Anal Bioanal Chem (2012) 402:983–987
DOI 10.1007/s00216-011-5518-x

TECHNICAL NOTE

Raman-spectroscopy-based noninvasive microanalysis of native lignin structure

Pradeep N. Perera · Martin Schmidt · Vincent L. Chiang · P. James Schuck · Paul D. Adams

Received: 20 July 2011 / Revised: 6 October 2011 / Accepted: 18 October 2011 / Published online: 10 November 2011
© Springer-Verlag (outside the USA) 2011

Abstract A new robust, noninvasive, Raman microspectroscopic method is introduced to analyze the structure of native lignin. Lignin spectra of poplar, *Arabidopsis*, and *Miscanthus* were recovered and structural differences were unambiguously detected. Compositional analysis of 4-coumarate-CoA ligase suppressed transgenic poplar showed that the syringyl-to-guaiacyl ratio decreased by 35% upon the mutation. A cell-specific compositional analysis of basal stems of *Arabidopsis* showed similar distributions of S and G monolignols in xylary fiber cells and interfascicular cells.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-5518-x) contains supplementary material, which is available to authorized users.

P. N. Perera (✉) · M. Schmidt
Energy Biosciences Institute, University of California,
Berkeley, CA 94720, USA
e-mail: pperera@lbl.gov

V. L. Chiang
Forest Biotechnology Group, Department of Forestry
and Environmental Resources, College of Natural Resources,
North Carolina State University,
Raleigh, NC 27695, USA

P. J. Schuck
Molecular Foundry, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, USA

P. D. Adams
Physical Biosciences Division, Lawrence Berkeley National
Laboratory,
Berkeley, CA 94720, USA

P. D. Adams
Department of Bioengineering, University of California Berkeley,
Berkeley, CA 94720, USA

Keywords Lignin · Entropy minimization · Chemometrics · Cell wall · Monolignol · Raman

Introduction

Lignin, a complex polyphenolic biomacromolecule, consisting primarily of coniferyl, sinapyl, and *p*-coumaryl alcohols (also known as the G, S, and H monolignols), is a major component of the cell walls of all vascular plants (Fig. 1). Its presence is crucial for the structural and mechanical integrity of the cell walls. It also provides hydrophobicity for water and nutrient transport and protects cell walls against pathogenic attacks. Lignin's stability is a consequence of its structure, which affects the rate of microbial decomposition of phytomass. Because of its interaction with cellulose and hemicelluloses in the plant cell wall, lignin is a major barrier to the utilization of biomass for energy and for papermaking. There has long been interest in understanding the structural properties of lignin, particularly since lignocellulosics have emerged as a potential sustainable feedstock for replacement of fossil carbon. Lignin has been found to be related to saccharification yield, thus provoking attempts to modify the content/composition of lignin to improve the efficiency of saccharification of lignocelluloses [1, 2].

Current techniques for lignin quantification (e.g., Klason [2], acetyl bromide [4]) and lignin compositional analysis (e.g., pyrolysis gas chromatography/mass spectrometry [5], thioacidolysis [6]) are very useful but labor-intensive and therefore are often not appropriate for high-throughput analysis. Also, methods involving extraction of lignin require large quantities of samples and may result in a biased estimation (due to preferential partial extraction) of the lignin structure while losing cellular specific-

