



CrossMark
click for updates

Cite this: *Analyst*, 2015, **140**, 4708

Synthesis of model humic substances: a mechanistic study using controllable H/D exchange and Fourier transform ion cyclotron resonance mass spectrometry†

Alexander Ya. Zherebker,^a David Airapetyan,^a Andrey I. Konstantinov,^a Yury I. Kostyukevich,^{b,c,d} Alexey S. Kononikhin,^{b,c} Igor A. Popov,^{b,c,e} Kirill V. Zaitsev,^a Eugene N. Nikolaev^{*b,c,d,e} and Irina V. Perminova^{*a}

The products of the oxidative coupling of phenols are frequently used as synthetic analogues to natural humic substances (HS) for biomedical research. However, their molecular compositions and exact structures remain largely unknown. The objective of this study was to develop a novel approach for the molecular-level analysis of phenolic polymerisates that is capable of inventorying molecular constituents and resolving their distinct structural formulas. For this purpose, we have synthesized the model HS using the oxidative coupling of a specifically designed phenylpropanoic monomer, 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid, to hydroquinone. We have characterized the synthesized model HS using high resolution Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS), ¹H NMR spectroscopy, and controllable hydrogen/deuterium (H/D) exchange. We succeeded in the molecular inventory of the model HS. The assigned molecular formulas occupied the substantial space of CHO compositions in the Van Krevelen diagram with a maximum density found in the regions of tannins and lignins, resembling those of natural HS. To identify the exact structural formulas of the individual constituents in the model HS, we have applied selective H/D exchange of non-labile backbone protons by a choice of basic or acidic catalytic conditions followed by FTICR MS. The determined formulas allowed us to verify the proposed pathways of hydroxylation and carboxylation in the course of the phenolic coupling and to identify the acetylation of aromatic rings as an important side reaction. We conclude that the proposed analytical approach may be used to identify the molecular carriers of biological activity within the phenolic polymerisates and eventually within natural HS.

Received 26th March 2015,
Accepted 5th May 2015

DOI: 10.1039/c5an00602c

www.rsc.org/analyst

^aDepartment of Chemistry, Lomonosov Moscow State University, Leninskie gory 1-3, 119991 Moscow, Russia. E-mail: iperm@org.chem.msu.ru; Fax: +7(495)9395546; Tel: +7(495)9395546

^bInstitute for Energy Problems of Chemical Physics of RAS, Leninskij pr. 38-2, 119334 Moscow, Russia

^cMoscow Institute of Physics and Technology, Dolgoprudnyi, 141700 Moscow Region, Russia

^dSkolkovo Institute of Science and Technology, Novaya St. 100, 143025 Skolkovo, Russia

^eEmanuel Institute of Biochemical Physics of RAS, Kosygina str. 4, 119991 Moscow, Russia. E-mail: ennikolaev@rambler.ru; Fax: +7(916)5491047; Tel: +7(916)5491047

† Electronic supplementary information (ESI) available: (1) Synthetic protocols for the intermediate reagents used for synthesis of the phenylpropanoic monomer; (2) ¹H NMR spectra of the synthetic products; (3) reaction pathways leading to the formation of the products from **9** to **11**; (4) mass lists with errors of identifications and the corresponding CHO assignments made for MHQ-HA and MHQ-FA based on FTICR MS data. See DOI: 10.1039/c5an00602c

Introduction

Humic substances (HS) are natural compounds that are formed during the oxidative decomposition of biomacromolecules constituting plant residues and other debris of living organisms.¹ As a result, they are comprised of versatile classes of chemical compounds with dominating contributions of oxidized aromatic moieties stemming from lignins and polyphenols due to their abundance and refractory character.^{2,3} These phenolic compounds are believed to contribute the most to the remarkable biological activity of HS including antiviral, antibiotic, and carcinostatic effects, which have been extensively reported in the literature.^{4,5} Hence, the products of phenolic oxidative coupling are frequently used for biomedical research as synthetic surrogates to natural HS.^{5–8} The structure of these synthetic HS may be considerably better controlled as compared

to natural HS by selecting phenolic precursors thereby improving the targeting of their therapeutic applications.⁹

The choice of synthetic strategy used for preparing the synthetic HS relies mostly on the oxidative coupling of phenols, which is the major process of the formation of HS in nature.¹⁰ As such, this process has been intensively studied since the beginning of the 20th century.^{11,12} Goh and Stevenson were the first to conduct a structural comparison of the phenolic polymeric products and natural HS using IR spectroscopy.¹³ They showed that the IR spectra of *p*-benzoquinone-based polymers only slightly resembled those of soil HS, whereas those of protocatechuic acid-based polymers looked very much like soil HS.¹³ The substantial similarity between the NMR spectra of synthetic phenolic analogues and natural HS was reported by Hanninen with co-workers¹⁰ as well as Cataldo who used *p*-benzoquinone, pyrogallol, and gallic acid as model phenolic compounds.¹⁴ The authors also noted that carboxylic groups were found in the resulting synthetic polymers regardless of their presence in the initial monomer due to the partial ring opening of the diphenolic compounds.^{10,13} As a result, it was concluded that the phenol-derived polymers were assembled through rigid polyphenylenic structures with the random incorporation of carboxylic units.

Further progress in this direction has been recently made by Drosos with co-workers¹⁵ who used carboxyl-containing phenolic precursors (gallic and protocatechuic acids) and maintained controllable redox conditions during the course of the polymerization process. The authors claimed that more condensed, higher molecular weight products were obtained under reducing conditions, whereas oxidizing conditions led to the formation of fulvic acid-like polymers, which were confirmed by the general structural features revealed from the NMR spectroscopic data. The authors proposed a molecular mechanism for the polymerization, leading to the formation of these humic-like products; however, they did not confirm it by the identification of either the reaction intermediates or reaction products. Hence, further advancements in this field are needed in regard to both the mechanistic studies, which would underpin the synthetic strategy used for preparing synthetic HS with the desired structure and activity and the missing analytical tools that could be used for controlling and inventorying their molecular compositions.

In this respect, high resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) deserves particular consideration. This method emerged at the end of the 20th century as an indispensable tool for exploring complex systems due to its unprecedented resolution capacity.^{16,17} As a result, it became the method of choice for investigating the molecular composition of natural organic matter (NOM) and HS.^{18–21} However, to the best of our knowledge, it has not been used to date for characterizing synthetic HS. We believe that the application of FTICR MS will contribute to inventorying the molecular composition of synthetic HS, while the use of specific isotopic labeling techniques^{22,23} might allow for the identification of the structural formulas of their individual molecular constituents.

In this study, we have synthesized model HS using the oxidative coupling of a specifically designed phenylpropanoic monomer, 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid, to hydroquinone. The use of this precursor was, firstly, to account for the substantial contribution of ligninic units in the aromatic compartments of natural HS. Secondly, the presence of protons with different chemical environments in this precursor was used for developing a controllable H/D exchange technique followed by FTICR MS analysis. This technique allowed the identification of the exact structural formulas of the individual molecules within the synthesized HS, which facilitated mechanistic conclusions with respect to the chemical transformations of the phenylpropanoic precursors during the oxidative coupling of phenols. Information on the exact structural formulas of the individual constituents in synthetic HS is also pivotal for the prognostication of their biological activities using structure–activity relationships and other drug candidate modeling.

