



Hydroxycinnamaldehyde-derived benzofuran components in lignins

Koichi Yoshioka ^{1,2} Hoon Kim ^{1,2,†} Fachuang Lu ^{1,2} Nette De Ridder ^{3,4} Ruben Vanholme ^{3,4} Shinya Kajita ^{5,6} Wout Boerjan ^{3,4} and John Ralph ^{1,2,7,*}

- 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 Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 71, Ghent 9052, Belgium
- 4 VIB Center for Plant Systems Biology, VIB, Technologiepark 71, Ghent 9052, Belgium
- 5 Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
- 6 Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
- 7 Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

*Author for correspondence: jralph@wisc.edu

[†]Present address: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, WI 53726, USA.

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.

Abstract

Lignin is an abundant polymer in plant secondary cell walls. Prototypical lignins derive from the polymerization of monolignols (hydroxycinnamyl alcohols), mainly 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 biosynthesis pathway that reduces hydroxycinnamaldehydes to monolignols, chain-incorporated aldehyde unit levels are elevated. The nature and relative levels of aldehyde components in lignins can be determined from their distinct and dispersed correlations in 2D ¹H–¹³C-correlated nuclear magnetic resonance (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 correlation peaks 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, the 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 that are coupled to coniferaldehyde can be distinguished from those coupled with a generic guaiacyl end-unit. These benzofuran peaks may now be annotated and reported and their structural ramifications further studied.

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, 2019; Vanholme et al. 2012). In this lignification process, the 2 major monomers, coniferyl alcohol (**G**^a) and sinapyl alcohol (**S**^a), give rise to guaiacyl (**G**) and syringyl (**S**) units in the polymer

(Fig. 1). Dimeric units in the polymer arise from the various possible radical-coupling modes. The coupling itself produces unstable intermediates that rearomatize to produce the final dimeric units that are conventionally identified by the characteristic coupling reactions that produced them and the new interunit bonds that result. In normal lignins derived from 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 (Fig. 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, 2019; Vanholme et al. 2012). Only 2 such units, the β -aryl ether (β -O-4) and the phenylcoumaran (β -5), are shown in Fig. 1, which 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.

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 postlignification by the oxidation of hydroxycinnamyl alcohol endgroups, it is being increasingly realized that the hydroxycinnamaldehydes are authentic lignin monomers, especially in plants deficient in the last monolignol biosynthetic pathway enzyme, coniferyl alcohol dehydrogenase (CAD), which 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 the 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 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'**, Fig. 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 Fig. 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 Fig. 3.

Hydroxycinnamaldehyde 8-O-4-coupled units (denoted as **A'** units here, Fig. 1) were originally identified in CAD-deficient plants (Ralph et al. 1997, 2001; Sederoff et al. 1999; Chabannes et al. 2001; Kim et al. 2002, 2003; Barrière et al. 2004; Lapiere 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 ^1H - ^{13}C -correlated [heteronuclear single-quantum coherence (HSQC)] NMR spectra (Anderson et al. 2015; Yan et al. 2019; Yamamoto et al. 2020).

In recent studies on a CAD-deficient mulberry (*Morus alba*) mutant (Yamamoto et al. 2020), we noted but did not report on the presence of well-dispersed aldehyde signals in 2D HSQC NMR spectra that had not been noted in prior literature. This peak is prominently shown with red-colored contours in Fig. 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 2D NMR correlation peak via synthetic lignification

We had previously observed the HSQC NMR signature of this component in synthetic lignins (DHPs) produced from monolignols by the peroxidase- H_2O_2 single-electron oxidation