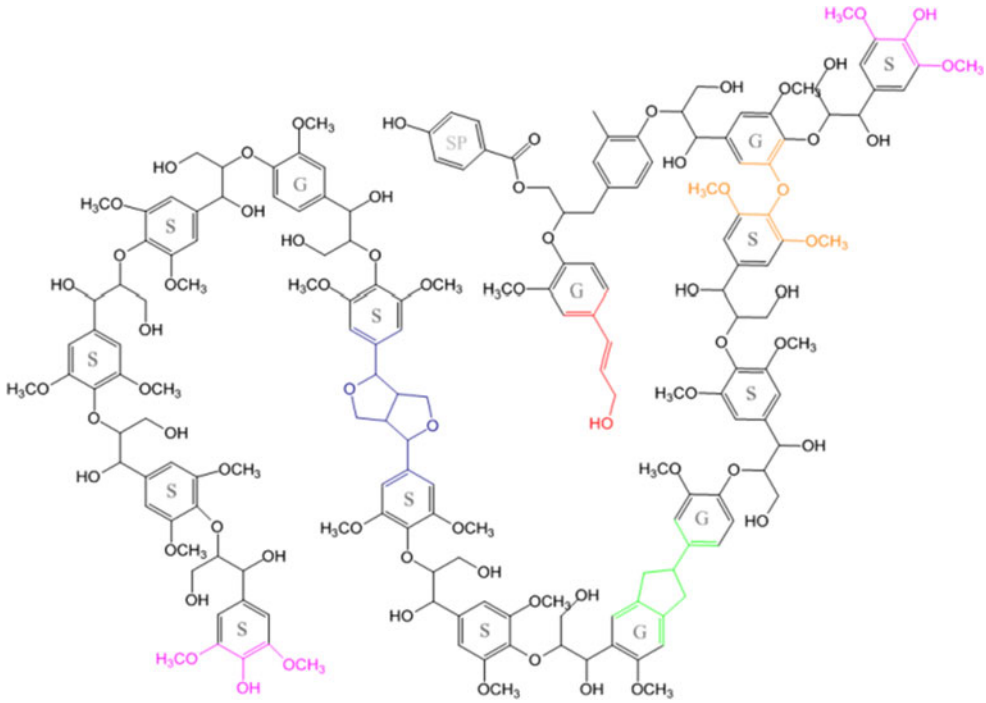

Fig. 1 Representation of a lignin moiety of poplar, as predicted by NMR analysis. *S* syringyl, *G* guaiacyl, *SP* sinapyl *p*-hydroxybenzoate, *black* β-aryl ether linkages, *green* phenylcoumaran, *blue* resinol, *orange* biphenyl ether, *red* cinnamyl alcohol end group, *violet* phenolic end group (From Stewart et al. [3])

ity. NMR-based methods have been routinely used to probe the structure of native lignin [7]. However, this technique also requires mechanical breakdown of the cell wall structures, thus resulting in loss of cell-specific information content.

New analytical tools for lignin analysis which are rapid, noninvasive, unbiased, and specific at the cellular level are highly sought after [8]. In this communication, we describe a new microspectroscopic method for compositional analysis of native lignin using a combined approach of Raman microspectroscopy and chemometrics. This method is unique in three important aspects: (1) it involves neither extraction nor breakdown of the polymeric structure, thus facilitating estimation of the structure of native lignin as it exists in the plant cell wall; (2) the submicron spatial resolution of confocal Raman microspectroscopy allows the estimation of the local lignin structure with cellular specificity; (3) no prior knowledge of the cell wall is needed (e.g., known spectra of isolated lignin), offering an efficient means of tracking varying lignin structure. The method is briefly outlined. Further computational details are given in the electronic supplementary material, and extensive discussion of the method and its merits has been given in a separate publication [9].

As the input, hyperspectral confocal Raman microspectroscopic maps from areas of interest of thin plant sections were collected. The noisy Raman spectra were first preprocessed to mitigate noise (with wavelet decomposition) and the fluorescence baseline was removed by converting the original spectra into their second derivatives. The second-derivative spectral matrix was compressed with principal component analysis. The resultant principal component analysis matrix was subjected to *k*-means clustering to identify compositionally different regions within the cell wall map (see Fig. S1). Spectral identities of these spatial regions were calculated by averaging all original spectra within each cluster. Each average spectrum is composed of a linear combination of the spectra of pure components (e.g., lignin, cellulose) of the cell wall. The average spectral matrix (**D**) was deconvolved to estimate the spectra of the pure components with an objective function based on spectral entropy minimization methodology and spectral constraints of nonnegativity of intensity and weights [10] with a stimulated annealing optimization algorithm [11]. The spectral entropy minimization methodology relies on the fact that the entropies (as defined in Eq. 4) of the spectra of the pure components are smaller than those of spectra arising from mixtures of the components. Spectra of pure components were determined as follows.