Experimental

Reagents

All reagents used in this study are commercially available. The solvents used in this study were purified using known techniques.²⁴ Amberlite resin XAD 8 (Rohm & Haas) was used for isolating the fulvic acid-like (FA-like) products. Ion exchanging resin, Amberlite IR 120 (H⁺) (the Dow Chemical Company), was used for desalting the alkaline fractions of the FA-like products.

Synthesis of the oxidized phenylpropanoic monomer

The synthesis of the oxidized phenylpropanoic monomer (**3**) was conducted using the three-step reaction pathway shown in Fig. 1. The synthesis of 4-ethoxycarbonyloxy-3-methoxybenzoic acid (**1**) was conducted in accordance with the study of Kaspar *et al.*²⁵ A detailed protocol and NMR characterization are provided in the ESI† Potassium ethylmalonate (EtOOCCH₂COOK) and anhydrous magnesium chloride (MgCl₂) were prepared as described by Strube²⁶ and Rieke *et al.*²⁷ (the details are given in the ESI†).

Synthesis of ethyl 3-(4-(ethoxycarbonyloxy)-3-methoxyphenyl)-3-oxopropionate (**2**). To a solution of **1** (12.48 g, 0.052 mol) in anhydrous THF (200 mL), carbonyldiimidazole (CDI) (9.30 g, 0.057 mol) was added at ambient temperature, and the mixture was stirred for 1 hour. To the obtained mixture, a solution of potassium ethylmalonate (8.84 g, 0.052 mol) and MgCl₂ (7.41 g, 0.078 mol) in THF (50 mL) was added dropwise. The obtained reaction mixture was stirred for 12 hours. Subsequently, the solvent was evaporated at reduced pressure, and the residue was dissolved in dichloromethane (DCM) and washed with 20% citric acid. The organic phase was dried over Na₂SO₄, and the solvent was evaporated *in vacuo*, followed by purification using flash chromatography (silica gel, *n*-hexane/ethyl acetate 1 : 1). Yield 8.23 g (53%). ¹H NMR (400 MHz, CDCl₃) δ: 7.85–7.16 (m, 3H, aromatic

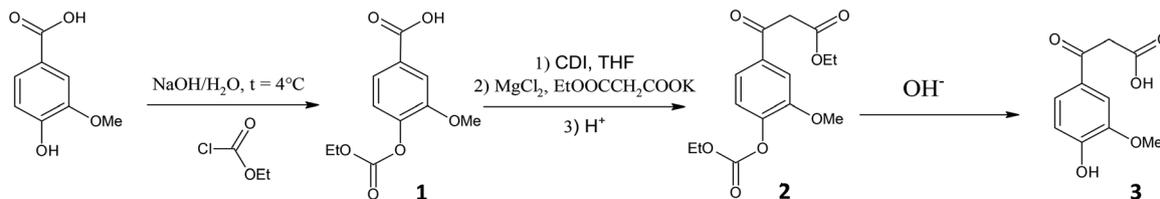


Fig. 1 Synthetic pathway for preparing an oxidized phenylpropanoic monomer M (3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid – compound 3) from vanillic acid.

protons), 4.36–4.19 (m, 4H, OCH_2CH_3), 3.98 (s, 2H, $\text{C}(\text{O})\text{CH}_2$), 3.91 (s, 3H, OCH_3), 1.40–1.23 (m, 6H, OCH_2CH_3); ^{13}C NMR (100 MHz, CDCl_3) δ : 191.3 ($\text{C}=\text{O}$), 167.3 (CH_2COOEt), 152.5 ($\text{C}_6\text{H}_3\text{OCOOEt}$), 151.6, 144.4, 134.8, 122.5, 122.0 and 111.9 (aromatic carbons), 65.3 (OCH_2CH_3), 61.6 (CH_2CH_3), 56.1 (OCH_3), 45.9 (CH_2COOEt), 14.1 (CH_2CH_3), 14.0 (CH_2CH_3). Elemental analysis %: found H 5.82, C 58.04. $\text{C}_{15}\text{H}_{18}\text{O}_7$, calc. H 5.85, C 58.06.

Sodium 3-(3-methoxy-phenyl)-3-oxopropionate (3) was synthesized by the hydrolysis of 2. For this purpose, an aliquot of 2 (1 g, 0.003 mol) was added to a 3 M solution of NaOH (100 mL) and refluxed for 2 hours. The reaction mixture was cooled down and diluted with water (1:3 by volume). The obtained compound 3 was used as a solution without isolation.

Synthesis of model humic substances (HS) using oxidative coupling

Hydroquinone (0.66 g, 0.006 mol) was added to the alkaline solution of 3, which was diluted three times. The reaction mixture was heated up to 60 °C. After one hour, potassium persulfate, $\text{K}_2\text{S}_2\text{O}_8$, in large excess (12.5 g, 0.046 mol) was added as an oxidant, as described by Eller,¹¹ and stirred for one more hour. Then, the reaction mixture was cooled down and the HS-like products were isolated as described below.

Isolation of the model HS

The humic acid (HA)-like fraction was precipitated from the obtained reaction mixtures by acidification with HCl up to pH 2 in accordance with the International Humic Substances Society (IHSS) protocol.²⁸ The precipitate was separated by centrifugation, washed with 0.1 M HCl and dried in vacuum oven. The HA-like product obtained was designated MHQ-HA. The residual acidic supernatant was discharged through Amberlite XAD8 resin, as described by Aiken *et al.*²⁹ The FA-like product was eluted using 0.1 M NaOH and desalted using cation-exchanging resin in the H-form. It was dried under reduced pressure. The corresponding product was designated MHQ-FA.

H/D exchange reaction of MHQ-FA

Solutions of 300 μL of 4 M NaOD or 16% DCl in D_2O and 5 mg of MHQ-FA were heated at 120 °C for 40 hours in sealed tubes.³⁰ After this step, the solvent was evaporated under vacuum in case of DCl. The solution of labeled compounds in

NaOD was acidified until pH 2 and was isolated using XAD 8 as described for MHQ-FA.

Elemental analyses (C, H) were performed using a Vario EL analyzer (Germany).

^1H and ^{13}C NMR spectroscopy

^1H and ^{13}C NMR spectra were acquired using a Bruker Avance 400 NMR spectrometer operating at 400 MHz proton frequency.

The ^1H NMR spectra of the synthetic compounds were acquired in a 5 mm tube using 90° excitation pulses ($90(^1\text{H}) = 9 \mu\text{s}$ relaxation delay, 100 scans). 15 mg of synthetic HS was dissolved in deuterated dimethylsulfoxide (DMSO-d_6) for ^1H NMR analysis. As a reference for proton assignments, the signal of the residual protons of DMSO-d_6 located at 2.5 ppm was used. Fourier transformation, phase correction and integration were performed using ACD-labs software Version 10 (Advanced Chemistry Development, Canada). Chemical shifts in the spectra are given in ppm relative to internal Me_4Si . To detect both exchangeable and backbone protons in the synthesized compounds, the original sample preparation technique was used.³¹ In brief, prior to analysis, hygroscopic water was removed from the samples under reduced pressure using a vacuum pipeline. This procedure is necessary because the HS samples readily absorb water from the air, and the content of this hygroscopic water may reach 12% depending on the air humidity. The dried samples were dissolved in an anhydrous aprotic solvent, DMSO-d_6 , and the ^1H NMR spectra were acquired before and after the addition of 20 μL of deuterated trifluoroacetic acid.

