ORGANIC CHEMISTRY

An "ideal lignin" facilitates full biomass utilization

Yanding Li[1,2], Li Shuai[1,3], Hoon Kim[1,4], Ali Hussain Motagamwala[1,5], Justin K. Mobley[1], Fengxia Yue[1,4], Yuki Tobimatsu[1,4,6], Daphna Havkin-Frenkel[7,8], Fang Chen[9,10], Richard A. Dixon[9,10], Jeremy S. Luterbacher[11], James A. Dumesic[1,5], John Ralph[1,2,4]*

Lignin, a major component of lignocellulosic biomass, is crucial to plant growth and development but is a major impediment to efficient biomass utilization in various processes. Valorizing lignin is increasingly realized as being essential. However, rapid condensation of lignin during acidic extraction leads to the formation of recalcitrant condensed units that, along with similar units and structural heterogeneity in native lignin, drastically limits product yield and selectivity. Catechyl lignin (C-lignin), which is essentially a benzodioxane homopolymer without condensed units, might represent an ideal lignin for valorization, as it circumvents these issues. We discovered that C-lignin is highly acid-resistant. Hydrogenolysis of C-lignin resulted in the cleavage of all benzodioxane structures to produce catechyl-type monomers in near-quantitative yield with a selectivity of 90% to a single monomer.

INTRODUCTION

Lignin is a polymeric material composed of phenylpropanoid subunits and is one of the largest sources of naturally produced aromatics on the planet. Because of its aromatic nature, lignin has a higher energy density than polysaccharide polymers, as well as a higher potential commercial value (*1*). However, because of lignin's complexity, its efficient utilization, either as a polymer or from its derivable small-molecule products, is currently problematic (*1*–*3*).

Although mild depolymerization methods, such as oxidative (*4*, *5*) and hydrogenolytic (*6*–*8*) procedures, have produced encouraging results in laboratory-scale experiments, their applicability in industrial processes has been limited. Direct hydrogenolysis, that is, the hydrogenation of unprocessed solid biomass by a heterogeneous metal catalyst, remains one of the most promising methods for cleaving lignin's ether bonds and producing aromatic monomers in high yields (*8*–*10*). However, hydrogenolysis still suffers from product complexity issues. In most wild-type biomass, the lignin polymer is composed of three phenylpropanoid subunits—*p*-hydroxyphenyl (**H**), guaiacyl (**G**), and syringyl (**S**)—derived by combinatorial radical coupling from the three main monolignols (*p*-coumaryl, coniferyl, and sinapyl alcohols). Although **H** units are typically at low-levels, this results in at least three different types of monomers (**H**, **G**, and **S**), each with a selection of side chains, as the primary hydrogenolysis products, which makes monomer separation and utilization difficult. Lignin's principal alkyl-aryl-ether units with their β–O–4 interunit bonds (45 to 85%) can be selectively cleaved, but other linkages including β–5 (1 to 12%), β–β (5 to 12%), 5–5 (1 to 9%), 4–O–5 (~2%), and β–1 (1 to 2%), which are also present in lignins, remain largely uncleaved (*8*); carbon-carbon (C–C) and diaryl ether (4–O–5) units typically result from dimeric or higher oligomeric products.

The use of extracted lignins rather than whole biomass has the advantage that the material can be fully dissolved in organic solvents, facilitating catalyst recovery and continuous processing. However, acidic industrial lignin fractionation is known to cause some β-ether cleavage and condensation between units via the electrophilic substitution of acid-generated benzylic carbocation intermediates on the electron-rich aromatic rings (*7*, *11*), limiting depolymerization yields (Scheme 1A) (*12*–*14*). There are some elegant solutions focusing on suppressing the condensation reaction, either using a capping agent (*7*, *15*) or using two-step strategies (*4*, *5*, *11*). However, extra chemicals or catalysts are needed to achieve this goal.

Bioengineered biomass could be used to achieve higher hydrogenolysis yields and simpler product mixtures. For example, the recent use of formaldehyde protection during lignin extraction from a high-**S** poplar lignin (*7*, *16*) that has up to 98% syringyl **S** units and ~90% β–O–4 linkages [from nuclear magnetic resonance (NMR) estimates] prevented condensation reactions and allowed an unprecedentedly high monomer yield (78%) under hydrogenolytic conditions (*7*). However, even in this high-**S** lignin, some 10% of the linkages are C–C bonds that do not cleave. The use of formaldehyde to protect the lignin from condensation reactions also resulted in some formaldehyde addition to the ring, complicating the hydrogenolysis products with methyl-substituted aromatics. Without formaldehyde, the lignin extracted under acidic conditions had significant condensation, thwarting the production of monomers and resulting in a hydrogenolysis monomer yield of only 26% (*7*). Although new methods for displacing formaldehyde for the protection from acid-catalyzed condensation reactions, retaining much of the yield (70%) and producing a simpler monomer mix, have recently been revealed (*17*), extra protection chemicals remain necessary during the lignin extraction.

RESULTS

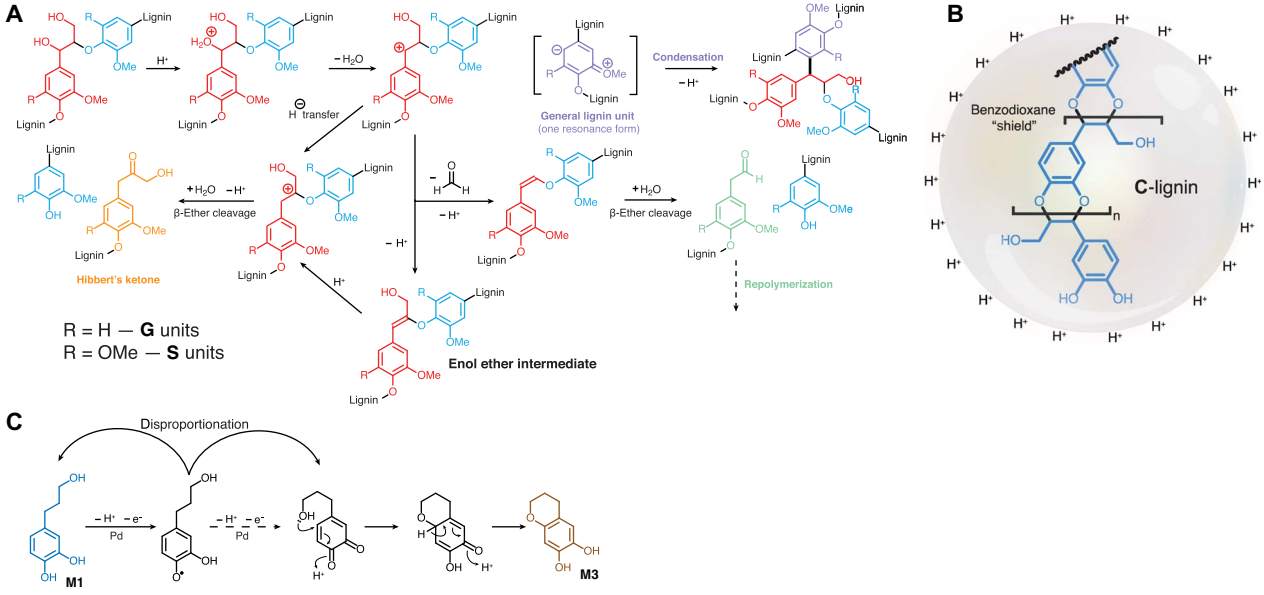
An "ideal lignin" archetype

On the basis of the plethora of information stemming from the lignin biosynthetic research community over the last decade, and with the revelations regarding lignins' structural malleability from studies on lignin pathway mutants and transgenics as well as on various "natural" plants discovered to have unusual lignins, researchers have been able to

[1]U.S. Department of Energy Great Lakes Bioenergy Research Center, and Wisconsin Energy Institute, University of Wisconsin–Madison, Madison, WI 53726, USA. [2]Department of Biological Systems Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. [3]Department of Sustainable Biomaterials, Virginia Tech, Blacksburg, VA 24061, USA. [4]Department of Biochemistry, University of Wisconsin–Madison, Madison, WI 53706, USA. [5]Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. [6]Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan. [7]Department of Plant Biology and Pathology, Rutgers, State University of New Jersey, New Brunswick, NJ 08901, USA. [8]Bakto Flavors LLC, 772 Cranbury Crossroad, North Brunswick, NJ 08092, USA. [9]BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, TX 76203, USA. [10]Center of Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [11]Laboratory of Sustainable and Catalytic Processing, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
*Corresponding author. Email: jralph@wisc.edu

contemplate designing lignins for improved utilization (*18*). It is now a realistic juncture to posit the characteristics for an ideal lignin archetype for biomass processing. For the depolymerization of the polymer to monomers, lignin should have at least the following three characteristics. First, if acidic pretreatment is used, then it should be stable under acidic conditions to prevent condensation and the generation of undesired new C–C bonds. Second, it should contain only ether (C–O) interunit linkages in its backbone so that it can be fully depolymerized. Finally, it should be generated *in planta* from a single phenylpropanoid monomer to allow the production of the simplest array of compounds.