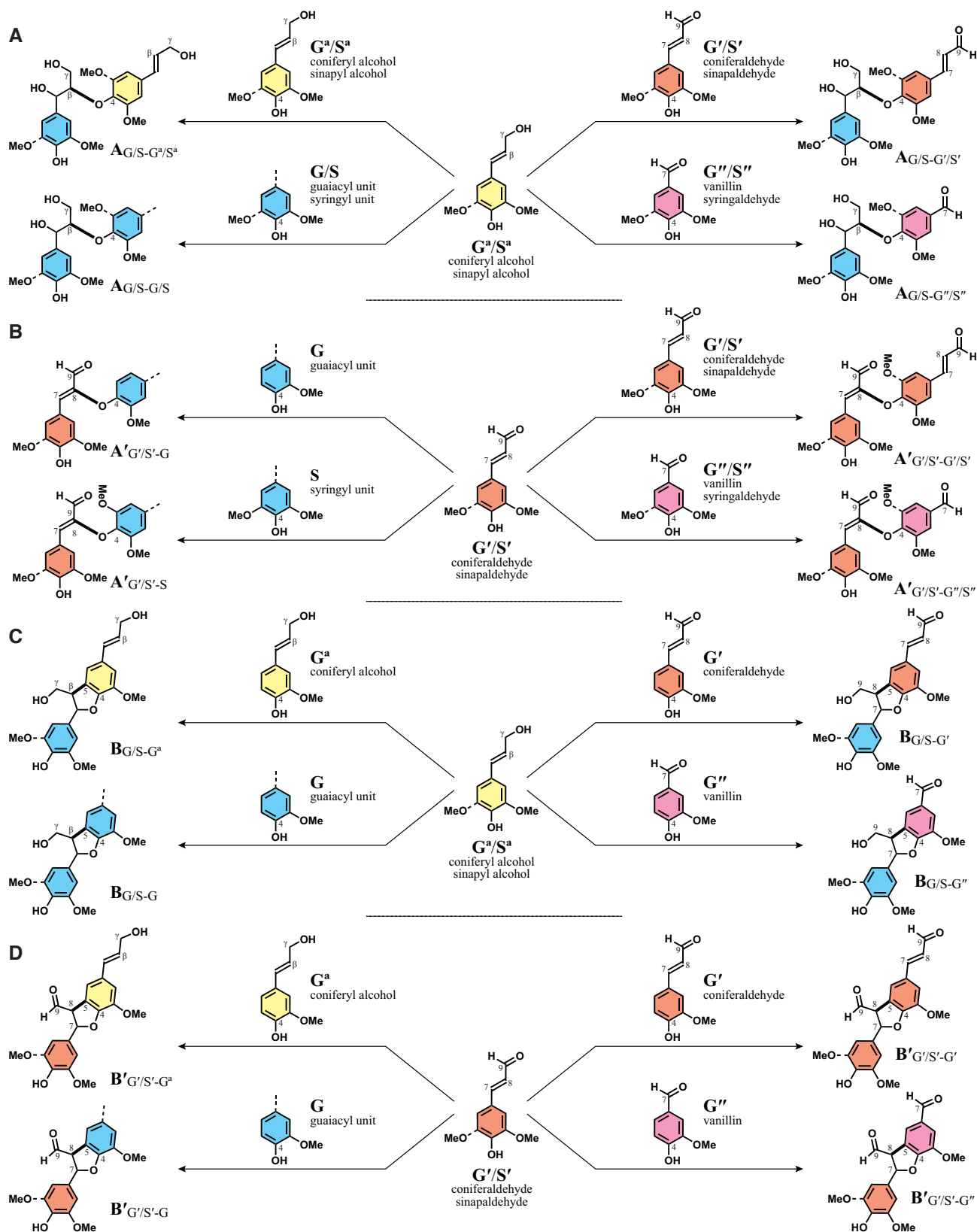


Figure 1. Lignification scheme for producing β -ether and phenylcoumaran units from monolignols and hydroxycinnamaldehyde monomers. **A)** The β -O-4/8-O-4 coupling or cross-coupling of a monolignol (coniferyl alcohol G^a or sinapyl alcohol S^a , center) with another monolignol (top left),
(continued)

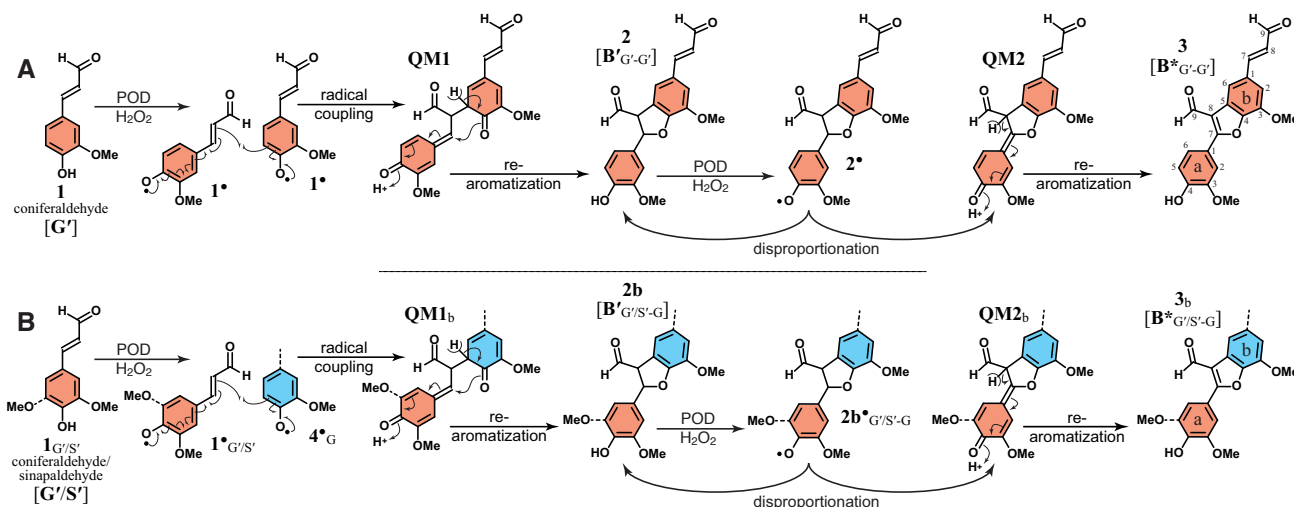


Figure 2. A scheme by which benzofuran dimer **3** is produced via a biomimetic in vitro system, and how analogous components **B*** are formed in natural and synthetic lignins. **A)** The dehydrodimerization and dehydrogenation of coniferaldehyde **1**. 8-5-Coupling of coniferaldehyde radical **1*** produces, following rearomatization and cyclization of the intermediate quinone methide **QM1**, the phenylcoumaran **2** (a **B'G'-G'** moiety). After generation of the phenolic radical **2***, 2 such radicals can disproportionate to one molecule of the original **2** and a new quinone methide **QM2** that, following rearomatization via loss of its acidic 8-proton, produces the benzofuran **3**, a **B*G'-G'** moiety. **B)** The analogous reaction from cross-coupling of coniferaldehyde **G'** (also represented here as compound **1G'** to show the consistency with Fig. 2A) or sinapaldehyde **S'** (**1S'**) with a generic **G** unit to produce **3G'/S'-G**, i.e. **B*G'/S'-G** moieties. Note that we accede that disproportionation of radicals **2*** need not require 2 such radicals; one could be another phenolic radical present during lignification. It is structures **3** (**B***) that we find incorporated into polymeric lignins, as evidenced from their NMR data. POD, peroxidase.

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 1-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** (Fig. 2A) using peroxidase-H₂O₂ in buffer, pH 6.5, produced a dimer in low yield that, after purification, had the characteristic ¹H–¹³C NMR correlation signature seen in the lignin and DHP

Figure 1. (Continued)

