Basis of a Humeomics Science: Chemical Fractionation and Molecular Characterization of Humic Biosuprastructures

Antonio Nebbioso and Alessandro Piccolo*

Dipartimento di Scienze del Suolo, della Pianta, dell’Ambiente e delle Produzioni Animali, Università di Napoli Federico II, Via Università 100, 80055 Portici, Italy, and Centro Interdipartimentale di Ricerca sulla Spettroscopia di Risonanza Magnetica Nucleare (CERMANU), Via Università 100, 80055 Portici, Italy

Supporting Information

ABSTRACT: We propose a mild stepwise fractionation of molecular components of a humic acid (HA) suprastructure and their structural identification by advanced analytical methods. This procedure may be the basis of a “Humeomics” approach to characterize natural humic molecules and clarify their relations with ecosystems functions. Sequential fractionation included: (1) organic solvent extraction, (2) transesterification with boron trifluoride in methanol (BF3-CH3OH), (3) methanolic alkaline hydrolysis (KOH-CH3OH), and (4) cleavage of ether and glycosidic bonds with HI. Structural identification of initial and final material, separated organo-soluble and hydrosoluble fractions, and subfractions was conducted by GC-MS, HPSEC-ESI-MS (high-resolution, Orbitrap), and solid- and liquid-state NMR. GC-MS revealed in organosoluble unbound fractions the presence of both saturated and unsaturated, linear and branched, alkanoic, hydroxalkanoic and alkandioic acids, n-alkanes, and n-alkanols. These components decreased progressively in fractions obtained after weak and strong ester cleavage. Unsubstituted alkanoic acids with variable chain length were ubiquitously detected in all fractions, thereby suggesting their fundamental function in the architecture of humic suprastructures. An important role in differentiating supramolecular associations should also be attributed to substituted alkanoic acids that were detected in variable amounts in different fractions. The content of aromatic acids and steroids was only noticed in the latter fractions. HPSEC-ESI-MS of initial and final solid fractions showed similar compounds, as indicated by GC-MS, whereas the hydrosoluble fraction after transesterification revealed fewer of these compounds but noticeable nitrogen-containing acids. A large amount of “cyclic” acids were identified by MS empirical formula in initial HA, and, to a lesser extent, in the final fractionation residue as well as in the hydrosoluble fraction. The predominant alkyl NMR signals in organosoluble extracts and those of CH-N, CH-O, and O-CH-O groups in hydrosoluble fraction confirmed mass spectrometry results. Homo- and heterocorrelated liquid-state NMR spectra indicated spin systems interactions varying with separated fractions. Solid-state and dipolar-dephasing NMR spectra of final residue showed predominance of sp2 carbons, 66% of which were quaternary carbons, and a significant increase in conformational rigidity with respect to initial HA. Separated fractions accounted for 60% of initial HA weight, and losses were attributed to hydration water, liberated volatile compounds, and decarboxylation. Quantization of analytes showed that the sum of compound classes in separated fractions was greater than that for the initial HA, thereby showing that stepwise fractionation increased significantly the analytical identification of humic molecules. Our results suggest this “Humeomics” approach as a valid path for mapping humic molecular composition and assess humus origin and formation.

INTRODUCTION

Humic substances (HSs) are ubiquitous natural compounds arising from the chemical and biological degradation of plant and animal residues1. Humic substances (HSs) are crucial in ecosystems because they regulate the global carbon and nitrogen cycles, the growth of plants and microorganisms, the fate and transport of anthropogenic compounds and heavy metals, and the stabilization of soil structure.1−5 However, manufactured technologies for controlling HS activity are not yet developed because of their large molecular complexity. Several analytical techniques have been applied to characterize HS, such as oxidative and reductive degradation reactions,6 UV and fluorescence spectroscopy,7 13C NMR spectroscopy,8−11 high-performance size exclusion chromatography (HPSEC),12 pyrolysis-gas chromatography/mass spectrometry (GC/MS),13−15 low- and high-resolution mass spectrometry with electrospray ionization (ESI/MS),16−19 solid sorption on several materials (XAD, PVP, and similar) coupled to tangential ultrafiltration,20 and others. Nevertheless, none of these techniques alone was found to be sufficient to elucidate the structural complexity of humic molecules.
A general consensus considers HSs as supramolecular associations of heterogeneous and relatively small (<1000 Da) molecules, which are held together in only apparently large molecular sizes by weak linkages, such as hydrogen and hydrophobic bonds. This novel concept suggests that humic molecular complexity may be reduced by progressively breaking the inter- and intramolecular interactions that stabilize the complex suprastructures and single humic molecules be isolated. Their structural identification may then be achieved by combining advanced analytical techniques.

