nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-019-0191-1

Determination of hydroxyl groups in biorefinery resources via quantitative ^{31}P NMR spectroscopy

Xianzhi Meng[1], Claudia Crestini[2*], Haoxi Ben[3], Naijia Hao[1], Yunqiao Pu[4], Arthur J. Ragauskas[1,4,5] and Dimitris S. Argyropoulos[6]

The analysis of chemical structural characteristics of biorefinery product streams (such as lignin and tannin) has advanced substantially over the past decade, with traditional wet-chemical techniques being replaced or supplemented by NMR methodologies. Quantitative ^{31}P NMR spectroscopy is a promising technique for the analysis of hydroxyl groups because of its unique characterization capability and broad potential applicability across the biorefinery research community. This protocol describes procedures for (i) the preparation/solubilization of lignin and tannin, (ii) the phosphitylation of their hydroxyl groups, (iii) NMR acquisition details, and (iv) the ensuing data analyses and means to precisely calculate the content of the different types of hydroxyl groups. Compared with traditional wet-chemical techniques, the technique of quantitative ^{31}P NMR spectroscopy offers unique advantages in measuring hydroxyl groups in a single spectrum with high signal resolution. The method provides complete quantitative information about the hydroxyl groups with small amounts of sample (~30 mg) within a relatively short experimental time (~30–120 min).

Introduction

Over the past decade, advances in genetic engineering and biotechnology have led to a new manufacturing concept directed at converting renewable lignocellulosics to valuable fuels and products, generally referred to as the biorefinery[1]. The development of biomass-based biorefineries, in turn, has been the main impetus for developing methods to assess the characteristics of the biorefinery feedstocks, process intermediates, and final products. A strong case can be made that substrate characterization in biorefinery has emerged as a research field in its own right. Characterization of biorefinery resources has been examined by several state-of-the-art analytical techniques including Fourier transform infrared spectroscopy (FTIR)[2], fluorescence spectroscopy[3], HPLC[4], gas chromatography–mass spectrometry (GC–MS)[5], NMR spectroscopy[6], gel permeation chromatography (GPC)[7], scanning electron microscopy[8], tunneling electron microscopy[9], atomic force microscopy[10], Raman spectrometry[11], time-of-flight secondary ion mass spectrometry (ToF-SIMS)[12], and small-angle neutron scattering[13] along with a host of wet-chemistry and biological assays.

Phosphitylation followed by ^{31}P NMR spectroscopy analysis is a promising technique for the analysis of hydroxyl groups. Since the initial publication series entitled "^{31}P NMR in wood chemistry"[14–21], over 500 research efforts have used or adopted the proposed technique for the analysis of lignins and other biomass-related products and product streams. However, a survey of the literature identified that on occasion there have been some iterations of the method (mainly related to internal standards (ISs), solvents used, and sometimes data interpretation) without clear experimental data justifying these choices, which may lead to misleading results. Hence, at this juncture, there is a clear need to publish a well-defined protocol to help maintain and promote the scientific accuracy and uniform application of this valuable methodology as currently used by both academia and industry. In this protocol, we describe how to perform quantitative ^{31}P liquid-state NMR spectroscopy analysis on various organosolv lignin and tannin samples[22,23]. In all cases, the samples need to be sulfur free, rich in hydroxyl groups, and of limited carbohydrate contamination. The same protocol can also be

UNCORRECTED PROOF

[1]Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN, USA. [2]Department of Molecular Sciences and Nanosystems, Ca' Foscary University of Venice, Mestre,, Italy. [3]State Key Laboratory of Bio-Fibers and Eco-Textiles, Qingdao University, Qingdao, China. [4]Center for Bioenergy Innovation (CBI), Joint Institute for Biological Sciences, Biosciences Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA. [5]Department of Forestry, Wildlife and Fisheries, Center of Renewable Carbon, The University of Tennessee Institute of Agriculture, Knoxville, TN, USA. [6]Departments of Chemistry and Forest Biomaterials, North Carolina State University, Raleigh, NC, USA.
*e-mail: claudia.crestini@unive.it; aragausk@utk.edu; dsargyro@ncsu.edu

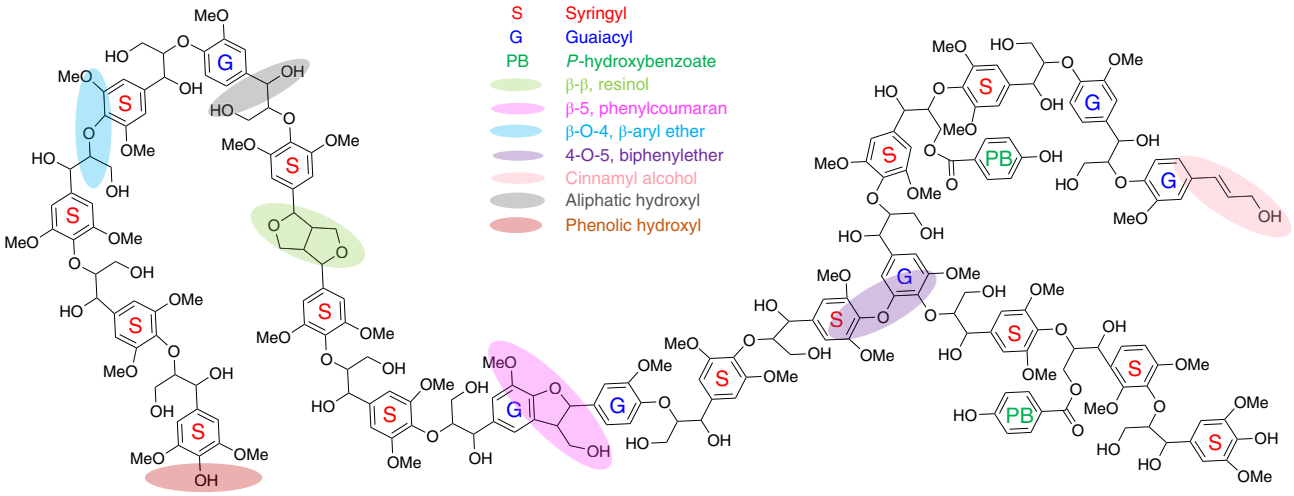

Fig. 1 | A representative structural model of hardwood lignin, as predicted from NMR-based lignin analysis. Adapted with permission of the American Society of Plant Biologists, from "Lignin biosynthesis and structure," Vanholme et al., *Plant Physiol.* Vol. 153, 2010.

extended to polymer chemistry[24–26] and biological systems such as lipid metabolites[27], as long as hydroxyl and carboxyl groups are presented.

Structure of lignin and tannin

Lignin, a 3D heterogeneous phenolic polymer in the plant cell wall, is generated in substantial amounts in most current biorefinery systems that focus mainly on converting plant polysaccharides to liquid fuels. In lignocellulosic plants, lignin is synthesized by combinatorial free radical polymerization of phenylpropanoid monomers, namely, coniferyl, sinapyl, and *p*-coumaryl alcohols, giving rise to guaiacyl (G), syringyl (S), and *p*-hydroxyphenyl (H) subunits[28]. In addition, other types of flavonoids, stilbenes, and hydroxyphenylpropanoid compounds including tricin[29], hydroxystilbenes[30], hydroxycinnamaldehydes[31], hydroxycinnamyl acetates[32], dihydrocinnamyl alcohols[33], and the catechol-based hydroxycinnamyl alcohols caffeyl and 5-hydroxyconiferyl alcohols[34], all derived from the monolignol biosynthetic pathway, have also been reported to be subunits of lignin in either wild-type or transgenic plants[35]. The C_9 phenylpropane units are primarily connected through ether (e.g., β-O-4, α-O-4, 4-O-5) and carbon–carbon (e.g., β-β, β-5, β-1, 5-5) linkages, with β-O-4 being the dominant linkage. The lignin macromolecule also contains various functional groups including hydroxyl, carbonyl, methoxy, and carboxyl groups that have an effect on lignin's reactivity. A representative structure of hardwood lignin is shown in Fig. 1, which does not depict the actual structure of lignin but outlines the common linkages and functional groups known to occur in lignin and their relative frequencies determined by NMR spectroscopy studies[36,37]. Despite high-value opportunities, lignin is still significantly underutilized in the current biorefinery systems, with the bulk of technical lignins being used to meet internal energy demands by combustion[28]. The effective use of lignins in a future integrated biorefinery process strongly depends on the fundamental understanding of their chemical structures, and consequently, the structural analysis of lignin has indeed become a subdiscipline of its own[28]. The inherent structural complexity of lignin brings substantial barriers to traditional analytical methods. Although a suite of methods have been developed, such as thioacidolysis[38], oximation[39], ozonation[40], alkaline nitrobenzene oxidation[41], permanganate oxidation[42], and derivatization followed by reductive cleavage (DFRC)[43], these techniques usually suffer from disadvantages such as being tediously laborious and involving many steps prone to significant error[44].

Proanthocyanidins are flavan-3-ol-based oligomers and polymers, also termed condensed tannins. They are ubiquitous in plants and constitute up to 40% of the dry weight of the bark in several species[45–48]. Their structures are depicted in Fig. 2[22,49,50]. Proanthocyanidins are currently exploited in a wide range of applications, including use in dyes for inks and textiles, waste water amendment, adhesives, binder formulations, biomedical and nutraceutical compounds, cosmetics, and food. The phenolic substitution patterns in rings A and B are characteristic of each proanthocyanidin species, as shown in Fig. 2c[22]. Ring B presents either a catechol or a pyrogallol hydroxylation pattern, whereas

UNCORRECTED PROOF

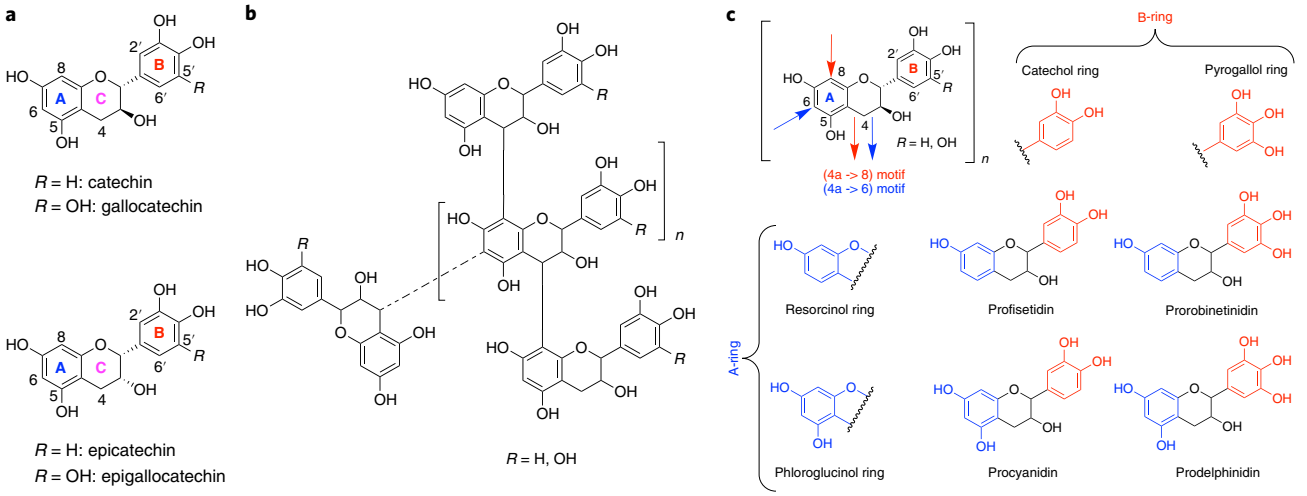

Fig. 2 | Structural features of proanthocyanidin and its subunits. a, Stereocenters. **b**, Regiochemistry of polymerization. **c**, Possible substitution patterns of A and B rings. Adapted with permission from ref. 22, American Chemical Society.

the phenolic distribution in ring A has a resorcinol or phloroglucinol hydroxylation pattern. These structural features determine the complexing, antioxidant, and biological properties of tannins[50–52]. The specific phenolic substitution patterns result in different responses to the commonly used analytical protocols based on the determination of the phenolic groups' content or protein precipitation[53–55]. This implies that the overall amount of proanthocyanidin or phenolic groups in a sample cannot be reported as a single value and does not allow the use of specific standards.