First, the **D** matrix was subjected to singular value decomposition:

$$D_{k\times v} = U_{k\times k} \times S_{k\times v} \times V^{T}_{v\times v}. \quad (1)$$

Different linear combinations of rows of the right-hand singular matrix (**V**) will reproduce all the spectra of the mixtures of the components as well as the spectra of the pure components in the system. Therefore, the goal is to find the combinations of weighting coefficients that generate the spectra of the pure components.

The first pure spectral estimate was calculated by multiplying each row of the singular matrix (**V**) with a random number vector ($\mathbf{T}_{1xj}$):

$$\widehat{a}_{1\times v} = T_{1\times j} \times V^{T}_{j\times v}. \quad (2)$$

The coefficient vector (**T**) is refined by minimizing an objective function based on spectral entropy (Eq. 3) and nonnegativity constraints of intensity and spectral weights (Eq. 5) to obtain the spectra of the pure components:

$$Obj = -\Sigma_v\, h_v \ln h_v + P. \quad (3)$$

The entropy term is

$$h_v = \left| {}^{d(\widehat{a}_v)}/_{dv^n} \right| \Big/ \sum \left| {}^{d(\widehat{a}_v)}/_{dv^n} \right|. \quad (4)$$

The penalty function is

$$P(\widehat{a}_{1\times v}, \widehat{c}_{k\times 1}) = F(\widehat{a}_{1\times v}) + F(\widehat{c}_{k\times 1}). \quad (5)$$

$F(\widehat{c}_{k\times 1})$ and $F(\widehat{a}_{1\times v})$ represent penalty functions for nonnegativity of spectral weights and intensity respectively.

This mathematical procedure allows us to separate carbohydrate spectral features from the lignin spectrum and vice versa, efficiently.

To date, Raman microspectroscopy has been utilized to map out the relative abundance of lignin in cell walls, in general, based on the lignin signature peaks between 1,590 and 1,660 cm^{-1}, the aryl ring stretching modes of lignin [12–14]. However, the inability of traditional data analysis methods to unambiguously detect the vibrational signatures of different monolignol units of lignin due to high spectral overlap with signals from cellulose and hemicelluloses has severely limited the utility of Raman microspectroscopy for in situ lignin structure analysis. Furthermore, chemical and extraction modifications of the lignin structure result in significant changes in the Raman spectra, thus preventing knowledge of the native structure being obtained [16]. In contrast, the ability of our data analysis method to recover the complete Raman lignin spectrum, including the peaks that are unique to different monolignol units, allows us to analyze the native lignin composition in situ without extraction or chemical or mechanical degradation of the polymeric structure of lignin as demonstrated in the remainder of this publication.

Estimated native lignin spectra from three different angiosperms, namely, fiber cells of poplar (*Populus trichocarpa*—a woody angiosperm), interfascicular cells of *Arabidopsis* (*Arabidopsis thaliana*—a nonwoody angiosperm), and a potential energy crop, the perennial grass *Miscanthus* (*Miscanthus giganteus*), are shown in Fig. 2. These spectra clearly show, other than the peak at approximately 1,600 cm^{-1} (aryl stretching mode [13]), peaks that are characteristic of different monolignol units.

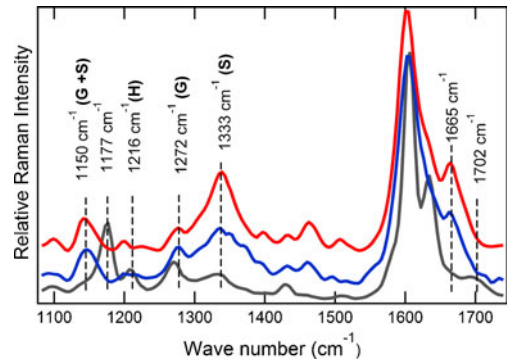

Fig. 2 Estimated Raman lignin spectra of *Arabidopsis* (*blue*), poplar (*red*), and *Miscanthus* (*black*). Spectra are vertically offset for clarity. Peak assignments are based on literature references [13, 15]

We have previously assigned all major peaks of the estimated lignin spectrum of *Populus trichocarpa* to known vibrational modes of lignin, based on the literature [9], thus eliminating the possibility of spectral contaminations from other components in the cell wall, including pigments.

