Multimethod Approach for Extensive Characterization of Gallnut Tannin Extracts

Aude Watrelot, Christine Le Guernevé, Hélène Hallé, Emmanuelle Meudec, Pascale Williams, Bertrand Robillard, François Garcia, Celine Poncet-Legrand, Veronique Cheynier

Multi-method approach for extensive characterization of gallnut tannin extracts

Journal:	*Journal of Agricultural and Food Chemistry*
Manuscript ID	jf-2019-082217.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Watrelot, Aude; INRAE, SPO; Iowa State University College of Agriculture and Life Sciences Le Guernevé, Christine; INRAE, SPO Halle, Hélène; INRAE, SPO Meudec, Emmanuelle; INRAE, SPO Veran, Frederic; INRAE, SPO Williams, Pascale; INRAE, SPO Robillard, Bertrand; Institut Œnologique de Champagne Garcia, François ; Univ Montpellier, SPO Poncet-Legrand, Céline; INRA, Umr Spo Cheynier, Veronique ; INRAE, SPO

Multi-method approach for extensive characterization of gallnut tannin extracts

Aude A. Watrelot[1†], Christine Le Guernevé[1], Hélène Hallé[1], Emmanuelle Meudec[1], Frédéric Véran[1], Pascale Williams[1], Bertrand Robillard[2], François Garcia[1], Céline Poncet-Legrand[1], Véronique Cheynier[1]*

[1]SPO, INRAE, Univ Montpellier, Montpellier SupAgro, 2 Place Pierre Viala, 34060 Montpellier, France.

[2]Institut Œnologique de Champagne, Z.I. de Mardeuil, BP 25, 51201 Epernay cedex, France

*corresponding author:

V. Cheynier

Phone: +33 (0)4 99 61 22 98

Fax: +33 (0)4 99 61 28 57

veronique.cheynier@inra.fr

[†]present address: Iowa State University, Department of Food Science and Human Nutrition, 536 Farm House Lane, Ames, IA 50011-1054, USA. (watrelot@iastate.edu)

ABSTRACT

Gallotannins extracted from gallnuts are commonly added to wine to improve its properties. They consist of mixtures of galloylester derivatives of glucose. However, their composition and properties are not well established. In this study, methods based on liquid chromatography coupled to UV-visible detection and mass spectrometry, size exclusion chromatography and 1D (^{31}P) and 2D (^{1}H DOSY, ^{31}P TOCSY, ^{1}H/^{13}C HSQC, HMBC) NMR spectroscopy have been implemented for extensive chemical characterization of three commercial gallnut tannin extracts. Differences in the proportions of the different constituents (gallic, digallic, trigallic acids, galloylglucose derivatives) and in the structure and molecular weight distributions of gallotannins were demonstrated between the three extracts, with chains containing 8.5, 12.2, and 12.4 galloyl groups in average for TAN A, B1, and B2, respectively. The antioxidant capacities of the extracts, evaluated using the ABTS method, were similar and related mostly to their total tannin content, with only limited impact of the tannin composition.

Keywords: enological tannins, gallotannins, polygalloylglucose, ^{31}P NMR, ^{1}H, 2D ^{1}H/^{13}C, ^{1}H DOSY NMR, UPLC-DAD-ESI-MS, size exclusion chromatography, molecular weight distribution, antioxidant properties.

INTRODUCTION

Oenological tannins are extracted from different botanical sources with water and/or alcohol. They are commonly used in winemaking to stabilize color, to facilitate the clarification of wine and musts, promote partial precipitation of proteins and prevent protein haze, and contribute to the antioxidant protection of wines, but they must not modify the wine color nor bring further aroma, according to the International Oenological Codex (OIV-Oeno 613-2019; http://www.oiv.int/public/medias/6881/oiv-oeno-613-2019-en.pdf). Among the large polyphenol group, tannins are defined by their capacity to precipitate alkaloids and proteins.[1] They comprise condensed tannins which are flavanol oligomers and polymers found in fruits such as grape, apple, or persimmon, as well as in tea and quebracho, and hydrolysable tannins. The latter are subdivided in two categories, both composed of a polyol core and galloyl groups. Gallotannins, extracted from gallnuts of *Quercus infectoria* (Turkish or Alep gall*) or Rhus semialata* (Chinese gall) and from tara (*Caesalpina spinosa*) consist of polygalloyl esters of glucose and quinic acid, respectively, whereas ellagitannins, found in oak and chestnut, contain hexahydroxydiphenoyl (HHDP) units formed by oxidative linkage of galloyl groups.[2,3]

According to the international code of enological practices (http://www.oiv.int/public/medias/6558/code-2019-en.pdf), tannin extracts sold as enological tannins must contain 65 % polyphenols. In the earlier versions of the code, the method imposed for estimation of tannin content was based on spectrophotometry measurement at 280 nm, converted to gallic acid equivalent, but this has been recently replaced by gravimetric analysis of the water soluble material adsorbed on a polyvinylpolypyrrolidone (PVPP) solid phase extraction column. PVPP is a cross-linked synthetic polymer of polyvinylpyrrolidone known to bind polyphenols through mechanisms similar to protein tannin interactions. This method thus relies on

the assumption that adsorption by PVPP is specific of tannins, like other common assays based on precipitation with protein such as bovine serum albumin[4] or with methylcellulose[5]. These properties are related to the use of tannins in preventing protein haze[6] and to astringency perception,[7,8] which is attributed to formation of insoluble complexes between tannins and salivary proteins.[9–11]

Comparison of the official OIV methods performed in a recent study confirmed that they provide different results and all lack specificity.[6] Liquid chromatography analysis of gallic acid or methyl gallate released after acid hydrolysis[12] or methanolysis,[13] respectively, is also commonly used to quantify gallotannins. However, these methods do not allow the determination of gallotannin composition in commercial tannin extracts.

Gallotannins extracted from Chinese galls were shown to be composed of a pentagalloylglucose core to which three or four galloyl groups were linked by depsidic linkages, forming up to a nonagalloylglucose.[14] The presence of longer polygalloyl chains in the tannin structure has also been suggested.[14,15] Chromatographic analysis provided evidence of heterogeneous mixtures of gallotannins composed of a glucose linked to up to 12 galloyl groups in Chinese galls,[16–18] and up to 8 in Turkish galls,[16,18] The latter was described as a mixture of galloylglucoses based on tetragalloylglucose cores[19] or on both tetragalloylglucose and pentagalloylglucose cores.[20] Moreover, depside gallic acid oligomers up to the octamer have also been detected in a commercial gallnut extract.[12]

Gallotannins from gallnut extracts thus seem to be very polydisperse in a given extract and the chemical structure of these tannins as well as the position of linkages between glucose and gallic acid residues is not fully identified and characterized. In addition, different gallnut extracts differ in their tannin composition and thus potentially in their properties. Indeed, protein precipitation

capacity was found to vary with the composition of the gallotannin mixture and generally to increase with the degree of galloylation.[21,22] Moreover, pentagalloylglucose is highly antioxidant compared to a monogalloylglucose, suggesting that the antioxidant activity increases with the number of galloyl units.[23,24] However, gallic acid itself showed higher antioxidant activity than pentagalloylglucose.[25] In a recent study, measurement of the antioxidant activity by diverse methods showed 2 to 3 fold differences between five commercial gallnut extracts but their chemical composition was not elucidated and no structure-activity relationship was identified.[6]

The aim of this study is to extensively characterize three commercial gallnut tannin extracts using a combination of analytical methods as well as determine their antioxidant properties. UPLC-DAD-MS analysis enabled detection of the species present in the sample. Structural and quantitative information was obtained by 1D ^{1}H, ^{31}P NMR and 2D $^{1}H/^{13}C$ spectroscopy and size distribution was obtained by 2D ^{1}H DOSY NMR spectroscopy and gel permeation chromatography (GPC). The antioxidant capacities of the gallnut tannin extracts were measured by an ABTS method.

MATERIAL AND METHODS

Chemicals

Acetonitrile, water, methanol, acetic acid and formic acid were LC-MS grade from VWR Prolabo (Fontenay sous Bois. France). Gallic acid, 2,2'-azino-bis(3-ethyl-benzothiazoline-6-sulfonic acid) diammonium salt (ABTS), L-tartaric acid, ethanol, sodium hydroxide, potassium persulfate, β-glucogallin, 1,2,3,4,6-pentagalloylglucose (PGG), methyl gallate, deuterated dimethylsulfoxide (DMSO-d6), deuterated pyridine (pyridine-d5), deuterated chloroform ($CDCl_3$), 2-chloro-4,4,5,5-

tetramethyl-1,3,2-dioxaphospholane (cTMDP), lithium chloride, and hydrochloric acid 37 % were provided by Sigma-Aldrich (Saint Louis, MO, USA), and N,N-dimethylformamide by Carl Roth GmbH (Karsruhe, Germany).

Three commercial gallotannin extracts (TAN A, TAN B1, TAN B2), available in the powder form and sold to the wine industry with claims related to their antioxidant properties were analyzed. Sample A was Tanal W2 (Ajioto Onmnichem, Wetteren, Belgium). Samples B1 and B2 correspond to two different batches of the same commercial brand (Tan'Activ GC, SilvaTeam, San Michele Mondovì, Italy). All three were stored in the dark in a desiccator until analysis.

Neutral Sugar Characterization by Gas Chromatography

Neutral sugar composition was determined as alditol acetates after hydrolysis, reduction and acetylation of 1 mg of gallnut tannin extracts as previously described.[26] The alditol acetates were quantified by gas chromatography with flam ionization detection (GC-FID) (GC 2010 Plus Shimadzu) using a DB225 (30 m × 0.25 mm ID, 0.25 µm film) capillary column and hydrogen 5.6 B50 as the carrier gas. Allose and myo-inositol were used as internal standards for the neutral sugar quantification.

