Lignins are abundant phenylpropanoid polymers produced primarily from oxidative polymerization of three 4-hydroxycinnamyl alcohols differing in their degrees of methoxylation (Fig. S1). Lignins occur mostly in vessels, tracheids, and fibrous tissues of vascular plants where they bind, strengthen, and waterproof cell walls to provide mechanical support, enhance water transport, and help ward off pathogens and pests. The biosynthesis and bioengineering of cell wall lignins, and their chemical and mechanical properties, have attracted significant attention because lignin hinders agro-industrial processes, such as chemical pulping of woody crops (1), forage digestion by livestock (2), and conversion of lignocellulosic plant biomass into liquid biofuels (3, 4). In addition, the variability of biosynthesis, and thereby the structures of various lignins, is considered to be closely correlated with the diversity and evolution of land plants (3, 5).

DURING lignin biosynthesis, the monolignol precursors are functionalized by aromatic hydroxylation and O-methylation (as well as successive side-chain reductions) to generate monolignols differing in their aromatic substitution patterns (Fig. L4 and Fig. S1). Natural lignins are generally composed of p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, that are biosynthesized by polymerization of the three primary monolignols, p-coumaryl, coniferyl, and sinapyl alcohols, respectively; natural angiosperm lignins have only low levels (≤2%) of H-units. Catechyl (C) and 5-hydroxyguaiacyl (5-OH-G) units that may derive from polymerization of the corresponding caffeyl and 5-hydroxy coniferyl aldehydes (3, 4) are found in “normal” lignins. Extensive studies have revealed the essential plasticity of lignin biosynthesis (6, 10, 12–15), and support the concept that lignin polymerization results from a combinatorial radical coupling process that is under simple chemical control (14, 16, 17). Thus, lignin monomer composition is largely determined by monolignol availability and, under certain circumstances, this permits incorporation of the “unusual” C and 5-OH-G monolignols into the polymer. For example, 5-hydroxyconiferyl alcohol participates in lignification in various angiosperm plants in which caffeic acid/5-hydroxyconiferaldehyde O-methyltransferase (COMT), the key enzyme for conversion of monolignol precursors from the 5-OH-G to the S aromatic level (18, 19), is down-regulated. The combination of a mutation in the gene encoding COMT with overexpression of 5-hydroxyferulate 5-hydroxylase, which catalyzes hydroxylation of G to 5-OH-G aromatic level precursors, generates lignins largely composed of 5-OH-G units in benzodioxane structures (20, 21). Similarly, down-regulation of caffeoyl-CoA O-methyltransferase (CCoAMT), for conversions from C to G aromatic-level precursors, introduces low levels (less than 10%) of C units into cell wall lignins in tracheary-element cultures of the gymnosperm Pinus radiata (22). However, down-regulation of CCoAMT in angiosperm species, such as Arabidopsis, alfalfa, poplar, and tobacco, does not result in the incorporation of C units into lignin (23–27), and neither does down-regulation of both monolignol methylation enzymes (28). Here we report a lignin in the monocotyledonous angiosperm Vanilla orchid (Vanilla planifolia) that is naturally biosynthesized from the unusual C monolignol, caffeyl alcohol. Similar polymers are found in the seeds of other vanilla species and several species of cacti (which are dicots). The V. planifolia polymer was structurally characterized by various chemical methods, 2D NMR spectroscopic techniques, and gel-permeation chromatography (GPC). All evidence indicates that the C-lignin is formed by combinatorial oxidative radical coupling under simple chemical control, a mechanism analogous to that occurring in classic lignification.

