Electronic Supplementary Information (ESI)

Impact of engineered lignin composition on biomass recalcitrance and ionic liquid pretreatment efficiency

Jian Shi[1,2,3], Sivakumar Pattathil[4, 5], Parthasarathi Ramakrishnan[1,2], Nickolas A. Anderson[6], Jeong Im Kim[6], Sivasankari Venketachalam[4, 5], Michael G. Hahn[4, 5], Clint Chapple[6], Blake A. Simmons[1,7], and Seema Singh[1,2,*]

Figure S1. pyro-GC/MS analysis of **a)** untreated and **b)** IL pretreated Arabidopsis genotypes including *COL* (wild type), *fah1-2* (G-lignin dominant), *C4H-F5H* (S-lignin dominant), *COMT1* (G/5-hydroxy G-lignin dominant) and *med5a med5b ref8* (H-lignin dominant) mutants.

Figure S2. Area-normalized SEC chromatograms of lignin extracted from different streams during IL pretreatment and enzymatic hydrolysis of different Arabidopsis genotypes including *COL* (wild type), *fah1-2* (G-lignin dominant), *C4H-F5H* (S-lignin dominant), *COMT1* (G/5-hydroxy G-lignin dominant) and *med5a med5b ref8* (H-lignin dominant) mutants. **a)** L_1: lignin from untreated biomass, **b)** L_2: solublized lignin in $[C_2C_1Im][OAc]$, **c)** L_3: lignin remaining in pretreated biomass. See **Table 3** for relative area of excluded and retained regions.

Figure S3. 2D HSQC NMR spectra of nonderivatized Arabidopsis cell walls of different genotypes including *COL* (wild type), *fah1-2* (G-lignin dominant), *C4H-F5H* (S-lignin dominant), *COMT1* (G/5-hydroxy G-lignin dominant) and *med5a med5b ref8* (H-lignin dominant) mutants: aliphatic (a-e), anomeric (f-j) and aromatic (k-o) regions of the HSQC spectrum. All contours are color-coded to match their respective structures in **Figure S3**. See **Table S2** for structural characteristics from integration of ^{13}C-^{1}H correlation peaks in the HSQC.

Figure S4. Main lignin structures present in Arabidopsis genotypes: (A) β-O-4 aryl ethers; (B) phenylcoumarans; (C) resinols; (D) dibenzodioxocins; (E) cinnamyl alcohol end-groups; (FA) ferulates; (*p*CA) *p*-coumarates; (H) *p*-hydroxyphenyl; (G) guaiacyl units; (S) syringyl units. Peak assignments are shown in **Table S2**.

Figure S5. Optimized geometries of inter-unit lignin linkages.