We have reported the discovery of an unusual catechyl lignin (**C**-lignin) present in the seed coats of vanilla (*Vanilla planifolia*) (*19*) and various members of the Cactaceae of the genus *Melocactus* (*20*). In this special case, the lack of *O*-methyltransferase (OMT) activity for conversion from catechyl **C** to guaiacyl **G** and, subsequently, on to syringyl **S**, aromatic-level precursors, results in 100% **C** units in the cell wall (CW). This **C**-lignin was, somewhat surprisingly, found to be essentially a homopolymer synthesized almost purely by β–O–4 coupling of caffeyl alcohol with the growing polymer chain, producing benzodioxanes as the dominant unit in the polymer (Fig. 1A). If it has particular stability toward biomass pretreatment conditions, then this



Scheme 1. Mechanisms for lignin condensation, C-lignin structure, and monomer M3 formation. (**A**) Mechanism of lignin acidolysis and condensation routes. (**B**) The benzodioxane structure acts as a "shield" that can protect **C**-lignin from unwanted acidolysis and condensation reactions. (**C**) Proposed mechanism for the cyclization reaction of **M1** to **M3**.

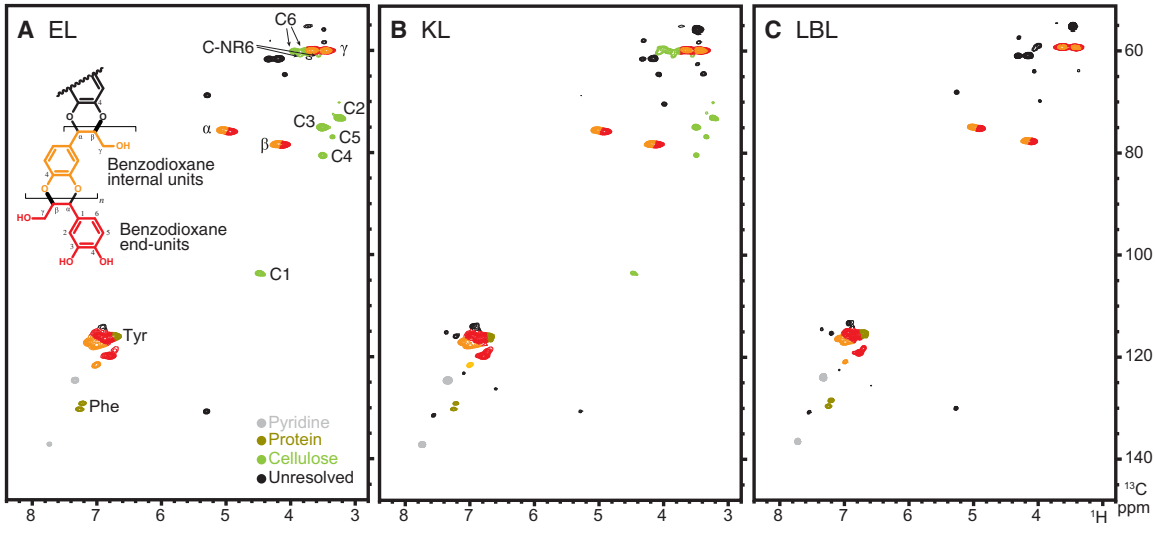


Fig. 1. NMR spectra. Partial 2D HSQC NMR spectra of (**A**) EL, (**B**) KL, and (**C**) LBL from vanilla (*V. planifolia*) seed coat. There are no obvious lignin structural changes after the acidic lignin extraction processes. Cellulose was labeled following the conventional monosaccharide nomenclature; NR is the nonreducing end of the cellulose. Protein residuals were labeled by the aromatic amino acid. Tyr, L-tyrosine; Phe, L-phenylalanine; ppm, parts per million.

C-lignin might therefore represent an example of such an ideal lignin that can, in principle, be depolymerized to a single product by hydrogenolysis. Furthermore, this substrate has the potential to produce valuable catechol monomers, whereas the large majority of monomers produced from lignin have been **S** or **G** derivatives (*1*, *2*). Expanding the arsenal of lignin-derived platform molecules could play an important role in the successful use of this fraction within future biorefineries. Here, we describe the ideal nature of this lignin via a revised compositional characterization of the vanilla seed coat fiber, new features of the **C**-lignin's reactivity and stability, and our successful attempts at converting it to monomers in near-quantitative yields.

Acid stability of C-lignin

Because of the lack of an accessible and eliminable benzylic hydroxyl group in **C**-lignin units (Scheme 1B), condensation reactions due to the formation of benzyl cations might be mitigated under acidic conditions. We therefore examined the acid stability of the polymer to determine whether acidolytic methods could be used to purify the lignin. Comparison of the two-dimensional (2D) heteronuclear single-quantum coherence (HSQC) NMR spectra from the enzyme lignin (EL) (derived by removing polysaccharides via crude cellulases treatment) (*21*) and Klason lignin (KL) showed no significant differences in the lignin structure (Fig. 1, A and B) (*22*). The **C**-lignin survives even the harshest of acidic pretreatment methods—the KL isolation procedure includes a 1-hour treatment in 72% (w/w) sulfuric acid, followed by dilution to 4% (w/w) sulfuric acid and autoclaving at 121°C for 1 hour—while retaining its original lignin structure. An efficient acidic lithium bromide (LiBr) pretreatment method was also used to purify the lignin. This treatment method is known for its quick and near-quantitative removal of the polysaccharides to give an LiBr lignin (LBL) (Fig. 1C) (*23*). The molecular weight of the LBL was shown to be similar to that of the EL (Fig. 2). The **C**-lignin polymer appeared to survive this pretreatment based on the retention of its key lignin structural features in its NMR spectra and little change in its molecular weight distribution. On the basis of these results, we can conclude that, unlike normal **S**-**G** lignins, polysaccharides can be removed via acid pretreatment from **C**-lignin without its suffering from unwanted condensation reactions. After removing the polysaccharides, the resulting lignins (EL, KL, and LBL) were completely soluble in various organic solvents [for example, acetone, dioxane, or tetrahydrofuran (THF)] mixed with water to match lignin solubility parameters (*24*). Efficient lignin solubilization should greatly facilitate continuous processing in an industrial setting.

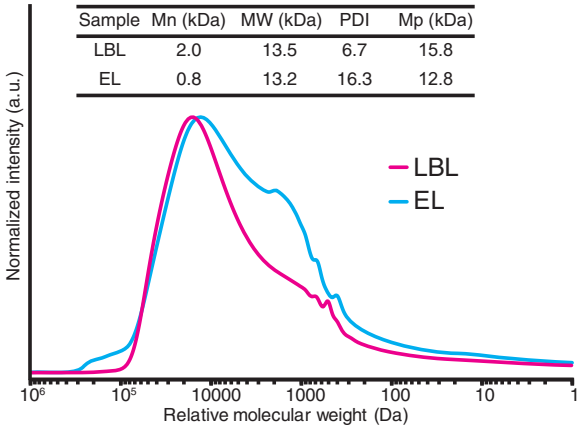


Fig. 2. Molecular weight profiles. Molecular weight profiles of EL (cyan) and LBL (magenta) from *V. planifolia* seed coat measured by gel-permeation chromatography (GPC). The *x* axis indicates the apparent molecular weight of individual lignin polymers and is shown as a log scale. The *y* axis shows the response of a UV-light detector (at 280 nm) normalized to the most abundant signal in each chromatogram. The most abundant signal in the each of the two samples corresponds to a molecular weight of ~13,000 Da (determined via polystyrene standards); comparison shows that there was no obvious lignin polymer degradation during the acid pretreatment. PDI is the polydispersity index. a.u., arbitrary units. MW, molecular weight.

Response of C-lignin to traditional degradative methods

To investigate the potential for **C**-lignin depolymerization, we applied two traditional lignin degradative analytical methods, alkaline nitrobenzene oxidation (NBO) and thioacidolysis, to a **C**-lignin model compound, the caffeyl alcohol dimer **D1** (**C**-dimer), and to the vanilla bean seed coat CW (Fig. 3 and fig. S3). Although relatively low yields of the corresponding monomeric products (30 to 60%) were obtained from the dimeric compound, the use of the CW gave monomeric products in extremely low yields (<1%). As discussed widely in the past, both thioacidolysis and alkaline oxidation need the involvement of a free benzylic hydroxyl group on the lignin side chain (*25*, *26*). It was therefore concluded that, because of the stability of the 1,4-benzodioxane structure, especially under the tested acidic and alkaline oxidative conditions, traditional lignin chemical degradation methods are ineffective for the depolymerization of **C**-lignin. A computational approach to evaluate the bond dissociation energy (BDE) of **C**-lignin using density functional theory models suggested that depolymerization of **C**-lignin is theoretically possible (*27*). Although the benzodioxane β–O–4 bond calculates to have a slightly higher BDE value than a conventional β–O–4 bond, it is still much lower than the BDEs of lignin's C–C bonds (*28*).