RESULTS

Identification of C-Lignin Signatures in V. planifolia. Initial studies on the lignin of mature beans of the vanilla orchid V. planifolia (Fig. 1C) by thioacidolysis (29, 30) revealed the presence of a small doublet in the gallic chromatography (GC)-MS profile at a retention time consistent with the catechyl (C) monomer, α,β,γ-trithioethyl-propylcatechol (Fig. 1B). The beans contained black-coated seeds (Fig. 1F), and thioacidolysis revealed that the lignin in the isolated seed coats was entirely composed of C units (Fig. 1H), with practically no release of α,β,γ-trithioethyl-propylguaiacol, from whole cell wall NMR

PLANT BIOLOGY

Lignins are complex phenylpropanoid polymers mostly associated with plant secondary cell walls. Lignins arise primarily via oxidative polymerization of the three monolignols, p-coumaryl, coniferyl, and sinapyl alcohols. Of the two hydroxycinnamyl alcohols that represent incompletely methylated biosynthetic products (and are not usually considered to be monolignols), 5-hydroxyconiferyl alcohol is now well established as incorporating into angiosperm lignins, but incorporation of caffeayl alcohol has not been shown. We report here the presence of a homopolymer of caffeyl alcohol in the seed coats of both monocot and dicot plants. This polymer (C-lignin) is deposited to high concentrations in the seed coat during the early stages of seed development in the vanilla orchid (Vanilla planifolia), and in several members of the Cactaceae. The lignin in other parts of the Vanilla plant is conventionally biosynthesized from coniferyl and sinapyl alcohols. Some species of cacti contain only C-lignin in their seeds, whereas others contain only classical guaiacyl/syringyl lignin (derived from coniferyl and sinapyl alcohols). NMR spectroscopic analysis revealed that the Vanilla seed-coat polymer was massively comprised of benzodioxane units and was structurally similar to the polymer synthesized in vitro by peroxidase-catalyzed polymerization of caffeyl alcohol. CD spectroscopy did not detect any optical activity in the seed polymer. These data support the contention that the C-lignin polymer is produced in vivo via combinatorial oxidative radical coupling that is under simple chemical control, a mechanism analogous to that theorized for classical lignin biosynthesis.
guaiacyl units, nor the syringyl analog (Fig. 1B). In contrast, thio-
acidolysis of the pod residue (after seed isolation), stem, leaf, and
aerial root released normal G and S monomers with essentially no
C monomers, indicating that the lignins present in these tissues
are typical G-rich G/S-lignins (Fig. 1–K). C-lignin signatures first
appeared in the seed coat at around 8 wk after pollination (Fig.
S2), when the coat turns from transparent white to brown (Fig. 1
D–F), and at least 2–3 mo before the appearance in the pods of the
flavor compound vanillin, which is biosynthetically related to lig-
nin (31). The absolute levels of the C-lignin in the vanilla seed are
substantially higher than the values estimated from the released
thioacidolysis monomers because of the unusual structure of the
polymer (see below).

C-Lignin Signatures in Seeds from Other Plant Species. Similar GC-
MS traces to that shown for the V. planifolia seed coat were obtained
from seed coats of two other Vanilla species, Vanilla pompona and
Vanilla tahitensis. However, C units were not detected in the seed
coats of Phalaenopsis orchid species, nor Asparagus and Agave, two
other members of the monocot order Asparagales, to which Vanilla
belongs. We did, however, observe strong C-lignin signatures in the
seed coats from several species of the family Cactaceae, specifically
members of the genera Astrophytum, Discocactus, Frailea, Melocactus
(Fig. 1L), Notocactus, Uebelmanna, and Wiggiosia. Interestingly,
all these species possessed black-coated seeds (Fig. 1G) that re-
leased only C monomers on thioacidolysis, whereas Mammillaria
(Fig. 1M) and Opuntia species possessed brown seeds that con-
tained normal G/S-lignin. During this limited survey, no seeds
were found that contained both C- and G/S-lignins.