FTICR mass spectrometry

FTICR mass spectra were acquired using a commercial 7 Tesla LTQ FT Ultra mass spectrometer equipped with an Ion Max Electrospray Ion source (Thermo Electron Corp., Bremen, Germany) located at the facilities of the Institute of Biochemical Physics of RAS (Moscow, Russia). The samples were dissolved in methanol at concentrations of 1 g L^{-1} . Electrospray ionization (ESI) was used with the following conditions: flow rate 1 $\mu\text{L min}^{-1}$, negative ion mode; needle voltage –3 kV; no sheath and auxiliary gas flow; tube lens voltage 130 V; heated capillary temperature 200 °C. Full-scan MS spectra (m/z 200–2000) were acquired in the FTICR with a resolution $R = 400\,000$ at m/z 400. The automatic gain control (AGC) target for

FTICR MS was set to 1×10^6 , corresponding to the number of ions accumulated in the linear ion trap and transferred to the ICR cell. The maximum injection time to fill the linear ion trap was set at 500 ms. The average FTICR mass spectrum was a sum of 400 consecutive scans. The LTQ FT tuning mix was used for external mass calibration. The FTICR MS data were processed using the lab-made “Transhumus” software designed by A. Grigoriev, which is based on the total mass difference statistics algorithm.^{32,33} Error threshold in formula assignments was set to ± 0.5 ppm. For all ions, the mass accuracy (measured as the root mean squared (rms) errors for the given mass) was below 1 ppm in the mass range from 300 to 900 m/z . The rms values for the assigned formulas are given in Table S1 in the ESI.†

Calculation of the H/D exchange series

Data processing was described in our previous work.^{22,23} In brief, using “Transhumus” software we arranged the data in the following tabular format:

$$T_{\text{original}} = \{m_i, I_i, c_i, (h-1)_i, o_i\},$$

where m_i is the mass of the i -th identified peak, I_i is the i -th peak intensity, c_i , $(h-1)_i$, and o_i are the elemental compositions of the identified ions. A neutral CHO molecule has a molecular composition of c_i , h_i , and o_i .

For each formula from T_{original} , there should be related peaks in the corresponding isotope-exchange spectra. To identify those peaks, for each mass m_i from T_{original} , we select all peaks M_n^i from T_{exchange} such that

$$(M_n^i - m_i) - kd < E,$$

where the integer k spans the region 0, 1, ..., K . Here k is the maximum possible number of exchanges, and d is the mass difference, which is equal to 1.006277 for H-D exchange. E is the error set by the user. In our calculations we used $E = 10^{-3}$, which is less than 1 ppm for the experimental mass range. For each m_i , we analyzed the extracted peaks $T_{\text{extracted}}^i = \{M_n^i, I_n^i\}$ to determine the maximum number of exchanges. We performed this step manually for the most abundant peaks by plotting spectra of $T_{\text{extracted}}^i$ and analyzing them visually.

Results and discussion

Preparation of the HS-like materials: synthesis and reaction pathways

Given the importance of lignins in the formation of terrestrial HS from plant debris, we have synthesized an oxidized phenylpropanoic monomer carrying a carboxyl group to comply with the general structural features of oxidized lignin fragments within the humic molecular ensemble.^{3,34,35} Such a monomer (**M**) was obtained from vanillic acid in three steps, as shown in Fig. 1. Intermediate compounds **1** and **2** were isolated as solids with confirmed structures (see Experimental section). A phenylpropanoic monomer, 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid (compound **3**), was obtained *in situ* by the hydrolysis of

compound **2** and as such represented an oxidized derivative of coniferyl monolignons constituting lignin of gymnosperms.³⁵

To prepare HS-like materials, the synthesized monomer (**M**) was oxidatively coupled to hydroquinone (HQ) under alkaline conditions using potassium persulfate in large excess as the oxidant.¹¹ The choice of hydroquinone as the counterpart was to circumvent any preferential recombination of the coniferyl radicals and thus to obtain humic-like products.³⁵ Based on the literature data and the results obtained in this work, we surmised the following reaction pathways leading to the formation of model HS under the conditions used in this study (Fig. 2).^{36–39}

Owing to the electrophilic nature of the sulfate radical, it was expected to attack the electron rich atoms, *e.g.*, oxygen carrying negative charge, as well as the aromatic carbons in the *ortho* and *para* positions to the OH group, as shown in Fig. 2A. Because the sulfate radical easily leaves the aromatic ring, it gets eliminated to form carbon-centered radicals *via* electron transfer from the substrate to the sulfate radical, and then hydrolysis leads to the formation of hydroxylated products.³⁸ This reaction is known as Elbs oxidation. However, in the presence of excess persulfate, the oxidative coupling of phenoxy radicals becomes the major reaction pathway, leading to the formation of polymeric humic-like products linked *via* both C–C and C–O–C bonds.³⁹ This oxidative coupling also includes phenoxy radical recombination, leading to the formation of dimers and oligomers. Given the strong oxidizing conditions, it is also accompanied by ring cleavage and decarboxylation processes, which are not shown in Fig. 2 due to their poorly predictable character. Together these processes lead to a very complex mixture of reaction products. Nonetheless, the reaction pathways, shown in Fig. 2, enable the identification of the major structural patterns, which might be present in the humic-like products obtained in this study.

The obtained model HS were fractionated into humic acid (MHQ-HA)- and fulvic acid (MHQ-FA)-like fractions by precipitating the acid-insoluble HA-like fraction and extracting the acid-soluble FA-like fraction on Amberlite XAD8 resin. The amount of MHQ-HA was 180 mg *versus* 730 mg for MHQ-FA, which is indicative of the low polymerization degree of the model HS obtained under the conditions used in this study. This can be connected to the relatively short reaction time (1 hour), which was used for the oxidative coupling. The elemental compositions of MHQ-HA and MHQ-FA are given in Table 1.

The results of the elemental analysis show the rather high aromaticity of both products, which is consistent with the type of precursors used. The MHQ-HA product is more aromatic and less oxidized when compared to MHQ-FA. This was to be expected from the lower solubility of MHQ-HA in an acidic solution. The same trend is valid for HA and FA from natural sources.

FTICR MS studies on the model HS

To characterize the molecular compositions of the obtained model HS, ESI FTICR MS was used. The corresponding mass-