Catalytic hydrogenolysis of C-lignin

We reasoned that hydrogenolysis had the potential to more efficiently depolymerize **C**-lignin. We first sought efficient methods for cleaving dimeric model **D1**, rationalizing that, although the corollary is not necessarily true, any reaction conditions that did not produce high yields from **D1** would have little chance of being effective on the polymer. When hydrogenolysis was applied to the **C**-dimer **D1** and vanilla seed coat CW, analysis by gas chromatography with flame-ionization detection (GC-FID) showed that the products were rather simple with dominant products **M1** (catechylpropanol), **M2** (catechylpropane), and **M3** (chroman-6,7-diol), together with some minor products (Fig. 3). The major products, **M1** and **M2**, were identified by comparison with authentic synthetic standards. The initially puzzling minor product **M3**, which is a cyclization product from **M1**, was separated from the product mixture by silica-gel chromatography, characterized, and structurally identified by NMR and high-resolution mass spectrometry (MS). Because it was not obvious whether the chromane ring oxygen originated from the lignin γ-OH or from water, the hydrogenolysis reaction was run in ^{18}O-labeled $H_2{}^{18}O$. No ^{18}O was detected in the product **M3**, so the cyclization mechanism was concluded to involve the γ-OH via a radical disproportionation reaction (Scheme 1C) (*29*). This is the first report of this lignin hydrogenolysis product. The minor impurity peaks displayed in the chromatograms from the CW materials (fig. S3) were derived from the solvent, polysaccharide, and fatty acid products, which were identified via GC-MS.

Monomer production data under different conditions are shown in Fig. 4. Yields are normalized to the total molar concentration of caffeyl alcohol in **C**-lignin determined from quantitative ^{13}C NMR (table S2). Not surprisingly, the monomer distributions were heavily affected by

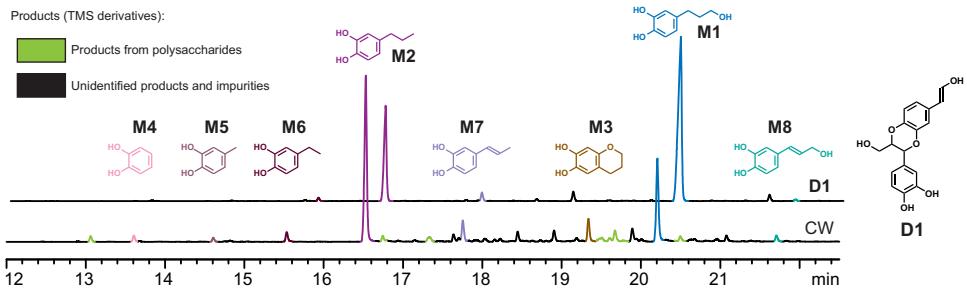


Fig. 3. GC-FID spectra of hydrogenolysis products from dimeric compound D1 and from CW. Hydrogenolysis condition: Pt/C, 200°C, 40-bar H_2, 15 hours. Coloring of peaks matches that of the structures for monomers **M1** to **M8**. Products from polysaccharide in the CW are colored light green, and unidentified products from other non-lignin compounds are left in black. TMS, trimethylsilyl. Note that the upper **D1** product chromatogram is offset by ~0.3 min.

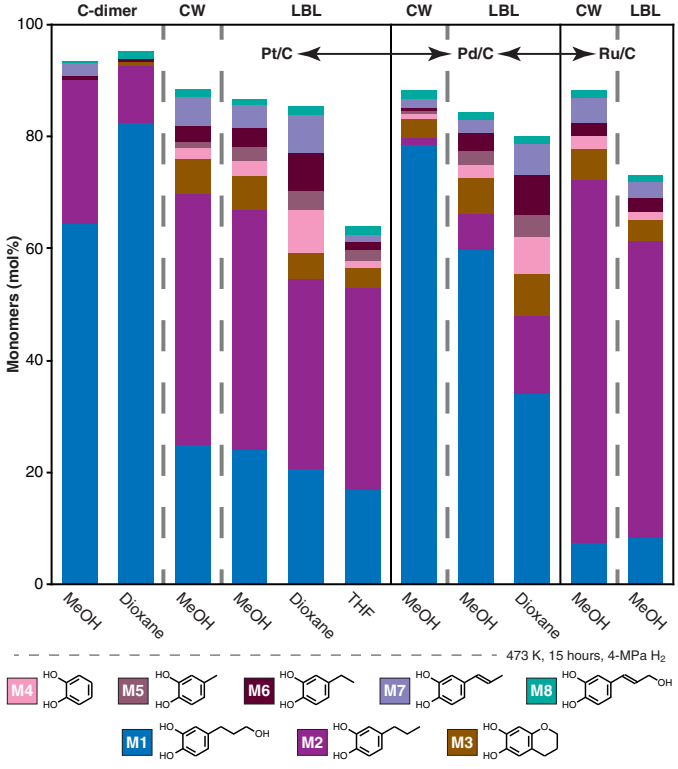


Fig. 4. Hydrogenolysis monomer yields from different catalyst and solvent combinations. Yields are on a **C**-lignin molar basis (see also table S3, from left to right: entries 1, 2, 3, 5, 7, 9, 10, 12, 14, 19, and 21).

the choice of catalyst and solvent (*17*, *30*, *31*). Here, we illustrate that Pt/C showed a slightly higher reactivity, whereas Pd/C and Ru/C showed a much better product selectivity. More side chain truncation products were obtained from **C**-LBL compared to that from vanilla seed coat CW, suggesting that a significant degree of side chain truncation occurred during the acid pretreatment stage or that the isolated lignin was more accessible to the catalyst. In terms of solvent effects, methanol produced a slightly higher monomer yield compared to dioxane, whereas THF gave a substantially lower yield. Both monomer yield and reaction selectivity were maximized using Pd/C or Ru/C as catalyst and methanol as the solvent. Retaining or losing the hydroxy group on the side chain can be controlled by simply changing the catalyst to satisfy the different intended purposes for using the catechyl monomers.

Thus, Pd/C produced the catechylpropanol monomer **M1** with 89% selectivity, whereas Ru/C produced the catechylpropane monomer **M2** with 74% selectivity. Increasing the hydrogenolysis reaction time from 3 to 15 hours (table S3, entries 16 to 19) led to an ~10% increase in lignin conversion and monomer yield. The resulting product oil mixture after vacuum drying was completely soluble in methanol, ethanol (EtOH), dioxane, pyridine, and other solvents but only partially soluble in acetone, ethyl acetate, and dichloromethane (DCM) due to the presence of products from degraded polysaccharides and other non-lignin components. The mass balance and total organic carbon (TOC) analyses (Table 1) indicated that volatile products were minimal or insignificant.

A 2D HSQC NMR spectrum of the total hydrogenolysis product, which was completely soluble in dimethyl sulfoxide (DMSO)/pyridine (4:1, v/v), demonstrated that the **C**-lignin had been completely depolymerized, that is, no detectable residual benzodioxane structures remained (fig. S4A). The major products were fully authenticated by comparison with synthetic compounds **M1** and **M2** and with authenticated isolated **M3**. No detectable products from side reactions or recondensation were detectable. The GPC molecular weight profile of the hydrogenolysis products mixture from **C**-LBL before and after the hydrogenolysis reactions showed a dominant monomer peak (fig. S4B). The high–molecular weight fractions were separated from monomer fractions, and the fractions were characterized by HSQC NMR (fig. S7). The data revealed that only traces of the original benzodioxane structures from the **C**-lignin remained in the product and that the high–molecular weight fractions contained only nonaromatic components present in the original sample and were therefore not from the lignin proper. It can therefore be safely concluded that essentially all of the **C**-lignin in the samples was depolymerized to monomeric compounds during hydrogenolysis. The non-lignin components in the lignin stream were nonextractable oils, waxes, or the other (difficult to remove) components in the sample that are not necessarily associated directly with the phenylpropanoid polymer.

DISCUSSION

Prospects for C-lignin and its derived catechylpropanoid monomers

Catechols in nature are remarkably biochemically active; because of the interaction of the vicinal phenolic hydroxyl groups, catechols play a vital role in both biomedical and biomimetic functional materials (*32*). Their synthesis is challenging because of the difficulty of transforming phenols

Table 1. Mass balance and TOC on hydrogenolysis of C-LBL and its resulting product oil.