Methods for characterization of lignin and tannin

Lignin and tannin characterization has advanced substantially over recent decades, with conventional methods being replaced or supplemented by chromatographic and spectroscopic approaches such as pyrolysis GC–MS[56], ToF-SIMS[12], FTIR[57], near-infrared spectroscopy (NIR)[58], and NMR[59]. Among these techniques, no other single technique has been more comprehensively used to offer structural insight into lignin than NMR technology, which has advanced the structural knowledge of lignin. Simple 1D ^{1}H and ^{13}C NMR have been widely used to characterize alkyl groups, aliphatic/phenolic hydroxyl groups, methoxy groups, aromatic O/C/H structures, aldehydes, ketones, and β-O-4 substructures in lignin[60,61]. 2D heteronuclear single-quantum coherence (HSQC) NMR is also attracting significant attention owing to its versatility in the determination of lignin interlinkages and subunits, lignin–carbohydrate complexes (LCCs)[62,63]. Both ^{13}C NMR (liquid state and cross-polarization magic angle spinning solid state) and HSQC NMR have been used to determine the compositional aspects of tannins[64–66]. Furthermore, some gel-state and liquid-state 2D whole-cell-wall NMR techniques using different solvent systems such as dimethyl sulfoxide (DMSO)-d_6, DMSO-d_6/pyridine-d_5, and DMSO-d_6/deuterated hexamethylphosphoramide have been introduced to provide insights into the full array of polymers that comprise the plant cell wall, and this procedure requires only fine grinding of the biomass[59,67–69]. In spite of the advances in these NMR approaches, both 1D (e.g., ^{1}H and ^{13}C NMR) and 2D correlation (e.g., HSQC) techniques have their own limitations: 2D NMR is typically not fully quantitative, and 1D NMR usually suffers from the spectral overlap of functionality. More recently, the elucidation of lignin structure by quantitative 2D HSQC (Q-HSQC) NMR pulse sequences has been developed and has provided invaluable contributions to the understanding of lignin's structural details, including the degree of polymerization and the presence and role of branch units[70–73]. However, Q-HSQC still cannot provide full quantitative information related to the nature and amount of the specific functional groups present on the lignin backbone.

An alternative approach to the abovementioned ^{13}C and ^{1}H NMR methodologies is to selectively label functional groups such as hydroxyl groups with a specific NMR-active nucleus and then analyze the derivatized substrate by NMR. These active nuclei normally have much broader chemical shift ranges with less signal overlap and less interference from the homonuclear coupling. Most important, however, is the fact that the solvent normally has no effect on the NMR signal and thus solvent suppression is not necessary[74]. Phosphitylation followed by ^{31}P NMR analysis was one of the first

UNCORRECTED PROOF

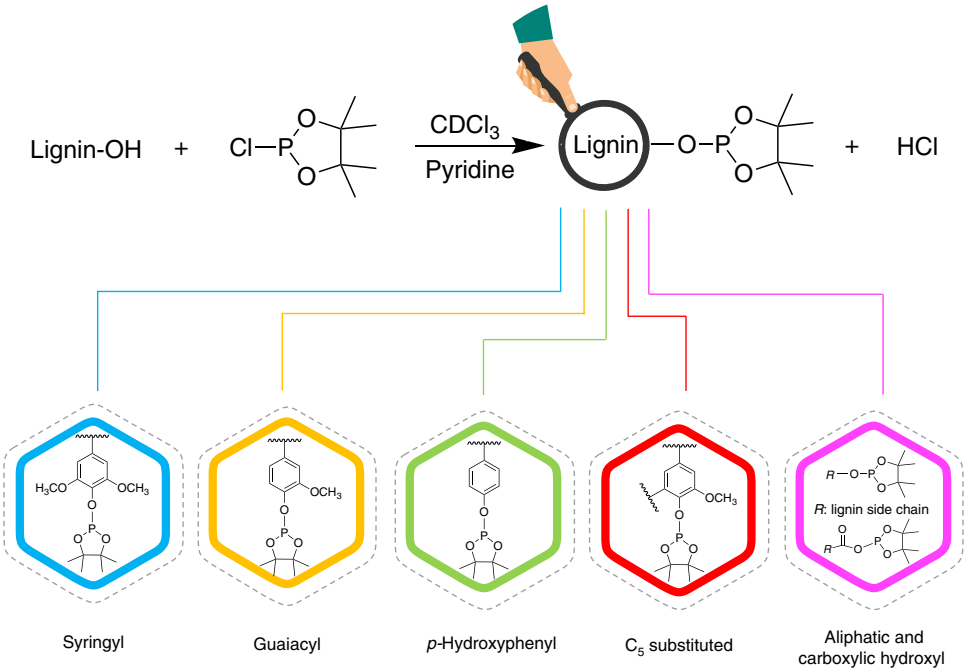

Fig. 3 | The phosphitylated products of the reaction between hydroxyl groups in lignin and TMDP in the presence of a solvent mixture of pyridine and deuterated chloroform.

UNCORRECTED PROOF

methodologies to use this approach to characterize the hydroxyl groups in coal pyrolysis condensates and lignins[14–16,75]. With appropriate phosphorus reagents, different hydroxyl groups including aliphatic, phenolic, and carboxylic, may now be readily quantified with ^{31}P NMR spectroscopy. Hydroxyl groups, particularly phenolic, are one critical functionality that affects the physical and chemical properties of lignin and tannins. The role of the phenolic hydroxyl groups in lignin has been reported to be extremely important for its overall reactivity[76]. For example, the existence of phenolic hydroxyl groups in lignin is known to promote the base-catalyzed cleavage of interunit linkages of lignin and oxidative degradation in the commercial pulping process[76]. Furthermore, the phenolic hydroxyl group is critical in determining antioxidant activity[77], thermal and oxidative characteristics[78,79], and the properties of the resulting materials[77,80]. A recent study also suggests that the condensed phenolic hydroxyl groups are a primary component in the nonproductive binding between lignin and cellulase enzymes, which detrimentally affects the biological conversion of biomass to simple sugars and subsequently the production of biofuels[62]. The content and regiochemical details of phenolic hydroxyl groups also regulate the protein-binding capacities and antioxidant activities of tannins[81]. Thus, the quantitative determination of hydroxyl groups in lignin and tannins is essential, offering vital information related to the chemical and biological reactivity of lignin and tannins in various biorefinery processes. In the following sections we provide an overview of the ^{31}P NMR analysis protocol, using lignin as an example, as well as the development and application of this NMR technique in other areas.

Overview of the ^{31}P NMR protocol

A typical ^{31}P NMR experiment involves the phosphitylation of hydroxyl groups in the substrate using an appropriate ^{31}P reagent (Steps 3–7) followed by quantitative NMR spectroscopy analysis (Steps 8–13), data processing (Steps 14–18), and calculation of the amounts of different hydroxyl groups (Steps 19–23). Since the introduction of 2-chloro-4,4,5,5-tetramethyl-1,3-2-dioxaphospholane (TMDP) in 1995, the reagent has seen wide applicability, and it is now the most commonly employed phosphitylating reagent for hydroxyl group analysis of biorefinery resources[82]. Figure 3 shows the phosphitylated products of the reaction between various hydroxyl groups (e.g., aliphatic, phenolic, and carboxylic acid) in lignin and TMDP in the presence of a solvent mixture of pyridine (major) and deuterated chloroform (minor). The HCl liberated during the reaction could cause decomposition to

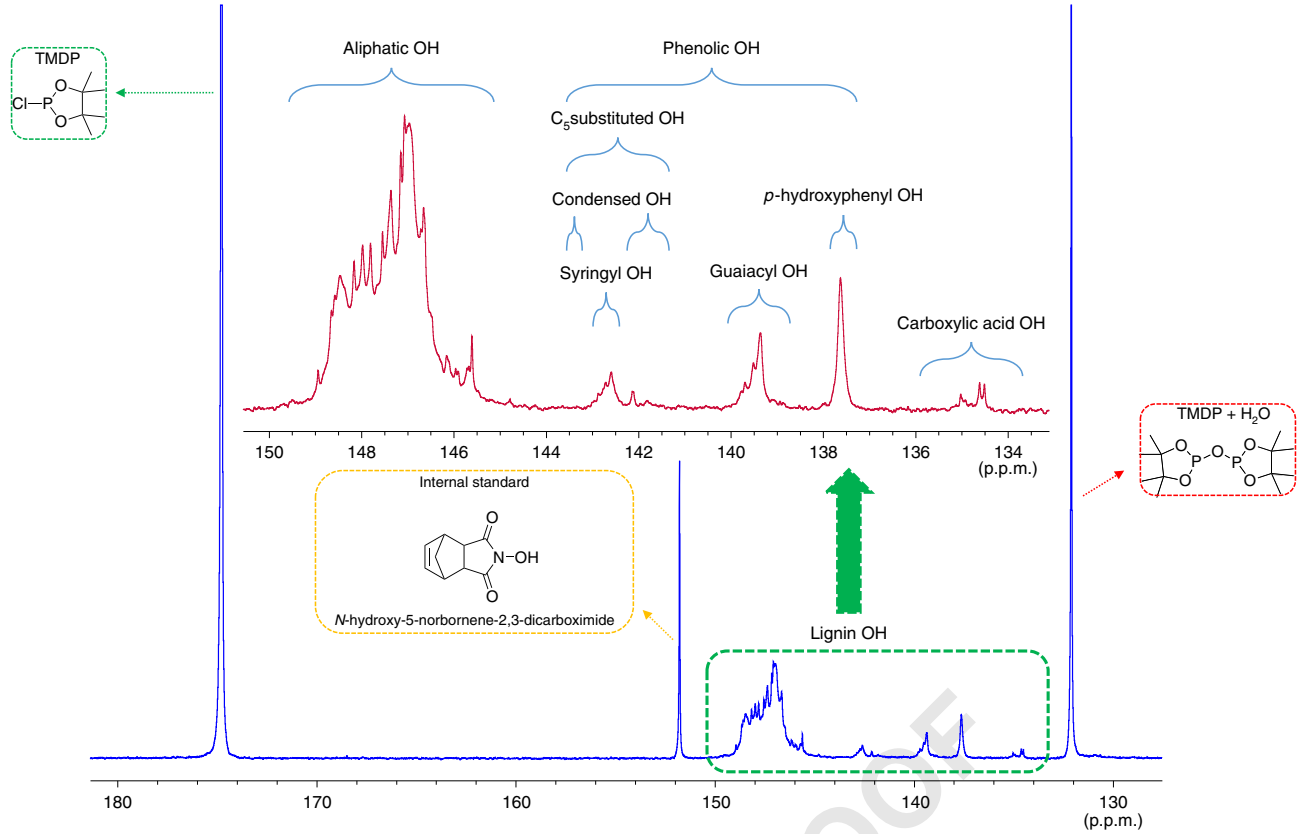

Fig. 4 | A quantitative ^{31}P NMR spectrum of a hardwood poplar lignin derivatized with TMDP using NHND as an IS.

the derivatized compounds, but it is captured by the excess amount of pyridine in the solvent mixture used, to form a pyridine–HCl salt. Deuterated chloroform is also added, offering a triple function: it helps dissolve the sample (along with the added pyridine), it prevents precipitation of the pyridine–HCl salt, and it provides the necessary deuterated signal for NMR signal locking. For the analysis of other types of polymers that have good solubility in $CDCl_3$, the amount of pyridine in the solvent mixture can be reduced, as it serves only as an HCl-trap reagent in that case. After phosphitylation, a quantitative ^{31}P NMR spectrum can then be recorded and the phosphitylated hydroxyl groups are subsequently quantitatively assessed against an internal standard (IS). Figure 4 shows a typical ^{31}P NMR spectrum of hardwood lignin derivatized with TMDP using *N*-hydroxy-5-norbornene-2,3-dicarboximide (NHND) as an IS. The selection of solvent, ^{31}P reagent, and IS, as well as the peak assignments and integration regions, is further discussed in detail in the following sections.

Development of the ^{31}P NMR protocol

In the early and mid-1990s, Argyropoulos's group published a series of papers with the general title "^{31}P NMR in Wood Chemistry" focused on determining the hydroxyl groups in lignin model compounds, carbohydrates, LCCs, and lignins derived from various processes[14–21]. In these papers, an extensive effort using 2-chlorine-1,3,2-dioxaphospholane (CDP) as the phosphorus reagent was made to prove that different hydroxyl groups in model compounds including carboxylic acid, guaiacyl, syringyl, and *p*-hydroxyphenyl offered notably well-separated ^{31}P NMR signals. Furthermore, significant differentiation also occurred between primary and secondary hydroxyl groups and between the erythro and threo conformations of β-O-4 structures, as they gave rise to sharp, single, and well-separated ^{31}P NMR signals[16,83]. Subsequent studies suggested that one of the drawbacks of using CDP as the phosphorus reagent is that signal overlap is usually observed between syringyl phenolic, primary hydroxyls, and C_5 substituted condensed phenolic groups[82,84]. To address this issue, Argyropoulos recommended the use of another phosphitylation reagent, namely, TMDP, which has now become the most commonly used phosphitylating reagent for hydroxyl group analysis in

UNCORRECTED PROOF

lignin[82], tannins[22,81,85], synthetic aromatic polymers[26], and edible oils[86–93]. This was found to be particularly effective in terms of providing considerably better signal resolution for the uncondensed and condensed phenolic moieties in lignin[82]. Furthermore, Jiang and Argyropoulos showed that the coupling of Mannich chemistry on the lignin followed by ^{31}P NMR offered an additional means to resolve even more the guaiacyl groups in softwood technical lignins from minor overlapping signals caused by condensed moieties[94]. Notably, the ^{31}P NMR technique has been further validated by an examination of a series of lignins that was the subject of an international analytical round-robin effort using independent methods of analysis[82,95].

