Lignin-Derived Thioacidolysis Dimers: Reevaluation, New Products, Authentication, and Quantification

Fengxia Yue,[a, b] Fachuang Lu,[a, b] Matt Regner,[b] Runcang Sun,[c] and John Ralph*[b]

Lignin structural studies play an essential role both in understanding the development of plant cell walls and for valorizing lignocellulosics as renewable biomaterials. Dimeric products released by selectively cleaving β-aryl ether linkages between lignin units reflect the distribution of recalcitrant lignin units, but have been neither absolutely defined nor quantitatively determined. Here, 12 guaiacyl-type thioacidolysis dimers were identified and quantified using newly synthesized standards. One product previously attributed to deriving from β-1-coupled units was established as resulting from β-5 units, correcting an analytical quandary. Another longstanding dilemma, that no β-β dimers were recognized in thioacidolysis products from gymnosperms, was resolved with the discovery of two such authenticated compounds. Individual GC response factors for each standard compound allowed rigorous quantification of dimeric products released from softwood lignins, affording insight into the various interunit-linkage distributions in lignins and thereby guiding the valorization of lignocellulosics.

Lignin, a major component of plant cell-wall (CW) complexes, plays an essential role in their development and strongly affects the utilization of sustainable lignocellulosic biomass.[3] Lignin valorization is becoming essential for the overall economics of a lignocellulosic biorefinery.[2] However, lignins have complex and heterogeneous structures that are neither absolutely definable nor determinable. The analysis of degradation products released by lignin depolymerization, especially dimeric and larger fragments, is valuable for characterizing lignins but is currently imperfect owing to the lack of authentic compounds and the resulting uncertainty regarding their structures.[2b,3] Such a status, to some extent, hinders the utilization of lignocellulosic biomass, especially lignins.

In softwood CWs, lignin consists primarily of guaiacyl (G) units, derived from coniferyl alcohol, along with a small proportion of p-hydroxyphenyl (H) units, derived from p-coumaryl alcohol.[4] It has been well documented that the majority of intrunit linkages in softwood lignins are, in order of decreasing abundance: β–O–4’ (β-aryl ether, henceforth just termed β-ether), β–5’ (phenylcoumaran), β–β (pinoresinol), and 5–5’ (biaryl, usually present in lignins as dibenzodioxocin units).[5] Also present at relatively minor levels are β–1’ (spirodienone), and 4–O–5’ (diaryl ether) units. Degradative methods, including acidolysis, thioacetylation, hydrolysis, and the widely used thioacidolysis and derivatization followed by reductive cleavage (DFRC) methods, have contributed extensively to our current knowledge of lignin structure.[6] All of these methods rely primarily on β-ether cleavage, such that any units or substructures linked by β-ethers (at either their 4–O- or β-positions) are released as low-molecular-weight compounds suitable for separation by gas or liquid chromatography (GC or LC), and analysis by mass (MS) or nuclear magnetic resonance (NMR) spectroscopy.

As an analytical method for the characterization of lignins, thioacidolysis has been successfully applied to a wide array of lignocellulosic materials and isolated lignins in various studies related to the biosynthesis of lignins, biological processing of plant biomass, and the pulping industry, as reviewed.[7] By capitalizing on its selective cleavage of β-ether units, diagnostic monomers and—after Raney-Ni desulfurization to reduce the number of products and improve the volatility—dimeric fragments can be analyzed by GC or GC–MS. However, although this approach has provided insight into lignin structure, particularly with regard to carbon–carbon and diaryl ether linkages between units,[8] it has also produced some dilemmas such as the large proportion of dimers that are β–1 (although only 1–2% of lignin units are typically so-linked) and the total absence, seemingly, of β–β-units that are known to be quite prominent in softwood G-lignins. Structural authentication, along with accurate quantification of lignin-derived thioacidolysis products, normally accomplished using flame-ionization detection (GC–FID) or GC–MS, would provide improved structural information. Historically, the most common thioacidolysis-
released dimers recovered after Raney-Ni desulfurization were quantified based on the single GC response factor (RF) for the 5–5 (biphenyl) dimer XII (Figure 1), using docosane, tetracosane, or octadecane as the internal standard.\textsuperscript{[9]}

