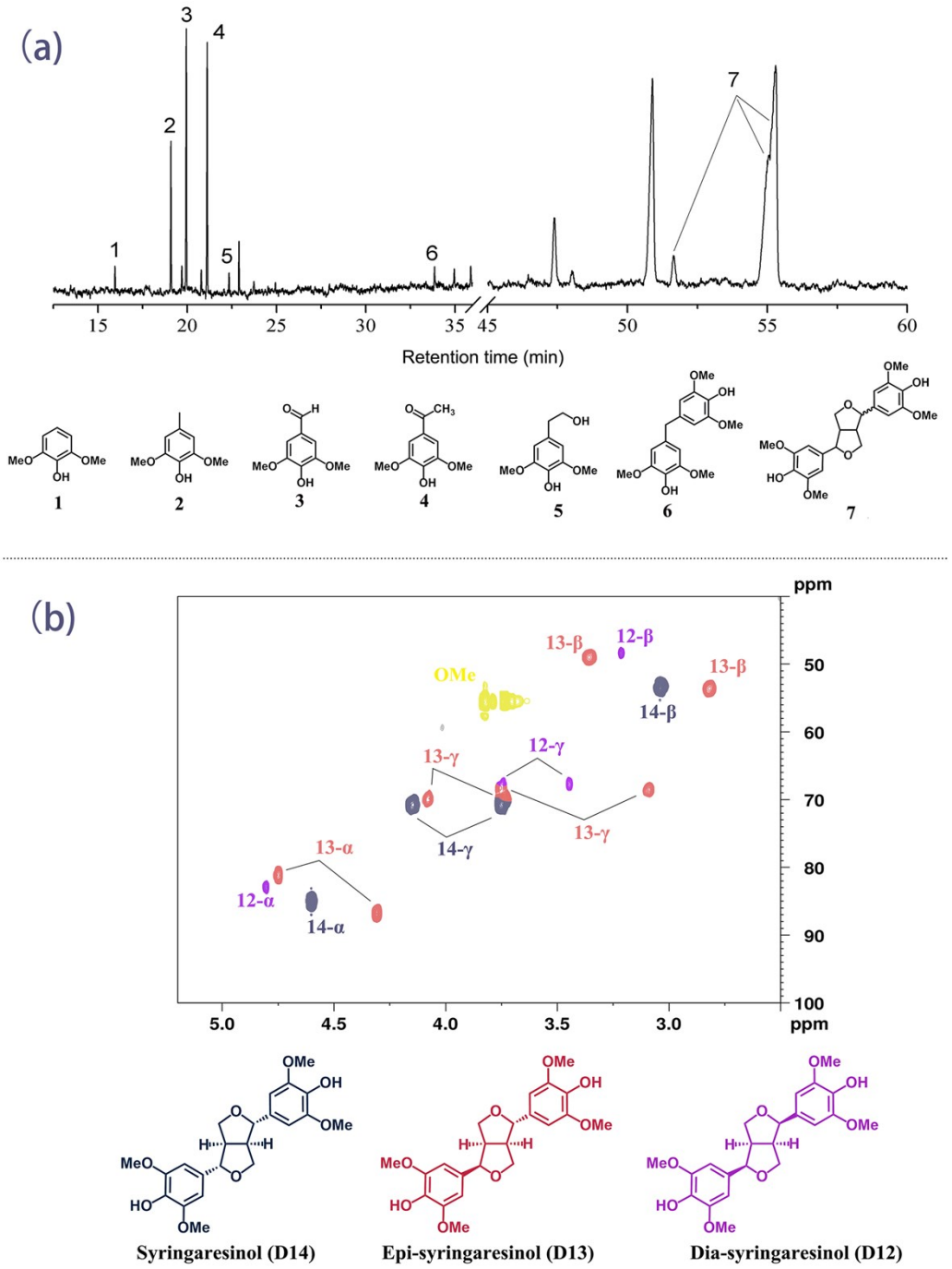


Figure S5 Analysis of the product of β-β model compound (syringaresinol, **D14**) after alkali treatment (1 mol/L NaOH, 170 ºC, 1 h). (a) GC-MS analysis of the products. Two signals with retention time of 52.0 min and 54.9 min in the chromatogram have similar MS information with the syringaresinol (retention time 55.1 min), which presumably come from the isomers of syringaresinol. (b) Partial 2D HSQC NMR spectrum of the syringaresinol product. According to the assignment of multiple signals in the spectrum,[7,8] two isomers of syringaresinol, dia-syringaresinol (**D12**) and epi-syringaresinol (**D13**), were proved to be generated in the product.