A comparison of the Raman spectra of extracted lignin from *Miscanthus* and poplar with the estimated lignin spectra of the respective plants further validated our procedure (see Figs. S3, S4). The lignin spectrum of *Miscanthus*, a monocot, is substantially different from the other two lignin spectra. Notably, the aryl ring stretching mode in the lignin spectrum of *Miscanthus* is narrower than the corresponding modes of the two dicots. Assuming that peak broadening is dominated by inhomogeneities of the local structure, we conclude that lignin structure in grass is relatively more homogeneous than that of the dicots. This observation is further supported by comparatively narrower peaks at 1,270 cm^{-1} and 1,177 cm^{-1}. Also, the lignin spectrum of *Miscanthus* displays fewer peaks than the spectra of the two dicots and is less congested between 1,240 cm^{-1} and 1,400 cm^{-1}. It is known from wet chemical analyses of bulk *Miscanthus* samples that more than 90% of the lignin exists as "noncondensed" lignin [17] compared with 40-50% in dicots, thus providing a rationale for the homogeneous nature. As indicated by the relatively intense peak at 1,211 cm^{-1} [15], *Miscanthus* lignin is also shown to contain measurable amounts of hydroxyphenyl (H) units, which is in contrast to the two dicots. These H units are generally difficult to track with lignin extraction methods as they are associated with the "condensed lignin" part of the polymer which is hard to extract from the cell wall matrix. Furthermore, the shoulder peak at 1,665 cm^{-1} which has been assigned to conjugated C=C and C=O modes of coniferyl alcohol and coniferaldehyde, respectively [13], is absent in *Miscanthus* lignin, and a new broad peak appears at 1,702 cm^{-1}. The latter peak could arise from the acyl groups of highly acylated lignin in *Miscanthus* [17]. Further structural differences are indicated by the phenolic mode of

Miscanthus which has narrowed and shifted to 1,177 cm^{-1}, compared with 1,154 cm^{-1} for the two dicots. A similar shift for the phenolic mode of guaiacyl (G) units of *Miscanthus* lignin was shown in a recent study [18].

Signature peaks of G units at approximately 1,270 cm^{-1} and syringyl (S) units at approximately 1,333 cm^{-1} can be used to calculate the relative abundance of these units. We calculated the S-to-G ratio for these lignin spectra as follows. First, the region between 1,240 cm^{-1} and 1,380 cm^{-1} was fitted to a set of four Gaussians (not shown) which include peaks at 1,272-1,276 cm^{-1} and 1,332-1,336 cm^{-1} to represent G and S units, respectively. S-to-G ratios were calculated by using the area under the G and S Gaussians peaks and the Raman cross-sectional ratios of those peaks (determined by using lignin model compounds). We calculated S-to-G molar ratios of 0.5±0.08, 0.6±0.1, and 1.9±0.2 for *Miscanthus*, *Arabidopsis*, and poplar, respectively. These ratios are comparable to reported values for stem cell tissues: 0.7 for *Miscanthus* (milled wood lignin obtained by 2D NMR spectroscopy and thioacidolysis) [17], 0.2 [19] to 0.3 [20] for *Arabidopsis*, and 1.8–2 for poplar (NMR spectroscopy and thioacidolysis) [21]. Although the results are in reasonably good agreement with results obtained by other methods, our measurements are specific to particular cell types, whereas the ratios determined by NMR spectroscopy and thioacidolysis represent average values for stem tissues. To highlight the differences, for example, interfascicular cells from floral stems of *Arabidopsis* have been shown to have an S-to-G ratio of 0.64 compared with 0.3 for whole-stem samples [20].