Among the many releasable thioacidolysis dimers, only the 5–5 (biphenyl) dimer has been authenticated by independent synthesis; the remaining dimers have been assigned based solely on analysis of their mass spectra. Accurate quantification of degradation fragments by GC strictly requires individual RFs between each analyte and the internal standard itself, but the authentic compounds required to obtain response factors have not been available. Thus, of the many laboratories using GC or GC–MS for analysis of thioacidolysis dimers, all have reported quantitative results based on the assumption that all dimers share the same GC–FID or GC–MS RF, which seems unlikely in light of their considerable structural diversity. Clearly, although lignin analysis by thioacidolysis has been improved over several decades by various researchers and, primarily, by the originator of the method, the complexity of lignins and the lack of authentic compounds still limits its efficacy to some degree.\textsuperscript{[9a, 10]} In particular, it has been difficult to quantify dimeric and oligomeric products owing to this lack of reference standards. Models for the original structures in lignin can be difficult to synthesize, as are the direct thioacidolysis products, the thioethyl ethers, owing to the abundance of regio- and stereoisomers that thioacidolysis releases. Unexpected side reactions also complicate the results. In some cases, as we show here, the synthesis of the products of thioacidolysis followed by Raney-Ni desulfurization can be somewhat easier.

Recognizing a need for identification and quantification of dimeric degradation products, here we have synthesized a variety of G-type dimeric thioacidolysis products (the same dimers can come from, for example, hydrogenolysis) and then used these G-type lignin-derived dimers I–XII (Figure 1) as standard authentic reference compounds to verify and quantify lignin dimers released by thioacidolysis. We have also synthesized a mono-methylated 5–5-linked dimer (XIII) to be used as an internal standard that, in comparison with the standards used previously, will have an equilibrium partitioning coefficient in organic and aqueous phases closer to those of the analytes and a more similar chemical response.

This newly synthesized library of compounds I–XII (but not yet compound VI, see below) contained the majority of lignin-derived dimeric products known, or expected (in the case of the previously unreported β–[β]-dimers IX and X), to be released from thioacidolysis of G-type lignin. To confirm the presence of these products in thioacidolysis-processed biomass, representative softwood (white spruce and loblolly pine) samples were subjected to the standard thioacidolysis method, followed by Raney-Ni desulfurization. After silylating the reaction mixture using N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA), the recovered degradation products underwent GC–FID and GC–MS analysis. Model compounds I–XII were also subjected to GC–MS and GC–FID analysis under the same conditions, and the resulting mass spectral data (Table S1 and Figure S5 in the Supporting Information) and GC retention times match those obtained from the white spruce sample (Figure 2), thus confirming these twelve synthesized compounds as authentic thioacidolysis products. All analyses were conducted in at least triplicate, with standard deviations all below 5%. 

---

Figure 1. Structures of synthetic G-type standard authentic compounds and the internal standard (IS) used in this study.

Figure 2. Partial (dimer region) GC–FID chromatogram of thioacidolysis products (after Raney-Ni desulfurization) released from white spruce, showing determined peak identities. \* β–[β]-linked dimers are expected but have escaped detection among the degradation products from acidolysis and thioacidolysis in the past. \*\* VI is the newly identified β–5-linked dimer that corrects an analytical quandary.
The quantification of individual “condensed” thioacidolysis-released products, with their characteristic interunit linkages, is used to detail differences among various units in lignins. In the past, however, the required reference compounds for such a quantification were not available. Consequently, individual GC–FID RFs of synthesized dimers I–XII versus compound XIII were measured. Compared to the long-chain hydrocarbons for which RFs have been reported,[9] the synthesized mono-methylated 5–5-linked dimer XIII is structurally closer to the desulfurized lignin-derived thioacidolysis dimers (I–XII) so that they have similar equilibrium partitioning concentrations between organic and water phases, thus providing more accurate results. The RFs determined for dimers I–XII varied from 0.87 to 1.17 (Table 1), rather than being invariant as previously assumed.[9a] These data also suggest that compound XIII is an appropriate internal standard for thioacidolysis dimers, as most of the RF values were close to 1. In other words, all of the dimers had appropriate RFs relative to the new internal standard by GC–FID, but displayed significant differences that, when taken into account, allow for more accurate quantification.

With these RFs in hand, four softwood samples, including two extract-free whole-CW samples and two cellulolytic enzyme lignins (CELs) from loblolly pine and white spruce, were analyzed by thioacidolysis (followed by Raney-Ni desulfurization), again using XIII as the internal standard. The yields of dimers I–XII (μmol g⁻¹ Klason lignin or μmol g⁻¹ CEL lignin) obtained from each sample were adjusted based on each dimer’s experimentally determined RF (Table 1).