Here we propose a chemical fractionation for a stepwise removal of molecules from complex humic matrices and application of up-to-date analytical techniques to determine their chemical structure. This approach may be defined as “Humeomics,” an analogy to modern terms describing genes, transcripts, proteins, and metabolite mapping technologies. However, whereas biomolecules are synthesized in living cells from specific precursors to serve defined cellular functions, HSs are formed from the entropy-driven decay of dead organic matter to encompass broader ecological and environmental functions. Nevertheless, the operational procedure of molecular separation, followed by characterization, is common to other “-omic” sciences. Humeomics may then be described as “a stepwise separation of molecules from humic suprastructures by progressively breaking intermolecular bonds and characterizing their structure by advanced analytical instrumentation.”

The humeomic fractionation adopted here begins with an organic solvent extraction of free or unbound humic molecules associated with the humic superstructure only by weak dispersive interactions, without breaking any covalent bond. Two subsequent steps include cleavage of covalent bonds in weakly bound esters by a mild boron trifluoride-methanol (BF₃-MeOH) transesterification and more strongly bound esters by an alkaline (KOH-MeOH) solvolysis. Biological polyesters in HS have been characterized, and both methods were previously described to cleave esters in soils, plants, and humic matter. Humic molecules released as a result of ester cleavage can be further fractionated by means of liquid/liquid or solid phase extraction (SPE). Finally, both strong ether and glycosidic bonds are cleaved by treatment with hydroiodic acid (HI) following a classic mechanism of protonation of the organic ether and subsequent nuclophilic substitution (SN) by iodide with an alcohol acting as a good leaving group. Qualitative and quantitative investigations on humic molecules released during this stepwise fractionation were conducted by either GC/MS or LC/MS, depending on their water solubility, whereas additional structural information was acquired by solid-state and liquid-state NMR spectroscopy.

### EXPERIMENTAL SECTION

**Humic Matter.** A humic acid (HA) was isolated from a volcanic soil (Allic Fulvudand) at Vico (near Rome (Italy)) and purified as described elsewhere. This humic extract (RES0) was oven-dried overnight at 40 °C before being submitted to sequential chemical fractionation (Figure 1). Elemental composition of humic samples was by a Fisons Instruments EA 1108 elemental analyzer. All reagents used here were by Sigma-Aldrich 99.9% pure and used without further purification. All measurements were carried out in triplicate.

**Sequential Chemical Fractionation.** Unbound Fraction (ORG1). Unbound humic molecules (ORG1) were extracted by stirring for 24 h at room temperature 1.0 g RES0 in 500 mL of a 2:1 v/v dichloromethane (DCM) and methanol (MeOH) solution. The supernatant was separated by centrifugation (15 min, 15,000 rpm) and filtered through a Whatman GFC (1.0 μm) filter. This procedure was repeated four times. The humic residue was air-dried before the next step.

