RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Full Paper

Molecular characterization of an end-residue of Humeomics applied to a soil humic acid

Nebbioso A.[a], Piccolo A.[a]*, Lamshöft M.[b], Spiteller M.[b]

[a]Centro Interdipartimentale di Ricerca per la Spettroscopia di Risonanza Magnetica Nucleare per l'Ambiente, l'Agro-alimentare e i nuovi Materiali (CERMANU), Università di Napoli Federico II, Portici, Italy

[b]Institut für Umweltforschung, Dortmund Technische Universität, Otto-Hahn Str. 6, Campus North D-44227 Dortmund, Germany

*** Corresponding author:** alessandro.piccolo@unina.it

Abstract

Humeomics encompasses a step-wise chemical fractionation and instrumental determination to fully characterize the heterogeneous molecular composition of Natural Organic Matter. Humeomics applied to a terrestrial humic acid produced a recalcitrant end-residue, that was again solubilised in alkali and size-fractionated by preparative High Performance Size Exclusion Chromatography (HPSEC). The separated ten size-fractions were then subjected to analytical HPSEC hyphenated with a high-resolution Electrospray Mass Spectrometer (ESI-MS). Total ion chromatograms (TIC) of size-fractions showed two eluting peaks, whose molecular masses were identified with empirical formulae by ESI-MS. Most empirical formulae were easily associated to linear alkanoic,

unsaturated, hydroxylated and hydroxy-unsaturated acids, as well as cyclic acids, whereas other compounds structure remained unclear. Tandem MS fragmentation was applied to one linear and three cyclic compounds, and allowed to elucidate their structures with hydroxyunsaturated hexanoic acid, two furane rings and a norbornane-like ring, respectively. The latter three compounds were never reported for terrestrial HA, but they resembled the Carboxyl-Rich Alicyclic Molecules (CRAM) earlier proposed for dissolved organic matter. Quantitative measurement of components indicated that long-chain saturated acids were present in large-sized fractions more than in short-chain homologues, while unsaturated, hydroxylated and most cyclic acids were more abundant in small-sized fractions. This suggests that long, saturated and unsubstituted linear acids enabled formation of large suprastructures, probably due to favourable intermolecular packing, as compared to the irregularly shaped cyclic, unsaturated or hydroxylated compounds. We showed that Humeomics clarify the molecular composition and conformational arrangement of Natural Organic Matter, and may contribute to elucidate the relationship between humic structure and its environmental activity.

Keywords: natural organic matter, humus, Humeomics, HPSEC, ESI-MS, tandem fragmentation.

1. Introduction

Natural Organic Matter or Humic Substances (HS) refer to a class of naturally occurring organic compounds that is commonly found in soils, sediments and natural bodies.[1] They play a pivotal role in environmental physical-chemistry and biology, and bear great influence on soil conservation, quality and fertility.[2-3] The heterogeneous and complex nature of humic matter represents an obstacle to reach a definite and rigorous structural characterization of its molecular components. Recently, HS began to be viewed as composed of relatively low molecular-weight compounds

organized in a supra-molecular structure that is held together by weak intermolecular bonds.[4-5] According to this view, a thorough characterization of single components can be achieved only if they are adequately separated before analysis. A strategy based on a step-wise chemical fractionation has proved highly efficient in reducing the humic chemical complexity.[6] As advances in biochemistry have resulted in modern Genomics, advances in the understanding of humic chemical nature produced a novel separation-characterization methods of HS that inspired the definition of "Humeomics", a term introduced in analogy with other "omic" sciences.[6] Moreover, application of size-exclusion chromatography to humic matter preliminarily separated in size-fractions has shown to increase analytical identification of humic molecules.[7-8] This was also proved by applying Humeomics to size-fractions obtained by fractionating a bulk humic acid and showed a much larger molecular identification for the size-separates than for the bulk material.[9]

Keypoints of "Humeomics" are: i. extensive fractionation of starting materials, and, ii. minimal rearrangement of carbon backbone of native HS. While the former ensures an increase in analytical yield,[6, 9] the latter prevents formation of artifacts. This is an important step forward with respect to old destructive methods such as $KMnO_4$ oxidation.[10] In fact, critics to such degradative methods have been raised because of the relevant modification of the carbon backbone.[11] "Humeomics" minimize such a drawback by preferring reactions that do not affect C-C bonds, such as hydrolysis of ester and ethers groups.[6]

While previous works succeeded to describe the type and amount of molecules separated in both organic solvents and aqueous media during Humeomics, a humic fraction remained unextractable at the end of the stepwise procedure, thereby baffling the complete detailed molecular characterization of humic matter consituents.[6, 9] The aim of this work was then to solubilize such recalcitrant end-product of Humeomics, decrease further its complexity by separation in different size-fractions, and characterize the resulting molecular components by high-resolution Electrospray Mass Spectrometry (ESI-MS).

2. Results and Discussion

2.1 Preparative HPSEC separation. As the final end-product of Humeomics, the humic residue RES4 is expected to be significantly less heterogeneous than the starting HA. In fact, RES4 was depleted of humic components solubilized as unbound molecules in solvent extraction or after hydrolysis of ester or ether bonds. Such separated molecules were shown through NMR and MS characterization to be prevalently composed of aliphatic and hydroxyl-alkyl compounds.[6] Thus, RES4 residue became enriched in aromatic compounds with a consequent enhancement of its chromophoric character. The increased aromaticity in RES4 was confirmed by CPMAS spectra, which showed a greater content of sp^2 carbon (110-160 ppm) than original HA and previous RES products of Humeomics.[6]

The enhanced light-absorbing capacity of RES4 due to enrichment in sp^2 carbons was revealed by the UV-detected preparative HPSEC chromatogram, whose elution profile consisted in three diffuse absorptions at around 28, 35 and 47 min of elution time. Such distribution of humic matter over the preparative HPSEC column allowed to separate and collect 10 different size-fractions (Fig. S1 of SI). The greatest intensity in the elution profile resulted in the largest abundance of humic matter in the corresponding size-fractions. In fact, the maximum of intensity at 35 min (Fig. S1 of SI) yielded most of the eluting humic mass, as shown by the quantity obtained for the combined 2-4 size-fractions (Tab. S1 of SI).

