Hydroxycinnamaldehyde-derived benzofuran components in lignins

Koichi Yoshioka,[1,2] Hoon Kim,[1,2,3] Fachuang Lu,[1,2] Nette De Ridder,[4,5] Ruben Vanholme,[4,5] Shinya Kajita,[6,7] Wout Boerjan,[4,5] and John Ralph[1,2,8*]

1 The US Department of Energy's Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53726, USA
2 The Wisconsin Energy Institute, University of Wisconsin, Madison, WI 53726, USA
3 Current address: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, WI 53726, USA
4 Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 71, 9052 Ghent, Belgium
5 VIB Center for Plant Systems Biology, VIB, Technologiepark 71, 9052 Ghent, Belgium
6 Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
7 Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
8 Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

*Author for correspondence: jralph@wisc.edu
University of Wisconsin-Madison, Wisconsin Energy Institute, 1552 University Avenue, Madison, WI 53726, USA. **Tel: 608-890-2429**

Research Area:
Biochemistry and Metabolism

Hydroxycinnamaldehyde-derived benzofuran components in lignins

Koichi Yoshioka,[1,2] Hoon Kim,[1,2,3] Fachuang Lu,[1,2] Nette De Ridder,[4,5] Ruben Vanholme,[4,5] Shinya Kajita,[6,7] Wout Boerjan,[4,5] and John Ralph[1,2,8*]

1 The US Department of Energy's Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53726, USA
2 The Wisconsin Energy Institute, University of Wisconsin, Madison, WI 53726, USA
3 Current address: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, WI 53726, USA
4 Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 71, 9052 Ghent, Belgium
5 VIB Center for Plant Systems Biology, VIB, Technologiepark 71, 9052 Ghent, Belgium
6 Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
7 Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
8 Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

*Author for correspondence: jralph@wisc.edu

Abstract

Lignin is an abundant polymer in plant secondary cell walls. Prototypical lignin derives from the polymerization of monolignols (hydroxycinnamyl alcohols), primarily coniferyl and sinapyl alcohol, via combinatorial radical coupling reactions, and primarily via the endwise coupling of a monomer with the phenolic end of the growing polymer. Hydroxycinnamaldehyde units have long been recognized as minor components of lignins. In plants deficient in Cinnamyl Alcohol Dehydrogenase, the last enzyme in the monolignol biosynthetic pathway that reduces the hydroxycinnamaldehydes to the monolignols, the levels of chain-incorporated aldehyde units are elevated. The nature and relative levels of aldehyde components in lignins can be determined from their distinct and dispersed correlations in 2D ^{1}H–^{13}C correlated NMR spectra. We recently became aware of aldehyde NMR peaks, well-resolved from others, that had been overlooked. NMR of isolated low-molecular-weight oligomers from biomimetic radical coupling reactions involving coniferaldehyde revealed that the new correlations belonged to

hydroxycinnamaldehyde-derived benzofuran moieties. Coniferaldehyde 8-5-coupling initially produces the expected phenylcoumaran structures but the derived phenolic radicals undergo preferential disproportionation rather than radical coupling to extend the growing polymer. As a result, hydroxycinnamaldehyde-derived phenylcoumaran units are difficult to detect in lignins, but the benzofurans are now readily observed by their distinct and dispersed correlations in the aldehyde region of NMR spectra from any lignin or monolignol dehydrogenation polymer. Hydroxycinnamaldehydes coupled to coniferaldehyde can be distinguished from those derived from coupling with a generic guaiacyl end-unit. These 'new' benzofuran peaks may now be annotated and reported, and their structural ramifications further studied.

Keywords: lignification, CAD, monolignol, coniferaldehyde, sinapaldehyde, phenylcoumaran, NMR

Introduction

Lignins derive from the polymerization of monolignols (*p*-hydroxycinnamyl alcohols) via combinatorial radical coupling reactions, and primarily via the endwise coupling of a monomer with the phenolic end of the growing polymer (Freudenberg and Neish, 1968; Ralph et al., 2004b; Vanholme et al., 2012; Ralph et al., 2019). In this lignification process, the two major monomers, coniferyl alcohol (**G**^{a}) and sinapyl alcohol (**S**^{a}), give rise to guaiacyl (**G**) and syringyl (**S**) units in the polymer (Figure 1). Dimeric units in the polymer arise from the various possible radical-coupling modes. The coupling itself produces unstable intermediates that re-aromatize to produce the final dimeric units that are conventionally identified by the characteristic coupling reactions that produced them and the new inter-unit bonds that result. In normal lignins derived from the monolignols the convention is to number the aromatic ring carbons in the usual fashion from 1 to 6 and the sidechain carbons with Greek letters (α, β, and γ) starting from the ring (Figure 1). As monolignols favor coupling at their β-positions, β-aryl ether (usually simply termed β-ether, β-O-4), phenylcoumaran (β-5), resinol (β-β), and spirodienone (β-1) units are common; biphenyl (5-5) and biaryl ether (4-O-5) units arise from the radical coupling of dimers or higher oligomers (Ralph et al., 2004b; Vanholme et al., 2012; Ralph et al., 2019). Only two such units, the β-aryl ether (β-O-4) and the phenylcoumaran (β-5), are shown in Figure 1 that also explains the lettering shorthand that has become rather standard, **A** for β-ethers from β-O-4-coupling, **B** for phenylcoumarans from β-5-coupling, etc.

Hydroxycinnamaldehyde moieties, as well as their derived hydroxybenzaldehydes, are typically minor structures in polymeric lignins. The most famous lignin stain, the Wiesner reaction with phloroglucinol, derives its characteristic color not from main-chain or "core-lignin" units but from hydroxycinnamaldehyde endgroups (Adler et al., 1948). Originally thought to be created post-

lignification by oxidation of hydroxycinnamyl alcohol endgroups, it is increasingly realized that the hydroxycinnamaldehydes are authentic lignin monomers, especially in plants deficient in the last monolignol biosynthetic pathway enzyme, coniferyl alcohol dehydrogenase (CAD), that converts the penultimate hydroxycinnamaldehydes to their hydroxycinnamyl alcohols (Higuchi et al., 1994; Kim et al., 2003; Anderson et al., 2015). Hydroxycinnamaldehydes originate either from the translocation to the plant cell wall of the hydroxycinnamaldehyde monomers themselves resulting from truncated monolignol biosynthesis or from minor oxidation of the monolignols under the oxidative lignification conditions and the H_2O_2 present in the cell wall itself (Kim et al., 2003; Ralph et al., 2023a); hydroxycinnamaldehydes and hydroxybenzaldehydes can both be noted in DHPs from pure monolignols, indicating that the oxidation is occurring in the system, and coniferaldehyde can indeed be produced from coniferyl alcohol in H_2O_2 solution.

In normal lignification in wild-type (WT) plants the main incorporation of hydroxycinnamaldehydes, as with their hydroxybenzaldehydes, is as endgroups. Such endgroups result from the favorable cross-coupling reactions with monolignols in which the monolignol dominantly couples at its β-position, relegating the hydroxycinnamaldehyde to coupling at its O-4- and 5- as well as perhaps its own 8-position. They are therefore usually found as pendent starting-end units on the lignin, but incorporate at low levels into the lignin backbone by cross-coupling with the phenolic end of the growing polymer. In CAD-deficient plants the levels of chain-incorporated aldehyde units are elevated and, in extreme cases, highly CAD-deficient plants may create their lignins almost entirely from hydroxycinnamaldehydes with scarcely any contribution from the normal monolignols (Sibout et al., 2005; Zhao et al., 2013; Anderson et al., 2015), a testament to the flexibility of the lignification system in allowing plants to survive even drastic mutations or engineering (Ralph et al., 2008).

Coniferaldehyde (**G'**, Figure 1) and sinapaldehyde (**S'**) can undergo peroxidase-H_2O_2 oxidation forming dehydrodimers by combinatorial radical coupling reactions in a manner similar to the formation of lignin dehydrodimers from coniferyl alcohol (**G^a**) and/or sinapyl alcohol (**S^a**), as has been long known (Higuchi et al., 1994; Kim et al., 2003). The final products differ in that the intermediate quinone methide, generated each time a hydroxycinnamaldehyde monomer **G'** or **S'** couples via its 8-position (analogous to the β-position in a monolignol), has a lower-energy path to rearomatization because of the ready elimination of the 8-proton that is particularly acidic due to its proximity to the aldehyde group (e.g., for the 8-5-dimer **2** in Figure 2A). Many of these hydroxycinnamaldehyde coupling modes are evident in the NMR spectra of lignins from CAD-deficient plants (Kim et al., 2003; Zhao et al., 2013; Anderson et al., 2015; Van Acker et al., 2017; Yan et al., 2019; Yamamoto et al., 2020), as will be seen in Figure 3.

Hydroxycinnamaldehyde 8-O-4-coupled units (denoted as **A'** units here, Figure 1) were originally identified in CAD-deficient plants (Ralph et al., 1997; Sederoff et al., 1999; Chabannes et al., 2001; Ralph et al., 2001; Kim et al., 2002; Kim et al., 2003; Barrière et al., 2004; Lapierre et al., 2004; Anderson et al., 2015; Yan et al., 2019; Liu et al., 2021) and can now be found at low levels in most lignins. The nature and relative levels of aldehyde components in lignins is revealed by their distinct and dispersed correlations in the aldehyde region of 2D $^{1}H–^{13}C$ correlation (HSQC, heteronuclear single-quantum coherence) NMR spectra (Anderson et al., 2015; Yan et al., 2019; Yamamoto et al., 2020).

In recent studies on a CAD-deficient mulberry mutant (Yamamoto et al., 2020), we noted but did not report on the presence of "new" well-dispersed aldehyde signals in 2D HSQC NMR spectra. This peak is prominently shown with red contours in Figure 3D. We had observed these peaks previously from many different lignins, including biomimetically-produced synthetic lignins (dehydrogenation polymers, DHPs), but they appear to have gone without comment in the literature. It was therefore timely to identify the product in the polymer and gain insight into the seemingly unusual structure and its ramifications.

Results and Discussion

Determining the nature of a new 2D NMR correlation peak via synthetic lignification

We had previously observed the HSQC NMR signature of this new component in synthetic lignins (dehydrogenation polymers, DHPs) produced from monolignols by the peroxidase-H_2O_2 single-electron oxidation system, and in particular from DHPs in which coniferaldehyde was also a component, or from a coniferaldehyde-only DHP. We therefore reasoned that we might be able to isolate one or more low-molecular-mass components from truncated DHP synthesis, i.e., by a limited-duration one-electron-oxidation of coniferaldehyde. Such products could then be purified and identified by the usual combination of MS and NMR analytical methods.

Radical coupling of coniferaldehyde **1** (Figure 2A) using peroxidase-H_2O_2 in buffer, pH 6.5, produced a dimer in low yield that, after purification, had the characteristic $^{1}H–^{13}C$ NMR correlation signature seen in the lignin and DHP spectra, Figure 3. It was readily identified from its distinct NMR correlation peak, and from full structural analysis by the usual array of NMR experiments, as compound **3** (Figure 2A, Supplemental Figure S1).