Thus, whenever methods are used that provide bulk measurements for parts of the plant (e.g., stem, leaves), such as NMR spectroscopy and thioacidolysis, information on cell-specific variations in lignin composition and structure may be masked by ensemble averaging. As shown by histochemical analysis [22] and recently also by laser capture microdissection [20], lignin distribution and monolignol composition show spatially and temporally variable behavior. We demonstrate the ability of our method to extract cell-specific lignin spectra and thereby to determine the lignin composition of interfascicular and xylary fiber (fibers plus vessels) cells in the vascular bundle of basal stems of *Arabidopsis thaliana* (see Figs. S5, S6). Our results indicate statistically similar S-to-G ratios for xylary fiber cells (0.76±0.14) and for interfascicular cells (0.64±0.12) of basal stems from senescent *Arabidopsis* plants. In contrast, compared with interfascicular cells, the vascular bundle of floral stems studied by laser capture microdissection followed by gas chromatography/mass spectrometry was found to have a factor of 2 lower S-to-G ratio [20]. However, a qualitative histochemical staining of basal stems has shown that both interfascicular cells and xylary fibers react positively to S and G staining, although no quantitative estimates were presented [22].

An important application of our method is the rapid compositional and structural screening of mutant phenotypes as demonstrated in the following example. A comparison between the lignin spectra from transgenic poplar deficient in lignin (transgenic suppression of a monolignol biosynthesis gene encoding 4-coumarate-CoA ligase [2]) and wild-type poplar is shown in Fig. 3. The lignin content was shown to be reduced by 28% and 40% (based on the area under the aryl stretching peak) for two transgenic lines. More importantly, by fitting the region between 1,240 cm^{-1} and 1,380 cm^{-1} to four Gaussians, including two at 1,272 cm^{-1} and 1,333 cm^{-1} (inset in Fig. 3), we found that the S-to-G ratio had decreased from 1.9±0.2 in the wild type to 1.2±0.05, which corresponds to an approximately 35% reduction for both transgenic lines relative to the wild type. Further evidence of structural changes in lignin caused by the transgenic suppression is confirmed by the shift of the 1,140-cm^{-1} peak of the wild type to 1,155 cm^{-1} for both transgenic lines.

In conclusion, a novel method that combines Raman microspectroscopy and chemometrics has been successfully employed to estimate the spectrum of native lignin in three different plant species as it occurs in the plant cell wall. The estimated Raman lignin spectrum of *Miscanthus* was clearly distinct from the spectra of *Arabidopsis* and poplar. A cell-specific examination of xylary fiber cells and interfascicular cells of senescent basal stems of *Arabidopsis* found a similar monolignol composition in contrast to the finding for floral stems. We have also shown that transgenic suppression of 4-coumarate-CoA ligase in poplar resulted not only in a reduction of lignin content in the plant cell wall but also in a significant change in the S-to-G ratio. Moreover, efforts are under way to identify Raman

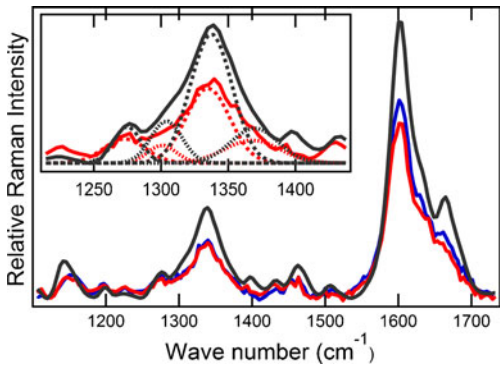

Fig. 3 Estimated Raman lignin spectra of wild-type poplar (*black*) and two lines of lignin-deficient transgenic 4-coumarate-CoA ligase (4CL) poplar (*red* and *blue*). *Inset*: Peak-fitted region of 1,235–1,400 cm^{-1} of wild-type and transgenic poplar. Each spectrum has been fitted with four Gaussians. *Dashed lines* represent the G signature peak (1,273 cm^{-1}) and the S signature peak (1,336 cm^{-1}) of wild-type polar (*black*) and 4CL poplar (*red*). The *dotted Gaussians* are due to other lignin peaks, which are inconsequential for determining the S-to-G ratio

signatures of other important lignin units (e.g., phenylcoumaran, resinol) and noncondensed lignin, i.e., Raman signatures of β-O-4 linkages, which are important for the recalcitrance of the cell wall. We will thus be able to unlock more structural details using the full lignin spectrum, extending the utility of our method in the future.