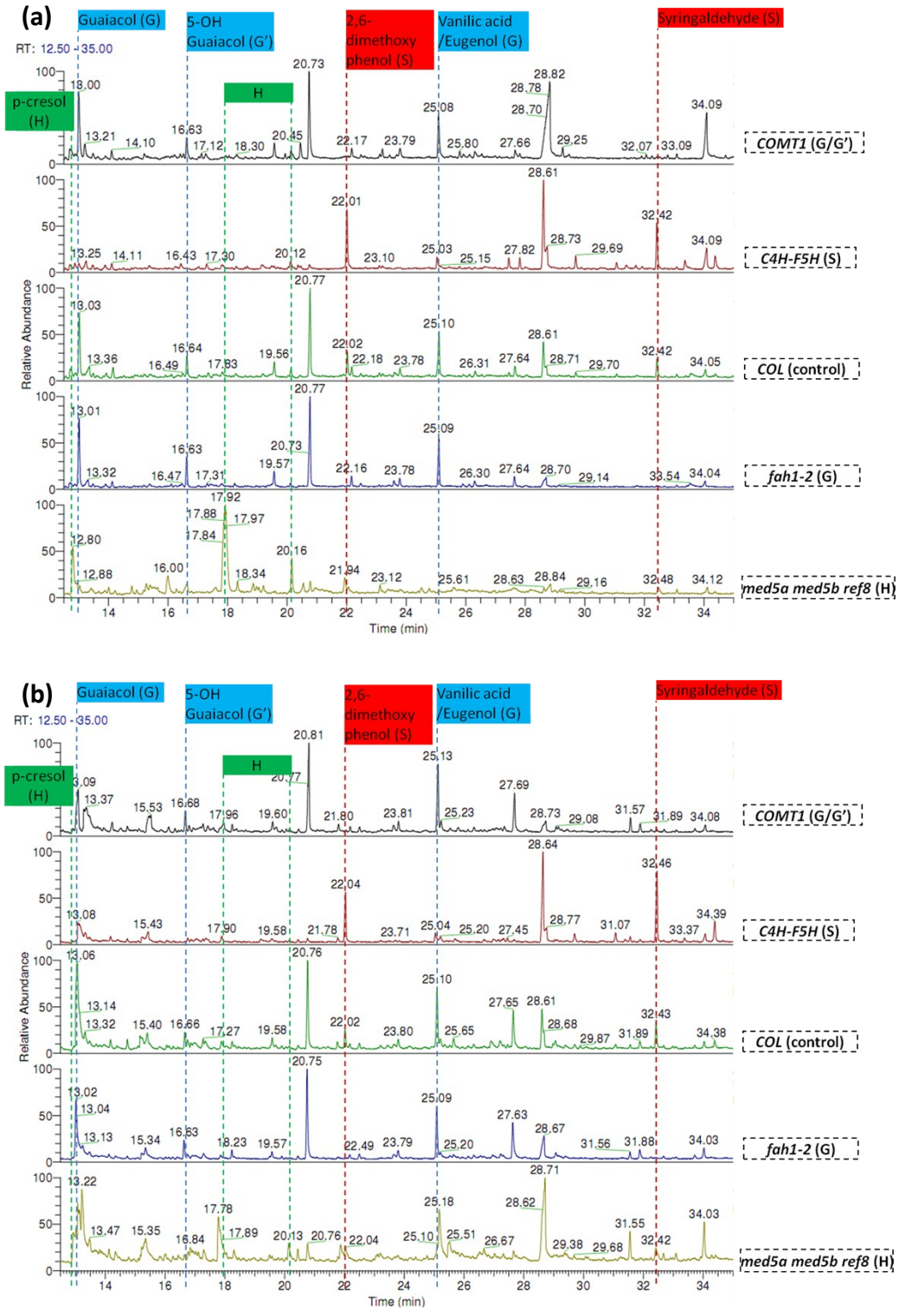


Figure S1. pyro-GC/MS analysis of **a)** untreated and **b)** IL pretreated Arabidopsis genotypes including *COL* (wild type), *fah1-2* (G-lignin dominant), *C4H-F5H* (S-lignin dominant), *COMT1* (G/5-hydroxy G-lignin dominant) and *med5a med5b ref8* (H-lignin dominant) mutants.