The origin of structure **3** was not initially obvious. It was clear (Figure 2A) that it was a dehydrodimerization product resulting from 8-5-coupling of radical **1•** from coniferaldehyde **1** (also denoted more generally as **G'**, Figure 1). Following rearomatization of the intermediate quinone methide **QM1**, the known phenylcoumaran dehydrodimer **2** is produced. But how did the 7-8 double-bond in **3** arise? The production of the (unsaturated) benzofuran ring required another dehydrogenation, i.e., the production of the phenolic radical from compound **2**. Radical **2•** can, via interaction with another such

radical, disproportionate such that one of the radicals becomes reduced to the original phenolic **2**, and one becomes oxidized (dehydrogenated) to a quinone methide **QM2** (Figure 2A). The quinone methide intermediate **QM2** readily rearomatizes to product **3** by loss of the acidic 8-proton. In the coniferaldehyde DHP, this interaction of two radicals **2**$^{\bullet}$ is the way to produce the dehydrogenated dimer **3**, but it is also reasonable to surmise that the coupling of one coniferaldehyde radical need not be with another such monomer – it could equally be with any general guaiacyl radical **4**$^{\bullet}_{G}$ (Figure 2B). An analogous product from sinapaldehyde **1**$_{S'}$ (or just **S'**) is also possible. Additionally, and not diagramed in Figure 2, disproportionation need not involve two radicals of exactly the same type.

After making this structural assignment for dimer **3** that was isolated from radical coupling of coniferaldehyde, and the apparent discovery of a new radical coupling pathway, we learned that dimer **3** had already been reported and characterized via dichloro-dicyano-benzoquinone (DDQ) oxidation of the β-5 dimer of coniferyl alcohol (Lahive et al., 2018). We therefore synthesized **3** in the same manner (Figure 4). The data matched perfectly and it was undoubtedly the same benzofuran. That dimer **3** results from dehydrogenation of coniferaldehyde (Figure 2) is also consistent with its synthesis from β-5-dehydroconferyl alcohol **7** by dehydrogenation/oxidation via DDQ (Figure 4) (Lahive et al., 2018). DDQ is useful for oxidizing benzylic and allylic alcohols to aldehydes (Alsharif et al., 2021), and was therefore expected to oxidize the phenylcoumaran **7** (Figure 4) to compound **2** (Figure 2). The direct production of the required benzofuran **3** is now eminently understandable as a product of a further dehydrogenation of phenylcoumaran **2** via radical disproportionation as noted for the peroxidase-H_2O_2 dehydrogenation in Figure 2.

DDQ-oxidation was further reported to produce the benzofuran moiety in lignin. However, this component has not been previously noted in native lignins, i.e., non-chemically-oxidized lignins. As there are important biosynthetic ramifications for these units, particularly in lignins from CAD-deficient plants, and as these structures represent a new sub-pathway in lignification, it is particularly important to identify and authenticate its presence in lignins.

Benzofurans in poplar xylem metabolites

Following the isolation and synthesis of benzofuran dimer **3**, we searched for benzofuran-containing oligomers in metabolic extracts. We routinely profile oligolignols from poplar xylem utilizing ultra-performance liquid chromatography–mass spectrometry (UPLC-MS) (Morreel et al., 2010). In the extract from a WT poplar from prior studies (Van Acker et al., 2017; De Meester et al., 2022), a minor peak with low intensity emerged, sharing the same retention time and m/z value as dimeric compound **3** (Supplemental Figure S2A). Limited signal intensity hindered the generation of an MS/MS spectrum to ascertain the veracity of this compound. Nonetheless, the analogous coupling product between

sinapaldehyde and coniferyl alcohol, labeled as **3**$_{S'\text{-}G^a}$ in Supplemental Figure S2B, and two distinct isomers of the trimer derived from it following further extension by coupling with a coniferyl alcohol radical, i.e. **G**(8-O-4)**3**$_{S'\text{-}G^a}$, were putatively identified in the xylem extract based on their accurate masses and MS/MS fragmentation (Supplemental Figure S2C, D). All of these putative benzofuran-containing components are slightly elevated in CAD-deficient transgenic poplars from earlier studies (Van Acker et al., 2017; De Meester et al., 2022). Absolute structural confirmation awaits the synthesis of the authentic reference compounds.

Aldehyde NMR profiles in the DHPs and lignins

The aldehyde region of HSQC NMR spectra is ideal for profiling the types of aldehydic components in lignins (Figure 3). In addition to the hydroxybenzaldehyde (**G"**/**S"**) and hydroxycinnamaldehyde (**G'**/**S'**) endgroups that are well-established in lignins, 8-8-dimerization results in structures **C'**, and 8-O-4-cross-coupling of hydroxycinnamaldehyde radicals with the radical from the phenolic end of the growing polymer produces structures **A'** in the lignin. The latter have been increasingly noted in most lignins since their discovery and structural elucidation from studies on CAD-deficient mutants and transgenics (Kim et al., 2003). Analytical thioacidolysis-generated CAD-markers result from these ether structures **A'** (Lapierre et al., 2004). Incidentally, the spectrum from the CAD-deficient mutant pine (Figure 3C) features an **A'** peak that, in other spectra, is labeled as **A'**$_{G'/S'\text{-}S/S'}$ but there are no S units in pine lignins. The analogous peak in the pine CAD spectrum (Figure 3C) arises from 8-O-4-coupling of a hydroxycinnamaldehyde with a 5-linked G unit, i.e., in structures labeled A'$_{G'\text{-}(5\text{-linked-}G'/G)}$. We have previously noted how 5-linked G-units can mimic S units in their reactivity. A relevant example is the discovery that 8-O-4-cross-coupling between coniferaldehyde and a guaiacyl unit does not readily occur, *in vivo* or *in vitro*, explaining why coniferaldehyde incorporation into CAD-deficient pine lignins is low, whereas the analogous cross-coupling of coniferaldehyde with an S unit is favored such that coniferaldehyde readily incorporates into G/S lignins (Ralph et al., 2001; Kim et al., 2003; Anderson et al., 2015).

In Figure 3A we show the region for the benzofuran dimer **3** isolated as described above (Figure 2) but also independently synthesized (Figure 4), along with that from a coniferaldehyde DHP (Figure 3B), and from a selection of lignins (Figures 3C-I) from various studies involving our labs over many years. The NMR spectrum from the benzofuran dimer conforms with its lignin counterpart. Slight peak offsets result from dimer **3**'s free-phenolic nature *vs* the usually-etherified corresponding units in lignin. Small NMR chemical shift variations are seen with DMSO-d_6/pyridine-d_5 as an NMR solvent due to the deviations from exactly 4:1 v/v, from the matrix and differing solute concentrations, and from the differing amounts of the water in the samples. The natural mulberry CAD-deficient mutant in which this

component most recently came to our attention (Figure 3D) along with lignins from studies involving various CAD-deficient mutants and transgenics are shown (Figures 3C-F). Then, to make it clear that similar structures are evident in lignins from wild-type plants, the bottom row shows the same regions for the control poplar (Figure 3G) from the studies that produced the CAD-deficient poplar in Figure 3E (Van Acker et al., 2017; De Meester et al., 2022), and the rather interesting observations noted below from two grasses, maize (Figure 3H) and ryegrass (Figure 3I).

Quantifying the degree of hydroxycinnamaldehyde incorporation into the lignin polymer backbone is currently difficult because of the number of structures involved and the peak overlap in NMR spectra. A useful measure is from the unique peak in the aromatics region arising from sinapaldehyde that has 8-O-4-coupled into lignin (Supplemental Figure S3). These $\mathbf{A'}_{S'}$ levels are reported on an S+G basis and are shown in the yellow box at the bottom-right of each spectrum in Figure 3. A slightly different method was required for the G-lignin from pine, as explained in the caption to Supplemental Figure S3.

As readily noted in the spectra in Figure 3, there may be two well-resolved correlation clusters for structures **B***. In principle, four major moieties should occur from the 8-5-coupling of either coniferaldehyde or sinapaldehyde with either coniferaldehyde or a generic **G** unit, i.e., in our shorthand, structures $\mathbf{B^*}_{G'\text{-}G'}$, $\mathbf{B^*}_{G'\text{-}G}$, $\mathbf{B^*}_{S'\text{-}G'}$, and $\mathbf{B^*}_{S'\text{-}G}$ (Figure 2). Another distinction could be those units with a coniferyl alcohol $\mathbf{G^a}$ endgroup, structures $\mathbf{B^*}_{G'\text{-}G^a}$, and $\mathbf{B^*}_{S'\text{-}G^a}$, but those are also indistinguishable from structures involving a generic **G** unit. Obviously, 8-5-coupling is not possible with an S unit of any kind as the 5-position is occupied by a methoxy group. It is apparent from the distributions of the two resolved **B*** contours in highly CAD-deficient plants, in which hydroxycinnamaldehydes predominate, *vs* those in wild-type plants in which they are minor, that the upper peak is from coupling with an aldehyde **G'** (or **G"** from vanillin) unit, whereas the lower one is from coupling with the generic **G** unit that has no conjugation to an aldehyde. This is the main resolvable distinction resulting in the two correlation peaks. Close examination of the data processed under resolution-enhancing conditions suggests that coniferaldehyde-derived $\mathbf{B^*}_{G'\text{-}G'}$ peaks may have slightly lower carbon and proton chemical shifts than sinapaldehyde-derived $\mathbf{B^*}_{S'\text{-}G'}$ ones, but the peaks are not resolved and the correlations are therefore labeled as combined $\mathbf{B^*}_{G'/S'\text{-}G'}$, i.e., as resulting from either hydroxycinnamaldehyde having coupled with a coniferaldehyde or vanillin unit for the dicots and monocots here (Figures 3D-I). Similarly, the lower peak deriving from the 8-5-coupling of either hydroxycinnamaldehyde with a generic **G** unit cannot be resolved into $\mathbf{B^*}_{G'\text{-}G}$ and $\mathbf{B^*}_{S'\text{-}G}$, so this correlation cluster is labeled as the combined $\mathbf{B^*}_{G'/S'\text{-}G}$, for the dicots and monocots (Figures 3D-I).