2.2 Analytical HPSEC-ESI-Orbitrap-MS

2.2.1 Total ion chromatograms (TIC). The unfractionated RES4 and its separated size-fractions were subjected to analytical HPSEC in order to characterize their humic components by the hyphenated mass spectrometry. The ammonia mobile phase chosen for this HPSEC-ESI-MS elution ensured reliable attribution of empirical formulae to molecular masses detected by the high-resolution mass spectrometry employed here.[6] The TIC resulting from the HPSEC-ESI-MS mass

detection for the original RES4 and its ten size-fractions are reported in Fig. S2 of SI. The charged ions produced by electrospray ionization yielded two major signals at about 20-24 and 26-30 minutes in TIC of eluting samples. A third signal was also observed at 24-26 minutes starting from size-fraction 4 up to 10. The latter may be possibly due to inorganic phosphate impurities carried over from the separation of size-fractions by preparative HPSEC. These traces may have been kept trapped into tightly associated humic domains, despite extensive dialysis in water. Interestingly, the first three size-fractions did not show such interfering signal (Fig. S2 of SI), probably because of their greater gravimetric yields of fractionation (Tab. S1 of SI), and consequent negligible influence of inorganic impurities in their TIC profiles.

The intensity of the two signals varied greatly in the different TIC of RES4 and its ten size-fractions (Fig. S2 of SI). The first signal (22-24 min) progressively decreased when going from the largest-sized to the smallest-size fractions, whereas the second signal (26-30 min) showed an opposite trend by steadily increasing its intensity with decreasing fractions size. This behaviour indicates that the preparative size-fractionation of RES4 correctly separated humic fractions according to their hydrodynamic volume. Nevertheless, the change in signals intensity observed over the size-fractionation, may suggest that the two TIC signals contained a different distribution of large- and small-sized humic associations.

2.2.2 High resolution mass spectra. The hyphenated HPSEC-ESI-Orbitrap system provided high resolution spectra for TIC signals, from which single empirical formulae were identified (Fig. 1). Since negative ionization of acidic groups produced [M-1] ions, it was generally possible to infer the ions structure based on the empirical formulae obtained by the mass-builder software of the high-resolution spectrometer. For example, the $C_{14}H_{27}O_2$, and $C_{14}H_{25}O_2$ empirical formulae, that were indicated by the software as components of the second TIC signals (Fig. 1B), could be confidently attributed to saturated and unsaturated alkanoic acids, respectively. Similarly, the $C_6H_9O_3$ and $C_6H_5O_4$ formulae under the first TIC signal (Fig.1A), were explained with an

unsaturated hydroxy alkanoic acid and a cyclic acid, respectively. The latter structure appears the only plausible explanation for an empirical formula that implies a highly unsaturated oxygen-rich compound. Unsaturated cyclic structures had been already inferred from CPMAS-NMR spectra which showed large content of sp^2 carbon in RES4.[6] The occurrence of unsaturated and hydroxy-unsaturated acids in the bulk and organosoluble fractions of this humic matter was previously highlighted.[6]

Other empirical formulae identified under TIC signals suggested the presence of still linear molecules but with less unsaturation and greater oxygenation than alkanoic acids. Double bonds together with oxygen-rich functions, such as carboxyl or hydroxyl groups, appeared concomitantly present in these molecules, although it was impossible to define their mutual substitution pattern only from empirical formulae. For example, the $C_{18}H_{33}O_4$ formula, that describes a mass eluted under the second TIC signal (Fig. 1B), may be attributed to either a saturated dicarboxylic acid or an unsaturated dihydroxy acid, and was therefore non-specifically defined as unsaturated oxygenated acid.

It is noteworthy that the molecular masses of identified empirical formulae did not follow an expected order of SEC elution time, by which analytes with large hydrodynamic radii elute before those with small radii (Tab. S2 of SI). In fact, molecules with formulae such as $C_7H_6O_8$ (most likely a cyclic compound), $C_6H_{10}O_3$ (a hydroxy-hexenoic acid), and $C_7H_{12}O_3$ (a hydroxy-heptenoic acid), were eluted before most other molecules with empirical formulae of larger masses. Nevertheless, probable cyclic structures having small masses were anyhow eluted at a much greater elution time (Tab. S2 of SI). An explanation of this contradictory behaviour resides in the association of humic molecules in large heterogeneous conformations, whose size and, thus, consequent HPSEC elution time, depends on the reciprocal arrangement of specific molecular structures.

A quantitative evaluation of empirical formulae related to other unsaturated and oxygenated acids, suggested a general progressive increase of unsaturation and oxygenation with decreasing size of fractions (Tab. 1). In fact, most of these molecules were less abundant in size-fractions

separated at small rather than at large elution time. For example, a compound with a $C_{17}H_{26}O_4$ empirical formula was found to be only 13.3, 60.2, and 20.9 μg g^{-1} of total HA in the first three size-fractions, respectively, whereas it was found to be as much as 162.5, 153.8 and 110.8 μg g^{-1} of total HA in the last three size fractions (8-10), respectively (Tab. 1). Such significant, though non-linear, increase of unsaturated oxygenated acids in small-size fractions suggests that these compounds were hardly stabilized in the large-size supramolecular associations eluting at earlier chromatographic time. This should be due to the poor associating capacity of such irregularly shaped carbon chains, since unsaturation and oxygen substitution rendered intermolecular aggregation more difficult than for the more hydrophobic unsubstituted alkanoic acids. In fact, the regular shape of the latter compounds favorably enabled their accommodation into larger size-fractions due to an easier intermolecular packing in a heterogeneous suprastructure.