a generic guaiacyl or syringyl lignin unit (**G** or **S**, bottom left), a hydroxycinnamaldehyde (**G'** or **S'**, top right), and a hydroxybenzaldehyde (**G''** or **S''**, bottom right), all yielding various β-ether structures **A**. The convention is that such β-ethers are labeled **A**, and the nature of the resulting dimer or higher oligomer is denoted via a subscript, first with the phenolic end-unit and, after the hyphen, the newly etherified unit; an **a** is used for the hydroxycinnamyl alcohols (the monolignols and monolignol endgroups), a prime for the hydroxycinnamaldehyde, and a double-prime for the hydroxybenzaldehyde (monomers or endgroups). **B)** The analogous 8-O-4 coupling or cross-coupling of a hydroxycinnamaldehyde (**G'** or **S'**, center) with a generic guaiacyl lignin end-unit (**G**, top left), a generic syringyl lignin end-unit (**S**, bottom left), another hydroxycinnamaldehyde (**G'** or **S'**, top right), and a hydroxybenzaldehyde (**G''** or **S''**, bottom right), all yielding various β-ether structures **A'**. The prime is required on the **A** to indicate that the unit is not the usual arylglycerol-β-aryl ether unit **A** but a cinnamaldehyde-2-aryl ether resulting from the loss of the acidic β-proton from the intermediate quinone methide (not shown, but see later in Fig. 2), i.e. it is still a product of 8-O-4-coupling but results in a different structure and therefore is differentiated by the prime to indicate that the new phenolic end-unit is derived from a hydroxycinnamaldehyde. Note that we do not show 8-O-4-cross-coupling of **G'** or **S'** with a monolignol (**G^a** or **S^a**) because the monolignol invariably couples at its β-position in these reactions, as above in **A**). **C)** The analogous β-5 coupling and cross-coupling of a monolignol (**G^a** or **S^a**, center) with coniferyl alcohol (**G^a**, top left), a generic lignin guaiacyl unit (**G**, bottom left), coniferaldehyde (**G'**, top right), and vanillin (**G''**, bottom right), all yielding various phenylcoumaran structures **B**. Note that no syringyl unit (**S^a**, a generic **S** unit, **S'**, or **S''**) is capable of 5-coupling because of the methoxy-group substitution. **D)** The analogous 8-5 coupling and cross-coupling of a hydroxycinnamaldehyde (**G'** or **S'**, center) with coniferyl alcohol (**G^a**, top left), a generic lignin guaiacyl unit (**G**, bottom left), coniferaldehyde (**G'**, top right), and vanillin (**G''**, bottom right), all yielding various phenylcoumaran structures **B'**. The prime in **B'** is required to indicate again that the new phenolic end-unit is derived from a hydroxycinnamaldehyde. The succinct shorthand nomenclature used here is mildly complex but hopefully rational. It is an extension to that already reasonably well adopted in metabolite profiling and NMR work. We could have used simple numbering specific to just this paper but wanted the labeling to be applicable to any research in which lignin NMR peaks need unambiguous and consistent annotation. We will be attempting to introduce a more expansive formalism to address the various structures anticipated in lignins, so that assignments among papers become coherent and the current confusion from inconsistent labeling may be obviated. Utilizing a part of that scheme here is hopefully useful for the plant research community. A more detailed description is provided in the Supplemental data.

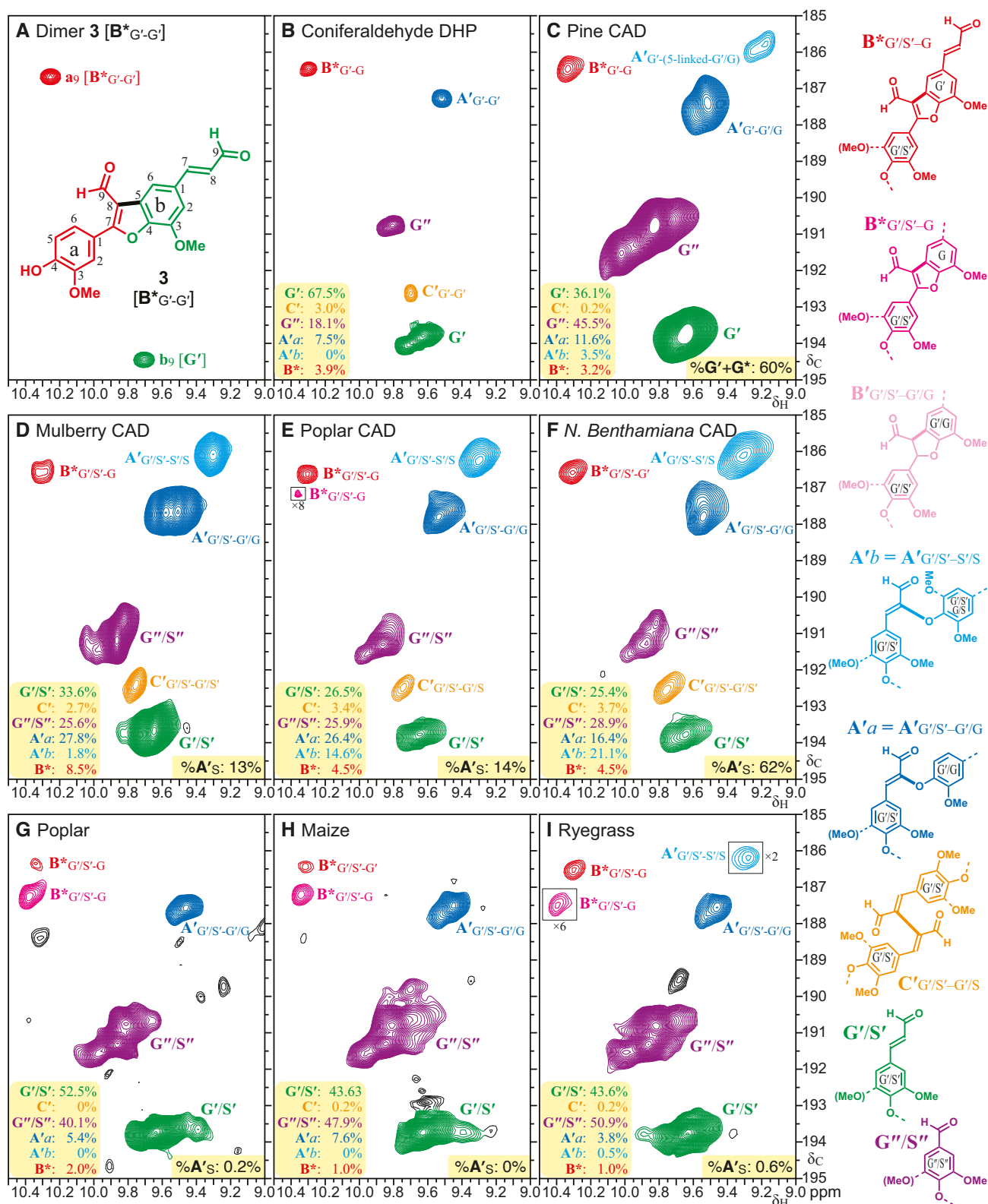


Figure 3. Aldehyde regions from the HSQC spectra of the dimeric benzofuran, a synthetic DHP, and various lignins. In addition to the hydroxybenzaldehyde (G''/S'') and hydroxycinnamaldehyde (G'/S') endgroups that are well established in lignins, 8-8-dimerization results in structures C' , and 8-O-4-cross-coupling of hydroxycinnamaldehyde radicals with the radical from the phenolic end of the growing polymer produces structures

(continued)

