Analyst

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

Synthesis of Model Humic Substances by Oxidative Coupling of Phenylpropanoic Monomer and Hydroquinone: 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], Irina V. Perminova[*a]

[a]Department of Chemistry, Lomonosov Moscow State University, Leninskie gory 1-3, 119991 Moscow, Russia;

[b]Institute for Energy Problems of Chemical Physics of RAS, Leninskij pr. 38 - 2, 119334 Moscow, Russia

[c]Moscow Institute of Physics and Technology, Dolgoprudnyi, 141700 Moscow Region, Russia

[d]Skolkovo Institute of Science and Technology, Novaya St. 100, 143025 Skolkovo, Russia

[e]Emanuel Institute of Biochemical Physics of RAS, Kosygina str. 4, 119991 Moscow, Russia,

*Phone/Fax: +7(495)9395546 ; E-mail: iperm@org.chem.msu.ru

*Phone/Fax: +7(916)5491047 ; E-mail: ennikolaev@rambler.ru

Abstract

The products of 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 novel approach to molecular-level analysis of phenolic polymerisates capable of inventorying molecular constituents and resolving their distinct structural formulas. For this purpose, we have synthesized the model HS using oxidative coupling of specifically designed phenylpropanoic monomer - 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid to hydroquinone. We have characterized thus synthesized model HS using high resolution Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS), ^{1}H NMR spectroscopy, and controllable hydrogen/deuterium (H/D) exchange. We succeeded in molecular inventory of the model HS. The assigned molecular formulas occupied substantial space of CHO compositions in Van Krevelen diagram with maximum density in the regions of tannins and lignins resembling those of natural HS. To identify exact structural formulas of individual constituents of 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 phenolic coupling and to identify acetylation of aromatic rings as an important side reaction. The conclusion was made that the proposed analytical approach might be used for identifying molecular carriers of biological activity within the phenolic polymerisates and, eventually, within the natural HS.

INTRODUCTION

Humic substances (HS) are natural compounds which are formed during oxidative decomposition of biomacromolecules constituting the 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 compartments are believed to contribute the most into remarkable biological activity of HS including antiviral, antibiotic, and carcinostatic effects, which have been numerically 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 much better controlled as compared to natural HS by selecting phenolic precursors thus improving targeting of their therapeutic application.[9]

A choice of synthetic strategy for preparing the synthetic HS relays mostly on the oxidative coupling of phenols as the major process of formation of HS in nature.[10] As such, this process has been intensively studied since the beginning of the 20-th century.[11,12] Goh and Stevenson were the first to conduct 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 the soil HS, whereas those of protocatechuic acid-based polymers looked very much alike to soil HS.[13] The substantial similarity between NMR spectra of synthetic phenolic analogues and those of natural HS was reported by Hanninen with coworkers[10] and 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 partial ring opening of diphenolic compounds.[10,13] As a result, it was concluded that the phenol-derived polymers were assembled by rigid polyphenylenic structures with random incorporation of carboxylic units.

Further progress in this direction has been recently made by Drosos with coworkers[15] who used carboxyl-containing phenolic precursors (gallic and protocatechuic acids) and maintained controllable redox conditions during the course of polymerization. The authors claimed that more condensed, higher molecular weight products were obtained under reducing conditions, whereas oxidizing conditions lead to formation of fulvic acid – like polymers as confirmed by the general structural features revealed by the data of NMR spectroscopy. The authors proposed molecular mechanism of polymerization leading to formation

of these humic-like products, however, they did not confirm it by identification of either reaction intermediates or reaction products. Hence, further advancements in this field are needed with regard both to mechanistic studies which would underpin synthetic strategy used for preparing synthetic HS with desired structure and activity, and to missing analytical tools 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 has 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 in investigating molecular compositions of natural organic matter (NOM) and HS.[18-21] However, to our knowledge, it has not been used so far for characterizing the synthetic HS. We believe that application of FTICR MS will contribute to inventorying molecular composition of synthetic HS, while a use of specific isotopic labeling techniques[22,23] might allow for identification of structural formulas of their individual molecular constituents.

In this study we have synthesized the model HS using oxidative coupling of the specifically designed phenylpropanoic monomer - 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid to hydroquinone. A 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 controllable H/D exchange technique followed by FTICR MS analysis. This technique allowed for identification of exact structural formulas of individual molecules within the synthesized HS which facilitated mechanistic conclusions with respect to chemical transformations of phenylpropanoic precursors during oxidative coupling to phenols. Information on exact structural formulas of the individual constituents of the synthetic HS is also pivotal for prognostication of their biological activities using structure – activity relationships and other drug candidate modeling.

EXPERIMENTAL

Reagents. All reagents used in this study were commercially available. Solvents used in this study were purified using known techniques.[24] Amberlite resin XAD 8 (Rohm & Haas) was used for isolating 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 FA-like products.

Synthesis of the Oxidized Phenylpropanoic Monomer. Synthesis of the oxidized phenylpropanoic monomer (**3**) was conducted using the three-step reaction pathway shown in Fig. 1. Synthesis of 4-ethoxycarbonyloxy-3-methoxybenzoic acid **(1)** was conducted in accordance with Kaspar et al.[25] The detailed protocol and NMR identifications are provided in the Supplementary material. Potassium ethylmalonate ($EtOOCCH_2COOK$) and anhydrous magnesium chloride ($MgCl_2$) were prepared as described by Strube[26] and Rieke et al. [27] (the details are given in ESI).

Synthesis of ethyl 3-(4-(ethoxycarbonyloxy)-3-methoxy-phenyl)-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, the mixture was stirred for 1 hour. To the mixture obtained the solution of potassium ethylmalonate (8.84 g, 0.052 mol) and $MgCl_2$ (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, the residue was dissolved in dichloromethane (DCM) and washed by 20% citric acid. The organic phase was dried over Na_2SO_4, the solvent was evaporated in vacuo followed by purification by flash chromatography (silicagel, *n*-hexane/ethyl acetate 1:1)**.** Yield 8.23 g (53%). ^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.85-7.16 (m, 3H, aromatic protons), 4.36-4.19 (m, 4H, $OC\underline{H}_2CH_3$), 3.98 (s, 2H, $C(O)C\underline{H}_2$), 3.91 (s, 3H, OCH_3), 1.40-1.23 (m, 6H, $OCH_2C\underline{H}_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 191.3 ($\underline{C}$=O), 167.3 ($CH_2\underline{C}OOEt$), 152.5 ($C_6H_3O\underline{C}OOEt$), 151.6, 144.4, 134.8, 122.5, 122.0 and 111.9 (aromatic carbons), 65.3 ($O\underline{C}H_2CH_3$), 61.6 ($\underline{C}H_2CH_3$), 56.1 (OCH_3), 45.9 ($\underline{C}H_2COOEt$), 14.1 ($CH_2\underline{C}H_3$), 14.0 ($CH_2\underline{C}H_3$). Elemental analysis %: found H 5.82, C 58.04. $C_{15}H_{18}O_7$, calc. H 5.85, C 58.06.

Sodium 3-(3-methoxy-phenyl)-3-oxopropionate **(3)** was synthesized by hydrolysis of **2.** For this purpose an aliquot of **2** (1 g, 0.003 mol) was added to a 3M 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 alkaline solution of **3** diluted threefold. The reaction mixture was heated up to 60 °C. After one hour, potassium persulfate $K_2S_2O_8$ 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. Humic acid (HA)-like fraction was precipitated from the obtained reaction mixtures by acidification with HCl 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 obtained HA-like product was designated MHQ-HA. The residual acidic supernatant was discharged through Amberlite XAD8 resin as described by Aiken et al.[29] FA-like product was eluted using 0.1 M NaOH and desalted using cation-exchanging resin in H-form. It was dried under reduced pressure. The corresponding product was designated MHQ-FA.