Similarly, the content of cyclic acids progressively ehnanced with decreasing fractions size (Tab. 1). In fact, the corresponding $C_7H_6O_2$ and $C_6H_4O_5$ empirical formulae accounted, respectively, for 4336 and 7888 μg g^{-1} of total HA in fraction 1, whereas both compounds significantly increased to 34382 and 53960 μg g^{-1} of total HA in fraction 10, thereby confirming the elution behavior noted for TIC data (Tab. S2 of SI). Conversely, the content under the first TIC signal of a highly oxygenated cyclic acid identified with a $C_7H_6O_8$ empirical formula (Tab. S2 of SI), first increased by passing from size-fraction 1 to 3, and, then, progressively decreased in smaller size-fractions (Tab. 1).

While the abundance of individual alkanoic acids, either saturated, unsaturated, or hydroxyl-unsaturated homologues, did not show any relation with size of eluting fractions (Tab. 1), a meaningful trend can be found by pooling into short- and long-chain groups, the short acids up to dodecanoic acid, and the acids longer than tridecanoic acid, respectively. In fact, the ratio of short- over long-chain acids did generally increase with decreasing fractions size (Fig. 2), thereby suggesting a prevalence of long chain acids in suprastructures of larger volumes, and an abundance of short chain acids in those of smaller volume. Since alkanoic acids were found largely in RES4,

their average length may thus be the main factor that controls the size of humic associations in this Humeomics end-product. Moreover, the ubiquitous massive presence of alkanoic acids already observed in either humic,[6] and fulvic acids,[12] as well as in humin,[13] further suggests that such a compounds class may be mostly responsible for the stability and size of humic supramolecular associations.

By pooling the content of unsaturated and hydroxylated alkanoic acids for each size fraction in one group, and that of unsubstituted alkanoic acids in an another group, we attributed a significance to the ratio of the first group over the second one. In fact, the variation of this ratio indicated a non-linear increase with progressive decrease of size-fractions (Fig. 2). As noted earlier, the abundance of unsubstituted acids in larger size-fraction could be attributed to their regular shape and consequent more ordered packing in large-sized supramolecular associations. Conversely, alkanoic acids bearing one or more unsaturation or hydroxyl substitution are less sterically capable to regularly associate with other humic compounds, and may therefore accumulate more preferably in small size-fractions.

A quantitative evaluation of all compounds identified in RES4 and its ten size-fractions, showed that the sum of each analyte in the combined size-fractions significantly exceeded that for the unfractionated bulk RES4 (Tab.1). This substantial discrepancy may be due to a less tight molecular association in the separated size-fractions than for the bulk material, with a consequent easier ionization and detection of molecules in size-fractions. In fact, the original intermolecular association in RES4 was disrupted during HPSEC separation of size-fractions. The reorganization of the separated smaller associations into less stable and less tightly aggregated superstructures facilitated the mass spectrometry analysis by significantly increasing analytical yields. This explanation well agrees with previous findings that showed that a preliminary size-fractionation of a humic acid enhanced the detection yield for molecules undergone a Humeomic procedure.[9]

2.2.3 Tandem Mass Spectrometric analysis. The identification of empirical formulae by high resolution mass-spectrometry for humic molecules generally enables to infer their structures. However, while a single unsaturation and oxygenation in detected analytes may easily convey the assignment to alkanoic acids and their homologues, an univocal assignment of other analytes, such as cyclic acids, or that of a hydroxyl substitution to a specific carbon in hydroxy-unsaturated acids, requires further structural information. This can be achieved by reaching a further molecular fragmentation of specific masses through application of tandem MS techniques.

We concentrated attention on the four masses with m/z 129, 141, 155 and 217, to which the instrument software attributed the empirical formulae of $C_6H_{10}O_3$, $C_6H_6O_4$, $C_6H_4O_5$, and $C_7H_6O_8$, respectively. The mass-mass (MS^2) fragmentation spectra for these four empirical formulae are reported in Fig. 3, where the most plausible structure is reconstructed based on the resulting fragmented masses.

The MS^2 fragmentation for the $C_6H_{10}O_3$ empirical formula (Fig. 3A) showed formation of a [M-44] daughter ion, that signifies a loss of CO_2, as commonly occurs for carboxylic acids. Furthermore, the [M-70] fragment results from a structural breakdown to leave an acetate CH_3COO^- group, thus confirming the presence of a carboxylic acid. A daughter ion at [M-18] implies a loss of water and, hence, separation of a hydroxyl group. Moreover, the large intensity of the 111 mass peak suggests the preferential cleavage of a weak chemical bond between the hydroxyl group and the carbon chain, such as that occurring at a benzylic or allylic position. The identified $-CH_2COO^-$ and -OH groups accounted for the partial empirical formula of $C_2H_3O_3$. The remaining atoms represented an empirical formula of C_4H_6, accountable to a butyl chain with a single unsaturation. Further indications were provided by the [M-58] fragment that can be attributed to loss of CH_2COO^- with charge retention on the alkyl chain. In fact, this fragment may probably be the end-part of a carbon chain bearing a carboxyl group, that, when placed in combination with the unsaturated butyl chain, would provide a hexenoic acid with a single hydroxyl substitution. The fragmented hydroxyl group should be placed on the C4 with respect to the carboxyl group, since the

C5 or C6 positions must be involved in a double bond, since the unlikelihood of an unstable tautomer such as a vinyl alcohol. The adjacent C3 position was also ruled out for hydroxylation, because this would have favoured a great intensity for the [M-41] ion as a stable fragment of hydroxypropanoic acid, that was instead absent in the fragmentation pattern. Thus, the allylic position for the hydroxyl group is most likely, since it explains the relatively strong intensity of the daughter ion caused from loss of water. The structure of the compound with a $C_6H_{10}O_3$ empirical formula mostly consistent with experimental data is thus that reported in Fig. 3A.