Feed	CW	Dissolved C-LBL
Solid recovery*	55–74%	~100%
Oil recovery†	23–35%	50–60%
TOC of **C**-LBL	—	62.66 ± 0.23%
TOC of product oil	—	61.44 ± 0.34%

*Solid includes recovered CW material and catalyst. †Oil yield on a CW and **C**-LBL mass basis.

to catechols; although researchers have recently developed several catechol synthetic methods (*33*), applying those methods at scale remains complicated. There is little reference to high-yielding biomass conversion to catechols, although catechols were reported as hydrogenolysis products from organosolv lignin of candlenut shells (using Cu-doped porous metal oxides) in which the cleaving of the aromatic methoxyl groups during the reaction was claimed (*34*). Catechols act as important intermediates for the conversion of lignin-derived monomers to value-added platform chemicals via the bacterial β-ketoadipate pathway (*35*). We therefore contend that it would be beneficial, and more energy-efficient for aromatic metabolism/catabolism, if high yields of catechols could be obtained directly from lignin.

Our study provides a new perspective for the production of catechols from a renewable biomass source rather than petroleum. Compounds **M1** and **M2** are currently not available in bulk, so their commercial value is not obvious. However, an enriched diversity of the raw materials from the catechol family would likely provide significant value. A LiBr pretreatment method is able to convert a fraction of β–O–4 units in **S**-**G**–type lignins into benzodioxanes (*36*). Large amounts of catechols are potentially producible if we could produce benzodioxane-type lignins in energy crops. We do not yet know if genetically engineered plants, including plantation trees such as pines and poplars, in which the production of lignin is on a large scale, will tolerate **C**-lignins in stem tissues; **C**-polymers have been evidenced in a gymnosperm tracheary element system, in which OMT activity was down-regulated (*37*), but have not yet been found in OMT–down-regulated dicots, and we suspect that additional activities will need to be suppressed for the synthesis and deposition of the **C**-lignin polymer. Given the unique acid-resistant property of **C**-lignin, the potential value of the monomeric products, the homogeneous nature of **C**-lignins that is already known to aid lignin fiber production (*38*), and the high conversion to catechol monomers by hydrogenolysis reported here, we suggest that continuing to pursue the means to produce **C**-lignins in planta is decidedly worthwhile.

C-lignin therefore has numerous compelling features for a biorefinery operation aimed at delivering value from its lignin component. It maintains its native structure after treatment under even strongly acidic conditions; acid pretreatment can therefore be applied to vanilla seed coats to recover the polysaccharide while retaining the native **C**-lignin structure. After sufficient pulverization followed by acid pretreatment, **C**-lignin could be dissolved in organic solvents, enabling both detailed NMR analysis and continuous processing schemes. **C**-lignin can be completely depolymerized by a hydrogenolytic method to produce simple monomeric catechols near-quantitatively and, by selecting the catalyst, with a single monomer accounting for 90% of the monomer product. The yield and selectivity for a single monomer are higher than for any other lignin or biomass to date (fig. S6). There is therefore considerable potential for economic hydrogenolysis of **C**-lignin–rich waste biomass resources only now being structurally characterized, such as Jatropha (*Jatropha curcas*) seed coats (*39*) and candlenut (*Aleurites moluccanus*) shells (*40*), and via genetic engineering if high levels of **C**-lignin could be expressed in traditional biomass sources. Such an approach toward significantly valorizing lignins and biomass in biorefining processes would aid process economics.

MATERIALS AND METHODS

C-lignin sample pretreatment

Processing of seed coat material

Vanilla seed and pod were received as a mixture from a natural vanilla processing plant (Bakto Flavors LLC). The mixture was sifted, and the lower-density remaining pod powder was blown away using a heat gun (set on cold). Preparation of vanilla seed coat NMR samples was via methods described previously (*22*). Briefly, isolated vanilla seed coats (4 × 300 mg) were ball-milled (30 × 10 min, 5-min cooling cycle) using a Retsch PM100 ball mill vibrating at 600 rpm with ZrO_2 vessels containing ZrO_2 ball bearings. Preground seed coat was extracted using a modified Bligh and Dyer extraction (*41*) to remove oils and extractives.

Modified Bligh and Dyer extraction

Vanilla seed material (100 g in total) was shaker-milled (MM400, Retsch) at 3600 rpm for 5 min using a 50-ml stainless steel jar and a single 20-mm ball bearing. The milled sample was transferred to a 1-liter volumetric flask, and a magnetic stir bar was added. Deionized (DI) water (80 ml), chloroform (100 ml), and methanol (200 ml) were added, and the mixture was stirred at 50°C for 30 min. To the mixture was then added 100 ml more of chloroform, and then, after another 30 min, 100 ml of DI water was added. The stirring was continued at 50°C for 24 hours, and the insoluble material was removed by centrifugation (3800 rpm for 15 min), retaining the solids by decanting off the solvent and keeping the filtrate as well. The residue was extracted again by the same method. The filtrates were combined, and the solvents were removed by rotary evaporation to produce the extractives fraction for analysis.

EL from vanilla seed coat

The ball-milled extract-free vanilla seed coat material (1 g) was placed in centrifuge tubes and digested at 35°C with crude cellulases [CELLULYSIN cellulases, *Trichoderma viride*; sample (50 mg/g) in acetate buffer (pH 5.0); two times over 3 days; fresh buffer and enzyme were added each time; catalog no. D00074989, Calbiochem], leaving all of the phenolic polymers and residual polysaccharides totaling 859 mg (85.9%) (table S1).

Acidic LiBr pretreatment of C-lignin from vanilla seed coat

C-LBL was prepared using the acidic LiBr trihydrate method described previously (*23*). Briefly, ball-milled extract-free vanilla seed coat material (1 g) was added into a 40-ml glass vial with a polytetrafluoroethylene (PTFE) lined cap, together with 4.50 ml of acidic 60 weight % (wt %) LiBr solution containing 0.04 M HCl. The vial was immersed into an oil bath preheated at 110°C under magnetic stirring. The mixture was filtered under vacuum and washed with water. The residues were dried at 40°C under reduced pressure (yield, 72.4%; table S1).

Compositional analysis

KL analysis was performed by the two-stage sulfuric acid hydrolysis following the National Renewable Energy Laboratory's standard protocol

(*42*). Briefly, 0.3 g of biomass (weighed to the nearest 0.1 mg) was treated in 72% (w/w) H_2SO_4 at room temperature for 60 min. The slurry was diluted to 4% (w/w) H_2SO_4 and autoclaved at 121°C for 60 min. After filtration, the acid-insoluble lignin (AIL = KL) and the acid-soluble lignin were quantitated gravimetrically and spectrophotometrically, respectively (table S1). Monosaccharides in the KL filtrates (hydrolysates) were quantitated using high-performance ion-chromatography on a Dionex ICS-3000 system equipped with an integrated amperometric detector and a CarboPac PA1 column (4 × 250 mm) at 30°C. DI water was used as an eluent at a flow rate of 0.7 ml/min according to the following gradient: 0 to 25 min, 100% water; 25.1 to 35 min, 30% water and 70% 0.1 M NaOH; and 35.1 to 42 min, 100% water. The post-run eluent of 0.5 M NaOH at a flow rate of 0.3 ml/min was used to purge remaining materials from the column to ensure baseline stability and detector sensitivity (*23*). Crude protein content was determined from the nitrogen (N) content using a 6.25 N-to-protein factor (table S1). The total N was determined using an elemental combustion system (model 4010, Costech Analytical Technologies). Samples (approximately 10 mg) were accurately weighed into tin combustion cups using a microbalance. After complete combustion, total N was measured as N_2 gas. The compositional analysis results are shown in table S1.

C-lignin characterization and quantification

Lignin characterization by 2D NMR spectroscopy

NMR spectra were acquired on a Bruker Biospin AVANCE III 700 MHz spectrometer fitted with a cryogenically cooled 5-mm QCI $^{1}H/^{31}P/^{13}C/^{15}N$ gradient probe with inverse geometry (proton coils closest to the sample), and spectral processing used Bruker's TopSpin 3.5pl6 (Mac) software. For NMR experiments, ball-milled whole vanilla seed coat material was swelled in DMSO-d_6/pyridine-d_5, isolated lignins and **C**–DHP (dehydrogenation polymer) were dissolved in 4:1 v/v DMSO-d_6/pyridine-d_5, and model compounds were dissolved in acetone-d_6. The central solvent peaks were used as the internal references (δ_C/δ_H: DMSO, 39.5/2.49; acetone, 29.84/2.05 ppm). Standard Bruker implementations of the traditional suite of 1D and 2D [gradient-selected and ^{1}H-detected; for example, correlation spectroscopy (COSY), ^{1}H–^{13}C HSQC (Fig. 1), and heteronuclear multiple-bond correlation (HMBC)] NMR experiments were used for structural elucidation and assignment authentication for monomers and dimers. Adiabatic 2D HSQC ("hsqcetgpsisp2.2") experiments for ball-milled seed coat material in a gel state were carried out using the parameters described previously (*22*). Processing used typical matched Gaussian apodization in F2 (LB = −0.5; GB = 0.001) and squared cosine-bell apodization in F1.