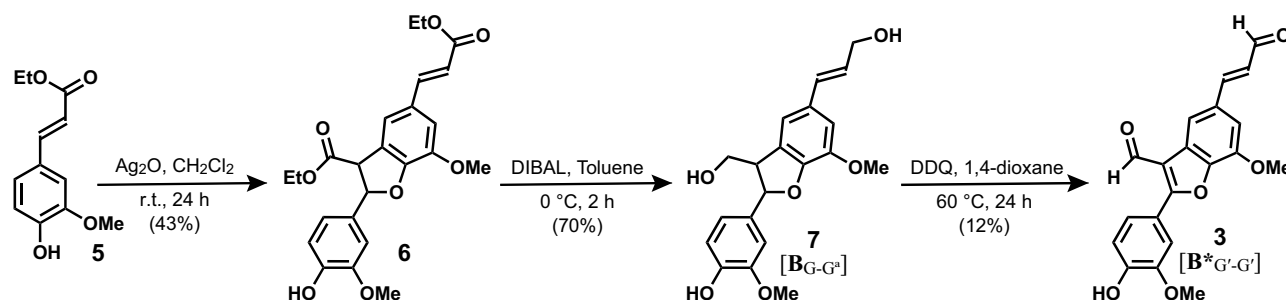


Figure 4. Synthesis of the benzofuran dimer **3** (a $B^*_{G'-G'}$ model) via the traditional β -5-coupled coniferyl alcohol dimer **7**. Yields for each step following purification are in parentheses. DIBAL, diisobutylaluminum hydride; DDQ, dichloro-dicyano-benzoquinone.

spectra (Fig. 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** (Fig. 2A, Supplemental Fig. S1).

The origin of structure **3** was not initially obvious. It was clear (Fig. 2A) that it was a dehydrodimerization product resulting from 8-5-coupling of radical 1^* from coniferaldehyde **1** (also denoted more generally as G' , Fig. 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** (Fig. 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 2 radicals 2^* 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^*_G (Fig. 2B). An analogous product from sinapaldehyde 1_S

(or just $5'$) is also possible. Additionally, and not diagrammed in Fig. 2, disproportionation need not involve 2 radicals of exactly the same type.

After making this structural assignment for dimer **3** that was isolated from the radical coupling of coniferaldehyde, and the apparent discovery of a previously unrecognized 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 (Fig. 4). The data matched perfectly, and it was undoubtedly the same benzofuran. That dimer **3** results from the dehydrogenation of coniferaldehyde (Fig. 2) is also consistent with its synthesis from β -5-dehydroconiferyl alcohol **7** by dehydrogenation/oxidation via DDQ (Fig. 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 phenylcoumaran **7** (Fig. 4) to compound **2** (Fig. 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 Fig. 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-

Figure 3. (Continued)

A' in the lignin; the latter has been increasingly well noted in most lignins because of its discovery and structural elucidation from studies on mutant and transgenic plants deficient in CAD (Kim et al. 2003). **A**) The dimeric benzofuran model **3** isolated from short-duration radical coupling of coniferaldehyde G' (Fig. 2) or synthesized independently (see Fig. 4). **B**) A DHP produced from coniferaldehyde G' . **C**) An EL from a CAD-deficient mutant pine (Ralph et al. 1997). Note that this spectrum features an A' peak that, in other spectra, is labeled as $A'_{G'/S'-S'/S'}$; here, it arises from β -ethers resulting from the coupling of a hydroxycinnamaldehyde with a 5-linked G unit, i.e. in structures labeled $A'_{G'-(5-linked-G'/G)}$. **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. 1998b). **G**) EL from the WT 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). Structures illustrate the shorthand formalism and are colored to match the resolved and identified NMR correlation peaks. The compositional box in each figure reports the relative integrals for the various classes of aldehydes, color-coded to match the structures. The box in the bottom-right corner of each spectrum provides a measure of the level of hydroxycinnamaldehyde incorporation into the lignin from integrating the well-resolved aldehyde-derived peaks and expressing the integrals relative to the $S + G$ total, i.e. on a total lignin basis—see Supplemental Fig. S3. Colored contours and labels match those of the structures. Boxes around a contour have the zoom factor noted, e.g. 8 $\times$ means the contour is plotted at an 8-fold higher level. DHP, synthetic dehydrogenation polymer; CAD, coniferyl alcohol dehydrogenase (deficient).

oxidized lignins. As there are important biosynthetic ramifications for these units, particularly in lignins from CAD-deficient plants, and as these structures represent a subpathway in lignification that was previously unrecognized, it is particularly important to identify and authenticate their 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 (*Populus tremula* × *alba*) xylem utilizing ultra-performance liquid chromatography–MS (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 Fig. 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'-G} in Supplemental Fig. S2B, and 2 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'-G}, were putatively identified in the xylem extract based on their accurate masses and MS/MS fragmentation (Supplemental Fig. S2, C and 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 (Fig. 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 has been increasingly noted in most lignins since its 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 (*Pinus taeda* L.) (Fig. 3C) features an **A'** peak that, in other spectra, is labeled as **A'**_{G'/S'-S'/S'}, but there are no **S** units in pine lignins. The analogous peak in the pine CAD spectrum (Fig. 3C) arises from 8-O-4-coupling of a hydroxycinnamaldehyde with a 5-linked **G** unit, i.e. in structures labeled **A'**_{G'-(5-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 Fig. 3A, we show the region for the benzofuran dimer **3** isolated as described above (Fig. 2) but also independently synthesized (Fig. 4), along with that from a coniferaldehyde DHP (Fig. 3B), and from a selection of lignins (Fig. 3, C to 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 because of the deviations from exactly 4:1 v/v, from the matrix and differing solute concentrations, and from the differing amounts of water in the samples. The natural mulberry CAD-deficient mutants in which this component most recently came to our attention (Fig. 3D) along with lignins from studies involving various CAD-deficient mutants and transgenics are shown (Fig. 3, C to F). Then, to make it clear that similar structures are evident in lignins from WT plants, the bottom row shows the same regions for the control poplar (Fig. 3G) from the studies that produced the CAD-deficient poplar (Fig. 3E; Van Acker et al. 2017; De Meester et al. 2022), and from 2 grasses, maize (*Zea mays*) (Fig. 3H) and ryegrass (*Lolium multiflorum* Lam. cv. Florida 80) (Fig. 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 Fig. S3). These **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 Fig. 3. A slightly different method was required for the **G**-lignin from pine, as explained in the caption of Supplemental Fig. S3.