Historically, benzoic acid[96], bisphenol[26], cholesterol[97], and cyclohexanol[82] have all been used as ISs in ^{31}P NMR analysis of lignin. However, these ISs overlapped with the resonance of lignin moieties, causing an incomplete baseline resolution and subsequently leading to underestimated values for lignin hydroxyl groups[84]. In the early 2000s, Ragauskas's group introduced NHND as a promising alternative IS in ^{31}P NMR analysis of lignin owing to its ability to be fully baseline resolved from lignin-derived resonances[98]. His group has applied ^{31}P NMR using this IS to diverse areas including the characterization of transgetic and pretreated biomass[99–101]. In addition, Argyropoulos's group carried out extensive studies to elucidate the spin-lattice relaxation times and solvent effects on the ^{31}P NMR chemical shifts and arrived at the now universally used relaxation additive (chromium acetylacetonate) and the mixture of pyridine and $CDCl_3$ at the ratio of 1.6:1 as the solvent[15,20]. The above developments have allowed Crestini's group to show that ^{31}P NMR allows the straightforward assignment of the structural details of both hydrolysable and condensed tannins[22,81,85]. Recent advances in NMR technology, including the development of cryogenic probes and pulse-shaping software, have made it possible to acquire a spectrum from the whole plant cell wall without any laborious isolation of individual biomass components[59,69]. Furthermore, Argyropoulos's group has developed various imidazole-type ionic liquids to dissolve and characterize cellulose and the whole plant cell wall at elevated temperatures (~80–100 °C) via quantitative ^{31}P NMR analysis[102–105]. In a recent study, Ben's group reported that a pyridine-based ionic liquid, 1-allyl-3-butylpyridinium chloride, was capable of dissolving the whole biomass at low temperature (50 °C) and could be used in conjunction with the ^{31}P NMR methodology, thus providing a promising approach to quantitatively assess the hydroxyl groups in biomass as their original structures[106]. The broad applicability of this technique within the biorefinery research community is demonstrated in the following section.

Applications of the ^{31}P NMR protocol

Nowadays, the quantitative ^{31}P NMR technique is widely used in characterizing hydroxyl groups of biorefinery resources. It has been used to determine the functional groups of different industrial technical lignins including herbaceous lignins (soda and organosolv wheat straw), hardwood lignin (Alcell and organosolv poplar), and softwood lignin (Indulin Kraft, organosolv spruce)[73,107]. Biomass pretreatment, an important step to overcome biomass recalcitrance, often requires elevated temperatures under acidic or alkaline conditions. A large variety of depolymerization or condensation reactions typically occur within the lignin structure under such conditions. Quantitative ^{31}P NMR analyses of lignins isolated from various biomass pretreatment technologies have been highlighted in recent studies[101,108–117]. As lignin has become a key genetic engineering target for the enhancement of wood quality and biofuel production[118,119], quantitative ^{31}P NMR analyses also have been carried out on transgenic lignin[97,120]. Use of the ^{31}P NMR methodology along with DFRC offers unique information about etherified and carbon–carbon linked bonding patterns in lignin[121]. Fu et al.[122] initially adopted quantitative ^{31}P NMR to understand the nature of pyrolysis oils. ^{31}P NMR was used by other researchers to track the water content of pyrolysis oil, as well as the chemical transitions of bio-oil that take place during pyrolysis of biomass and the ensuing catalytic upgrading processes[123–128]. Following these efforts, a laboratory analytical procedure for the quantification of hydroxyls in bio-oil has been developed by National Renewable Energy Laboratory[129]. Functional groups of LCCs in softwood and hardwood might be also analyzed by quantitative ^{31}P NMR analysis[130,131]. Biomass impurities, tricin, and tricin-like flavonoid derivatives were also successfully identified by this ^{31}P NMR technique in recent studies[132,133]. ^{31}P NMR was also used to analyze the structural details of both hydrolysable and condensed tannins[22,81,85]. More specifically, the ^{31}P NMR technique provides a deep analysis of the substitution patterns in rings A, B, and C of the proanthocyanidins (Fig. 2), the acquisition of a specific fingerprint, and the determination of the purity of samples in complex matrices[22]. The principles and applications of this highly useful methodology involving quantitative ^{31}P NMR in biomass, lignin, and biofuel precursor characterization have been summarized in two

UNCORRECTED PROOF

critical reviews[21,134]. Furthermore, in a series of food-chemistry-related studies, Dais's group successfully applied this ^{31}P NMR technique on olive oils to quantify several minor compounds including phenolic compound, sterols, glycerol, and free acidity. Their studies have concluded that the application of this technique could be seamlessly extended to other edible oils, foods, and beverages for the purposes of quality control and authentication[86–93].

Advantages and limitations of the ^{31}P NMR protocol

To date, many approaches have been developed to determine the hydroxyl groups in lignin and tannins. These methodologies have been compared in several publications[135–138]. It was reported that the values of hydroxyl groups in lignins determined by ^{31}P NMR are in good agreement with the values obtained by other independent techniques such as FTIR/pyrolysis, the conductometric titration method, and ^{1}H and ^{13}C NMR spectroscopy[96,139]. Compared with traditional wet-chemical techniques, ^{31}P NMR offers unique advantages in determining hydroxyl groups in a single spectrum with great signal resolution. It allows the discrimination of the phenolic hydroxyl groups attached to syringyl, guaiacyl, and *p*-hydroxyphenyl units instead of simply offering the total aromatic hydroxyl groups. The method requires only small amounts of sample (~30 mg) and can be carried out within a relatively short experimental time (~30–120 min).

Nevertheless, the technique has its limitations. One of these is that the phosphorus reagent TMDP is pricey (€107/g in Europe, or $133/g in the US) and currently not widely commercially available among the common chemical suppliers. Its availability is not always guaranteed, which makes the use of ^{31}P NMR for the 'screening' of process parameters in biorefineries quite limited. However, Argyropoulos's group has arrived at a detailed protocol for its synthesis, which can be supplied upon request. As an NMR analysis, it always requires expensive hardware and specialized technical support. ^{31}P NMR in particular also requires high-purity samples (free of moisture, ash, sulfur, and carbohydrates) and sufficient solubility of samples in appropriate solvents. Signal overlap in phenolic regions caused by similar aromatic ring substitution patterns is another issue. Specifically, hydroxyl groups in lignin syringyl units are typically difficult to separate from the condensed guaiacyl units, which can cause biased integration by simply splitting unresolved resonance signals[84]. Derivatized samples and some ISs are not stable after a long period of storage, which requires an almost instant ^{31}P NMR acquisition. Finally, the existence of amine groups can also interfere with the quantification of the hydroxyl groups in some cases, such as preparations including large amounts of enzyme contamination.

Experimental design

Substrate solubilization

The complete solubilization of the substrate in an appropriate solvent is an essential prerequisite for an accurate quantitative liquid-state NMR measurement. A mixture of deuterated chloroform and pyridine is the suggested solvent, based on the early identified principles for ^{31}P NMR analysis of lignin, bio-oil, tannins, and other types of biorefinery resources[15]. The solvent used certainly can be adjusted on the basis of the properties or solubility of the substrates being analyzed. For lignin samples with limited solubility in chloroform and pyridine, *N*,*N*-dimethylformamide (DMF) could be introduced as a third solvent to help facilitate solubilization[134,140,141]. In a recent study, Stücker et al.[142] reported that a novel solvent system containing DMF, deuterated DMF, and pyridine (4.5:1:1 (vol/vol/vol)) is capable of fully dissolving sulfonated lignins. Pre-swelling of lignosulfonates in DMSO also proved to be an effective way to overcome the solubility obstacle of lignosulfonates in the conventional deuterated chloroform and pyridine solvent mixture[140]. The same study also showed that although sonication can improve the solubility of lignin, it can also cause a dramatic increase of the aliphatic and phenolic OH groups, which is suggestive of depolymerization of lignin. Thus sonication is not recommended. When new solvents are introduced to help facilitate such analyses, one should always ensure their inert nature toward the phosphitylating reagents and the ^{31}P NMR chemical shift dependency principles developed in the literature[15]. In this respect, completely new solvents or solvent mixtures may necessitate extensive verification of chemical shift integration ranges, as those originally proposed by the McGill team might not be valid in the new solvent system[14–17,26].

Internal standards

NHND has proved to be one of the most effective ISs in ^{31}P NMR analysis of lignin; however, it is not stable after a long period of storage, and another type of IS such as cholesterol or triphenylphosphine oxide (TPPO) should be used instead if a long-term experiment or extended sample storage is

UNCORRECTED PROOF

required[143]. For example, cholesterol has proved to be an excellent IS in ^{31}P NMR analysis of tannins owing to its long-time stability. The chemical shift of phosphitylated cholesterol (~144.8 p.p.m.) also does not overlap with other tannin functional groups[81]. In the case of pyrolysis oils, TPPO was considered as the best IS[129,143]. The derivatization reagent used in ^{31}P NMR analysis can react with water almost instantly, and thus a dry sample is essential. In addition, it is critical to ensure dry sample preparation conditions. All solvents to be used in the analysis should be of high purity and anhydrous. It is thus highly recommended that a fresh solvent mixture be prepared for each analysis set and that activated molecular sieves be used to minimize the moisture content.

NMR acquisition

In terms of the NMR acquisition, the inverse gated decoupling pulse needs to be routinely used to minimize the nuclear Overhauser effect enhancement for quantitative measurements[21,82]. Chromium (III) acetylacetonate is usually introduced into the solvent system as a relaxation agent to shorten the spin-lattice relaxation time of the phosphorus nuclei[82,97]. A relatively long interpulse delay (≥10 s) is also required in order to ensure complete spin relaxation before the application of a subsequent radiofrequency pulse[134,140]. The spectra quality should be judged on the basis of characteristics such as flat baseline, sharp and phased peaks, and the absence of spectral artifacts. The chemical shift is usually internally calibrated relative to the sharp peak at around 132.2 p.p.m. (Fig. 4) arising from the phosphitylation product of TMDP with water and ~121.1 p.p.m. for the case of CDP[82,95].

Materials

▲ **CRITICAL** None of the equipment and reagents described below have to be obtained from specific suppliers. Alternative similar equipment and identical chemicals from different suppliers may certainly be used; however, it is highly recommended to use the same product specifications listed below to ensure accuracy. The use of appropriate personal protective equipment to minimize exposure to hazardous materials is also essential.

Biological materials

- Lignin ▲ **CRITICAL** Lignin samples that have been analyzed successfully using this protocol include organosolv lignin derived from aqueous ethanol pretreated (180 °C, 1.25% H_2SO_4, 60% ethanol, 60 min) hardwood poplar, herbaceous switchgrass, and softwood pine. In all cases, the organosolv lignin, which is sulfur free, is rich in OH functionality, is low ash, and has limited carbohydrate contamination, is precipitated and recovered from the concentrated pretreatment liquor according to procedures described in the literature[23].
- Tannin ▲ **CRITICAL** For this work, Quebracho Colorado (*Schinopsis balansae*) was collected in the Chaco region between Argentina and Paraguay, and black wattle (*Acacia mearnsii*) was collected in Tanzania. The plant material was collected, identified, and certified by SilvaTeam. The extracts studied are commercial samples obtained from *S. balansae* wood and *A. mearnsii* bark kindly supplied by SilvaTeam. Epigallocatechin gallate was supplied by Sigma-Aldrich.