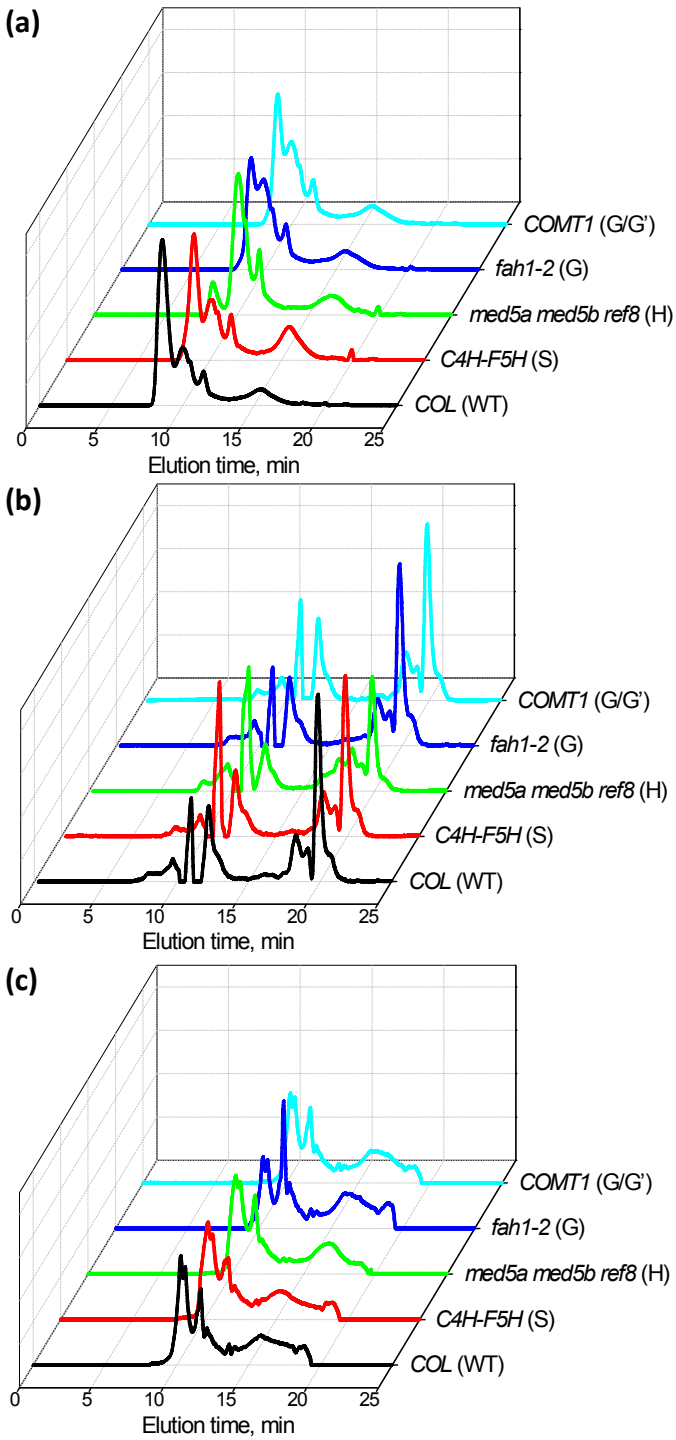


Figure S2. Area-normalized SEC chromatograms of lignin extracted from different streams during IL pretreatment and enzymatic hydrolysis of different Arabidopsis genotypes including *COL* (wild type), *fah1-2* (G-lignin dominant), *C4H-F5H* (S-lignin dominant), *COMT1* (G/5-hydroxy G-lignin dominant) and *med5a med5b ref8* (H-lignin dominant) mutants. **a)** L_1: lignin from untreated biomass, **b)** L_2: solubilized lignin in $[C_2C_1Im][OAc]$, **c)** L_3: lignin remaining in pretreated biomass. See **Table 3** for relative area of excluded and retained regions.