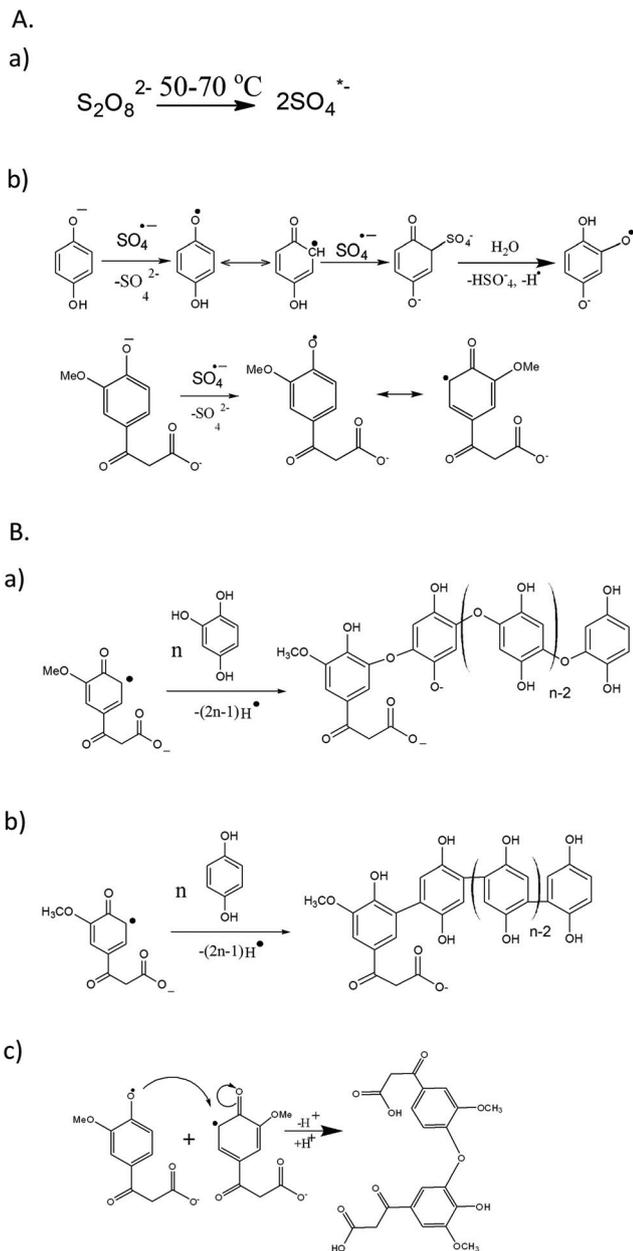


Fig. 2 The possible reaction pathways for the oxidative coupling of 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid (M) and hydroquinone (HQ) in the presence of a large excess of persulfate ion in an alkaline medium. (A) Elbs oxidation of hydroquinone and 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid: (a) thermal decomposition of potassium persulfate in alkaline medium: (b) formation of free phenoxy radicals and their hydroxylation. (B) Oxidative coupling of phenoxy radicals leading to the formation of the humic-like products: (a) *via* C–O–C bonding: (b) *via* C–C bonding: (c) *via* recombination (e.g., dimer formation).

Table 1 Content of elements (% mass) on the ash free basis and atomic ratios in the synthesized HS

Sample	%, C	%, H	%, O	H/C	O/C
MHQ-HA	67.45	5.16	27.38	0.92	0.30
MHQ-FA	52.67	4.26	43.07	0.97	0.61

spectra of the MHQ-HA and MHQ-FA samples are shown in Fig. 3A and B, respectively.

It can be seen that the obtained FTICR mass-spectra of MHQ-HA and MHQ-FA are characterized with a high peak density within the range of m/z values from 300 and 900, reaching its maximum at 400. The observed broad distributions of peaks are characteristic of the spectra reported for heterogeneous mixtures such as synthetic polyelectrolytes and natural humic materials. The spectra of the samples under study were composed of peaks with $z = 1$ and 2, which is in line with the patterns observed in natural HS.^{20,21} To avoid false identifications, we deployed the filtration of ions using S/N ratio >10. This allowed us to exclude poorly resolved peaks (some of them are clearly seen in the mass scale-expanded segments of the full FTICR MS spectra shown in Fig. 3) from further consideration.

To identify the molecular compositions of the model HS obtained in this study, the acquired FTICR MS data were used for formula assignments, which yielded about 3000 formulas (CHO-only) for each product. A full list of the corresponding assignments is given in Table S1 in the ESI.† They were further used for plotting Kendrick diagrams,⁴⁰ which are shown in Fig. 4. The left panel shows the diagrams for MHQ-HA and the right ones show those for MHQ-FA.

The CH₂-based diagrams for both samples (Fig. 4A and B) demonstrate a lack of the CH₂-homologues series, whereas the diagonals produced in the CO₂-series can be clearly seen. That is why we plotted the corresponding CO₂-based diagrams, which is shown in Fig. 4C and D. The extended CO₂-series are indicative of the intense decarboxylation processes, which took place during the oxidative coupling under the conditions used in this study. To account for hydroxylation reactions, which were expected here, we plotted oxygen-based Kendrick diagrams (Fig. 4E and F). They are characterized with the most extended series. This confirms that intense hydroxylation occurs during the oxidative polymerization of hydroquinone.⁴¹ It should be noted that both the CO₂- and O-homologues were more abundant in MHQ-FA when compared to MHQ-HA, which is in agreement with their solubility properties and elemental compositions. MHQ-FA is considerably more oxidized when compared to MHQ-HA.

To visualize the molecular space of the synthesized HS, the assigned formulas were used to calculate the H/C and O/C atomic ratios, which were plotted in the Van Krevelen diagrams shown in Fig. 5.

It can be seen that the major portion of compounds consisting of the HA-like products is located in the region of condensed tannins, whereas considerably smaller portions occupy areas assigned to lignins, hydrolysable tannins and polyhydroxy aromatic acids.^{42–44} This is indicative of the highly hydrophobic character of this fraction, which is consistent with its low solubility at acidic pH. On the other side, the major portion of compounds consisting of FA-like products is located in the area of lignins and polyhydroxy aromatic acids, which is consistent with the considerably more hydrophilic character of this fraction. The larger O/C ratios for the

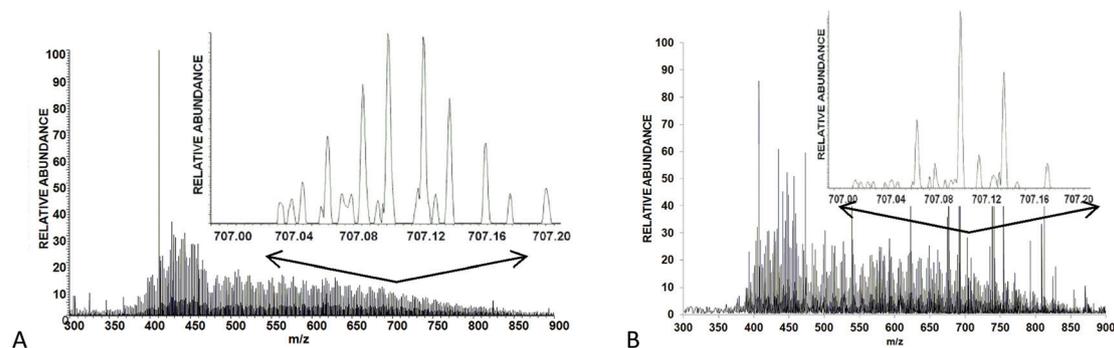


Fig. 3 ESI FTICR mass spectra of the model HS obtained *via* the oxidative coupling of the oxidized phenylpropanoic monomer (M) to hydroquinone (HQ): (A) fulvic acid-like sample (MHQ-FA) and (B) humic acid-like sample (MHQ-HA). The corresponding mass scale-expanded segments allow for visual resolution in the range of m/z from 707.000 to 707.200.