The β–5, β–1, and 5–5 dimers (in order of decreasing abundance) are the three main dimeric products released by thioacidolysis from softwood CWs or CELs (Table 1). Although 2D HSQC NMR spectroscopy (Figure 3) showed that the contents of β–5 units B in both white spruce and loblolly pine lignins are close to each other, the total yield of β–5 dimers (I–V) obtained by thioacidolysis of the white spruce sample was much lower than that obtained from the loblolly pine sample (78.6 and 107.7 μmol g⁻¹ Klason lignin, respectively), suggesting a possible issue with the current thioacidolysis product assignments. The total yield of β–5 dimers appeared to vary based on the yields of V, which was authenticated by our model compound, and compound VI (m/z = 488) (Figure 2), which was at that time reported to be a β–6 dimer (but see below).[9b] Indeed, the data suggested an inverse relationship between the yields of V and VI, with a similar combined yield for the two in loblolly pine and white spruce (37.9 and 38.4 μmol g⁻¹ Klason lignin, respectively). In other words, if peak VI was included as a β–5 product (consistent with our NMR analysis), the total yield of β–5 products obtained from the white spruce sample was close to that of loblolly pine. As shown in Table 1, similar results could be observed from the thioacidolysis of CELs. Based on these results and the mass spectral characteristics of VI, this compound appeared to have originated from β–5 structures (somewhat related to V) during

### Table 1. GC-FID RFs of thioacidolysis/Raney-Ni dimers (I–XII) and their yields (μmol g⁻¹ Kl or CEL lignin) from softwood samples.

| Dimers | RFs | Loblolly pine | White spruce | Loblolly pine | White spruce | Spruce MWL |
|--------|-----|---------------|--------------|---------------|--------------|------------|
|        |     | CW            | CW           | CEL           | CEL           |            |
| β–5    |     |               |              |               |              |            |
| I      | 0.92 | 55.7          | 49.8         | 40.0          | 45.0         |            |
| II     | 0.92 | 5.4           | 6.8          | 3.6           | 5.4          |            |
| III    | 0.87 | 7.8           | 7.1          | 3.2           | 4.9          |            |
| IV     | 1.07 | 2.0           | 1.6          | 1.1           | 1.1          |            |
| V      | 0.95 | 36.8          | 13.3         | 15.9          | 23.8         |            |
| VI     | 1.05 | 1.1           | 25.1         | 15.0          | 2.1          |            |
| total  |      | 108.8         | 103.7        | 78.6          | 82.3         | 65         |
| RR [%] |      | 44.5          | 45.2         | 44.7          | 46.5         | 31         |
| β–1    |     |               |              |               |              |            |
| VII    | 0.92 | 51.2          | 48.3         | 35.0          | 32.5         |            |
| VIII   | 0.94 | 8.8           | 10.3         | 11.1          | 7.4          |            |
| total  |      | 60.0          | 58.6         | 46.1          | 39.9         | 66         |
| RR [%] |      | 24.5          | 25.5         | 26.2          | 22.5         | 31         |
| β–β    |     |               |              |               |              |            |
| IX     | 1.15 | (1.26)        | 0.9          | 0.9           | 1.1          | –          |
| X      | 1.17 | (1.02)        | 0.3          | 0.4           | 0.1          | –          |
| total  |      | 1.2           | 1.6          | 1.0           | 1.3          | –          |
| RR [%] |      | 0.5           | 0.7          | 0.6           | 0.7          | –          |
| 4–5    |     |               |              |               |              |            |
| XI     | 1.05 | 19.4          | 15.6         | 15.2          | 16.1         | 16         |
| RR [%] |      | 8.0           | 6.8          | 8.6           | 9.1          | 8          |
| 5–5    |     |               |              |               |              |            |
| XII    | 0.87 | 55.1          | 49.9         | 35.1          | 37.5         | 62         |
| RR [%] |      | 22.5          | 21.8         | 19.9          | 21.2         | 30         |
| total  |      | 244.5         | 229.4        | 176.0         | 177.1        | 209        |

[a] KL = Klason (acid-insoluble) Lignin; KL contents for loblolly pine CW, white spruce CW, loblolly pine CEL, and white spruce CEL are 27.5, 28.6, 90, and 88%, respectively; [b] RFs of IX and X were determined from their selected-ion chromatograms; [c] Relative ratios (RR) for each dimer are their percentages relative to the total identified dimers in this study. In addition, the GC–FID RF of XIII versus docosane is 1.12.
thioacidolysis or desulfurization. The peak for this compound VI (Figure 2) was originally assigned to be a β–6 dimer but, because such coupling is chemically impossible, it was later suggested to be consistent with a β–1 isochroman structure. From MS analysis and the herein established derivation of compound VI from a β–5 dimeric model compound, it is now clear that the structure shown in Figure 1 and Figure S1 is simply derived from an unopened phenylcoumaran, as will be described more fully elsewhere. Its more logical assignment as the β–5 dimer VI shown in Figure 1 not only partially helps to reduce the level of compounds purporting to be β–1-derived, but also solves the apparent discrepancy in yields for β–5 thioacidolysis products between softwood samples. Accordingly, we included the yield of compound VI in the total yield of β–5 dimers for all samples.