A single new NMR correlation peak, even though it does have two simultaneous constraints in that both the ^{1}H and the ^{13}C chemical shifts have to match those in the component of interest, is insufficiently diagnostic to confidently assign a new structure in a complex polymer such as lignin. The assignment

would be more compelling if there were other correlation peaks for the proposed component that simultaneously matched. This is one of the major advantages of NMR for structural assignments – if a structure is suspected because of a single correlation peak, all of the other carbon-proton pairs in that structure should also be apparent; should even one peak be missing, then it is almost certainly not the conjectured structure. However, in the case of the new **B*** units there are no other protonated sidechain carbons in the moiety that gives rise to the proton a_9 to carbon a_9 correlation peaks highlighted here in Figure 3 and shown for the isolated and synthesized model compound **3** (Figure 3A, Supplemental Figure S1). Correlation experiments using the proton coupling network such as COSY, TOCSY, and the usually powerful 2D HSQC-TOCSY experiment, therefore provide no additional data for authentication of units **B***. Long-range correlations over 2-3 bonds from proton a_9 (see Figure 2 and Supplemental Figure S1 for the labeling of the two moieties) would be particularly diagnostic. Such correlations could be expected from proton a_9 to carbons a_8 and a_7 and, crucially (to establish the radical coupling mode), carbon b_5 in model **3**. Observing the corresponding correlations for structures $\mathbf{B^*}_{G'/S'\text{-}G'}$ in a lignin would provide compelling evidence for the veracity of structures **B***. Such evidence for all three correlations is shown in the heteronuclear multiple-bond correlation (HMBC) spectra from the model **3** and in a lignin from CAD-deficient tobacco in Supplemental Figure S4. Again, unknown to us at the time, such structures **B*** were authenticated via HMBC spectra on a DDQ-oxidized lignin (Lahive et al., 2018); we additionally note the weak but also diagnostic proton a_9 to carbon a_7 correlation here (Supplemental Figure S4) that is not seen in their Figure 3.

Implications of finding benzofuran moieties B* in lignins

Regardless of whether a hydroxycinnamaldehyde 8-5-couples with a coniferaldehyde (or a similarly conjugated unit in the lignin) or a generic **G** unit, the product of that coupling is the usual phenylcoumaran **B'** (Figure 2). The endgroup aldehyde correlation would be in the green **G'/S'** NMR contour (Figure 3), but the correlation from the aldehyde in the phenylcoumaran moiety **B'** itself, has a much higher carbon shift (197.6 ppm in acetone-d_6 for the authentic coniferaldehyde dimer $\mathbf{2}_{G'\text{-}G'}$) and is extremely weak in actual lignins (Kim et al., 2003). Supplemental Figure S5 shows plots of the extended aldehyde regions for all 9 samples used in Figure 3. A small **B'** peak is seen in spectra from the coniferaldehyde DHP (Supplemental Figure S5B) and perhaps the CAD-deficient pine (Supplemental Figure S5C), both of which are estimated to be ~4% of the total **B'**+**B*** level, but cannot be discerned above the noise level in any of the other spectra. It is for this reason that this extended aldehyde region was not plotted in the main Figure 3.

The formation of structures **B*** implies that phenolic **B'** phenylcoumaran end-units not only form radicals but must have sufficient lifetimes to then encounter others radicals to undergo preferential

disproportionation reactions before the product benzofurans **B*** themselves undergo radical coupling to extend the polymer and bury the **B*** unit within the lignin chain, i.e., the phenylcoumaran units **B'** (or the model **2**) undergo radical disproportionation that competes with radical coupling. We have previously evidenced this competition between radical coupling and radical transfer to a more stable unit to explain how *p*-coumarate units remain free-phenolic (unetherified) in grass lignin polymers that derive, in part, from monolignol *p*-coumarate conjugates. Although *p*-coumarates form radicals, they preferentially undergo radical transfer over radical coupling and, in a radical-limited system, essentially all remain as pendent free-phenolic units (Takahama and Oniki, 1997; Ralph et al., 2004a; Hatfield et al., 2008; Ralph, 2010; Ralph et al., 2023a). In synthetic lignification with an overabundance of radical-generating capability, *p*-coumarate radicals will readily form radical-coupling products and become integrated more fully into the polymer backbone. Ferulate units may also favor this radical-transfer pathway. Although ferulates on arabinoxylan will dehydrodimerize (and form higher oligomers) by radical coupling, and will similarly cross-couple with monolignols and oligolignols to incorporate into the polymer backbone (Ralph, 2010), a substantial proportion of the ferulate on monolignol ferulate conjugates that are involved in lignification also remain as pendent units (Fanelli et al., 2021; Smith et al., 2022).

Not all observations from the lignin spectra have a ready explanation. Normal maize has, as would be expected because of the low level of hydroxycinnamaldehydes incorporated, most of its **B*** component as the one arising initially from coupling of a hydroxycinnamaldehyde with a generic **G** unit, i.e. **B***$_{G'/S'-G}$. Only a small amount of its **B*** component is derived from the analogous coupling with another coniferaldehyde (or at least another conjugated, unsaturated unit), i.e., **B***$_{G'/S'-G'}$. It is not clear why ryegrass had predominantly the **B***$_{G'/S'-G'}$ correlation peak (Figure 3I, Supplemental Figure S5I), making its profile appear more like those from CAD-deficient plants. The S/G is lower in ryegrass (0.78 vs 1.7 in the maize, Supplemental Figure S3), maize has nearly 5-fold higher levels of lignin derived from sinapyl *p*-coumarate conjugates, but ryegrass has over four-fold higher levels of tricin. None of these factors rationalize the CAD-deficient profile, but we note (Figure 3I, Supplemental Figure S5I), that the **A'** levels are also higher, and the usually extremely weak or undetectable **A'**$_{G'/S'-S/S'}$ is also higher in this material. Whether this apparent CAD limitation is peculiar to the sampled line (Ralph et al., 1995), or ryegrass in general, may be interesting for future enquiry.

Conclusions

Hydroxycinnamaldehyde-derived phenylcoumaran structures **B'** are difficult to detect in lignins (Kim et al., 2003) (Supplemental Figure S5). The new benzofuran products **B*** here are, however, readily evidenced in essentially any lignin NMR spectrum, Figure 3, from normal as well as at enhanced levels from CAD-deficient plants. It can be concluded that the phenolic radicals from phenylcoumaran

structures **B'** preferentially undergo disproportionation reactions rather than undergoing radical coupling themselves (Figure 2). The products **B*** then apparently readily undergo radical coupling reactions, incorporating such moieties into the backbone of the polymer. Products $\mathbf{B^*}_{G'\text{-}G'}$ arising from the homodimerization of coniferaldehyde, the presumed cross-coupling of sinapaldehyde with coniferaldehyde (i.e., units $\mathbf{B^*}_{S'\text{-}G'}$), and the cross-coupling of either sinapaldehyde or coniferaldehyde with a generic G unit (i.e., units $\mathbf{B^*}_{G'/S'\text{-}G}$), all appear to be viable pathways to produce the **B*** units in lignins. A major implication from this work is that we can now rather reliably assign one more set of correlation peaks in NMR spectra. Such hard-won assignments represent a significant challenge but become valuable additions for lignin characterization and evaluation in future studies that may also elucidate how these **B*** moieties will behave under various biomass processing conditions.

Materials and methods

General

Acetone-d_6, dimethyl sulfoxide-d_6, (DMSO-d_6) pyridine-d_5, 4-hydroxy-3-methoxycinnamaldehyde (coniferaldehyde), horseradish peroxidase (Sigma Type VI, 250 units/mg), and all the solvents were purchased from Sigma-Aldrich (St. Louis, MO, U.S.A.). Ethyl ferulate was obtained from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). Analytical thin-layer chromatography (TLC) was performed on Merck TLC silica gel 60 F_{254} aluminum sheets. Preparative TLC was conducted on Merck PLC silica gel plates 60 F_{254} (1 mm thickness) and UNIPLATE™ Analtech Silica gel GF (1 mm thickness). Silica gel column chromatography was carried out on an Isolera One (Biotage, Charlottesville, VA) equipped with Biotage silica gel cartridges (SNAP KP-Sil, SNAP Ultra or Sfär Silica D).

NMR

NMR spectra were obtained on a Bruker Biospin (Rheinstetten, Germany) NEO 700-MHz spectrometer fitted with a cryogenically-cooled 5-mm quadruple-resonance (QCI) $^{1}H/^{31}P/^{13}C/^{15}N$ gradient probe with inverse geometry, or an AVANCE 500-MHz spectrometer fitted with a cryogenically-cooled 5-mm Triple Resonance Inverse (TCI) $^{1}H/^{13}C/^{15}N$ gradient probe with inverse geometry. $^{1}H–^{13}C$ HSQC experiments were carried out and processed as reported previously (Kim and Ralph, 2010; Mansfield et al., 2012). Chemical shifts (δ) are given in ppm relative to (residual) solvent peaks for acetone (δ_H 2.04, δ_C 29.8) or DMSO (δ_H 2.49, δ_C 39.50).

High-resolution mass spectrometry

Mass spectra of isolated and synthetic compounds were acquired on a quadrupole time-of-flight mass spectrometer, Impact II (Bruker Daltonics, Bremen, Germany) equipped with Shimadzu Nexera X2

system (Kyoto, Japan) using Kinetex XB-C18 column (50 × 2.1 mm, Particle diameter: 1.7 μm, 100 Å) (Phenomenex, Torrance, CA, U.S.A.).

UPLC-MS analysis of benzofuran-containing metabolites in poplar xylem

The stem (~20 cm) from greenhouse-grown *Populus tremula* × *alba*, both the WT control and a CAD-deficient line, was harvested, debarked and flash-frozen. Frozen milled wood (30 mg) was extracted with 1 mL methanol (15 min, ambient temperature). The extract was dried in a speed-vac, and resuspended in 100 μL water/methanol 90/10 (v/v) and then transferred to a solid-phase extraction (SPE) Extract Clean™ SPE C18 50 mg column (S*Pure). After the sample solvent was eluted from the cartridge, an additional 1 mL water/methanol 95/5 (v/v) was added. The combined eluates were collected and dried in a speed-vac. The pellet was resuspended in 100 μL water/methanol 90/10 (v/v) and subjected to UPLC-MS at the VIB Metabolomics Core Ghent (VIB-MCG). An aliquot (10 μL) was injected on a Waters Acquity UHPLC device connected to a Vion HDMS Q-TOF mass spectrometer (Waters, Manchester, UK), with settings as previously reported (Van Acker et al., 2017). Data recording and processing were performed using Unifi Workstation v2.0 (Waters) and Progenesis QI v2.4 (Waters).

Preparation of coniferaldehyde dehydrogenation polymer (DHP) (Figure 3B)

Three different solutions A, B, and C were prepared for coniferaldehyde-DHP synthesis. Solution A: Coniferaldehyde (178 mg) was dissolved in acetone (24 mL) and 0.1 M Na-phosphate buffer (216 mL, pH = 6.5). Solution B: Thirty percent hydrogen peroxide aqueous solution (118 μL, 1.2 mmol) was dissolved in Milli-Q water (240 mL). Solution C: Horseradish peroxidase (5 mg) in 0.1 M Na-phosphate buffer (60 mL). Solutions A and B were co-added to solution C via an ISMATEC IPC peristaltic pump at a flow rate of 0.2 mL/min over 20 h. The reaction mixture was stirred for an additional 4 h. The reaction mixture was transferred into 50 mL Falcon tubes and centrifuged for 10 min at 8,000 *g* using a Thermo Scientific Sorvall Primo centrifuge. Precipitates were collected and washed with water (100 mL × 3). After the precipitates were lyophilized, the DHP was obtained as a red solid (150 mg, 85%). The DHP (15 mg) was dissolved in 0.75 mL of DMSO-d_6/pyridine-d_5 (4:1, v/v) for various NMR measurements.