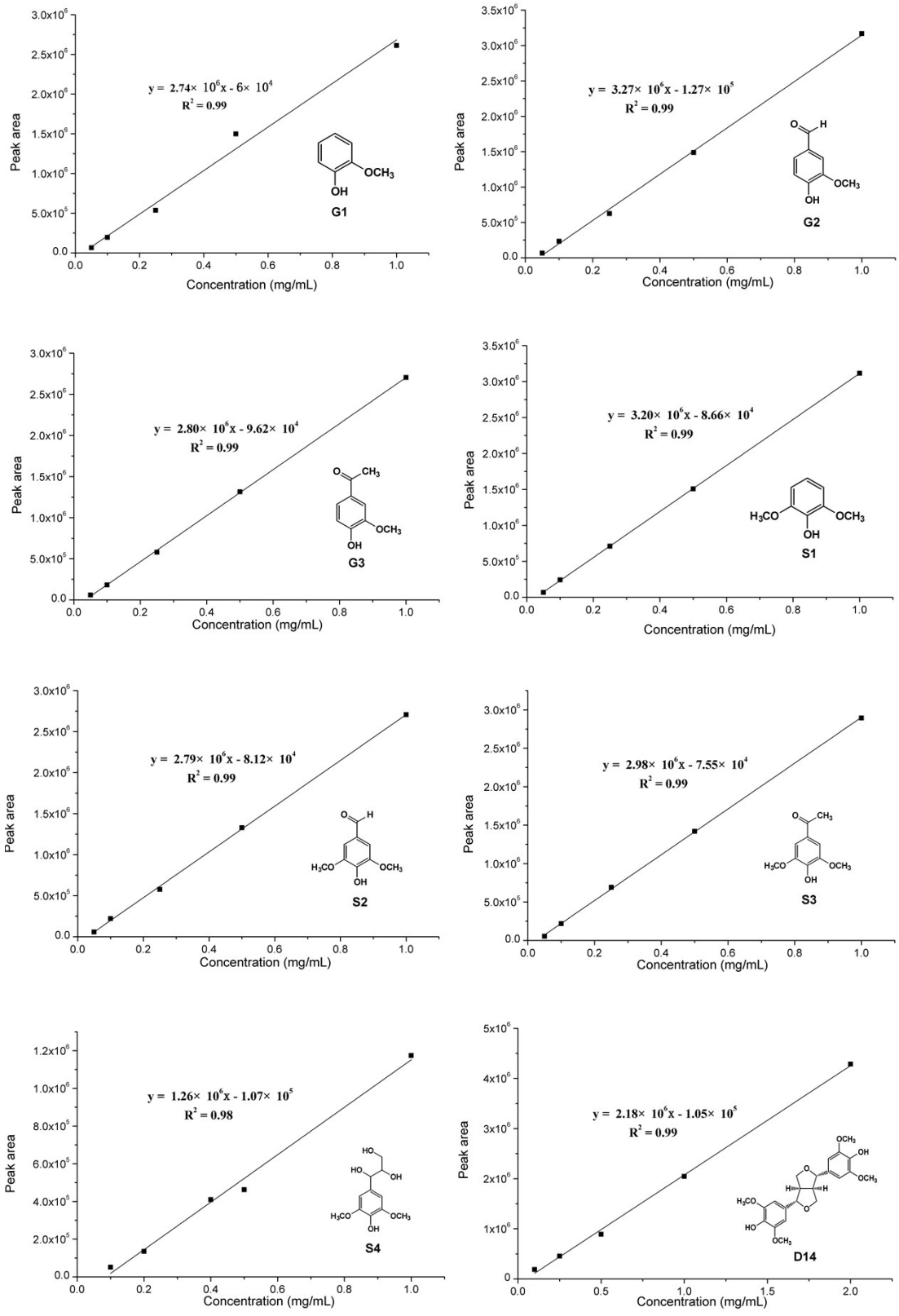


Figure S6 Calibration curves for quantification of compounds in GC chromatograms.