Weakly Bound Ester Fractions (ORG2 and AQU2). The residue was suspended in a Teflon tube overnight with 12% BF₃-MeOH under a N₂
Table 1. Percentage (%) of Gravimetric Yields and Elemental Composition (Standard Deviations in Parentheses) of Humic Fractions Separated by Sequential Chemical Fractionationa

<table>
<thead>
<tr>
<th>sample</th>
<th>yieldb</th>
<th>C</th>
<th>O</th>
<th>N</th>
<th>H</th>
<th>S</th>
<th>C/H</th>
<th>C/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>RES0</td>
<td>100</td>
<td>50.3±0.9</td>
<td>37.3±0.1</td>
<td>4.1±0.1</td>
<td>4.6±0.3</td>
<td>3.6±1.0</td>
<td>0.91</td>
<td>1.79</td>
</tr>
<tr>
<td>ORG1</td>
<td>13.30</td>
<td>58.2±0.9</td>
<td>30.2±0.1</td>
<td>2.3±0.1</td>
<td>6.4±0.3</td>
<td>3.0±1.0</td>
<td>0.75</td>
<td>2.56</td>
</tr>
<tr>
<td>ORG2</td>
<td>11.40</td>
<td>54.0±9.8</td>
<td>39.6±0.3</td>
<td>1.6±0.3</td>
<td>4.8±0.5</td>
<td>0.0</td>
<td>0.93</td>
<td>1.82</td>
</tr>
<tr>
<td>AQU2</td>
<td>14.20</td>
<td>34.7±5.5</td>
<td>58.2±0.2</td>
<td>3.5±0.5</td>
<td>3.6±0.2</td>
<td>0.0</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>ORG3</td>
<td>0.90</td>
<td>42.7±0.9</td>
<td>53.0±0.1</td>
<td>1.2±0.1</td>
<td>3.0±0.3</td>
<td>0.1±1.0</td>
<td>1.19</td>
<td>1.07</td>
</tr>
<tr>
<td>RES4</td>
<td>20.0</td>
<td>42.7±0.9</td>
<td>53.0±0.1</td>
<td>1.2±0.1</td>
<td>3.0±0.3</td>
<td>0.1±1.0</td>
<td>1.19</td>
<td>1.07</td>
</tr>
</tbody>
</table>

a See Scheme in Figure 1. b Unaccounted weight = 40.20% (calculated by mass difference from RES0). c Oxygen was obtained by difference.

Atmosphere at 90 °C. A ratio of 0.075 mL of solution per milligram of residue was used. This transesterification was repeated three times, and the supernatants were centrifuged (15 min, 7000 rpm) and combined. The residual BF3 was quenched with water, and the solution was extracted three times with a total of 80 mL (50:50, v/v) of chloroform/water mixture. The organic phase was separated (ORG2), dried with anhydrous Na2SO4, filtered on a Whatman 41 filter, and rotovaporated. The aqueous phase (AQU2) was ultrafiltered over Amicon C membranes (1000 Da cutoff) against distilled water until it was chloride-free and freeze-dried. The remaining solid residue was water-washed and air-dried before the next step.

Strongly Bound Ester Fractions (ORG3 and AQU3). The residue was suspended with a ratio of 0.25 mL per mg of substance in 1 M KOH-MeOH solution and refluxed for 2 h at 70 °C under a N2 atmosphere. After cooling, the reaction mixture was centrifuged (10 min, 3000 rpm) and the supernatant was recovered. The residue was washed with MeOH (2 × 10 mL) and DCM (2 × 10 mL) and centrifuged. The supernatants were combined, their pH adjusted to 2.0 with 37% HCl, and then liquid-liquid extracted three times with total of 100 mL (50:50, v/v) of DCM/water mixture. The organosoluble (ORG3) and hydrophilic (AQU3) extracts were purified as for ORG2 and AQU2. The solid residue was water-washed and air-dried before the next step.

Strongly Bound Ether Fraction and Final Residue (ORG4 & RES4). A suspension of 0.25 mL of 47% HI aqueous solution per milligram of residue was stirred for 48 h at 75 °C under a N2 atmosphere. After cooling, the reaction mixture was neutralized by a saturated NaHCO3 solution. The organohydrocarbon (ORG4) was dried with anhydrous Na2SO4. The solid humic residue (RES4) was dialyzed (3500 Da cutoff) against water and freeze-dried. The separation by solid phase extraction.