The characterization of the vanilla seed coat **C**-lignin was initially consistent with the previous report (*19*) but belied some issues. For both the CW and its EL (derived by removing polysaccharides via crude cellulases treatment) (*21*), characterization revealed each lignin to be an almost 100% benzodioxane polymer with only a trace level of the resinol (β–β) structure. Although not as high as we previously reported (~80%) (*19*), the seed coat sample had a very high KL value (~65%). However, the 2D HSQC NMR of the so-purified lignins contained many peaks in the aliphatic region that were not from the lignin itself (fig. S1). An alternative method (below) was therefore required for lignin quantification in these materials.

C-lignin quantification by ^{13}C NMR

Samples for quantitative ^{13}C NMR analysis were prepared by accurately weighing predried **C**-LBL samples (100 mg) dissolved in 1-ml internal standard [1,3,5-trioxane, DMSO-d_6 (3.12 mg/ml)] solution. The **C**-LBL concentration was also 100.0 mg/ml. Relaxation reagent chromium(III) acetylacetonate [$Cr(acac)_3$; ~2 mg] was added to the samples to facilitate the relaxation of the magnetization. Quantitative ^{13}C NMR spectroscopy was performed as previously described (*43*). The NMR spectra were acquired on the 700-MHz spectrometer described above. Relaxation delays were set to be ~5 times the longest T1 values of carbon signals (for inverse-gated proton decoupled ^{13}C NMR spectra); in our case, d1 = 12.5 s was used to fully relax of all of the carbons with the aid of the relaxation reagent. For the inverse-gated proton-decoupled ^{13}C spectrum, at least 38 hours (10K scans) were required. Spectral processing used both Bruker's TopSpin 3.5pl6 (Mac) and MestreNova 11.0 (Mac) software. The acquired FIDs were processed typically with a 5-Hz line broadening. The central solvent peaks were used as the internal references (δ_C/δ_H: DMSO, 39.5/2.49 ppm). Baseline was corrected manually over the 50- to 100-ppm region using TopSpin.

^{13}C NMR is mostly used to quantify low–molecular weight technical lignins (such as kraft lignin and organosolv lignin) or milled wood lignins (*43*, *44*). It is difficult to quantify native lignin with ^{13}C NMR for two reasons. One is the poor solubility of lignin, and the other is the overlapping peaks from the lignin side chain with polysaccharide peaks. However, **C**-LBL is a perfect sample for ^{13}C NMR analysis. First, the lignin structure is simple; there is only one type of structure in the lignin backbone—the benzodioxane derived from β–O–4-coupling. The chemical shifts of the benzodioxane carbons are unique (75 to 80 ppm), so that there is little chance of signal overlap with other components. Second, **C**-lignin is acid-resistant. Unlike the **S**-**G**–type lignins, harsh acid pretreatment can be applied to **C**-lignin without destroying the benzodioxane structure. Thus, we can easily remove the polysaccharides by acid pretreatment, further minimizing the signal overlap problem. According to the 2D HSQC spectrum of **C**-lignin (fig. S1), Cα and Cβ have the potential to allow ^{13}C NMR quantification of the phenylpropanoid unit derived from caffeyl alcohol in the **C**-lignin (fig. S2). Cγ cannot be used for the quantification because of the signal overlap with the unknown peaks (δ_H, 4.00 to 4.35 ppm; δ_C, 60.0 to 62.5 ppm). The aromatic region of **C**-lignin cannot be used for the quantification because of the overlap with signals from protein residues (tyrosine and phenylalanine) (*45*). Cα and Cβ may seem equally good for the **C**-lignin quantification; however, when looking at the HSQC spectrum at a lower contour level, peaks from polysaccharide residues cannot be completely ignored even after the acidic LiBr pretreatment; the residual C_3 and C_5 of the cellulose overlap with the Cβ of the **C**-lignin. Because the relaxation reagent $Cr(acac)_3$ was added to reduce the experiment time, the line broadening caused by the relaxation reagent made the overlap between Cβ and the cellulose residues even worse. As a result, Cα was chosen for the quantification as it had minimal peak overlap issues. Assuming that **C**-lignin is derived from pure caffeyl alcohol, the detailed calculation was as shown below (table S2)

$$c_{C\beta} = \frac{c_{IS} \times 3}{A_{IS}} \times A_{C\beta}$$

$$Y_{CA} = \frac{c_{C\beta}}{\rho_{LBL}}$$

$$W_{lignin(LBL)} = Y_{CA} \times Mw_{CA} \times 100\%$$

$$W_{lignin(CW)} = \frac{W_{lignin(LBL)}}{LBL\%}$$

In the equations, c_{IS} (mmol/ml) is the molar concentration of internal standard (IS; 1,3,5-trioxane), A_{IS} is the peak integral of internal standard in the quantitative ^{13}C NMR spectrum, $c_{C\beta}$ (mmol/ml) is the molar concentration of caffeyl alcohol unit in the **C**-lignin polymer, $A_{C\beta}$ is the peak integral of Cβ in the quantitative ^{13}C NMR spectrum, ρ_{LBL} (mg/ml) is the

mass concentration of **C**-LBL sample, Y_{CA} (mmol/mg) is the mole amount of caffeyl alcohol (CA) per milligram of **C**-LBL, Mw_{CA} (mg/mmol) is the molecular weight of caffeyl alcohol, $W_{lignin(LBL)}$ is the weight percentage of **C**-lignin in **C**-LBL, LBL% is the weight percentage of **C**-LBL obtained from whole CW, and $W_{lignin(CW)}$ is the weight percentage of **C**-lignin in whole CW.

Lignin depolymerization methods

Alkaline NBO

NBO was performed as previously described (*46*). Dimeric model compound **D1** (5 mg) or extracted vanilla seed coat (40 mg) was mixed with nitrobenzene (0.4 ml) and 2 M NaOH (7 ml) in a 10-ml stainless steel reactor vessel (Taiatsu Techno Co.) and heated at 170°C for 2 hours in an oil bath. The reactor was then cooled in ice water, and 1 ml of freshly prepared ethyl vanillin (3-ethoxy-4-hydroxybenzaldehyde**;** 5 mg/ml) in 0.1 M NaOH solution was added to the reaction mixture as an internal standard. The mixture was transferred to a 100-ml separatory funnel and washed three times with 15 ml of DCM. The remaining aqueous layer was acidified with 2 M HCl until the pH was below 3.0 and extracted with 2 × 20 ml of DCM and 20 ml of diethyl ether. The combined organic layers were washed with DI water (20 ml) and dried over $MgSO_4$. After filtration, the filtrate was collected in a 100-ml pear-shaped flask and dried under reduced pressure. For the TMS derivatization step, NBO products were transferred with pyridine (3 × 200 μl) into a GC vial, and *N*,*O*-bis(trimethylsilyl)trifluoroacetamide [BSTFA; 100 μl] was added. The mixture was heated to 50°C for 30 min. The silylated NBO products were analyzed by GC-MS and quantified by GC-FID using calibration curves (fig. S3, A and B).

Thioacidolysis followed by Raney nickel desulfurization

Thioacidolysis was performed as previously described (*47*). The thioacidolysis reagent was prepared freshly by adding 2.5 ml of EtSH and 0.7 ml of BF_3 etherate to a 25-ml volumetric flask containing 20 ml of distilled 1,4-dioxane and then complemented with dioxane to exactly 25 ml. Freshly made thioacidolysis reagent (4.0 ml) was added to a 5-ml screw-cap reaction vial containing extractive-free CW (40 mg) or model compound (15 mg) and a magnetic stir bar. The vial cap was screwed on tightly, and the vial was kept in an oil bath containing a heating block at 100°C for 4 hours with stirring. After the reaction, the vial was cooled in an ice-water bath for 2 min. A solution of 4,4′-ethylidenebisphenol in dioxane was prepared and used as an internal standard. The product mixture was transferred to a separatory funnel and 10 ml of saturated $NaHCO_3$ solution, along with internal standard solution, was added. Then, 5 ml of 1 M HCl solution was added to adjust the pH of the solution to below 3. The aqueous layer was extracted three times with 20 ml of DCM, and the combined organic phase was washed with saturated NH_4Cl, dried over anhydrous $MgSO_4$, and evaporated under reduced pressure at 40°C. The resulting products were desulfurized via Raney nickel. Briefly, the thioacidolysis products were dissolved in 3 ml of distilled dioxane with 1 ml of Raney nickel 3202 (Sigma-Aldrich) slurry. The mixture was heated at 80°C for 2 hours. After the reaction, nickel powder was removed using a magnet, and the reaction mixture was transferred quantitatively with DCM into a separatory funnel charged with 10 ml of NH_4Cl and 10 ml of DCM. Then, 5 ml of 1 M HCl solution was added to adjust the pH of the solution to below 3. The aqueous layer was extracted twice with 10 ml of DCM, and the combined organic phase was washed with brine, dried over anhydrous $MgSO_4$, and evaporated under reduced pressure at 40°C. For the TMS derivatization step, products were transferred with pyridine (3 × 200 μl) into a GC vial, and BSTFA (100 μl) was added. The mixture was heated to 50°C for 30 min. The silylated thioacidolysis products were analyzed by GC-MS and quantified by GC-FID using calibration curves (fig. S3, C and D).