The molecule with a m/z 141 mass and $C_6H_6O_4$ as empirical formula was also characterized by MS^2 fragmentation (Fig. 3B). The precursor compound produced [59] and [M-18] as main daughter ions, which were interpreted with formation of acetic acid and neutral loss of water, respectively. With the same logic used above (Fig.3A), these daughter ions may be explained with a hydroxyl group and a carboxyl group with an unsubstituted alpha position. Then, by subtracting an acetic and a hydroxyl group from the original empirical formula, the residual C_4H_3O may well be explained with an unsaturated furane ring. Arrangement of the identified hydroxyl and acetic groups on the furane ring, yielded the structure shown in Fig. 3B. This structure accounts for the lower intensity of the [M-18] daughter ion, since the energy required to break the OH-furane bond and release water, is greater than for the hydroxyhexenoic acid (Fig. 3A).

Similarly, the [59] and [M-18] daughter ions obtained by MS^2 fragmentation of the empirical formula related to the m/z 155 mass peak, were attributed to acetic acid and neutral loss of water, respectively (Fig. 3C). The $C_4H_1O_2$ empirical formula, remaining after subtraction of acetic acid and water from the initial $C_6H_4O_5$, may be justified by an oxidized oxo-furane ring (Fig. 3C). While it could be argued that a loss of water from oxo- substituent is unlikely, it is true that its tautomeric form is capable of such fragmentation (Fig. 3C). Since the energy required to break such a vinyl OH from carbon chain would be significant, this explains why the intensity of its consequent [M-18] daughter ion was relatively lower than for the corresponding fragmentation in hydroxyhexenoic acid (Fig. 3A). The furane structures with empirical formulae with m/z 141 and

155, may also suggest a possible carbohydrate source. In fact, an extensive oxidation of a carbohydrate structure may produce a carboxyl group and condensation into a furane ring.

Finally, for the empirical formula $C_7H_6O_8$ with m/z 217 (Fig. 3D), the MS^2 fragmentation. produced strong peaks with m/z [M-42] and [M-44] and a weaker one with m/z [M-18], attributable to loss of C_2H_2O (ketene), CO_2 and H_2O groups, respectively. A loss of a saturated C_3H_7 propyl group was ruled out as alternative explanation for [M-42] daughter ion, due to too many H atoms even larger than for the parent $C_7H_6O_8$ compound. Conversely, loss of ketene was a plausible reason for the formation of [M-42], since this fragmentation was already reported for other cyclic terpene molecules, such as esters and ketones with endocyclic -CH_2-CO-.[14] The subtraction of ketene, CO_2,and water from the parental molecule produced the residual $C_4H_2O_4$ empirical formula, that could not be accounted by a furane ring, due to instability of endocyclic structures with -CH_2-CO- group. A 1,4 dioxane ring with two oxygen substituents proved a better candidate to provide endocyclic substitution, as the compound suggested in Fig. 3D. In such a structure, a -CH_2-CO- chain binds together the C2 and C5 carbons in a norbornane-like ring (Fig. 3D), and bears both a carboxyl and a hydroxyl group as additional substituents. This structure fits with the experimental data and satisfy both its empirical formula and unsaturation index. Furthermore, it may also explain the unusually intense [M-44] fragmentation peak, since the related CO_2 loss may easily occur from the breakdown of endocyclic ester as by the same mechanism discussed above for the ketene loss from an endocyclic ketone (Fig. 3D). The structural complexity of this molecule is unusual when compared to the large abundance of linear compounds found in the bulk RES4 and its size fractions. The possible origin of such norbornane-type structure resides in the decay of secondary metabolites released by plants.

It is noteworthy that such a norbornane-type structure proposed here to explain a mass contained in the recalcitrant RES4 end-product, may well be related to the so-called Carboxyl-Rich Alicyclic Molecules (CRAM). This class of compounds are defined as fused carboxylated alicyclic structures, with a carboxyl-C/aliphatic-C ratio between 1:2 and 1:7, and were generally reported in

dissolved organic matter (DOM) by NMR and ultra high resolution mass spectrometry.[9, 15-17] Although CRAM are generally regarded as refractory components of dissolved organic matter in marine environments,[16] the presence in NOM of terrestrial origin of a structure with great similarity to CRAMs is reported here for the first time. Thus, this finding suggests that cyclic acids and CRAM are formed by similar biogeochemical pathways, and their different environmental fate may be ultimately determined by their degree of hydrophobicity.

3. Experimental Section

3.1 Humeomic fractionation. All reagents were from Sigma-Aldrich (99.9% pure), and used without further purification. A humic acid (HA) was isolated from a volcanic soil (Allic Fulvudand) sampled at Vico lake, near Rome (Italy), and purified as described elsewhere.[18] This HA was then submitted to the step-wise Humeomics procedure as previously reported in details.[6, 9] The procedure removed from humic matrix either unbound or ester-bound molecules, as well as ether-bound molecules, and left a final solid residue defined as RES4. This was extensively washed with Milli-Q deionized water until disappearance of iodine and freeze-dried. The freeze dried solid RES4 humic residue was suspended in water and dissolved with a 0.50 M NaOH solution by automatically titrating under N_2 to pH 7.2 (VIT 909 21 Videotitrator, Radiometer, Copenhagen) until pH remained constant for 120 minutes, and reaching a final concentration of 0.20 g L^{-1}. Possible microbial growth was prevented by adding 0.3 g/L NaN_3. This RES4 solution was then filtered through a 1 μm glass microfiber filter (Whatman GF/C) and kept refrigerated under N_2 atmosphere, until subjected to a preparative High Performance Size Exclusion Chromatography (HPSEC).