H/D exchange reaction of MHQ-FA. The solutions of 300 µl of 4M NaOD or 16% DCl in D_2O and 5 mg of MHQ-FA were heated at 120 ˚C during 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 it was isolated using XAD 8 as it is described for MHQ-FA.

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

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

The ^{1}H NMR spectra of synthetic compounds were acquired in a 5 mm tube using 90 excitation pulses (90(^{1}H) = 9 μs relaxation delay, 100 scans). 15 mg of synthetic HS were dissolved in deuterated dimethylsulfoxide (DMSO-d_6) for ^{1}H NMR analysis. As a reference for proton assignments, a signal of 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 vacuum pipeline. This procedure is necessary while HS samples readily absorb water from air, and the content of this hygroscopic water may reach 12% depending on air humidity. The dried samples were dissolved in anhydrous aprotic solvent – DMSO-d_6, and ^{1}H NMR spectra were acquired before and after 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 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 $g \cdot L^{-1}$. Electrospray ionization (ESI) was used at the following conditions: flow rate 1 $\mu L \cdot min^{-1}$, 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 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. Maximum injection time to fill the linear ion trap was set to 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 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 H/D exchange series. Data processing was described in our previous work.[22,23] In brief, using "Transhumus" software we arranged data in the following tabular format:

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

where

m_i is the mass of the i-th identified peak,

I_i is the i-th peak intensity,

c_i, $(h\text{-}1)_i$, and o_i are the elemental compositions of the identified ions. A neutral CHO molecule has a molecular composition of c_i, h_i, 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 $T_{original}$, choose all peaks M^i_n from $T_{exchange}$ such that:

$(M^i_n - m_i) - k \cdot d < E$,

where the integer k spans the region 0, 1,…, K. Here K is the maximum possible number of exchanges, 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 $T^i_{extracted} = \{M^i_n, I^i_n\}$ to determine the maximum number of exchanges. We performed this step manually for most abundant peaks, plotting a spectra of $T^i_{extracted}$ and analyzing them visually.

RESULTS AND DISCUSSION

Preparation of the HS-Like Materials: Synthesis and Reaction Pathways. Given importance of lignins in formation of terrestrial HS from the plant debris, we have synthesized an oxidized phenylpropanoic monomer carrying carboxyl group to comply with general structural features of oxidized lignin fragments within humic molecular ensemble.[3,34,35] Such a monomer (M) was obtained from vanillic acid in three steps as it is shown in Fig. 1. Intermediate compounds **1** and **2** were isolated as solids with confirmed structure (see Experimental section). A phenylpropanoic monomer - 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid (Compound 3) was obtained *in situ* by hydrolysis of compound **2** and as such, represented an oxidized derivative of coniferylic monolignons constituting lignin of gymnosperms.[35]

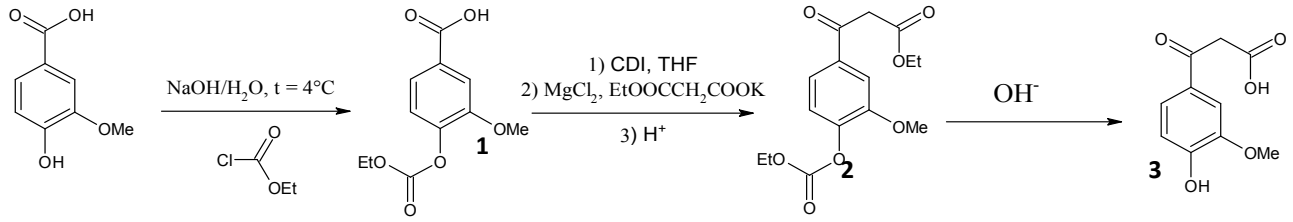
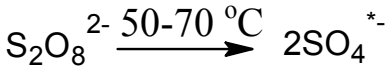


Fig. 1 Synthetic pathway for preparing an oxidized phenylpropanoic monomer M (3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid – compound **3**) from vanillic acid.

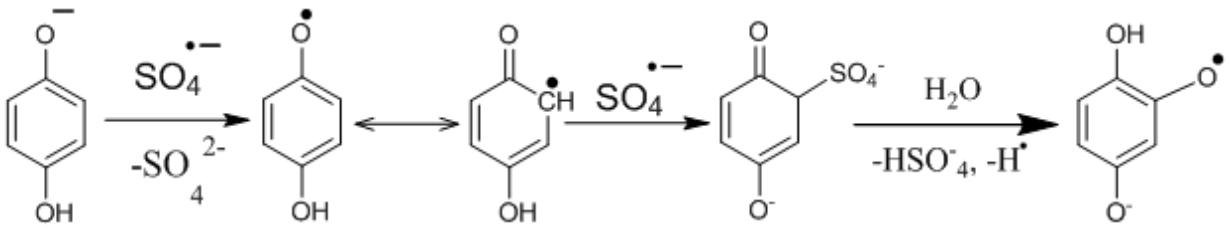
To prepare HS-like materials, the synthesized monomer (M) was oxidatively coupled to hydroquinone (HQ) under alkaline conditions using potassium persulfate in large excess as an oxidant.[11] A choice of hydroquinone as a counterpart was to circumvent preferential recombination of the coniferylic radicals, 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 formation of model HS under conditions used in this study (Fig. 2)[36-39].

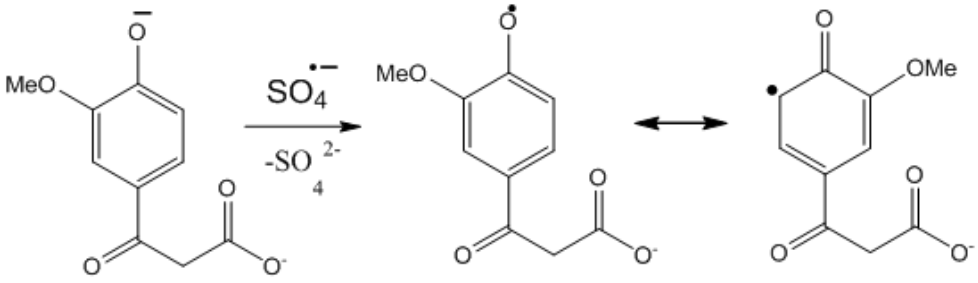
A. Elbs oxidation of hydroquinone and of 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid:

a) Thermal decomposition of the potassium persulfate in alkaline medium:



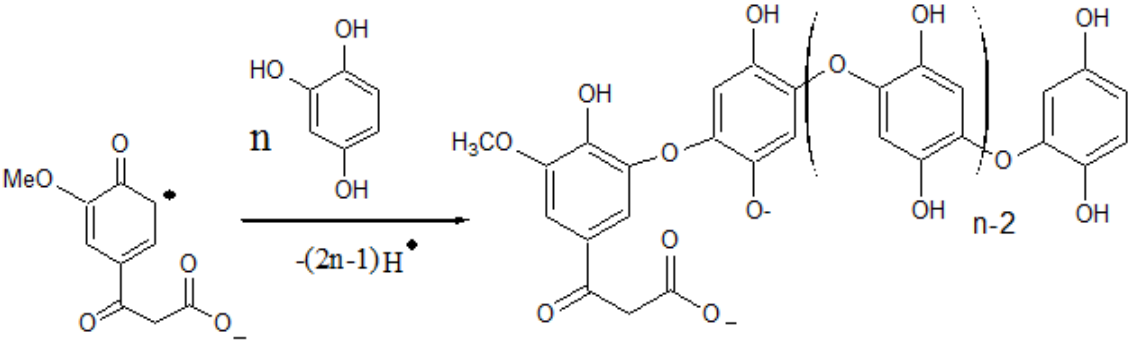
b) Formation of free phenoxy radicals and their hydroxylation:



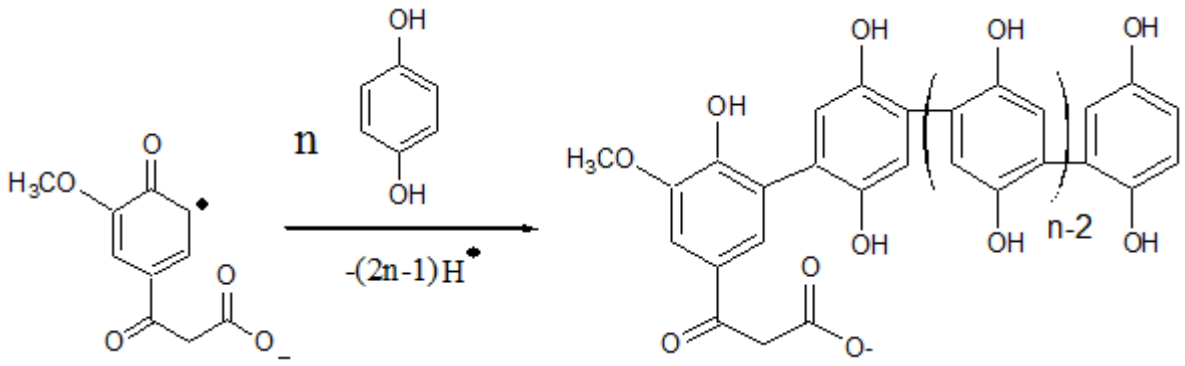


B. Oxidative coupling of phenoxy radicals leading to formation of humic-like products:

a) via C-O-C bonding:



b) via C-C bonding:



c) via recombination (e.g., dimer formation):

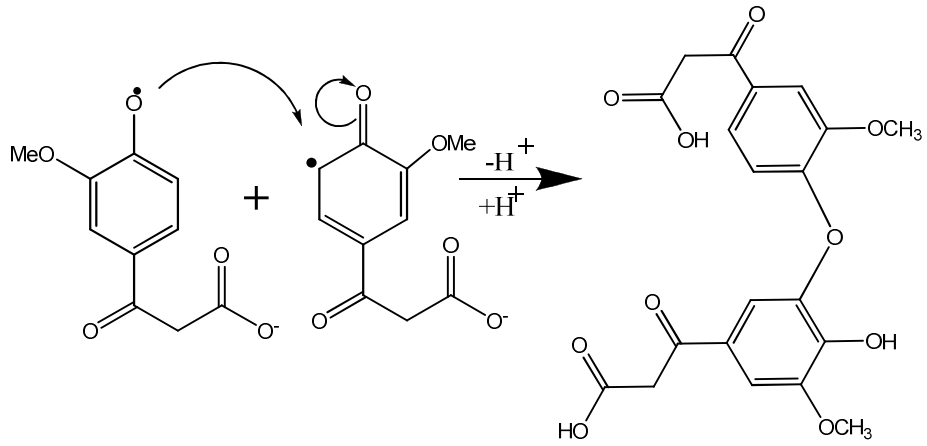


Fig. 2 The possible reaction pathways of oxidative coupling of 3-(4-hydroxy-3-methoxyphenyl)-3-oxopropionic acid (M) and hydroquinone (HQ) in the presence of large excess of persulfate ion in alkaline medium.

Owing to the electrophilic nature of sulfate radical, it was expected to attack on electron rich atoms, e.g., oxygen carrying negative charge, as well as aromatic carbon in ortho and para positions to OH group, as shown in Fig. 2A. Because the sulfate radical easily leaves the aromatic ring, it eliminates to form the carbon-centered radicals via electron transfer from the substrate to the sulfate radical, and then the hydrolysis leads to formation of hydroxylated products.[38] This reaction is known as Elbs oxidation. However, in the presence of persulfate excess, oxidative coupling of phenoxy radicals becomes the major reaction

pathway leading to formation of polymeric humic-like products linked via both C-C and C-O-C.[39] This oxidative coupling includes also phenoxy radical recombination leading to formation of dimers and oligomers. Given strong oxidyzing conditions, it is also accompanied by ring cleavage and decarboxylation processes, which are not shown in Fig. 2 due to their poorly predictable character. All together these processes lead to very complex mixture of reaction products. Nonetheless, the reaction pathways shown in Fig. 2 enable definition 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 acid-insoluble HA-like fraction and extracting 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 low polymerization degree of the model HS obtained under conditions used in this study. This can be connected to relatively short reaction time (1 hour), which was used for oxidative coupling. Elemental compositions of MHQ-HA and MHQ-FA are given in Table 1.

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

The results of elemental analysis show 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 as compared to MHQ-FA. This was to expect from lesser solubility of MHQ-HA in acidic solutions. The same trend is valid for HA and FA from natural sources.

FTICR MS studies on the model HS. To characterize molecular compositions of the obtained model HS, ESI FTICR MS was used. The corresponding mass-spectra of MHQ-HA and MHQ-FA samples are shown in Figs. 3A and B, respectively.

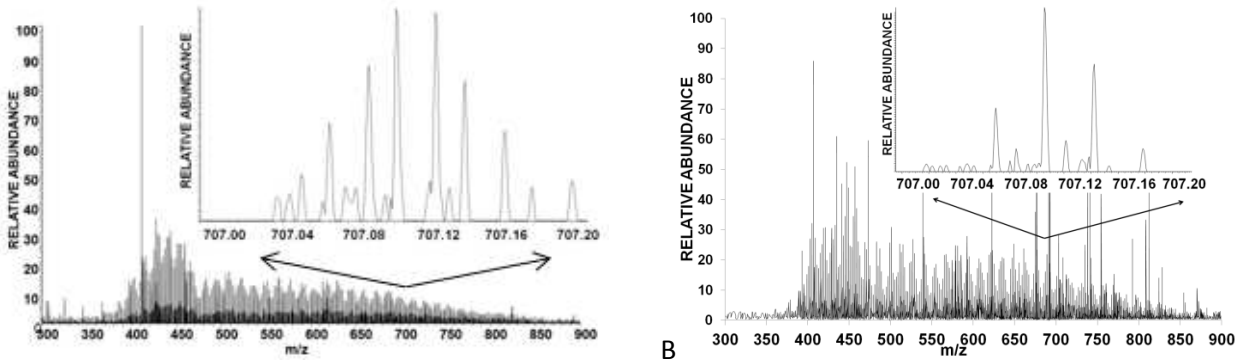


Fig. 3. ESI FTICR mass spectra of the model HS obtained via oxidative coupling of the oxidized phenylpropanoic monomer (M) to hydroquinone (HQ): A) fulvic acid-like sample (MHQ-FA), and B) humic acid-like sample (MHQ-HA), and the corresponding mass scale-expanded segments allowing for visual resolution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

It can be seen that the obtained FTICR mass-spectra of MHQ-HA and MHQ-FA are characterized with 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 spectra reported for heterogeneous mixtures such as synthetic polyelectrolytes and natural humic materials. 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 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 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. Left panel shows diagrams for MHQ-HA, and right ones – for MHQ-FA.