As readily noted in the spectra in Fig. 3, there may be 2 well-resolved correlation clusters for structures **B***. In principle, 4 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 **B***_{G'-G'}, **B***_{G'-G}, **B***_{S'-G'}, and **B***_{S'-G} (Fig. 2). Another distinction could be those units with a coniferyl alcohol **G^a** endgroup, structures **B***_{G'-G^a} and **B***_{S'-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 2 resolved **B*** contours in highly CAD-deficient plants, in which hydroxycinnamaldehydes predominate vs those in WT 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 2 correlation peaks. A close examination of the data processed under resolution-enhancing conditions suggests that coniferaldehyde-derived **B***_{G'-G'} peaks may have slightly lower carbon and proton chemical shifts than sinapaldehyde-derived **B***_{S'-G'} ones, but the peaks are not resolved and the correlations are, therefore, labeled as combined **B***_{G'/S'-G'}, i.e. as resulting from either hydroxycinnamaldehyde having coupled with a coniferaldehyde or vanillin unit for the dicots and monocots here (Fig. 3, D to I). Similarly, the lower peak deriving from the 8-5-coupling of a hydroxycinnamaldehyde with a generic **G** unit cannot be resolved into **B***_{G'-G} and **B***_{S'-G}, so this correlation cluster is labeled as the combined **B***_{G'/S'-G} for the dicots and monocots (Fig. 3, D to I).

A single previously unreported NMR correlation peak, even though it does have 2 simultaneous constraints in that both the ¹H and the ¹³C chemical shifts have to match those in the component of interest, is insufficiently diagnostic to confidently assign a 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 **B*** units, there are no other protonated side-chain carbons in the moiety that gives rise to the proton a₉ to carbon a₉ correlation peaks highlighted here in Fig. 3 and shown for the isolated and synthesized model compound **3** (Fig. 3A, Supplemental Fig. S1). Correlation experiments using the proton coupling network such as COSY and TOCSY and the usually powerful 2D HSQC-TOCSY experiment therefore provide no additional data for the authentication of units **B***. Long-range correlations over 2 to 3 bonds from proton a₉ (see Fig. 2 and Supplemental Fig. S1 for the labeling of 2 moieties) would be particularly diagnostic. Such correlations could be expected from proton a₉ to carbons a₈ and a₇ and, crucially (to establish the radical coupling mode), carbon b₅ in model **3**. Observing the corresponding correlations for structures **B***_{G'/S'-G'} in a lignin would provide compelling evidence for the veracity of structures **B***. Such evidence for all 3 correlations is shown in the heteronuclear multiple-bond correlation (HMBC) spectra from the model **3** and in a lignin from CAD-deficient *Nicotiana benthamiana* in Supplemental Fig. 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 a₉ to carbon a₇ correlation here (Supplemental Fig. S4), which is not seen in their Fig. 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'** (Fig. 2). The endgroup aldehyde correlation would be in the green **G'/S'** NMR contour (Fig. 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 **2**_{G'-G'}) and is extremely weak in actual lignins (Kim et al. 2003). Supplemental Fig. S5 shows plots of the extended aldehyde regions for all 9 samples used in Fig. 3. A small **B'** peak is seen in spectra from the coniferaldehyde DHP (Supplemental Fig. S5B) and perhaps the CAD-deficient pine (Supplemental Fig. 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 Fig. 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 other 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, 2023a; Hatfield et al. 2008; Ralph 2010). 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 conifer-aldehyde (or at least another conjugated, unsaturated unit), i.e. $\mathbf{B}^*_{\text{G}'/\text{S}'-\text{G}'}$. It is not clear why ryegrass had predominantly the $\mathbf{B}^*_{\text{G}'/\text{S}'-\text{G}'}$ correlation peak (Fig. 3I, Supplemental Fig. S5I), making its profile appear more like those from CAD-deficient plants. The $\mathbf{S}/\mathbf{G}$ is lower in ryegrass (0.78 vs 1.7 in the maize, Supplemental Fig. S3); maize has nearly 5-fold higher levels of lignin derived from sinapyl *p*-coumarate conjugates, but ryegrass has over 4-fold higher levels of tricin. None of these factors rationalize the CAD-deficient profile, but we note (Fig. 3I, Supplemental Fig. S5I) that the $\mathbf{A}'$ levels are also higher, and the usually extremely weak or undetectable $\mathbf{A}'_{\text{G}'/\text{S}'-\text{S}'/\text{S}'}$ 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 $\mathbf{B}'$ are difficult to detect in lignins (Kim et al. 2003; Supplemental Fig. S5). The benzofuran products $\mathbf{B}^*$ here are, however, readily evidenced in essentially any lignin NMR spectrum (Fig. 3), from normal as well as at enhanced levels from CAD-deficient plants. It can be concluded that the phenolic radicals from the phenylcoumaran structures $\mathbf{B}'$ preferentially undergo disproportionation reactions rather than undergoing radical coupling themselves (Fig. 2). The products $\mathbf{B}^*$ then apparently readily undergo radical coupling reactions, incorporating such moieties into the backbone of the polymer. Products $\mathbf{B}^*_{\text{G}'-\text{G}'}$ arising from the homodimerization of coniferaldehyde, the presumed cross-coupling of sinapaldehyde with coniferaldehyde (i.e. units $\mathbf{B}^*_{\text{S}'-\text{G}'}$), and the cross-coupling of either sinapaldehyde or coniferaldehyde with a generic $\mathbf{G}$ unit (i.e. units $\mathbf{B}^*_{\text{G}'/\text{S}'-\text{G}}$), all appear to be viable pathways to produce the $\mathbf{B}^*$ 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 substantial challenge but become valuable additions for lignin characterization and evaluation in future studies that may also elucidate how these $\mathbf{B}^*$ 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, USA). 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, USA) equipped with Biotage silica gel cartridges (SNAP KP-Sil, SNAP Ultra or Sfär Silica D).

Nuclear magnetic resonance

NMR spectra were obtained on a Bruker Biospin (Rheinstetten, Germany) NEO 700-MHz spectrometer fitted with a cryogenically cooled 5-mm quadrupole-resonance (QCI) $^1\text{H}/^{31}\text{P}/^{13}\text{C}/^{15}\text{N}$ gradient probe with inverse geometry or an AVANCE 500-MHz spectrometer fitted with a cryogenically cooled 5-mm Triple Resonance Inverse (TCI) $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ gradient probe with inverse geometry. ^1H - ^{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 MS

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 $\times$ 2.1 mm, particle diameter: 1.7 μm , 100 Å) (Phenomenex, Torrance, CA, USA).