Wet-Chemical and NMR Characterization of V. planifolia Tissues.
Separated tissues from mature V. planifolia were further charac-
terized by wet-chemical methods and by 2D NMR using direct
dissolution/swelling (32, 33). Klason analysis of the seed coat indi-
cated a very high level (~80%) of acid-insoluble lignin polymer
(Table S1). Butanol-HCl assay (34) did not detect proanthocya-
nidins, which are major components in some seed coats (e.g.,
Arabidopsis and Medicago) (35, 36). The majority of the remaining
material in the seed coat was crystalline cellulose (16%); very little
noncellulosic sugars (2%) were detected. The chemical composi-
tions of the pod remaining after seed isolation and the stem were
similar overall (Table S1); these tissues were most rich in cellulose,
with modest levels of hemicellulosic and pectic sugars, and Klason
lignins.

Two-dimensional gel-state NMR spectra of vanilla tissues were
acquired with samples prepared by swelling whole-plant materials,
after fine milling, in DMSO-<i>d</i><sub>6</sub>/pyridine-<i>d</i><sub>5</sub> (4:1, vol/vol) (33) (Fig.
2). The spectra of the pod and stem displayed typical G/S type
lignins with an array of hemicellulosic and pectic sugar units (Fig. 2
B and C; see also Fig. S3 for expanded polysaccharide anomeric
regions). However, the overwhelming signals in the spectrum of
the seed coat were from the C-lignin polymer (Fig. 2A), as evi-
denced by the striking similarity to the spectrum of a synthetic
dehydrogenation polymer (C-DHP) generated by horseradish
peroxidase-catalyzed polymerization of caffeal alcohol (Fig. 2D).
The aliphatic regions of the spectrum indicated a massive pres-
ence of benzodioxane units V, for which the α-, β-, and γ-correla-
tions from trans-benzodioxane rings V<sub>r</sub>, as well as lower-level
contributions from cis-benzodioxane rings V<sub>c</sub>, were resolved and
readily assigned by comparison with the data from the synthetic
model dimers (Fig. S4 A and B). Conventional lignin aromatic
signals were predictably absent and, instead, the dominant signals
were from C units: compelling confirmation of these assignments
could be made via comparison with the C-DHP and model com-
ponds (Fig. S4 A and B). The benzodioxane polymer in the seed
coat is therefore derived from the polymerization, almost exclu-
sively, of caffeal alcohol (see Fig. 3D for the mechanism).

Ball-milled vanilla seed coat was also analyzed by normal solu-
tion-state NMR via complete dissolution/acetolysis using the
DMSO/N-methylimidazole (NMI) solvent system (Fig. S5) (32).
A massive presence of C-lignins was again firmly established by
diagnostic catechyl aromatic signals and benzodioxane units; these
signals were predictably shifted following acetolysis and could be
assigned by comparison with the data from acetolylated model
dimers and C-DHP (Fig. S4 C and D). In addition, the signals from
cellulosic glucans, which were not significant in the gel-state NMR
spectra (Fig. 2) because of the incomplete gelation of crystalline
cellulose (33), were then clearly observed, but the signals from
hemicellulosic and pectic sugars remain practically absent.

NMR Characterization of Isolated C-Lignin from V. planifolia Seed
Coat. Representative fractions enriched in the C-lignin polymer
were isolated in 24% or 16% yield by cellulase treatment of ball-
milled V. planifolia seed coat followed by extraction with DMSO
or with 96% dioxane/water solution (37–39) (Table S1). Solution-
state 13C-1H correlation heteronuclear single quantum coherence
(HSQC) spectra of the isolated fractions in DMSO-<i>d</i><sub>6</sub> indi-
cated successful removal of the polysaccharides and concurrent
enrichment of C-lignin (Fig. S6). The isolated seed-coat lignin was then acetylated so as to be soluble in chloroform-\textit{d}, which facilitates long-range $^{13}$C–$^1$H correlation NMR experiments (e.g., heteronuclear multiple bond correlation, HMBC) (40, 41).