Preparation of coniferaldehyde dehydrogenation products

Coniferaldehyde (2.5 g, 14 mmol) in acetone (10 mL) was added to distilled water (250 mL) in a 1 L flask. After the addition of horseradish peroxidase (type II, 0.5 mg, 150-250 U/mg) in water (13 mL) into the flask, the hydrogen peroxide solution (3%, 10 mL) was added into the mixture. The mixture was stirred for an additional 3 h. The yellow solution changed to red-brown during the reaction. The reaction mixture was extracted with EtOAc (150 mL × 4), and the combined organic layer was dried over

anhydrous Na_2SO_4 to give crude products (2.51 g). The crude products were dissolved in DMSO-d_6 and assayed for the compounds of interest by NMR.

Purification of the aldehyde dimer 3, a B*$_{G'-G'}$ model, from coniferaldehyde dehydrogenation products

The crude products (2.51 g) were fractionated on an Isolera One flash column chromatography instrument with hexane and EtOAc via gradient elution. Fractions obtained were dried *in vacuo* and assayed by NMR to check for the peaks of interest at δ_C/δ_H 186.7/10.24 ppm in the HSQC spectrum. Fraction 3 (670 mg), Fraction 4 (438 mg), Fraction 5 (559 mg), Fraction 6 (204 mg) and Fraction 7 (85 mg) included the peak. Fraction 5 (100 mg) was further purified using preparative TLC plates with EtOAc:hexane (2:1, v/v) and EtOAc:hexane (3:1, v/v). The new aldehyde dimer (1.3 mg) was obtained, i.e., in 13% yield from Fraction 5; we made no attempt to estimate the yield across all fractions and these certainly included oligomers larger than this dimer.

The NMR data in four solvents, $CDCl_3$, acetone-d_6, DMSO-d_6, and 4:1 v/v DMSO-d_6:pyridine-d_5 (as here, to match the solvent used for the lignins), are deposited in Record #3073 in the latest release of the "NMR database of lignin and cell wall model compounds" (Ralph et al., 2023b). Benzofuran dimer **3**, (*E*)-2-(4-hydroxy-3-methoxyphenyl)-7-methoxy-5-(3-oxoprop-1-en-1-yl)benzofuran-3-carbaldehyde, NMR (700 MHz, 4:1 DMSO-d_6/pyridine-d_5): δ_H 3.79 (3H, s, a_3-OMe), 3.95 (3H, s, b_3-OMe), 6.90 (1H, dd, J = 15.8, 7.7 Hz, b_8), 6.97 (1H, d, J = 8.2 Hz, a_5), 7.39 (1H, dd, J = 8.2, 2.1 Hz, a_6), 7.42 (1H, d, J = 1.4 Hz, b_2), 7.44 (1H, d, J = 2.1 Hz, a_2), 7.78 (1H, d, J = 15.8 Hz, b_7), 7.96 (1H, d, J = 1.3 Hz, b_6), 9.63 (1H, d, J = 7.7 Hz, b_9), 10.20 (1H, s, a_9).; δ_C 55.8 (a_3-OMe), 56.2 (b_3-OMe), 107.1 (b_2), 112.4 (a_2), 115.4 (b_5), 116.0 (b_6), 116.2 (a_5), 118.5 (a_1), 122.9 (a_6), 127.4 (a_8), 128.4 (b_8), 132.3 (b_1), 143.8 (b_4), 145.1 (b_3), 148.2 (a_3), 150.5 (a_4), 153.6 (b_7), 165.8 (a_7), 186.6 (a_9), 194.2 (b_9); HRMS: *m/z* $[M-H]^-$ calculated for $C_{20}H_{15}O_6$: 351.0869, found: 351.0871.

Synthesis of the aldehyde dimer 3, a B*$_{G'-G'}$ model

The authentic dimeric compound **3** (Figures 3A, 4, and Supplemental Figure S1) was also independently synthesized (Figure 4, in which yields for the 3 steps are provided) for additional confirmation, for use as a model, and to obtain NMR data. First, (β-5)-dehydrodiconiferyl alcohol **7** was synthesized from ethyl ferulate **5** via its radical coupling product diethyl (8-5)-dehydrodiferulate **6**, as it has been numerous times in our group (Ralph et al., 1994; Ralph et al., 1998a), and most recently in a paper reassessing the claimed cytokinin-substituting activity of dehydrodiconiferyl alcohol glucoside (Witvrouw et al., 2023).

The conversion of **7** to the required compound **3** was achieved in a single oxidative step as in a published method (Lahive et al., 2018), producing both conjugated aldehyde moieties. 1,4-Dioxane (0.84

mL) was added into (β-5)-dehydrodiconiferyl alcohol (**7**, 10 mg, 28 μmol) and DDQ (19 mg, 84 μmol) in a 15 mL ace glass pressure tube (Ace glass, Inc., Vineland, NJ). The mixture was stirred for 24 h at 60 °C. The reaction mixture was evaporated to yield the crude product. The residue was purified by preparative TLC, yielding compound **3** (1.2 mg, 12%, Figure 4). No attempt was made to optimize yields as we only needed the authentic compound for structural characterization and for its NMR data.

Benzofuran dimer **3**: data were identical to those provided above (for the isolated compound).

Plant materials and extract-free (EF) wood preparations

All but one of the plant materials used here were from previous studies over many years: A pine CAD mutant (Ralph et al., 1997); Sekizaisou, the mulberry CAD mutant (Yamamoto et al., 2020); a CAD-deficient transgenic poplar and its wild-type control (Van Acker et al., 2017; De Meester et al., 2022); a CAD-deficient tobacco transgenic grown in 10% $^{13}CO_2$ (Ralph et al., 1998b); ryegrass grown in 10% $^{13}CO_2$ (Ralph et al., 1995). The maize sample was from IsoLife, Wageningen, the Netherlands, P-60501 P-^{13}C Maize stem (10 atom% ^{13}C).

Enzyme lignin (EL) preparation

A fraction of the dried EF wood (750 mg) was ball-milled in 20 mL agate jars with 10 × 10 mm agate ball-bearings using a Fritsche Pulverisette 7 planetary ball mill operating at 600 rpm for 35 cycles of 10-min grinding, followed by 5-min breaks to avoid excessive heating. The ball-milled samples (650 mg) were suspended in 25.5 mM acetate buffer pH 5.0 (45 mL), crude cellulases (Cellulysin, Calbiochem, 20 mg) were added, and the samples were shaken at 250 rpm for 3 days at 35 °C. They were then pelletized (10 min at 1777 rcf on an Eppendorf 5810R). The acetate buffer was decanted, and the enzymatic digestion was repeated. After two digestion cycles, the samples were washed three times by suspending the solids in the water, pelleting, and decanting the wash water. Finally, the samples were lyophilized to produce enzyme lignin (EL).

2D HSQC (^{1}H–^{13}C) and HMBC NMR spectroscopy of isolated enzyme lignins

The 2D HSQC NMR experiments (Kupče and Freeman, 2007) were performed on enzymatically isolated lignins (EL) as previously described (Kim et al., 2008). The lignin (10–20 mg) was dissolved in 500 μL of DMSO-d_6/pyridine-d_5 (4:1, v/v) for NMR. The peak assignments were performed manually based on previously reported correlation peaks (Kim and Ralph, 2010; Kim et al., 2017). Correlation peaks are color-coded for the structures shown. The HMBC spectrum of the tobacco lignin in DMSO-d_6/pyridine-d_5 (4:1, v/v) for Supplemental Figure S4 was used a long-range coupling delay of 65 ms (corresponding to a long-range $J_{C–H}$ of 7.7 Hz).

Supplemental Materials

The following materials are available in the online version of this article.

Shorthand formalism. An expanded explanation of the shorthand formalism used for the various structures.

Supplemental Figure S1 NMR spectra for the benzofuran dimer.

Supplemental Figure S2 Negative ionization mode MSMS spectra of **3** and three metabolites in poplar xylem that are likely benzofuran-containing.

Supplemental Figure S3 Partial aromatics regions for all samples in Figure 3.

Supplemental Figure S4 HMBC data establishing the benzofuran structure in lignin.

Supplemental Figure S5 Extended aldehyde regions from the HSQC spectra in Figure 3.

Acknowledgements

We thank Steve Karlen (WEI) for the hi-res MS and for valuable discussions, Sarah Lu (WEI) for the preparation of recent samples, and Geert Goeminne of the VIB Metabolomics Core Ghent (VIB-MCG) for the UPLC-MS analysis. We are also grateful to many collaborators over the years who provided the original plant materials used in the cited works (in the Plant Materials section) from which spectral data were mined for Figure 3 and Supplemental Figures S3 and S5.

Funding

KY, HK, FL, and JR were supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. SK was partially supported by the Japan Science and Technology Agency, and by the Japan Society for the Promotion of Science. NDR was supported by the Research foundation Flanders (FWO) for a predoctoral fellowship. WB acknowledges support from FWO project G008116N and from the interuniversity Bijzonder Onderzoeksfonds iBOF “Next-Bioref” (BOFIBO2021000902-iBOF).

Author contributions

All authors designed and performed the research; KY, HK, and FL performed the synthetic work and, with JR, analyzed the data and wrote the article; NDR and RV screened the poplar metabolites. All authors contributed to and proofed the manuscript.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is John Ralph.

References

Adler E, Björkquist KJ, Häggroth S (1948) Über die Ursache der Farbreaktionen des Holzes. Acta Chemica Scandinavica **2:** 93-94

Alsharif MA, Raja QA, Majeed NA, Jassas RS, Alsimaree AA, Sadiq A, Naeem N, Mughal EU, Alsantali RI, Moussa Z, Ahmed SA (2021) DDQ as a versatile and easily recyclable oxidant: a systematic review. RSC Adv **11:** 29826-29858

Anderson NA, Tobimatsu Y, Ciesielski PN, Ximenes E, Ralph J, Donohoe BS, Ladisch M, Chapple C (2015) Manipulation of guaiacyl and syringyl monomer synthesis in an Arabidopsis Cinnamyl Alcohol Dehydrogenase mutant results in atypical lignin biosynthesis and modified cell wall structure. Plant Cell **27:** 2195-2209

Barrière Y, Ralph J, Méchin V, Guillaumie S, Grabber JH, Argillier O, Chabbert B, Lapierre C (2004) Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants. Comptes Rendus Biologies **327:** 847-860

Chabannes M, Barakate A, Lapierre C, Marita J, Ralph J, Pean M, Danoun S, Halpin C, Grima-Pettenati J, Boudet A-M (2001) Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. The Plant Journal **28:** 257-270

De Meester B, Van Acker R, Wouters M, Traversari S, Steenackers M, Neukermans J, Van Breusegem F, Dejardin A, Pilate G, Boerjan W (2022) Field and saccharification performances of poplars severely downregulated in CAD1. New Phytologist **236:** 2075-2090

Fanelli A, Rancour DM, Sullivan M, Karlen SD, Ralph J, Riaño-Pachón DM, Vicentini R, da Franca Silva T, Ferraz AL, Hatfield RD, Romanel E (2021) Overexpression of a sugarcane BAHD acyltransferase alters hydroxycinnamate content in maize cell wall. Frontiers in Plant Science **12:** 626168: 1-14