Reagents

- Acetone (Sigma-Aldrich, cat. no. 650501) ! **CAUTION** Acetone is an extremely flammable liquid and an eye irritant. Inhalation and contact with skin should be prevented.
- Chloroform-d (Sigma-Aldrich, cat. no. 151823) ! **CAUTION** Chloroform-d is a skin/eye irritant, carcinogenic, and toxic if inhaled. Inhalation and contact with skin should be prevented.
- Cholesterol (Sigma-Aldrich, cat. no. C8667)
- Chromium(III) 2,4-pentanedionate (Alfa Aesar, cat. no. 12538)
- Drierite (Sigma-Aldrich, cat. no. 238988)
- Pyridine anhydrous (Sigma-Aldrich, cat. no. 270970) ! **CAUTION** Pyridine anhydrous is an extremely flammable liquid and a skin/eye irritant. Inhalation and contact with skin should be prevented.
- 2-chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane (TMDP; Sigma-Aldrich, cat. no. 447536) ! **CAUTION** TMDP can cause severe skin burns and eye damage. Inhalation and contact with skin should be prevented.
- Molecular sieves (type 3A; Sigma-Aldrich, cat. no. MX1583C)
- *N*-hydroxy-5-norbornene-2,3-dicarboxylic acid imide (NHND; 97%; Sigma-Aldrich, cat. no. 226378)
- *N*,*N*-dimethylformamide (DMF; Sigma-Aldrich, cat. no. 227056) ! **CAUTION** DMF is a flammable liquid and a skin/eye irritant. Inhalation and contact with skin should be prevented.
- Calcium sulfate

UNCORRECTED PROOF

Equipment

- Analytical balance (Sartorius, model no. BP 210 S, or equivalent)
- Drierite gas-drying unit (Sigma-Aldrich, cat. no. Z112879, or equivalent)
- Hamilton glass gastight syringes (Hamilton, cat. nos. 81243 and 81000, or equivalent)
- Glass vials (Sigma-Aldrich, cat. no. 27345, or equivalent)
- Glass Pasteur pipette (Sigma-Aldrich, cat. no. Z628018, or equivalent)
- Glass desiccator (Sigma-Aldrich, cat. no. SLW1591/02D, or equivalent)
- NMR tubes (Sigma-Aldrich, cat. no. Z272019, or equivalent)
- Stirring plate (Sigma-Aldrich, cat. no. CLS6795420D, or equivalent)
- Stir bars (Sigma-Aldrich, cat. no. Z126942, or equivalent)
- NMR spectrometer (e.g., Bruker Avance III HD 500-MHz with 5-mm BBO probe, capable of ^{31}P detection, or equivalent)
- Vacuum oven (VWR, model no. 1400E, or equivalent)

Software

- NMR acquisition and processing software (Bruker Topspin 3.5pl7, or equivalent)
- MestReNova and VnmrJ

Reagent setup

Solvent A

Prepare 10.0 mL of a solvent mixture (solvent A) composed of deuterated chloroform and anhydrous pyridine at a volume ratio of 1:1.6 (vol/vol). Solvent A can be stored at room temperature (20–25 °C) for up to 4 weeks over molecular sieves in a sealed container that has a hole cap with a PTFE-lined silicone septum. Wrap the cap of the container with moisture-resistant Parafilm. ▲ **CRITICAL** Anhydrous pyridine is normally stored in a crown-cap bottle that has a hole in the cap and a PEFE-faced rubber liner under the crown-cap. It needs to be dispensed from the reagent bottle under inert atmosphere (e.g., N_2). Insert a needle connected to a Schlenk line or regulated low-pressure N_2 source equipped with a laboratory Drierite gas-drying unit into the septum to fill the space above the liquid with the inert gas inside the bottle. Use another glass gastight syringe as an outlet to withdraw the liquid from the container.

IS solution

Prepare a solution of chromium(III) 2,4-pentanedionate ($Cr(acac)_3$) by using solvent mixture A at a concentration of ~5.0 mg/mL, sealed from the atmosphere. Add NHND to the $Cr(acac)_3$ solution at a concentration of ~18.0 mg/mL (~0.1 M). Record the actual weight of NHND. This solution will be referred to as the IS solution. Record the actual weight of the entire IS solution (containing both $Cr(acac)_3$ and NHND). Store the IS solution over molecular sieves in a sealed container equipped with a PTFE-lined silicone septum, and wrap the cap of the container with moisture-resistant Parafilm. ▲ **CRITICAL** For ^{31}P NMR analysis of tannins or other types of substrate that need a long-term experiment or extended sample storage, it is recommended to use cholesterol as the IS. In that case, add cholesterol to the $Cr(acac)_3$ solution at a concentration of ~38.67 mg/mL (~0.1 M). Record the actual weight of cholesterol and the entire IS solution.

Procedure

Sample setup ● Timing ~24 h

1 Place the lignin or the tannin sample into a vacuum oven at ~45 °C and allow it to dry until a constant weight is attained (~24 h).

2 Cool the samples to 25 °C in a glass desiccator over anhydrous calcium sulfate.

NMR solution setup ● Timing ~30 min–12 h

3 Transfer ~0.1 mL of the IS solution (see 'Reagent setup' section) into a 4-mL glass vial equipped with a PTFE-lined silicone septum. Record the actual weight of the 0.1 mL of IS solution.

4 Add ~30 mg of pre-dried lignin or tannin sample from Step 2 into the same vial. Record the actual weight of the samples to the nearest 0.1 mg.

5 Use a glass gastight syringe to add ~0.5 mL of solvent A (see 'Reagent setup' section) into the same vial with constant stirring, using a magnetic stirrer. Note: stir the solution overnight (~12 h) to fully dissolve the lignin or tannin samples if necessary (depending on the nature of the sample).

UNCORRECTED PROOF

▲ CRITICAL STEP Treat solvent A as an anhydrous solvent and transfer the solvent under an inert atmosphere as described in the 'Reagent setup' section. Make sure the substrate is fully dissolved in the solvent mixture, as this is the key step for an accurate quantitative analysis.

? TROUBLESHOOTING

6 Add ~0.1 mL of TMDP to the homogeneous lignin or tannin solution by using a glass gastight syringe, and wash the syringe with acetone immediately.

▲ CRITICAL STEP Excess amounts of TMDP need to be added so that all the hydroxyl group can be phosphitylated. ~0.1 mL of TMDP is more than enough for typical lignin or tannin analyses. However, for samples such as pyrolysis oil that may contain a significant amount of moisture and hydroxyl groups, up to 0.2 mL of TMDP should be added.

▲ CRITICAL STEP TMDP should not be added until a homogeneous solution is formed.

▲ CRITICAL STEP TMDP reacts with water almost instantly, producing a yellow precipitate in the reagent bottle. Thus TMDP should be transferred into a vial equipped with a PTFE-lined silicone septum so that it can be withdrawn from the container without the need to remove the cap.

■ PAUSE POINT The vial and original TMDP container should be wrapped with Parafilm and can be kept in a freezer at −20 °C for several days.

7 Shake the mixture for ~30 s to several minutes and confirm that no precipitates are formed.

▲ CRITICAL STEP As long as a homogeneous solution is formed, the solution should be analyzed by NMR immediately. The phosphitylation derivative of NHND is not stable over a long period of time, and long-term storage will therefore lead to inaccurate measurements. If cholesterol or triphenylphosphine oxide is used as the IS (see 'Experimental design' section), the sample can be stored for up to several hours.

? TROUBLESHOOTING

NMR measurement ● Timing ~30 min to 2 h

8 Transfer all the phosphitylated lignin or tannin solution from Step 7 into a 5-mm NMR tube via a glass Pasteur pipette.

9 Load the sample tube into the NMR spectrometer with an appropriate probe capable of ^{31}P detection.

10 Set up the NMR parameters using the following recommended conditions:

Spectrometer	Bruker Avance III HD 500 MHz
Pulse program	Inverse gated decoupling pulse (zgig)
Nucleus	^{31}P
Spectral width (SW)	100 p.p.m.
Acquisition time (AQ)	~0.8 s
Relaxation delay (D1)	≥10 s
Number of scans (NS)	64 or more
Center of spectrum (O1P)	140 p.p.m.

▲ CRITICAL STEP More scans (128 or 256) could lead to an increased signal-to-noise ratio (S/N) at the cost of a longer experiment time. For 300-MHz spectrometers, a minimum of 128 scans is needed.

▲ CRITICAL STEP Bruker Avance III HD 500 MHz and Topspin 3.5 were used for NMR data acquisition and processing in this study. Equivalent experiments could also be performed on other NMR systems such as Agilent/Varian and JEOL, and the nomenclature for setting and parameters could be different. The obtained data could be processed via different software (e.g., MestReNova and VnmrJ) as well.

11 Lock the spectrometer frequency to the deuterium resonance arising from the NMR solvent ($CDCl_3$).

12 Shim the sample and tune the NMR probe using either manual or automated methods following the instructions provided by the manufacturer.

13 Acquire the ^{31}P spectra by applying the standard Bruker inverse gated decoupling pulse sequence.

Data processing ● Timing ~20 min

14 Apply NMR processing commands to process the FID and calculate the spectrum (Bruker command: efp) and to perform automatic phase correction (Bruker command: apk).

UNCORRECTED PROOF

Table 1 | Typical integration regions for lignins in a ^{31}P NMR spectrum[134]

Lignin functional group	Chemical shift (p.p.m.)
Aliphatic OH	~145.4–150.0
Phenolic OH	~137.6–144.0
C_5 substituted	~140.0–144.5
5-5	~141.2
4-O-5	~142.3
Syringyl	~142.7
β-5	~143.5
Guaiacyl	~139.0–140.2
p-Hydroxyphenyl	~137.8
Carboxylic acid OH	~133.6–136.0

Table 2 | Typical integration regions for tannins in a ^{31}P NMR spectrum[22]

Tannin functional group	Chemical shift (p.p.m.)
Ring A	
o-Unsubstituted phenolic OH	137.9–137.4
o-Substituted	138.8–137.9
Ring B	
Catechol	140.2–138.8
Pyrogallol	144.0–140.2
Ring C	
Aliphatic OH	150.0–145.5

15 Make sure there is a sharp signal around 175 p.p.m. representing the extra TMDP peak. **? TROUBLESHOOTING**

16 Calibrate the chemical shift by assigning the sharp TMDP + H_2O peak at 132.2 p.p.m. (Fig. 4).

17 Adjust the zero and first order phasing of the spectra to make all the peaks as symmetrical as possible.

18 Integrate the NMR peak regions with respect to the IS (NHND) peak based on the assignments described in Table 1 (for lignins) and Table 2 (for tannins).

▲CRITICAL STEP Certain ^{31}P NMR spectra may show a distorted baseline that could prevent accurate integration of the peaks. A baseline correction should always be applied. Automatic baseline correction works in the majority of cases. The user can also carry out baseline correction manually by defining a polynomial function to subtract from the spectrum. Peak simulation or advanced line shape fitting is not necessary for the ^{31}P NMR protocol.

▲CRITICAL STEP The chemical shift assignments described in Tables 1 and 2 are based on ^{31}P NMR analysis of lignin or tannin model compounds using TMDP and pyridine/$CDCl_3$ as the phosphitylating reagent and solvent, respectively. A completely new solvent system and ^{31}P reagent may necessitate extensive verification of chemical shift integration ranges, as those originally proposed might not be valid in the new solvent system.

Calculation of the amount of different hydroxyl groups ● Timing ~30 min

19 Calculate the mole quantity of IS (NHND, with a purity of 97%) in IS solution:

$$\text{NHND in IS solution (mmol)} = \frac{\text{Mass of NHND (g)}}{179.17\ (\text{g/mol})} \times 97\% \times 1,000$$

UNCORRECTED

Box 1 | Determining the fine structural details of condensed tannins

1. Amount of proanthocyanidins = 1/2 × catechols.
2. Average phenols in ring B (b) = (B ring OH)/(1/2×catechols).
3. Ring B substitution pattern = (pyrogallol OH)/(1/2×catechols).
4. Average phenols in ring A (a) = (A ring OH)/(1/2×catechols) = (2 ring A = phloroglucinol; 1 ring A = resorcinol).
5. Average monomeric unit molecular weight (M_w) = $C_{15}H_{12}O_2$ + O ($a + b$), where a = average phenols on ring A and b = average phenols on ring B.
6. Sample purity (flavan-3-ol content) (%) = 1/2 × catechols × M_w × 0.1.

20 Calculate the mole quantity of IS (NHND) in the NMR sample:

$$\text{NHND in NMR sample (mmol)} = \frac{\text{NHND in IS solution (mmol)}}{\text{Total mass of IS solution (g)}} \times \text{Mass of 0.1 mL of IS solution (g)}$$

21 Calculate the ratio (R) of the integration of the spectral region of interest (I_{OH}) over the IS region (I_{NHND}):

$$R = I_{OH}/I_{NHND} = \frac{\text{Integration of spectral region of interest}}{\text{Integration of NHND region}}$$

22 Calculate the amount of different hydroxyl groups in lignin/tannin samples, that is, mmol OH/g lignin:

$$\text{mmol of different types of OH per g of lignin/tannin} = \frac{\text{R mmol of NHND in NMR sample(mmol)}}{\text{Dry weight of lignin/tannin sample(g)}}$$

23 Determine the fine structural details of condensed tannins using the scheme in Box 1.

Troubleshooting

Troubleshooting advice can be found in Table 3.