Table S2 ^{1}H-^{13}C correlations of the structural moieties in eucalyptus alkali lignins.[2, 3,5,9-12]

Labels	Substructures	δC/δH correlations (ppm)
A	β-O-4	α: 70.9/4.77; β: 84.1/4.30; γ: 59.9/3.60 and 3.26
A’	γ-etherified β-O-4	α: 70.9/4.77; β: 80.8/4.52; γ: 64.09/4.06 and 4.49
A’’	α-carbonylated β-O-4	β: 82.6/5.11; γ: 62.39/4.17
B	phenylcoumaran (β-5)	α: 86.8/5.49; β: 53.2/3.47; γ: 62.9/3.73 and 3.62
C	Resinol (β-β)	α: 85.1/4.63; β: 53.6/3.07; γ: 70.9/4.16 and 3.76
C’	Epiresinol (β-β)	α: 87.0/4.34; α': 81.2/4.77; β: 53.7/2.84; β': 49.2/3.35; γ: 70.2/4.06 and 3.75; γ': 68.8/3.77 and 3.12
C’’	Diaresinol (β-β)	α: 83.1/4.82; β: 48.5/3.19; γ: 67.5/3.78 and 3.45
D	Spirodienone (β-1)	α: 81.2/5.01; α': 83.6/4.68; β: 59.5/2.75; β': 79.4/4.10; G2': 113.0/6.17; G5': 127.9/6.06; G6': 151.2/7.07; S2': 113.5/6.25; S6': 118.9/6.06
SB1	*trans*-stilbene (β-1)	α: 125.6/6.97
SB5	*trans*-stilbene (β-5)	α: 128.2/7.07; β: 120.1/7.22
DE	Diarylethane (β-1)	α: 39.8/2.74-2.90
E(G-G)	enol ether	α-*trans*: 112.0/6.14; α-cis: 109.1/5.56
E(S-G)	enol ether	α-*trans*: 112.10/6.10; α-cis: 109.45/5.57
E(G-S)	enol ether	α-*trans*: 108.20/5.72; α-cis: 106.42/5.27
E(S-S)	enol ether	α-*trans*: 108.70/5.77; α-cis: 106.80/5.28
Gly	Arylglycerol	α: 73.0/4.45; β: 75.5/3.48; γ: 62.3/3.16-3.36
CA	Cinnamyl alcohol	α: 128.6/6.49; β: 128.4/6.26; γ: 61.4/4.10
P	Aryl propylalcohol	α: 31.1/2.51; β: 34.4/1.70; γ: 60.0/3.42
Aldehyde in vanillin and syringaldehyde		191.8/9.71-9.84
Ketone in vanillone and acetosyringone		26.4/2.52
G	Guaiacyl	G2:109-113.5/6.62-7.55; G5: 113.9/116.4/6.52-7.16; G6: 117-122/6.5-7.1
S	Syringyl	S2,6: 101-108/6.06-7.43

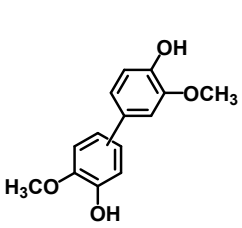
Table S3 Quantitative analysis of lignin-derived monomers and dimers in low-Mw fraction (DCM-soluble fraction) by using GC-MS.