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

The stem (~20 cm) from greenhouse-grown poplar (*P. tremula* $\times$ *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

Three different solutions A, B, and C were prepared for coniferaldehyde-DHP synthesis (Fig. 3B). 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: hydrogen peroxide aqueous solution (30% v/v, 118 μ L, 1.2 mmol) was dissolved in Milli-Q water (240 mL). Solution C: horseradish peroxidase (5 mg) was dissolved in 0.1 M Na-phosphate buffer (60 mL). Solutions A and B were coadded 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 $\times$ g using a Thermo Scientific Sorvall Primo centrifuge. Precipitates were collected and washed with water (100 mL $\times$ 3). After the precipitates were lyophilized, the DHP was obtained as a red solid (150 mg, 85% yield). 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 to 250 U/mg) in water (13 mL) into the flask, the hydrogen peroxide solution (3% v/v, 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 $\times$ 4), and the combined organic layer was dried over anhydrous Na₂SO₄ 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 aldehyde dimer 3 (1.3 mg) was obtained, i.e. in 13% yield from Fraction 5; we did not attempt to estimate the yield across all fractions, and these certainly included oligomers larger than this dimer.

The NMR data in 4 solvents, CDCl₃, 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, a₃-OMe), 3.95 (3H, s, b₃-OMe), 6.90 (1H, dd, J = 15.8, 7.7 Hz, b₈), 6.97 (1H, d, J = 8.2 Hz, a₅), 7.39 (1H, dd, J = 8.2, 2.1 Hz, a₆), 7.42 (1H, d, J = 1.4 Hz, b₂), 7.44 (1H, d, J = 2.1 Hz, a₂), 7.78 (1H, d, J = 15.8 Hz, b₇), 7.96 (1H, d, J = 1.3 Hz, b₆), 9.63 (1H, d, J = 7.7 Hz, b₉), 10.20 (1H, s, a₉); δ_C 55.8 (a₃-OMe), 56.2 (b₃-OMe), 107.1 (b₂), 112.4 (a₂), 115.4 (b₅), 116.0 (b₆), 116.2 (a₅), 118.5 (a₁), 122.9 (a₆), 127.4 (a₈), 128.4 (b₈), 132.3 (b₁), 143.8 (b₄), 145.1 (b₃), 148.2 (a₃), 150.5 (a₄), 153.6 (b₇), 165.8 (a₇), 186.6 (a₉), 194.2 (b₉); HRMS: m/z [M-H][−] calculated for C₂₀H₁₅O₆: 351.0869, found: 351.0871.

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

The authentic dimeric compound 3 (Figs. 3A and 4, Supplemental Fig. S1) was also independently synthesized (Fig. 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, 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 μ mol) and DDQ (19 mg, 84 μ mol) in a 15 mL ace glass pressure tube (Ace Glass, Inc., Vineland, NJ, USA). 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% yield, Fig. 4). No attempt was made to optimize the 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 wood preparations

All but one of the plant materials used in this study were from previous studies over many years: a pine (*P. taeda* L.) CAD mutant (Ralph et al. 1997); Sekizaisou, the mulberry (*M. alba*) CAD mutant (Yamamoto et al. 2020); a CAD-deficient transgenic poplar (*P. tremula* $\times$ *alba*) and its WT control (Van Acker et al. 2017; De Meester et al. 2022); a CAD-deficient *N. benthamiana* transgenic grown in 10% ¹³CO₂ (Ralph et al. 1998b); and ryegrass (*L. multiflorum* Lam. cv. Florida 80) grown in 10% ¹³CO₂ (Ralph et al. 1995). The maize (*Z. mays*) sample was from IsoLife (Wageningen, The Netherlands) P-60501 P-¹³C Maize stem (10 atom% ¹³C).

Enzyme lignin preparation

A fraction of the dried extract-free wood (750 mg) was ball-milled in 20 mL agate jars with 10 $\times$ 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 d at 35°C. They were then pelletized (10 min at 1,777 rcf on an Eppendorf 5810R). The acetate buffer was decanted, and the enzymatic digestion was repeated. After 2 digestion cycles, the samples were washed 3 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 (^1H – ^{13}C) and HMBC NMR spectroscopy of isolated ELs

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 to 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 *N. benthamiana* lignin in DMSO- d_6 /pyridine- d_5 (4:1, v/v) for Supplemental Fig. S4 has used a long-range coupling delay of 65 ms (corresponding to a long-range $J_{\text{C-H}}$ of 7.7 Hz).

Accession numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers Potri.009G095800 for the poplar (*P. tremula* $\times$ *alba*), GenBank AF003349 and AF003350 for the pine (*P. taeda* L.), GenBank LC476972 for the mulberry (*M. alba*), and Potri.009G095800 for poplar (*P. tremula* $\times$ *alba*).

Acknowledgments

We thank Steve Karlen (WEI) for the high-resolution 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 and extract-free wood preparations* section) from which spectral data were mined for Fig. 3 and Supplemental Figs. S3 and S5.

Author contributions

All authors designed and performed the research; K.Y., H.K., and F.L. performed the synthetic work and, with J.R., analyzed the data and wrote the article; N.D.R. and R.V. screened the poplar metabolites. All authors contributed to and proofed the manuscript.

Supplemental data

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

Supplemental Note S1. An expanded explanation of the shorthand formalism used for the various structures.

Supplemental Figure S1. ^1H – ^{13}C correlation NMR spectra for the benzofuran dimer 3.

Supplemental Figure S2. Tandem MS of benzofuran dimer 3 and putative benzofuran-containing poplar xylem metabolites.

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

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

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

Funding

K.Y., H.K., F.L., and J.R. were supported by the Great Lakes Bioenergy Research Center, US Department of Energy, Office of Science, and Office of Biological and Environmental Research under Award Number DE-SC0018409. S.K. was partially supported by the Japan Science and Technology Agency and by the Japan Society for the Promotion of Science. N.D.R. was supported by the Research Foundation Flanders (FWO) for a predoctoral fellowship. W.B. acknowledges support from FWO project G008116N and from the interuniversity Bijzonder Onderzoeksfonds iBOF “Next-Bioref” (BOFIBO2021000902-iBOF).