3.2 Preparative HPSEC. The RES4 solution was eluted through a Phenomenex Biosep SEC-S-2000 column (21.2 mm diameter x 300 mm length) and precolumn (21.2 mm diameter x 78 mm length). A Gilson 305 pump, a Gilson auto-sampler model 231 equipped with a 5.0 mL loop, a

Gilson FC205 fraction collector, and a Gilson 116 UV detector set at 280 nm were used to automatically fractionate humic fractions in continuous. Chromatographic runs and profiles were monitored with a Gilson Unipoint software. A mobile phase consisting of aqueous (Milli-Q Millipore deionized water) 0.3 g L^{-1} NaN_3, 0.01 M AcONa, and 3.5 mM of NaH_2PO_4/Na_2HPO_4 buffer solution at pH 7.0 was eluted through columns at a flow of 1.5 mL min^{-1}. A total of 50 mg of RES4 were injected into the preparative sysem and ten fractions were separated at different elution times. Yields for the 10 fractions are summarized in Table S1 of Supporting Information (SI). The 10 fractions (F1-F10) were acidified to pH 2.0 with 1.0 M HCl, and the precipitated humic matter was dialyzed against Milli-Q deionized water in Spectrapore 3 membranes (cut-off 3500 Da), and freeze-dried.

3.3 Hyphenated HPSEC-ESI-MS analysis. A 0.4 g L^{-1} solution of the original RES4 and its 10 size separates were made in 0.01 M NH_3, transferred in LC vials and injected into a HPSEC-MS system by a 50 µL Rheodyne loop. The flow was generated by a Dionex P 580 pump working at 0.3 mL min^{-1} elution rate. The mobile phase was a mixture of two solution A/B in a 55/45 proportion (A: 5 mM $AcONH_4$ in Milli-Q water and 5% MeCN, pH 7; B: 100% MeCN) and pumped through Phenomenex Bio-Sep SEC-S 2000 analytical column (300 x 7.8 mm) and precolumn (30 x 7.8 mm), both thermostatted at 30°C. This HPSEC system was directly connected to a LTQ Orbitrap (Thermo Electron, Waltham, MA) mass-spectrometer, that acquired spectra with negative ESI mode, a mass range of 100-1000m/z, and a 1.0 s scan time. N_2 was used as sheath gas (45 AU) and He as collision gas (7.99 AU). Spray voltage was set at 4.00 kV, spray current at 2.05 µA, capillary temperature at 260 C°, and capillary voltage at 14.93 V. Tandem MS experiments were conducted in LTQ mode at a resolution of 7500 m/z. For each compound, a scan window ranging from 50 m/z to the mass of the molecular ion was programmed. A collision energy of 45 units for MS^2 and 35 for MS^3 was applied. Average scan time was set at 30 ms for each scan event.

External standard solutions were prepared with isotopically labelled compounds (Cambridge Isotope Labs, 99%), such as ω-deuterated hexadecanoic acid (16-d-3) for linear compounds and ^{13}C

341 labelled 4-OH-benzoic acid (ring ^{13}C-6) for cyclic compounds. A stock solution of 1.0 g L^{-1} in 0.01
342 M NH_3 was prepared for each standard and subsequently diluted to reach a concentration of 10 mg
343 L^{-1}. Standards were obtained by diluting this stock solution at the following concentrations: 10 and
344 100 µg L^{-1}, and 1 and 10 mg L^{-1}. Additionally, a blank was prepared without standard. A calibration
345 curve covering variations by five orders of magnitude was built with instrumental response on the
346 basis of standards and blank.

347 Internal standard solutions were prepared by first dissolving aliquots of RES4 and its size-
348 fraction in 0.01 M NH_3 to reach a 0.40 g L^{-1} final concentration. After centrifugation, the resulting
349 supernatant was spiked with the previously described stock solution of isotopically labelled
350 compounds (1.0 g L^{-1}), in order to reach in humic solutions the same concentration as the external
351 standards: blank (no standard), 10 and 100 µg L^{-1}, and 1 and 10 mg L^{-1}. A calibration curve from
352 instrumental response of internal standards was built accordingly.

353

354 **4. Conclusions**

355

356 Our results showed that even a recalcitrant end-product (RES4) of Humeomics applied to a
357 soil humic acid could be successfully solubilized in alkaline solution, and its heterogeneity reduced
358 by separating 10 different size-fractions by preparative HPSEC. A further decrease of the complex
359 intermolecular humic associations was obtained by eluting the size-fractions through an analytical
360 HPSEC column before their characterization be a hyphenated a high-resolution Orbitrap ESI-MS.
361 We found that the molecular components of size-fractions consisted in alkanoic acids, unsaturated
362 alkanoic and hydroxyalkanoic acids, other linear unsaturated oxygenated acids, and oxygenated
363 cyclic acids. Moreover, application of tandem high-resolution mass spectrometry on masses of
364 uncertain structure, allowed their most plausible identification as unsaturated hydroxy carboxylic
365 acids, furane rings, and complex norbornane-like compounds. The molecular elucidation of such
366 highly unsaturated and oxidized chemical structures is unprecedented for terrestrial humic matter,

and suggests that extensive application of tandem mass spectrometry may contribute to enhance our knowledge on the elusive quaternary-carbon components of humic matter. While a lack of characterization of recalcitrant humic matter had been for long preventing the progress of a molecular understanding of natural organic matter, we confirm here that Humeomics is capable to overcome this limitation and promises to further enhance knowledge on the structural composition of Natural Organic Matter and its dynamics in environmental compartments.

Acknowledgement

The first author is grateful for the collaboration provided by the Institut für Umweltforschung of the Technical University of Dortmund.