Compound Identification and Quantification by UPLC-DAD-MS

UPLC-DAD-MS analyses were performed using an Acquity UPLC-DAD system (Waters, Milford, MA) equipped with an Amazon X mass spectrometer (Bruker Daltonics, Bremen, Germany), piloted by Hystar software. For direct analysis of the gallnut tannin extracts, they were dissolved in triplicate in a solution of water: methanol (50:50, v/v) containing 1 % formic acid at a concentration of 5 g/L and vortexed. The extracts were centrifuged at 15000 rpm (Mikro 220R, Hettich, Germany) for 15 min at 15 ºC prior to be injected (0.5 µL) to the UPLC system. The

gallnut tannin extracts were also submitted to an acidic methanolysis. 5 mg of extract were dissolved in 2.5 mL of acidified methanol 0.6 N HCl. The methanolysis reaction was performed at 120 ºC for 260 min then samples were placed in ice to cool down to room temperature. The samples were transferred into 5 mL volumetric flasks and the volume was adjusted with water. The depolymerized samples were then injected to the UPLC system after centrifugation at 15,000 rpm for 15 min at 15 ºC. The column was a BEH C18 (1.7 µm×1 mm×150 mm) equipped with a 0.2 µm prefilter (Waters, Milford, MA) placed in an oven at 35 ºC. The flow rate was 0.08 mL/min and the run time was 45 min. The gradient conditions consisting in solvent A (water with 1 % formic acid) and solvent B (acetonitrile with 1 % formic acid) was as follows: 0 min (1.3 % B), 6.5 min (10 % B), 12 min (20 % B), 27 min (50 % B), 32 min (99.5 % B), 40 min (1.3 % B). The detection was both with the DAD from 250 to 600 nm and with the MS equipped with an electrospray-ionization source and an ion trap analyzer. The MS detection was operated in the negative ion mode with the full scan mode under the following conditions: a scan between m/z 150 to 2000, the target mass set at m/z 400, a capillary tension of 4500 V, the nebulizer gas (N_2) at 10 psi, and N_2 as the desolvation gas at 5 L/min and 200 °C. The collision energy for fragmentation used for the MS2 experiments was set at 1 eV.

The quantification of the constituents in the crude extracts and in the extracts after methanolysis was based on the calibration curve of standards based upon the peak area at 280 nm. Digallic and trigallic acids and mono-, di-, tri-, and poly-galloylglucose were expressed as gallic acid equivalent. After methanolysis, methyl gallate and methyl di-gallate were expressed as methyl gallate equivalent.

Determination of Molecular Size Distribution by Gel Permeation Chromatography

The molecular size distribution of gallnut tannin extracts was carried out by gel permeation chromatography according to a previously published method[27] slightly modified as follows. The extracts were dissolved in the mobile phase (2.5 g/L) consisting in dimethylformamide with 1 % acetic acid (v/v), 5 % water (v/v) and 0.15 M lithium chloride, in triplicate. After vortex then centrifugation at 15000 rpm for 15 min at 15 °C, 20 µL of sample were injected to the high-performance liquid chromatography system. The Agilent HPLC 1260 Infinity II system equipped with a diode array detector consisted in two Phenogel columns connected in series (300 mm × 7.8 mm, 5 µm 50 Å and 300 mm × 7.8 mm, 5 µm 1000 Å) protected by a guard column (50 mm × 7.8 mm) from the same material placed at 60 °C. The isocratic solvent was dimethylformamide containing 1 % acetic acid (v/v), 5 % water (v/v) and 0.15 M lithium chloride filtered at 0.2 µm, at a flow rate of 0.8 mL/min, and the run time was 35 min. The detection was carried out at 280 nm and commercial gallic acid, β-glucogallin, and 1,2,3,4,6-pentagalloylglucose were injected under the same conditions and used as size distribution standards.

NMR Spectroscopy

All NMR experiments were conducted on an Agilent DD2 500 MHz spectrometer (Agilent Technologies, Santa Clara, Calif., USA), operating at 500.02, 202.43 and 175.74 MHz for proton, phosphorus-31, and carbon-13 nuclei respectively.

- ^{1}H 1D, 2D DOSY, and ^{1}H /^{13}C 2D NMR analysis

1D and 2D experiments were performed using a 5mm indirect detection probe equipped with a gradient coil, at 25°C, on tannin extracts (20 mg) dissolved in 500 µL DMSO-d6. 1D ^{1}H and 2D heteronuclear ^{1}H/^{13}C HSQC and HMBC were obtained using classical pulse sequences. DOSY measurements were carried out using the DgcsteSL pulse sequence of VNMRJ4.1 library. Data were acquired with an array of 32 gradients with amplitudes ranging from 3.0 to 50.0 Gcm^{-1} in

equal steps of gradient squared, a total diffusion encoding gradient of 3 ms and a diffusion time of 0.2 s, giving a residual signal intensity of about 5% of the original intensity. For each gradient step, 512 transients were recorded with 64 K complex data points. Free induction decays were zero-filled to 128k points and apodized with a combination of exponential and Gaussian functions set at -0.8 and +0.8 Hz respectively to enhance spectrum resolution. After phase correction, 2D DOSY spectra were constructed from peak height measurement, with a discrete multiexponential fitting at two components and a correction for the effects of pulsed field gradient non-uniformity, using the DOSY module of VNMRJ4.2 software.

- Phosphitylation procedure

The phosphitylation procedure was adapted from that described earlier.[28,29] An amount of 1 to 5 mg of model compounds (gallic acid, β-glucogallin, pentagalloylglucose) depending on their molecular weights or 20 mg of gallnut tannin samples were dissolved in 400 µL of anhydrous NMR solvent consisting in pyridine-d5:$CDCl_3$ (1.6: 1; v/v), containing cholesterol (100 mM) as an internal standard (IS) for the absolute quantification and chromium acetylacetonate, Cr- (acac)3 (3 µM) as a relaxation agent allowing to shorten the relaxation times of the phosphorus nuclei. After complete dissolution, 100 µL of cTMDP was added dropwise under continuous agitation. The mixture was then transferred to a NMR tube. For each tannin extract, three independent phosphitylated samples were prepared and analyzed. ^{31}P spectra were recorded using a 5mm direct gradient probe.

- Quantitative 1D ^{31}P spectra

1D ^{31}P NMR spectra were obtained using a standard pulse sequence with a 90° pulse and inverse-gated ^{1}H decoupling. Spin-lattice relaxation time experiment realized on phosphitylated tannin

extract has shown that the highest T1 value was 1.5 s (measured for the carboxyl group of gallic acid), which was taken into account to set the recycle delay to 15 s, i.e. more than 7 times higher than the longest T1 value, allowing a quite complete relaxation of spins between scans. For the tannin samples, 1D spectra of 512 scans were recorded to quantify with enough precision small peaks. The NMR spectra were processed using Mnova 14.1.0 software. The data were zero-filled to 64 K. A line broadening of 0.1 Hz and a drift correction were applied prior to Fourier transformation and phase correction. The baseline was corrected using a cublic spline function before peak area integration. The phosphitylated hydroxyl signal area of cholesterol was used as internal reference for the determination of the absolute concentrations. All chemical shifts were reported relative to the reaction product of water with Cl-TMDP, which gave a sharp signal in pyridine/$CDCl_3$ at 132.2 ppm.

- 2D ^{31}P TOCSY NMR spectra

2D ^{31}P TOCSY spectra were obtained with a classic TOCSY pulse sequence using a mixing time of 80 ms, 128 increments of 1 K complex data points and 64 scans per increments. The data set was zero-filled to 256 x 1 K points prior to Fourier transformation. A squared sinusoidal window function was used in both dimensions.

Determination of the Antioxidant Capacity using the ABTS Method

The antioxidant capacity of the three extracts was determined using the ABTS method.[30] Briefly, a solution of 7 mM ABTS was mixed with a solution of 2.45 mM potassium persulfate in ethanol: water (1:1, v/v) to produce $ABTS^{+\cdot}$. The solution was then placed at 22 °C for 24 hours, and then diluted in ethanol: water (1:1, v/v) to produce an absorbance of 0.70 ± 0.02 at 734 nm. For tannin analysis, 1 mL of diluted $ABTS^{+\cdot}$ solution and 0.1 mL of tannin solution sample (0.01 g/L in

acidified hydroalcoholic solution (12% (v/v) ethanol, 33 mM L-tartaric acid and pH 3.6 whose value was adjusted with 1 M NaOH)) were mixed and incubated for 6 min before measuring the absorbance at 734 nm. The percentage of inhibition was calculated by the following equation for each tannin extract sample:

$$\%\ inhibition = \frac{(A734\ ABTS^{+\cdot} - A734\ Tannin)}{A734\ ABTS^{+\cdot}} \times 100$$

Where A734 corresponds to the absorbance at 734 nm.

All the experimental measurements were performed in triplicate using a SAFAS UV mc2 spectrophotometer (Monaco).

Statistical Analysis

ANOVA type II sum of squares analyses were performed using RStudio, version 3.5.1, with the FactoMineR, agricolae and car packages. A Tukey LSD test with a p value of 0.05 was performed using XlStat for the ABTS results.