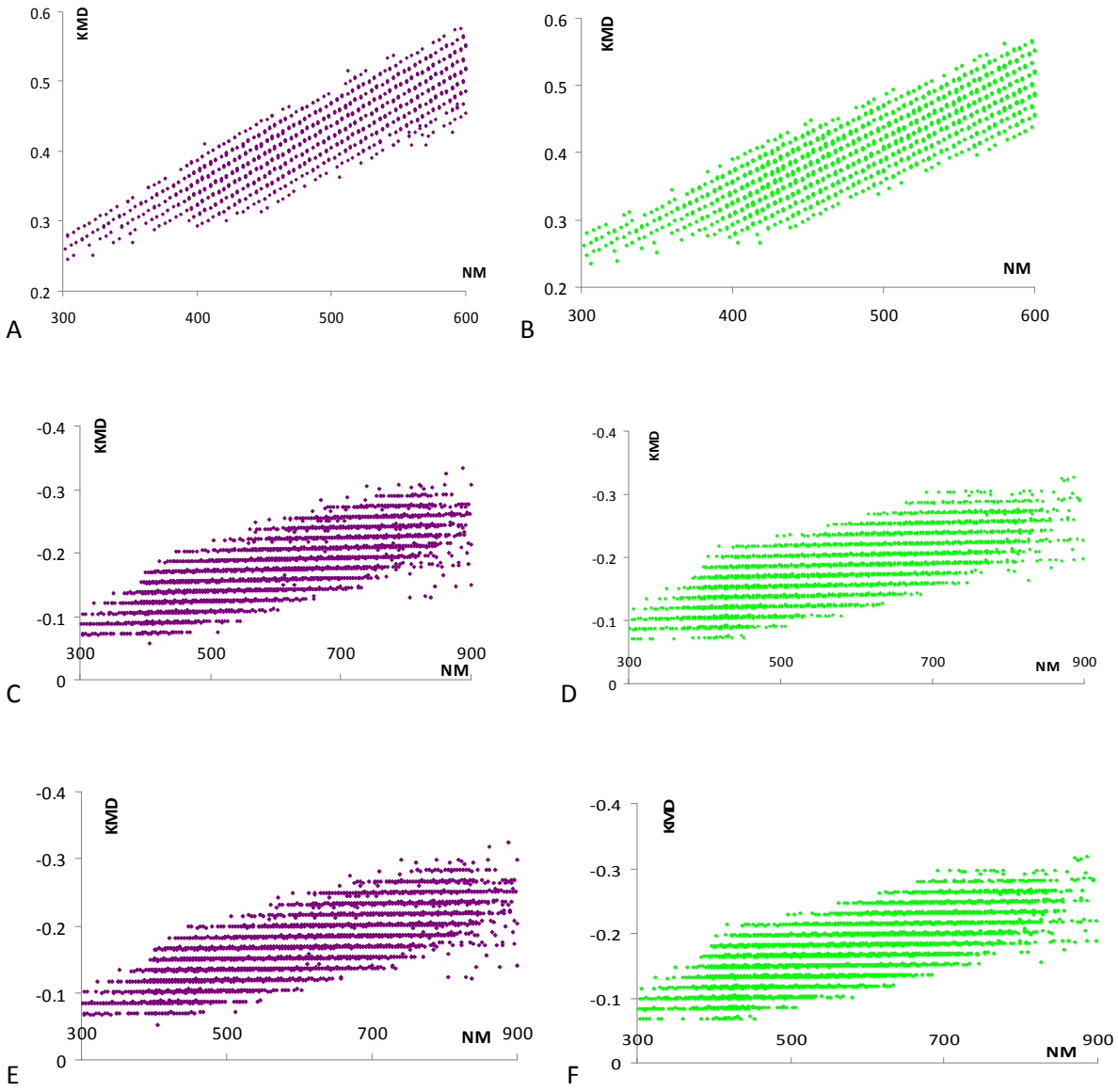


Fig. 4 Kendrick diagrams plotted using different bases for the model HS (left panel, violet dots – for MHQ-HA, right panel, green dots – for MHQ-FA): A, B - CH_2-based diagrams, C, D - CO_2-based diagrams, E, F - oxygen-based diagrams. KMD stands for Kendrick mass defect, NM stands for nominal mass.

The CH_2-based diagrams for the both samples (Fig. 4A,B) demonstrate a lack of the CH_2- homologues series, whereas the diagonals produced by the CO_2-series can be clearly seen. That is why we plotted the

corresponding CO_2 –based diagrams shown in Fig. 4C,D. The extended CO2-series are indicative of intense decarboxylation processes which took place during oxidative coupling under conditions used in this study. To account for hydroxylation reactions which were to expect here, we plotted oxygen-based Kendrick diagrams (Fig. 4E,F). They are characterized with the most extended series. This confirms intense hydroxylation occurring during oxidative polymerization of hydroquinone.[41] It should be noted that both CO_2- and O-homologues were more abundant in MHQ-FA as compared to MHQ-HA which is in agreement with their solubility properties and elemental compositions: MHQ-FA is much more oxidized as compared to MHQ-HA.

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

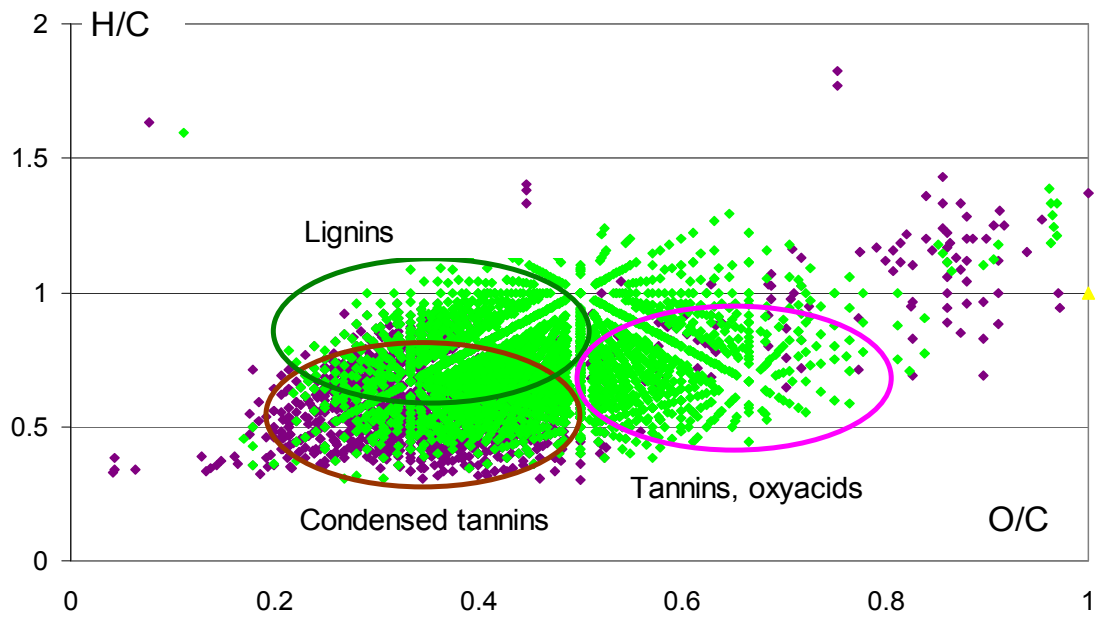
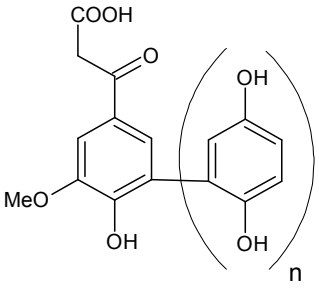


Fig. 5 Van Krevelen diagrams for the model HS obtained in this study: purple dots represent CHO formulas belonging to HA-like product (MHQ-HA), green dots represent CHO formulas belonging to FA-like products (MHQ-FA). Brown circle shows location of condensed tannins on Van Krevelen diagram, dark green circle –

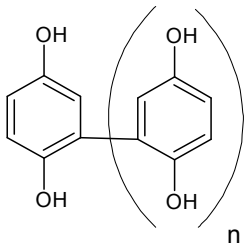
location of lignins, and lilac circle – location of hydrolysable tannins and polyhydroxyl carbonic aromatic acids[42-44].