Acknowledgements We thank Andrew Carroll for providing *Arabidopsis* samples, Purbasha Sarkar for providing *Miscanthus* samples, and Lan Sun and Stefan Bauer for assistance with extracting lignin from *Miscanthus*. This work was supported by the Energy Biosciences Institute, University of California, Berkeley, CA, USA. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH1123. The work on transgenic *Populus trichocarpa* was supported by a grant from the National Research Initiative of the USDA CSREES #2006-35504-17233 to V.L.C.

References

1. Chen F, Dixon RA (2007) Nat Biotechnol 25(7):759–761. doi:10.1038/nbt1316
2. Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Nat Biotechnol 17(8):808–812
3. Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD (2009) Plant Physiol 150(2):621–635. doi:10.1104/pp.109.137059
4. Iiyama K, Wallis AFA (1990) J Sci Food Agric 51(2):145–161
5. Ralph J, Hatfield RD (1991) J Agric Food Chem 39(8):1426–1437
6. Lapierre C, Monties B, Rolando C (1986) Holzforschung 40 (2):113–118
7. Wagner A, Ralph J, Akiyama T, Flint H, Phillips L, Torr K, Nanayakkara B, Kiri LT (2007) Proc Natl Acad Sci USA 104 (28):11856–11861. doi:10.1073/pnas.0701428104
8. Anterola AM, Lewis NG (2002) Phytochemistry 61(3):221–294
9. Perera PN, Schmidt M, Schuck PJ, Adams PD (2011) Anal Chim Acta 702(2):172–177. doi:10.1016/j.aca.2011.06.021
10. Widjaja E, Garland M (2008) Anal Chem 80(3):729–733
11. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Science 220 (4598):671–680
12. Schmidt M, Schwartzberg AM, Perera PN, Weber-Bargioni A, Carroll A, Sarkar P, Bosneaga E, Urban JJ, Song J, Balakshin MY, Capanema EA, Auer M, Adams PD, Chiang VL, Schuck PJ (2009) Planta 230(3):589–597. doi:10.1007/s00425-009-0963-x
13. Gierlinger N, Schwanninger M (2006) Plant Physiol 140(4):1246–1254. doi:10.1104/pp.105.066993
14. Agarwal UP (2006) Planta 224(5):1141–1153. doi:10.1007/s00425-006-0295-z
15. Sun L, Simmons BA, Singh S (2011) Biotechnol Bioeng 108 (2):286–295. doi:10.1002/bit.22931
16. Agarwal UP, McSweeny JD, Ralph SA (2011) J Wood Chem Technol 31:324–344
17. Villaverde JJ, Li JB, Ek M, Ligero P, de Vega A (2009) J Agric Food Chem 57(14):6262–6270. doi:10.1021/jf900483t
18. Chu LQ, Masyuko R, Sweedler JV, Bohn PW (2010) Bioresour Technol 101(13):4919–4925. doi:10.1016/j.biortech.2009.10.096
19. Vanholme R, Ralph J, Akiyama T, Lu FC, Pazo JR, Kim H, Christensen JH, Van Reusel B, Storme V, De Rycke R, Rohde A, Morreel K, Boerjan W (2010) Plant J 64(6):885–897. doi:10.1111/j.1365-313X.2010.04353.x
20. Ruel K, Berrio-Sierra J, Derikvand MM, Pollet B, Thevenin J, Lapierre C, Jouanin L, Joseleau JP (2009) New Phytol 184(1):99–113. doi:10.1111/j.1469-8137.2009.02951.x
21. Leple JC, Dauwe R, Morreel K, Storme V, Lapierre C, Pollet B, Naumann A, Kang KY, Kim H, Ruel K, Lefebvre A, Joseleau JP, Grima-Pettenati J, De Rycke R, Andersson-Gunneras S, Erban A, Fehrle I, Petit-Conil M, Kopka J, Polle A, Messens E, Sundberg B, Mansfield SD, Ralph J, Pilate G, Boerjan W (2007) Plant Cell 19(11):3669–3691. doi:10.1105/tpc.107.054148
22. Patten AM, Jourdes M, Cardenas CL, Laskar DD, Nakazawa Y, Chung BY, Franceschi VR, Davin LB, Lewis NG (2010) Mol Biosyst 6(3):499–515. doi:10.1039/b819206e