Conflict of interest statement. None declared.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Adler E, Björkquist KJ, Häggroth S. Über die Ursache der Farbreaktionen des Holzes. *Acta Chem Scand.* 1948;2:93–94. <https://doi.org/10.3891/acta.chem.scand.02-0093>
- Alsharif MA, Raja QA, Majeed NA, Jassas RS, Alsimaree AA, Sadiq A, Naeem N, Mughal EU, Alsantali RI, Moussa Z, et al. DDQ as a versatile and easily recyclable oxidant: a systematic review. *RSC Adv.* 2021;11(47):29826–29858. <https://doi.org/10.1039/D1RA04575J>
- Anderson NA, Tobimatsu Y, Ciesielski PN, Ximenes E, Ralph J, Donohoe BS, Ladisch M, Chapple C. 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.* 2015;27(8):2195–2209. <https://doi.org/10.1105/tpc.15.00373>
- Barrière Y, Ralph J, Méchin V, Guillaumie S, Grabber JH, Argillier O, Chabbert B, Lapierre C. Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants. *C R Biol.* 2004;327(9-10):847–860. <https://doi.org/10.1016/j.crvi.2004.05.010>

- Chabannes M, Barakate A, Lapierre C, Marita J, Ralph J, Pean M, Danoun S, Halpin C, Grima-Pettenati J, Boudet A-M. 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. *Plant J.* 2001;**28**(3):257–270. <https://doi.org/10.1046/j.1365-3113X.2001.01140.x>
- De Meester B, Van Acker R, Wouters M, Traversari S, Steenackers M, Neukermans J, Van Breusegem F, Dejardin A, Pilate G, Boerjan W. Field and saccharification performances of poplars severely downregulated in CAD1. *New Phytol.* 2022;**236**(6):2075–2090. <https://doi.org/10.1111/nph.18366>
- Fanelli A, Rancour DM, Sullivan M, Karlen SD, Ralph J, Riaño-Pachón DM, Vicentini R, da Franca Silva T, Ferraz AL, Hatfield RD, et al. Overexpression of a sugarcane BAHD acyltransferase alters hydroxycinnamate content in maize cell wall. *Front Plant Sci.* 2021;**12**:626168. <https://doi.org/10.3389/fpls.2021.626168>
- Freudenberg K, Neish AC. Constitution and biosynthesis of lignin. Berlin, Heidelberg, New York: Springer-Verlag; 1968.
- Hatfield RD, Ralph J, Grabber JH. A potential role of sinapyl *p*-coumarate as a radical transfer mechanism in grass lignin formation. *Planta.* 2008;**228**(6):919–928. <https://doi.org/10.1007/s00425-008-0791-4>
- Higuchi T, Ito T, Umezawa T, Hibino T, Shibata D. Red-brown color of lignified tissues of transgenic plants with antisense CAD gene: wine-red lignin from coniferyl aldehyde. *J Biotechnol.* 1994;**37**(2): 151–158. [https://doi.org/10.1016/0168-1656\(94\)90006-X](https://doi.org/10.1016/0168-1656(94)90006-X)
- Kim H, Padmakshan D, Li Y, Rencoret J, Hatfield RD, Ralph J. Characterization and elimination of undesirable protein residues in plant cell wall materials for enhancing lignin analysis by solution-state NMR. *Biomacromolecules.* 2017;**18**(12):4184–4195. <https://doi.org/10.1021/acs.biomac.7b01223>
- Kim H, Ralph J. Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-*d*₆/pyridine-*d*₅. *Org Biomol Chem.* 2010;**8**(3): 576–591. <https://doi.org/10.1039/B916070A>
- Kim H, Ralph J, Akiyama T. Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-*d*₆. *Bioenergy Res.* 2008;**1**(1):56–66. <https://doi.org/10.1007/s12155-008-9004-z>
- Kim H, Ralph J, Lu F, Pilate G, Leplé JC, Pollet B, Lapierre C. Identification of the structure and origin of thioacidolysis marker compounds for cinnamyl alcohol dehydrogenase deficiency in angiosperms. *J Biol Chem.* 2002;**277**(49):47412–47419. <https://doi.org/10.1074/jbc.M208860200>
- Kim H, Ralph J, Lu F, Ralph SA, Boudet A-M, MacKay JJ, Sederoff RR, Ito T, Kawai S, Ohashi H, et al. NMR analysis of lignins in CAD-deficient plants. Part 1. Incorporation of hydroxycinnamaldehydes and hydroxybenzaldehydes into lignins. *Org Biomol Chem.* 2003;**1**(2):268–281. <https://doi.org/10.1039/b209686b>
- Kupče E, Freeman R. Compensated adiabatic inversion pulses: broadband INEPT and HSQC. *J Magn Reson.* 2007;**187**(2):258–265. <https://doi.org/10.1016/j.jmr.2007.05.009>
- Lahive CW, Lancefield CS, Codina A, Kamer PCJ, Westwood NJ. Revealing the fate of the phenylcoumaran linkage during lignin oxidation reactions. *Org Biomol Chem.* 2018;**16**(11):1976–1982. <https://doi.org/10.1039/C7OB03087H>
- Lapierre C, Pilate G, Pollet B, Mila I, Leplé JC, Jouanin L, Kim H, Ralph J. Signatures of cinnamyl alcohol dehydrogenase deficiency in poplar lignins. *Phytochemistry.* 2004;**65**(3):313–321. <https://doi.org/10.1016/j.phytochem.2003.11.007>
- Liu X, Van Acker R, Voorend W, Pallidis A, Goeminne G, Pollier J, Morreel K, Kim H, Muylle H, Bosio M, et al. Rewired phenolic metabolism and improved saccharification efficiency of a *Zea mays cinnamyl alcohol dehydrogenase 2 (zmcad2)* mutant. *Plant J.* 2021;**105**(5):1240–1257. <https://doi.org/10.1111/tpj.15108>
- Mansfield SD, Kim H, Lu F, Ralph J. Whole plant cell wall characterization using solution-state 2D-NMR. *Nat Protoc.* 2012;**7**(9):1579–1589. <https://doi.org/10.1038/nprot.2012.064>
- Morreel K, Kim H, Lu F, Dima O, Akiyama T, Vanholme R, Niculaes C, Goeminne G, Inzé D, Messens E, et al. Mass-spectrometry-based fragmentation as an identification tool in lignomics. *Anal Chem.* 2010;**82**(19):8095–8105. <https://doi.org/10.1021/ac100968g>
- Ralph J. Hydroxycinnamates in lignification. *Phytochem Rev.* 2010;**9**(1): 65–83. <https://doi.org/10.1007/s11101-009-9141-9>
- Ralph J, Brunow G, Harris PJ, Dixon RA, Schatz PF, Boerjan W. Lignification: are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template replication? In: Daayf F, El Hadrami A, Adam L, Ballance GM, editors. Recent advances in polyphenol research. Vol. 1. Oxford (UK): Wiley-Blackwell Publishing; 2008. p. 36–66.
- Ralph J, Bunzel M, Marita JM, Hatfield RD, Lu F, Kim H, Schatz PF, Grabber JH, Steinhart H. Peroxidase-dependent cross-linking reactions of *p*-hydroxycinnamates in plant cell walls. *Phytochem Rev.* 2004;**3**(1-2):79–96. <https://doi.org/10.1023/B:PHYT.0000047811.13837.fb>
- Ralph J, Garcia-Conesa MT, Williamson G. Simple preparation of 8-5-coupled diferulate. *J Agric Food Chem.* 1998a;**46**(7):2531–2532. <https://doi.org/10.1021/jf980123r>
- Ralph J, Grabber JH, Hatfield RD. Lignin-ferulate crosslinks in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydr Res.* 1995;**275**(1):167–178. [https://doi.org/10.1016/0008-6215\(95\)00237-N](https://doi.org/10.1016/0008-6215(95)00237-N)
- Ralph J, Hatfield RD, Piquemal J, Yahiaoui N, Pean M, Lapierre C, Boudet A-M. NMR characterization of altered lignins extracted from tobacco plants down-regulated for lignification enzymes cinnamyl-alcohol dehydrogenase and cinnamoyl-CoA reductase. *Proc Natl Acad Sci U S A.* 1998b;**95**(22):12803–12808. <https://doi.org/10.1073/pnas.95.22.12803>
- Ralph J, Kim H, Lu F, Smith RA, Karlen SD, Nuoendagula, Yoshioka K, Eugene A, Liu S, Sener C, et al. Lignins and lignification: new developments and emerging concepts. In: Quideau S, Salminen J-P, Wähälä K, de Freitas V, editors. Recent advances in polyphenol research. Vol. 8. Oxford (UK): Wiley-Blackwell; 2023a. p. 1–50.
- Ralph SA, Landucci LL, Ralph J. NMR database of lignin and cell wall model compounds. [accessed 2023 Oct 6]. Available from https://www.glbrc.org/databases_and_software/nmrdbase/.
- Ralph J, Lapierre C, Boerjan W. Lignin structure and its engineering. *Curr Opin Biotechnol.* 2019;**56**:240–249. <https://doi.org/10.1016/j.copbio.2019.02.019>
- Ralph J, Lapierre C, Marita J, Kim H, Lu F, Hatfield RD, Ralph SA, Chapple C, Franke R, Hemm MR, et al. Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. *Phytochemistry.* 2001;**57**(6):993–1003. [https://doi.org/10.1016/S0031-9422\(01\)00109-1](https://doi.org/10.1016/S0031-9422(01)00109-1)
- Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, et al. Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochem Rev.* 2004b;**3**(1-2):29–60. <https://doi.org/10.1023/B:PHYT.0000047809.65444.a4>
- Ralph J, MacKay JJ, Hatfield RD, O'Malley DM, Whetten RW, Sederoff RR. Abnormal lignin in a loblolly pine mutant. *Science.* 1997;**277**(5323):235–239. <https://doi.org/10.1126/science.277.5323.235>
- Ralph J, Quideau S, Grabber JH, Hatfield RD. Identification and synthesis of new ferulic acid dehydrodimers present in grass cell walls. *J Chem Soc Perkin Trans.* 1994;**1**(23):3485–3498. <https://doi.org/10.1039/p19940003485>
- Sederoff RR, MacKay JJ, Ralph J, Hatfield RD. Unexpected variation in lignin. *Curr Opin Plant Biol.* 1999;**2**(2):145–152. [https://doi.org/10.1016/S1369-5266\(99\)80029-6](https://doi.org/10.1016/S1369-5266(99)80029-6)
- Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Seguin A. Cinnamyl Alcohol Dehydrogenase-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. *Plant Cell.* 2005;**17**(7):2059–2076. <https://doi.org/10.1105/tpc.105.030767>