Compounds	Peak area	Content (mg/g DCM-soluble fraction) [a]
G1	105807	1.50
G2	152313	2.13
G3	75398	1.50
S1	181697	2.10
S2	773353	7.75
S3	528807	5.25
S4	131413	4.75
D1	60528	1.90
D2	168049	3.13
D3	452090	6.39
D4	210223	3.61
D5	283066	4.45
D6	452949	6.40
D7	641086	8.56
D8	295879	4.60
D9	813445	10.53
D10	1416846	17.45
D11	2899708	34.46
D12	311054	5.69
D13	1846519	23.3
D14	3028718	36.85

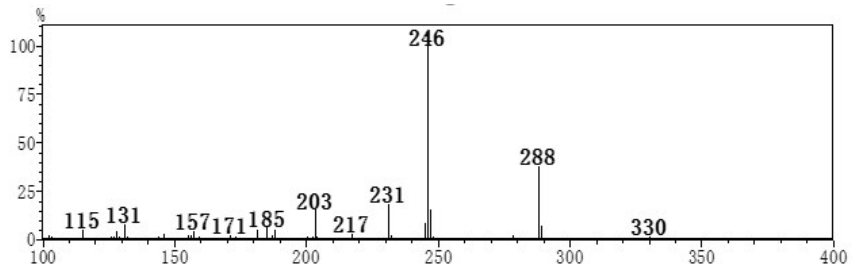
[a] a The content of G1, G2, G3, S1, S2, S3, S4, D12, D13, and D14 in DCM-soluble fraction was determined by external standard method, in which standard curves (peak area vs. concentration) of synthetic reference compounds (seen in Figure S6) were plotted for calibration; the content of D1-D11 were determined with the calibration of a same reference compound, syringaresinol.

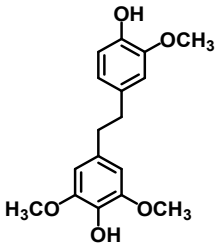
2.2 MS information

Compounds and the MS information of their acetylated forms are listed below:

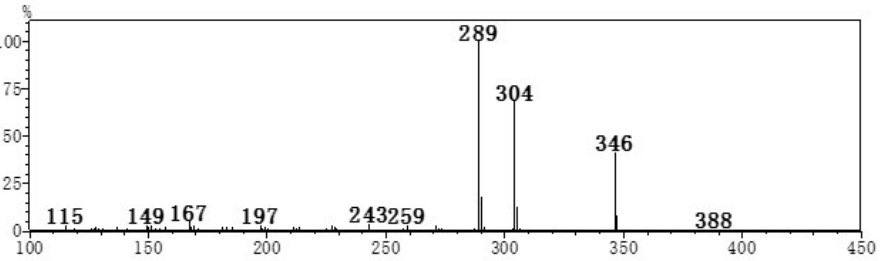


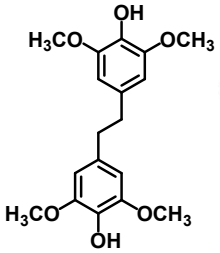
D1



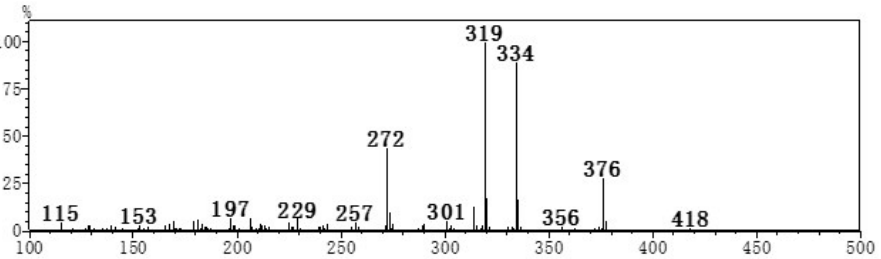


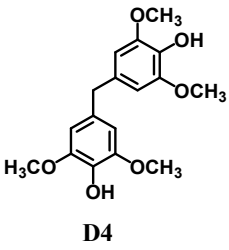
D2



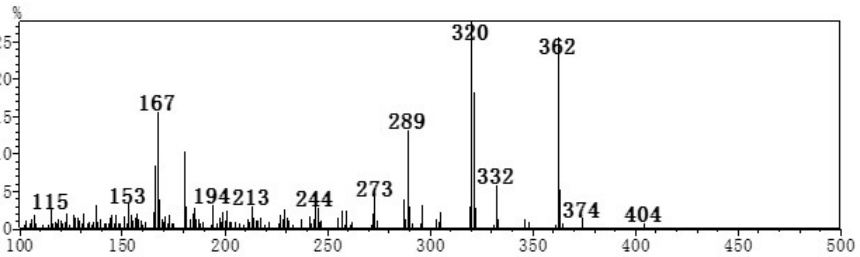


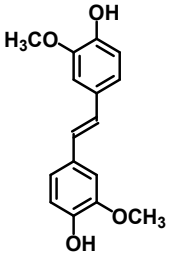
D3



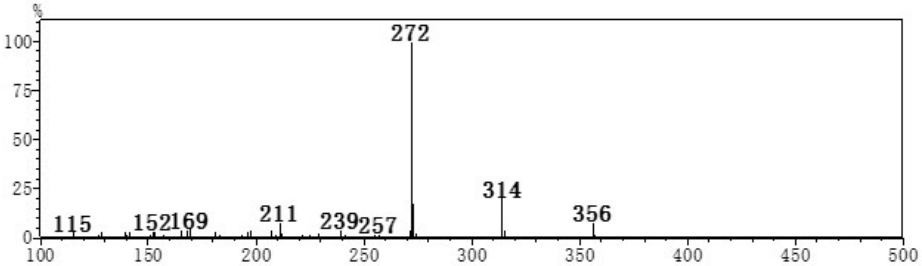


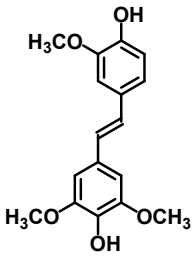
D4



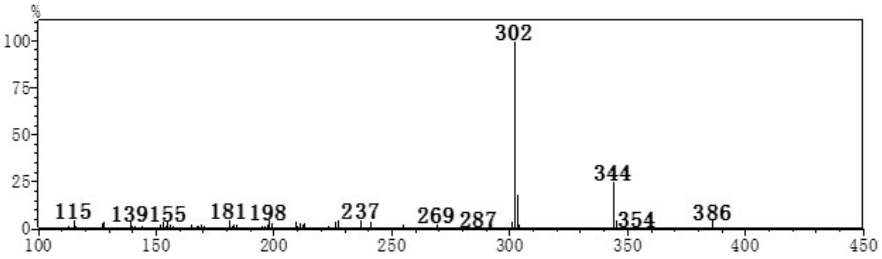


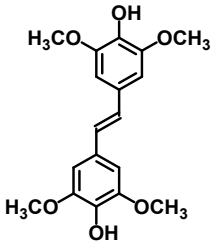
D5



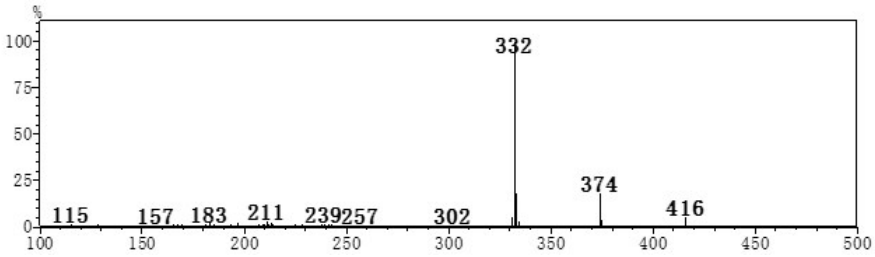


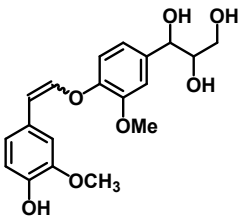
D6



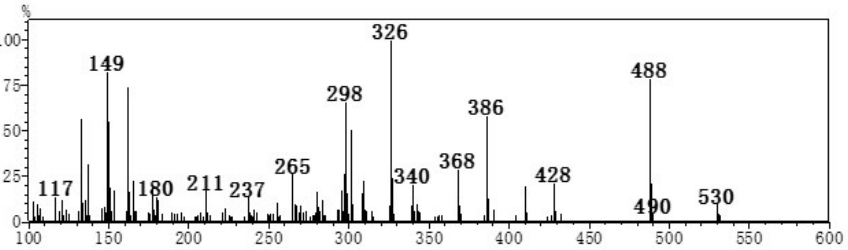


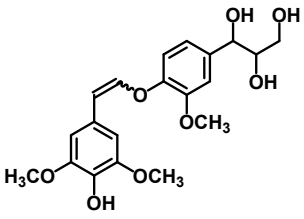
D7



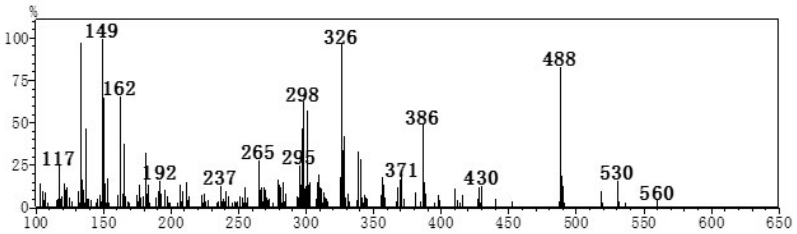


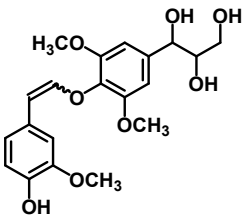
D8



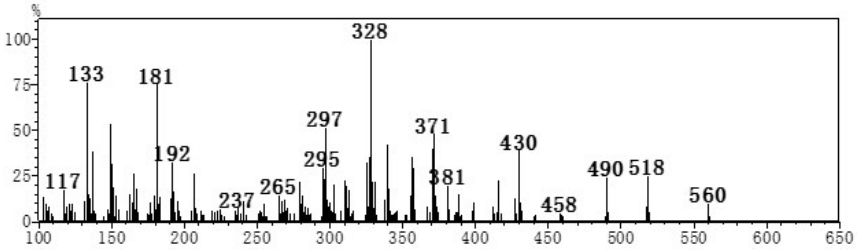


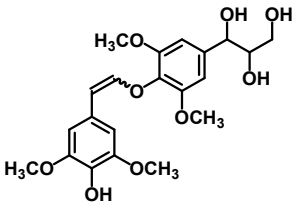
D9



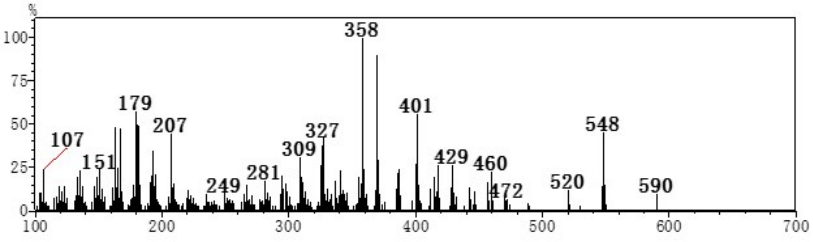


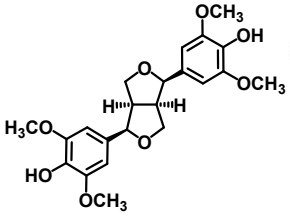
D10



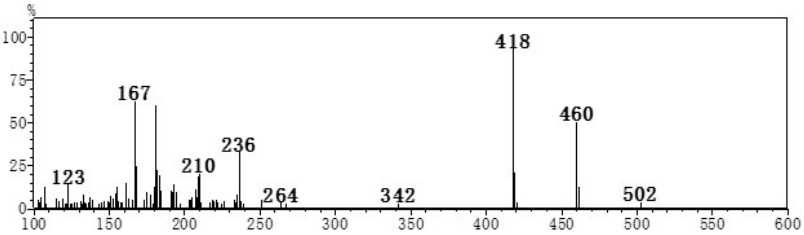


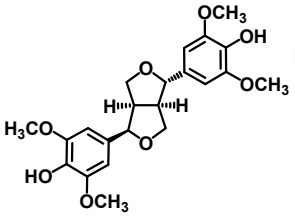
D11



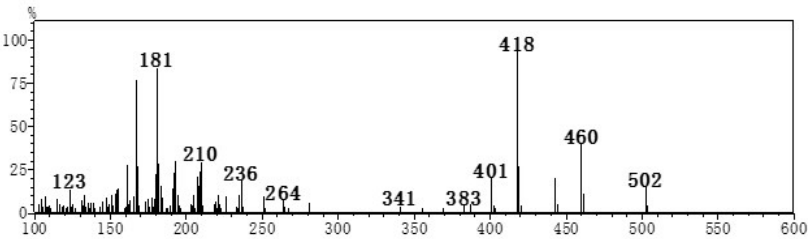


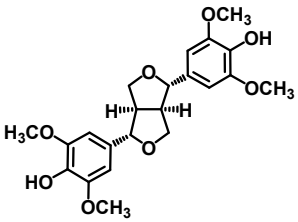
D12



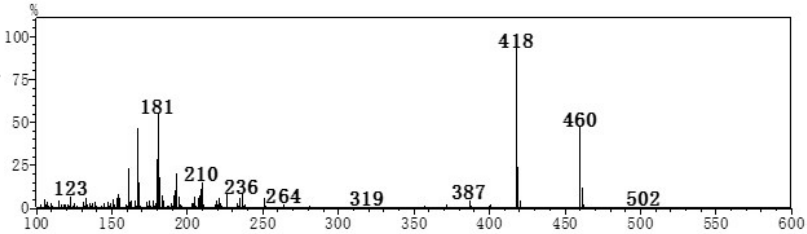


D13





D14



References

1. F. Yue, F. Lu, R. C. Sun and J. Ralph, *Journal of Agricultural and Food Chemistry*, 2012, **60**, 922-928.
2. C. Zhao, J. Huang, L. Yang, F. Yue and F. Lu, *Industrial & Engineering Chemistry Research*, 2019, **58**, 5707-5714.
3. F. Lu and J. Ralph, *Organic & Biomolecular Chemistry*, 2008, **6**, 3681.
4. M. Habib, M. Trajkovic and M. W. Fraaije, *ACS Catalysis*, 2018, **8**, 5549-5552.
5. S. Constant, H. L. J. Wienk, A. E. Frissen, P. d. Peinder, R. Boelens, D. S. van Es, R. J. H. Grisel, B. M. Weckhuysen, W. J. J. Huijgen, R. J. A. Gosselink and P. C. A. Bruijnincx, *Green Chemistry*, 2016, **18**, 2651-2665.