It can be seen that the major portion of compounds consisting the HA-like product is located in the region of condensed tannins, while much smaller portions occupy areas assigned to lignins and hydrolysable tannins and polyhydroxy aromatic acids.[42-44] This is indicative of 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 FA-like product is located in the area of lignins and polyhydroxy aromatic acids which is consistent with much more hydrophilic character of this fraction. The larger O/C ratios of the MHQ-FA sample might be indicative of progressive hydroxylation of aromatic rings characteristic to this product. An increase in H/C ratio compared to the monomers observed in both copolymers could be explained by a cleavage of aromatic rings that is followed by formation of the oxidized aliphatic products.[14]

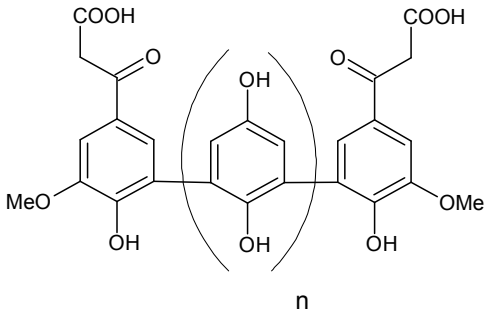
To characterize major structural features inherent within the synthesized humic-like compounds, ^{1}H NMR spectroscopy was used. This method allows for fast characterization of different types of protons in complex mixtures such as HS.[31] The ^{1}H NMR spectra (shown in Fig. S1 in the ESI) obtained for both type 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 (in ppm): 3.81 (OCH_3-groups), 3.83 (C(O)$\underline{CH_2}$COOH), 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 protons of α-CH_3 group). From the data obtained we could conclude that the side chain of the lignin monomer M used in this study remained intact during oxidative coupling (the presence of strong resonances at 3.83 ppm). Hence, formation of polymeric chains occurred mostly through coupling of the aromatic rings as it was suggested in Fig. 2, and could be exemplified by the structures shown below:



4

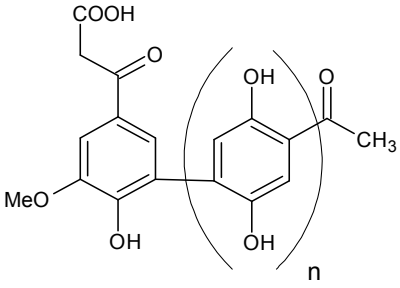


5

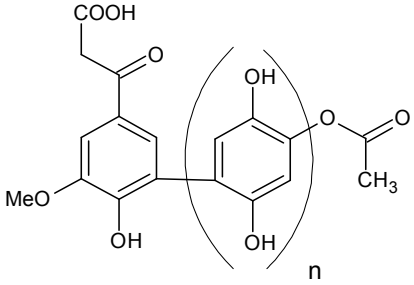


6

At the same time, the presence of α-CH_3 group in the synthesized HS could be accommodated by the structures below:



7



8

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

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

To get this information, we have undertaken controlled H/D exchange of 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 conditions of basic catalysis, and they, hence, may be substituted with deuterium; whereas under conditions of acidic catalysis, α-CH and benzyl protons remain intact, but all protons of the aromatic ring could be exchanged with deuterium.[30,45] Hence, it gets feasible to discern between backbone protons constituting aromatic ring and alpha-CH moieties of the same molecule as it is shown in Fig. 6 on the example of the phenylpropanoic monomer used in this study.

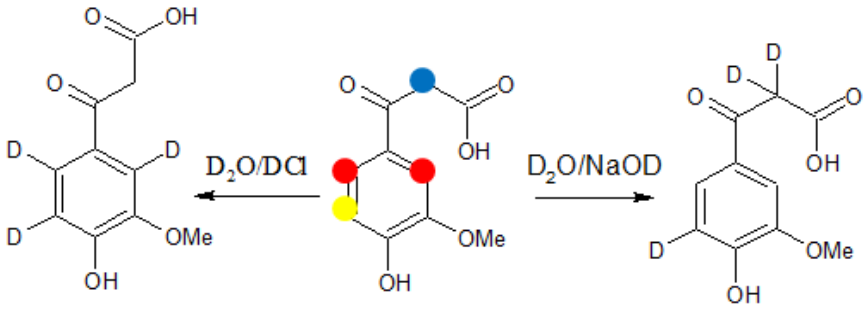


Fig. 6 H/D exchange of the non-labile backbone protons of 3 - (4-hydroxy-3-methoxyphenyl) - 3-oxopropionic acid, which takes place under conditions of acidic catalysis (in the presence of DCl, the H/D exchange positions are shown with red dots), basic catalysis (in the presence of NaOD, the H/D exchange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

It can be seen that under conditions of acidic catalysis (left panel in Fig. 6), all H-C_{ar} protons should undergo exchange with D-atoms, whereas α-CH protons of the methylene group in the propanoic moiety remain intact. At the same time, under conditions of basic catalysis (right panel in Fig. 6), only aromatic protons in the ortho-position to phenolic group might undergo exchange with D-atoms as well as α-CH protons of methylene group in the propanoic moiety; but 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 a length of exchange series to three under conditions of either acidic or basic catalysis.

Number of exchanged protons for each molecule, in its turn, can be determined by counting a number of mass shifts (Δm) equal to the mass difference between deuterium and protium (1.00628), which are related to the 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 it is shown in Fig. 7 for m/z = 441.08299, and described in detail in the experimental section.