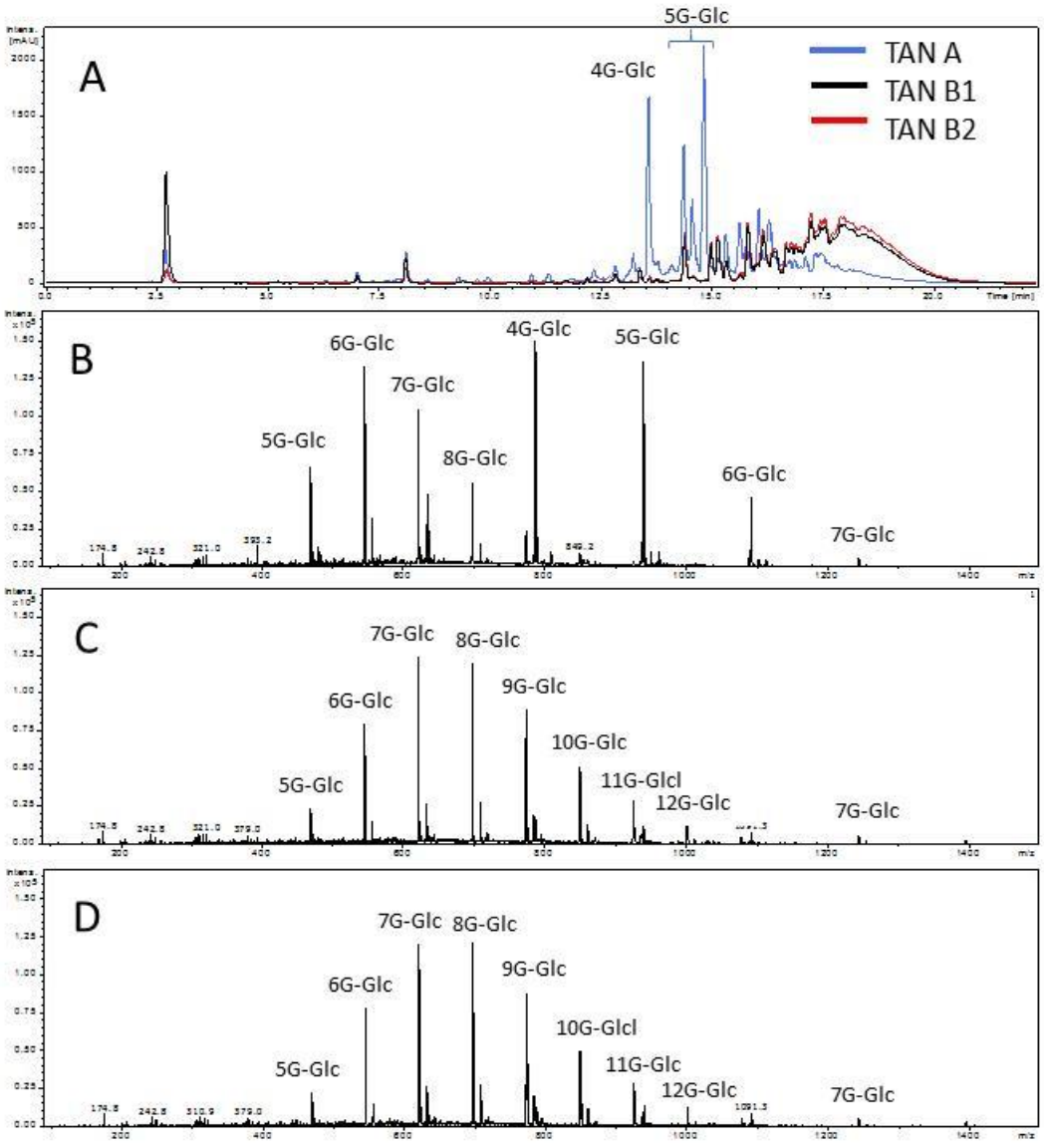
RESULTS AND DISCUSSION

Identification and Quantification of the Constituents of Gallnut Tannin Extracts

- UPLC-DAD-MS

Concentrations evaluated by integration of the total peak area at 280 nm after HPLC separation were 1.06 ± 0.01, 1.14 ± 0.02, 1.13 ± 0.01 g gallic acid equivalent per g of extract (corresponding to 790, 822 and 875 mg/g pentagalloylglucose equivalent per g of sample), for TAN A, B1 and B2, respectively. The LC-UV profile of TAN A appeared very different from those of TAN B1 and B2, the former showing a number of well resolved peaks, while the latter are characterized by a large hump eluted later (Figure 1). Such differences have been observed earlier by normal phase

HPLC analysis, showing that Turkish gall extracts are mixtures of penta and hexagalloylglucose, while Chinese gall extracts contain larger gallotannins.[16,18] Information on constitutive molecules has also been obtained by mass spectrometry. In Turkish gall tannin, gallotannins composed of a single glucose linked to up to 8 gallic acids were detected by MALDI-TOF analysis, with the most abundant signals corresponding to tri- and tetra-galloylglucose.[31]. However, it should be emphasized that relative intensities in the mass spectra do not reflect proportions of tannins of different molecular weights, as peak intensities decrease when molecular weight increases[32–34] which explains the discrepancy between LC-UV profiles and MS data. The most polar compounds identified by LC-MS in all samples were gallic acid (m/z 169), digallic acid (m/z 321, two isomers eluted at 7.2 and 8.5 min), and trigallic acid (m/z 473). Low intensity signals detected at *m/z* 331, 483, 635, 625, and 777 were attributed to monogalloylglucose (tentatively identified as β-glucogallin from its retention time), digalloylglucose, trigalloylglucose, tetragallic acid and pentagallic acid, respectively. Additional minor signals at m/z 495, 647 and 799 detected with higher intensity in TAN B1 were tentatively attributed to di-, tri- and tetra-galloylquinic acids which are present in Tara tannins but usually not detected in gallnut extracts.



256

257 *Figure 1. HPLC trace at 280 nm of TAN A (blue), TAN B1 (black) and TAN B2 (red) (A). Negative*
258 *mode ESI MS spectra (150-1500 m/z) recorded after HPLC-DAD-MS analysis of TAN A (B), TAN*
259 *B1 (C) and TAN B2 (D). G-Glc: galloyl-glucose*

260 The concentration of gallic acid was much higher in TAN B1 than in TAN A and TAN B2 while
261 monogalloylglucose was found in the highest concentration in TAN A (Table 1). The
262 concentrations of digallic acid and trigallic acids were also different between the three samples

263 with TAN A > TAN B1 > TAN B2, but digallic acid was overestimated by LC-UV since its major
264 isomer was coeluted with digalloylglucose.

265 Major peaks detected in TAN A were identified as tetragalloylglucose (eluted at 13.7 min and
266 detected as singly charged $[M-H]^-$ ion at *m/z* 787 and doubly charged $[M-2H]^{2-}$ ion at *m/z* 393) and
267 pentagalloylglucose ($[M-H]^-$ at *m/z* 939, $[M-2H]^{2-}$ at *m/z* 469, 3 isomers, eluted at 14.6, 14.8, and
268 15 min), followed by lower intensity peaks attributed to hexa- ($[M-H]^-$ at *m/z* 1091, $[M-2H]^{2-}$ at
269 *m/z* 545.5), hepta- (*m/z* 1243 and 621.5) and octa- ($[M-2H]^{2-}$ at *m/z* 697.5) galloylglucose. Some
270 of these compounds were also present in TAN B1 and TAN B2 while others, including
271 tetragalloyglucose, were found in significant amounts only in TAN A. The first
272 pentagalloylglucose isomer, eluted at 14.5 minutes, present in all samples and detected mostly as
273 the doubly charged ion at m/z 469, was identified as 1,2,3,4,6-penta-galloylglucose by comparison
274 of its retention time and spectral characteristics with those of the standard. The other two isomers,
275 specific of TAN A, were detected mostly as the singly charged ion at *m/z* 939. Their UV spectra
276 exhibited a shoulder at 300 nm, which is characteristic of depside bonds,[35] suggesting that they
277 are based on a tetra-galloylglucose core substituted by a fifth galloyl moiety through a depside
278 bond. The different pentagalloylglucose isomers also showed different fragmentation patterns.
279 Fragmentation of 1,2,3,4,6-penta-galloylglucose yielded fragments at *m/z* 769, and 617,
280 corresponding to successive losses of gallic acid and water molecules (-152-18) through cleavage
281 of the glucose ester bond. The same fragmentation was obtained from the ion at m/z 787 in TAN
282 A (617, 787-152-18), consistent with the tetragalloylglucose core. In contrast, the major fragment
283 ion obtained from the other pentagalloylglucose isomers was at m/z 787 (-152), arising from
284 cleavage of a depside bond. Similarly, the tetragalloylglucose isomers present in trace amounts in

TAN B1 and B2 yielded a fragment ion at m/z 635 (787-152), suggesting that they are mono-, di-, or tri-galloylglucose derivatives containing depside bonds.

Signals corresponding to hexa- and hepta- galloylglucose were also more polydisperse and less polar in TAN A than in TAN B, indicating that they are based on different structures. Signals corresponding to larger galloylglucose polymers (from 8 to 10 galloyl groups) were more polydisperse in all samples, reflecting a diversity of substitution patterns.

The total concentration of gallotannins determined by UPLC-UV was significantly higher in TAN B2 than in B1 and A (Table 1). The proportion of lower molecular weight oligomers (tetra- and penta-galloylglucose) was much higher in TAN A (Fig 1A) while TAN B1 and B2 contained mostly higher molecular weight tannins eluted as a hump at the end of the chromatogram. The MS spectrum of the hump showed singly charged, doubly charged and triply charged ions that correspond to a polymeric series consisting of a glucose molecule linked to up to twelve (m/z 1001 $[M-2H]^{2-}$), eighteen (m/z 1457 $[M-2H]^{2-}$) and seventeen (m/z 1381 $[M-2H]^{2-}$) galloyl units in TAN A, TAN B1, and TAN B2, respectively (Figure S1 in the Supporting Information). Another set of masses (Δ22 amu, Δ11 amu, respectively for monocharged and doubly charged ions) which likely correspond to sodium adducts was observed with lower intensity for each galloylglucose polymer.

- UPLC-DAD-MS after methanolysis

After methanolysis under acidic conditions, the majority of gallic acids involved in gallotannins formed methyl gallate (around 600 mg/g) and methyl digallate (around 50 mg/g) (data not shown). Detection of methyl digallate and of m/z values corresponding to di-, tri- and tetra- and penta-galloylglucose (*m/z* 483, 635, 787, 939) indicated that methanolysis. was incomplete. Moreover, di-, tri- and tetra- and penta-galloylglucose methyl esters (*m/z* 345, 497, 649, 801, 953) were also

detected, showing that some glucose was linked to gallic acid through ether rather than ester bonds, as proposed earlier.[14,36] Further analysis of the LC-MS data showed that several isomers of the methyl esters of galloylglucose ethers were present in all samples. Their proportions in TAN A were different from those observed in TAN B1 and B2. Moreover, these structures were significantly more abundant in TAN B than in TAN A and in TAN B2 than in TAN B1 (Figures S2 and S3 in supporting information).

- NMR Spectroscopy

Further results upon composition of the tannin extracts were obtained from ^{31}P NMR spectroscopy analysis after phosphitylation of the free hydroxylic groups. This method has been shown to give both qualitative and quantitative information on samples from various origins such as tannins, lignins, olive oils, biomass and the chemical shifts of functionalized groups are well documented.[28,29,37–42] It has also provided evidence of both *meta* and *para*-depside linkages in Chinese and Turkish gall extracts.[29] ^{31}P NMR signals arising from the three classes of phosphitylated functional groups appear in three well-separated chemical shift regions: the aliphatic hydroxyl groups are observed in the most deshielded region between ~146-149 ppm, the carboxyl groups give the most shielded signals, between~134.5-135.5 ppm, and the aromatic hydroxyl groups cover a wider range between ~137-143 ppm.

More precise assignment of the aromatic hydroxyl peaks was performed with the help of ^{31}P TOCSY scalar correlations using a mixture of gallotannin standards (Figure 2). P-P long-range scalar correlations are indeed observed between the mono-*ortho*-substituted non-equivalent OP3 / OP5 located in the sub-region ~137.5 to 138.7 ppm and between the mono-*ortho*-substituted OP3 / OP5 and the di-*ortho*-substituted OP4 found between ~140.5-141.5 ppm. These attributions are in agreement with those indicated by Melone et al.[28] As expected, phosphitylated glucose hydroxyl

peaks appear in the most deshielded region between ~146.0-148.2 ppm. It is worth noting that correlations occurred between glucose OP2 and aromatic phosphitylated hydroxyls OP3/OP5 and OP4 of β-glucogallin.