Freudenberg K, Neish AC (1968) Constitution and Biosynthesis of Lignin. Springer-Verlag, Berlin-Heidelberg-New York

Hatfield RD, Ralph J, Grabber JH (2008) A potential role of sinapyl *p*-coumarate as a radical transfer mechanism in grass lignin formation. Planta **228:** 919-928

Higuchi T, Ito T, Umezawa T, Hibino T, Shibata D (1994) Red-brown color of lignified tissues of transgenic plants with antisense CAD gene: Wine-red lignin from coniferyl aldehyde. Journal of Biotechnology **37:** 151-158

Kim H, Padmakshan D, Li Y, Rencoret J, Hatfield RD, Ralph J (2017) Characterization and elimination of undesirable protein residues in plant cell wall materials for enhancing lignin analysis by solution-state NMR. Biomacromolecules **18:** 4184–4195

Kim H, Ralph J (2010) Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d_6/pyridine-d_5. Organic & Biomolecular Chemistry **8:** 576-591

Kim H, Ralph J, Akiyama T (2008) Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d_6. BioEnergy Research **1:** 56-66

Kim H, Ralph J, Lu F, Pilate G, Leplé JC, Pollet B, Lapierre C (2002) Identification of the structure and origin of thioacidolysis marker compounds for cinnamyl alcohol dehydrogenase deficiency in angiosperms. Journal of Biological Chemistry **277:** 47412-47419

Kim H, Ralph J, Lu F, Ralph SA, Boudet A-M, MacKay JJ, Sederoff RR, Ito T, Kawai S, Ohashi H, Higuchi T (2003) NMR Analysis of Lignins in CAD-deficient Plants. Part 1. Incorporation of hydroxycinnamaldehydes and hydroxybenzaldehydes into lignins. Organic and Biomolecular Chemistry **1:** 268-281

Kupče E, Freeman R (2007) Compensated adiabatic inversion pulses: Broadband INEPT and HSQC. Journal of Magnetic Resonance **187:** 258-265

Lahive CW, Lancefield CS, Codina A, Kamer PCJ, Westwood NJ (2018) Revealing the fate of the phenylcoumaran linkage during lignin oxidation reactions. Organic & Biomolecular Chemistry **16:** 1976-1982

Lapierre C, Pilate G, Pollet B, Mila I, Leplé JC, Jouanin L, Kim H, Ralph J (2004) Signatures of cinnamyl alcohol dehydrogenase deficiency in poplar lignins. Phytochemistry **65:** 313-321

Liu X, Van Acker R, Voorend W, Pallidis A, Goeminne G, Pollier J, Morreel K, Kim H, Muylle H, Bosio M, Ralph J, Vanholme R, Boerjan W (2021) Rewired phenolic metabolism and improved saccharification efficiency of a *Zea mays cinnamyl alcohol dehydrogenase 2* (*zmcad2*) mutant. The Plant Journal **105:** 1240-1257

Mansfield SD, Kim H, Lu F, Ralph J (2012) Whole plant cell wall characterization using solution-state 2D-NMR. Nature Protocols **7:** 1579-1589

Morreel K, Kim H, Lu F, Dima O, Akiyama T, Vanholme R, Niculaes C, Goeminne G, Inzé D, Messens E, Ralph J, Boerjan W (2010) Mass-spectrometry-based fragmentation as an identification tool in lignomics. Analytical Chemistry **82:** 8095-8105

Ralph J (2010) Hydroxycinnamates in lignification. Phytochemistry Reviews **9:** 65-83

Ralph J, Brunow G, Harris PJ, Dixon RA, Schatz PF, Boerjan W (2008) Lignification: Are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template

replication? *In* F Daayf, A El Hadrami, L Adam, GM Ballance, eds, Recent Advances in Polyphenol Research, Vol 1. Wiley-Blackwell Publishing, Oxford, UK, pp 36-66

Ralph J, Bunzel M, Marita JM, Hatfield RD, Lu F, Kim H, Schatz PF, Grabber JH, Steinhart H (2004a) Peroxidase-dependent cross-linking reactions of *p*-hydroxycinnamates in plant cell walls. Phytochemistry Reviews **3:** 79-96

Ralph J, Garcia-Conesa MT, Williamson G (1998a) Simple preparation of 8-5-coupled diferulate. Journal of Agricultural and Food Chemistry **46:** 2531-2532

Ralph J, Grabber JH, Hatfield RD (1995) Lignin-ferulate crosslinks in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. Carbohydrate Research **275:** 167-178

Ralph J, Hatfield RD, Piquemal J, Yahiaoui N, Pean M, Lapierre C, Boudet A-M (1998b) NMR characterization of altered lignins extracted from tobacco plants down-regulated for lignification enzymes cinnamyl-alcohol dehydrogenase and cinnamoyl-CoA reductase. Proceedings of the National Academy of Sciences **95:** 12803-12808

Ralph J, Kim H, Lu F, Smith RA, Karlen SD, Nuoendagula, Yoshioka K, Eugene A, Liu S, Sener C, Ando D, Chen M-j, Li Y, Landucci LL, Ralph SA, Timokhin VI, Lan W, Rencoret J, del Río JC (2023a) Lignins and lignification: New developments and emerging concepts. *In* S Quideau, J-P Salminen, K Wähälä, V de Freitas, eds, Recent Advances in Polyphenol Research, Vol 8. Wiley-Blackwell, Oxford, UK, pp 1-50

Ralph J, Lapierre C, Boerjan W (2019) Lignin structure and its engineering. Current Opinion in Biotechnology **56:** 240-249

Ralph J, Lapierre C, Marita J, Kim H, Lu F, Hatfield RD, Ralph SA, Chapple C, Franke R, Hemm MR, Van Doorsselaere J, Sederoff RR, O'Malley DM, Scott JT, MacKay JJ, Yahiaoui N, Boudet A-M, Pean M, Pilate G, Jouanin L, Boerjan W (2001) Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. Phytochemistry **57:** 993-1003

Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, Boerjan W (2004b) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. Phytochemistry Reviews **3:** 29-60

Ralph J, MacKay JJ, Hatfield RD, O'Malley DM, Whetten RW, Sederoff RR (1997) Abnormal lignin in a loblolly pine mutant. Science **277:** 235-239

Ralph J, Quideau S, Grabber JH, Hatfield RD (1994) Identification and synthesis of new ferulic acid dehydrodimers present in grass cell walls. Journal of the Chemical Society, Perkin Transactions 1**:** 3485-3498

Ralph SA, Landucci LL, Ralph J (2023b) NMR database of lignin and cell wall model compounds. *In* Available over Internet at https://www.glbrc.org/databases_and_software/nmrdatabase/ (updated sporadically since 1993). S. A. Ralph, L. L. Landucci, J. Ralph

Sederoff RR, MacKay JJ, Ralph J, Hatfield RD (1999) Unexpected variation in lignin. Current Opinion in Plant Biology **2:** 145-152

Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Seguin A (2005) *Cinnamyl Alcohol Dehydrogenase-C* and *-D* are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. Plant Cell **17:** 2059-2076

Smith RA, Beebe ET, Bingman CA, Vander Meulen KA, Eugene A, Steiner A, Karlen SD, Ralph J, Fox BG (2022) Identification and characterization of a set of monocot BAHD monolignol transferases. Plant Physiology **189:** 37-48

Takahama U, Oniki T (1997) Enhancement of peroxidase-dependent oxidation of sinapyl alcohol by an apoplastic component, 4-coumaric acid ester isolated from epicotyls of *Vigna angularis* L. Plant Cell Physiology **38:** 456-462

Van Acker R, Déjardin A, Desmet S, Hoengenaert L, Vanholme R, Morreel K, Laurans F, Kim H, Santoro N, Foster C, Goeminne G, Légée F, Lapierre C, Pilate G, Ralph J, Boerjan W (2017) Different routes for conifer- and sinapaldehyde and higher saccharification upon deficiency in the dehydrogenase CAD1. Plant Physiology **175:** 1018-1039

Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W (2012) Metabolic engineering of novel lignin in biomass crops. New Phytologist **196:** 978-1000

Witvrouw K, Kim H, Vanholme R, Goeminne G, Ralph J, Boerjan W, Vanholme B (2023) Reassessing the claimed cytokinin-substituting activity of dehydrodiconiferyl alcohol glucoside. Proceedings of the National Academies of Sciences **120:** e2123301120: 1-7

Yamamoto M, Tomiyama H, Koyama A, Okuizumi H, Liu S, Vanholme R, Goeminne G, Hirai Y, Shi H, Nuoendagula, Takata N, Ikeda T, Uesugi M, Kim H, Sakamoto S, Mitsuda N, Boerjan W, Ralph J, Kajita S (2020) A century-old mystery unveiled: Sekizaisou is a natural lignin mutant. Plant Physiology **182:** 1821-1828

Yan X, Liu J, Kim H, Liu B, Huang X, Yang Z, Lin Y-C, Chen H, Yang C, Wang J, Muddiman D, Ralph J, Sederoff R, Li Q, Chiang V (2019) CAD1 and CCR2 protein complex formation in monolignol biosynthesis in *Populus trichocarpa*. New Phytologist **222:** 244-260

Zhao Q, Tobimatsu Y, Zhou R, Pattathil S, Gallego-Giraldo L, Fu C, Jackson LA, Hahn MG, Kim H, Chen F, Ralph J, Dixon RA (2013) Loss of function of *Cinnamyl Alcohol Dehydrogenase 1* causes accumulation of an unconventional lignin and a temperature-sensitive growth defect in

Medicago truncatula. Proceedings of the National Academy of Sciences of the United States of America **110:** 13660-13665

Supplementary Materials for
Hydroxycinnamaldehyde-derived benzofuran components in lignins

Koichi Yoshioka, Hoon Kim, Fachuang Lu, Nette De Ridder,
Ruben Vanholme, Shinya Kajita, Wout Boerjan, and John Ralph

TABLE OF CONTENTS

Shorthand formalism

In the shorthand formalism, **G** is used to designate guaiacyl units derived from lignification from the monolignol coniferyl alcohol, **S** for syringyl units from sinapyl alcohol, and **H** (not used in this paper) for *p*-hydroxyphenyl units derived from *p*-coumaryl alcohol.

For the monomers themselves, the guaiacyl monomer, coniferyl alcohol, is designated as **G**^{a}, the syringyl monomer, sinapyl alcohol, as **S**^{a}. A prime is used for the hydroxycinnamaldehyde monomers coniferaldehyde **G'** and sinapaldehyde **S'**. A double-prime is used for hydroxybenzaldehydes vanillin **G"** and syringaldehyde **S"**.

For the dimeric units with their characteristic interunit linkages, we use a bold **A** for β-ethers (β-O-4-linked = 8-O-4-linked; Greek letters are usually used for the monolignols and the units derived from them in lignins, whereas Arabic numerals 7-9 have traditionally been used for hydroxycinnamate and hydroxycinnamaldehyde sidechains), **B** for β-5/8-5-linked units, **C** for β-β/8-8-linked units and, not used in this paper, **D** for biphenyl (5-5-linked) units, **E** for β-1-linked units, and **F** for biphenyl ether (5-O-4-linked) units.