References

1. Stevenson, F. J. Humus chemistry: genesis, composition, and reactions. 2nd ed. Wiley-Interscience, New York. 1994
2. Piccolo, A. Humus and soil conservation. In *Humic Substances in Terrestrial Ecosystems* Piccolo A.; Elsevier: Amsterdam, Netherlands, 1996; pp 225-264.
3. Nardi, S.; Pizzeghello, D.; Muscolo, A.; Vianello A. Physiological effects of humic substances on higher plants. *Soil Biol. Biochem*. **2002**, *34*, 1527-1536.
4. Piccolo, A. The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science. *Adv. Agron.* **2002**, *75*, 57-134.
5. Šmejkalová, D.; Piccolo, A. Aggregation and disaggregation of humic supramolecular assemblies by NMR diffusion ordered spectroscopy (DOSY-NMR). *Environ. Sci. Technol.* **2008**, *42*, 699-706.
6. Nebbioso, A.; Piccolo, A. Basis of a humeomics science: chemical fractionation and molecular characterization of humic biosuprastructures. *Biomacromolecules*, **2011**, *12*, 1187–1199.
7. Piccolo, A.; Conte, P.; Trivellone, E.; Van Lagen, B.; Buurman, P. Reduced heterogeneity of a lignite humic acid by preparative HPSEC following interaction with an organic acid. Characterization of size-separates by Pyr-GC-MS and ^{1}H-NMR spectroscopy. *Environ. Sci. Technol.*, **2002**, *36*, 76-84.

8. Conte, P.; Spaccini, R.; Piccolo, A. Advanced CPMAS-^{13}C NMR techniques for molecular characterization of size-separated fractions from a soil humic acid *Anal. Bioanal. Chem.* **2006**, *386*, 382–390.
9. Nebbioso, A.; Piccolo, A. Advances in humeomics: Enhanced structural identification of humic molecules after size fractionation of a soil humic acid. *Anal. Chim. Acta* **2012**, *720*, 77-90.
10. Matsuda, K.; Schnitzer, M. The permanganate oxidation of humic acids extracted from acid soils. *Soil Sci.*, **1972**, *114*, 185-193.
11. Maximov, O.B.; Shvets, T.V.; Elkin, Y. .N. On permanganate oxidation of humic acids. *Geoderma*, **1977**, *19*, 63-78.
12. Schnitzer, M.; Neyroud, J. A. Alkanes and fatty acids in humic substances. *Fuel*, **1975**, *54*, 17-19.
13. Nebbioso, A.; Piccolo, A. The molecular composition of soil Humin is unravelled by Humeomics. *Eur. J. Soil Sci.* **2014**. Under review.
14. Donovan, T.; Brodbelt, J. Characterization of the dissociation behavior of gas-phase protonated and methylated lactones. *J. Amer. Soc. Mass Spectrom.* **1992**, *3*, 47-59.
15. McCaul, M. V., Sutton, D., Simpson, A. J.; Spence, A.; McNally, D. J.; Moran, B. W.; Goel, A.; O'Connor, B.; Hart, K.; Kelleher, B. P. Composition of dissolved organic matter within a lacustrine environment. *Environ. Chem.* **2011**, *8*, 146–154.
16. Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a major refractory component of marine dissolved organic matter *Geochim. Cosmochim. Acta* **2006**, *70*, 2990–3010.
17. Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences*, **2013**, *10*, 1583–1624.
18. Nebbioso, A.; Piccolo, A. Molecular rigidity and diffusivity of Al^{3+} and Ca^{2+} humates as revealed by NMR spectroscopy. *Environ. Sci. Technol.*, **2009**, *43*, 2417–2424.

FIGURES CAPTIONS

Figure 1. Mass peaks and related empirical formulae under two main peaks in HPSEC-ESI-MS elutions for bulk RES4

Figure 2. Correlations between elution time of ten size fractions from RES4 and length (black) and substitution degree (gray) of carbon chain in molecules identified in size-fractions.

Figure 3. Tandem MS fragmentation analysis and plausible structure interpretation for four compounds with empirical *formulae* $C_6H_{10}O_3$ (a); $C_6H_6O_4$ (b), $C_6H_4O_5$ (c), $C_7H_6O_8$ (d).