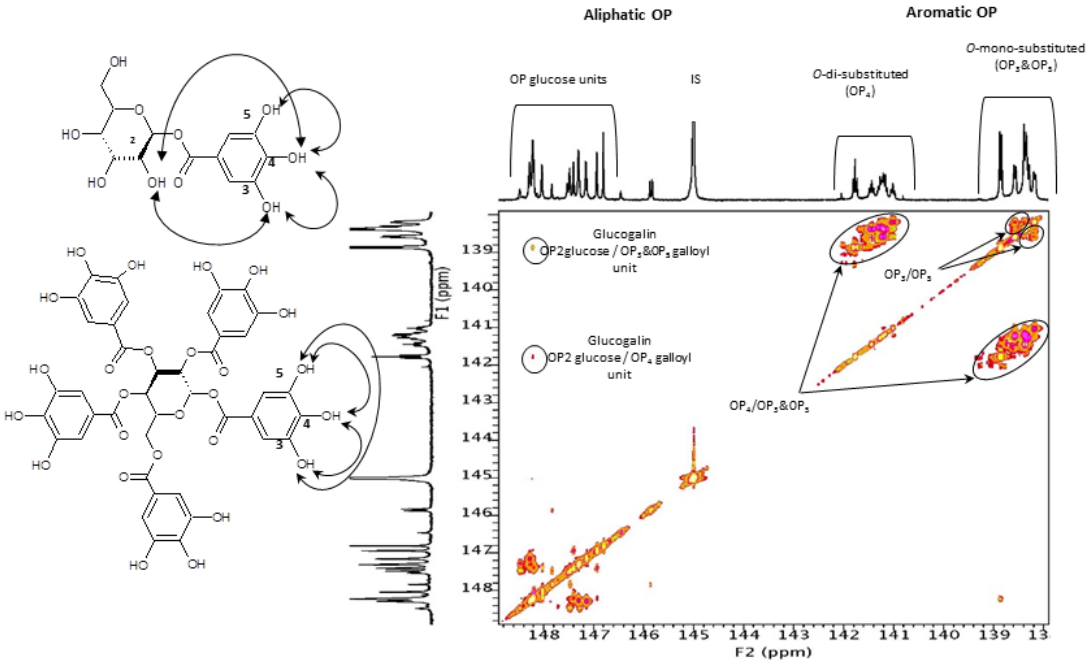


Figure 2: ^{31}P 1D and 2D TOCSY NMR spectra of a mixture of glucose, β-glucogalin and PGG in pyridine-d5:$CDCl_3$ after phosphitylation. The numbering system of phenolic compounds and the TOCSY correlations are shown on the left.

The aromatic region of the ^{31}P NMR spectra of gallnut tannin extracts after phosphitylation reaction was much more crowded (Figure 3 A). Signals were attributed using both chemical shifts and ^{31}P TOCSY correlations (Figure 3 B). Terminal unit OP3 / OP5 and OP4 were easily attributed since their chemical shifts are similar to those of gallotannin standards (between ~137.8 to 138.6 ppm and ~140.7-141.55 ppm respectively) and TOCSY correlations are observed between non-equivalent OP3 / OP5 and between OP3 / OP5 and OP4. Additional more deshielded signals found

in the aromatic region from ~138.6 to 140.7 ppm and from 141.55 to 142.4 ppm were unambiguously attributed respectively to OP3 /OP5 and OP4 of internal units, in accordance to the presence of correlations between these two groups of signals. It is worth noting that the integral area of internal OP3/OP5 signals exceeded that of internal OP4 signals, meaning that some of the former correspond to OP3/OP5 of internal units substituted in the 4 position. In the chemical shift range of carboxylic groups, three peaks were observed in all samples. Beside the carboxylic peak of free gallic acid easily identified at 134.45 ppm, two other peaks appeared at 135.02 and 135.07 ppm. A fourth peak was also present in TAN B1 spectrum at 134.73 ppm (Figure 3 A). The peaks arising from the aliphatic groups were very different depending on the tannin extract. The ^{31}P spectrum of TAN A exhibited mainly a broad and intense phosphitylated aliphatic peak whereas the peaks in this region in TAN B1 and B2 spectra were numerous and sharper, with higher intensity in TAN B1 than in TAN B2.

The signal areas were integrated to get more insight on the composition of the tannin extracts (Table 2). Free gallic acid concentrations were about 0.1, 0.36 and 0.04 mmol/g (17, 61, 7 mg/g) in TAN A, TAN B1, and TAN B2, respectively, consistent with the values determined by UPLC-DAD (23, 73, and 9 mg/g). The other carboxyl signals at 135.7 and 135.02 ppm integrated for lower concentrations (about 0.01 to 0.04 mmol/g) likely arise from the di and tri-gallic acids also detected by UPLC-DAD-MS. Concentrations calculated for trigallic acid (8.5, 7.2 and 4.2 mg/g, in TAN A, B1 and B2, respectively) are consistent with those determined by HPLC (Table 1). Concentrations calculated for digallic acid (14.2, 13.5 and 7.4 mg/g, in TAN A, B1 and B2, respectively) are lower than those determined by HPLC, due to coelution of digallic acid with digalloylglucose (Table 1).