The large differences from classical lignins are most readily visualized from the aromatic regions of HSQC spectra of the isolated and acetylated seed-coat lignin (Fig. 3A, Left). Volume integration of the contour signals confirmed that this lignin is almost exclusively composed of C units; typical G and S lignin aromatics are virtually nonexistent in this polymer. Other correlations (gray) are currently unassigned, but do not seem to arise from caffeyl alcohol because they are not seen in the acetylated C-DHP (Fig. 3B).

High-field HSQC spectra of the side-chain (aliphatic) regions resolved most of the correlations for the various linkage types, revealing more clearly the manner in which the monomeric units are assembled (Fig. 3A, Right). The major contours in the spectrum were almost identical to those of the C-DHP (Fig. 3B). Benzodioxanes, resulting from $\beta$-O-4-coupling of a monomer with a caffeyl unit, were the dominant units in both the seed-coat lignin and C-DHP, accounting for over 98% of the total identifiable dimeric units. The \textit{trans/cis} compositions of the benzodioxane rings (\textit{Vt}/\textit{Vc}) in the seed polymer and C-DHP were similar (\textit{Vt}:\textit{Vc} = 97:3 and 96:4, respectively). The normal acyclic $\beta$-aryl ether units \textbf{I}, which are the predominant linkages in typical natural lignins, were absent in these polymers. Small amounts of phenylcoumaran \textbf{II} and resinoI \textbf{III} units were present in both the seed-coat lignin and C-DHP. HMBC experiments revealed the expected long-range correlations between the side-chain $\alpha$-protons and the 1-, 2-, and 6 carbons of the catechyl aromatic rings in the polymer, supporting the contention that all of these units derive from caffeyl alcohol (Fig. S7). In addition, at least four unassigned correlations (colored in pink) were observed in the spectrum of the seed polymer (Fig. 3A). Because these were also identified in the spectrum of C-DHP (Fig. 3B), they presumably represent new structures resulting from radical coupling reactions of caffeyl alcohol. The isolated stem lignin is, in contrast, a G-rich G/S type lignin rich in $\beta$-aryl ether units \textbf{I}, with more modest amounts of phenylcoumaran \textbf{II} and resinoI \textbf{III} units, and also with more minor amounts of dibenzodioxacin units \textbf{IV} (Fig. 3C, and Figs. S6 and S7), as is typical for angio-derm stem lignins (40, 41).

**Molecular Weights of C-Lignin in \textit{V. planifolia} Seed Coat.** Molecular-weight distributions of acetylated samples of ball-milled whole \textit{V. planifolia} seed coat and extracted C-lignins were determined by GPC with UV detection (Fig. 4A and Table S2). The molecular-weight profiles showed broad distributions spanning a range of up to $10^5$ Da. The number-average degree of polymerization (\textit{Dn}, based on the molecular weight of the catechyl benzodioxane internal unit) of the whole seed coat was $\sim$30. The \textit{Dn} of the isolated lignins extracted with DMSO and 96% dioxane: water solution were $\sim$18 and 13, respectively. It is likely that the insoluble fractions that were left after extractions of isolated seed coat lignins contain C-lignins with higher molecular masses. As expected from the end-unit analysis (Fig. 3A and B), the \textit{Dn} of the in vitro lignin (C-DHP) was even lower than those of isolated lignins, $\sim$8. All these values are comparable to literature values for various isolated and synthetic lignins (42–44).

**Fig. 2.** A 2D NMR characterization of separated \textit{V. planifolia} tissues. (A–C) Gel-state short-range $^{13}$C–$^1$H correlation (HSQC) spectra of whole vanilla tissues, (A) seed coat, (B) pod, and (C) stem in 4:1 dimethyl sulfoxide-\textit{d}$_6$/pyridine-\textit{d}$_5$. (D) Solution-state HSQC spectra of an in vitro polymer (C-DHP), synthesized via peroxidase-catalyzed polymerization of caffeyl alcohol, in dimethyl sulfoxide-\textit{d}$_6$. 