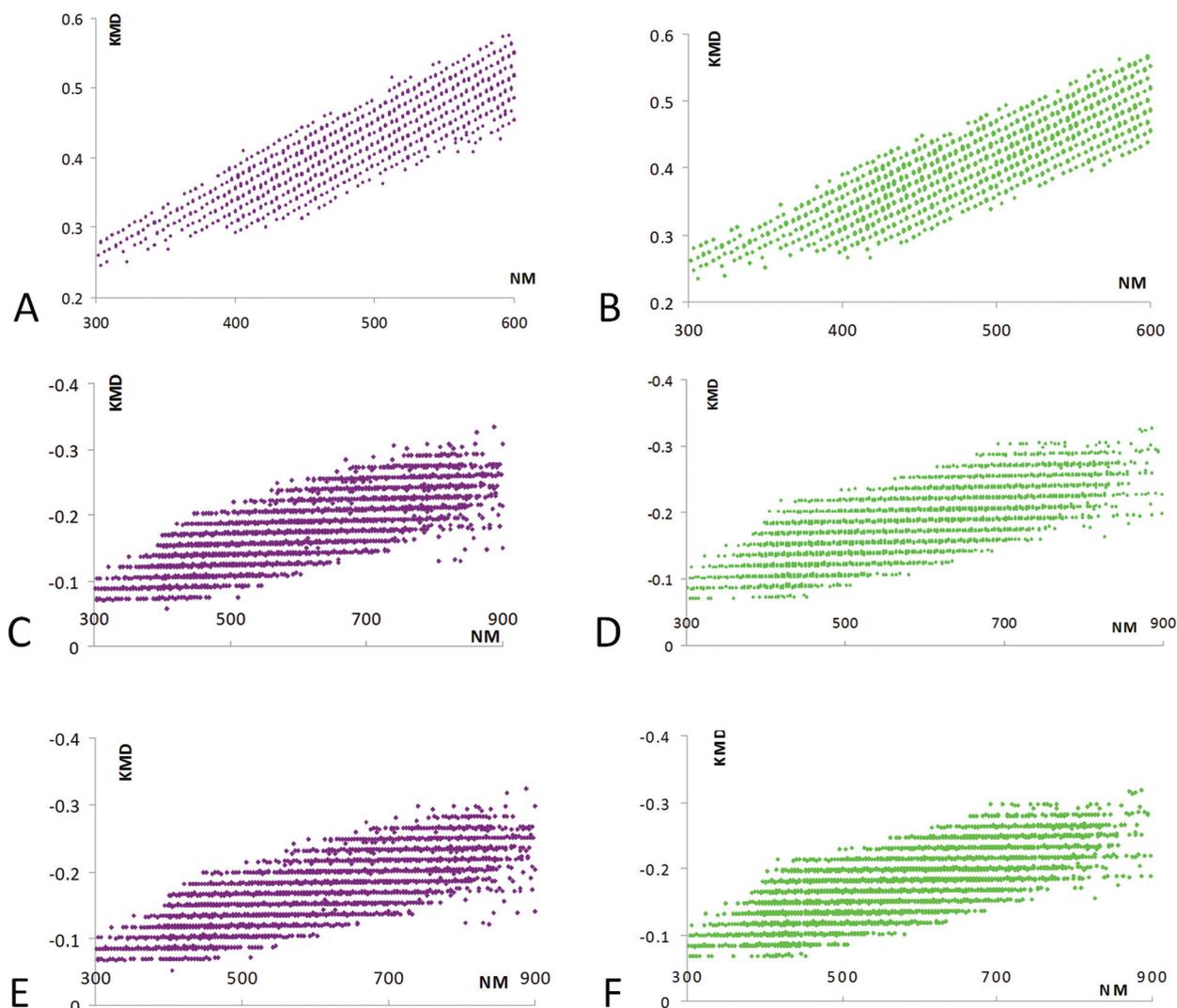


Fig. 4 Kendrick diagrams plotted using different bases for the model HS (left panel, violet dots – for MHQ-HA, right panel, green dots – for MHQ-FA): A, B – CH_2 -based diagrams, C, D – CO_2 -based diagrams, and E, F – oxygen-based diagrams. KMD stands for Kendrick mass defect and NM stands for nominal mass.

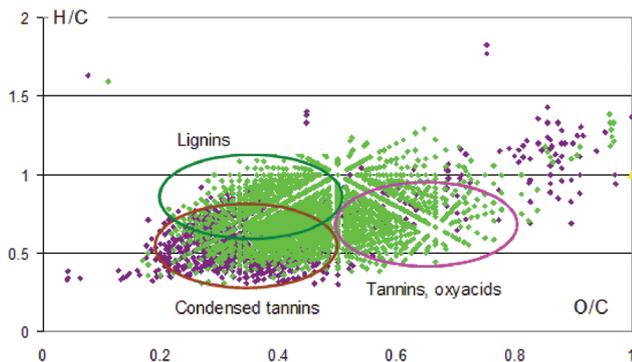
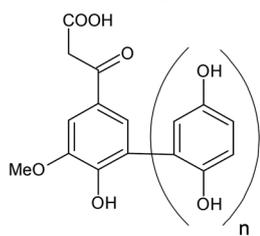


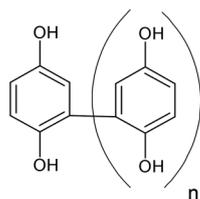
Fig. 5 Van Krevelen diagrams for the model HS obtained in this study: purple dots represent the CHO formulas belonging to the HA-like products (MHQ-HA) and green dots represent the CHO formulas belonging to the FA-like products (MHQ-FA). The brown circle shows the location of condensed tannins on the Van Krevelen diagram, the dark green circle – the location of lignins, and the lilac circle – the location of hydrolysable tannins and polyhydroxyl carbonic aromatic acids.^{42–44}

MHQ-FA sample might be indicative of the progressive hydroxylation of the aromatic rings characteristic to this product. An increase in the H/C ratio compared to the monomers observed in both copolymers could be explained by a cleavage of the aromatic rings that is followed by the formation of the oxidized aliphatic products.¹⁴

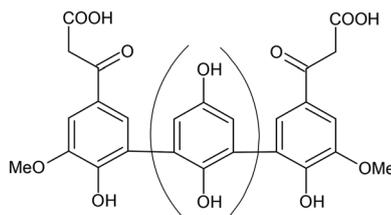
To characterize the major structural features inherent within the synthesized humic-like compounds, ¹H NMR spectroscopy was used. This method allows for the fast characterization of the different types of protons in complex mixtures such as HS.³¹ The ¹H NMR spectra (shown in Fig. S1 in the ESI†) obtained for both types of the humic-like compounds were characterized with the presence of broadened “humps” in the region of aromatic and α -CH protons, which are typical for polymers. In addition, the resolved signals were observed at following values (in ppm): 3.81 (OCH₃-groups), 3.83 (C(O)CH₂COOH), 6.93 (protons of “terminal” hydroquinone groups), 7.5 (aromatic protons of monomer M and of hydroquinone), and 2.43 (an intense singlet, which was assigned to the protons of an α -CH₃ group). From the data obtained, we could conclude that the side chain of the lignin monomer M used in this study remained intact during the oxidative coupling (the presence of strong resonances at 3.83 ppm). Hence, the formation of polymeric chains occurred mostly through the coupling of the aromatic rings, as shown in Fig. 2, and could be exemplified by the structures shown below:



4

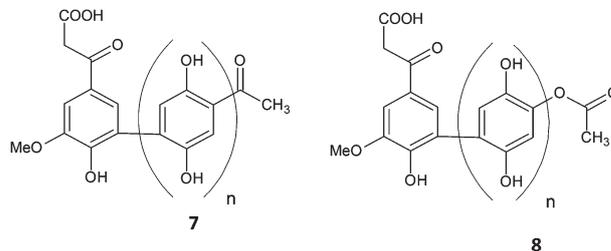


5



6

At the same time, the presence of an α -CH₃ group in the synthesized HS could be accommodated by the structures below:



8

We can suggest that these acetylated structures have been formed as a result of decarboxylation indicated by Kendrick diagrams in Fig. 4C and D.