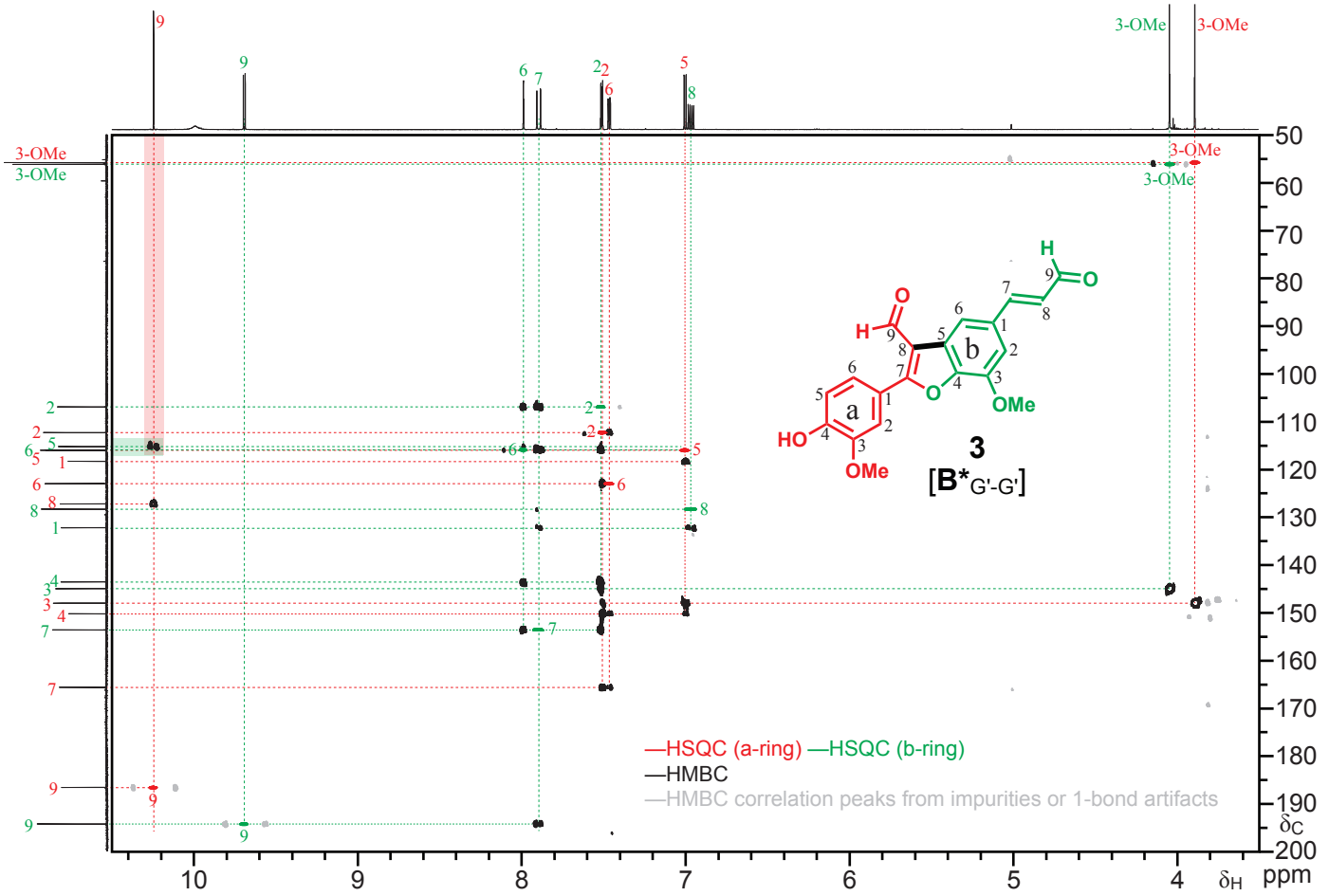
For the main dimeric moieties, we use the same convention to indicate a structural variation…

- **X'**: A prime is for units that derive from hydroxycinnamaldehydes.
  - **A'** for β-ether derived from 8-O-4-coupling of a hydroxycinnamaldehyde. These units have an unsaturated sidechain because rearomatization of the quinone methide intermediate produced by radical coupling involves loss of the acidic 8-proton (as opposed to water addition in the analogous QM derived from monolignol coupling).
  - **B'** for the phenylcoumaran derived from 8-5-coupling of a hydroxycinnamaldehyde.
  - **C'** designates the butadienes resulting from 8-8-coupling of hydroxycinnamaldehydes.
- **X***: An asterisk is used to designate that something quite unusual has occurred to produce a structural variant.
  - **B*** here when the normal phenylcoumaran ring in structures **B'** has been dehydrogenated (contains a double-bond) to a benzofuran. As the prime focus of this paper, the designation of 8-5-coupled units in lignins, the asterisk in **B*** indicates that it is a variant of the normal phenylcoumaran units **B** derived from the monolignols and the normal phenylcoumaran units **B'** derived from hydroxycinnamaldehydes.

Indicating the nature of each of the units involved in these various structures is via non-bolded subscripts. Examples using various **B** (8-5-coupled) units, also illustrated in Figure 1 in the main paper, illustrate this formalism…

- **B**$_{G\text{-}G}$ denotes the **G** nature of the two units involved in a normal phenylcoumaran (8-5-coupled) dimeric unit derived from cross-coupling coniferyl alcohol **G**^{a} with a generic **G** unit (at the phenolic end of the growing oligomer).
- **B**$_{S\text{-}G}$ is from the analogous cross-coupling of sinapyl alcohol **S**^{a} with a generic **G** unit.
- **B**$_{G\text{-}G^{a}}$ is from the analogous cross-coupling (dehydrodimerization) of coniferyl alcohol **G**^{a} with another coniferyl alcohol **G**^{a}.
- **B**$_{S\text{-}G^{a}}$ is from the analogous cross-coupling of sinapyl alcohol **S**^{a} with coniferyl alcohol **G**^{a}.
- **B'**$_{G'\text{-}G}$ is the 8-5-coupled unit derived from cross-coupling coniferaldehyde **G'** with a generic **G** unit; this still creates a normal phenylcoumaran structure but is denoted as **B'**$_{G'\text{-}G}$, because it results from hydroxycinnamaldehyde coupling.
- **B'**$_{S'\text{-}G}$ is from the analogous cross-coupling of sinapaldehyde with a generic **G** unit.
- **B'**$_{G'\text{-}G^{a}}$ is from the analogous cross-coupling of coniferaldehyde **G'** with coniferyl alcohol **G**^{a}.
- **B'**$_{S'\text{-}G^{a}}$ is from the analogous cross-coupling of sinapaldehyde **S'** with coniferyl alcohol **G**^{a}.
- **B'**$_{G'\text{-}G'}$ is from the analogous coupling of coniferaldehyde **G'** with another coniferaldehyde **G'** or a coniferaldehyde-derived unit.
- **B'**$_{S'\text{-}G'}$ is from the analogous cross-coupling of sinapaldehyde **S'** with coniferaldehyde **G'** or a coniferaldehyde-derived unit.
- **B***$_{G'\text{-}G}$ denotes the 8-5-coupled unit derived from cross-coupling coniferaldehyde **G'** with a generic **G** unit in which the initial phenylcoumaran coupling product **B'**$_{G'\text{-}G}$ has undergone a further dehydrogenation and then disproportionation to produce the benzofuran **B***.
  As above, for the structures that are the main topic of this paper, the **B** unit is no longer that standard 8-5-coupled phenylcoumaran unit **B**, nor is it the hydroxycinnamaldehyde-derived phenylcoumaran **B'** – it has undergone further transformation to a benzofuran such that it is labeled as a **B*** unit to clearly differentiate it. Such units initially derive from cross-coupling of coniferaldehyde **G'** with a generic **G** unit, followed by further dehydrogenation, would then be denoted as **B***$_{G'\text{-}G}$.
- **B***$_{S'\text{-}G}$ is from the analogous process involving sinapaldehyde **S'** and a generic **G** unit.
- **B***$_{G'\text{-}G'}$ is from the analogous process involving coniferaldehyde **G'** with another coniferaldehyde **G'** or a coniferaldehyde-derived unit.
- **B***$_{S'\text{-}G'}$ is from the analogous process involving sinapaldehyde **S'** with coniferaldehyde **G'** or a coniferaldehyde-derived unit.
- **B'**$_{G'\text{-}S}$, **B'**$_{S'\text{-}S}$, **B'**$_{G'\text{-}S^{a}}$, **B'**$_{G'\text{-}S'}$, **B'**$_{S'\text{-}S'}$, **B***$_{G'\text{-}S}$, **B***$_{S'\text{-}S}$, **B***$_{G'\text{-}S^{a}}$, **B***$_{G'\text{-}S'}$, **B***$_{S'\text{-}S'}$, are all impossible – coupling to the 5-position of a syringyl unit is impossible because of the methoxy substitution at carbon-5.

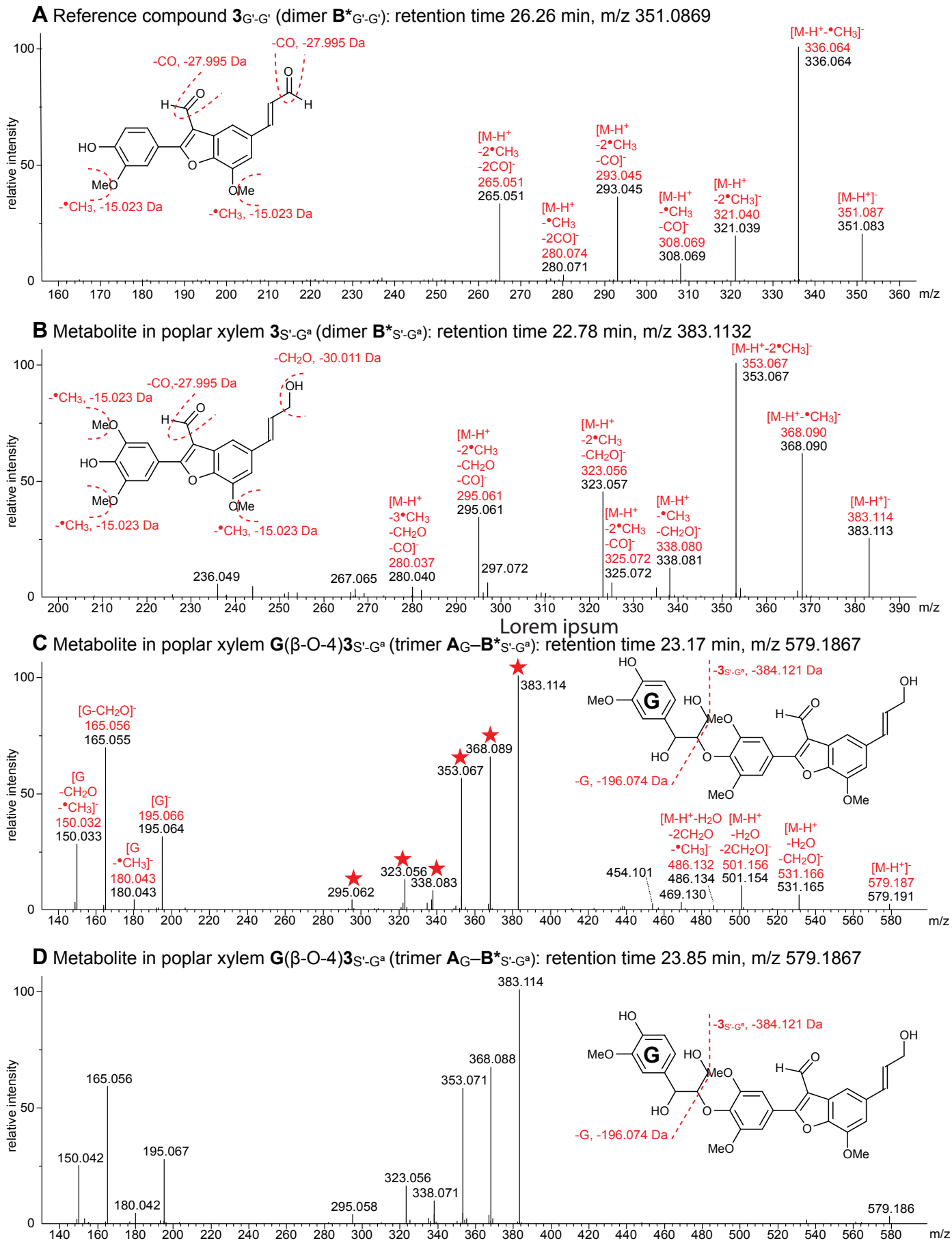
For the pendant units acylating lignin polymers, we simply use ***p*HB** for the *p*-hydroxybenzoates on poplar and ***p*CA** for the *p*-coumarates on the grasses. These features of some lignins are not discussed in this paper.