Chen et al.
Optical Activity of C-Lignin in V. planifolia Seed Coat. Natural lignins are optically inactive (45, 46). The optical activity of the isolated seed-coat lignin was investigated by CD spectroscopy. In addition, chiral benzodioxane dimers 1a and 1b were separated from a racemic mixture of synthetic dimer 1 (Fig. S4) by chiral HPLC and also subjected to CD for comparison (Fig. 4B and C). The CD spectrum of the seed-coat lignin indicated no detectable optical activity, whereas the chiral dimer 1a displayed clear positive Cotton effects in the region of 240–320 nm under the same analytical conditions (Fig. 4D). The other enantiomer 1b had essentially a mirror image spectrum (Fig. 4C). Optical activity was readily seen by spiking (46) a preparation of the seed polymer with as little as 5% of chiral dimer 1a (Fig. 4D). Benzodioxane units in the vanilla seed are therefore, within the limits of detection by the current method, optically inactive.

Discussion

Seeds of both monocot and dicot species contain previously unsuspected lignin polymers constructed almost entirely from catechyl (C) units. This C-polymer is a major component of the seed coat of V. planifolia, whereas the stem, leaf, and aerial root have only typical angiosperm G/S lignins. Thioacidolysis and 2D-NMR data clearly indicate that the C-polymer is essentially a homopolymer synthesized purely from caffeoyl alcohol, and with benzodioxanes as essentially the only intermonomer unit in the polymer. Based on preliminary thioacidolysis data, similar C-lignins are found in the seed coats of certain cactus species. Our data so far suggest that lignins present in seed coat of cactus species are either of the C- or G/S-types, but not both.

It is premature to speculate about the possible distribution of the C-lignin polymer within the plant kingdom. Currently only observed in the Asparagales (Orchidaceae) and Caryophyllales (Cactaceae), it is likely that the polymer has wider distribution, as it is found in both monocots and dicots. The taxonomy of the Cactaceae is constantly under revision, but in a recent analysis *Astrophytum* (C-lignin in seed) and *Mammillaria* (G/S-lignin in seed) are closely related in the same clade (47). This finding suggests that the formation of C-lignin is not an ancient trait, but has occurred recently and probably frequently within the plant kingdom. Thus, the genetic/biochemical mechanisms that allow for the monolignol pathway to be derailed into production of high concentrations of caffeoyl alcohol are probably relatively simple.

The C-polymer appears in *V. planifolia* beans at least 2–3 mo before the appearance of the flavor compound vanillin, which is synthesized in hair cells within the pod (48) and is likely derived from an initial phenylpropanoid precursor by side-chain shortening (31). The high concentration (>80%) by the Klason lignin method) of C-polymer in the seed coat may imply a lignin-like structural role in addition to a tannin-like role for seed protection.

Fig. 3. Short-range $^{13}$C–$^1$H correlation (HSQC) spectra of acetylated isolated lignins from *V. planifolia* seed coat (A) and stem (C) (extracted with dioxane-water, 96:4, vol/vol), and (B) an acetylated in vitro polymer (C-DHP), synthesized via peroxidase-catalyzed polymerization of caffeoyl alcohol. For abbreviations for signal assignments, see Fig. 2. (D) Scheme for generation of benzodioxane units via radical cross-coupling reactions between caffeoyl alcohol monomer Mc and catechyl (C)-polymer (P$_c$) end-units.
The vanilla seed polymer is strikingly similar to the in vitro polymer synthesized via peroxidase-catalyzed polymerization of caffeoyl alcohol. Both polymers are massively composed of benzodioxane units, which are uniquely formed via radical coupling of the monomer (at its β-position) with a C-polymere end-unit (at its 4-O-position) followed by internal trapping of the quinone methide intermediates (QM) by the o-hydroxyl (3-hydroxyl) group in the C unit to form the six-membered ring (Fig. 3). Similar benzodioxanes were recently identified in cell cultures of CCoAOMT-deficient P. radiata (22), and analogous benzodioxanes are products of lignification with 5-hydroxyconiferyl alcohol in COMT-deficient angiosperms (20, 21, 26, 49, 50). As the only β-O-4-type units were benzodioxanes, the postcoupling rearomatization of QM seems to be exclusively via the efficient internal trapping by the o-hydroxyl group in C units (Fig. 3D). The main benzodioxane backbones in the seed polymer are trans/cis-isomeric mixtures, as in the in vitro polymer and synthetic dimer, suggesting that the stereochimistry of postcoupling rearomatization of the QM is under simple kinetic control. Therefore, it is most plausible that caffeoyl alcohol is enzymatically oxidized, presumably by plant oxidoreductases such as peroxidases and laccases initially, but is cross-coupled onto the growing polymer in a chemically controlled fashion, independent of enzymes or other proteins, in the same way as conventional monolignols are during lignin polymerization (14, 16, 17).