Table 3 | Troubleshooting table

Step	Problem	Possible reason	Solution
5	A heterogeneous mixture is formed	Poor solubility of the substrate in chloroform and pyridine solvent mixture	Add more solvent. If the substrate is still insoluble, a small amount of a third solvent such as DMF could be added to help dissolve the samples. Always keep in mind that the third solvent introduced into the system needs to be inert so as not to interfere with the sought functionality analysis and not significantly affect the chemical shifts[15,20]. Stir the solution overnight if necessary with a magnetic stir bar
7	Precipitates are formed	The substrate has too much moisture and/or the TMDP-solvent mixture has been contaminated by water	The sample needs to be dried again before being dissolved in freshly prepared pyridine–chloroform-d and derivatized by fresh TMDP
		There is not enough deuterated chloroform to dissolve the pyridine-HCl salt	Add an extra amount of solvent A (deuterated chloroform and pyridine) until the precipitate is no longer visible. The amount of solvent being added will not affect the ensuing quantitative analysis
15	The TMDP peak is missing in the ^{31}P NMR spectra	There is not enough TMDP in the NMR sample solution, and thus the phosphitylation reaction is incomplete	If the TMDP peak is missing, repeat Step 14 by adding extra TMDP and perform the NMR experiment again until a sharp TMDP peak shows up

CORRECTED PROOF

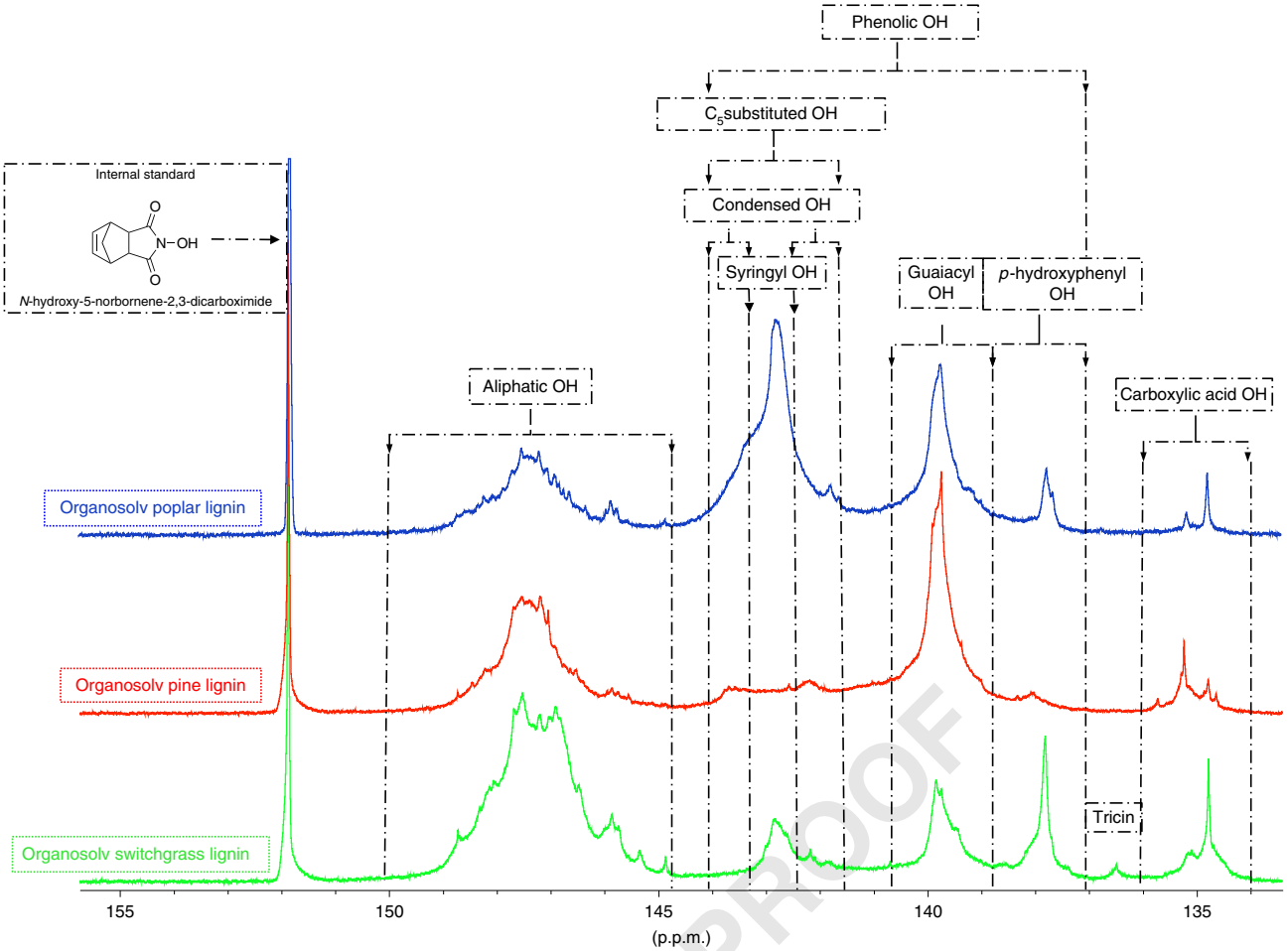

Fig. 5 | Quantitative ^{31}P NMR spectra of organosolv poplar, pine, and switchgrass lignin derivatized with TMDP using NHND as the IS.

UNCORRECTED PROOF

Timing

Steps 1 and 2, sample setup: ~24 h
Steps 3–7, NMR solution setup: ~30 min–12 h
Steps 8–13, NMR measurement: ~30 min–2 h
Steps 14–18, data processing: ~20 min
Steps 19–23, calculation of the amount of different hydroxyl groups: ~30 min

Anticipated results

Figure 5 shows three quantitative ^{31}P NMR spectra of phosphitylated organosolv poplar, pine, and switchgrass lignin derivatized with TMDP using NHND as the IS. A cryoprobe-equipped spectrometer was used in this study; room temperature probes are sufficient for the ^{31}P NMR protocol, but an increased number of scans (128 or 256) is needed to achieve the equivalent S/N. As shown in Fig. 5, the IS used in this study, NHND, has a chemical shift that is well separated from lignin-derived components. Different TMDP-derivatized OH groups including aliphatic, syringyl, guaiacyl, *p*-hydroxyphenyl, and carboxylic acid OH have substantially different chemical shifts, which allows them to be distinguished and quantified by ^{31}P NMR. It is recommended that the syringyl and other types of condensed 5-substituted units (e.g., 5-5 and β-5) be combined into C_5 substituted phenolic OH because of the signal overlap issue, in order to prevent possible overestimation of the syringyl OH and underestimation of condensed units[84,107]. This normally will not be a problem for softwood lignins, as the C_5-substituted OH is composed exclusively of condensed guaiacyl units as shown in Fig. 5. Taking into account the comments above, we show the calculated contents of five different

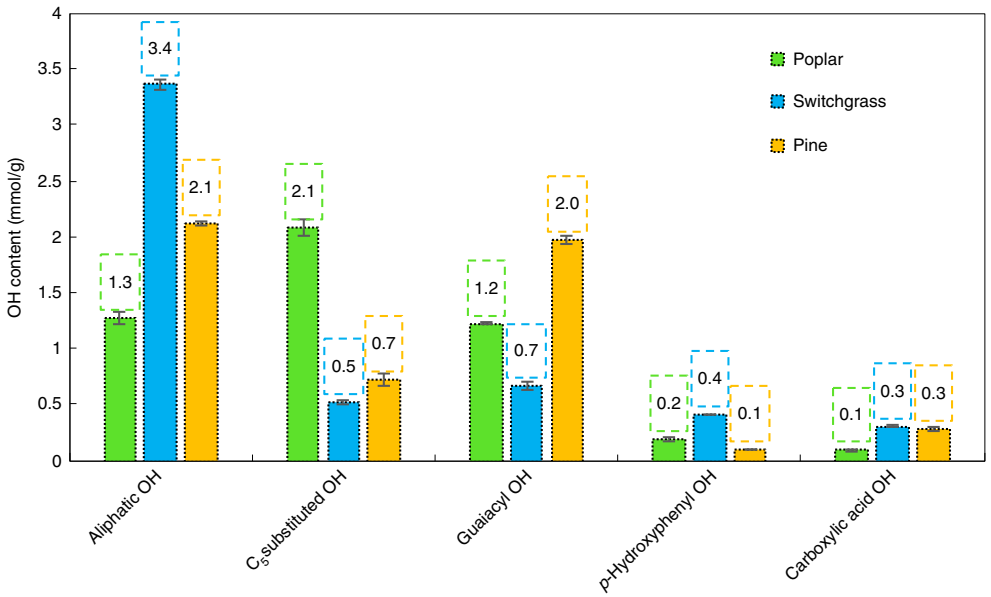

Fig. 6 | Contents of different hydroxyl groups in organosolv poplar, switchgrass, and pine lignin determined by ^{31}P NMR analysis. Error bars represent ±s.d. of three individual measurements. Q14

types of hydroxyl groups expressed in mmol of OH groups per g of organosolv lignin samples in Fig. 6. Measurements were done in triplicate on distinct samples, and the results show great reproducibility, with the coefficient of variation ranging from 1.1% to 9.8%. Quantitative data clearly show that softwood lignin and herbaceous lignin were the richest in guaiacyl and *p*-hydroxyphenyl OH groups, respectively, while hardwood lignin had the greatest amount of C_5-substituted OH. Tricin, a flavonoid metabolite, was also identified in switchgrass lignin.

The ^{31}P NMR spectrum of proanthocyanidins can be divided into three main regions: 138.1–137.5, 142.5–137.9, and 146.0–145.0 p.p.m. relative to ring A (*o*-unsubstituted phenolic OH-groups), ring B (catechols/pyrogallols), and ring C (aliphatic OH), respectively. The absorbance area of ring B is further divided into two specific regions because of the *o*-disubstituted phenolics in 3,4,5-trihydroxyphenyl units (142.5–141.8/141.5–141.0 p.p.m.) and the *o*-monosubstituted hydroxyphenyl groups (139.4–137.9 p.p.m.) characteristic of the prorobinetinidin/prodelphinidin and procyanidin/profisetinidin subunits, respectively. Integration of the catecholic area allows the quantitative determination of the amount of flavan-3-ol units; in fact, each subunit in condensed tannins contains two catecholic groups in ring B (Fig. 2). Figure 7 shows representative spectra of commercial samples Q15 of condensed tannins along with A, B, and C rings in chemical shift regions. When the B ring belongs to the procyanidin/profisetinidin family, only signals in the catecholic area are present (Fig. 7a). In the case of samples containing prorobinetinidin/prodelphinidin units, an additional signal is present in the pyrogallol area (Fig. 7b). The ratio between the pyrogallol and the catechol content gives the ring B substitution pattern ratio.

The specific nature of ring A is determined by the integration of the *o*-unsubstituted phenolic groups (138.1–137.5 p.p.m.). The ratio of the integrals of the regions related to B and A rings provides Q16 the average amount of phenolic units in the A ring per flavan-3-ol unit, and thus its specific resorcinol or phloroglucinol nature (Fig. 2). Technical samples can contain mixtures of different proanthocyanidins. In this case, the ratio of the different A rings' patterns is easily determined. The ^{31}P NMR analysis of tannins also provides a determination of the sample purity. More specifically, it can be calculated from the ratio between the amount of flavan-3-ol units as determined by the integration of the catechol region (divided by 2, as each flavan-3-ol unit contains two catecholic OH groups) and the average M_w of the flavan-3-ol subunits as determined from the substitution patterns of A and B rings according to the following equation [22]:

$$\text{Flavan}-3-\text{ol content}(\%) = (\text{Flavan}-3-\text{ol content}/\text{Theoretical flavan}-3-\text{ol content}) \times 100$$

where the flavan-3-ol content is determined by the mmol/g of catechols divided by 2, and the theoretical flavan-3-ol content is determined by 1/average flavan-3-ol M_w. When complex tannins are

UNCORRECTED PROOF

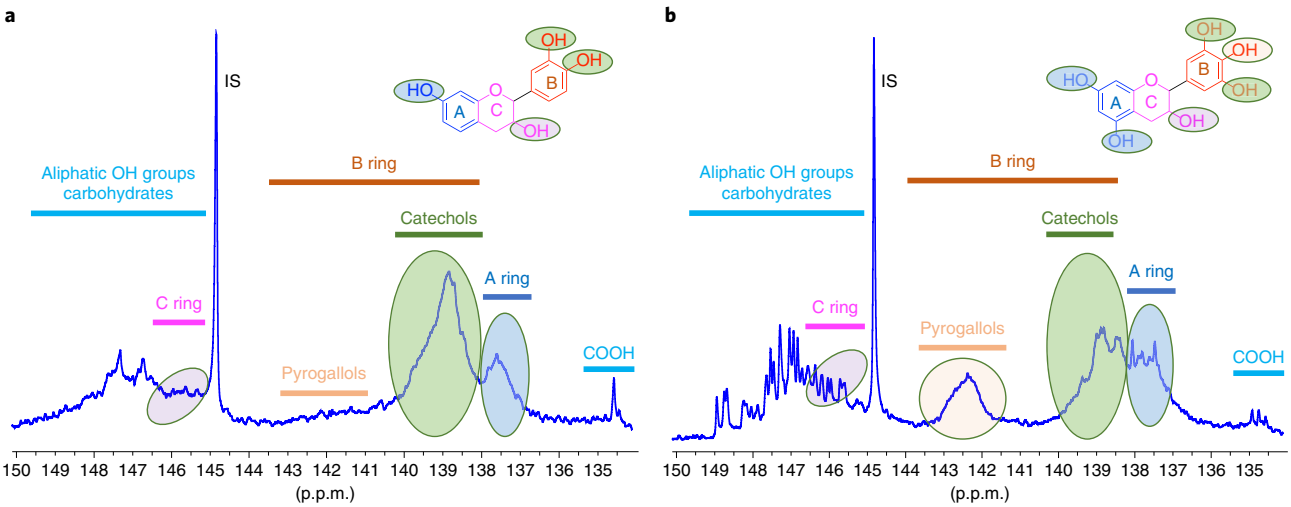

Fig. 7 | Representative ^{31}P NMR spectra of condensed tannins using cholesterol as the IS. a, *Schinopsis balansae* wood extract. **b**, *Acacia mearnsii* wood extract. Reproduced with permission from ref. 22, American Chemical Society.