Hydrogenolysis

Hydrogenolysis was performed as previously described (*7*). In cases in which isolated **C**-LBL was used as a feedstock, 200 mg of **C**-LBL was dissolved in 30 ml of methanol or dioxane/water (9:1, v/v) or THF/water (96:4, v/v) in a 100-ml high-pressure Parr reactor along with 100 mg of catalyst (5 wt % Pt/C, Pd/C, or Ru/C). The reactor was stirred with a mechanical propeller and heated via a high-temperature heating jacket. Once closed, the reactor was purged three times and then pressurized with H_2 (40 bar, 4 MPa). The reactor was heated to the desired temperature and then held at that temperature for the specified residence time. After the reaction was completed, the reactor was cooled in a water bath to room temperature. The resulting liquid was filtered through a nylon membrane filter (0.8 μm, 47 mm; Whatman) and washed with EtOH. The solvent was removed under reduced pressure at 40°C with a rotary evaporator. The crude products were dissolved in EtOH and made up to 10 ml in a volumetric flask. A 1 ml of aliquot was transferred into three 5-ml vials and then dried under reduced pressure. The dried samples were used for GC, GPC, and NMR analyses. For GC sample preparation, the sample was dissolved in 0.9 ml of pyridine and 0.1 ml of BSTFA, incubated at 50°C for 30 min, and then subjected to GC-FID and GC-MS. For NMR sample preparation (fig. S4A), the sample was dissolved in 0.6 ml of DMSO-d_6/pyridine-d_5 (4:1, v/v) and then transferred to a 5-mm NMR tube for NMR. For GPC sample preparation (fig. S4B), the sample was dissolved in 1 ml of dimethylformamide (DMF) containing 0.1 M LiBr.

For the cases in which hydrogenolysis was performed directly on the CW material, 200 mg of preextracted vanilla seed coat was mixed with 30 ml of methanol and 100 mg of the catalyst (5 wt % Pt/C, Pd/C, or Ru/C). The remaining procedure was performed as described above.

For the cases in which the lignin model compound was used as the feedstock, a solution of 50 mg of dimer **D1** in 30 ml of methanol or dioxane/water (9:1, v/v) was mixed with 50 mg of the catalyst (5 wt % Pt/C). The remaining procedure was performed as described above.

Analytical methods

GC-MS qualitative analysis of low–molecular weight products

Samples were dissolved in pyridine, and BSTFA was added for TMS derivatization. The mixture was heated to 50°C for 30 min. An aliquot of the sample (1 μl) was injected by an autosampler into a GC-MS (GC2010/PARVUM2, IC-1 column, Shimadzu Co.) equipped with a fused silica capillary column (30-m × 0.25-μm film; SHR5XLB capillary column, Shimadzu Co.) operating in split mode (split ratio of 20:1) to identify the products. The products were identified by comparison with the peak retention times and mass spectra of the authentic compounds and (or) by comparing with entries in the National Institute of Standards and Technology mass spectral library (fig. S5).

GC-FID quantitative analysis of low–molecular weight products

The identified major products were quantified by GC-FID (GC-2014, Shimadzu Co.) using calibration curves derived from authentic synthetic compounds (table S3). The yields of major hydrogenolysis products catechylpropanol **M1** and catechylpropane **M2** were quantified by using the calibration curves generated from their authentic synthetic standards. The yields of minor products without a primary hydroxy group [chroman-6,7-diol **M3**, catechol **M4**, 4-methylcatechol **M5**, 4-ethylcatechol **M6**, and 4-(1-propenyl)catechol **M7**] were calculated by the effective carbon number (ECN) method based on the yield of catechylpropane **M2**,

whereas the minor product with a primary hydroxy group (caffeyl alcohol **M8**) was calculated on the basis of the yield of catechylpropanol **M1**. The theoretical ECN of TMS-derivatized catechol **M4** (10.0), 4-methylcatechol **M5** (11.0), 4-ethylcatechol **M6** (12.0), catechylpropane **M2** (13.0), 4-(1-propenyl)catechol **M7** (12.9), chroman-6,7-diol **M3** (12.0), catechylpropanol **M1** (15.5), and caffeyl alcohol **M8** (15.4) was used for the calculation. The ECN contribution of aliphatic carbon 1.0, aromatic carbon 1.0, olefinic carbon 0.95, primary alcohol −0.5, and TMS 3.0 was used as described (*7*, *17*, *48*). The detailed calculation was as follows

$$n_{\text{monomer}} = \frac{A_{\text{monomer}}}{A_{\text{M1 or M2}}} \times n_{\text{M1 or M2}} \times \frac{\text{ECN}_{\text{M1 or M2}}}{\text{ECN}_{\text{monomer}}}$$

$$n_{\text{CA}} = Y_{\text{CA}} \times m_{\text{LBL}}$$

$$Y_{\text{monomer}} = \frac{n_{\text{monomer}}}{n_{\text{CA}}} \times 100\%$$

In the equations, n_{monomer} (mmol) is the molar amount of monomer in each analyzed sample, A_{monomer} is the peak area of monomer in the GC-FID chromatogram, $n_{\text{M1 or M2}}$ (mmol) is the molar amount of **M1** or **M2** in each analyzed sample based on its calibration curve, $A_{\text{M1 or M2}}$ is the peak area of **M1** or **M2** in the GC-FID chromatogram, $\text{ECN}_{\text{monomer}}$ is the effective carbon number of monomer, $\text{ECN}_{\text{M1 or M2}}$ is the effective carbon number of **M1** or **M2**, n_{CA} (mmol) is the molar amount of caffeyl alcohol in the feedstock, Y_{CA} (mmol/mg) is the mole amount of caffeyl alcohol per milligram of **C**-LBL from the quantitative ^{13}C NMR analysis (table S2), m_{LBL} (in milligrams) is the weight of **C**-LBL in the feedstock, and Y_{monomer} is the yield of monomer based on the molar amount of caffeyl alcohol in the feedstock.

Analytical GPC

Molecular weight distributions of lignins were determined by GPC using a Shimadzu LC20-AD LC pump equipped with a Shimadzu SPD-M20A UV-vis detector set at 280 nm and a Polymer Standard Services GPC column and guard column [PSS PolarSil analytical Linear S, 8-mm inner diameter (ID) × 5 cm and 5-μm particle size → PSS PolarSil analytical Linear S, 8-mm ID × 30 cm and 5-μm particle size]. The samples and column compartment were held at 40°C during analysis. The mobile phase was DMF with 0.1 M LiBr, and the flow rate was 1 ml/min. Molecular weight distributions were determined using Wyatt ASTRA 7 software (Wyatt Technology Corporation) via a conventional calibration curve using a ReadyCal polystyrene kit from Sigma-Aldrich [catalog no. 76552, M(p) 250-70000].

GPC fractionation of hydrogenolysis product mixtures

Using LBL as a hydrogenolysis feedstock and dioxane/water as the solvent, the product mixture was dried in vacuo, redissolved in pure dioxane with sonication, filtered through a PTFE membrane (0.2 μm), and then subjected to GPC. The GPC conditions here were slightly different from those in the analytical GPC method. For the fractionation, dioxane was used as the mobile phase instead of 0.1 M DMF/LiBr solution at a slower flow rate (0.3 ml/min) to achieve better fractionation. Four fractions were separated and collected (fig. S7A). The ultraviolet (UV) absorption contour map showed that different molecular weight fractions had completely different UV absorption properties. Because of peak overlap, each fraction was characterized by using its 2D HSQC NMR spectra and subtracting the overlapped fractions' spectra (fig. S7, B to E; note that f2 was characterized by subtracting f1 and f3 from f2, f3 was characterized by subtracting f2 and f4 from f3, and f4 was characterized by subtracting f3 from f4]. As seen in the NMR spectra, peaks from some nonaromatic components appear in all fractions (f1 to f4). The molecular weight of these nonaromatic components cannot be measured accurately because of the low GPC resolution and peak tailing, and/or the possibility that these nonaromatic components have a wide molecular weight distribution. Fractions f1 and f2 were almost identical to each other and contained only traces of aromatic peaks. The highest molecular weight component(s) in the product mixture was therefore not from lignin but from other components in the seeds. Fraction f3 contained the major hydrogenolysis products **M1** and **M3**. This fraction exhibited the strongest UV absorption in the UV contour map, which means that it was the dominant aromatic-containing mixture in the product. Fraction f4 was the other major hydrogenolysis product **M2**, which has a slightly lower molecular weight compared with **M1** and **M3**. It is inferred that there was a large amount of high–molecular weight products (the products in f1 and f2), which are distributed from f1 to f4 because of peak overlap. As these products lack aromatic rings, they are not from the caffeyl alcohol–derived phenylpropanoid polymer. Thus, they must be produced from other components existing in the seed, such as waxes, fatty acids, etc. These observations support our conclusion that the lignin content of vanilla seed coats is not determined accurately by KL and other traditional lignin analytical methods because of these nonextractable, nonaromatic components.