- Smith RA, Beebe ET, Bingman CA, Vander Meulen KA, Eugene A, Steiner A, Karlen SD, Ralph J, Fox BG.** Identification and characterization of a set of monocot BAHD monolignol transferases. *Plant Physiol.* 2022;**189**(1):37–48. <https://doi.org/10.1093/plphys/kiac035>
- Takahama U, Oniki T.** 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 Physiol.* 1997;**38**(4):456–462. <https://doi.org/10.1093/oxfordjournals.pcp.a029189>
- Van Acker R, Déjardin A, Desmet S, Hoengenaert L, Vanholme R, Morreel K, Laurans F, Kim H, Santoro N, Foster C, et al.** Different routes for conifer- and sinapaldehyde and higher saccharification upon deficiency in the dehydrogenase CAD1. *Plant Physiol.* 2017;**175**(3):1018–1039. <https://doi.org/10.1104/pp.17.00834>
- Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W.** Metabolic engineering of novel lignin in biomass crops. *New Phytol.* 2012;**196**(4):978–1000. <https://doi.org/10.1111/j.1469-8137.2012.04337.x>
- Witvrouw K, Kim H, Vanholme R, Goeminne G, Ralph J, Boerjan W, Vanholme B.** Reassessing the claimed cytokinin-substituting activity of dehydrodiconiferyl alcohol glucoside. *Proc Natl Acad Sci U S A.* 2023;**120**(9):e2123301120. <https://doi.org/10.1073/pnas.2123301120>
- Yamamoto M, Tomiyama H, Koyama A, Okuizumi H, Liu S, Vanholme R, Goeminne G, Hirai Y, Shi H, Nuoendagula, et al.** A century-old mystery unveiled: Sekizaisou is a natural lignin mutant. *Plant Physiol.* 2020;**182**(4):1821–1828. <https://doi.org/10.1104/pp.19.01467>
- Yan X, Liu J, Kim H, Liu B, Huang X, Yang Z, Lin Y-C, Chen H, Yang C, Wang J, et al.** CAD1 and CCR2 protein complex formation in monolignol biosynthesis in *Populus trichocarpa*. *New Phytol.* 2019;**222**(1):244–260. <https://doi.org/10.1111/nph.15505>
- Zhao Q, Tobimatsu Y, Zhou R, Pattathil S, Gallego-Giraldo L, Fu C, Jackson LA, Hahn MG, Kim H, Chen F, et al.** Loss of function of *Cinnamyl Alcohol Dehydrogenase 1* causes accumulation of an unconventional lignin and a temperature-sensitive growth defect in *Medicago truncatula*. *Proc Natl Acad Sci U S A.* 2013;**110**(33):13660–13665. <https://doi.org/10.1073/pnas.1312234110>