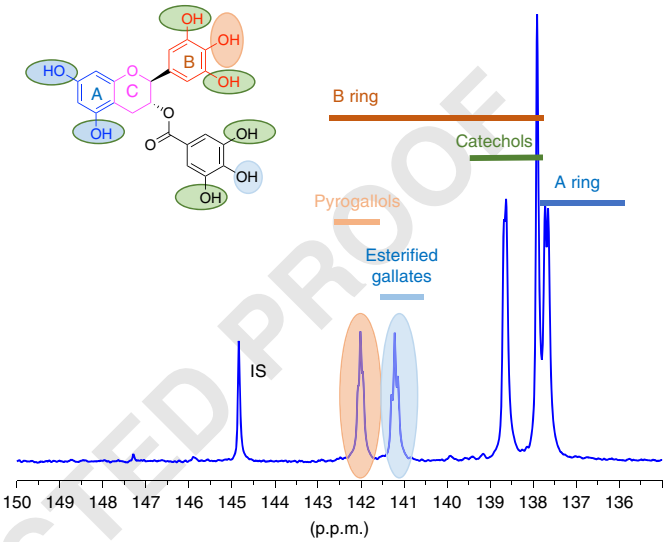

Fig. 8 | Representative ^{31}P NMR of a complex tannin (epigallocatechin-3-O-gallate) using cholesterol as the IS: signals of esterified gallates and B rings are excellently resolved. Adapted with permission from ref. 22, American Chemical Society.

UNCORRECTED PROOF

present where both pyrogallol groups and esterified gallates are present, it is possible to specifically integrate the different units owing to the excellent signal resolution, as shown in Fig. 8.

In conclusion, the quantitative ^{31}P NMR technique presented in this protocol offers a rapid but reliable analytical tool for analysis of hydroxyl groups, and we believe it will be continually used to explore the functionality of biomass-based green materials in future integrated biorefinery studies and plant sciences.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated during this study are included in this published article. The NMR integration data are available upon request. The software used for NMR data analysis is freely available (see 'Software').

References

1. Ragauskas, A. J. et al. The path forward for biofuels and biomaterials. *Science* **311**, 484–489 (2006).
2. Traoré, M., Kaal, J. & Martínez Cortizas, A. Application of FTIR spectroscopy to the characterization of archeological wood. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **153**, 63–70 (2016).
3. Wu, W. et al. Thermo-responsive and fluorescent cellulose nanocrystals grafted with polymer brushes. *J. Mater. Chem. A* **3**, 1995–2005 (2015).
4. DeMartini, J. D., Studer, M. H. & Wyman, C. E. Small-scale and automatable high-throughput compositional analysis of biomass. *Biotechnol. Bioeng.* **108**, 306–312 (2011).
5. Long, C. P. & Antoniewicz, M. R. Quantifying biomass composition by gas chromatography/mass spectrometry. *Anal. Chem.* **86**, 9423–9427 (2014).
6. Foston, M. Advances in solid-state NMR of cellulose. *Curr. Opin. Biotechnol.* **27**, 176–184 (2014).
7. Tolbert, A., Akinosho, H., Khunsupat, R., Naskar, A. K. & Ragauskas, A. J. Characterization and analysis of the molecular weight of lignin for biorefining studies. *Biofuels Bioprod. Biorefin.* **8**, 836–856 (2014).
8. Kim, J. E. & Lee, J.-W. Enzyme adsorption properties on dilute acid pretreated biomass by low vacuum-scanning electron microscopy and structural analysis of lignin. *Bioresour. Technol* **262**, 107–113 (2018).
9. Hinkle, J. D., Ciesielski, P. N., Gruchalla, K., Munch, K. R. & Donohoe, B. S. Biomass accessibility analysis using electron tomography. *Biotechnol. Biofuels* **8**, 212/211–212/216 (2015).
10. Ye, C., Malak, S. T., Hu, K., Wu, W. & Tsukruk, V. V. Cellulose nanocrystal microcapsules as tunable cages for nano- and microparticles. *ACS Nano* **9**, 10887–10895 (2015).
11. Chiu, L.-d., Ho, S.-H., Shimada, R., Ren, N.-Q. & Ozawa, T. Rapid in vivo lipid/carbohydrate quantification of single microalgal cell by Raman spectral imaging to reveal salinity-induced starch-to-lipid shift. *Biotechnol. Biofuels* **10**, 9/1–9/9 (2017).
12. Jung, S., Foston, M., Kalluri, U. C., Tuskan, G. A. & Ragauskas, A. J. 3D chemical image using TOF-SIMS revealing the biopolymer component spatial and lateral distributions in biomass. *Angew. Chem. Int. Ed.* **51**, 12005–12008 (2012).
13. Harton, S. E. et al. Evidence for complex molecular architectures for solvent-extracted lignins. *ACS Macro Lett.* **1**, 568–573 (2012).
14. Archipov, Y., Argyropoulos, D. S., Bolker, H. I. & Heitner, C. ^{31}P NMR spectroscopy in wood chemistry. I. Model compounds. *J. Wood Chem. Technol.* **11**, 137–157 (1991).
15. Argyropoulos, D. S., Bolker, H. I., Heitner, C. & Archipov, Y. ^{31}P NMR spectroscopy in wood chemistry. Part IV. Lignin models: spin lattice relaxation times and solvent effects in ^{31}P NMR. *Holzforschung* **47**, 50–56 (1993).
16. Argyropoulos, D. S., Bolker, H. I., Heitner, C. & Archipov, Y. ^{31}P NMR spectroscopy in wood chemistry. Part V. Qualitative analysis of lignin functional groups. *J. Wood Chem. Technol.* **13**, 187–212 (1993).
17. Archipov, Y., Argyropoulos, D. S., Bolker, H. & Heitner, C. ^{31}P-N.m.r. spectroscopy in wood chemistry. Part II. Phosphite derivatives of carbohydrates. *Carbohydr. Res.* **220**, 49–61 (1991).
18. Argyropoulos, D. S., Heitner, C. & Morin, F. G. ^{31}P NMR spectroscopy in wood chemistry. Part III. Solid state ^{31}P NMR of trimethyl phosphite derivatives of chromophores in mechanical pulp. *Holzforschung* **46**, 211–218 (1992).
19. Argyropoulos, D. S. & Heitner, C. ^{31}P NMR spectroscopy in wood chemistry. Part VI. Solid state ^{31}P NMR of trimethyl phosphite derivatives of chromophores and carboxylic acids present in mechanical pulps; a method for the quantitative determination of ortho-quinones. *Holzforschung* **48**, 112–116 (1994).
20. Mazur, M. & Argyropoulos, D. S. ^{31}P NMR spectroscopy in wood chemistry. VII. Studies toward elucidating the phosphorus relaxation mechanism of phosphitylated lignins. *Cellul. Chem. Technol* **29**, 589–601 (1995).
21. Argyropoulos, D. S. ^{31}P NMR in wood chemistry: a review of recent progress. *Res. Chem. Intermed.* **21**, 373–395 (1995).
22. Crestini, C., Lange, H. & Bianchetti, G. Detailed chemical composition of condensed tannins via quantitative ^{31}P NMR and HSQC analyses: *Acacia catechu*, *Schinopsis balansae*, and *Acacia mearnsii*. *J. Nat. Prod.* **79**, 2287–2295 (2016).
23. Pan, X. et al. Bioconversion of hybrid poplar to ethanol and co-products using an organosolv fractionation process: optimization of process yields. *Biotechnol. Bioeng.* **94**, 851–861 (2006).
24. Debuissy, T., Pollet, E. & Averous, L. Synthesis of potentially biobased copolyesters based on adipic acid and butanediols: kinetic study between 1,4- and 2,3-butanediol and their influence on crystallization and thermal properties. *Polymer* **99**, 204–213 (2016).
25. Debuissy, T., Pollet, E. & Averous, L. Synthesis and characterization of biobased poly(butylene succinate-ran-butylene adipate). Analysis of the composition-dependent physicochemical properties. *Eur. Polym. J* **87**, 84–98 (2017).
26. Chan, K. P., Argyropoulos, D. S., White, D. M., Yeager, G. W. & Hay, A. S. Facile quantitative analysis of hydroxyl end groups of poly(2,6-dimethyl-1,4-phenylene oxide)s by ^{31}P NMR spectroscopy. *Macromolecules* **27**, 6371–6375 (1994).
27. DeSilva, M. A. et al. Application of ^{31}P NMR spectroscopy and chemical derivatization for metabolite profiling of lipophilic compounds in human serum. *Magn. Reson. Chem.* **47**, S74–S80 (2009).
28. Ragauskas, A. J. et al. Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843 (2014).

UNCORRECTED PROOF

29. Li, M., Pu, Y., Yoo, C. G. & Ragauskas, A. J. The occurrence of tricin and its derivatives in plants. *Green Chem.* **18**, 1439–1454 (2016).
30. del Rio, J. C., Rencoret, J., Gutierrez, A., Kim, H. & Ralph, J. Hydroxystilbenes are monomers in palm fruit endocarp lignins. *Plant Physiol.* **174**, 2072–2082 (2017).
31. Zhao, Q. et al. Loss of function of cinnamyl alcohol dehydrogenase 1 leads to unconventional lignin and a temperature-sensitive growth defect in *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* **110**, 13660–13665 (2013).
32. Ralph, J. An unusual lignin from kenaf. *J. Nat. Prod.* **59**, 341–342 (1996).
33. Ralph, J., Kim, H., Peng, J. & Lu, F. Arylpropane-1,3-diols in lignins from normal and CAD-deficient pines. *Org. Lett.* **1**, 323–326 (1999).
34. Chen, F., Tobimatsu, Y., Havkin-Frenkel, D., Dixon, R. A. & Ralph, J. A polymer of caffeyl alcohol in plant seeds. *Proc. Natl. Acad. Sci. USA* **109**, 1772–1777 (2012).
35. Lan, W. et al. Tricin, a flavonoid monomer in monocot lignification. *Plant Physiol.* **167**, 1284–1295 (2015).
36. Vanholme, R., Demedts, B., Morreel, K., Ralph, J. & Boerjan, W. Lignin biosynthesis and structure. *Plant Physiol.* **153**, 895–905 (2010).
37. Stewart, J. J., Akiyama, T., Chapple, C., Ralph, J. & Mansfield, S. D. The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiol.* **150**, 621–635 (2009).
38. Harman-Ware, A. E. et al. A thioacidolysis method tailored for higher-throughput quantitative analysis of lignin monomers. *Biotechnol. J.* **11**, 1268–1273 (2016).
39. Faix, O., Andersons, B. & Zakis, G. Determination of carbonyl groups of six round robin lignins by modified oximation and FTIR spectroscopy. *Holzforschung* **52**, 268–274 (1998).
40. Santos, R. B., Capanema, E. A., Balakshin, M. Y., Chang, H.-m. & Jameel, H. Lignin structural variation in hardwood species. *J. Agric. Food Chem.* **60**, 4923–4930 (2012).
41. Tarabanko, V. E. & Tarabanko, N. Catalytic oxidation of lignins into the aromatic aldehydes: general process trends and development prospects. *Int. J. Mol. Sci.* **18**, 2421/2421–2421/2429 (2017).
42. Bose, S. K., Wilson, K. L., Hausch, D. L. & Francis, R. C. Lignin analysis by permanganate oxidation. Part 2. Lignins in acidic organosolv pulps. *Holzforschung* **53**, 603–610 (1999).
43. Guerra, A., Mendonça, R., Ferraz, A., Lu, F. & Ralph, J. Structural characterization of lignin during *Pinus taeda* wood treatment with *Ceriporiopsis subvermispora*. *Appl. Environ. Microbiol.* **70**, 4073–4078 (2004).
44. Lupoi, J. S., Singh, S., Parthasarathi, R., Simmons, B. A. & Henry, R. J. Recent innovations in analytical methods for the qualitative and quantitative assessment of lignin. *Renew. Sustain Energy Rev.* **49**, 871–906 (2015).
45. Tian, L., Pang, Y. & Dixon, R. A. Biosynthesis and genetic engineering of proanthocyanidins and (iso) flavonoids. *Phytochem. Rev.* **7**, 445–465 (2008).
46. Xie, D.-Y. & Dixon, R. A. Proanthocyanidin biosynthesis—still more questions than answers? *Phytochemistry* **66**, 2127–2144 (2005).
47. Vek, V., Oven, P., Piljansek, I. & Ters, T. Contribution to understanding the occurrence of extractives in red heart of beech. *BioResources* **10**, 970–985 (2015).
48. Weißmann, G. & Ayla, C. Die Verwendung von natürlichen Polyphenolen zur Herstellung von Holzleimen. *Holz als Roh- und Werkstoff* **38**, 245–249 (1980).
49. Haslam, E. *Plant Polyphenols: Vegetable Tannins Revisited* (Cambridge University Press, 1989).
50. Ferreira, D., Steynberg, J. P., Roux, D. G. & Brandt, E. V. Diversity of structure and function in oligomeric flavanoids. *Tetrahedron* **48**, 1743–1803 (1992).
51. Roux, D. G., Ferreira, D., Hundt, H. K. L. & Malan, E. Structure, stereochemistry, and reactivity of natural condensed tannins as basis for their extended industrial application. *Appl. Polym. Symp.* **28**, 335–353 (1975).
52. Roux, D. G., Ferreira, D., Botha, J. J. & Garbutt, D. C. F. Heartwood extracts of the black wattle (*Acacia mearnsii*) as a possible source of resorcinol. *Appl. Polym. Symp.* **28**, 1365–1376 (1976).
53. Price, M. L. & Butler, L. G. Rapid visual estimation and spectrophotometric determination of tannin content of sorghum grain. *J. Agric. Food Chem.* **25**, 1268–1273 (1977).
54. Krishnaiah, D., Bono, A., Sarbatly, R. & Anisuzzaman, S. M. Antioxidant activity and total phenolic content of an isolated *Morinda citrifolia* L. methanolic extract from poly-ethersulphone (PES) membrane separator. *J. King Saud. Univ. Eng. Sci.* **27**, 63–67 (2015).
55. Hagerman, A. E. & Butler, L. G. Protein precipitation method for the quantitative determination of tannins. *J. Agric. Food Chem.* **26**, 809–812 (1978).
56. van Erven, G. et al. Quantification of lignin and its structural features in plant biomass using ^{13}C lignin as internal standard for pyrolysis-GC-SIM-MS. *Anal. Chem.* **89**, 10907–10916 (2017).
57. Zhou, G., Taylor, G. & Polle, A. FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods* **7**, 9 (2011).
58. Li, X., Sun, C., Zhou, B. & He, Y. Determination of hemicellulose, cellulose and lignin in moso bamboo by near infrared spectroscopy. *Sci. Rep.* **5**, 17210 (2015).
59. Mansfield, S. D., Kim, H., Lu, F. & Ralph, J. Whole plant cell wall characterization using solution-state 2D NMR. *Nat. Protoc.* **7**, 1579–1589 (2012).
60. Li, S. & Lundquist, K. A new method for the analysis of phenolic groups in lignins by ^{1}H-NMR spectrometry. *Nord. Pulp Pap. Res. J.* **9**, 191–195 (1994).
61. Hallac, B. B., Pu, Y. & Ragauskas, A. J. Chemical transformations of *Buddleja davidii* lignin during ethanol organosolv pretreatment. *Energy Fuels* **24**, 2723–2732 (2010).