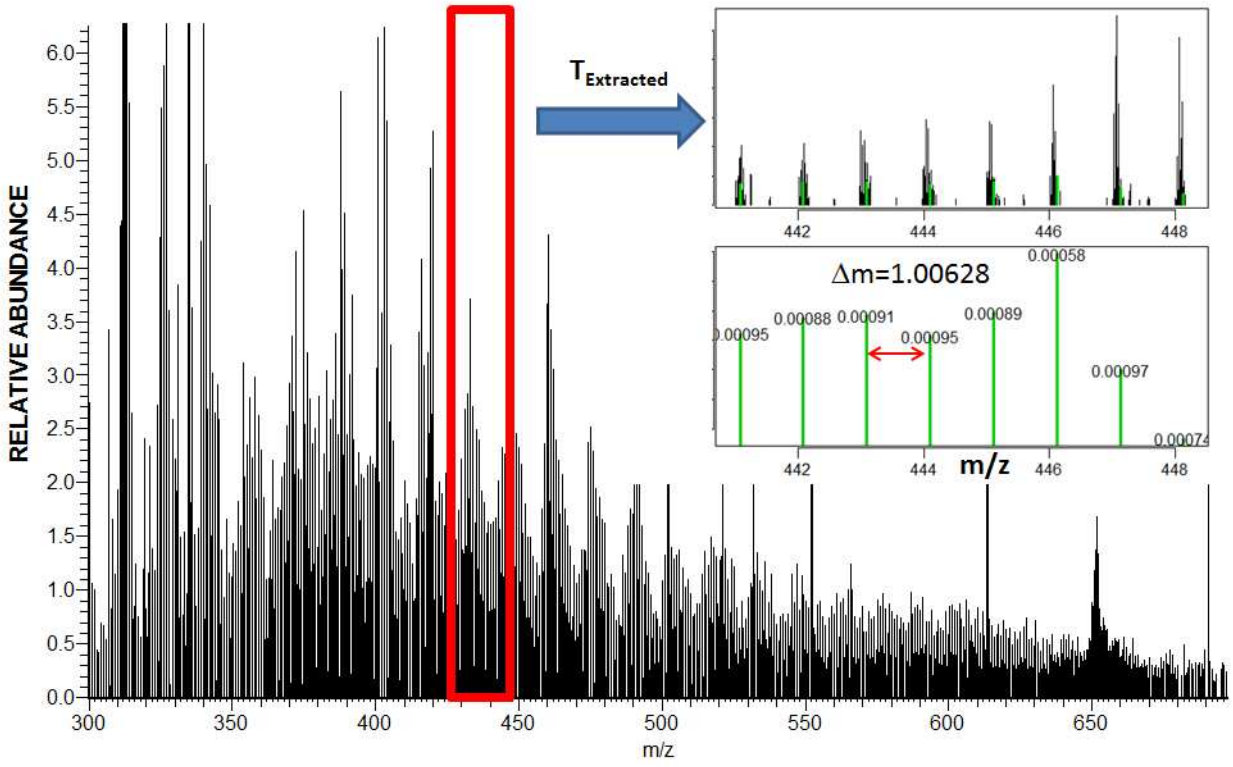


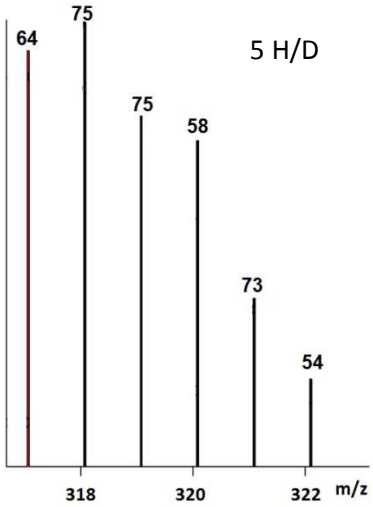
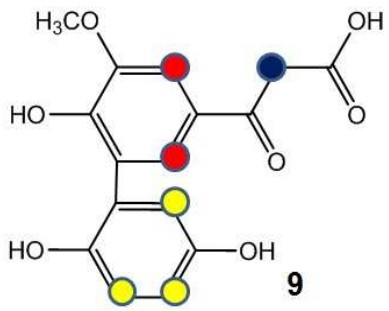
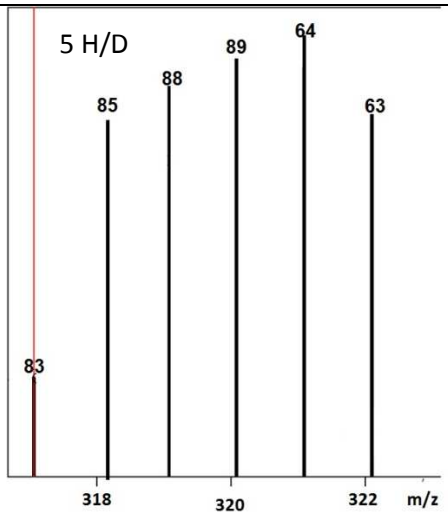
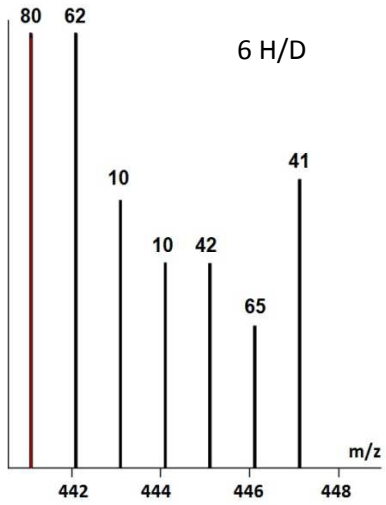
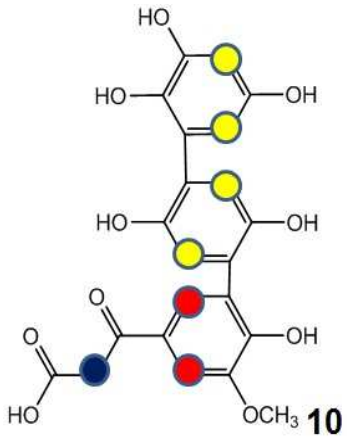
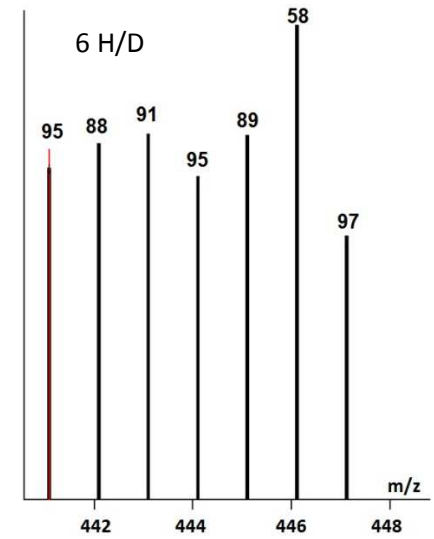
Fig. 7 FT ICR mass-spectrum of the H/D-exchanged MHQ-FA under conditions of basic catalysis (in the presence of NaOD). The insertions show mass scale-expanded segment of the full range spectrum highlighted with red color with obvious periodicity at every 1 nominal mass unit, and the extracted subspectrum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

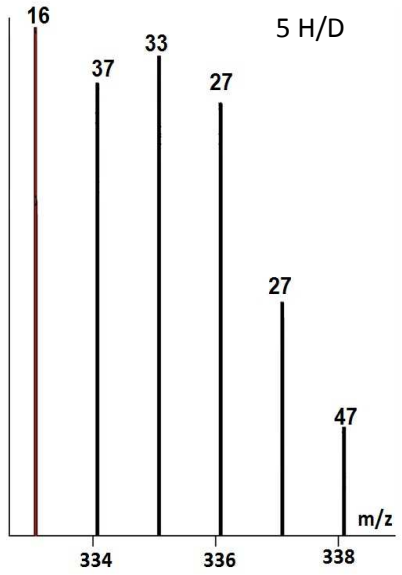
The obtained information on the length of exchange series under conditions of basic and acidic catalysis can be thus easily converted into the amount and positions of protons in the backbone of the investigated structures and can be further used to discern between the structural isomers. To demonstrate how does it work in practice, we have examined the general structural patterns identified using ^{1}H NMR spectroscopy and 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. Specifically, we searched for the formulas which would fit the elemental compositions of the patterns from **4** to **8** by varying a number of monomeric units (n) from 1 to 2. Some examples of the found molecular formulas which fit above requirements are shown in Table 2. To assign exact chemical structures to these formulas, we have extracted 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 conditions of acidic and basic catalysis. The extracted H/D exchange series are shown in Table 2.

The identified structures from **9** to **13,** which are shown in Table 2, refer to empirical formulas of $C_{16}H_{14}O_7$, $C_{22}H_{18}O_{10}$, $C_{16}H_{14}O_8$, $C_{20}H_{18}O_{10}$, and $C_{18}H_{18}O_8$, respectively. One can deduce that the compound **9** is a recombination product of M and HQ radicals, 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 surmised in Fig. 2 with exception of formation of acetylated products.

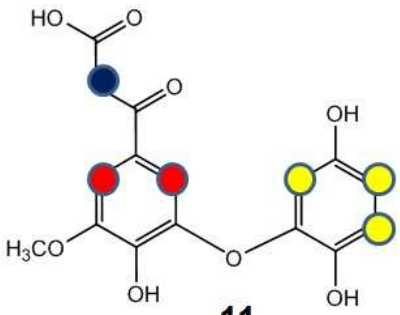
To explain the presence of acyl-substituent in compound **13**, we suggested a cleavage of aromatic ring of monomer M under oxidative conditions followed by decarboxylation of α-keto acids which leads to unstable acyl-carbanion[46].

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

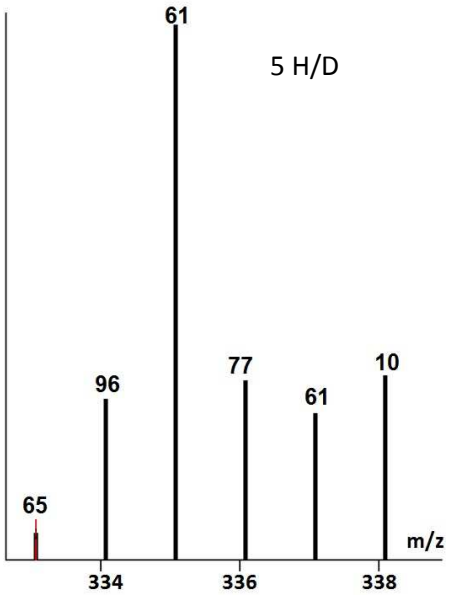
Extracted spectra of H/D exchange in DCl	Identified Structure	Extracted spectra of H/D exchange in NaOD
	**$C_{16}H_{14}O_7$** (m/z = 317.06601)  **9**	
	**$C_{22}H_{18}O_{10}$** (m/z = 441.08206)  **10**	