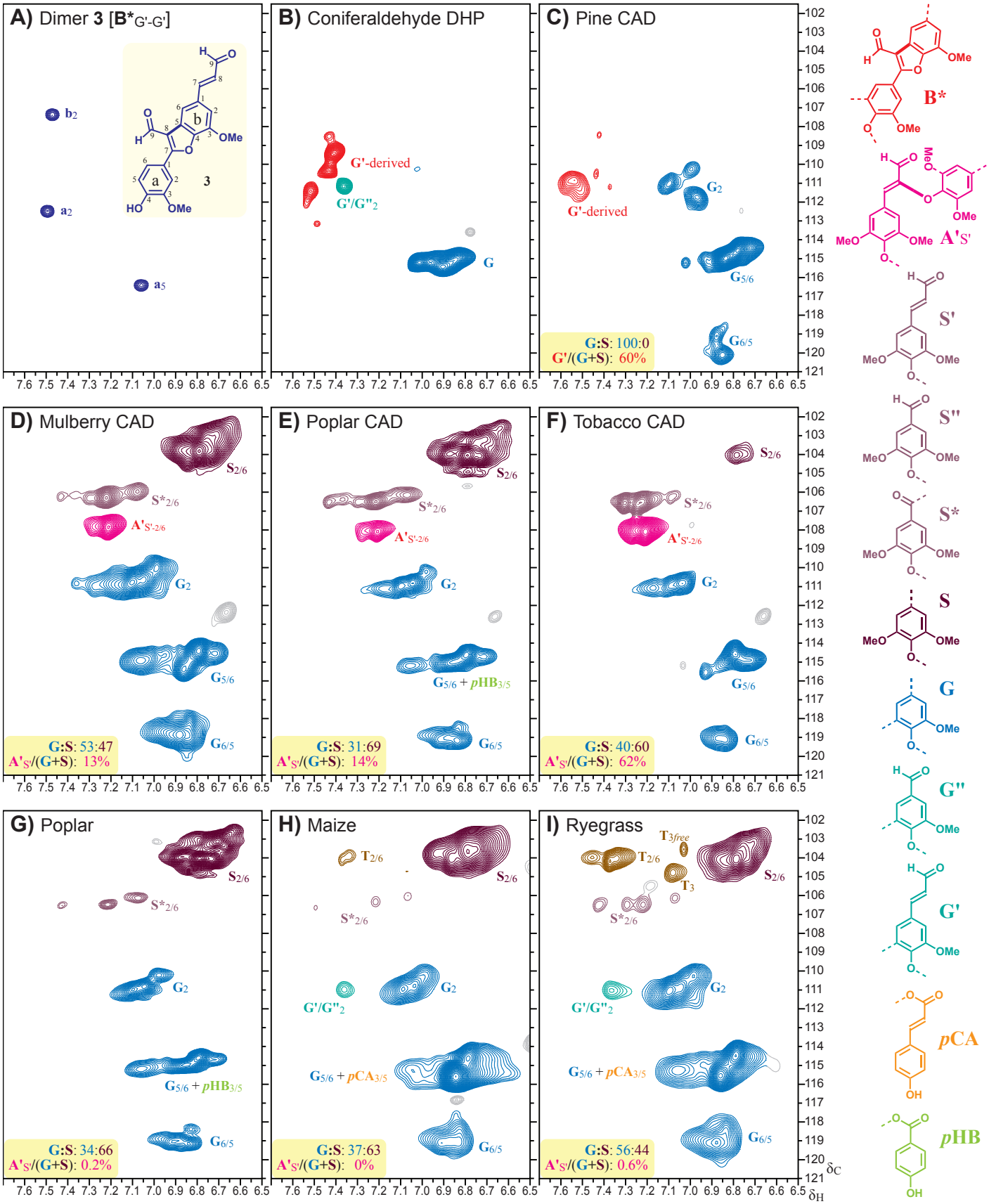
Supplemental Figure S1 ^{1}H–^{13}C correlation NMR spectra for the benzofuran dimer **3** (700 MHz, in DMSO-d_6). The HSQC spectrum (red and green contours and assignments) is overlaid (from a separate spectrum/experiment) on HMBC spectra (black correlations) showing colored correlation lines (between protons and carbons within 3 bonds of each other) for the a and b units (in red and green, as for the structure and assignments). Of major significance, as highlighted, is the correlation between proton a_9 and carbon b_5 establishing the 8-5-bonding between the two units. Contours colored light gray are from minor impurities or 1-bond-coupled peaks. The NMR data in four solvents, $CDCl_3$, acetone-d_6, DMSO-d_6 (as here), and 4:1 v/v DMSO-d_6:pyridine-d_5, are deposited in Record #3073 in the latest release of the "NMR database of lignin and cell wall model compounds" (Ralph et al., 2023).

Next page

Supplemental Figure S2 Negative-ionization mode MSMS spectra of **3** (a **B***$_{G'\text{-}G'}$ model) and three metabolites in poplar xylem that are likely benzofuran-containing. The red-font numbers represent the theoretical m/z values for the provided fragmentations, whereas the black-font numbers correspond to the experimentally-derived m/z values. For generating MS/MS spectra, electrospray ionization (ESI) was used under the following conditions: capillary voltage, 3 kV; source temperature, 120 °C; desolvation gas temperature, 550 °C; desolvation gas flow, 800 L h^1; and cone gas flow, 50 L h^1. The collision energy was ramped from 20 to 70 eV. The mass range was set from 50 to 1500 Da, and the scan time was 0.1 s. Nitrogen (greater than 99.5%) was employed as a desolvation and cone gas. A, MSMS spectrum of the aldehyde dimer **3** reference compound. B, MS/MS spectrum of poplar xylem metabolite annotated as **3**$_{S'\text{-}G^a}$. C, MS/MS spectrum of a poplar xylem metabolite of the putative β-O-4-coupling product of coniferyl alcohol with the compound in B, annotated as **G**(β-O-4)**3**$_{S'\text{-}G^a}$. The molecular ion loses H_2O and CH_2O, which is a signature for 8-O-4-coupled oligolignols (Morreel et al., 2010). The additional CH_2O loss, is derived from the aliphatic alcohol of the **3**$_{S'\text{-}G}$ moiety. The six m/z signals with a red star are those that are also seen in the spectrum B and refer to the **3**$_{S'\text{-}G^a}$ moiety of the molecule. The four annotated signals with m/z 195.066 and lower, are signatures of an 8-O-4-linked G unit (Morreel et al., 2010). D, Diastereomer

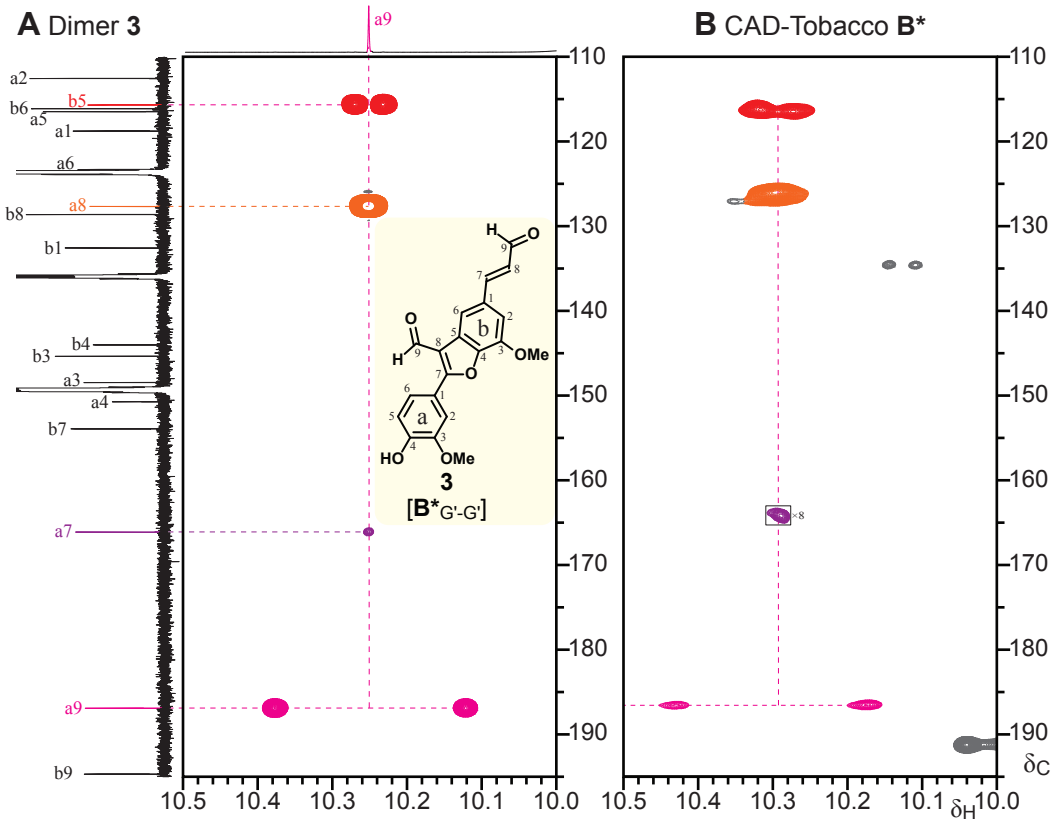


Supplemental Figure S2 *ctd.*
of the compound shown in panel C. For annotations of the detected m/z signals, see panel C.