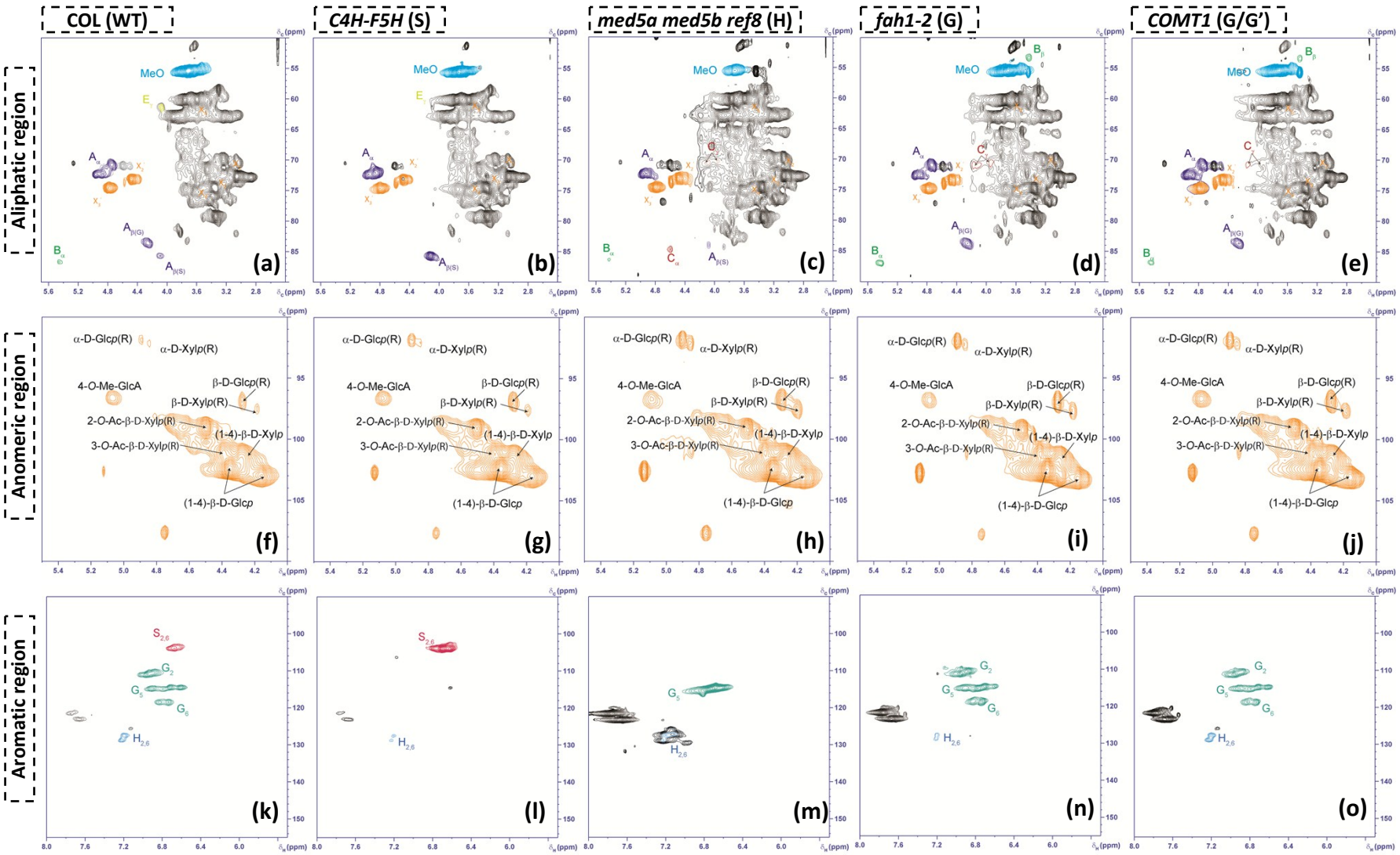


Figure S3. 2D HSQC NMR spectra of nonderivatized Arabidopsis cell walls of different genotypes including *COL* (wild type), *fah1-2* (G-lignin dominant), *C4H-F5H* (S-lignin dominant), *COMT1* (G/5-hydroxy G-lignin dominant) and *med5a med5b ref8* (H-lignin dominant) mutants; aliphatic (a-e), anomeric (f-j) and aromatic (k-o) regions of the HSQC spectrum. All contours are color-coded to match their respective structures in **Figure S4**. See **Table S2** for structural characteristics from integration of ^{13}C-^{1}H correlation peaks in the HSQC.