UNCORRECTED PROOF

62. Sun, S., Huang, Y., Sun, R. & Tu, M. The strong association of condensed phenolic moieties in isolated lignins with their inhibition of enzymatic hydrolysis. *Green Chem.* **18**, 4276–4286 (2016).
63. Du, X., Gellerstedt, G. & Li, J. Universal fractionation of lignin-carbohydrate complexes (LCCs) from lignocellulosic biomass: an example using spruce wood. *Plant J.* **74**, 328–338 (2013).
64. Low, J. H., Rahman, W. A. W. A. & Jamaluddin, J. Structural elucidation of tannins of spent coffee grounds by CP-MAS ^{13}C NMR and MALDI-TOF MS. *Industr. Crops Products* **69**, 456–461 (2015).
65. Pizzi, A. & Stephanou, A. A comparative C^{13} NMR study of polyflavonoid tannin extracts for phenolic polycondensates. *J. Appl. Polym. Sci.* **50**, 2105–2113 (1993).
66. Zeller, W. E. et al. ^{1}H-^{13}C HSQC NMR spectroscopy for estimating procyanidin/prodelphinidin and cis/trans-flavan-3-ol ratios of condensed tannin samples: correlation with thiolysis. *J. Agric. Food Chem.* **63**, 1967–1973 (2015).
67. Kim, H., Ralph, J. & Akiyama, T. Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d6. *BioEnergy Res.* **1**, 56–66 (2008).
68. Kim, H. & Ralph, J. Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d6/pyridine-d5. *Organ. Biomol. Chem.* **8**, 576–591 (2010).
69. Yoo, C. G., Pu, Y., Li, M. & Ragauskas, A. J. Elucidating structural characteristics of biomass using solution-state 2D NMR with a mixture of deuterated dimethylsulfoxide and hexamethylphosphoramide. *ChemSusChem* **9**, 1090–1095 (2016).
70. Crestini, C., Melone, F., Sette, M. & Saladino, R. Milled wood lignin: a linear oligomer. *Biomacromolecules* **12**, 3928–3935 (2011).
71. Sette, M., Wechselberger, R. & Crestini, C. Elucidation of lignin structure by quantitative 2D NMR. *Chemistry* **17**, 9529–9535 (2011).
72. Sette, M., Lange, H. & Crestini, C. Quantitative HSQC analyses of lignin: a practical comparison. *Comput. Struct. Biotechnol. J.* **6**, e201303016 (2013).
73. Crestini, C., Lange, H., Sette, M. & Argyropoulos, D. S. On the structure of softwood kraft lignin. *Green Chem.* **19**, 4104–4121 (2017).
74. Okaru, A. O. et al. Application of ^{19}F NMR spectroscopy for content determination of fluorinated pharmaceuticals. *J. Anal. Methods Chem.* **2017**, 7 (2017).
75. Wroblewski, A. E., Lensink, C., Markuszewski, R. & Verkade, J. G. Phosphorus-31 NMR spectroscopic analysis of coal pyrolysis condensates and extracts for heteroatom functionalities possessing labile hydrogen. *Energy Fuels* **2**, 765–774 (1988).
76. Gierer, J. Chemistry of delignification. Part 1: general concept and reactions during pulping. *Wood Sci. Technol.* **19**, 289–312 (1985).
77. Sadeghifar, H. & Argyropoulos, D. S. Correlations of the antioxidant properties of softwood kraft lignin fractions with the thermal stability of its blends with polyethylene. *ACS Sustain. Chem. Eng.* **3**, 349–356 (2015).
78. Sen, S., Patil, S. & Argyropoulos, D. S. Thermal properties of lignin in copolymers, blends, and composites: a review. *Green Chem.* **17**, 4862–4887 (2015).
79. Cui, C., Sadeghifar, H., Sen, S. & Argyropoulos, D. S. Toward thermoplastic lignin polymers; part II: thermal & polymer characteristics of kraft lignin & derivatives. *Bioresources* **8**, 864–886 (2013).
80. Sadeghifar, H., Cui, C. & Argyropoulos, D. S. Toward thermoplastic lignin polymers. Part 1. Selective masking of phenolic hydroxyl groups in kraft lignins via methylation and oxypropylation chemistries. *Industr. Eng. Chem. Res.* **51**, 16713–16720 (2012).
81. Melone, F., Saladino, R., Lange, H. & Crestini, C. Tannin structural elucidation and quantitative ^{31}P NMR analysis. 2. Hydrolyzable tannins and proanthocyanidins. *J. Agric. Food Chem.* **61**, 9316–9324 (2013).
82. Granata, A. & Argyropoulos, D. S. 2-Chloro-4,4,5,5-tetramethyl-1,3,2-dioxaphospholane, a reagent for the accurate determination of the uncondensed and condensed phenolic moieties in lignins. *J. Agric. Food Chem.* **43**, 1538–1544 (1995).
83. Jiang, Z. H. & Argyropoulos, D. S. The stereoselective degradation of arylglycerol-β-aryl ethers during kraft pulping. *J. Pulp Pap. Sci.* **20**, 183–188 (1994).
84. Balakshin, M. & Capanema, E. On the quantification of lignin hydroxyl groups with ^{31}P and ^{13}C NMR spectroscopy. *J. Wood Chem. Technol.* **35**, 220–237 (2015).
85. Melone, F., Saladino, R., Lange, H. & Crestini, C. Tannin structural elucidation and quantitative ^{31}P NMR analysis. 1. Model compounds. *J. Agric. Food Chem.* **61**, 9307–9315 (2013).
86. Dais, P. & Spyros, A. ^{31}P NMR spectroscopy in the quality control and authentication of extra-virgin olive oil: a review of recent progress. *Magn. Reson. Chem.* **45**, 367–377 (2007).
87. Vigli, G., Philippidis, A., Spyros, A. & Dais, P. Classification of edible oils by employing ^{31}P and ^{1}H NMR spectroscopy in combination with multivariate statistical analysis. A proposal for the detection of seed oil adulteration in virgin olive oils. *J. Agric. Food Chem.* **51**, 5715–5722 (2003).
88. Spyros, A. & Dais, P. Application of ^{31}P NMR spectroscopy in food analysis. 1. Quantitative determination of the mono- and diglyceride composition of olive oils. *J. Agric. Food Chem.* **48**, 802–805 (2000).
89. Fragaki, G., Spyros, A., Siragakis, G., Salivaras, E. & Dais, P. Detection of extra virgin olive oil adulteration with lampante olive oil and refined olive oil using nuclear magnetic resonance spectroscopy and multivariate statistical analysis. *J. Agric. Food Chem.* **53**, 2810–2816 (2005).
90. Fronimaki, P., Spyros, A., Christophoridou, S. & Dais, P. Determination of the diglyceride content in Greek virgin olive oils and some commercial olive oils by employing ^{31}P NMR spectroscopy. *J. Agric. Food Chem.* **50**, 2207–2213 (2002).

UNCORRECTED PROOF

91. Dais, P. & Hatzakis, E. Quality assessment and authentication of virgin olive oil by NMR spectroscopy: a critical review. *Anal. Chim. Acta* **765**, 1–27 (2013).
92. Hatzakis, E., Agiomyrgianaki, A. & Dais, P. Detection and quantification of free glycerol in virgin olive oil by ^{31}P-NMR spectroscopy. *J. Am. Oil Chem. Soc* **87**, 29–34 (2010).
93. Hatzakis, E. & Dais, P. Determination of water content in olive oil by ^{31}P NMR spectroscopy. *J. Agric. Food Chem.* **56**, 1866–1872 (2008).
94. Jiang, Z.-H. & Argyropoulos, D. S. Coupling P-31 NMR with the Mannich reaction for the quantitative analysis of lignin. *Can. J. Chem.* **76**, 612–622 (1998).
95. Argyropoulos, D. S. Quantitative phosphorus-31 NMR analysis of lignins, a new tool for the lignin chemist. *J. Wood Chem. Technol.* **14**, 45–63 (1994).
96. Argyropoulos, D. S. Quantitative phosphorus-31 NMR analysis of six soluble lignins. *J. Wood Chem. Technol.* **14**, 65–82 (1994).
97. Akim, L. G. et al. Quantitative ^{31}P NMR spectroscopy of lignins from transgenic poplars. *Holzforschung* **55**, 386–390 (2001).
98. Zawadzki, M. & Ragauskas, A. N-hydroxy compounds as new internal standards for the ^{31}P-NMR determination of lignin hydroxy functional groups. *Holzforschung* **55**, 283–285 (2001).
99. Li, M. et al. Downregulation of pectin biosynthesis gene GAUT4 leads to reduced ferulate and lignin-carbohydrate cross-linking in switchgrass. *Commun. Biol.* **2**, 22 (2019).
100. Meng, X. et al. Characterization of fractional cuts of co-solvent enhanced lignocellulosic fractionation lignin isolated by sequential precipitation. *Bioresour. Technol.* **272**, 202–208 (2019).
101. Meng, X. et al. Chemical transformations of poplar lignin during cosolvent enhanced lignocellulosic fractionation process. *ACS Sustain. Chem. Eng.* **6**, 8711–8718 (2018).
102. Filpponen, I. & Argyropoulos, D. S. Determination of cellulose reactivity by using phosphitylation and quantitative ^{31}P NMR spectroscopy. *Ind. Eng. Chem. Res.* **47**, 8906–8910 (2008).
103. King, A. W. T. et al. In situ determination of lignin phenolics and wood solubility in imidazolium chlorides using ^{31}P NMR. *J. Agric. Food Chem.* **57**, 8236–8243 (2009).
104. Sadeghifar, H., Dickerson, J. P. & Argyropoulos, D. S. Quantitative ^{31}P NMR analysis of solid wood offers an insight into the acetylation of its components. *Carbohydr. Polym.* **113**, 552–560 (2014).
105. King, A. W. T. et al. A new method for rapid degree of substitution and purity determination of chloroform-soluble cellulose esters, using ^{31}P NMR. *Anal. Methods* **2**, 1499–1505 (2010).
106. Ben, H. et al. Characterization of whole biomasses in pyridine based ionic liquid at low temperature by ^{31}P NMR: an approach to quantitatively measure hydroxyl groups in biomass as their original structures. *Front. Energy Res.* **6** (2018). Q19
107. Constant, S. et al. New insights into the structure and composition of technical lignins: a comparative characterisation study. *Green Chem.* **18**, 2651–2665 (2016).
108. Chen, T. et al. Effects of hydrothermal pretreatment on the structural characteristics of organosolv lignin from *Triarrhena lutarioriparia*. *Polymers* 10, 1157/1151–1157/1114 (2018).
109. Wang, C. et al. Revealing the structure and distribution changes of *Eucalyptus* lignin during the hydrothermal and alkaline pretreatments. *Sci. Rep.* **7**, 593 (2017).
110. Wan, X. et al. Fractionating wheat straw via phosphoric acid with hydrogen peroxide pretreatment and structural elucidation of the derived lignin. *Energy Fuels* **32**, 5218–5225 (2018).
111. Li, T. et al. Deep eutectic solvents (DESs) for the isolation of willow lignin (*Salix matsudana* cv. Zhuliu). *Int. J. Mol. Sci.* **18**, 2266/2261–2266/2211 (2017).
112. Jin, L. et al. Fast dissolution pretreatment of the corn stover in gamma-valerolactone promoted by ionic liquids: selective delignification and enhanced enzymatic saccharification. *Bioresour. Technol.* **270**, 537–544 (2018).
113. Wang, H. et al. Structural changes of bagasse during the homogeneous esterification with maleic anhydride in ionic liquid 1-allyl-3-methylimidazolium chloride. *Polymers* 10, 433/431–433/415 (2018).
114. Thierry, M. et al. Imidazolium-based ionic liquids as efficient reagents for the C-O bond cleavage of lignin. *ChemSusChem* **11**, 439–448 (2018).
115. Lagerquist, L. et al. Structural characterization of birch lignin isolated from a pressurized hot water extraction and mild alkali pulped biorefinery process. *Ind. Crops Prod.* **111**, 306–316 (2018).
116. Zhang, Y. et al. Revealing the structure of bamboo lignin obtained by formic acid delignification at different pressure levels. *Ind. Crops Prod.* **108**, 864–871 (2017).
117. Dutta, T. et al. Survey of lignin-structure changes and depolymerization during ionic liquid pretreatment. *ACS Sustain. Chem. Eng.* **5**, 10116–10127 (2017).
118. Verma, S. R. & Dwivedi, U. N. Lignin genetic engineering for improvement of wood quality: applications in paper and textile industries, fodder and bioenergy production. *South Afr. J. Botany* **91**, 107–125 (2014).
119. Fu, C. et al. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA* **108**, 3803–3808 (2011).
120. Samuel, R. et al. Structural characterization of lignin in wild-type versus COMT down-regulated switchgrass. *Front. Energy Res.* **1**, 1–9 (2014).
121. Tohmura, S.-i & Argyropoulos, D. S. Determination of arylglycerol-β-aryl ethers and other linkages in lignins using DFRC/^{31}P NMR. *J. Agric. Food. Chem.* **49**, 536–542 (2001).