6. D. P. Ojha and K. R. Prabhu, *The Journal of Organic Chemistry*, 2012, **77**, 11027-11033.
7. F. Tran, C. S. Lancefield, P. C. J. Kamer, T. Lebl and N. J. Westwood, *Green Chemistry*, 2015, **17**, 244-249.
8. F.-R. Chang, Y.-C. Chao, C.-M. Teng and Y.-C. Wu, *Journal of Natural Products*, 1998, **61**, 863-866.
9. F. Alonso, P. Riente and M. Yus, *Tetrahedron*, 2009, **65**, 10637-10643.
10. C. S. Lancefield, H. L. J. Wienk, R. Boelens, B. M. Weckhuysen and P. C. A. Bruijnincx, *Chem Sci*, 2018, **9**, 6348-6360.
11. S. Thangaraj, W.-S. Tsao, Y.-W. Luo, Y.-J. Lee, C.-F. Chang, C.-C. Lin, B.-J. Uang, C.-C. Yu, J.-H. Guh and C.-M. Teng, *Tetrahedron*, 2011, **67**, 6166-6172.
12. A. Lourenço, J. Rencoret, C. Chemetova, J. Gominho, A. Gutiérrez, H. Pereira and J. C. del Río, *BioEnergy Research*, 2015, **8**, 1946-1955.