TOC analysis

A TOC analyzer (TOC-VCPH, Shimadzu Co.) with a solid sample module (SSM-5000A, Shimadzu Co.) was used to determine the total carbon content of the vanilla seed coat material and its hydrogenolysis products, its fractions, and the nonvolatile products. The hydrogenolysis products were dried at 50°C for 30 min to remove EtOH and then dried at 50°C in a vacuum oven for 30 min to completely remove water and EtOH. The dried solid samples (20.00 ± 1.00 mg) and hydrogenolysis products were measured as solids.

Using LBL as a hydrogenolysis feedstock and dioxane/water as the solvent, there was no significant change in the carbon content before (62.66 ± 0.23 wt %) and after (61.64 ± 0.34 wt %) the reaction (±SD, n = 2). Solvent degradation products (for example, ethylene glycol, diethylene glycol, etc.) were detected in the product mixture and identified by GC-MS when dioxane was used as solvent. It is still possible that some components in the **C**-LBL can either become volatile or attach to the catalyst. However, considering that the volatile products (for example, methane, ethane, and hexane) have much higher carbon contents (~75 to 85 wt %) compared with the solvent degradation products (~35 to 45 wt %), the loss of volatile products while introducing solvent degradation products should cause a significant decrease of carbon content. In our experiment, we did not observe any carbon content decrease nor did we observe any weight increase of the catalyst. This result suggested that the loss of volatile products during work-up and the effect of the solvent degradation products were negligible and also implied that most of the carbon-containing compounds were retained in the product mixture.

Synthetic model compounds and compound authentication

Synthetic methods are fully described in the Supplementary Materials.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/9/eaau2968/DC1

Synthetic model compounds and compound authentication.

Calibration curves and NMR spectra.

Fig. S1. 2D HSQC NMR.

REFERENCES AND NOTES

1. R. Rinaldi, R. Jastrzebshi, M. T. Clough, J. Ralph, M. Kennema, P. C. A. Bruijnincx, B. M. Weckhuysen, Paving the way for lignin valorisation: Recent advances in bioengineering, biorefining and catalysis. *Angew. Chem. Int. Ed. Engl.* **55**, 8164–8215 (2016).
2. Z. Sun, B. Fridrich, A. de Santi, S. Elangovan, K. Barta, Bright side of lignin depolymerization: Toward new platform chemicals. *Chem. Rev.* **118**, 614–678 (2018).
3. W. Schutyser, T. Renders, S. Van den Bosch, S.-F. Koelewijn, G. T. Beckham, B. F. Sels, Chemicals from lignin: An interplay of lignocellulose fractionation, depolymerisation, and upgrading. *Chem. Soc. Rev.* **47**, 852–908 (2018).
4. A. Rahimi, A. Ulbrich, J. J. Coon, S. S. Stahl, Formic-acid-induced depolymerization of oxidized lignin to aromatics. *Nature* **515**, 249–252 (2014).
5. C. S. Lancefield, O. S. Ojo, F. Tran, N. J. Westwood, Isolation of functionalized phenolic monomers through selective oxidation and C–O bond cleavage of the β-O-4 linkages in lignin. *Angew. Chem. Int. Ed. Engl.* **54**, 258–262 (2015).
6. E. Feghali, G. Carrot, P. Thuéry, C. Genre, T. Cantat, Convergent reductive depolymerization of wood lignin to isolated phenol derivatives by metal-free catalytic hydrosilylation. *Energ. Environ. Sci.* **8**, 2734–2743 (2015).
7. L. Shuai, M. T. Amiri, Y. M. Questell-Santiago, F. Héroguel, Y. Li, H. Kim, R. Meilan, C. Chapple, J. Ralph, J. S. Luterbacher, Formaldehyde stabilization facilitates lignin monomer production during biomass depolymerization. *Science* **354**, 329–333 (2016).
8. S. Van den Bosch, W. Schutyser, R. Vanholme, T. Driessen, S.-F. Koelewijn, T. Renders, B. De Meester, W. J. J. Huijgen, W. Dehaen, C. M. Courtin, B. Lagrain, W. Boerjan, B. F. Sels, Reductive lignocellulose fractionation into soluble lignin-derived phenolic monomers and dimers and processable carbohydrate pulps. *Energ. Environ. Sci.* **8**, 1748–1763 (2015).
9. Y. Shao, Q. N. Xia, L. Dong, X. Liu, X. Han, S. F. Parker, Y. Cheng, L. L. Daemen, A. J. Ramirez-Cuesta, S. H. Yang, Y. Q. Wang, Selective production of arenes via direct lignin upgrading over a niobium-based catalyst. *Nat. Commun.* **8**, 16104 (2017).
10. L.-P. Xiao, S. Wang, H. Li, Z. Li, Z.-J. Shi, L. Xiao, R.-C. Sun, Y. Fang, G. Song, Catalytic hydrogenolysis of lignins into phenolic compounds over carbon nanotube supported molybdenum oxide. *ACS Catal.* **7**, 7535–7542 (2017).
11. C. Zhang, H. Li, J. Lu, X. Zhang, K. E. MacArthur, M. Heggen, F. Wang, Promoting lignin depolymerization and restraining the condensation via an oxidation–hydrogenation strategy. *ACS Catal.* **7**, 3419–3429 (2017).
12. D. S. Argyropoulos, H. I. Bolker, Condensation of lignin in dioxane-water-HCl. *J. Wood Chem. Technol.* **7**, 1–23 (1987).
13. T. Yokoyama, Revisiting the mechanism of β-O-4 bond cleavage during acidolysis of lignin. Part 6: A review. *J. Wood Chem. Technol.* **35**, 27–42 (2015).
14. P. J. Deuss, M. Scott, F. Tran, N. J. Westwood, J. G. de Vries, K. Barta, Aromatic monomers by in situ conversion of reactive intermediates in the acid-catalyzed depolymerization of lignin. *J. Am. Chem. Soc.* **137**, 7456–7467 (2015).
15. V. M. Roberts, V. Stein, T. Reiner, A. Lemonidou, X. Li, J. A. Lercher, Towards quantitative catalytic lignin depolymerization. *Chemistry* **17**, 5939–5948 (2011).
16. R. Franke, C. M. McMichael, K. Meyer, A. M. Shirley, J. C. Cusumano, C. Chapple, Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase. *Plant J.* **22**, 223–234 (2000).
17. W. Lan, M. Talebi Amiri, C. M. Hunston, J. S. Luterbacher, Protection group effects during α,γ-diol lignin stabilization promote high-selectivity monomer production. *Angew. Chem. Int. Ed.* **57**, 1356–1360 (2018).
18. Y. Mottiar, R. Vanholme, W. Boerjan, J. Ralph, S. D. Mansfield, Designer lignins: Harnessing the plasticity of lignification. *Curr. Opin. Biotechnol.* **37**, 190–200 (2016).
19. F. Chen, Y. Tobimatsu, D. Havkin-Frenkel, R. A. Dixon, J. Ralph, A polymer of caffeyl alcohol in plant seeds. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1772–1777 (2012).
20. F. Chen, Y. Tobimatsu, L. Jackson, J. Nakashima, J. Ralph, R. A. Dixon, Novel seed coat lignins in the Cactaceae: Structure, distribution and implications for the evolution of lignin diversity. *Plant J.* **73**, 201–211 (2013).
21. H.-M. Chang, E. B. Cowling, W. Brown, E. Adler, G. Miksche, Comparative studies on cellulolytic enzyme lignin and milled wood lignin of sweetgum and spruce. *Holzforschung* **29**, 153–159 (1975).
22. H. Kim, J. Ralph, Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d_6/pyridine-d_5. *Org. Biomol. Chem.* **8**, 576–591 (2010).
23. N. Li, X. J. Pan, J. Alexander, A facile and fast method for quantitating lignin in lignocellulosic biomass using acidic lithium bromide trihydrate (ALBTH). *Green Chem.* **18**, 5367–5376 (2016).
24. Z. M. Xue, X. Zhao, R. C. Sun, T. C. Mu, Biomass-derived γ-valerolactone-based solvent systems for highly efficient dissolution of various lignins: Dissolution behavior and mechanism study. *ACS Sustainable Chem. Eng.* **4**, 3864–3870 (2016).
25. V. E. Tarabanko, D. V. Petukhov, G. E. Selyutin, New mechanism for the catalytic oxidation of lignin to vanillin. *Kinetics Catal.* **45**, 569–577 (2004).
26. C. Lapierre, B. Monties, C. Rolando, Preparative thioacidolysis of spruce lignin: Isolation and identification of main monomeric products. *Holzforschung* **40**, 47–50 (1986).
27. L. Berstis, T. Elder, M. Crowley, G. T. Beckham, Radical nature of C-lignin. *ACS Sustainable Chem. Eng.* **4**, 5327–5335 (2016).
28. S. Kim, S. C. Chmely, M. R. Nimos, Y. J. Bomble, T. D. Foust, R. S. Paton, G. T. Beckham, Computational study of bond dissociation enthalpies for a large range of native and modified lignins. *J. Phys. Chem. Lett.* **2**, 2846–2852 (2011).
29. J. Ralph, P. F. Schatz, F. Lu, H. Kim, T. Akiyama, S. F. Nelsen, in *Quinone Methides*, S. Rokita, Ed. (Wiley-Blackwell, 2009), vol. 1, pp. 385–420.
30. X. Y. Wang, R. Rinaldi, Solvent effects on the hydrogenolysis of diphenyl ether with Raney nickel and their implications for the conversion of lignin. *ChemSusChem* **5**, 1455–1466 (2012).
31. W. Schutyser, S. Van den Bosch, T. Renders, T. De Boe, S.-F. Koelewijn, A. Dewaele, T. Ennaert, O. Verkinderen, B. Goderis, C. M. Courtin, B. F. Sels, Influence of bio-based solvents on the catalytic reductive fractionation of birch wood. *Green Chem.* **17**, 5035–5045 (2015).
32. J. Sedó, J. Saiz-Poseu, F. Busqué, D. Ruiz-Molina, Catechol-based biomimetic functional materials. *Adv. Mater.* **25**, 653–701 (2013).
33. Q. Wu, D. Yan, Y. Chen, T. Wang, F. Xiong, W. Wei, Y. Lu, W.-Y. Sun, J. J. Li, J. Zhao, A redox-neutral catechol synthesis. *Nat. Commun.* **8**, 14227 (2017).
34. K. Barta, G. R. Warner, E. S. Beach, P. T. Anastas, Depolymerization of organosolv lignin to aromatic compounds over Cu-doped porous metal oxides. *Green Chem.* **16**, 191–196 (2014).
35. W. Wu, F. Liu, S. Singh, Toward engineering *E. coli* with an autoregulatory system for lignin valorization. *Proc. Natl. Acad. Sci U.S.A.* **115**, 2970–2975 (2018).
36. N. Li, Y. Li, C. G. Yoo, X. Yang, X. Lin, J. Ralph, X. Pan, An uncondensed lignin depolymerized in the solid state and isolated from lignocellulosic biomass: A mechanistic study. *Green Chem.* 10.1039/C2018GC00953H (2018).
37. A. Wagner, Y. Tobimatsu, L. Phillips, H. Flint, K. Torr, L. Donaldson, L. Piers, J. Ralph, *CCoAOMT* suppression modifies lignin composition in *Pinus radiata*. *Plant J.* **67**, 119–129 (2011).
38. R. Dixon, N. D'souza, F. Chen, M. Nar, U.S. Patent US9890480B2 (2018).
39. Y. Tobimatsu, F. Chen, J. Nakashima, L. Jackson, L. L. Escamilla-Treviño, R. A. Dixon, J. Ralph, Coexistence but independent biosynthesis of catechyl and guaiacyl/syringyl lignins in plant seeds. *Plant Cell* **25**, 2587–2600 (2013).
40. M. Montazeri, M. J. Eckelman, Life cycle assessment of catechols from lignin depolymerization. *ACS Sustainable Chem. Eng.* **4**, 708–718 (2016).
41. E. G. Bligh, W. J. Dyer, A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
42. A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, D. Crocker, "Determination of structural carbohydrates and lignin in biomass" (Technical Report, NREL/TP-510-42618, National Renewable Energy Laboratory, 2012).
43. Z. Xia, L. G. Akim, D. S. Argyropoulos, Quantitative ^{13}C NMR analysis of lignins with internal standards. *J. Agric. Food Chem.* **49**, 3573–3578 (2001).
44. K. M. Holtman, H.-M. Chang, H. Jameel, J. F. Kadla, Quantitative ^{13}C NMR characterization of milled wood lignins isolated by different milling techniques. *J. Wood Chem. Technol.* **26**, 21–34 (2006).
45. H. Kim, D. Padmakshan, Y. Li, J. Rencoret, R. D. Hatfield, J. Ralph, Characterization and elimination of undesirable protein residues in plant cell wall materials for enhancing lignin analysis by solution-state NMR. *Biomacromolecules* **18**, 4184–4195 (2017).
46. Y. Li, T. Akiyama, T. Yokoyama, Y. Matsumoto, NMR assignment for diaryl ether structures (4–O–5 structures) in pine wood lignin. *Biomacromolecules* **17**, 1921–1929 (2016).
47. C. Lapierre, B. Pollet, B. Monties, C. Rolando, Thioacidolysis of spruce lignin: Gas chromatography-mass spectroscopy analysis of the main dimers recovered after Raney nickel desulfurization. *Holzforschung* **45**, 61–68 (1991).
48. J. T. Scanion, D. E. Willis, Calculation of flame ionization detector relative response factors using the effective carbon number concept. *J. Chromatogr. Sci.* **23**, 333–340 (1985).