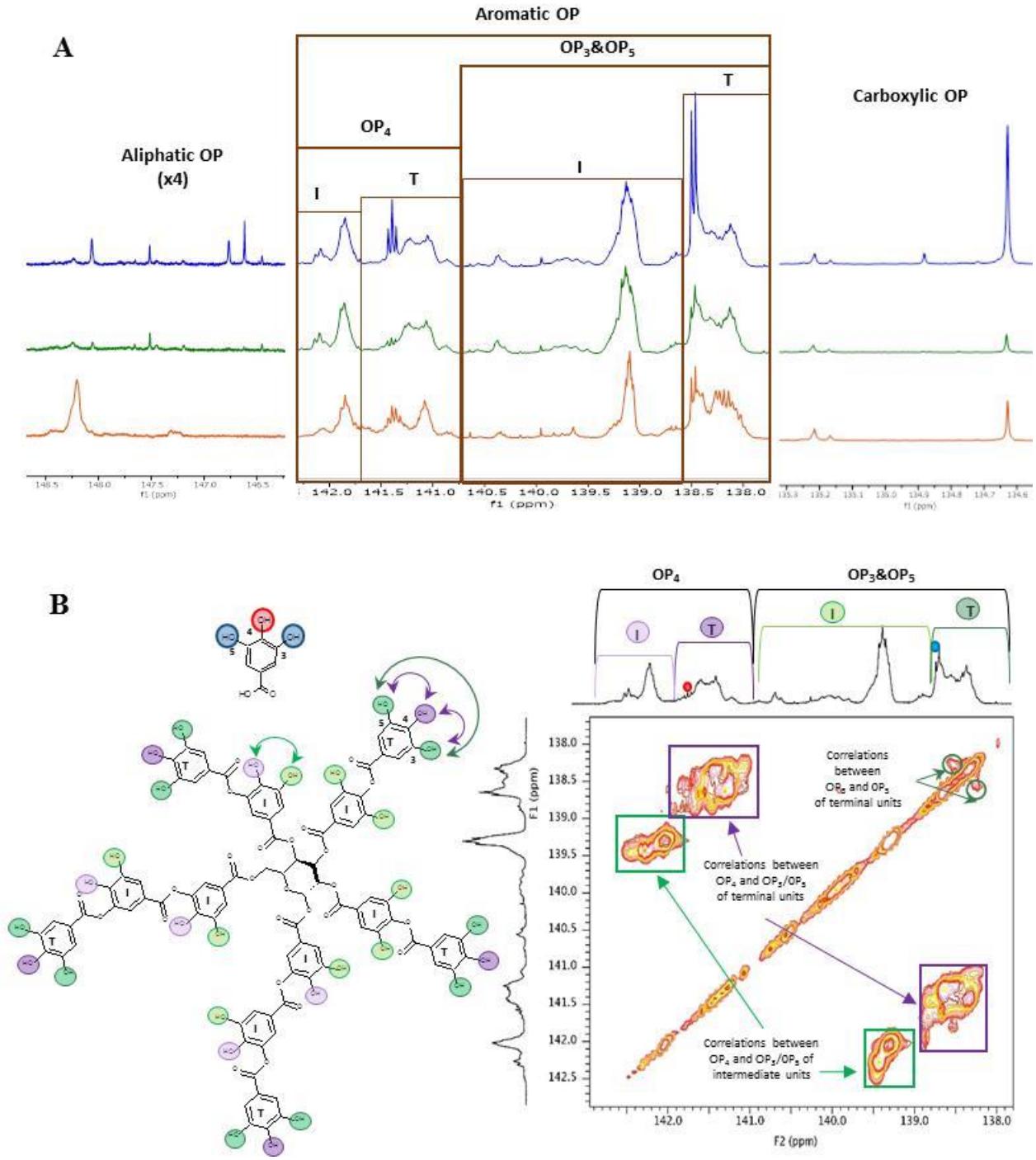


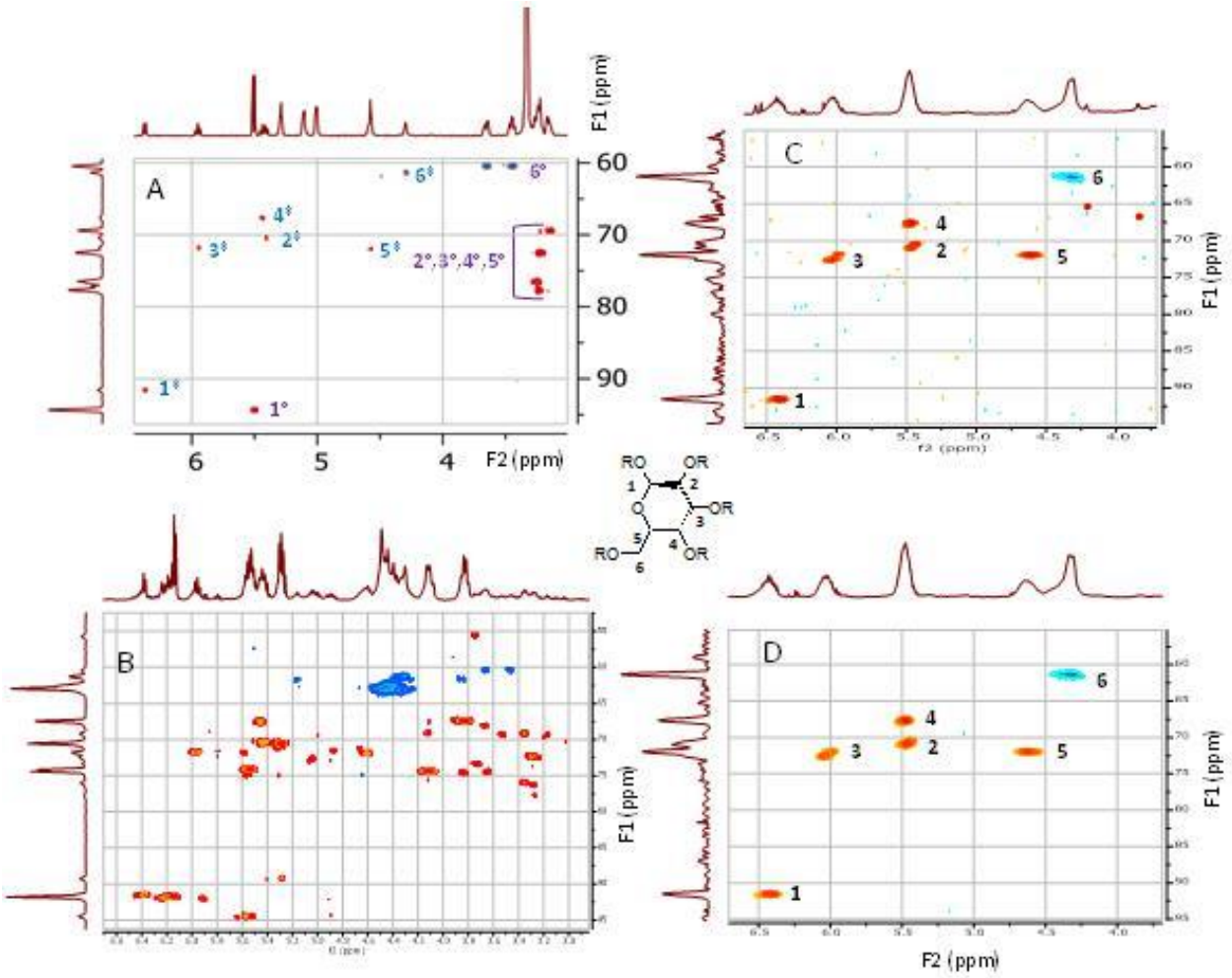
Figure 3: (A) ^{31}P 1D NMR spectra of TAN B1 (blue), TAN B2 (green) and TAN A (red) in pyridine-d5: CDCl3 after phosphitylation. (I= Internal units, T=Terminal units). (B) ^{31}P 1D and 2D TOCSY NMR spectra of TAN B2 in pyridine-d5: $CDCl_3$ after phosphitylation, region of aromatic signals.

The numbering system of polyphenol compounds and the TOCSY correlations are shown on the left.

Terminal and internal units were quantified from the integration of phosphitylated phenolic signals. The molarity of terminal units was calculated from the sum of OP4 and OP3 & OP5 signal integrations divided by three whereas that of intermediate units was determined from the sum of OP4 and OP3 & OP5 signal integrations divided by two to take into account the presence of both *para*- and *meta*-depside bonds. The molarity of *meta*-depside bonds is directly obtained from the integration value of the internal unit OP4 signals and that of *para*-depside bonds is deduced from the difference between the concentration of intermediate units and that of *meta*-depside bond internal units. The percentage of *para*-depside bonds calculated for TAN B samples was ~39% whereas it was only about 21% for TAN A. Another difference between these extracts was the ratio between the internal and the terminal units of the poly-galloylglucose molecules which was close to 1.1 for TAN A whereas it was about 1.5 for both TAN B extracts. This means that the gallotannin molecules of TAN A were mainly constituted of glucose substituted by chains of two galloyl units, whereas the TAN B extracts contain longer galloyl chains. The molarity calculated from the aliphatic signals of TAN B extracts was very low (less or equal to 0.01mmol/g), meaning that glucotannin molecules of these two tannins were mainly constituted of penta-O-substituted glucose moieties. In contrast, the intense broad signal observed in TAN A spectrum integrates for about 0.5 mmol/g, indicating that the glucose hydroxyls were only partially substituted by galloyl units. The ratio between the poly-galloylglucose terminal units and the hydroxyl glucose concentrations was close to four, confirming that the gallotannin molecules of TAN A were mainly constituted by tetra-O-substituted glucose moieties, in agreement with the UPLC-DAD-MS data.

The average number of gallic units per glucose residue was also calculated from the sum of the total unit concentration divided by the concentration of the terminal units, then multiplied by the number of substituted glucose hydroxyls, that was four in the case of TAN A and five for TAN B1 and B2. It provided a mean number of galloyl units per glucose moiety of 8.5, 12.2 and 12.4 for TAN A, TAN B1 and TAN B2, respectively, consistent with chains of two galloyl units in TAN A and longer in TAN B. The total concentrations of gallotannins calculated from the ^{31}P NMR data were 725, 749 and 780 mg/g in TANA, TAN B1, and TANB2, respectively, amounting to 765, 829, and 799 mg/g, when taking into account gallic, digallic and trigallic acids.

1D ^{1}H and 2D ^{1}H/^{13}C analysis performed on the three samples confirmed the results obtained by ^{31}P NMR. 2D ^{1}H/^{13}C HSQC spectra of the three tannin extracts showed typical chemical shifts of galloyl and glucose units. As expected, ^{1}H and ^{13}C aromatic signals appear in more deshielded regions than those of PGG (not shown) which indicate the presence of galloyl units linked by depside bonds.[25] Aliphatic signals in the HSQC spectra of TAN B1 and B2 are characteristic of glucoside moieties substituted at their five hydroxyl positions as observed for PGG (Figure 4 C, D).



411

412 *Figure 4. Aliphatic region of 1D ^{1}H and 2D HSQC $^{1}H/^{13}C$ spectra. (A) Standard mixture of β-*
413 *glucogallin (°) and PGG (*); (B) Tan A; (C) TAN B1; (D) TAN B2. Attributed protons and carbons*
414 *of glucoside moieties are noted in the spectra according to the glucoside numbering scheme.*

415

416 In contrast, aliphatic signals of TAN A (Figure 4 B) are much more numerous and dispersed in
417 accordance with the presence of glucose core units partly and non-uniformly substituted. In the
418 2D $^{1}H/^{13}C$ HMBC spectra (not shown), most of the carbonyl signals are located in the region
419 characteristic of ester groups (~160-166ppm). Four COOH groups are observed between 166.5-
420 168ppm in TAN B1 spectrum and three in both TAN B2 and TAN A spectra in accordance with

signals detected in the ^{31}P spectra. The COOH signal of free gallic acid is easily identified by comparison with a standard. Due to the low intensity of the other COOH peaks and the spectra crowding, it was not possible to get more insights upon the structure of the molecules corresponding to these minor peaks.

- GC analysis of neutral sugars

Glucose, arising from galloylglucose derivatives, was the only monosaccharide detected by GC analysis in all samples. Its concentration was evaluated at 47.3 ± 2.2, 23.4 ± 5.2, and 16.1 ± 1.4 mg/g respectively in TAN A, TAN B1 and TAN B2.

Characterization of the Molecular Weight Distribution of Tannins

- Gel Permeation Chromatography

Gel permeation chromatography performed on the three gallotannin extracts showed the presence of three peaks detected at 280 nm (Figure 5). The last one, eluted at 21.5 min, corresponds to the retention time of gallic acid. Its peak area was higher in TAN B1 than in the two other tannin extracts, as observed by LC-DAD-MS and NMR. The second peak, also present in small amount in the three extracts, coeluted with β-glucogallin (20.3 min) and thus likely corresponds to digallic acid as β-glucogallin was detected by LC-MS only in trace amounts. The larger peak detected in TAN A showed the same retention time as pentagalloylglucose (18.5 min), while the major peak detected in TAN B1 and B2, was eluted much earlier (17.7 min), thus showing a molecular size larger than that of pentagalloylglucose. These GPC profiles are similar to those reported earlier for Turkish and Chinese gall extracts.[29] The peak intensity was slightly higher in B2 than in B1 but the shapes looked similar, suggesting similar size distributions.

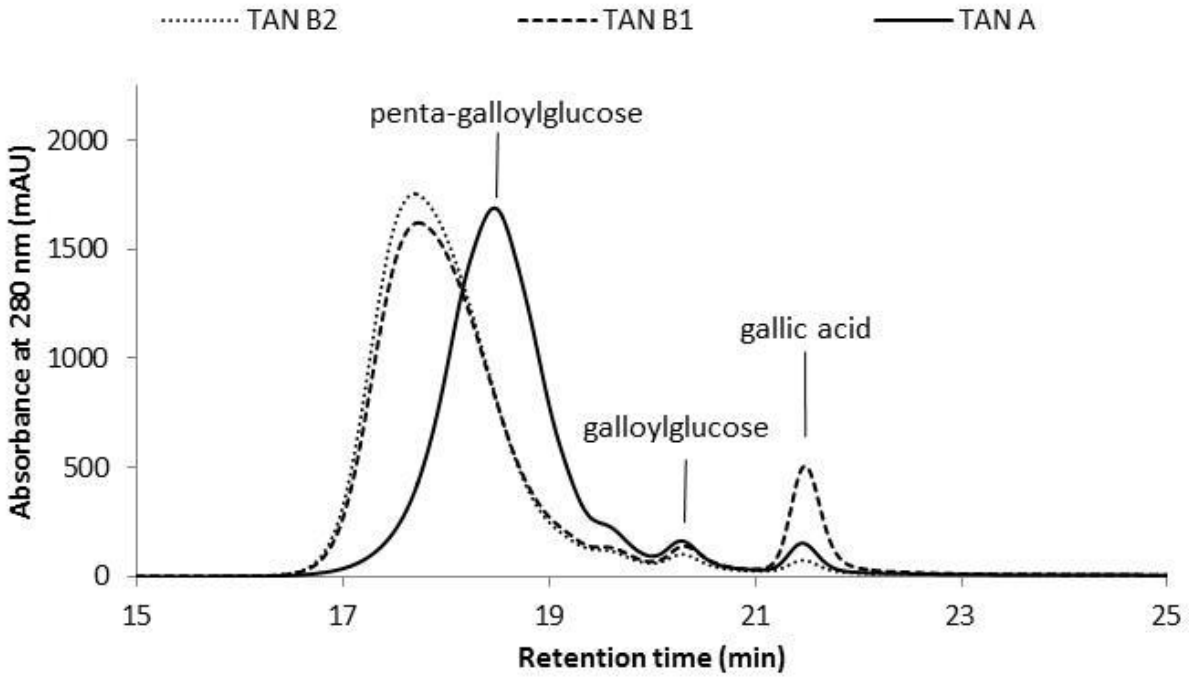


Figure 5 Molecular size distribution of gallnut tannin extracts (TAN A and TAN B2 vs. B1 lots at 2.5 g/L) detected at 280 nm by gel permeation chromatography. The retention times of gallic acid, β-glucogallin, and pentagalloylglucose standards are reported on the chromatogram.