FIGURE 1

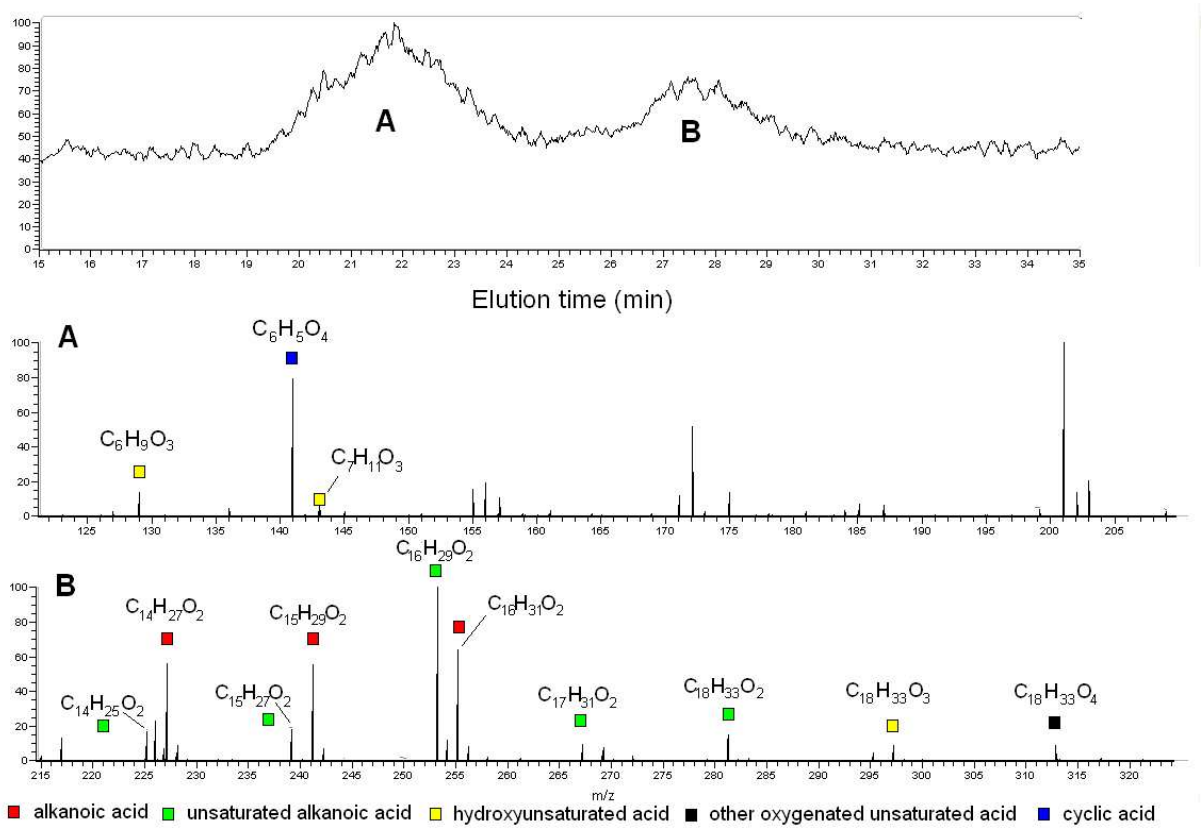
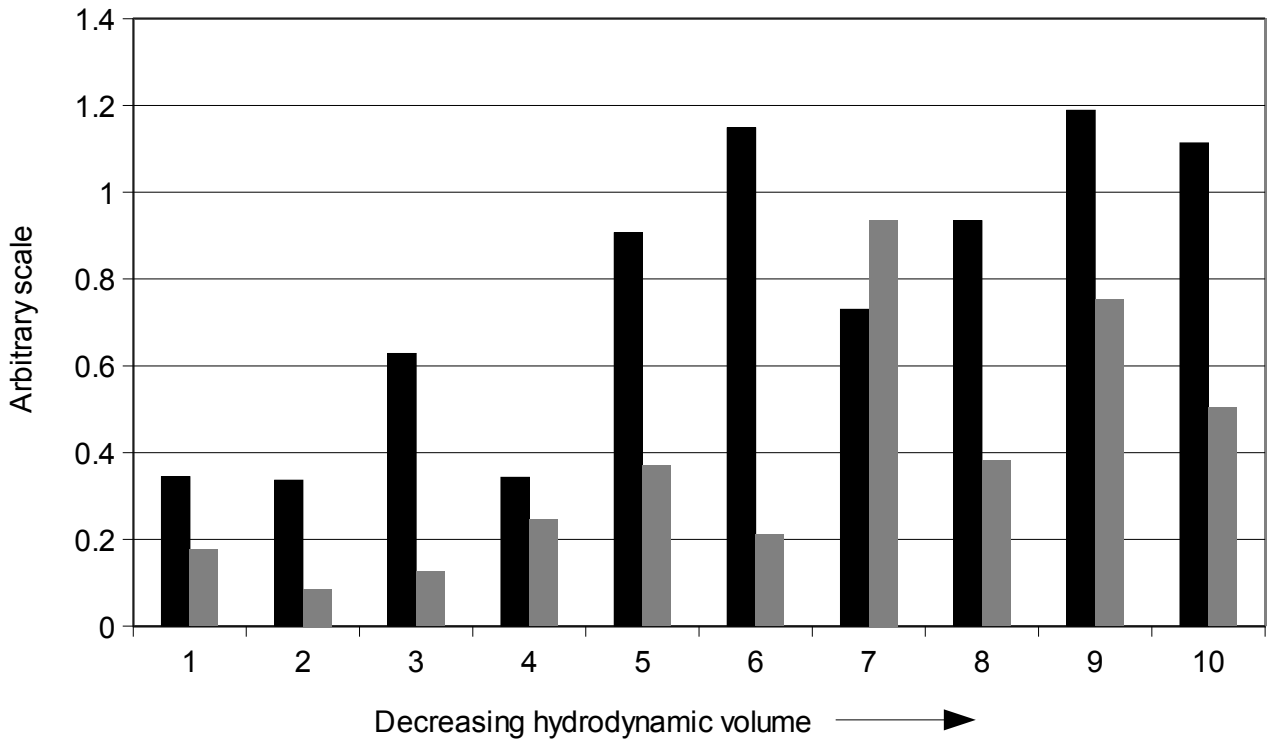


FIGURE 2



Values are based on quantitative assessment (Tab. 1). Black filling: total amount (µg g^{-1} of total HA weight) of short chain acids divided by total amount (µg g^{-1} of total HA weight) of long chain acids. Gray filling: total amount (µg g^{-1} of total HA weight) of both hydroxy unsaturated and unsaturated acids divided by total amount (µg g^{-1} of total HA weight) of unsubstituted linear alkanoic acids