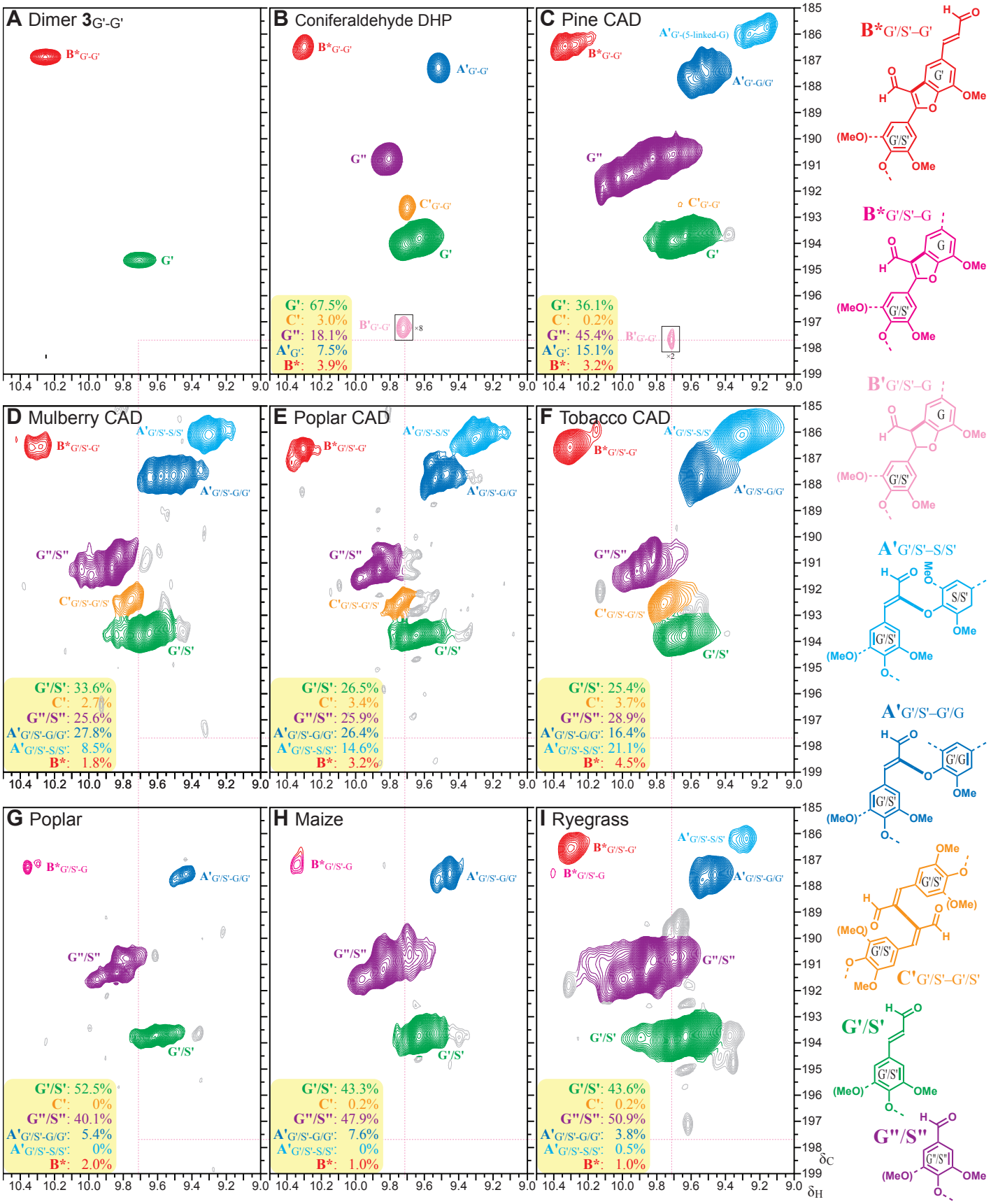


Supplemental Figure S3 Partial aromatics regions for all samples in Figure 3 in the main text, to indicate the G:S nature of the samples and to provide an indication of the levels of hydroxycinnamaldehyde incorporation into the lignin polymer chain. As noted in the main text, quantifying the degree of hydroxycinnamaldehyde incorporation into the lignin polymer backbone is currently difficult because of the number of structures

Supplemental Figure S3 *ctd*
involved and the peak overlap in NMR spectra. A useful measure is from the unique peak in the aromatics region arising from sinapaldehyde that has 8-O-4-coupled into lignin, the dispersed diagnostic **A'**$_{S'}$ component (magenta), a major cross-coupling product between the predominant sinapaldehyde (in dicots and monocots) and S or G units in lignin. These **A'**$_{S'}$ levels are reported on an S+G basis and are also shown in the yellow box at the bottom-right of each spectrum in Figure 3. A slightly different method was required for the G-lignin from pine; we simply ratio the shifted **G'**-derived peaks to the sum of those plus the peaks in the usual **G**$_2$ region. A more expansive and serious study is required (and planned) for understanding and assigning all these peaks. *p*-Coumarate (***p*CA**) and tricin (**T**) on the grasses, and *p*-hydroxybenzoate (***p*HB**) on the poplar, are labeled and/or structures are shown, but these are not a focus here.



Supplemental Figure S4 HMBC data establishing the benzofuran structure in lignin. A, Long-range ^{1}H–^{13}C correlation (HMBC) spectra (700 MHz, 4:1 DMSO-d_6:pyridine-d_5) of benzofuran dimer **3**, a **B***$_{G'-G'}$ moiety, showing correlations over 2-3 bonds from proton a_9 to carbons a_8, a_7, and, crucially (to establish the radical coupling mode), carbon b_5. B, The analogous HMBC data for the CAD-deficient tobacco lignin used for Figure 3F in the main paper, run in the same solvent but using the long-range coupling delay optimized for a long-range $J_{C–H}$ of 7.7 Hz (65 ms delay) instead of the 6.25 Hz (80 ms delay) used for model **3**. The corresponding correlations for structures **B***$_{G'/S'-G'}$ in the lignin provide compelling evidence for the veracity of the benzodioxane structure **B*** assignment. Coloring of 1D spectral peaks in this figure is reserved for the ones involved in the correlations. Again, unknown to us at the time, such structures were authenticated via HMBC spectra on a DDQ-oxidized lignin (Lahive et al., 2018); we additionally note the weak but also diagnostic proton a_9 to carbon a_7 correlation here not seen in their Figure 3.



Supplemental Figure S5 Extended aldehyde regions from the HSQC spectra of benzofuran dimer **3**, a DHP, and various lignins – the same samples and from the same raw data as used in Figure 3. The data are plotted at different levels, not necessarily with the same processing and, importantly, over an extended range to show the particularly minor-to-undetectable levels of the traditionally-assumed phenylcoumaran unit **B'** from coupling

Supplemental Figure S5 *ctd.*
of hydroxycinnamaldehyde monomers **G'**/**S'** with generic **G** units. A, The dimeric benzofuran model **3** isolated from short-duration radical coupling of coniferaldehyde **G'** (Figure 2), or synthesized independently (Figure 4). B, A DHP produced from coniferaldehyde **G'**; the **B'**/(**B'**+**B***) level is ~3.9%. C, An enzyme lignin (EL) from a CAD-deficient mutant pine (Ralph et al., 1997); the **B'**/(**B'**+**B***) level is ~4.3%. D, EL from a CAD-deficient mutant mulberry (Yamamoto et al., 2020). E, EL from a CAD-deficient transgenic poplar (De Meester et al., 2022). F, EL from a CAD-deficient transgenic tobacco, grown in 10%-enriched $^{13}CO_2$ to improve NMR sensitivity (Ralph et al., 1998). G, EL from a the wild-type control poplar for E (Van Acker et al., 2017; De Meester et al., 2022). H, EL from maize (stems), grown in 10%-enriched $^{13}CO_2$ to improve NMR sensitivity (IsoLife). I, EL from ryegrass (stems), grown in 10%-enriched $^{13}CO_2$ to improve NMR sensitivity (Ralph et al., 1995). The **B'** level in lignins D-I is essentially zero. Structures illustrate the shorthand formalism, and are colored to match the resolved and identified NMR correlation peaks.

References

De Meester B, Van Acker R, Wouters M, Traversari S, Steenackers M, Neukermans J, Van Breusegem F, Dejardin A, Pilate G, Boerjan W (2022) Field and saccharification performances of poplars severely downregulated in CAD1. New Phytologist **236:** 2075-2090

Lahive CW, Lancefield CS, Codina A, Kamer PCJ, Westwood NJ (2018) Revealing the fate of the phenylcoumaran linkage during lignin oxidation reactions. Organic & Biomolecular Chemistry **16:** 1976-1982

Morreel K, Kim H, Lu F, Dima O, Akiyama T, Vanholme R, Niculaes C, Goeminne G, Inzé D, Messens E, Ralph J, Boerjan W (2010) Mass-spectrometry-based fragmentation as an identification tool in lignomics. Analytical Chemistry **82:** 8095-8105

Ralph J, Grabber JH, Hatfield RD (1995) Lignin-ferulate crosslinks in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. Carbohydrate Research **275:** 167-178

Ralph J, Hatfield RD, Piquemal J, Yahiaoui N, Pean M, Lapierre C, Boudet A-M (1998) NMR characterization of altered lignins extracted from tobacco plants down-regulated for lignification enzymes cinnamyl-alcohol dehydrogenase and cinnamoyl-CoA reductase. Proceedings of the National Academy of Sciences **95:** 12803-12808

Ralph J, MacKay JJ, Hatfield RD, O'Malley DM, Whetten RW, Sederoff RR (1997) Abnormal lignin in a loblolly pine mutant. Science **277:** 235-239

Ralph SA, Landucci LL, Ralph J (2023) NMR database of lignin and cell wall model compounds. *In* Available over Internet at https://www.glbrc.org/databases_and_software/nmrdatabase/ (updated sporadically since 1993). S. A. Ralph, L. L. Landucci, J. Ralph

Van Acker R, Déjardin A, Desmet S, Hoengenaert L, Vanholme R, Morreel K, Laurans F, Kim H, Santoro N, Foster C, Goeminne G, Légée F, Lapierre C, Pilate G, Ralph J, Boerjan W (2017) Different routes for conifer- and sinapaldehyde and higher saccharification upon deficiency in the dehydrogenase CAD1. Plant Physiology **175:** 1018-1039

Yamamoto M, Tomiyama H, Koyama A, Okuizumi H, Liu S, Vanholme R, Goeminne G, Hirai Y, Shi H, Nuoendagula, Takata N, Ikeda T, Uesugi M, Kim H, Sakamoto S, Mitsuda N, Boerjan W, Ralph J, Kajita S (2020) A century-old mystery unveiled: Sekizaisou is a natural lignin mutant. Plant Physiology **182:** 1821-1828