Hence, it can be concluded that both the FTICR MS and ¹H NMR studies are in general agreement with the reaction pathways shown in Fig. 2, which lead to the formation of the humic-like compounds synthesized in this study. However, the obtained data are insufficient for the identification of the exact molecular constituents of the model HS, which was a specific goal of this study. To solve this problem, a more advanced technique should be used, which provides information on both the exact molecular mass of the compound and the positioning of its constituting atoms (*e.g.*, protons) within the backbone structures.

To get this information, we have undertaken controlled H/D exchange of the non-labile backbone protons using conditions of basic and acidic catalysis followed by FTICR MS. In designing this approach, we relied on the information known from the literature that these are α -CH, benzyl, *ortho*- and *para*-protons of the aromatic ring, which become labile under the conditions of basic catalysis, and hence they may be substituted with deuterium, whereas under the conditions of acidic catalysis, α -CH and benzyl protons remain intact, but all protons in the aromatic ring could be exchanged with deuterium.^{30,45} Hence, it gets feasible to discern between the backbone protons constituting the aromatic rings and α -CH moieties in the same molecule, as shown in Fig. 6, using the example of the phenylpropanoic monomer used in this study.

It can be seen that under the conditions of acidic catalysis (left panel in Fig. 6), all H-C_{ar} protons should undergo exchange with D-atoms, whereas the α -CH protons in the methylene group of the propanoic moiety remain intact. At the same time, under the conditions of basic catalysis (right panel

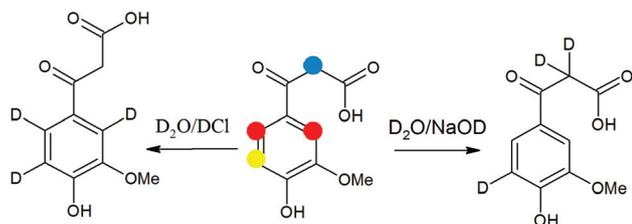


Fig. 6 H/D exchange of the non-labile backbone protons of 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid, which takes place under the conditions of acidic catalysis (in the presence of DCI, the H/D exchange positions are shown with red dots), basic catalysis (in the presence of NaOD, the H/D exchange positions are shown with blue dots), and under both acidic and basic catalysis (the H/D exchange positions are shown with yellow dots).

in Fig. 6), only the aromatic protons in the *ortho*-position to the phenolic group might undergo exchange with D-atoms as well as the α -CH protons in the methylene group of the propionic moiety; however, two aromatic protons in the *meta*-position to phenolic group remain intact. For the molecule under study, the feasible number of exchanged protons in both cases is three. This sets the length of exchange series to three under the conditions of either acidic or basic catalysis.

The number of exchanged protons for each molecule, can be determined by counting the number of mass shifts (Δm) equal to the mass difference between deuterium and protium (1.00628), which are related to a certain molecular peak, using

FTICR MS measurements. For this purpose, we compared the FTICR mass spectra of the samples under study before and after H/D exchange and inspected the length of exchange series for the selected molecular peak, as shown in Fig. 7, for $m/z = 441.08299$, and it is described in detail in the Experimental section.

The information obtained on the length of exchange series under the conditions of basic and acidic catalysis can be easily converted into the amount and positions of protons in the backbone of the investigated structures and can be further used to discern between the structural isomers. To demonstrate how it works in practice, we have examined the general structural patterns identified using ^1H NMR spectroscopy and are designated above as compounds 4 to 8. For this, we have searched the lists of molecular formulas assigned on the basis of FTICR MS data on MHQ-FA and MHQ-HA samples, which are given in Table S1 in the ESI.† In particular, we searched for the formulas that would fit the elemental compositions of the patterns from 4 to 8 by varying the number of monomeric units (n) from 1 to 2. Some examples of the molecular formulas found, which fit above the requirements, are shown in Table 2. To assign the exact chemical structures to these formulas, we have extracted the H/D exchange series related to the mass peak of the corresponding compound within the FTICR mass spectra of MHQ-HA or MHQ-FA exchanged under the conditions of acidic and basic catalysis. The extracted H/D exchange series are shown in Table 2.

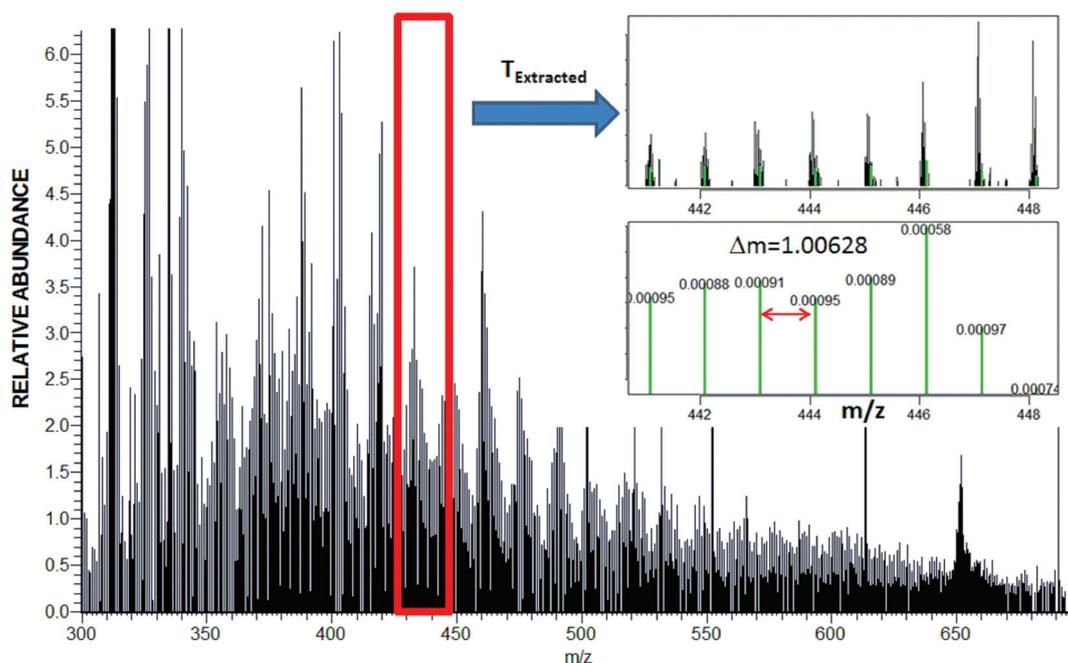


Fig. 7 FT ICR mass-spectrum of the H/D-exchanged MHQ-FA under the conditions of basic catalysis (in the presence of NaOD). The insets show the mass scale-expanded segment of the full range spectrum highlighted in red color with evident periodicity at every 1 nominal mass unit and the extracted sub-spectrum of the H/D series ($T_{\text{extracted}}$) for $m/z = 441.08299$ that has the length of the H/D exchange series equal to 6. The latter was determined by counting a number of mass shifts (Δm) of 1.00628 equal to the difference between the exact masses of deuterium and protium.