FIGURE 3

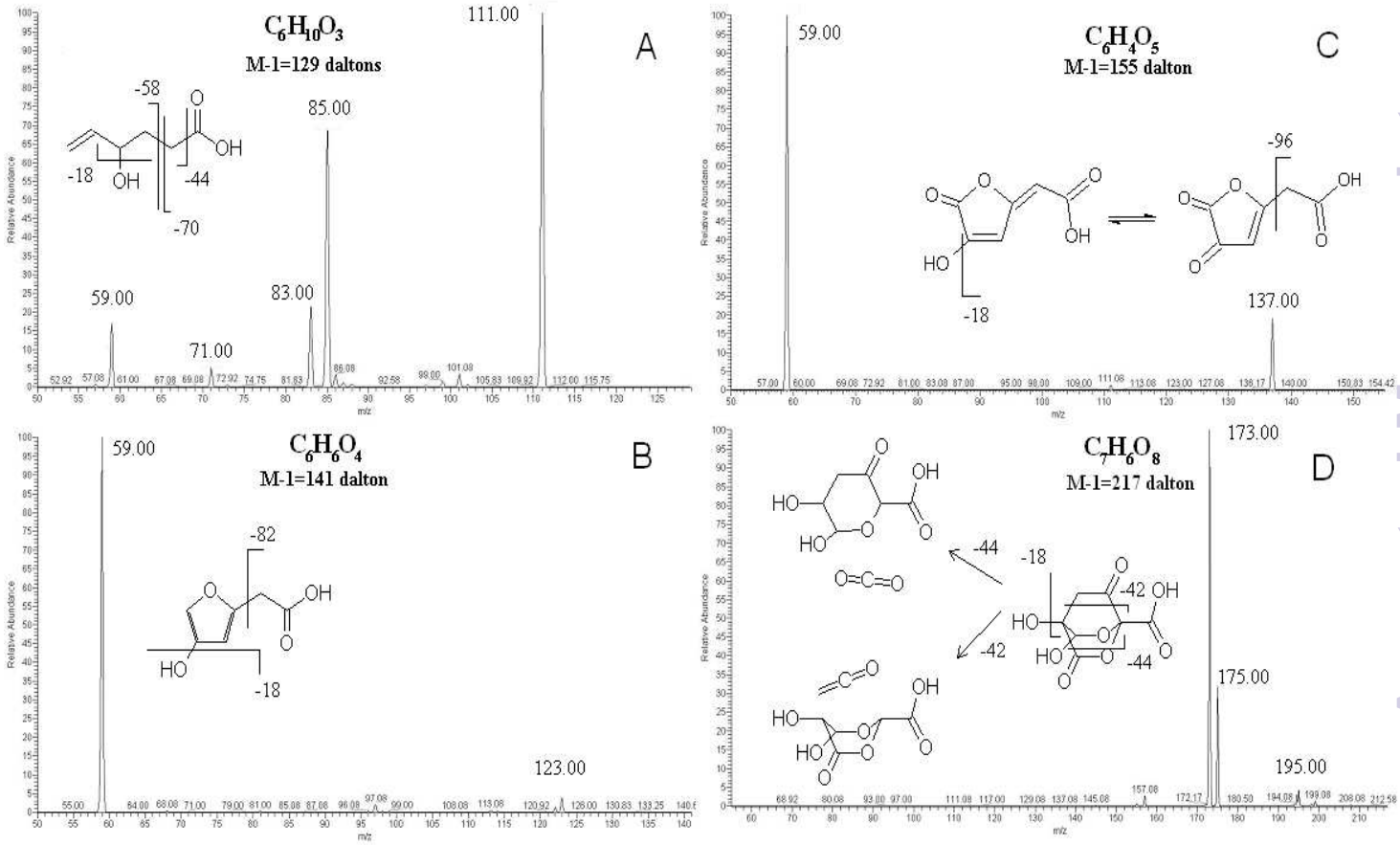


Table 1. Amount (μg g^{-1} of total HA) of molecular components found by hyphenated HPSEC-ESI-MS for RES4, ten size-fractions, and sum of size-fractions, as empirical formulae. Standard deviations were within 10% of value

Empirical formula	Bulk RES4	Size-Fractions										Sum of fractions
		1	2	3	4	5	6	7	8	9	10	
Alkanoic acids												
$C_{10}H_{20}O_2$	77	18	23	27	35	28	34	31	37	45	201	479
$C_{12}H_{24}O_2$	48	27	29	35	47	33	83	54	40	61	103	512
$C_{14}H_{28}O_2$	179	256	298	289	146	187	667	484	266	343	607	3544
$C_{15}H_{30}O_2$	191	389	411	317	431	230	791	545	229	372	400	4113
$C_{16}H_{32}O_2$	324	1000	1190	655	908	341	784	896	450	468	1090	7782
$C_{17}H_{34}O_2$	25	144	242	98	132	57	179	143	53	96	134	1278
$C_{18}H_{36}O_2$	8	122	145	104	422	25	67	160	50	59	99	1254
Unsaturated alkanoic acids												
$C_{14}H_{26}O_2$	46	37	52	64	80	36	168	106	53	98	117	812
$C_{16}H_{30}O_2$	326	537	708	828	1032	518	2054	1290	496	1035	899	9397
$C_{18}H_{34}O_2$	59	341	713	423	13108	168	507	415	150	678	625	17127
Unsaturated hydroxyacids												
$C_6H_{10}O_3$	33	15	24	24	38	58	34	65	32	46	176	511
$C_7H_{12}O_3$	23	17	28	13	33	35	38	42	27	21	93	346
$C_{18}H_{34}O_3$	22	46	21	50	72	29	49	41	18	106	243	675
Other unsaturated oxygenated acids												
$C_{15}H_{22}O_4$	9	22	24	12	73	67	186	368	71	217	426	1467
$C_{17}H_{26}O_4$	13	60	21	44	37	41	59	450	163	154	111	1138
$C_{17}H_{26}O_5$	19	12	9	6	40	36	65	233	59	103	80	642
$C_{18}H_{34}O_4$	22	9	5	9	13	3	14	12	6	23	16	110
$C_{18}H_{36}O_4$	4	33	8	8	43	2	7	5	3	22	27	157
$C_{18}H_{34}O_5$	4	31	4	3	125	2	1	3	2	51	34	257
$C_{19}H_{30}O_5$	2	28	10	8	10	19	23	391	9	103	12	611
$C_{19}H_{28}O_4N_2$	0	7	0	0	5	0	0	0	2	7	4	26
$C_{19}H_{30}O_6$	8	7	7	2	23	25	54	256	26	125	52	576
$C_{20}H_{24}O_3$	9	25	21	8	6	4	8	24	6	29	30	161
$C_{21}H_{26}O_3$	8	19	16	3	4	2	5	12	3	15	16	94
$C_{21}H_{34}O_6$	0	15	3	3	3	12	8	265	4	67	7	385
Cyclic acids												
$C_7H_6O_2$	4336	16336	17068	19570	10345	9587	12093	14422	47210	12619	34382	193632
$C_6H_4O_5$	7888	27484	9568	54330	28015	19636	23654	24775	55316	18970	53960	315708
$C_7H_6O_8$	9158	11842	9711	65224	16093	9249	9092	7834	8826	5427	3518	146816