- ^{1}H DOSY NMR

2D ^{1}H DOSY spectra which separate NMR signals of compounds in mixtures according to their diffusion coefficients have been performed to get additional information on the polydispersity of the molecules in the tannin extracts. These data also allowed calculation of the molecular weight of unknown molecules using a calibration curve obtained from gallotannin standards (gallic acid, β-glucogallin, pentagalloylglucose) added to the extracts in solution. If peaks are intense and resolved enough, molecular weight of small gallotannin molecules (from ~170 to ~940) can be calculated with accuracy since the correlation coefficient of the calibration curve obtained from

the logarithm of diffusion coefficients (log D) and the logarithm of the molecular weights (log PM) of the standards is > 0.9 (Figure 6 A). Molecular weights of larger polymers were also calculated from this regression line but with uncertainties as no reference molecule of higher molecular weight gallotannins was available.[43] Molecular weights evaluated from the diffusion coefficients (Figure 6 B, C, D) ranged broadly from 500 (~digalloylglucose) to 1860 (~11 galloylglucose) for all tannin extracts.

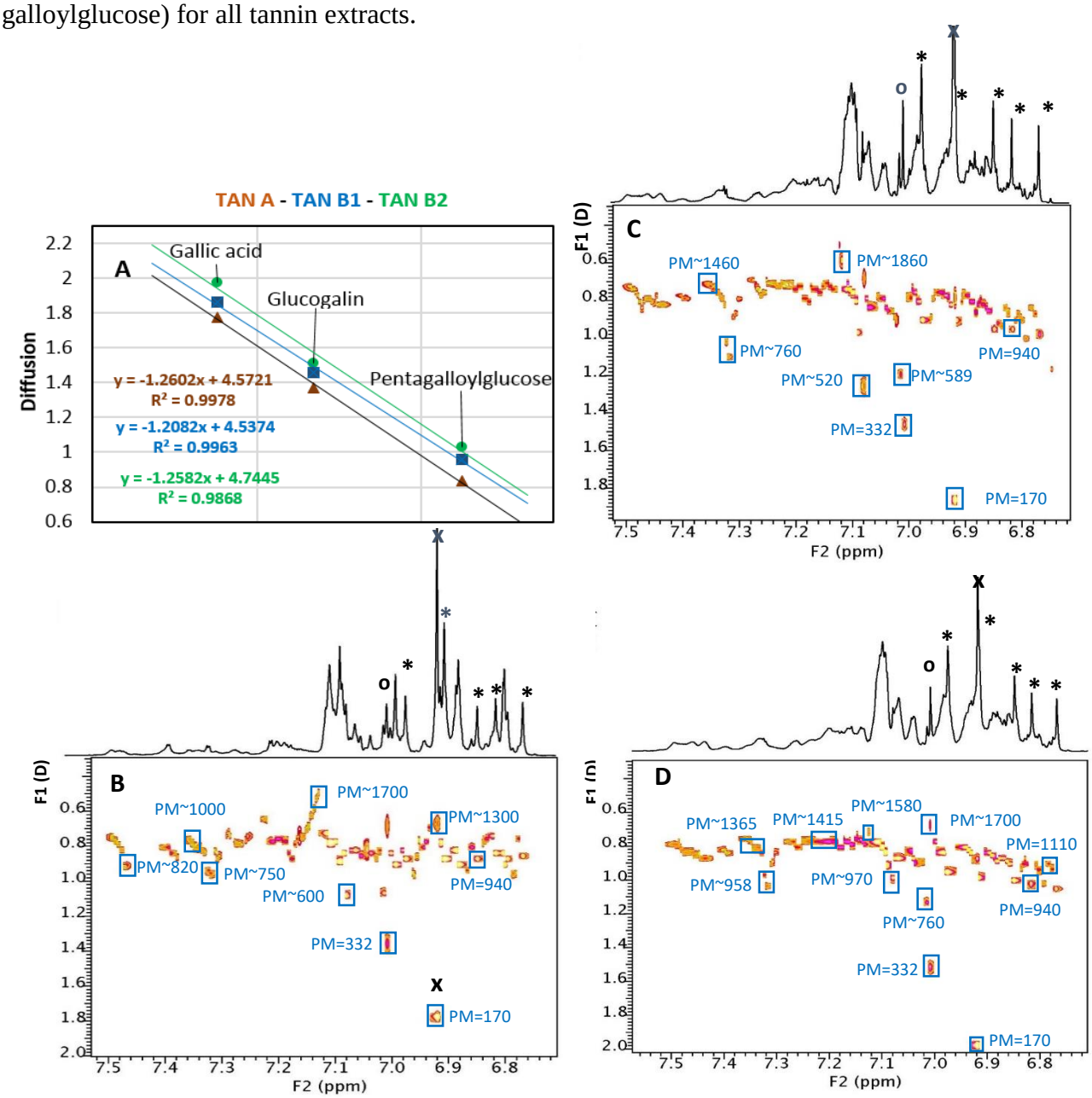


*Figure 6. A, calibration curves obtained from gallotannin standards (gallic acid, glucogallin and pentagalloylglucose) added to the extracts in solution. Aromatic region of ^{1}H 2D DOSY NMR spectra of TAN A (B), TAN B1 (C) and TAN B2 (D) extracts with added gallotannin standards in DMSO d6 (x: gallic acid, o : β-glucogallin, * : pentagalloylglucose signals).*

Taken together, these data indicate that the three gallnut tannin extracts contain similar levels of gallic acid derivatives but have different compositions, as summarized in Table 3. The concentrations determined by UPLC-DAD using gallic acid as the calibration standard were higher than 1000 mg/g in all three samples and thus overestimated. However, when expressed as pentagalloylglucose equivalent, the concentrations determined by UPLC-DAD were only slightly higher than those determined by ^{31}P NMR spectroscopy after phosphitylation, and very good congruence was obtained between the two methods both for the total gallic acid derivatives and for gallotannins.

The large signal corresponding to an aliphatic OH group in the ^{31}P NMR spectrum as well as UPLC-DAD-MS data showing the presence of tetragalloylglucose and its derivatives indicate that TAN A is mostly based on a tetra-O-substituted glucose core while TAN B1 and B2 are pentagalloylglucose derivatives. These structural characteristics have been described earlier for Turkish and Chinese gallnut extracts, respectively.[29,44] LC-MS, GPC, and DOSY NMR analysis enabled detection of a large range of molecular weights in all tannin samples. GPC analysis showed that polymers were larger in TAN B1 and B2 than in TAN A, in agreement with the average molecular weight calculated from integration of the ^{31}P NMR data (average number of galloyl groups evaluated at of 8.5, 12.2 and 12.4 respectively, for TAN A, TAN B1, and TAN B2). The average chain length determined for TANA confirms the results obtained earlier for Turkish

493 gallnut extracts.[29] To the best of our knowledge, that of Chinese tannins is reported here for the
494 first time. The average degrees of galloylation of 12.2 and 12.4 determined for TAN B1 and TAN
495 B2, as well as the detection by UPLC-MS of galloylglucose structures containing up to 12, 18 and
496 17 galloyl groups, respectively in TAN A, TAN B1, and TAN B2, indicate the presence of long
497 polygalloyl chains in the tannin structure, as proposed earlier.[14,15] This is also supported by
498 previous work showing the presence of gallic polymers up to the octamer,[12] likely arising from
499 cleavage of the glucose ester bonds of the original galloylglucose structure. Nevertheless, higher
500 molecular weight tannins were poorly detected by UPLC-ESI-MS, as shown earlier for
501 proanthocyanidin polymers[33] and by DOSY NMR. Sugar analysis by GC after hydrolysis and
502 derivatization also indicated that TAN A contains higher glucose contents than TAN B, consistent
503 with the lower degree of galloylation shown by the other methods. However, values obtained using
504 this method were not coherent with those determined by NMR. Reaction yields of 72% and 34%
505 were obtained respectively for β-glucogallin and pentagalloylglucose standards, confirming that
506 the method is not suitable for glucose ester derivatives and especially polymers.

507 Some differences in the composition of TAN B1 and B2 were also demonstrated, suggesting
508 differences in the preparation of these two batches. Thus, TAN B1 contained significantly higher
509 concentration of gallic, digallic, and trigallic acids and lower amounts of gallotannins than TAN
510 B2. It also contained an unknown low molecular weight aliphatic compound detected by ^{1}H DOSY
511 NMR (data not shown). The UPLC-DAD-MS data also showed significantly higher intensities of
512 tri- and tetra-galloylglucose and of mass signals attributed to minor compounds such as
513 galloylquinic derivatives in TAN B1.

514 Finally, the structures appeared more complex than expected. In particular, NMR provided
515 evidence of para-substituted gallic acid units in addition to the well-established meta-depside

structures. These structures were more common in TAN B than in TAN A, which may be related to their longer galloyl chains. Detection of methyl galloylglucose derivatives by UPLC-MS analysis after methanolysis indicated that some of the glucose units were linked to gallic acid through ether rather than ester bonds, as proposed earlier.[14,36] Different isomers of these structures were predominant after methanolysis of TAN A and TAN B, reflecting further structural differences.

Antioxidant Capacity

The antioxidant capacity of the three extracts was measured by the ABTS method. The percentage of inhibition was significantly lower for TAN B1 at 35.82 ± 0.63 compared to TAN A and B2 (39.95 ± 0.77 and 38.84 ± 0.63, respectively). This might be related to differences in the proportions of gallotannins and lower molecular weight compounds such as gallic acid, reflecting some gallotannin degradation during the tannin production process in TAN B1. However, the large structural differences observed between TAN A and TAN B2 did not impact their antioxidant properties measured by the ABTS method, which seem mostly related to the total tannin content.

The three tannin extracts analyzed in the present study showed much higher tannin contents than some other commercial gallnut extracts described in the literature (100% vs 46 to 51% in Vignault et al,[6] as determined by UV spectrophotometry). The presence of other compounds in the gallnut tannin extracts may explain the discrepancies between tannin composition and antioxidant properties reported by these authors.[6].

Thus, comparison of our results with the literature data confirm that commercial gallnut extracts exhibit large variations in their tannin content and composition and emphasize the need for detailed characterization of these products in the frame of quality control and selection for specific

applications. The large structural differences evidenced in our study had very little impact on radical scavenging activity which seems primarily driven by gallotannin concentration. However, their impact on other properties, including organoleptic properties and color stabilizing effect remain to be investigated.

ACKNOWLEDGEMENTS

The authors would like to thank Stéphanie Carrillo for technical assistance in the GPC analysis.

FUNDING SOURCES

Funding from FEDER and Région Grand-Est, in the frame of "Programme opérationnel FEDER/FSE/IEJ 2014-2020" is gratefully acknowledged.

Supporting information

Zooms of the negative ion mode ESI MS spectrum recorded after UPLC-DAD-MS analysis of TAN B2 showing the region 500-1500 m/z and mass spectra of the doubly charged 16G-glc, 17G-glc and triply charged 18G-glc.

Negative ion mode ESI MS spectra (90-2000 m/z) recorded after UPLC-DAD-MS analysis of TAN A, TAN B1, and TAN B2, after methanolysis. UPLC profile at 280 and extracted ion chromatograms at m/z 497 (Me-2G-glc), 649 (Me-3G-glc), and 801 (Me-4G-glc).

REFERENCES

(1) Haslam, E.; Cai, Y. Plant Polyphenols (Vegetable Tannins): Gallic Acid Metabolism. *Nat. Prod. Rep.* **1994**, *11* (0), 41–66.

(2) Quideau, S.; Feldman, K. S. Ellagitannin Chemistry. *Chem. Rev.* **1996**, *96* (1), 475–504.

(3) Arapitsas, P. Hydrolyzable Tannin Analysis in Food. *Food Chem.* **2012**, *135* (3), 1708–1717.

(4) Harbertson, J. F.; Kennedy, J. A.; Adams, D. O. Tannin in Skins and Seeds of Cabernet Sauvignon, Syrah, and Pinot Noir Berries during Ripening. *Am. J. Enol. Vitic.* **2002**, *53* (1), 54–59.

(5) Sarneckis, C. J.; Dambergs, R. G.; Jones, P.; Mercurio, M.; Herderich, M. J.; Smith, P. A. Quantification of Condensed Tannins by Precipitation with Methyl Cellulose: Development and Validation of an Optimised Tool for Grape and Wine Analysis. *Aust. J. Grape Wine Res.* **2006**, *12* (1), 39–49.

(6) Vignault, A.; González-Centeno, M. R.; Pascual, O.; Gombau, J.; Jourdes, M.; Moine, V.; Iturmendi, N.; Canals, J. M.; Zamora, F.; Teissedre, P.-L. Chemical Characterization, Antioxidant Properties and Oxygen Consumption Rate of 36 Commercial Oenological Tannins in a Model Wine Solution. *Food Chem.* **2018**, *268*, 210–219.

(7) Kennedy, J. A.; Ferrier, J.; Harbertson, J. F.; Peyrot des Gachons, C. Analysis of Tannins in Red Wine Using Multiple Methods: Correlation with Perceived Astringency. *Am. J. Enol. Vitic.* **2006**, *57* (4), 481–485.

(8) Mercurio, M. D.; Smith, P. A. Tannin Quantification in Red Grapes and Wine: Comparison of Polysaccharide- and Protein-Based Tannin Precipitation Techniques and Their Ability to Model Wine Astringency. *J. Agric. Food Chem.* **2008**, *56* (14), 5528–5537.

(9) de Freitas, V.; Mateus, N. Protein/Polyphenol Interactions: Past and Present Contributions. Mechanisms of Astringency Perception. *Curr. Org. Chem.* **2012**, *16* (6), 724–746.

(10) Scollary, G. R.; Pasti, G.; Kallay, M.; Blackman, J.; Clark, A. C. Astringency Response of Red Wines: Potential Role of Molecular Assembly. *Trends Food Sci. Technol.* **2012**, *27* (1), 25–36.

(11) Soares, S.; Brandao, E.; Mateus, N.; de Freitas, V. Sensorial Properties of Red Wine Polyphenols: Astringency and Bitterness. *Crit. Rev. Food Sci. Nutr.* **2017**, *57* (5), 937–948.

(12) Moutounet, M.; Leaute, B.; Delbos, C.; Doco, T.; Meudec, E.; Souquet, J.-M. Analyse de la composition de tanins oenologiques. *Rev. Fr. Oenologie* **2004**, No. 208, 22–27.

(13) Lei, Z.; Jervis, J.; Helm, R. F. Use of Methanolysis for the Determination of Total Ellagic and Gallic Acid Contents of Wood and Food Products. *J. Agric. Food Chem.* **2001**, *49* (3), 1165–1168.

(14) Armitage, R.; Bayliss, G. S.; Gramshaw, J. W.; Haslam, E.; Haworth, R. D.; Jones, K.; Rogers, H. J.; Searle, T. 360. Gallotannins. Part III. The Constitution of Chinese, Turkish, Sumach, and Tara Tannins. *J. Chem. Soc.* **1961**, 1842–1853.

(15) Haslam, E.; Haworth, R. D.; Mills, S. D.; Rogers, H. J.; Armitage, R.; Searle, T. 359. Gallotannins. Part II. Some Esters and Depsides of Gallic Acid. *J. Chem. Soc.* **1961**, 1836–1842.

(16) Beasley, T.; Ziegler, H.; Bell, A. Determination and Characterization of Gallotannin by High-Performance Liquid-Chromatography. *Anal. Chem.* **1977**, *49* (2), 238–243.

(17) Nishizawa, M.; Yamagishi, T.; Nonaka, G.; Nishioka, I. Tannins and Related-Compounds .5. Isolation and Characterization of Polygalloylglucoses from Chinese Gallotannin. *J. Chem. Soc.-Perkin Trans. 1* **1982**, No. 12, 2963–2968.

(18) Verzele, M.; Delahaye, P. Analysis of Tannic Acids by High-Performance Liquid-Chromatography. *J. Chromatogr.* **1983**, *268* (3), 469–476.

(19) Armitage, R.; Haworth, R.; Haslame; Searle, T. Gallotannins .6. Turkish Gallotannin. *J. Chem. Soc.* **1962**, No. SEP, 3808-.

(20) Nishizawa, M.; Nishioka, I.; Yamagishi, T.; Nonaka, G. Tannins and Related-Compounds .9. Isolation and Characterization of Polygalloylglucoses from Turkish Galls (Quercus-Infectoria). *J. Chem. Soc.-Perkin Trans. 1* **1983**, No. 5, 961–965.

(21) Verzele, M.; Delahaye, P.; Vandamme, F. Determination of the Tanning Capacity of Tannic Acids by High-Performance Liquid-Chromatography. *J. Chromatogr.* **1986**, *362* (3), 363–374.

(22) Hagerman, A.; Robbins, C.; Weerasuriya, Y.; Wilson, T.; Mcarthur, C. Tannin Chemistry in Relation to Digestion. *J. Range Manag.* **1992**, *45* (1), 57–62.

(23) Potier, P.; Maccario, V.; Giudicelli, M.-B.; Queneau, Y.; Dangles, O. Gallic Esters of Sucrose as a New Class of Antioxidants. *Tetrahedron Lett.* **1999**, *40* (17), 3387–3390.

(24) Karas, D.; Ulrichová, J.; Valentová, K. Galloylation of Polyphenols Alters Their Biological Activity. *Food Chem. Toxicol.* **2017**, *105*, 223–240.

(25) Luo, F.; Fu, Y.; Xiang, Y.; Yan, S.; Hu, G.; Huang, X.; Huang, G.; Sun, C.; Li, X.; Chen, K. Identification and Quantification of Gallotannins in Mango (Mangifera Indica L.) Kernel and Peel and Their Antiproliferative Activities. *J. Funct. Foods* **2014**, *8*, 282–291.

(26) Albersheim, P.; Nevins, D. J.; English, P. D.; Karr, A. A Method for the Analysis of Sugars in Plant Cell-Wall Polysaccharides by Gas-Liquid Chromatography. *Carbohydr. Res.* **1967**, *5* (3), 340–345.

(27) Kennedy, J. A.; Taylor, A. W. Analysis of Proanthocyanidins by High-Performance Gel Permeation Chromatography. *J. Chromatogr. A* **2003**, *995* (1), 99–107.

(28) Melone, F.; Saladino, R.; Lange, H.; Crestini, C. Tannin Structural Elucidation and Quantitative P-31 NMR Analysis. 1. Model Compounds. *J. Agric. Food Chem.* **2013**, *61* (39), 9307–9315.

(29) Melone, F.; Saladino, R.; Lange, H.; Crestini, C. Tannin Structural Elucidation and Quantitative 31P NMR Analysis. 2. Hydrolyzable Tannins and Proanthocyanidins. *J. Agric. Food Chem.* **2013**, *61* (39), 9316–9324.

(30) Catarina Guedes, A.; Amaro, H. M.; Giao, M. S.; Xavier Malcata, F. Optimization of ABTS Radical Cation Assay Specifically for Determination of Antioxidant Capacity of Intracellular Extracts of Microalgae and Cyanobacteria. *Food Chem.* **2013**, *138* (1), 638–643.

(31) Pizzi, A.; Pasch, H.; Rode, K.; Giovando, S. Polymer Structure of Commercial Hydrolyzable Tannins by Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry. *J. Appl. Polym. Sci.* **2009**, *113* (6), 3847–3859.

(32) Mane, C.; Sommerer, N.; Yalcin, T.; Cheynier, V.; Cole, R. B.; Fulcrand, H. Assessment of the Molecular Weight Distribution of Tannin Fractions through MALDI-TOF MS Analysis of Protein-Tannin Complexes. *Anal. Chem.* **2007**, *79* (6), 2239–2248.

(33) Fulcrand, H.; Mané, C.; Preys, S.; Mazerolles, G.; Bouchut, C.; Mazauric, J.-P.; Souquet, J.-M.; Meudec, E.; Li, Y.; Cole, R. B.; et al. Direct Mass Spectrometry Approaches to Characterize Polyphenol Composition of Complex Samples. *Phytochemistry* **2008**, *69* (18), 3131–3138.

(34) Yanagida, A.; Kanda, T.; Shoji, T.; Ohnishi-Kameyama, M.; Nagata, T. Fractionation of Apple Procyanidins by Size-Exclusion Chromatography. *J. Chromatogr. A* **1999**, *855* (1), 181–190.

(35) Arapitsas, P.; Menichetti, S.; Vincieri, F. F.; Romani, A. Hydrolyzable Tannins with the Hexahydroxydiphenoyl Unit and the M-Depsidic Link: HPLC-DAD-MS Identification and Model Synthesis. *J. Agric. Food Chem.* **2007**, *55* (1), 48–55.

(36) Verzele, M.; Delahaye, P.; Dewaele, C. A New Structural Formula for the Polygalloyl Glucose Tannic Acids. *Bull. Soc. Chim. Belg.* **1984**, *93* (2), 151–152.

(37) Balakshin, M.; Capanema, E. On the Quantification of Lignin Hydroxyl Groups with P-31 and C-13 Nmr Spectroscopy. *J. Wood Chem. Technol.* **2015**, *35* (3), 220–237.