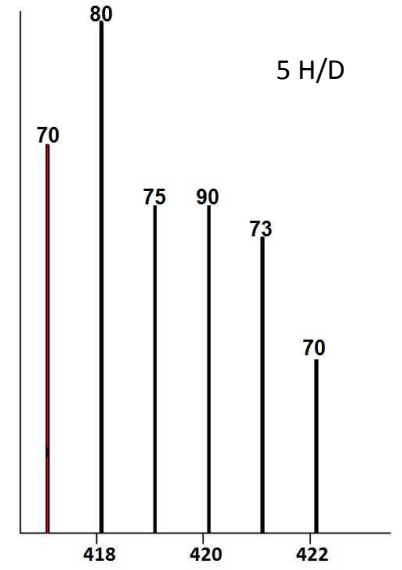


$C_{16}H_{14}O_8$ (m/z = 333.06093)

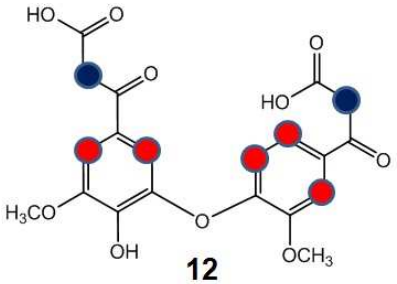


11

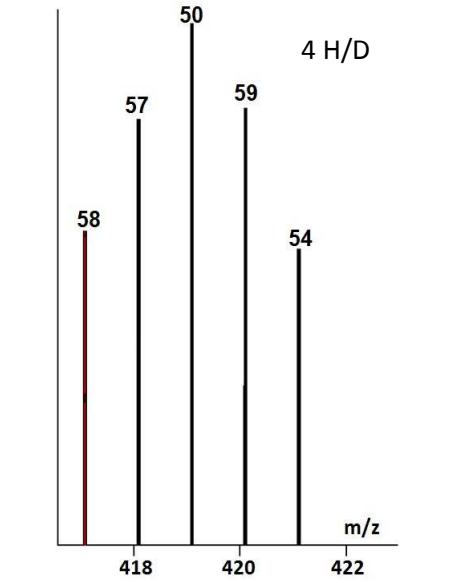


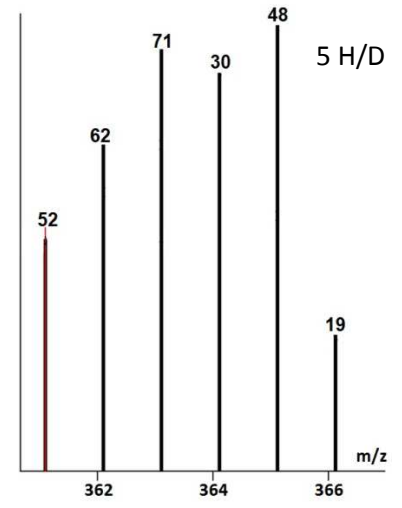


$C_{20}H_{18}O_{10}$ (m/z = 417.08206)

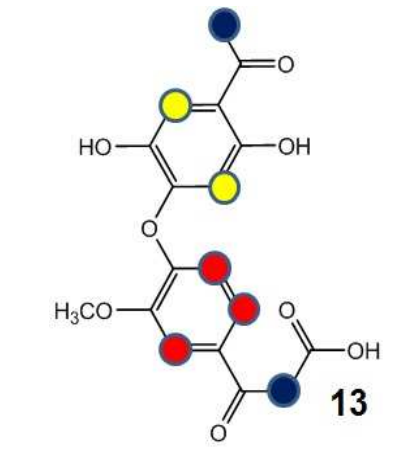


12

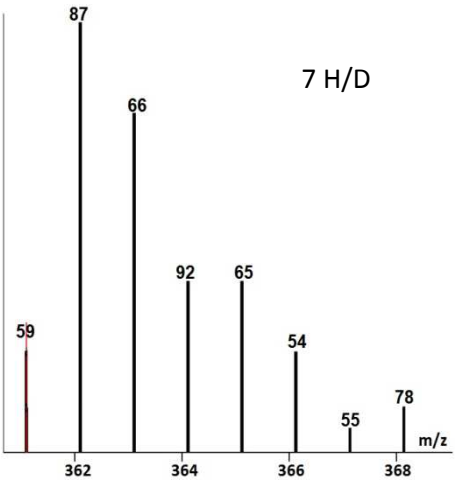




$C_{18}H_{18}O_8$ (m/z = 361.09223)



13



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

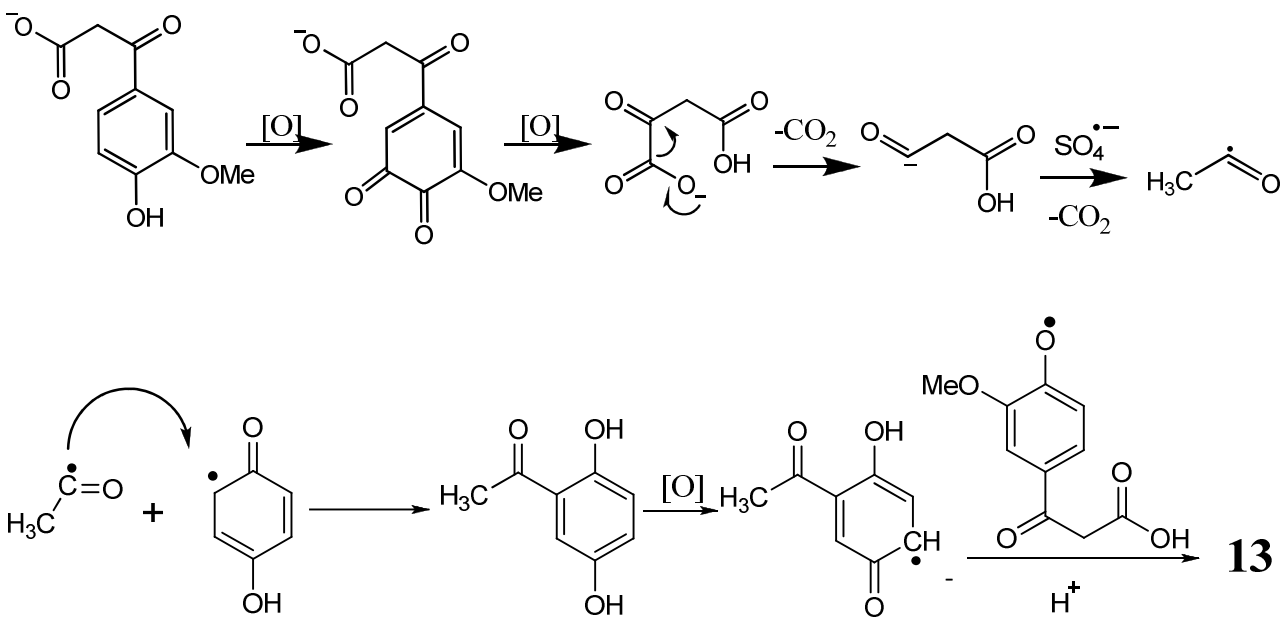


Fig. 8. The proposed reaction pathway for formation of the acetylated compound **13**.

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

This shows that a use of well-defined phenylpropanoic monomer for the oxidative coupling to hydroquinone combined with a use of high resolution isotopic exchange mass-spectrometry enabled us to identify structural formulas of individual constituents of the synthesized HS. The identified formulas, in turn, were used for refining the reaction pathways occurring during oxidative coupling in the presence of excess amount of persulfate ion. Nominally, they have revealed acetylation as an important side reaction resulting from ring cleavage of the phenylpropanoic monomer used in this study, which lead to formation of

acetylated aromatic rings. The proposed mechanism of this reaction corroborated well intense decarboxylation processes revealed by Kendrick diagrams plotted from FTICR MS data.