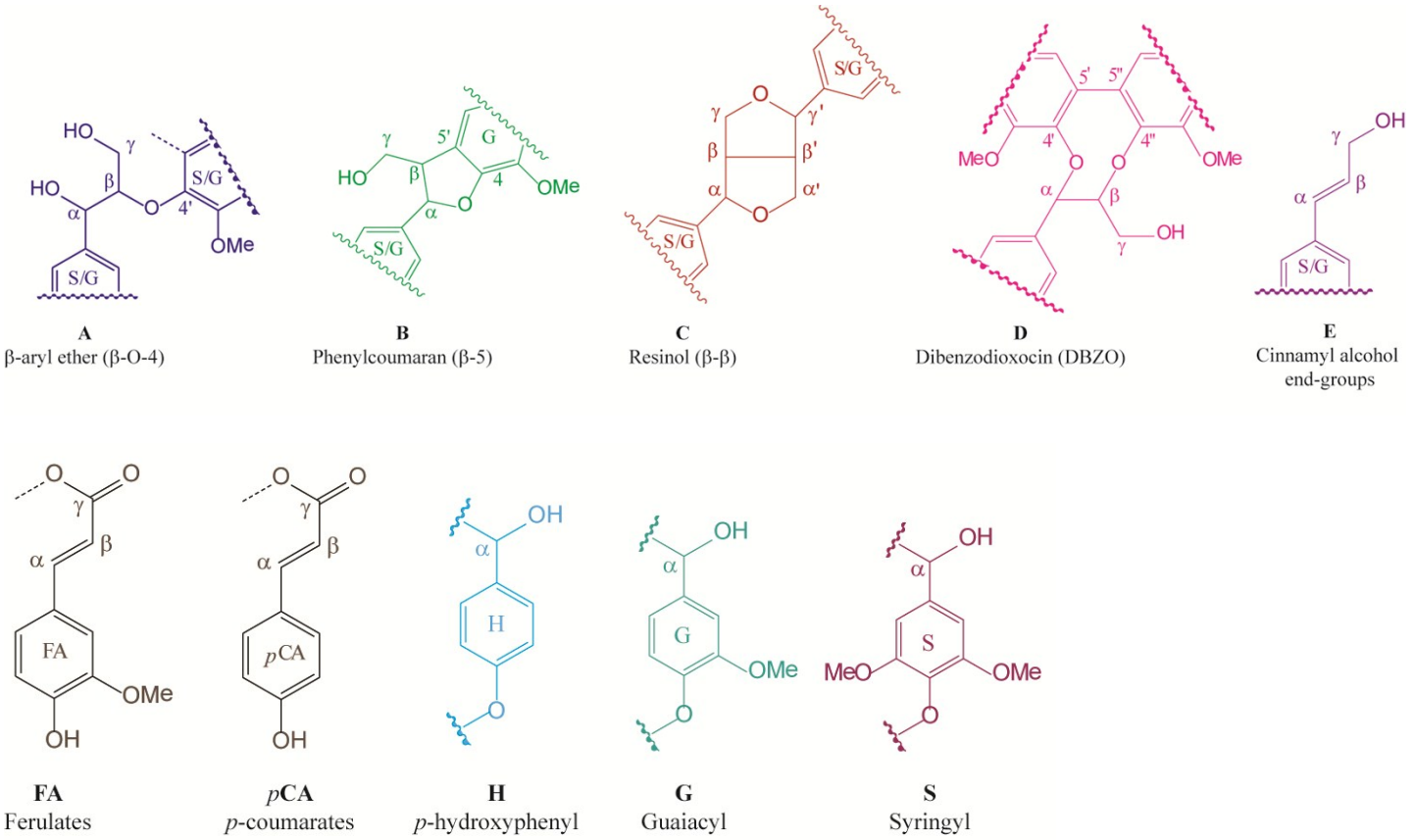


Figure S4. Main lignin structures present in Arabidopsis genotypes: (A) β-O-4 aryl ethers; (B) phenylcoumarans; (C) resinols; (D) dibenzodioxocins; (E) cinnamyl alcohol end-groups; (FA) ferulates; (*p*CA) *p*-coumarates; (H) *p*-hydroxyphenyl; (G) guaiacyl units; (S) syringyl units. Peak assignments are shown in **Table S2**.