Table 2 The extracted H/D exchange sub-spectra for five selected molecular formulas and the corresponding identified structures of the individual compounds. Blue, red, and yellow dots indicate the unique exchanging centers under acidic, basic and both catalysis, respectively. The number above the peak designates the corresponding value of the root mean squared (rms) error multiplied by 10^5

Extracted spectra of H/D exchange in DCl	Identified structure	Extracted spectra of H/D exchange in NaOD
	<p>C₁₆H₁₄O₇ (m/z = 317.06601)</p> <p>9</p>	
	<p>C₂₂H₁₈O₁₀ (m/z = 441.08206)</p> <p>10</p>	
	<p>C₁₆H₁₄O₈ (m/z = 333.06093)</p> <p>11</p>	

Table 2 (Contd.)

Extracted spectra of H/D exchange in DCl	Identified structure	Extracted spectra of H/D exchange in NaOD
	<p>C₂₀H₁₈O₁₀ (m/z = 417.08206)</p>	
	<p>C₁₈H₁₈O₈ (m/z = 361.09223)</p>	

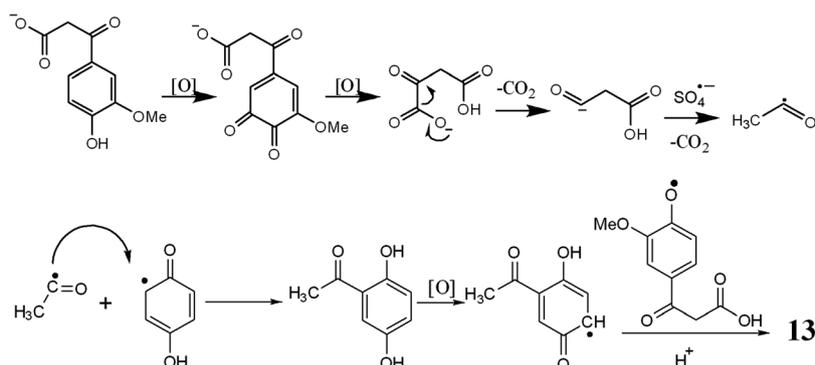


Fig. 8 The proposed reaction pathway for the formation of the acetylated compound 13.

The identified structures from 9 to 13, which are shown in Table 2, refer to the empirical formulas of C₁₆H₁₄O₇, C₂₂H₁₈O₁₀, C₁₆H₁₄O₈, C₂₀H₁₈O₁₀, and C₁₈H₁₈O₈, respectively.

One can deduce that the compound 9 is a recombination product of M and HQ radicals, and the compounds 10 and 11 are the hydroxylated isomers of the compound 4 with $n = 1$

and 2, respectively. The compound **12** is a dimer of 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid, which was used as a ligninic monomer in our studies. The compound **13** is a structural isomer of the acetylated compound **7** with $n = 1$. As a result, a set of identified compounds corroborated well with the reaction pathway of oxidative coupling, shown in Fig. 2, with the exception of the formation of acetylated products.

To explain the presence of the acyl-substituent in compound **13**, we suggested the cleavage of the aromatic ring of monomer **M** under oxidative conditions followed by the decarboxylation of the α -keto acids, which leads to an unstable acyl-carbanion.⁴⁶

The carbanion might undergo further oxidation by potassium persulfate to an acyl-radical, which recombines with the semiquinone radical. 1-(2,5-Dihydroxyphenyl)ethanone might be further coupled to **M** by phenolic radical formation reactions, as shown in Fig. 2B. The surmised reaction pathway for 1-(2,5-dihydroxyphenyl)ethanone formation is shown in Fig. 8.

The reaction pathways for the formation of the other identified products are provided in the ESI (Fig. S2†).

This shows that the use of the well-defined phenylpropanoic monomer in the oxidative coupling to hydroquinone combined with a use of high resolution isotopic exchange mass spectrometry enabled us to identify the structural formulas of the individual constituents of the synthesized HS. The identified formulas, in turn, were used for refining the reaction pathways occurring during the oxidative coupling in the presence of an excess amount of persulfate ions. In addition, they have revealed acetylation as an important side reaction resulting from the ring cleavage of the phenylpropanoic monomer used in this study, which leads to the formation of the acetylated aromatic rings. The proposed mechanism of this reaction corroborated well with intense decarboxylation processes revealed by Kendrick diagrams plotted from the FTICR MS data.

Conclusions

The synthesis of model humic substances using a specifically designed phenylpropanoic monomer and traditional synthetic strategy, which implied its oxidative coupling to hydroquinone, leads to a complex mixture of reaction products. The application of high resolution FTICR MS to the characterization of the synthesized HS showed that their molecular constituents occupied both lignin- and tannin-regions on the Van Krevelen diagram resembling closely the location of aromatic compartments in natural HS. This demonstrated that the use of the phenylpropanoic monomer contributed substantially to approach the structural patterns exhibited by natural HS in mimicking the ligninic part of their supramolecular ensemble. To make the obtained results more meaningful in the context of biomedical research, our further task was to propose an analytical tool, which would be able to identify distinct structural formulas of the molecular constituents present within this model HS. We believed that in this case the unimolecular

biosignatures might be revealed, which can be further connected to biological properties of HS. For this purpose we coupled the unprecedented resolution capacity of FTICR MS to the controllable selectivity of H/D exchange of the backbone protons constituting the humic-like molecules. This allowed us to come up with a powerful approach, which is capable of identifying the individual components present within the model HS. Despite the modest number of the identified formulas (dozens out of thousands), they already provided substantial information on the reaction mechanism under study and allowed us to refine a final step in the decarboxylation pathway leading to acetylation of the aromatic rings.

We believe that validation of the proposed approach with the use of individual compounds, as well as the use of deuterium NMR and a combination of different methods for selective isotopic labeling such as H/D and ¹⁶O/¹⁸O exchange, will yield reliable structural information, which will be sufficient to determine the individual molecular components in such complex mixtures as natural HS.^{22,23,47} These new analytical techniques along with a use of more sophisticated synthetic strategies such as controllable polymerization might start a new era in the biomedical research on HS and HS-like phenolic polymerisates based on structure–activity relationships and other drug candidate modeling.

Acknowledgements

We acknowledge the valuable comments of the anonymous reviewer, which allowed us to improve the manuscript. This study was partially supported by the Russian Foundation for Basic Research (Grant # 13-04-01853). The part of research related to FTICR mass spectrometry measurements was supported by the Russian Science Foundation (Grant # 14-24-00114).