Solid-state CPMAS-13C NMR (cross-polarization magic angle spinning-13C nuclear magnetic resonance) spectra were acquired with a Bruker AV 300 instrument equipped with a 4 mm wide bore MAS probe. Samples were fitted in 4 mm Zirconia rotors with Kel-F caps and spun at 13 000–15 000 rpm. Variable spin lock (VSL) experiments were acquired with a 3 s recycle delay, 30 ms acquiring time, 2200 scans, and 1510 points were acquired for each spectrum. The scan number ranged between 500 and 700. Variable spin lock (VSL) experiments were acquired with a 3 s recycle delay, 30 ms acquiring time, 2200 scans, and a VSL RAMP sequence with 1 ms contact time. An average spin lock frequency of 60 MHz was applied during the ramped cross-polarization time. Spin lock was varied in intervals from 0.01 to 7.50 ms in 21 increments. Dipolar dephasing (DD) experiments were set with a 15 μs delay and 15 000 scans. Data were processed with Mestre-C software 4.9.9.9, and all FID spectra were transformed with 100 Hz line broadening exponential type filter function and 2k zero filling.

Liquid Chromatography—Mass Spectrometry (LC/MS). RES0, AQU2, and RES4 samples were dissolved into LC vials at 0.4 g/L using a 0.01 M NH3 solution and injected with a 50 μL Rheodyne loop in a HPSEC system connected to the LC/MS system. HPSEC comprised a Phenomenex Bio-Sep SEC-S 2000 column (300 × 7.8 mm) and precolumn (30 × 7.8 mm), both thermostatted at 30 °C. A Dionex P 580 pump ensured a 0.3 mL·min⁻¹ elution of a 55/45 A/B solution (A: 5 mM AcONH4 in Milli-Q water and 5% MeCN, pH 7; B: 100% MeCN) for a total of 70 min. Mass spectra were obtained with a LTQ Orbitrap (Thermo Electron, Waltham, MA) and negative ESI at 100–1000 m/z mass range, and 1.0 s scan time. N2 was the sheath gas (45 AU) and He was the collision gas (7.99 AU). Spray voltage was at 4.00 kV, spray current was at 2.05 μA, capillary temperature was 260 °C, and capillary voltage was 14.93 V. Hexadecanoic acid (16 d-3) and 4-hydroxybenzoic acid (ring 13C-6) (Cambridge Isotope Laboratories, 99%) were dissolved in a 0.01 M NH3 solution and added as internal standards for quantitative evaluation.

NMR Spectroscopy. Solid-state CPMAS-13C NMR (cross-polarization magic angle spinning-13C nuclear magnetic resonance) spectra were acquired with a Bruker AV 400 instrument equipped with a 5 mm inverse broadband, z-gradient coil, actively shielded probe. About 5.0 mg of sample was placed in 5.0 mm quartz tubes and dissolved with 1:1 CD3OD/CDC13 for ORG1; DMSO-d6 for ORG2, AQU2, and ORG3; and D2O/NaOD for RES4. Monodimensional spectra were acquired with presaturation of water signal and 1000 scans. Bidimensional spectra were set up as follows: COSY experiments were multi quantum filter (MQF) magnitude type with pulse gradient, 60–88 scans, 256/2k data points set over F1/F2, and a 2.0 s delay; TOCSY experiments were TPPI phase-sensitive type with a mixing time of 80 ms, 112–128 scans, 128/2k data points set over F1/F2,
and a 2.0 s delay; HSQC experiments were TPPI phase-sensitive, echo-antiecho type with short-range coupling frequency set at 145 Hz, 128–256 scans, 256/2k data points set over F1/F2, and a 2.0 s delay; HMBC experiments were low-pass magnitude J-filtered pulse gradient type with a long-range coupling frequency set at 5 Hz, 100–140 scans, 256/4k data points set over F1/F2, and a 2.0 s delay.
Table 2. Amount (μg·g⁻¹ of Original HA Weight) and Standard Deviation of Compound Classes Detected by GC-MS and HPSEC–ESI–MS in Humic Fractions Separated by Sequential Chemical Fractionation