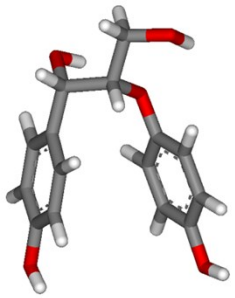
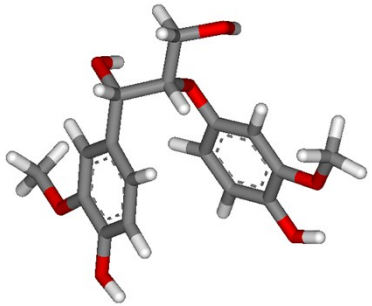
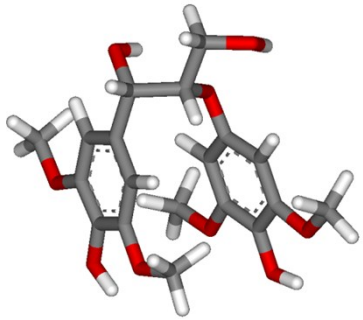
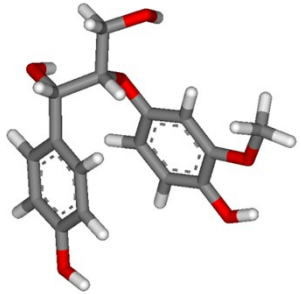
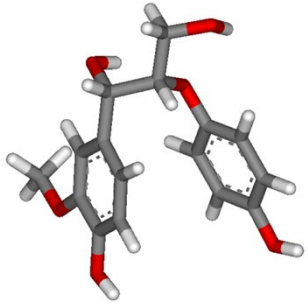
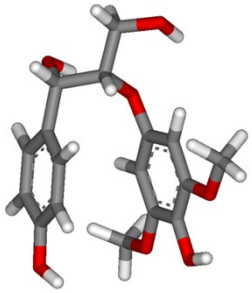
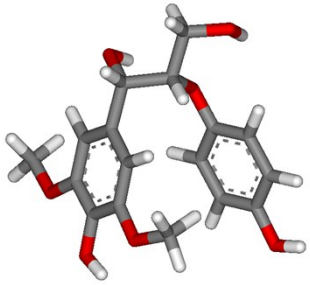
		
[β-*O*-4-H-H]	[β-*O*-4-G-G)]	[β-*O*-4-S-S]
		
[β-*O*-4-H-G)]	[β-*O*-4-G-H)]	
		
[β-*O*-4-H-S)]	[β-*O*-4-S-H)]	

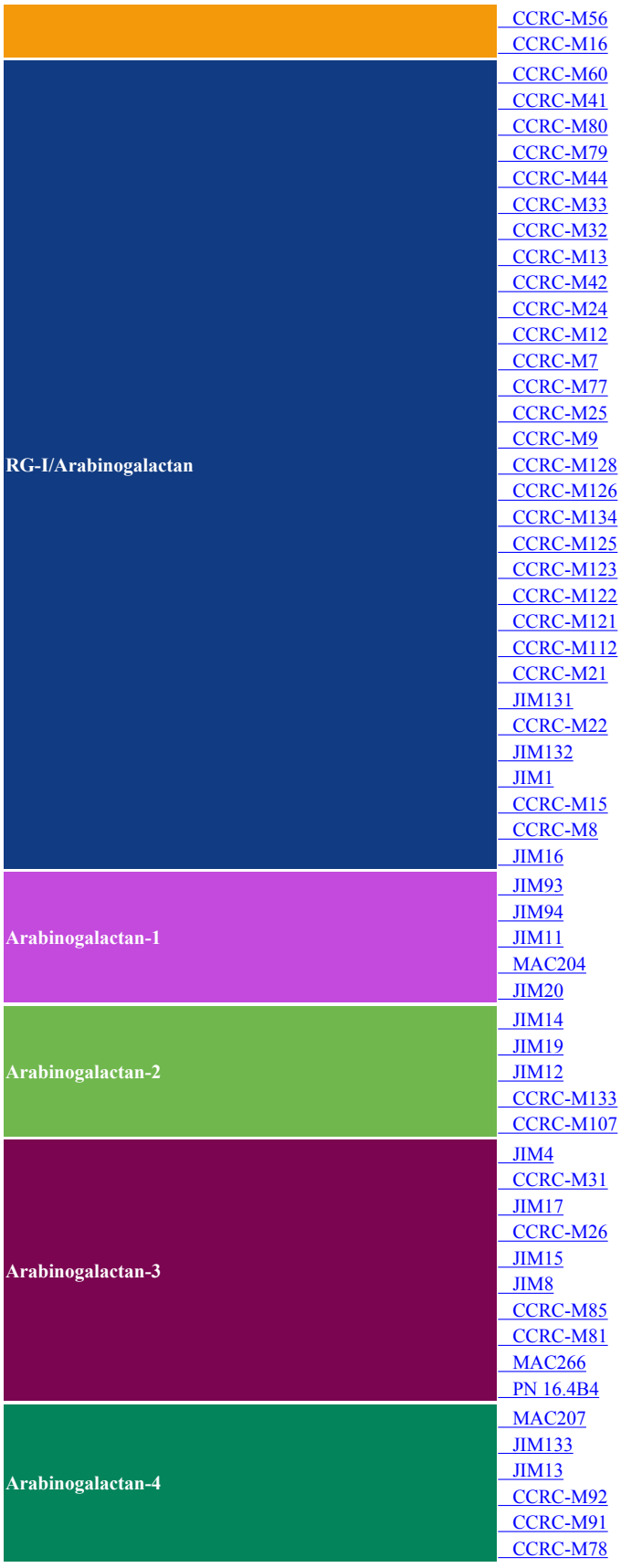
Figure S5. Optimized geometries of inter-unit lignin linkages.