CONCLUSIONS

Synthesis of model humic substances using the specifically designed phenylpropanoic monomer and traditional synthetic strategy which implied its oxidative coupling to hydroquinone, lead to complex mixture of reaction products. Application of high resolution FTICR MS to characterization of the synthesized HS showed that their molecular constituents occupied both lignin- and tannin- regions on Van Krevelen diagram resembling closely location of aromatic compartments of natural HS. This demonstrated that a use of the phenylpropanoic monomer contributed substantially to approaching structural patterns exhibited by natural HS in mimicking 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 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 unprecedented resolution capacity of FTICR MS to 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 capable of identification of individual components present within the model HS. Despite the modest number of the identified formulas (dozen out of thousands), they already provided substantial information of reaction mechanism under study and allowed us to refine a final step in decarboxylation pathway leading to acetylation of aromatic rings.

We believe that validation of the proposed approach with a use of individual compounds, as well as a use of deuterium NMR and combination of different methods of selective isotopic labeling, such as H/D and $^{16}O/^{18}O$ exchange, will yield reliable structural information, which will be sufficient to determine individual molecular components of 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.

Electronic Supporting Information (ESI) available: 1) synthetic protocols for the intermediate reagents used for synthesis of the phenylpropanoic monomer; 2) ^{1}H NMR spectra of the synthetic products; 3) reaction pathways leading to 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**.**

Acknowledgements: We acknowledge 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,* 2nd Ed.; John Wiley & Sons: New York, 1994.

3 G.W. Bailey, L.G. Akim, S.M. Shevchenko, In: *Humic substances and chemical contaminants*, eds: C.E. Clapp, M.H.B. Hayes, N. Senesi, P.R. Bloom, P.M. Jardine, Humic Substances Eds. Soil Science Society of America, Madison, WI, 2001, p 41-73.

4 N. A., Kulikova, E. V. Stepanova, O. V. Koroleva, In *Use of humic substances to remediate polluted environments: from theory to practice,* eds: I. V. Perminova, K. Hatfield, N. Hertkorn, NATO Science Series: IV: Earth and Environmental Sciences, Springer Verlag: Dordrecht, The Netherlands, vol. 52, 2004, p 285.

5 R. Kloecking, B. Helbig. In *Biopolymers for medical and pharmaceutical applications,* eds: A. Steinbüchel, R.H. Marchessault. Wiley-VCH Verlag GmbH: Weinheim, 2005, p. 3.

6 F.J. Lu, S.N. Tseng, M.L. Li, S.R. Shih. *Arch Virol.* 2002, **147**, 273-284.

7 R. Kloecking, B. Helbig, G. Schotz, M. Schacke, P. Wultzer. *Antivir. Chem. Chemother.* 2002, **13**, 241-249.

8 J. Schneider, R. Weis, C. Manner, B. Kary, A. Werner, B.J. Seubert, U.N. Riede. *Virology*. 1996, **218**(2), 389-395.

9 Ph. Schmitt-Kopplin, D. Freitag, A. Kettrup, N. Hertkorn, U. Schoen, R. Klöcking, B. Helbig, F. Andreux, A.W. Garrison. *Analusis*. 1999, **27**, 391-396.

10 K. I. Hanninen, R. Kloecking, B. Helbig, *Sci. Total Environ.,* 1987**, 62**, 201-210.

11 W. Eller, *Liebigs Ann. Chem.,* 1923, **431**, 133-161.

12 H. Erdtman, M. Granath, *Acta Chem. Scand.,* 1954, **8**, 811-816.

13 K.M. Goh, F.J. Stevenson, *Soil Sci.,* 1971**, 112**, 392-400.

14 F. Cataldo, *Polym. Int*., 1998, **46**, 263-268.

15 M. Drosos, M. Jerzykiewicz, M. Louloudi, Yi. Deligiannakis, *Colloids Surf., A: Physicochem. Eng. Aspects,* 2011, **389**, 254-265.

16 N. Hertkorn, C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E.M. Perdue, M. Witt, P. Schmitt-Kopplin, *Analyt. Bioanalyt. Chem.,* 2007, **389**, 1311-1327.

17 A.G. Marshall, C.L. Hendrickson, G.S. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1–35.

18 A.C. Stenson, W.M. Landing, A.G. Marshall, W.T. Cooper, *Anal. Chem.*, 2002, **74**, 4397–4409.

19 R.W. Kramer, E.B. Kujawinski, P.G. Hatcher, *Environ. Sci. Technol.*, 2004, **38**, 3387–3395.

20 N. Hertkorn, M. Harir, B. Koch, B. Michalke, P. Schmitt-Kopplin, *Biogeosciences*, 2013, **10**, 1583-1624.

21 I.V. Perminova, I.V. Dubinenkov, A.S. Kononikhin, A.I. Konstantinov, A.Ya. Zherebker, M.A. Andzhushev, V.A. Lebedev, E. Bulygina, R.M. Holmes, Y.I. Kostyukevich, I.A. Popov, E.N. Nikolaev, *Environ. Sci. Technol.*, 2014, **48**, 7461–7468.

22 Y. Kostyukevich, A. Kononikhin, I. Popov, O. Kharybin, I. Perminova, A. Konstantinov, E. Nikolaev, *Anal. Chem.,* 2013**, 85***,* 11007-11013.

23 Y. Kostyukevich, A. Kononikhin, A. Zherebker, I. Popov, I.V. Perminova, E.N. Nikolaev, *Anal. Bioanal. Chem*., 2014, **406**, 6655-6664.

24 X. Becker, G. Domshke, E. Fanghenel, *Organikum.* Mir Publishing Company, Moscow, 1992.

25 M. Kaspar, V. Hamplova, V. Novotna, M. Glogarova, 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, G.S. Poindexter, *Org. Synth***.,** 1988, **6**, 845.

28 R.S. Swift, In *Methods of soil analysis*. Part 3. *Chemical methods,* eds: D.L. Sparks, A.L. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnson, M.E. Sumner, SSSA Book Series 5. SSSA, Madison, WI, 1996, p. 1036.

29 G.R. Aiken, E.M. Thurman, R.L. Malcolm, H.F. Walton, *Anal. Chem.,* 1979, **51**, 1799-1803.

30 B. Lindstrom, B. Sjoquist, E. Anggard, *J. Labelled Compd.,* 1974, **10**, 187-194.

31 D.V. Kovalevskii, A.B. Permin, I.V. Perminova, D.V. Konnov, V.S. Petrosyan, *Vest. 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, E.N. Nikolaev, *Anal. Chem.,* 2009**, 81**, 10106–10115.

33 A. Gaspar, E.V. Kunenkov, R. Lock, M. Desor, I. Perminova, P. Schmitt-Kopplin, *Rapid Commun. Mass Spectrom*., 2009, **23**, 683-688.

34 R. Kim, W. Kramer, P.G. Hatcher, *Anal. Chem.,* 2003, **75**, 5336–5344.

35 R. Hatfield, W. Vermerris, *Plant Physiol.,* 2001, **126,** 1351–1357.

36 G. Moad, D.H. Solomon, The Chemistry of Radical Polymerization. 2th ed. Elsevier BV, Netherlands, 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, J.C. Salfield, *Naturwissenschaften*, 1960, **47**, 516.

42 W.C. Hockaday, J.M. Purcell, A.G. Marshall, J.A. Baldock, P.G. Hatcher, *Limnol. Oceanogr. Meth.,* 2009, **7**, 81–95.

43 E. Kujawinski, M. Behn, *Anal. Chem.,* 2006, **78**, 4363-4373.

44 M.M. Tfaily, S. Hodgkins, D.C. Podgorski, J.P. Chanton, W.T. Cooper, *Anal. Bioanal. Chem.,* 2012, **404**, 447–457.

45 J. Atzrodt, V. Derdau, T. Fey, J. Zimmermann, *Angew. Chem., Int. Ed.,* 2007, **46**, 7744–7765.

46 R.W. Hanson, *J. Chem. Educ*., 1987, **64**, 591-595.

47 M. Byrn, M. Calvin, *J. Am. Chem. Soc.,* 1966, **88**, 1916–1922.