References

- 1 P. MacCarthy, *Soil Sci.*, 2001, **166**, 738–751.
- 2 F. J. Stevenson, *Humus Chemistry: Genesis, Composition, Reactions*, John Wiley & Sons, New York, 2nd edn, 1994.
- 3 G. W. Bailey, L. G. Akim and S. M. Shevchenko, Humic Substances, in *Humic substances and chemical contaminants*, ed. C. E. Clapp, M. H. B. Hayes, N. Senesi, P. R. Bloom and P. M. Jardine, Soil Science Society of America, Madison, WI, 2001, pp. 41–73.
- 4 N. A. Kulikova, E. V. Stepanova and O. V. Koroleva, in *Use of humic substances to remediate polluted environments: from theory to practice*, ed. I. V. Perminova, K. Hatfield, N. Hertkorn, NATO Science Series: IV: Earth and Environmental Sciences, Springer Verlag, Dordrecht, The Netherlands, 2004, vol. 52, p. 285.
- 5 R. Kloeking and B. Helbig, in *Biopolymers for medical and pharmaceutical applications*, ed. A. Steinbüchel and

- R. H. Marchessault, Wiley-VCH Verlag GmbH, Weinheim, 2005, p. 3.
- 6 F. J. Lu, S. N. Tseng, M. L. Li and S. R. Shih, *Arch. Virol.*, 2002, **147**, 273–284.
- 7 R. Kloecking, B. Helbig, G. Schotz, M. Schacke and P. Wultzer, *Antiviral Chem. Chemother.*, 2002, **13**, 241–249.
- 8 J. Schneider, R. Weis, C. Manner, B. Kary, A. Werner, B. J. Seubert and U. N. Riede, *Virology*, 1996, **218**(2), 389–395.
- 9 P. Schmitt-Kopplin, D. Freitag, A. Kettrup, N. Hertkorn, U. Schoen, R. Klöcking, B. Helbig, F. Andreux and A. W. Garrison, *Analisis*, 1999, **27**, 391–396.
- 10 K. I. Hanninen, R. Kloecking and B. Helbig, *Sci. Total Environ.*, 1987, **62**, 201–210.
- 11 W. Eller, *Liebigs Ann. Chem.*, 1923, **431**, 133–161.
- 12 H. Erdtman and M. Granath, *Acta Chem. Scand.*, 1954, **8**, 811–816.
- 13 K. M. Goh and F. J. Stevenson, *Soil Sci.*, 1971, **112**, 392–400.
- 14 F. Cataldo, *Polym. Int.*, 1998, **46**, 263–268.
- 15 M. Drosos, M. Jerzykiewicz, M. Louludi and Yi. Deligiannakis, *Colloids Surf., A*, 2011, **389**, 254–265.
- 16 N. Hertkorn, C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E. M. Perdue, M. Witt and P. Schmitt-Kopplin, *Anal. Bioanal. Chem.*, 2007, **389**, 1311–1327.
- 17 A. G. Marshall, C. L. Hendrickson and G. S. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1–35.
- 18 A. C. Stenson, W. M. Landing, A. G. Marshall and W. T. Cooper, *Anal. Chem.*, 2002, **74**, 4397–4409.
- 19 R. W. Kramer, E. B. Kujawinski and P. G. Hatcher, *Environ. Sci. Technol.*, 2004, **38**, 3387–3395.
- 20 N. Hertkorn, M. Harir, B. Koch, B. Michalke and P. Schmitt-Kopplin, *Biogeosciences*, 2013, **10**, 1583–1624.
- 21 I. V. Perminova, I. V. Dubinenkov, A. S. Kononikhin, A. I. Konstantinov, A. Ya. Zhrebker, M. A. Andzhushev, V. A. Lebedev, E. Bulygina, R. M. Holmes, Y. I. Kostyukevich, I. A. Popov and E. N. Nikolaev, *Environ. Sci. Technol.*, 2014, **48**, 7461–7468.
- 22 Y. Kostyukevich, A. Kononikhin, I. Popov, O. Kharybin, I. Perminova, A. Konstantinov and E. Nikolaev, *Anal. Chem.*, 2013, **85**, 11007–11013.
- 23 Y. Kostyukevich, A. Kononikhin, A. Zhrebker, I. Popov, I. V. Perminova and E. N. Nikolaev, *Anal. Bioanal. Chem.*, 2014, **406**, 6655–6664.
- 24 X. Becker, G. Domshke and E. Fanghenel, *Organikum*, Mir Publishing Company, Moscow, 1992.
- 25 M. Kaspar, V. Hamplova, V. Novotna, M. Glogarova and P. Vanek, *J. Mater. Chem.*, 2002, **12**, 2221–2224.
- 26 R. E. Strube, *Org. Synth.*, 1963, **4**, 417.
- 27 R. D. Rieke, S. E. Bales, P. M. Hudnall, T. P. Burns and G. S. Poindexter, *Org. Synth.*, 1988, **6**, 845.
- 28 R. S. Swift, in *Methods of soil analysis. Part 3. Chemical methods*, ed. D. L. Sparks, A. L. Page, P. A. Helmke, R. H. Loeppert, P. N. Soltanpour, M. A. Tabatabai, C. T. Johnson and M. E. Sumner, SSSA Book Series 5. SSSA, Madison, WI, 1996, p. 1036.
- 29 G. R. Aiken, E. M. Thurman, R. L. Malcolm and H. F. Walton, *Anal. Chem.*, 1979, **51**, 1799–1803.
- 30 B. Lindstrom, B. Sjoquist and E. Anggard, *J. Labelled Compd.*, 1974, **10**, 187–194.
- 31 D. V. Kovalevskii, A. B. Permin, I. V. Perminova, D. V. Konnov and V. S. Petrosyan, *Vestn. Mosk. Univ., Ser. 2: Khim.*, 2000, **41**, 39–42.
- 32 E. V. Kunenkov, A. S. Kononikhin, I. V. Perminova, N. Hertkorn, A. Gaspar, P. Schmitt-Kopplin, I. A. Popov, A. V. Garmash and E. N. Nikolaev, *Anal. Chem.*, 2009, **81**, 10106–10115.
- 33 A. Gaspar, E. V. Kunenkov, R. Lock, M. Desor, I. Perminova and P. Schmitt-Kopplin, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 683–688.
- 34 R. Kim, W. Kramer and P. G. Hatcher, *Anal. Chem.*, 2003, **75**, 5336–5344.
- 35 R. Hatfield and W. Vermerris, *Plant Physiol.*, 2001, **126**, 1351–1357.
- 36 G. Moad and D. H. Solomon, *The Chemistry of Radical Polymerization*, Elsevier BV, Netherlands, 2nd edn, 2006.
- 37 H. Musso, *Angew. Chem.*, 1963, **75**, 965.
- 38 G. R. Peyton, *Mar. Chem.*, 1993, **41**, 91–103.
- 39 E. J. Behrman, *Organic Reactions*, John Wiley & Sons, Inc., 2004.
- 40 E. A. Kendrick, *Anal. Chem.*, 1963, **35**, 2146–2154.
- 41 W. Flaig and J. C. Salfeld, *Naturwissenschaften*, 1960, **47**, 516.
- 42 W. C. Hockaday, J. M. Purcell, A. G. Marshall, J. A. Baldock and P. G. Hatcher, *Limnol. Oceanogr.: Meth.*, 2009, **7**, 81–95.
- 43 E. Kujawinski and M. Behn, *Anal. Chem.*, 2006, **78**, 4363–4373.
- 44 M. M. Tfaily, S. Hodgkins, D. C. Podgorski, J. P. Chanton and W. T. Cooper, *Anal. Bioanal. Chem.*, 2012, **404**, 447–457.
- 45 J. Atzrodt, V. Derau, T. Fey and J. Zimmermann, *Angew. Chem., Int. Ed.*, 2007, **46**, 7744–7765.
- 46 R. W. Hanson, *J. Chem. Educ.*, 1987, **64**, 591–595.
- 47 M. Byrn and M. Calvin, *J. Am. Chem. Soc.*, 1966, **88**, 1916–1922.