Supplementary Table S1: List of cell wall glycan-directed monoclonal antibodies (mAbs) used for glycome profiling analyses.

The groupings of antibodies are based on a hierarchical clustering of ELISA data generated from a screen of all mAbs against a comprehensive panel of plant polysaccharide preparations (Pattathil et al., 2010; Pattathil et al., 2012) that clusters mAbs according to the predominant polysaccharides that they recognize. The majority of listings link to the Wall*Mab*DB plant cell wall monoclonal antibody database (http://www.wallmabdb.net) that provides detailed descriptions of each mAb, including immunogen, antibody isotype, epitope structure (to the extent known), supplier information, and related literature citations.

Glycan Group Recognized	mAb Names
Non-Fucosylated Xyloglucan-1	CCRC-M95 CCRC-M101
Non-Fucosylated Xyloglucan-2	CCRC-M104 CCRC-M89 CCRC-M93 CCRC-M87 CCRC-M88
Non-Fucosylated Xyloglucan-3	CCRC-M100 CCRC-M103
Non-Fucosylated Xyloglucan-4	CCRC-M58 CCRC-M86 CCRC-M55 CCRC-M52 CCRC-M99
Non-Fucosylated Xyloglucan-5	CCRC-M54 CCRC-M48 CCRC-M49 CCRC-M96 CCRC-M50 CCRC-M51 CCRC-M53
Non-Fucosylated Xyloglucan-6	CCRC-M57
Fucosylated Xyloglucan	CCRC-M102 CCRC-M39 CCRC-M106 CCRC-M84 CCRC-M1
Xylan-1/XG	CCRC-M111 CCRC-M108 CCRC-M109
Xylan-2	CCRC-M119 CCRC-M115 CCRC-M110 CCRC-M105
Xylan-3	CCRC-M117 CCRC-M113

	CCRC-M120 CCRC-M118 CCRC-M116 CCRC-M114
Xylan-4	CCRC-M154 CCRC-M150
Xylan-5	CCRC-M144 CCRC-M146 CCRC-M145 CCRC-M155
Xylan-6	CCRC-M153 CCRC-M151 CCRC-M148 CCRC-M140 CCRC-M139 CCRC-M138
Xylan-7	CCRC-M160 CCRC-M137 CCRC-M152 CCRC-M149
Galactomannan-1	CCRC-M75 CCRC-M70 CCRC-M74
Galactomannan-2	CCRC-M166 CCRC-M168 CCRC-M174 CCRC-M175
Acetylated Glucomannan	CCRC-M169 CCRC-M170
β-Glucan	LAMP BG1
HG Backbone-1	CCRC-M131 CCRC-M38 JIM5
HG Backbone-2	JIM136 JIM7
RG-I Backbone	CCRC-M69 CCRC-M35 CCRC-M36 CCRC-M14 CCRC-M129 CCRC-M72
Linseed Mucilage RG-I	JIM3 CCRC-M40 CCRC-M161 CCRC-M164
Physcomitrella Pectin	CCRC-M98 CCRC-M94
RG-Ia	CCRC-M5 CCRC-M2
RG-Ib	JIM137 JIM101 CCRC-M61 CCRC-M30
RG-Ic	CCRC-M23 CCRC-M17 CCRC-M19 CCRC-M18