In conclusion, the identification of this unique polymer provides compelling evidence for flexibility in the construction of lignin polymers in nature. The mechanisms that allow for formation of caffeoyl alcohol in developing vanilla and cactus seeds, and the question of whether such catechyl polymers are much more widespread in nature, remain to be determined. Such studies might contribute to the development of new avenues in lignin bioengineering, and may also provide new insights into the diversity and evolution of land plants.

Materials and Methods

Plant Materials. Mature vanilla beans were provided by Bakto Flavors, separated into the seed and pod (residue left after seed isolation), and processed as described in SI Materials and Methods. Vanilla stem material was obtained from vines growing in the greenhouse at the Noble Foundation, Ardmore, OK. Seeds of all cactus species were obtained from flowering plants in the collection of one of the authors (R.A.D.).

Isolations of Vanilla Lignins. Vanilla lignin samples for NMR, GPC, and CD analyses were prepared via methods largely described previously (22, 33, 39), and as further described in SI Materials and Methods.

Synthetic Model Dimers. The benzodioxane dimer 1 was synthesized from radical coupling reactions of caffeoyl alcohol via silver carbonate (Ag2CO3) oxidation, and dimer 2 was via methylation of dimer 1 with methyl iodide; detailed synthetic protocols and complete NMR and MS spectroscopic data are described in SI Materials and Methods.

Dehydrogenation Polymer from Caffeoyl Alcohol. A dehydrogenation polymer from caffeoyl alcohol (C-DHP) was generated via HRP-catalyzed polymerization, as previously described (22, 51, 52), and further outlined in SI Materials and Methods.

Chemical Analyses. Determination of Klasson lignin, crystalline cellulose, amorphous sugars, protein content, proanthocyanidins and lignin composition (by thioacidolysis) were as described in SI Materials and Methods.

NMR Spectroscopy. The NMR methods used were largely described previously (22, 32, 33, 51), and as further described in SI Materials and Methods.

GPC. GPC was performed on a Shimadzu LC-20A LC system (Shimadzu) as further described in SI Materials and Methods.

Chiral HPLC. Analytical and preparative chiral HPLC for enantiomeric separation of benzodioxane dimer 1 was performed on a Shimadzu LC-20A LC system as described in SI Materials and Methods.

CD Spectroscopy. CD spectra were run on an Model 202SF CD spectrophotometer (Aviv Biomedical) as described in SI Materials and Methods.
ACKNOWLEDGMENTS. We thank Cliff Foster for carbohydrate analysis; Heike Hofstetter for elemental analysis; Darrel McCaslin for assistance with CD spectrometry; Li Shuai, and Sasikumar Elumalai for assistance with Klasson lignin analysis; Lisa Jackson for assistance with thioacidolysis analysis; and Hoon Kim for helpful suggestions and assistance with NMR spectroscopy. This work was supported in part by the US Department of Energy Bioenergy Science Center and Great Lakes Bioenergy Centers, supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science (BER DE-AC05-00OR22725 and DE-FC02-07ER64940, respectively), and by a postdoctoral fellowship from the Japan Society for the Promotion of Science (to Y.T.).