Acknowledgments
Funding: Funding was provided by the U.S. Department of Energy (DOE) Great Lakes Bioenergy Research Center (DOE Biological and Environmental Research Office of Science

DE-FC02-07ER64494 and DE-SC0018409), the DOE Center of Bioenergy Innovation (DE-AC05-000R22725) and the Swiss Competence Center for Energy Research: Biomass for a Swiss Energy Future, through the Swiss Commission for Technology and Innovation grant KTI.2014.0116. We are also grateful to N. Li and X. Pan (University Wisconsin–Madison) for help with LiBr solubilization methods and to Y. Mottiar and S. Mansfield (University of British Columbia, Vancouver, Canada) and R. Vanholme and W. Boerjan (Vlaams Instituut voor Biotechnologie, Gent, Belgium) for discussions on designing lignins that resulted in recent papers on this topic. We are also grateful to valuable *Science Advances'* reviewer comments. **Author contributions:** J.R., J.S.L., H.K., Y.L., L.S., and J.A.D. were responsible for the conception, planning, and organization of the experiments. D.H.-F. provided the vanilla seed coat material. Y.T., F.C., and R.A.D. provided information and were involved in discussions relating to it. Y.L. performed the synthesis of dimer **D1** and the synthetic **C**-lignin and isolated lignins, performed the hydrogenolysis experiments with L.S., and isolated, quantified, and identified the products. Y.L. also performed NBO reactions and analysis and sugars analysis. L.S. and A.H.M. aided in the hydrogenolysis experiments. F.Y. performed the thioacidolysis analysis and helped with the model compound synthesis. H.K. helped Y.L. perform the quantitative ^{13}C NMR. J.K.M. helped with the catalyst considerations and carried out the GPC and molecular weight analysis. Y.L. and J.R. were responsible for NMR and MS data and interpretation. The manuscript was primarily written by Y.L. and J.R. with critical input from all coauthors. Figures were prepared by Y.L. and J.R. with support from H.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 26 May 2018
Accepted 22 August 2018
Published 28 September 2018
10.1126/sciadv.aau2968

Citation: Y. Li, L. Shuai, H. Kim, A. H. Motagamwala, J. K. Mobley, F. Yue, Y. Tobimatsu, D. Havkin-Frenkel, F. Chen, R. A. Dixon, J. S. Luterbacher, J. A. Dumesic, J. Ralph, An "ideal lignin" facilitates full biomass utilization. *Sci. Adv.* **4**, eaau2968 (2018).

Science Advances

An "ideal lignin" facilitates full biomass utilization

Yanding Li, Li Shuai, Hoon Kim, Ali Hussain Motagamwala, Justin K. Mobley, Fengxia Yue, Yuki Tobimatsu, Daphna Havkin-Frenkel, Fang Chen, Richard A. Dixon, Jeremy S. Luterbacher, James A. Dumesic and John Ralph

Sci Adv **4** (9), eaau2968.
DOI: 10.1126/sciadv.aau2968