The β–5 products released from loblolly pine CW, white spruce CW, loblolly pine CEL, and white spruce CEL accounted for approximately 44.5, 45.2, 44.7, and 46.5% of the total identified dimeric products, respectively, although the yield of individual β–5 products varied between samples, especially for V and VI (Table 1). The concordance in relative abundances is remarkable, but the diversity of products released indicates that thioacidolysis of β–5 structures involves many complicated and competing pathways that might be difficult to exactly replicate between sample runs. The only measurement of real value is the total amount of product derived from a given structure in lignin, and the sum of the β–5 units now has a rather consistent value across these four samples, providing confidence in both the assignment and in the quantification of total β–5 dimers released.

Similar results were observed for the β–1 products VII and VIII. β–1 dimers, the second most abundant of the thioacidolysis dimeric products, arise from the breakdown of the native lignin spirodienones during the reaction. They accounted for 24.5, 25.5, 26.2, and 22.5% of the total identified dimers in loblolly pine CW, white spruce CW, loblolly pine CEL, and white spruce CEL, respectively. Explaining the prevalence of β–1 products in the dimer fraction from various acidolytic degradative methods (acidolysis, thioacidolysis, and DFRC) has always been somewhat difficult especially because it was shown that traditional β–1 dimeric units cannot be found in the lignin polymer and because the levels of the more recently discovered spirodienones that explain how the products of such β–1-coupling arise in the polymer are typically <2% of the units in softwoods. Without knowing the relative propensities for the various coupling modes of the various phenolic end-units in lignin, a partial explanation is that β–1 units in spirodienones can only be etherified at one (latent) phenolic end during lignification, so the chance of that being a 4–O–β-ether is high, and release of such β–1 units might be expected to be higher than from the other condensed units in which, typically, linkages to two other lignin units are present, thus requiring that they both must be β-ethers to release the dimer.

It is well documented that β-β (resinol) structures are present in lignins and can be readily observed by NMR spectroscopy (Figure 3). It has been estimated that pinoresinol structures...
comprise at least 2% of softwood lignins, based on the total C₆ units.[5a, 15] In the past, failure to detect the expected degradation products by acidolysis and thiaoacidolysis (e.g., dimer IX from thioacidolysis) led to the suggestion that pinoresinol structures were connected to the rest of the lignin chain by at least one 5-linkage, which can be explained to some extent,[16] but could mean that they formed in a manner different from the traditional theory of lignification.[15b, 17] The non-observation of (any) β-β dimers from acidolysis or thiaoacidolysis became one of the cornerstones for the championed dismissal of the current combinatorial coupling theory of lignification and was used in strong support of a notion of absolute structural control over lignin assembly by dirigent proteins (or dirigent sites on a putative protein),[18] an over-zealous hypothesis that has now lost a basis for its support.[16] The dilemma from the apparent non-release of β-β dimers by ether-cleaving analytical procedures had already been refuted by studies of DFRC products,[5a] but it was never clear why DFRC would release such dimers whereas thiaoacidolysis would not. In this study we detected the expected β-β degradation product IX, further supporting the idea that either both phenolic ends on pinoresinol units can be involved in 4-O-β-ethers, or that one can while the other phenol remains free, according to the traditional free-radical theory of lignification in softwoods.[5c] Furthermore, dimer X, another pinoresinol degradation product, is identified here for the first time. In total, then, β-β dimers (IX and X) accounted for below 1% (0.5, 0.7, 0.6, and 0.7%) of the dimers released from loblolly pine CW, white spruce CW, loblolly pine CEL, and white spruce CEL, respectively, but can no longer be claimed to be absent from the array of thiaoacidolysis dimers. As shown in Table 1, the total-ion chromatogram and its selected-ion spectrum was used for a better calculation of β-β dimers (IX and X), owing to the possible peak overlap with other products in the GC chromatogram.