UNCORRECTED PROOF

122. Fu, Q., Argyropoulos, D. S., Tilotta, D. C. & Lucia, L. A. Products and functional group distributions in pyrolysis oil of chromated copper arsenate (CCA)-treated wood, as elucidated by gas chromatography and a novel ^{31}P NMR-based method. *Industr. Eng. Chem. Res.* **46**, 5258–5264 (2007).
123. Mohan, D., Pittman, C. U. & Steele, P. H. Pyrolysis of wood/biomass for bio-oil: a critical review. *Energy Fuels* **20**, 848–889 (2006).
124. Hao, N., Ben, H., Yoo, C. G., Adhikari, S. & Ragauskas, A. J. Review of NMR characterization of pyrolysis oils. *Energy Fuels* **30**, 6863–6880 (2016).
125. David, K. et al. Chemical characterization and water content determination of bio-oils obtained from various biomass species using ^{31}P NMR spectroscopy. *Biofuels* **3**, 123–128 (2012).
126. Sharifzadeh, M., Richard, C. J. & Shah, N. Modelling the kinetics of pyrolysis oil hydrothermal upgrading based on the connectivity of oxygen atoms, quantified by ^{31}P-NMR. *Biomass Bioenergy* **98**, 272–290 (2017).
127. Richard, C. J., Patel, B., Chadwick, D. & Hellgardt, K. Hydrothermal deoxygenation of pyrolysis oil from Norwegian spruce: *Picea abies*. *Biomass Bioenergy* **56**, 446–455 (2013).
128. David, K. et al. ^{31}P-NMR analysis of bio-oils obtained from the pyrolysis of biomass. *Biofuels* **1**, 839–845 (2010).
129. Olarte, M. V. et al. *Determination of hydroxyl groups in pyrolysis bio-oils using 31P-NMR: laboratory analytical procedure (LAP). Technical Report No. NREL/TP-5100-65887*. (National Renewable Energy Laboratory, Golden, CO, USA, 2016).
130. Lawoko, M. Unveiling the structure and ultrastructure of lignin carbohydrate complexes in softwoods. *Int. J. Biol. Macromol.* **62**, 705–713 (2013).
131. Zhao, B.-C. et al. Selective precipitation and characterization of lignin–carbohydrate complexes (LCCs) from *Eucalyptus*. *Planta* **247**, 1077–1087 (2018).
132. Li, M., Pu, Y., Tschaplinski, T. J. & Ragauskas, A. J. ^{31}P NMR characterization of tricin and its structurally similar flavonoids. *ChemistrySelect* **2**, 3557–3561 (2017).
133. Li, M., Yoo, C. G., Pu, Y. & Ragauskas, A. J. ^{31}P NMR chemical shifts of solvents and products impurities in biomass pretreatments. *ACS Sustain. Chem. Eng.* **6**, 1265–1270 (2018).
134. Pu, Y., Cao, S. & Ragauskas, A. J. Application of quantitative ^{31}P NMR in biomass lignin and biofuel precursors characterization. *Energy Environ. Sci.* **4**, 3154–3166 (2011).
135. Cateto, C. A. et al. Lignins as macromonomers for polyurethane synthesis: a comparative study on hydroxyl group determination. *J. Appl. Polym. Sci.* **109**, 3008–3017 (2008).
136. Gartner, A., Gellerstedt, G. & Tamminen, T. Determination of phenolic hydroxyl groups in residual lignin using a modified UV-method. *Nord. Pulp Pap. Res. J* **14**, 163–170 (1999).
137. Hu, L. et al. Chemical groups and structural characterization of lignin via thiol-mediated demethylation. *J. Wood Chem. Technol.* **34**, 122–134 (2014).
138. Serrano, L. et al. Fast, easy, and economical quantification of lignin phenolic hydroxyl groups: comparison with classical techniques. *Energy Fuels* **32**, 5969–5977 (2018).
139. Faix, O., Argyropoulos, D. S., Robert, D. & Neirinck, V. Determination of hydroxyl groups in lignins evaluation of ^{1}H-, ^{13}C-, ^{31}P-NMR, FTIR and wet chemical methods. *Holzforschung* **48**, 387–394 (1994).
140. Korntner, P., Sumerskii, I., Bacher, M., Rosenau, T. & Potthast, A. Characterization of technical lignins by NMR spectroscopy: optimization of functional group analysis by ^{31}P NMR spectroscopy. *Holzforschung* **69**, 807–814 (2015).
141. Safou-Tchiama, R., Barhe, T. A., Souloungangа, P., Akagah, A. G. & De Jeso, B. A comparative study of the syringyl, guaiacyl and hydroxyl groups units distribution in some African tropical hardwoods' lignin by Py-GC/MS and spectroscopic techniques. *J. Mater. Environ. Sci.* **8**, 2530–2540 (2017).
142. Stücker, A., Podschun, J., Saake, B. & Lehnen, R. A novel quantitative ^{31}P NMR spectroscopic analysis of hydroxyl groups in lignosulfonic acids. *Anal. Methods* **10**, 3481–3488 (2018).
143. Ben, H. & Ferrell Iii, J. R. In-depth investigation on quantitative characterization of pyrolysis oil by ^{31}P NMR. *RSC Adv.* **6**, 17567–17573 (2016).

Acknowledgements

This research used resources of the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, under contract no. DE-AC05-00OR22725 with the US Department of Energy (DOE). This study was supported and performed as part of the Center for Bioenergy Innovation (CBI). The CBI is a DOE Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The research effort that established the quantitative ^{31}P NMR in wood chemistry protocol was supported by FP Innovations (formerly Pulp and Paper Research Institute of Canada) and the Natural Sciences and Engineering Research Council of Canada.

Author contributions

A.R. proposed and designed the manuscript. X.M. wrote the manuscript with major input from D.A., C.C., H.B., Y.P., and A.R. X.M. and N.H. performed the organosolv pretreatment. X.M. and Y.P. performed the NMR experiments and data processing on lignins. C.C. carried out the NMR experiments and data processing on tannins. For specific questions regarding the conception and the foundations of the protocol, including selection of an appropriate solvent, relaxation time, and synthesis of phosphorus reagent, contact D.A. Questions about the characterization of tannin should be referred to C.C. For selection of an appropriate IS and the application of ^{31}P NMR for the characterization of native/transgenic plants and untreated or pretreated biomass, please contact A.R. All authors approved the final version of the manuscript.

UNCORRECTED PROOF

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-019-0191-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.C.

Peer review information: *Nature Protocols* thanks Luc Avérous and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 4 February 2019; Accepted: 9 May 2019;

Related links

Key references using this protocol

Gioia, C. et al. *J. Am. Chem. Soc.* **140**, 4054–4061 (2018): https://pubs.acs.org/doi/abs/10.1021/jacs.7b13620
Li, M. et al. *Commun. Biol.* **2**, 22 (2019): https://www.nature.com/articles/s42003-018-0265-6
Granata, A. & Argyropoulos, D. S. *J. Agric. Food Chem.* **43**, 1538–1544 (1995): https://pubs.acs.org/doi/10.1021/jf00054a023
Crestini, C., Lange, H. & Bianchetti, G. *J. Nat. Prod.* **79**, 2287–2295 (2016): https://pubs.acs.org/doi/abs/10.1021/acs.jnatprod.6b00380

UNCORRECTED PROOF

QUERY FORM

NPROT	
Manuscript ID	**[Art. Id: 191]**
Author	
Editor	
Publisher	

Journal: NPROT

Author :- The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Description	Author's Response
AQ1	Please note the change made to the article title ("nuclear magnetic resonance" shortened to NMR). OK?	
AQ2	Please carefully check the spelling and numbering of all author names and affiliations.	
AQ3	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.	
AQ4	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.	
AQ5	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).	
AQ6	Please note that after the paper has been formally accepted you can only provide amended Supplementary Information files for critical changes to the scientific content, not for style. You should clearly explain what changes have been made if you do resupply any such files.	

UNCORRECTED PROOF

QUERY FORM

NPROT	
Manuscript ID	**[Art. Id: 191]**
Author	
Editor	
Publisher	

Journal: NPROT

Author :- The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Description	Author's Response
AQ7	Per journal style, we reserve use of "significant" for mentions of statistical significance. At all uses of this word in the text, if you are referring to statistical significance, please provide a P value. If not, please revise to "substantial," etc.	
AQ8	"1,3,2-dioxaphospholane" above was changed to 2-chlorine-1,3,2-dioxaphospholane in keeping with the standard expansion of CDP; is this correct?	
AQ9	Please confirm the edits made to the Reagents section are ok.	
AQ10	Per journal requirements, all items listed in the "Reagents" and "Equipment" sections should include the supplier name and catalog and/or model number (aside from materials such as regular water that would not be obtained from an outside supplier). Please check the Equipment section carefully and add this information where applicable.	
AQ11	Please confirm the edits made to the Equipment section are ok.	
AQ12	Please define PEFE at its first mention above.	
AQ13	Please confirm the edits made to the Timing section are ok.	
AQ14	Editor's comment: Please add the data points of the individual replicates to this figure. Individual data points should be provided for graphs that are composed of fewer than 6 data points.	

UNCORRECTED Proof

QUERY FORM

NPROT	
Manuscript ID	**[Art. Id: 191]**
Author	
Editor	
Publisher	

Journal: NPROT

Author :- The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Description	Author's Response
AQ15	Please check whether the edit made to the sentence 'Figure 7 shows...' retains the intended meaning.	
AQ16	Please check the edits made to the sentence 'The ratio of the...'.	
AQ17	Please check the edits made to the related links section.	
AQ18	Please check that all funders have been appropriately acknowledged and that all grant numbers are correct.	
AQ19	Please provide page range/article number in reference 106.	

UNCORRECTED PROOF

nature research

Corresponding author(s): Arthur J. Ragauskas

Last updated by author(s): 4/2/2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Bruker TopSpin 3.5 pl 6 was used to collect NMR data
Data analysis	Bruker Topspin 3.5 pl 7 and Microsoft Excel 2016 were used for NMR data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated during this study are included in this published article. The NMR integration data are available upon request. The software used for NMR data analysis is freely available ('Equipment').

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Lignin and tannins were characterized by 31P NMR technique. The sample size (n) of each experiment is provided in the figure/table legends in the main manuscript. Three technical replicates were performed on each lignin sample which results in reproducible data that support meaningful conclusions.
Data exclusions	No data are excluded
Replication	All replicate analysis attempts were successful.
Randomization	Not applicable.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging