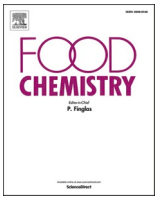
Food Chemistry 421 (2023) 136150

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem



A multiplatform metabolomics/reactomics approach as a powerful strategy to identify reaction compounds generated during hemicellulose hydrothermal extraction from agro-food biomasses

Andrea Fuso [a,*], Laura Righetti [a,b,c], Franco Rosso [d], Ginevra Rosso [d], Ileana Manera [d], Augusta Caligiani [a]

[a] Food and Drug Department, University of Parma, Via Parco Area delle Scienze 17/A, 43124 Parma, Italy
[b] Wageningen Food Safety Research (WFSR), Wageningen University & Research, P.O. Box 230, Wageningen 6700 AE, Netherlands
[c] Laboratory of Organic Chemistry, Wageningen University, Stippeneng 4, Wageningen 6708 WE, Netherlands
[d] Soremartec Italia Srl, Ferrero Group, 12051 Alba, CN, Italy

ARTICLE INFO

Keywords:
Hydrothermal treatment
Hazelnut shells
Degradation compounds
Ion mobility-mass spectrometry
Metabolomics
Hemicellulose

ABSTRACT

Hydrothermal treatment is commonly used for hemicelluloses extraction from lignocellulosic materials. In this study, we thoroughly investigated with a novel approach the metabolomics of degradation compounds formed when hazelnut shells are subjected to this type of treatment. Three different complementary techniques were combined, namely GC–MS, ^{1}H NMR, and UHPLC-IM-Q-TOF-MS. Organic acids, modified sugars and aromatic compounds, likely to be the most abundant chemical classes, were detected and quantified by NMR, whereas GC- and LC-MS-based techniques allowed to detect many molecules with low and higher Mw, respectively. Furans, polyols, *N*-heterocyclic compounds, aldehydes, ketones, and esters appeared, among others. Ion mobility-based LC-MS method was innovatively used for this purpose and could allow soon to create potentially useful datasets for building specific databases relating to the formation of these compounds in different process conditions and employing different matrices. This could be a very intelligent approach especially in a risk assessment perspective.

1. Introduction

In the last decades much attention has been paid in combatting the phenomenon of food waste, which represents a big issue for ethical, environmental, and economic reasons. Among all foodstuffs, it is well known that products of vegetable origin, such as cereals, tubers, roots, fruit, and vegetables, are connected to great quantities of waste. The composition of these wastes can be disparate, but according to a recent study about 90% of plant biomass is represented by lignocellulosic material (Saini, Saini, & Tewari, 2015), which means that it is mainly composed of lignin, hemicellulose and cellulose, in variable proportions. The valorization of these by-products, due to the extraordinary structural complexity of the polymers present within them, cannot disregard the principle of biorefinery, i.e. the separation of the individual biomass fractions for the production of different components, such as biomolecules, biomaterials, bioenergy and biofuels (Villacís-Chiriboga, Vera, Van Camp, Ruales, & Elst, 2021). In a historical period in which consumer's food choices are changing, moving towards healthier foods, dietary fibres naturally play a fundamental role. For this reason, hemicelluloses in particular acquire an even greater global interest. Indeed, hemicelluloses have a wide variety of applications: for instance, they can be easily transformed into oligosaccharides with interesting biofunctional properties or further depolymerized into pentoses and hexoses for subsequent conversion into bioethanol and chemical substances (Wang et al., 2016). Hemicelluloses are relatively small heteropolysaccharides (degree of polymerization 100–200 units), branched and made of five- and six-carbon monosaccharide units, of which the most frequent are xylose, mannose, arabinose, galactose, glucose, and glucuronic acid, as well as acetyl groups (Luo et al., 2019). These complex polymers were found to be linked by hydrogen bonds and van der Waals interactions with cellulose, forming highly resistant networks (Luo et al., 2019), and at the same time they interact with lignin through

* Corresponding author.
E-mail addresses: andrea.fuso@unipr.it (A. Fuso), laura.righetti@wur.nl (L. Righetti), franco.rosso@ferrero.com (F. Rosso), ginevra.rosso@ferrero.com (G. Rosso), ileana.manera@ferrero.com (I. Manera), augusta.caligiani@unipr.it (A. Caligiani).

https://doi.org/10.1016/j.foodchem.2023.136150
Received 7 November 2022; Received in revised form 4 April 2023; Accepted 10 April 2023
Available online 17 April 2023

complex interactions, which are radical coupling of ferulate substitutions and incorporation of hemicellulosic glycosyl residues by re-aromatization of lignification intermediates (Terrett & Dupree, 2019). Hence, to extract and isolate hemicelluloses, "strong" methods are needed, and the most used combine high temperatures and high pressures. The most common in this sense is the hydrothermal treatment (HT), also called autohydrolysis treatment, which consists in subjecting the matrix of interest to extraction at 160–220 °C in water, kept in the liquid state thanks to the high pressures (Kumari & Singh, 2018). When such treatment is applied, however, it is very likely to get the formation in the reaction medium of innumerable undesired compounds, such as degradation products derived from lignin, sugars, or proteins, as well as monosaccharides, furans and organic acids. The mechanism of formation of these compounds is extremely complicated, since it is strictly related to the time/temperature conditions set, to the pH and to the matrix. Depending on the reaction conditions, for example, glucose can be converted into 5-(hydroxymethyl)-2-furaldehyde (HMF) and/or levulinic acid, formic acid and various phenolics at high temperatures, while xylose can follow different reaction mechanisms originating furan-2-carbaldehyde (furfural) and/or various C-1 and C-4 compounds (Rasmussen, Sørensen, & Meyer, 2014). Furthermore, monosaccharides can further react to form pseudo-lignin, humine, aldehydes, ketones, organic acids, or aromatic compounds (Rasmussen, Sørensen, & Meyer, 2014). Several studies have shown how it was possible to vary the composition of the autohydrolysis liquor after HT in different conditions, mainly in terms of furfural, HMF, acetic, formic, and lactic acid content (Wang et al., 2016). At the same time, other research has pointed out that although some mechanisms are to date well understood, many other metabolic pathways remain unknown (Gao et al., 2016). Therefore, since one of the main aims of reusing hemicellulose from vegetable by-products is its transformation into healthy ingredients for food companies, it is of enormous importance to further investigate the presence and the formation of all these compounds that originate following HT, especially to evaluate them in terms of potential toxicity. This investigation must be done both through the study and understanding of the reaction mechanisms, and through the improvement of analytical techniques that allow their identification and quantification. In this complex scenario, where multiple reaction pathways not fully understood led to very complicated mixtures of neoformed compounds, there is the need of combining different and complementary analytical approaches, to further unravelling the reaction mechanisms and better characterize the composition of lignocellulose hydrothermal extracts. To this aim, in the present work the molecular composition of the hydrothermal extracts of hazelnut shells (HS) was studied using different metabolomics platform, namely ^{1}H NMR, GC–MS and UHPLC-IM-Q-TOF-MS.

2. Materials and methods

2.1. Chemicals and reagents

HPLC-grade acetonitrile, ammonium formiate, trifluoroacetic acid (TFA), N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA), 3-(trimethylsilyl)propionate-*d4* (TSP), D-glucose, D-fructose, D-galactose, D-mannose, D-rhamnose, D-ribose, D-xylose, D-fucose, D-galacturonic acid, D-glucuronic acid, D-glucosamine, D-galactosamine and phenyl-β-D-glucopyranoside were purchased from Sigma-Aldrich (Taufkirchen, Germany); bidistilled water was obtained using Milli-Q System (Millipore, Bedford, MA, USA); diethyl ether, hydrochloric acid and ethanol were purchased from Carlo Erba (Milan, Italy); D_2O was bought from VWR (Radnor, PA, USA); MS-grade formic acid from Fisher Chemical (Thermo Fisher Scientific Inc., San Jose, CA, USA) was also used. Leucine-enkephalin [186006013] used as lock mass solution and Major Mix IMS/TOF Calibration Kit [186008113] for mass and CCS calibration were purchased from Waters (Manchester, UK).

2.2. Hydrothermal treatment of hazelnut shells

Hazelnut shells (HS) were kindly provided by Ferrero S.p.A. (Cuneo, Italy), finely ground and sieved with grain size <500 μm. The hydrothermal treatment (HT) was performed in a stainless-steel Parr reactor 4566 (Parr Instrument Company, Moline, Illinois, USA) with an internal volume of 300 mL and internal cooling loop, equipped with Parr PDI for temperature control (model 4848). 5 g of ground sample were extracted with 125 mL of bidistilled water, and HT was carried out under isothermal conditions for 60 min at 175 °C. The time to reach the working temperature was 30 min and the final relative pressure was 8.1 bar. After treatment, the obtained solid and liquid phases were separated by centrifugation at 3900 rpm for 30 min at 4 °C. Finally, the supernatant was collected and freeze-dried. All the following analyses described in the next sections were performed on this sample.

2.3. Proximate composition of HS extract

Proximate composition of HS extract was investigated using standard procedures (AOAC, 2002). Moisture was determined by drying in oven at 105 °C for 24 h. Total ash was determined through mineralization at 550 °C in two steps, each one lasting 5 h. Total nitrogen was determined with a Kjeldahl system (DKL heating digester and UDK 139 semi-automatic distillation unit, VELP SCIENTIFICA) using 6.25 as a nitrogen-to-protein conversion factor. Lipid content was determined using a Soxhlet extractor (SER 148/3 VELP SCIENTIFICA, Usmate Velate, Italy) employing diethyl ether as extracting solvent.

For the determination of total sugars and monosaccharide distribution, 10 mg of sample were dissolved in 3 mL of 2 N trifluoroacetic acid (TFA) and hydrolysed at 110 °C for 2 h under stirring. Then, 900 μL of the solution were withdrawn and put together with 150 μL of 1000 ppm phenyl-β-D-glucopyranoside, used as internal standard, and then evaporated by rotavapor. The obtained dried hydrolysate was washed with 1 mL of methanol to remove the residue of TFA and evaporated again. 1 mL of 0.5 M NH_4OH was subsequently added to delactonize the eventually present acid sugar lactones in the sample, and again evaporated by rotavapor. Finally, the dried hydrolysate was dissolved in 800 μL of dimethylformamide (DMF) and 200 μL of N,O-Bis(trimethylsilyl) trifluoroacetamide (BSTFA), the latter used as derivatizing agent. The reaction was held for 1 h at 60 °C and finally the derivatized sample was injected in gaschromatography. GC–MS analysis of monosaccharides was performed with a 6890 N gas chromatograph coupled to a 5973 N mass selective detector (Agilent technologies, Santa Clara, CA). A SLB-5 ms, 30 m × 0.25 mm, 0.25 μm thickness column (Supelco, Bellafonte, PA, USA) was used. The chromatogram was recorded in the scan mode (40–500 *m/z*) with a programmed temperature from 60 °C to 270 °C. The initial temperature was 60 °C, held for 2 min, then increased to 160 °C at a rate of 10 °C/min, held isothermal for 5 min, increased to 220 °C at a rate of 10 °C/min, kept for 5 min, increased to 270 °C at a rate of 20 °C/min and maintained for 5 min. Quantification was performed with a response factor, considering the area and concentration ratios between the internal standard (phenyl-β-D-glucopyranoside), and the following monosaccharides: D-glucose, D-fructose, D-galactose, D-mannose, D-rhamnose, D-ribose, D-xylose, D-fucose, D-galacturonic acid, D-glucuronic acid, D-glucosamine and D-galactosamine.

2.4. Gaschromatography-mass spectrometry (GC–MS) analysis

The protocol for samples preparation before GC–MS analysis is represented in Fig. 1. Briefly, 300 mg of sample were weighted and added to 6 mL of water and the extraction was carried out at 50 °C for 2 h. Later, the sample was split in three different 2 mL aliquots: the first ("H_2O") was simply analyzed after putting 200 μL in a round-bottomed flask containing 65 μL of 1160 ppm phenyl-β-glucopyranoside, used as standard. In both the second and third aliquot, 8 mL of ethanol were firstly added to make the fibres to precipitate. The solutions were thoroughly

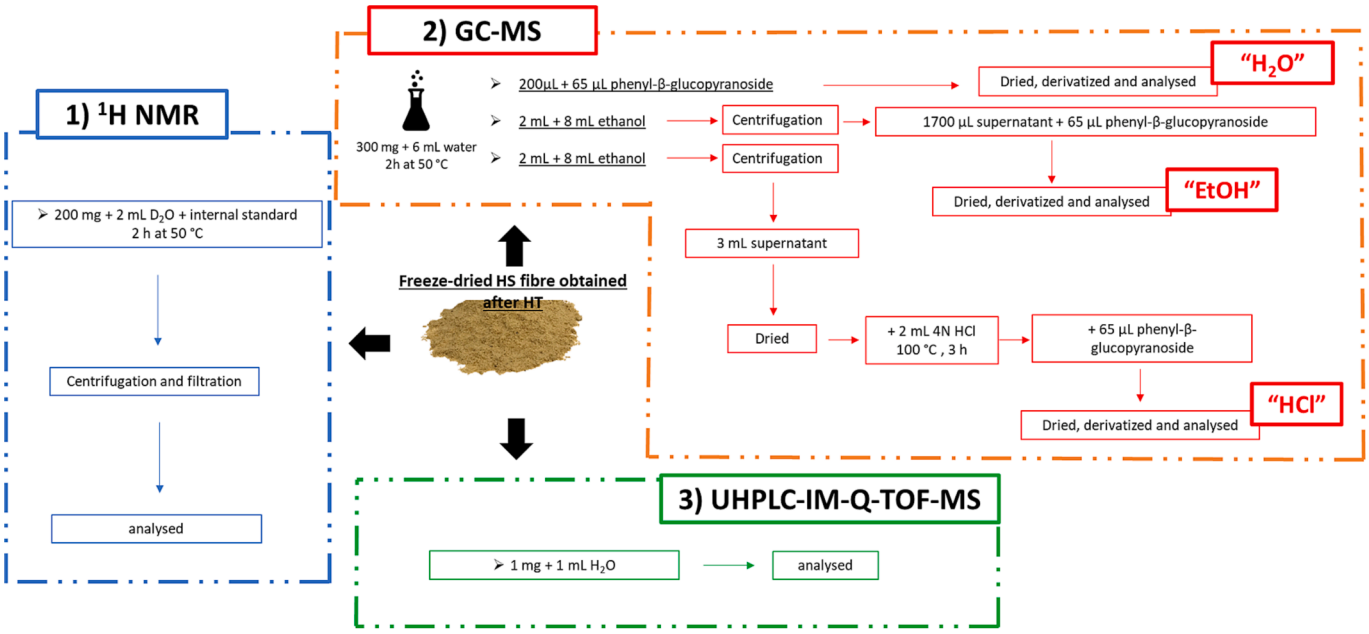


Fig. 1. Representation of samples preparation before 1H NMR, GC–MS, and UHPLC-IM-Q-TOF-MS analysis.

mixed and centrifuged at 4 °C and 3900 rpm for 25 min. Later, in the second aliquot ("EtOH") about 1700 µL of the supernatant were withdrawn and put together with 65 µL of 1160 ppm phenyl-β-glucopyranoside in a new round-bottomed flask. In the third aliquot ("HCl"), 3 mL of the supernatant obtained after centrifugation were withdrawn, put in a new round-bottomed flask, and dried by rotavapor, then here 2 mL of 4 N HCl were added, and the solution was allowed to stand at 100 °C for 3 h under stirring. At the end of the reaction, the same amount of phenyl-β-glucopyranoside used in the other experiments was added. Finally, the three solutions were dried by rotavapor, then derivatized by adding 800 µL of DMF and 200 µL of N,O-Bis(trimethylsilyl) trifluoroacetamide (BSTFA), the latter used as derivatizing agent, and allowing the reaction to stand for 1 h at 60 °C before being injected in GC–MS.

GC–MS analysis of the degradation products originated from the HT was performed with a 6890 N gas chromatograph coupled to a 5973 N mass selective detector (Agilent technologies, Santa Clara, CA). A SLB-5 ms, 30 m × 0.25 mm, 0.25 µm thickness column (Supelco, Bellafonte, PA, USA) was used. The mass spectrometer operated in the electron impact (EI) ionization mode (70 eV) and the ion source temperature was set at 230 °C. The chromatogram was recorded in the scan mode (50–800 *m/z*) with a programmed temperature from 60 °C to 270 °C. The initial temperature was 60 °C, held for 2 min, then increased to 160 °C at a rate of 10 °C/min, held isothermal for 10 min, increased to 220 °C at a rate of 10 °C/min, kept for 5 min, increased to 270 °C at a rate of 20 °C/min and maintained for 15 min. A semi-quantification of each analyte identified was performed, according to the equation (1) below:

$$\mathrm{ppm}_{\mathrm{analyte}} = \frac{\text{Area analyte} * \text{ppm phenyl - } \beta \text{ - D - glucopyranoside}}{\textit{Area phenyl} - \beta - D - \text{glucopyranoside}} \quad (1)$$

The identification of the various compounds was performed comparing mass spectra with WILEY library data.

2.5. ^{1}H NMR analysis

An aqueous extract of the freeze-dried extract obtained after hydrothermal treatment of HS was analysed through ^{1}H NMR. Briefly, 200 mg of sample were added to 2 mL D_2O together with 100 µL of 2210 ppm TSP dissolved in D_2O, and the extraction was performed under stirring at 50 °C for 2 h. Then, the solution was centrifuged at 4 °C and 3900 rpm for 25 min and subsequently filtered through a 0.45 µm nylon membrane in a 5 mm NMR tube (Fig. 1).

^{1}H NMR spectra were recorded on a Bruker Avance III 400 MHz NMR Spectrometer (Bruker BioSpin, Rheinstetten, Karlsruhe, Germany) operating at a magnetic field-strength of 9.4 T. A 1D 1H NOESY sequence, previously optimized to remove the residual water signal with minimal affection of the baseline, was utilized for water suppression (Musio et al., 2020). Spectra were acquired at 298 K, with 32 K complex points, using a 90° pulse length and 5 s of relaxation delay (d1). 128 scans were acquired with a spectral width of 20 ppm, an acquisition time of 1.707 s, 8 dummy scans and a mixing time of 0.010 s. The complete relaxation of the protons obtained during acquisition time and relaxation delay allowed to use integrals for quantitative purposes. Different spectral zones were identified and assigned to diverse classes of compounds found in hydrothermal extracts of hazelnut shells. The same zones were integrated and used for a preliminary crude quantification of the different compounds. Acetyl groups deriving from free acetate and substituted sugars (mainly acetylated xylans) were quantified integrating the spectral zone 1.92–2.34 ppm. Total aromatic compounds were determined from the integration of the whole zone 5.97–7.78 and expressed as phenylpropanoids. The signal of formic acid at 8.45 was integrated as determinant indicative of a part of degraded sugars according to the mechanism of levulinic acid formation (Pyo, Glaser, Rehnberg, & Hatti-Kaul, 2020).

2.6. Ultra-high-performance liquid chromatography – ion mobility – Quadrupole time-of-flight – mass spectrometry (UHPLC-IM-Q-TOF-MS) analysis

The reaction products originated from hydrothermal treatment of HS were further identified through UHPLC-IM-Q-TOF-MS. The sample was simply dissolved in water and analysed (Fig. 1). Specifically, 1 mg of the freeze-dried extract obtained after HT was dissolved in 1 mL of MilliQ water.

2.6.1. UHPLC-TWIMS-QTOF analysis

ACQUITY I-Class UHPLC separation system coupled to a VION IMS QTOF mass spectrometer (Waters, Wilmslow, UK) equipped with an electrospray ionization (ESI) interface was employed. Samples were

injected (2 μL) and chromatographically separated using a reversed-phase ACQUITY Premier HSS T3 column 2.1 × 100 mm, 1.8 μm particle size (Waters, Milford, MA, USA). A gradient profile was applied using water 5 mM ammonium formiate (eluent A) and acetonitrile (eluent B) both acidified with 0.1% formic acid as mobile phases. Initial conditions were set at 2% B, after 3 min of isocratic step, a linear change to 100% B was achieved in 12 min and holding for 5 min to allow for column washing before returning to initial conditions. Column recondition was achieved over 3 min, providing a total run time of 23 min. The column was maintained at 45 °C and a flow rate of 0.35 mL/min used.

Mass spectrometry data were collected in positive and negative electrospray mode over the mass range of *m/z* 50–1100. Source settings were maintained using a capillary voltage, 2.5 kV and 2 kV for positive and negative ESI modes, respectively; source temperature, 150 °C; desolvation temperature, 500 °C and desolvation gas flow, 950 L/h. Lock mass correction was performed by infusing a solution of leucine-enkephalin [M + H]+ (*m/z* 556.2766, calibration kit from Waters) at a concentration of 200 pg/μL (infusion rate 10 μL/min) and acquired every 2.5 min to provide a real-time single-point mass calibration. The TOF analyzer was operated in sensitivity mode and data acquired using HDMSE, which is a data independent approach (DIA) coupled with ion mobility. Low-energy scan (CE 6 V) and a high-energy scan (CE ramp 15–50 V) were alternatively acquired during the run. The optimized ion mobility settings included: nitrogen flow rate, 90 mL/min (3.2 mbar); wave velocity 650 m/s and wave height, 40 V. Device within the Vion was calibrated using the Major Mix IMS calibration kit (Waters, Wilmslow, UK) to allow for collisional cross section (CCS) values to be determined in nitrogen. The calibration covered the CCS range from 130 to 306 $Å^2$. The TOF was also calibrated prior to data acquisition and covered the mass range from 151 Da to 1013 Da. TOF and CCS calibrations were performed for both positive and negative ion mode. Data acquisition was conducted using UNIFI 1.8 (Waters, Wilmslow, UK) which enable to generate CCS values for all the detected species from the automatic CCS calibration.

2.6.2. Data processing

Data processing and compound identification were conducted using Progenesis QI Informatics (Nonlinear Dynamics, Newcastle, UK). Each UHPLC-MS run was imported as an ion-intensity map, including *m/z* and retention time, that were then aligned in the retention-time direction (0–20 min). From the aligned runs, an aggregate run representing the compounds in all samples was used for peak picking. This aggregate was then compared with all runs, so that the same ions were detected in every run. Isotope and adduct deconvolution were applied, to reduce the number of features detected. Metabolites were identified by publicly available database searches including PubChem as well as by fragmentation patterns, using Progenesis QI.

3. Results and discussion

3.1. Proximate composition of HS extract

After undergoing the HT and before studying the degradation compounds originated from it, both proximate composition and monosaccharide distribution of the freeze-dried extract was investigated. Results are reported in Table 1.

As expected, this sample was predominantly made of sugars, which accounted for half of its fresh weight (see Table 1). As concerns monosaccharide distribution of the fibres extracted, they were mainly made of xylose (63%). Other sugars, such as arabinose and glucuronic acid, were also present probably as substituents along the xylose-based chain, as often happens in xylans extracted from lignocellulosic matrices. Finally, galacturonic acid and glucose could be derived from pectin and cellulose fractions, while rhamnose and galactose could derive from pectin as well.

A relevant amount of ash, equal to 9%, was also found, while proteins and lipids were quite lower (3% and 1%, respectively). Interestingly, the sum of moisture, proteins, lipids, ash, and total sugars accounted for only 65%. By ^{1}H NMR analysis it was also possible to investigate the amount of acetyl groups attached to the xylan backbone, and to estimate both the aromatic compounds, derived from lignin and calculated as phenylpropanoids, as well as the sugars chemically modified by thermal degradation, overall reaching a total sum of about 80%. However, about one-fifth of the percentage by weight of the gross composition remained uncharacterized and most likely constituted by neoformation compounds risen from polysaccharides and lignin during the thermal process. In the effort of characterizing as much as possible these process-derived compounds, different analytical methods have been applied, as reported in the following paragraphs.

Table 1
Proximate composition and fibre's monosaccharide distribution of the freeze-dried extract obtained following ht of hs.

	Proximate composition (g/100 g of freeze-dried HT extract)	**Method**
Moisture	2.4 ± 0.1	AOAC official method (AOAC, 2002)
Lipids	1.0 ± 0.4	
Proteins	3.0 ± 0.5	
Ash	9.3 ± 0.2	
Total monosaccharides	49.3 ± 3.1	GC–MS and ^{1}H NMR
Total acetyl groups	8.0 ± 0.7	^{1}H NMR
Total aromatic compounds (expressed as phenylpropanoids)	5.0 ± 0.7	^{1}H NMR
Degraded sugars	1.1 ± 0.1	^{1}H NMR
	Monosaccharide distribution (g/100 g monosaccharides)	
Arabinose	8.1 ± 1.2	GC–MS
Rhamnose	6.0 ± 0.8	
Xylose	63.4 ± 1.3	
Galactose	4.4 ± 0.4	
Glucose	1.4 ± 0.3	
Galacturonic acid	8.0 ± 1.6	
Glucuronic acid + 4-O-methylglucuronic acid	8.7 ± 1.5	

3.2. GC–MS analysis

Following GC–MS analysis carried out on the sample treated in three different ways (Fig. 1), compounds identification was carried out by comparing the mass spectra of the trimethylsilyl (TMS) derivatives with Wiley275 library. A total of 54 compounds were selected and integrated as they are likely to be derived from HT (Table 2).

As reported in Table 2, the set of identified and semi-quantified compounds constituted 1.4%, 2.3% and 5.9% of the initial sample when semi-quantified in the simple aqueous extract (H_2O), in the aqueous extract followed by precipitation of the fibre by ethanol (EtOH), and in the aqueous extract followed by fibre precipitation and acid hydrolysis (HCl), respectively. It is important to emphasize how the acid hydrolysis carried out on the supernatant after fibre precipitation allows to quantify a total content of metabolites much higher than that obtained with a simple aqueous extract. This indicates the complex nature of the compounds extracted from the raw material or originated after the hydrothermal treatment, which are probably present conjugated to other compounds and not directly detectable by GC–MS due to the high molecular weight (Mw). The compounds were putatively identified comparing their mass spectra with Wiley 275 library and are reported in Table 2. They can be categorized into 5 main classes, namely organic acids, aromatic compounds, modified sugars (these three are the classes identified with ^{1}H NMR technique as well), fatty acids and polyols. Then, other compounds were also detected, including a couple of *N*-heterocyclic compounds and various unidentified signals. Organic acids

Table 2
List of compounds originated after ht, identified and semi-quantified by gc–MS.

Peak number	RT (min)	Compound	Main MS peaks of TMS derivatives (*m/z*)	Identification score	Pathway	Semi-quantitative amount (g/100 g initial sample)		
						Water Extract ("H_2O")	*Water Extract after Fibre Precipitation ("EtOH")*	*Acid Hydrolysed-Water Extract after Fibre Precipitation ("HCl")*
Organic acids								
1	6.88	Lactic acid	73,117,147,219,190	74	Glucose oxidation/ raw material	0.054	0.029	0.136
2	7.15	Glycolic acid	73,147,177,205	78	From sugars	0.045	0.004	0.113
3	8.02	2-ketogluconic acid	73,103,117,147,189,204	59	From sugar oxidation	0.019	0.013	ND
4	8.11	Levulinic acid	73,75,131,145,173	93	From sugars	ND	ND	0.012
6	8.24	3-hydroxypropionic acid	73,147,177,219	59	Aromatic compounds oxidation	ND	0.007	ND
10	10.74	5-hydroxyvaleric acid	73,75,147,172,247,	86	From sugars	0.011	ND	0.020
11	10.94	2,3-dihydroxy propanoic acid	73,103,147,189,205,292,307	83	Glycerol derivative	ND	ND	0.008
12	11.02	2-hydroxyhexanoic acid	73,147,159,190,233,261	78	From sugars	ND	ND	0.011
Total organic acids						**0.129**	**0.053**	**0.300**
Aromatic compounds								
7	9.83	Benzoic acid	51,77,105,135,179,194	86	lignin	0.320	0.420	0.161
24	23.56	3-Vanillyl propanol	73,179,192,206,221,236,311,326	83	lignin	0.010	0.024	0.011
25	23.67	3,4-dihydroxy benzoic acid	73,193,223,267,281,311,355,370	99	lignin	ND	ND	0.062
28	26.27	4-hydroxy mandelic acid	73,147,179,193,237,251,267	59	lignin	0.001	0.002	0.001
32	26.88	Gallic acid	73,179,281,311,355,399,443,458	99	raw material?-released after acid hydrolysis	0.022	0.006	0.091
34	27.70	Vanilethanediol	73,147,223,267,297	90	lignin	0.179	0.237	ND
37	28.73	*trans*-diethylstilbestrol	73,217,368,383,397,412	47	lignin	ND	ND	0.043
41	30.07	7,8-dihydroxyflavone	73,208,281,310,383,398	68	lignin	ND	ND	0.010
49	36.26	Vanillyl derivate	73,193,225,297,355		lignin	0.007	ND	ND
50	36.37	Hydroferulic acid	73,192,209,297	74	lignin	0.104	0.215	ND
51	36.53	Vanillyl mandelic acid	73,192,209,297	56	lignin	0.120	0.215	ND
Total aromatic compounds						**0.763**	**1.122**	**0.379**
Polyols								
9	10.14	Glycerol	73,103,117,147,205,218	93	raw material/ lipolysis	0.191	0.197	0.267
17	13.47	Erythritol	73,103,129,147,189,205	60		0.006	0.008	0.000
21	17.91	Arabitol	73,103,147,205,217,307	74	raw material	ND	0.010	0.000
22	19.26	Xylitol	73,103,147,205,217,307,319	89	raw material	ND	0.035	0.022
23	19.28	Ribitol	73,103,147,189,205,217,243,307,319	74	raw material	0.035	ND	ND
29	26.30	Sorbitol	73,103,147,205,217,319	59	sugar	0.008	0.014	0.011
31	26.75	Inositol1	73,147,191,217,265,291,305,318	98	raw material	0.035	0.042	0.080
38	28.84	Inositol2	73,103,147,191,204,217,265,291,305,318,367	87	raw material	0.050	0.055	0.158
Total polyols						**0.325**	**0.361**	**0.538**
Fatty acids								
8	10.05	Octanoic acid	73,75,117,132,145,201,216	96	Triglycerides hydrolysis	ND	ND	0.006
13	11.41	Nonanoic acid	73,75,117,129,132,215,230	59	Triglycerides hydrolysis	ND	ND	0.007
35	28.29	Palmitic acid	73,117,129,145,313	98	Triglycerides hydrolysis	ND	ND	0.449
40	29.93	Eptadecanoic acid	73,117,132,145,327,342	90	Triglycerides hydrolysis	ND	ND	0.007
42	31.96	Stearic acid	73,117,132,145,341,356	58	Triglycerides hydrolysis	ND	ND	0.027
Total fatty acids						**ND**	**ND**	**0.496**
Modified sugars								
18	14.66	2-O-methyl-xylose	73,89,131,146,159,191,204	83	Pectin	ND	ND	0.019
19	14.97	2-O-Methyl glucose	73,89,146,159,191,204	64		ND	ND	0.018
26	24.10	Glucuronic acid lactone	73,129,147,230	83	Glucuronic acid	ND	ND	0.019

(*continued on next page*)

Table 2 (*continued*)

Peak number	RT (min)	Compound	Main MS peaks of TMS derivatives (*m/z*)	Identification score	Pathway	Semi-quantitative amount (g/100 g initial sample)		
						Water Extract ("H_2O")	*Water Extract after Fibre Precipitation ("EtOH")*	*Acid Hydrolysed-Water Extract after Fibre Precipitation ("HCl")*
43	34.08	Pentose derivative	73,191, 217,259,281,341,356,428	–	Disaccharide?	0.018	0.032	ND
44	34.24	Pentose derivative	73,103,147,191,217,243,259	–	Disaccharide?	0.009	0.021	ND
45	34.31	Pentose derivative	73,103,147,191,217,230,259	–	Arabinose disaccharide?	0.032	0.057	ND
46	34.74	Hexose derivative	73,131,204,217,246,273,363	–	Disaccharide?	0.025	ND	ND
47	34.97	Pentose derivative	73,191, 217,259,281,341	–	Disaccharide?	0.005	0.011	ND
48	35.13	Pentose derivative	73,103,147,191,217,230,259	–	Arabinose disaccharide?	0.011	0.018	ND
52	36.93	Hexose derivative	73,131,204,217,259,283,341	–	Disaccharide?	0.013	0.099	ND
53	38.27	Disaccharide	73,103,147,191,217,243,271,361		Sucrose-like disaccharide	ND	ND	0.021
54	34.00 - –> 38.00	Sugar derivatives	73,103,129,147,191,204,217,259,349	–		ND	0.460	3.899
Total modified sugars						**0.113**	**0.698**	**3.976**
Other								
5	8.14	3-hydroxy-pyridine	73,152,167	57	Maillard reaction	ND	0.008	0.013
14	12.47	Unknown1	73,243,258	–		ND	ND	0.034
15	12.71	2,4(1H,3H)-Pyrimidinedione, dihyro-1,3-dimethyl	73,99,147,243,258	52	Maillard reaction	0.020	ND	ND
16	13.20	Unknown2	73,147,243,258	–		0.006	ND	ND
20	17.19	Unknown3	73,159, 204		sugar	0.026	0.014	ND
27	26.07	Unknown4	73,223,297,307,323,338	–	lignin	ND	ND	0.077
30	26.54	Unknown5	73,235,267,293,309,324	–	lignin	0.018	0.037	ND
33	27.51	Unknown6	73,103,192,236,325,428	–		ND	ND	0.006
36	28.45	Unknown7	73,103,129,147,175,205,217	–	sugars	ND	ND	0.098
39	28.98	Unknown8	73,117,147,208,282,327	–		0.002	0.008	ND
Total "other"						**0.072**	**0.067**	**0.228**
Total						**1.402**	**2.301**	**5.917**

constituted, in terms of percentage of the initial sample, a rather small quantity, varying between 0.05% and 0.3%. In all three extracts, the most abundant ones were found to be lactic acid and glycolic acid, originating from glucose or xylose as a result of high temperatures (Onda, Ochi, Kajiyoshi, & Yanagisawa, 2008). Levulinic acid, that has been classified as one of the top 12 promising bio-based building blocks for the synthesis of fuels and chemicals, was found in very small concentrations, and this is not surprising since it is formed from hexoses or directly from cellulose (Pyo et al., 2020), which are quite scarce in our extract. On the other hand, the higher amount of formic acid quantified by ^{1}H NMR could indicate the formation of a larger amount of levulinic acid which is further degraded during thermal treatment. 3-hydroxypropionic (3-HP) acid was found in very low concentrations too, and although its main origin is from glucose fermentation, in this case the formation is likely attributable to the oxidation of aromatic compounds (Shende & Levee, 1999). Finally, 5-hydroxyvaleric acid has been reported to derive from oxidation of furfural derivatives, such as cyclopentanone or 1,5-pentanediol, and therefore from xylose (Asano, Takagi, Nakagawa, Tamura, & Tomishige, 2019), while 2,3-dihydroxypropanoic acid (glyceric acid) is derived from glycerol oxidation (Yan et al., 2021) and 2-hydroxyhexanoic acid from sugar degradation in acid condition (Esteves et al., 2022). Even not detected by GC–MS, acetic and formic acids were also present, as evidenced by ^{1}H NMR, originating from autohydrolysis of acetylated xylans and degradation of hexose, respectively. As a whole, the majority of organic acids detected, originated from sugar modification pathways.

Aromatic compounds as a class constituted a significant fraction of the extract, if compared to organic acids: their total quantity was indeed semi-quantified with values ranging from 0.4% to 1.1%. These compounds are formed from the aromatic residues of lignin and are degraded to many types of phenolic structures, depending on the monomeric units in the native lignin (Almeida et al., 2007). In fact, lignin is a polymer made of various amounts of three different monolignols, that are p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. For this reason, most of the identified aromatic compounds might be divided in three subclasses, namely p-coumaryl, coniferyl, or sinapyl alcohol derivatives.

Previous studies showed how the increase in temperature during HT leads to methoxy group (-O-CH_3) hydrolysis with the consequent increase in p-coumaryl derivatives, making it harder to identify the mechanism of formation (Ramachandran et al., 2020). Among the compounds identified in our work by GC–MS, four of them were classifiable as coniferyl alcohol derivatives: 3-vanilyl propanol, vanilethanediol, hydroferulic acid and vanillylmandelic acid. The latter two were always found to be among the compounds present in the highest amount, even though not detected in "HCl" sample. The most abundant aromatic compound turned out to be benzoic acid, one of the most known lignin-derived molecules, and this may be explained by its great stability to high temperatures in liquid water (Dunn, Burns, Hunter, & Savage, 2003), while gallic acid, 7,8-dihydroxyflavone and diethylstilbestrol are likely derived from the raw material and were found only or especially in "HCl" sample, suggesting their presence in a complex with other structures.

Polyols were always found, as sum, in rather similar quantities in the three samples and varied between 0.3% and 0.5% of the initial sample, indicating their presence mainly as free form in the extract. An exception was glycerol, the most abundant one, its presence could be due to the thermal hydrolysis of triacylglycerols in HS and it is not surprising that

the highest amount has been found in "HCl" sample, being even easier to release it following acid hydrolysis. Sorbitol probably derived from the reduction of glucose, while the pentitols xylitol, arabitol and ribitol from the reduction of xylose and arabinose (Heuel, Shakeri-Garakani, Turgut, & Lengeler, 1998; Verduyn et al., 1985) in the pentose and glucuronate interconversions pathway in the raw material. Erythritol has been reported as a compound mainly produced from glucose through yeast fermentation (Kobayashi & Fukuoka, 2013) while inositol, on the other hand, may be produced from the hydrolysis of phytic acid, and it has been recently demonstrated how its degradation through HT is effective already with low temperatures (Sheikh & Saini, 2022). Then, many fatty acids were found, only in HCl sample because deriving from hydrolysis of triglycerides present in the raw material, and among them the most abundant was palmitic acid. Their total amount correspond to about 0.5% of the initial sample and adding up this value to glycerol's one the result turns out to be in accordance with the total lipid content reported in Table 1. One of the most numerous classes was that of modified sugars, that was the greatest in terms of abundance as well, reaching 4% of the initial sample in "HCl", showing again the complex nature of these compounds, likely bound to each other and detectable only after acid hydrolysis. Some rare sugars, namely 2-O-methyl xylose and 2-O-methyl glucose were also found: the first may have been obtained as a result of a simultaneous extraction of pectin fraction present in HS, because it has been reported as a rhamnogalacturonan-II component (Reuhs et al., 2004). Glucuronic acid lactone can be obtained from degradation of D-glucuronic acid when subcritical water is used, while a lot of sugar-derivative peaks appeared especially in "HCl" sample, in the last part of the chromatographic run, but without being able to identify them. These can be released from the matrix after hydrolysis, but it is not excluded their ex-novo formation during acid treatment. Finally, regarding "other" compounds, not classifiable in any of the previous classes, different abundant compounds were found. The maximum quantity of this substances peaked to 0.2% of the freeze-dried extract obtained from HT of HS. Both 3-hydroxy-pyridine and 2,4(1H,3H)-Pyrimidinedione, dihydro-1,3-dimethyl can be originated from Maillard reaction (Yu-Chiang, Chi-Tang, & Hartman, 1992).

Despite gas chromatographic methods are very often used for the characterization of HT degradation compounds (Rasmussen et al., 2014), their indisputable limits related to the necessity of matrix hydrolysis and analyte derivatization hinder a complete molecular characterization of the hydrothermal extract. As a consequence, the same sample was also submitted as it is to UHPLC-IM-Q-TOF-MS analysis.

3.3. UHPLC-IM-Q-TOF-MS analysis

The freeze-dried extract obtained after HT was also analysed by UHPLC-IM-Q-TOF-MS, dissolving it in water (see Fig. 1). This approach was applied to characterize in detail various degradation products with higher Mw that may originate from the thermal process. To the best of our knowledge, this is the first time that such a technique is used for this purpose. The sample was analysed in triplicate both in positive and in negative ion modes to enlarge the metabolite coverage. At first, 4852 and 3941 features were aligned in ESI+ and ESI− modes, respectively, and among these features a total of 212 compounds were putatively annotated. The selection of the compounds of interest was carried out at first checking their abundance and their fragmentation score. Then, only those compounds whose experimental fragmentation pattern matched with the theoretical and/or with the one present in the online databases were retained. Finally, duplicate metabolites such as those ionizing in both polarities were checked. Some features with even good abundance but low fragmentation score have therefore not been considered, even though they could constitute compounds with relevance for bioactivity or risk assessment. The workflow applied is represented in Fig. 2. As a further confirmation, *m/z* values of each compound were plotted against experimental TWIMS CCS (Fig. S4 of the supplementary material). The compounds presenting a mismatch between *m/z* and CCS have been reported as marked with an asterisk within the table containing the whole list of annotations (Table S1, supplementary material). These ions with low *m/z* values and high CCS may be due to post-mobility fragmentation. They may correspond to fragments formed after the TWIMS separation due to the low collision energy of 6 V applied to ensure a proper ion transmission through the collision cell. Most of these ions are annotated as acids and they could be originated from higher *m/z* value compounds such as acyl-quinic acids.

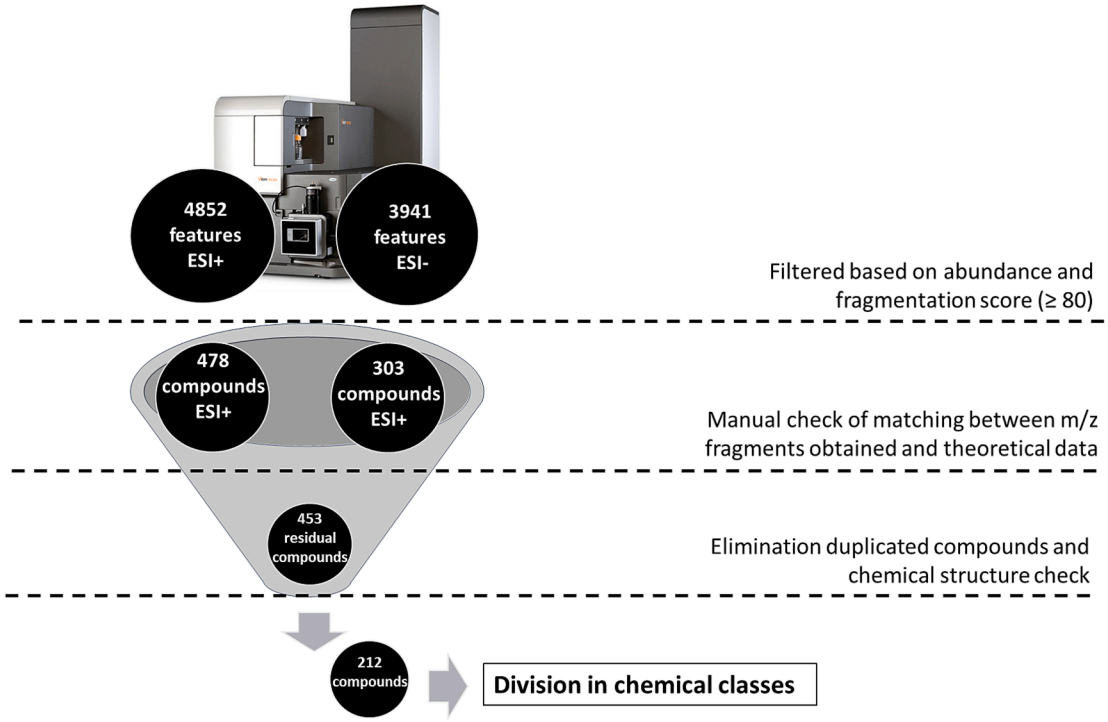


Fig. 2. Representation of funnel filtration of the compounds identified by UHPLC-IM-QTOF-MS.

Several examples of identification confirmation are reported in Fig. S1 of the supplementary material. The whole list of putatively identified compounds and their respective chemical structures are available in Table S1 of the supplementary material. All the selected compounds were finally classified in chemical classes, according to the main functional group. The inclusion of a compound in a specific class was in some cases arbitrary, being many molecules complex structures with more than one functional group. However, the secondary functional groups have been specified in Table S1.

Overall, all the compounds were classified into twenty-two classes, as reported in Fig. 3.

Some classes of compounds ionized both in ESI- and ESI+, and they are anhydro oligosaccharides, ketodeoxyoctonic acids (kdos), deoxy oligosaccharides, sugars conjugated with a large range of phenols, sugar conjugated with phenyl/diphenyl groups, sugars conjugated with various compounds (as for example pyran, fatty acid, polyol, ketone, purine), then organic acids, aldehydes, phenols, benzene-containing compounds, peptides, ketones, and furans. Then, in ESI- analysis one fatty acid and some pyran-derivatives were also detected, whilst carrying out ESI+ analysis even some sugar pyrimidinones, amino sugars, polyols and esters were comprised in the list. It is important to specify that many oligosaccharides and acetylated sugars were also putatively identified, but not selected nor reported in the list because not definable as degradation compounds. Indeed, both acetylated or unsubstituted sugars and oligosaccharides are formed from the hydrothermal extraction of hemicellulose, and they usually are the target compounds of this kind of process.

Generally, it can be noted that the main classes of molecules previously identified by GC–MS (i.e., organic acids, phenols, polyols, fatty acids, and modified sugars) came out with UHPLC-IM-Q-TOF-MS as well. As expected, many different types of modified sugars were detected, each of them with degrees of polymerization ranging from 2 to a maximum of 8 (Table S1). Each of the seven anhydro oligosaccharides was identified to be built up by agarobiose residues, and therefore they can be classified as agaro-oligosaccharides (AOS). Here, different DPs were detected, ranging from 2 (neoagarobiose) to 6. Agar/agarose is usually extracted from red seaweeds and basing on our knowledge no reports of its presence within lignocellulosic biomasses are present in literature. For this reason, their origin may need to be found from galactose degradation following the high temperatures employed in the HT. A recent work has proposed the mechanism of formation of levoglucosan and cellobiosan (i.e., glucose-based anhydro sugars) starting from cellulose by pyrolysis through hydrolysis and transglycosylation reactions (Leng et al., 2018), thus this may suggest the presence of galactose in HS in the form of galactan polymers. Seven compounds were also found and classified as Kdo sugars, which means that contain at least one 3-deoxy-d-*manno*-oct-2-ulosonic acid (or ketodeoxyoctonic acid, Kdo) molecule. Although Kdo is often just considered a characteristic component of bacterial lipopolysaccharide, it was also found to be a component of the cell walls of higher plants and in particular it is present only in the rhamnogalacturonan II structure (Temple, Saez-Aguayo, Reyes, & Orellana, 2016). This may suggest the formation of these compounds starting from pectin fractions, and this may be confirmed by the presence in relevant amounts of galacturonic acid among the total monosaccharides of HS (Table 1). Indeed, the UHPLC-IM-Q-TOF-MS analysis revealed the presence of Kdo sugars with different structures and degrees of polymerization, in some cases bound to each other, in others linked to monosaccharides or sometimes even to deoxy-sugars or dehydro-sugars. Some structures are reported in Fig. S2 of the supplementary material. In general, all these compounds (Kdos, anhydro oligosaccharides, deoxy oligosaccharides) accounted for about 8% of the total compounds identified (Fig. 3). Among them, deoxy sugars were the most abundant in terms of ion intensity, albeit numerically less. Although this value cannot be considered totally accurate from a quantitative point of view, it gives an indication that it is likely that other classes of molecules are more present in the hydrolysate. In fact, apart from the aforementioned compounds, coming from hydrolysis and modification of cellulose and hemicellulose, many other molecules were detected. Among them, organic acids, which represented almost 5% of the total number and only 1.3% in terms of relative abundance. Within this group, it can be noted that none of the molecules detected by GC–MS or ^{1}H NMR were found here, and this could be ascribable to the difference in their molecular weights. Indeed, all the organic acids identified by GC–MS or ^{1}H NMR had a low Mw, with a maximum of 132 g/mol, while the ones detected by UHPLC-IM-Q-TOF-MS had Mw in the range 188 – 290 g/mol. Within this class, monocarboxylic (e.g., 2-deoxypentonic acid, quinic acid), dicarboxylic (e.g. 3-oxoheptanedioic acid) and tri-carboxylic acids (1-Butene-1,2,4-tricarboxylic acid and 1-Hexene-1,2,6-tricarboxylic acid), but also more complex structures have been putatively identified. For example, the carboxylic group was also found in complex molecules, containing other functional groups and cyclic structures in the carbon skeleton, as a propanoic acid containing an indene-dioxol group, a methyl-butenoic acid substituted with and epoxycyclohexenoic acid, a 4-oxo-pentanoic acid containing a cyclopentane ring. The complexity of these structures makes any hypothesis about their formation speculative, also considering the lack of analytical standards. However, their presence in the hydrothermal extract suggests the need to reach a better level of understanding about the complex molecular profile of thermally treated biomasses.

Another class of compounds which confirms the result obtained with GC–MS and ^{1}H NMR involves lignin-derived aromatic molecules. Several sub-classes can be identified in it, depending on the functional groups and the molecular residues present bound together. Specifically, these sub-classes were phenols, sugar phenols (i.e., phenols conjugated with mono-/oligosaccharides), phenyls, phenyls/diphenyls sugars (i.e.,

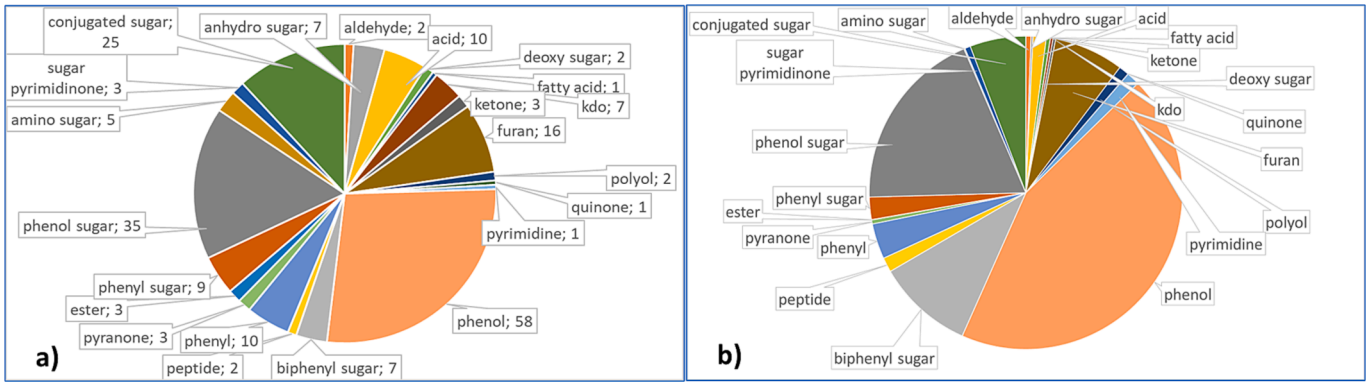


Fig. 3. Distribution of the compounds identified in different chemical classes, in terms of number of compounds per class (a) and relative abundance (b). Relative abundances were calculated dividing the peak area of each metabolite by the sum of the areas of all the metabolites annotated in that sample.

phenyls or diphenyls conjugated with mono-/oligosaccharides). The most present in this sense were by far phenols and sugar-phenols, comprising in total 93 compounds out of 212, corresponding to 44% of the total number of compounds and to 63% of the total relative ion abundance. These compounds come from complex mechanisms involving lignin. Compared to the aromatic molecules identified through GC–MS analysis, the ones found here were once again of greater complexity and with a Mw on average greater, also because constituted by complexes between phenols and other molecules. In fact, as reported in Table S1, very different molecules were found within the phenolic class, for instance containing functional groups typical of anthraquinones, naphtalenones, anthracenes, benzoquinones, stilbenes, tetraphenes, and again sulphate groups, furans and so on, depending on the number of rings that were present in the structure and on how they were bound together. Most of the features identified have never been previously reported for hydrothermal extracts of lignocellulosic materials, therefore it is difficult to extract information about their mechanism of formation, potential toxicity or, on the other hand, beneficial effects. Only few molecules among the identified ones were previously discussed in the literature. The presence of these compounds mainly in natural sources suggests that they (or some analogues) probably derive from the raw material. For example, a good relative abundance was detected for C-2′-decoumaroyl-aloeresin G, a phenolic glycoside previously isolated from Aloe vera (Dias & Rauter, 2015). Then, 6,8-Dihydroxy-3,4-bis (hydroxymethyl)-1H-isochromen-1-one is an isocoumarin previously obtained from a sponge-derived fungus on the genus *Penicillium* (Qi et al., 2013), and in the same way (2R,3S,4aS,9aR,10R)-2,3,5,8-Tetrahydroxy-6,10-dimethoxy-3-methyl-1,3,4,4a,9a,10-hexahydro-9(2H)-anthracenone and methyl (1S,2S,3S)-2,3,8-trihydroxy-3-methyl-9-oxo-2,3,4,9-tetrahydro-1H-xanthene-1-carboxylate were associated with fungi in literature as well (Yang et al., 2012; Zhang et al., 2012). Chlorogenic acid butyl ester was previously isolated from the leaves of a species of tree and flowering plant in the family *Anacardiaceae* (Corthout, Pieters, Claeys, Berghe, & Vlietinck, 1992). 5,7-Dihydroxy-2-(4-methoxyphenyl)-8-(3-methyl-2-buten-1-yl)-4-oxo-4H-chromen-3-yl-6-deoxy-β-D-mannopyranoside, better known as icariside II, is a prenylated flavonoid glycoside generally found in *Epimedii* herbs (Szabó, Pál, & Dulf, 2022). On the other hand, Fig. 4 reports as an example the chemical structures of two putatively identified phenolic compounds not found in natural sources. They correspond to complex cyclic molecules, having Mw ranging from 328 to 468 g/mol and despite no specific information about their mechanism of formation is reported, the presence of coumaryl and sinapyl alcohols' skeletons suggest their direct formation from lignin cleavage and intramolecular recondensation. At the same time, some of these structures seem distantly reminiscent of the skeleton of some diarylheptanoids. These molecules are frequently found in plants belonging to the *Betulaceae* family and have been described in depth in a recent review about hazelnut's phytochemicals (Bottone et al., 2019).

On the other hand, compounds carrying simple phenyl groups were numerically much less, and this is in line with the lignin origin of the most of these neoformed compounds. Nineteen compounds were detected and classified as phenyls or phenyl sugars, depending on the eventual presence of sugars residues, and even some biphenyl sugars were putatively identified in the extract: this is not surprising, since it is a typical bond that can be found in lignin structure. As in the other classes of compounds, it can be noted how different combinations of chemical structures were present again, such as furan-containing or acetic acid-containing molecules, and how different sugar moieties can occur. In general, the aromatic compounds detected showed extraordinary structural complexity. In a recent study, Kim and colleagues evaluated which are the main reactions lignin undergoes during a heat treatment. The authors concluded that these reactions concerned a methoxyl groups cleavage in C3 and C5, a propane side chain cleavage, a cleavage of the β-O-4 bond (depolymerization) and some simultaneous condensation reactions between released fragments, causing high molecular weight lignin (Kim et al., 2014). Although in our study the temperature of the HT was a little bit lower, our findings suggest that most of these condensation reactions may have taken place. Moreover, in the aforementioned work the lignin was extracted and thermally treated as it was, while in our case many other compounds, mainly derived from hemicellulose, had the chance to interact with lignin fragments, making even more complex the reaction pathways.

Another predominant class both in terms of relative abundance and number of compounds corresponded to furans, and sixteen molecules were detected and classified in this category. The most abundant one in negative ion mode was found to be a furan complexed with a hydrocarbon and containing a thio group, namely 5-(Tetradecylsulfanyl)-2-furoic acid. Some furancarboxylic acids have been reported to derive from oxidation of HMF (An, Sun, & Xia, 2019) and the presence of sulfate may therefore be due to some previous Maillard reactions with sulfur-containing amino acids. A derivative of maleic anhydride, namely 3-(4-Methyl-2,5-dioxo-2,5-dihydro-3-furanyl)propanoic acid was also found. In general, the putatively identified furans were once again found within very complex molecular structures, namely linked to other fragments of molecules containing other functional groups, such as

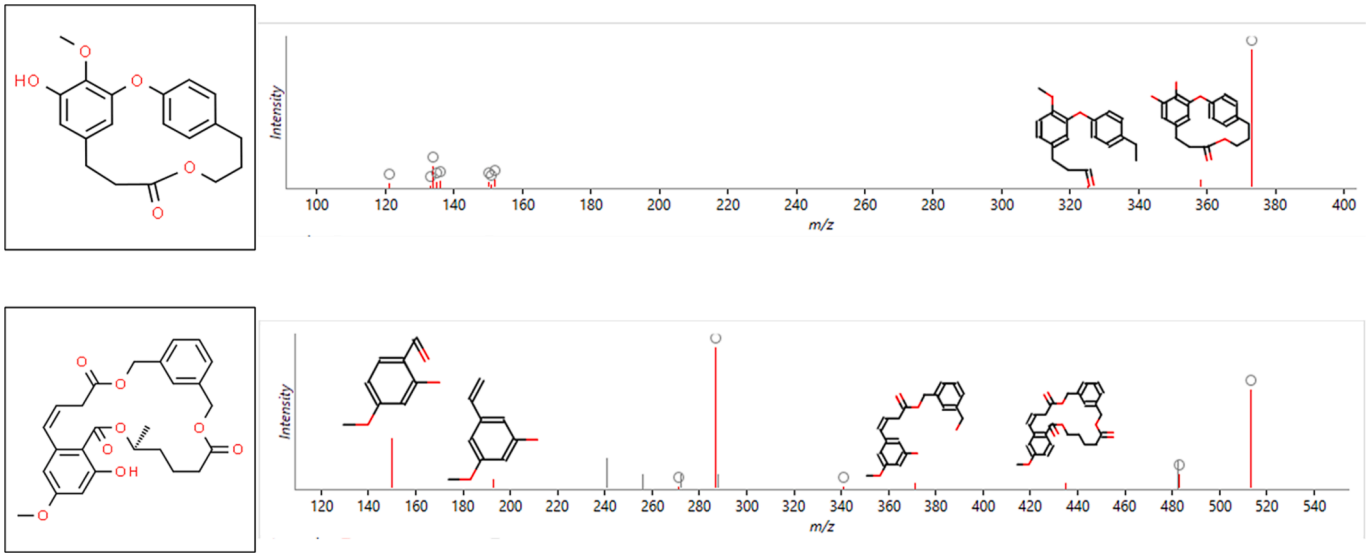


Fig. 4. Chemical structures of some annotated phenolic compounds and matching of their fragmentation spectrum.

phenols, phenyls, organic acids, and fatty acids (Fig. S3 of the supplementary material). Interestingly, simple furans as HMF or furanoic acids have not been detected with any of the analytical techniques used in this work, likely due to a loss in the most volatile molecules during HT extract freeze-drying process. The basic mechanism of formation of these furan-based molecules starts from sugars dehydratation, both starting from hexoses to form 5-HMF and/or from pentoses to form furfural, also depending on temperature and pH. Then, they may react in different ways, as dienes in the Diels–Alder reaction, but also through radical, electrophilic or polymerization reactions. A recent interesting review has been published about furans mechanisms of reaction (Gandini et al., 2022).

4. Conclusions

Since hydrothermal treatment is a very commonly used method to extract hemicellulose from lignocellulosic biomasses, it is essential to deeply investigate which are the degradation compounds that are simultaneously formed. Most of the literature available so far employed GC–MS analysis to identify the molecules originated from this kind of extraction. However, this technique presents some limitations in detecting high molecular weight and more complex compounds that may result from the degradation, hydrolysis and condensation reactions between lignin, hemicellulose, and other constituents of the matrix. In this work, three different techniques, namely ^{1}H NMR, GC–MS and UHPLC-IM-Q-TOF-MS have been employed to investigate and identify the degradation compounds after hydrothermal treatment of hazelnut shells, producing diverse but complementary information. These approaches led to the detection of some common chemical classes of compounds, likely the most abundant, such as organic acids, degraded/modified sugars and aromatic compounds, and ^{1}H NMR permitted to quantify them too. The two chromatographic-MS based techniques also allowed to detect different classes of furans, polyols, *N*-heterocyclic compounds, aldehydes, ketones, esters, among others. However, deciphering the chemical mechanism involved in the formation of all these compounds is still a major challenge. On the one hand several variables, primarily the HT severity, in terms of time and temperature, strongly influence their formation. On the other hand, also the initial matrix plays a fundamental role, since the proportion and the structure of the various polymers within the lignocellulosic complex can have a huge impact on the reaction mechanisms and products. However, sophisticated techniques such as UHPLC-IM-Q-TOF-MS are emerging and enabling to bridge this gap, allowing researchers to create datasets like ours, which can hopefully be used for the construction of specific metabolomics/"reactomics" databases, so that more light can be shed for the understanding of the formation of all these compounds and also for the building of predictive models.

Funding

This work was supported by Soremartec Italia Srl, Alba (CN, Italy).

CRediT authorship contribution statement

Andrea Fuso: Formal analysis, Investigation, Writing – original draft. **Laura Righetti:** Conceptualization, Methodology, Software, Investigation, Writing – review & editing. **Franco Rosso:** . **Ginevra Rosso:** Conceptualization, Investigation, Resources, Funding acquisition, Project administration, Supervision. **Ileana Manera:** Funding acquisition, Project administration, Supervision. **Augusta Caligiani:** Conceptualization, Validation, Resources, Writing – review & editing, Project administration, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

The authors are grateful to Simona Scarpella for helping in instrumental analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2023.136150.

References

Almeida, J. R. M., Modig, T., Petersson, A., Hähn-Hägerdal, B., Lidén, G., & Gorwa-Grauslund, M. F. (2007). Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. *Journal of Chemical Technology and Biotechnology, 82*(4), 340–349. https://doi.org/10.1002/jctb.1676

An, J., Sun, G., & Xia, H. (2019). Aerobic oxidation of 5-hydroxymethylfurfural to high-yield 5-hydroxymethyl-2-furancarboxylic acid by poly(vinylpyrrolidone)-capped Ag nanoparticle catalysts. *ACS Sustainable Chemistry and Engineering, 7*(7), 6696–6706. https://doi.org/10.1021/acssuschemeng.8b05916

AOAC. (2002). *Official method of analysis* (16th ed.). Washington DC: Association of Official Analytical.

Asano, T., Takagi, H., Nakagawa, Y., Tamura, M., & Tomishige, K. (2019). Selective hydrogenolysis of 2-furancarboxylic acid to 5-hydroxyvaleric acid derivatives over supported platinum catalysts. *Green Chemistry, 21*, 6133–6145. https://doi.org/10.1039/c9gc03315g

Bottone, A., Cerulli, A., Durso, G., Masullo, M., Montoro, P., Napolitano, A., & Piacente, S. (2019). Plant specialized metabolites in hazelnut (Corylus avellana) kernel and byproducts: An update on chemistry, biological activity, and analytical aspects. *Planta Medica, 85*(11/12), 840–855. https://doi.org/10.1055/a-0947-5725

Corthout, J., Pieters, L., Claeys, M., Berghe, D. V., & Vlietinck, A. (1992). Antiviral caffeoyl esters from Spondias mombin. *Phytochemistry, 31*(6), 1979–1981. https://doi.org/10.1016/0031-9422(92)80344-E

Dias, C., & Rauter, A. P. (2015). *Carbohydrates and glycomimetics in Alzheimer's disease therapeutics and diagnosis – Chapter 8*. RSC Drug Discovery Series.

Dunn, J. B., Burns, M. L., Hunter, S. E., & Savage, P. E. (2003). Hydrothermal stability of aromatic carboxylic acids. *Journal of Supercritical Fluids, 27*(3), 263–274. https://doi.org/10.1016/S0896-8446(02)00241-3

Esteves, B., Ayata, U., Cruz-Lopes, L., Brás, I., Ferreira, J., & Domingos, I. (2022). Changes in the content and composition of the extractives in thermally modified tropical hardwoods. *Maderas. Ciencia y Tecnología, 24*(22), 1–14. https://doi.org/10.4067/s0718-221x2022000100422

Gandini, A., & Lacerda, M. (2022). Furan polymers: State of the art and perspectives. *Macromolecular Materials and Engineering, 307*(6), 2100902. https://doi.org/10.1002/mame.202100902

Gao, Y., Wang, H., Guo, J., Peng, P., Zhai, M., & She, D. (2016). Hydrothermal degradation of hemicelluloses from triploid poplar in hot compressed water at 180–340 °c. *Polymer Degradation and Stability, 126*, 179–187. https://doi.org/10.1016/j.polymdegradstab.2016.02.003

Heuel, H., Shakeri-Garakani, A., Turgut, S., & Lengeler, J. W. (1998). Genes for D-arabinitol and ribitol catabolism from Klebsiella pneumoniae. *Microbiology, 144*(6). https://doi.org/10.1099/00221287-144-6-1631

Kim, J. Y., Hwang, H., Oh, S., Kim, Y. S., Kim, U. J., & Choi, J. W. (2014). Investigation of structural modification and thermal characteristics of lignin after heat treatment. *International Journal of Biological Macromolecules, 66*, 57–65. https://doi.org/10.1016/j.ijbiomac.2014.02.013

Kobayashi, H., & Fukuoka, A. (2013). Synthesis and utilisation of sugar compounds derived from lignocellulosic biomass. *Green Chemistry, 15*, 1740–1763. https://doi.org/10.1039/c3gc00060e

Kumari, D., & Singh, R. (2018). Pretreatment of lignocellulosic wastes for biofuel production: A critical review. *Renewable and Sustainable Energy Reviews, 90*, 877–891. https://doi.org/10.1016/j.rser.2018.03.111

Leng, E., Costa, M., Peng, Y., Zhang, Y., Gong, X., Zheng, A., Huang, Y., & Xu, M. (2018). Role of different chain end types in pyrolysis of glucose-based anhydro-sugars and oligosaccharides. *Fuel, 234*, 738–745. https://doi.org/10.1016/j.fuel.2018.07.075

Luo, Y., Li, Z., Li, X., Liu, X., Fan, J., Clark, J. H., & Hu, C. (2019). The production of furfural directly from hemicellulose in lignocellulosic biomass: A review. *Catalysis Today, 319*, 14–24. https://doi.org/10.1016/j.cattod.2018.06.042

Musio, B., Ragone, R., Todisco, S., Rizzuti, A., Latronico, M., Mastrorilli, P., … Gallo, V. (2020). A community-built calibration system: The case study of quantification of metabolites in grape juice by qNMR spectroscopy. *Talanta, 214*, Article 120855. https://doi.org/10.1016/j.talanta.2020.120855

Onda, A., Ochi, T., Kajiyoshi, K., & Yanagisawa, K. (2008). A new chemical process for catalytic conversion of d-glucose into lactic acid and gluconic acid. *Applied Catalysis A: General, 343*(1–2), 49–54. https://doi.org/10.1016/J.APCATA.2008.03.017

Pyo, S. H., Glaser, S. J., Rehnberg, N., & Hatti-Kaul, R. (2020). Clean production of levulinic acid from fructose and glucose in salt water by heterogeneous catalytic dehydration. *ACS Omega, 5*(24), 14275–14282. https://doi.org/10.1021/acsomega.9b04406

Qi, J., Shao, C. L., Li, Z. Y., Gan, L. S., Fu, X. M., Bian, W. T., Zhao, H. Y., & Wang, C. Y. (2013). Isocoumarin derivatives and benzofurans from a sponge-derived Penicillium sp. fungus. *Journal of Natural Products, 76*(4), 571–579. https://doi.org/10.1021/np3007556

Ramachandran, V., Ismail, F. S., Noor, M. J. M. M., Akhir, F. N. M. D., Othman, N., Zakaria, Z., & Hara, H. (2020). Extraction and intensive conversion of lignocellulose from oil palm solid waste into lignin monomer by the combination of hydrothermal pretreatment and biological treatment. *Bioresource Technology Reports, 11*, Article 100456. https://doi.org/10.1016/j.biteb.2020.100456

Rasmussen, H., Sørensen, H. R., & Meyer, A. S. (2014). Formation of degradation compounds from lignocellulosic biomass in the biorefinery: Sugar reaction mechanisms. *Carbohydrate Research, 385*, 45–57. https://doi.org/10.1016/j.carres.2013.08.029

Reuhs, B. L., Glenn, J., Stephens, S. B., Kim, J. S., Christie, D. B., Glushka, J. G., Zablackis, E., Albersheim, P., Darvill, A. G., & O'Neill, M. A. (2004). L-galactose replaces L-fucose in the pectic polysaccharide rhamnogalacturonan II synthesized by the L-fucose-deficient mur1 Arabidopsis mutant. *Planta, 219*, 147–157. https://doi.org/10.1007/s00425-004-1205-x

Saini, J. K., Saini, R., & Tewari, L. (2015). Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: Concepts and recent developments. *3 Biotech, 5*, 337–353. https://doi.org/10.1007/s13205-014-0246-5

Sheikh, M. A., & Saini, C. S. (2022). Combined effect of microwave and hydrothermal treatment on anti-nutritional factors, antioxidant potential and bioactive compounds of plum (Prunus domestica L.) kernels. *Food Bioscience, 46*, Article 101467. https://doi.org/10.1016/j.fbio.2021.101467

Shende, R. V., & Levee, J. (1999). Kinetics of wet oxidation of propionic and 3-hydroxypropionic acids. *Industrial and Engineering Chemistry Research, 38*(7), 2557–2563. https://doi.org/10.1021/ie9900061

Szabó, R., Pál, C. P., & Dulf, F. V. (2022). Bioavailability improvement strategies for icariin and its derivates: A review. *International Journal of Molecular Sciences, 23*(14), 7519.

Temple, H., Saez-Aguayo, S., Reyes, F. C., & Orellana, A. (2016). The inside and outside: Topological issues in plant cell wall biosynthesis and the roles of nucleotide sugar transporters. *Glycobiology, 26*(9), 913–925. https://doi.org/10.1093/glycob/cww054

Terrett, O. M., & Dupree, P. (2019). Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. *Current Opinion in Biotechnology, 56*, 97–104. https://doi.org/10.1016/j.copbio.2018.10.010

Verduyn, C., van Kleef, R., Frank, J., Schreuder, H., Van Dijken, J. P., & Scheffers, W. A. (1985). Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast Pichia stipitis. *Biochemical Journal, 226*, 669–677. https://doi.org/10.1042/bj2260669

Villacís-Chiriboga, J., Vera, E., Van Camp, J., Ruales, J., & Elst, K. (2021). Valorization of byproducts from tropical fruits: A review, Part 2: Applications, economic, and environmental aspects of biorefinery via supercritical fluid extraction. *Comprehensive Reviews in Food Science and Food Safety, 20*(3), 2305–2331. https://doi.org/10.1111/1541-4337.12744

Wang, Z. W., Zhu, M. Q., Li, M. F., Wang, J. Q., Wei, Q., & Sun, R. C. (2016). Comprehensive evaluation of the liquid fraction during the hydrothermal treatment of rapeseed straw. *Biotechnology for Biofuels, 9*, 142. https://doi.org/10.1186/s13068-016-0552-8

Yan, H., Yao, S., Zhao, S., Liu, M., Zhang, W., Zhou, X., Zhang, G., Jin, X., Liu, Y., Feng, X., Chen, X., Chen, D., & Yang, C. (2021). Insight into the basic strength-dependent catalytic performance in aqueous phase oxidation of glycerol to glyceric acid. *Chemical Engineering Science, 230*, Article 116191. https://doi.org/10.1016/j.ces.2020.116191

Yang, K. L., Wei, M. Y., Shao, C. L., Fu, X. M., Guo, Z. Y., Xu, R. F., Zheng, C. J., She, Z. G., Lin, Y. C., & Wang, C. Y. (2012). Antibacterial anthraquinone derivatives from a sea anemone-derived fungus Nigrospora sp. *Journal of Natural Products, 75*(5), 935–941. https://doi.org/10.1021/np300103w

Yu-Chiang, O., Chi-Tang, H., & Hartman, T. G. (1992). Volatile compounds generated from the Maillard Reaction of Pro-Gly, Gly-Pro, and a mixture of glycine and proline with glucose. *Journal of Agricultural and Food Chemistry, 40*, 1878–1880. https://doi.org/10.1021/jf00022a030

Zhang, F., Li, L., Niu, S., Si, Y., Guo, L., Jiang, X., & Che, Y. (2012). A thiopyranchromenone and other chromone derivatives from an endolichenic fungus, Preussia africana. *Journal of Natural Products, 75*(2), 230–237. https://doi.org/10.1021/np2009362