(38) Ben, H.; Chen, X.; Han, G.; Shao, Y.; Jiang, W.; Pu, Y.; Ragauskas, A. J. Characterization of Whole Biomasses in Pyridine Based Ionic Liquid at Low Temperature by P-31 NMR: An Approach to Quantitatively Measure Hydroxyl Groups in Biomass As Their Original Structures. *Front. Energy Res.* **2018**, *6*, UNSP 13.

(39) Crestini, C.; Lange, H.; Bianchetti, G. Detailed Chemical Composition of Condensed Tannins via Quantitative P-31 NMR and HSQC Analyses: Acacia Catechu, Schinopsis Balansae, and Acacia Mearnsii. *J. Nat. Prod.* **2016**, *79* (9), 2287–2295.

(40) Dais, P.; Spyros, A. P-31 NMR Spectroscopy in the Quality Control and Authentication of Extra-Virgin Olive Oil: A Review of Recent Progress. *Magn. Reson. Chem.* **2007**, *45* (5), 367–377.

(41) Filpponen, I.; Argyropoulos, D. S. Determination of Cellulose Reactivity by Using Phosphitylation and Quantitative P-31 NMR Spectroscopy. *Ind. Eng. Chem. Res.* **2008**, *47* (22), 8906–8910.

(42) Meng, X.; Crestini, C.; Ben, H.; Hao, N.; Pu, Y.; Ragauskas, A. J.; Argyropoulos, D. S. Determination of Hydroxyl Groups in Biorefinery Resources via Quantitative (PNMR)-P-31-N- Spectroscopy. *Nat. Protoc.* **2019**, *14* (9), 2627–2647.

(43) Neufeld, R.; Stalke, D. Accurate Molecular Weight Determination of Small Molecules via DOSY-NMR by Using External Calibration Curves with Normalized Diffusion Coefficients. *Chem. Sci.* **2015**, *6* (6), 3354–3364.

(44) Buziashvili, I. Sh.; Komissarenko, N. F.; Kovalev, I. P.; Gordienko, V. G.; Kolesnikov, D. G. The Structure of Gallotannins. *Chem. Nat. Compd.* **1973**, *9* (6), 752–755.

665 FIGURE CAPTIONS

666 Figure 2. UPLC trace at 280 nm of TAN A (blue), TAN B1 (black) and TAN B2 (red) (A).
667 Negative ion mode ESI MS spectra (150-1500 m/z) recorded after UPLC-DAD-MS analysis of
668 TAN A (B), TAN B1 (C) and TAN B2 (D). G-Glc: galloyl-glucose

669 Figure 2: ^{31}P 1D and 2D TOCSY NMR spectra of a mixture of glucose, β-glucogalin and PGG in
670 pyridine-d5:$CDCl_3$ after phosphitylation. The numbering system of phenolic compounds and the
671 TOCSY correlations are shown on the left.

672 Figure 3: (A) ^{31}P 1D NMR spectra of TAN B1 (blue), TAN B2 (green) and TAN A (red) in
673 pyridine-d5: CDCl3 after phosphitylation. (I= Internal units, T=Terminal units). (B) ^{31}P 1D and
674 2D TOCSY NMR spectra of TAN B2 in pyridine-d5: $CDCl_3$ after phosphitylation, region of
675 aromatic signals. The numbering system of polyphenol compounds and the TOCSY correlations
676 are shown on the left.

677 Figure 4. Aliphatic region of 1D ^{1}H and 2D HSQC ^{1}H/^{13}C spectra. (A) Standard mixture of β-
678 glucogallin (°) and PGG (*); (B) Tan A; (C) TAN B1; (D) TAN B2. Attributed protons and carbons
679 of glucoside moieties are noted in the spectra according to the glucoside numbering scheme.

680 Figure 5. Molecular size distribution of gallnut tannin extracts (TAN A and TAN B2 vs. B1 lots
681 at 2.5 g/L) detected at 280 nm by gel permeation chromatography. The retention times of gallic
682 acid, β-glucogallin, and pentagalloylglucose standards are reported on the chromatogram.

683 Figure 6. A, calibration curves obtained from gallotannin standards (gallic acid, glucogallin and
684 pentagalloylglucose) added to the extracts in solution. Aromatic region of ^{1}H 2D DOSY NMR
685 spectra of TAN A (B), TAN B1 (C) and TAN B2 (D) extracts with added gallotannin standards in
686 DMSO d6 (x: gallic acid, o: β-glucogallin, *: pentagalloylglucose signals).

687

688 *Table 1. Concentration of gallic acid, digallic acid, trigallic acid, β-glucogallin and poly-*

689 *galloylglucose expressed in mg/g gallic acid equivalent.*

	gallic acid	digallic acid[1]	trigallic acid	β-glucogallin	gallotannins[2]
TAN A	22.92 ± 0.67 [b]	26.92 ± 0.88 [c]	9.83 ± 0.26 [c]	0.71 ± 0.02 [c]	987.38 ± 38.27 [a]
TAN B1	73.27 ± 0.90 [c]	20.27 ± 0.30 [b]	7.17 ± 0.09 [b]	0.12 ± 0.01 [a]	1027.69 ± 18.65 [a]
TAN B2	8.91 ± 0.10 [a]	12.54 ± 0.26 [a]	4.73 ± 0.14 [a]	0.17 ± 0.01 [b]	1094.17 ± 10.16 [b]

690 *[1]two isomers, the least polar coeluted with digalloylglucose. [2]estimated from total peak area at*

691 *280 nm minus peak areas of gallic acid, digallic acid, trigallic acid, and β-glucogallin.*

692 *Different superscripts indicate statistical differences between samples.*

693

694 *Table 2. Integral values and compound contents of gallnut tannin extracts calculated from ^{31}P 1D*

695 *NMR spectra after phosphitylation*

	TAN A	TAN B1	TAN B2	Integration ranges in ppm
Integral values in mmol/g				
(OP3+OP5)T	4.30 ± 0.22	4.49 ± 0.19	3.80 ± 0.16	138.71-138.00
(OP4)T	2.12 ± 0.10	2.24 ± 0.12	2.09 ± 0.11	141.75-140.88
(OP3+OP5)I	2.83 ± 0.14	3.87 ± 0.16	4.00 ± 0.01	140.87-138.72
(OP4)I	1.83 ± 0.03	1.70 ± 0.09	1.74 ± 0.05	142.57-141.75
Concentrations in mmol/g				
Gallic acid	0.10 ± 0.01	0.35 ± 0.02	0.04 ± 0.00	134.70-134.54
Digallic acid	0.044 ± 0.002	0.042 ± 0.002	0.023 ± 0.002	135.09-135.05
Trigallic acid	0.018 ± 0.002	0.015 ± 0.001	0.009 ± 0.000	135.04-135.00
OP of glucose	0.52 ± 0.02			148.20-147.85
T [a]	2.14 ± 0.10	2.24 ± 0.05	1.96 ± 0.09	
I [b]	2.33 ± 0.08	2.79 ± 0.12	2.87 ± 0.03	
T*	1.98 ± 0.1	1.83 ± 0.04	1.89 ± 0.09	
I*	2.19 ± 0.08	2.66 ± 0.12	2.79 ± 0.03	
I *meta* [c]	1.83 ± 0.03	1.70 ± 0.09	1.74 ± 0.05	
I *para* [d]	0.5 ± 0.06	1.09 ± 0.05	1.13 ± 0.02	
Glucose [e]	0.50 ± 0.16	0.37 ± 0.01	0.38 ± 0.02	
Number of galloyl units per glucose moieties [f]	8.48 ± 1.00	12.24 ± 0.05	12.39 ± 0.27	
gallotannins in mg/g [g]	725 ± 28	749 ± 26	780 ± 20	
Mass balance in mg/g [h]	765 ± 30	829 ± 27	799 ± 20	

696 *Means and standard deviations are calculated from the results of three independent experiments.*

697 *The formula used are**: a**: ((OP3+OP5)T+(OP4)T))/3, **b**: ((OP3/OP5)I+(OP4)I))/2, **c**: (OP4)I ,*

698 ***d**: I - I meta, **e**: T/4 for TAN A, T/5 for TANB, **f**: 4(I+T)/(T) for TAN A, 5(I+T)/(T) for TAN B, **g**:*
699 *Sum of (molarity*molecular weight) of T, I and glucose moieties taking into account the loss of*
700 *H_2O in the ester bonds, **h**: Sum of gallotannins, mono, -di- and tri-gallic acids. I= Internal units,*
701 *T=Terminal units, T*=T minus mono-, di- and tri-gallic acid terminal units, I*= I minus di- and*
702 *tri-gallic acid intermediate units*

703

704 *Table 3. Summary of the data generated by the different methods*

		UPLC-DAD[1]	UPLC-MS	NMR	GPC
Gallic acid (mg/g)	A B1 B2	22.9 ± 0.7 73.3 ± 0.9 8.9 ±0.1		17 ± 1.7 59.5 ± 3.4 6.8 ± 0.1	B1>>A>B2
Digallic acid (mg/g)	A B1 B2	26.9 ± 0.9 20.3 ± 0.3 12.5 ± 0.3		14.2 ± 0.6 13.5 ± 0.6 7.4 ± 0.6	
Trigallic acid (mg/g)	A B1 B2	9.83 ± 0.26 7.17 ± 0.09 4.73 ± 0.14		8.5 ± 0.9 7.1 ± 0.5 4.3 ± 0.2	
β-glucogallin (mg/g)	A B1 B2	0.71 ± 0.02 0.12 ± 0.01 0.17 ± 0.01			
Gallotannins (mg/g)	A B1 B2	790 ± 66 822± 14 875 ± 8		725 ± 28 749 ± 26 780 ± 20	B2 > B1
Glucose substitution	A B1 B2	tetragalloylglucose pentagalloylglucose pentagalloylglucose			
Average number of galloyl groups	A B1 B2			8.5 12.2 12.4	5G-glc > 5G-glc > 5G-glc
Max number of galloyl groups detected	A B1 B2		12 17 18		
Major species detected	A B1 B2	4G-glc, 5G-glc 7G-glc, 8G-glc 7G-glc, 8G-glc (different isomers)			
Additional information	A B1 B2		galloylglucose ethers (as methyl esters detected after methanolysis)	unknown low molecular weight aliphatic compound in B1 proportion of p-depside bonds > in B	

705 [1] Gallic, digallic, trigallic acids, and β -glucogallin in mg/g gallic acid equivalent; gallotannins in mg/g
706 pentagalloylglucose equivalent.

707

708

709 Graphic for Table of Content

710

711

712