CCRC-M56
CCRC-M16
RG-I/Arabinogalactan
CCRC-M60
CCRC-M41
CCRC-M80
CCRC-M79
CCRC-M44
CCRC-M33
CCRC-M32
CCRC-M13
CCRC-M42
CCRC-M24
CCRC-M12
CCRC-M7
CCRC-M77
CCRC-M25
CCRC-M9
CCRC-M128
CCRC-M126
CCRC-M134
CCRC-M125
CCRC-M123
CCRC-M122
CCRC-M121
CCRC-M112
CCRC-M21
JIM131
CCRC-M22
JIM132
JIM1
CCRC-M15
CCRC-M8
JIM16
Arabinogalactan-1
JIM93
JIM94
JIM11
MAC204
JIM20
Arabinogalactan-2
JIM14
JIM19
JIM12
CCRC-M133
CCRC-M107
Arabinogalactan-3
JIM4
CCRC-M31
JIM17
CCRC-M26
JIM15
JIM8
CCRC-M85
CCRC-M81
MAC266
PN 16.4B4
Arabinogalactan-4
MAC207
JIM133
JIM13
CCRC-M92
CCRC-M91
CCRC-M78

Miscellaneous	MAC265 CCRC-M97

Table S2. Assignments of the lignin ^{13}C-^{1}H correlation peaks in the 2D HSQC spectra of Arabidopsis samples

Region	Label	δ_C/δ_H (ppm)	Assignment
Aliphatic	A_α	71.8/4.83	C_α–H_α in β-O-4' substructures (**A**)
	$A_{\beta(G)}$	83.4/4.27	C_β–H_β in β-O-4' substructures (**A**) linked to a G unit
	$A_{\beta(S)}$	85.9/4.10	C_β–H_β in β-O-4' substructures linked (**A**) to a S unit
	B_α	86.8/5.43	C_α–H_α in β-5 phenylcoumaran substructures (**B**)
	B_β	53.1/3.43	C_β–H_β in β-5 phenylcoumaran substructures (**B**)
	C_α	84.8/4.65	C_α–H_α in β–β' resinol substructures (**C**)
	C_β	53.5/3.05	C_β–H_β in β–β' resinol substructures (**C**)
	C_γ	71.0/4.17	C_γ–H_γ in β–β' resinol substructures (**C**)
	D_α	83.3/4.81	C_α-H_α in dibenzodioxocin substructures (**D**)
	E_γ	61.3/4.08	C_γ–H_γ in cinnamyl alcohol end-groups (**E**) overlaps with carbohydrate signals
	MeO (–OCH_3)	55.6/3.73	C–H in methoxyls
Aromatic	$H_{2,6}$	127.8/7.22	$C_{2,6}$–$H_{2,6}$ in *p*-hydroxyphenyl units (**H**)
	G_2	110.9/6.99	C_2–H_2 in guaiacyl units (**G**)
	G_5/G_6	114.9/6.72 and 6.94	C_5–H_5 and C_6–H_6 in guaiacyl units (**G**)
	G_5	118.7/6.77	C_5–H_5 in guaiacyl units (**G**)
	$S_{2,6}$	103.8/6.69	C_2–H_2 and C_6–H_6 in etherified syringyl units (**S**)
	$pCA_{2,6}$	130.1/7.45	C_2–H_2 and C_6–H_6 in *p*-coumarate (***p*CA**)
	FA_2	110.9/7.33	C_2–H_2 in ferulate (**FA**)
	$T_{2',6'}$	103.3/7.19	$C_{2'}$–$H_{2'}$ in tricin (**T**)