The 4-O-5 dimer XI accounted for 8.0, 6.8, 8.6, and 9.1 of the total identified dimers in loblolly pine CW, white spruce CW, loblolly pine CEL, and white spruce CEL, respectively. Until recently, it was not clear why such products were detected at such moderate levels from the various degradative methods although 4-O-5-linked structures had never been detected in NMR spectra of softwood lignins. Two studies have now reported their detection and authentication at low levels, estimated to be no more than approximately 1% of lignin units.[19] These studies also now partially explain why they are prevalent in the thiaoacidolysis dimers beyond their actual proportion in the lignin; both research groups reporting on the findings could find no evidence that 4-O-5 dimeric units in lignin were further etherified, which also implies that they do not form the branching structure typically envisioned.[19]

The 5-5-linked dimer XII comprised 22.5, 21.8, 19.9, and 21.2%, of the total identified dimers in loblolly pine CW, white spruce CW, loblolly pine CEL, and white spruce CEL. Although other biphenyl structures have been reported in lignin,[5a, 11b, 20] dibenzodioxocins D (Figure 3) are evidenced to be the main biphenyl structures present,[5a, 21] so the 5-5 dimers can be considered mostly as dibenzodioxocin degradation products. No estimate of the conversion yield from the eight-membered ring dibenzodioxocins is yet available, and this might not even be accessible from model compounds given that attempted acidolytic cleavage of model dibenzodioxocins can yield stable seven-membered-ring oxepine products.[21a, 22]

CELS from white spruce and loblolly pine were isolated and used for this study. We chose this isolation method instead of using milled wood lignin (MWL), which has commonly been used in previous studies, because CEL preparations have higher yields (ca. 55%)[19a, 23] and should therefore be more representative. In contrast, MWLS obtained from softwoods are typically obtained in approximately 15% yield based on Klason lignin.[20b] Interestingly, the total relative yields of 5-5 and β-1 products released from MWL (as measured in previous studies) were much higher (31 and 30%, respectively) than those from the CELs (19.9–22.5 and 22.5–26.2%, respectively) we studied in this work (Table 1).[20b] With regard to 5-5 dimers, this discrepancy could be explained by the fact that we have included only dimer XII as a 5-5 product, whereas previous studies have included an additional (although unidentified/un-authenticated) 5-5 thiaoacidolysis dimer.[7b] As for the β-1 products, the difference in yields could be attributed to the fact that β-1 structures can be selectively enriched in MWLS, as previously reported.[10a] Nevertheless, the relatively lower abundance of β-1 products detected in CELs may suggest that the CELs used in this study are more representative than the MWLS used in previous studies, as hypothesized above.[7a, 10a] Finally, the total relative yield of β-5 products released from MWL was much lower (31%) than the level we found from the CELs (44.5–46.5%, Table 1); a discrepancy that likely arose either from the omission of dimers III and V in the calculation of β-5 products or the lack of compound-specific RFs in previous studies. It should also be noted that the results obtained for the CELs were calculated based on the CEL weight, rather than on the actual lignin content of the isolate, which may have led to an underestimation of the yields reported in Table 1. The similar relative ratios of dimers recovered after thiaoacidolysis between the CEL and whole cell wall samples suggests that CEL is indeed representative,[23b] and that either material can be used in future studies.

In summary, a total of twelve guaiacyl-type dimeric compounds, including six β-5, two β-1, two β-β, one 4-O-5, and one 5-5, released by thiaoacidolysis from softwood lignins have been verified and quantified with synthesized standard compounds. With the use of these authenticated compounds, β-β products from pinoresinol structures in softwood lignins were verified and quantified for the first time, supporting the traditional theory of lignification that β-β (resinol) structures of softwood lignins are formed through the usual combinatorial β-β-coupling of coniferyl alcohol. Additionally, we have assigned VI as a β-5 product, in contrast to previous studies that assumed it to be a β-6 dimer or a β-1 isochroman. Moreover, the quantification of various thiaoacidolysis-released dimeric products by using individual response factors (RFs) obtained in this work provides the most accurate information about carbon–carbon and diaryl ether units in various lignins that can be released by β-ether cleaving reactions, and is also beneficial to our efforts toward lignocellulosics valorization.
Acknowledgements

The authors would like to thank the China Scholarship Council, State Education Department for supporting F.Y. as a visiting student in the Department of Biochemistry and the Great Lakes Bioenergy Research Center (GLBRC) at the University of Wisconsin. This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science, DE-FC02-07ER64494).

Conflict of interest

The authors declare no conflict of interest.

Keywords: analytical methods • GC-FID • lignin dimers • response factor • synthesis design