<table>
<thead>
<tr>
<th>GC-MS</th>
<th>ORG1</th>
<th>ORG2</th>
<th>ORG3</th>
</tr>
</thead>
<tbody>
<tr>
<td>fatty acids</td>
<td>14179 (±119)</td>
<td>2203 (±123)</td>
<td>632 (±55.4)</td>
</tr>
<tr>
<td>alcohols</td>
<td>1591 (±97.0)</td>
<td>1964 (±107)</td>
<td>305 (±57.2)</td>
</tr>
<tr>
<td>di/trihydroxyacids</td>
<td>4454 (±672)</td>
<td>10811 (±557)</td>
<td>988 (±0.00)</td>
</tr>
<tr>
<td>ω-hydroxyacids</td>
<td>10997 (±714)</td>
<td>8784 (±472)</td>
<td>271 (±1.00)</td>
</tr>
<tr>
<td>α,β-hydroxyacids</td>
<td>2978 (±577)</td>
<td>10145 (±623)</td>
<td>335 (±0.00)</td>
</tr>
<tr>
<td>α,ω-diacids</td>
<td>13282 (±461)</td>
<td>2418 (±211)</td>
<td>323 (±0.00)</td>
</tr>
<tr>
<td>aromatic acids</td>
<td>0.00</td>
<td>1157 (±78.5)</td>
<td>24.0 (±2.83)</td>
</tr>
<tr>
<td>steroids</td>
<td>0.00</td>
<td>183 (±23.9)</td>
<td>65.0 (±0.00)</td>
</tr>
<tr>
<td>hydrocarbons</td>
<td>175 (±28.0)</td>
<td>0.00</td>
<td>19.0 (±0.00)</td>
</tr>
<tr>
<td>total</td>
<td>47481 (±1237)</td>
<td>37666 (±999)</td>
<td>2943 (±79.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HPSEC–ESI–MS</th>
<th>RES0</th>
<th>RES4</th>
<th>AQU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>even-numbered saturated acids</td>
<td>3420 (±786)</td>
<td>635 (±127)</td>
<td>13.5 (±12.8)</td>
</tr>
<tr>
<td>odd-numbered saturated acids</td>
<td>1264 (±284)</td>
<td>216 (±8.81)</td>
<td>7.24 (±7.32)</td>
</tr>
<tr>
<td>hydroxyl-unsaturated acids</td>
<td>848 (±313)</td>
<td>0.00</td>
<td>14.6 (±3.80)</td>
</tr>
<tr>
<td>hydroxylated acids</td>
<td>1452 (±59.0)</td>
<td>55.8 (±2.67)</td>
<td>0.00</td>
</tr>
<tr>
<td>diacids</td>
<td>2680 (±111)</td>
<td>0.00</td>
<td>1.27 (±0.07)</td>
</tr>
<tr>
<td>cyclic acids</td>
<td>285899 (±10551)</td>
<td>21383 (±829)</td>
<td>343 (±24.0)</td>
</tr>
<tr>
<td>unsaturated acids</td>
<td>2462 (±692)</td>
<td>431 (±19.5)</td>
<td>23.6 (±18.1)</td>
</tr>
<tr>
<td>hydroxy-diacids</td>
<td>458 (±36.7)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>other acids</td>
<td>815 (±323)</td>
<td>151 (±24.0)</td>
<td>0.00</td>
</tr>
<tr>
<td>nitrogen containing acids</td>
<td>0.00</td>
<td>0.00</td>
<td>36.7 (±1.35)</td>
</tr>
<tr>
<td>total</td>
<td>299298 (±10617)</td>
<td>22290 (±0.00)</td>
<td>450 (±29.8)</td>
</tr>
</tbody>
</table>

RESULTS

The sequential fractionation of this HA yielded substantial amounts of ORG1, ORG2, ORG3, AQU2, and RES4 fractions, whereas AQU3 and ORG4 fractions yielded too small amounts for meaningful instrumental analysis. Approximately 40% of the weight of fractionated material was unaccounted for (Table 1), perhaps because of losses of occluded water or small volatile organic compounds and because of partial decarboxylation induced by heating.

Original Humic Acid (RES0). Elemental analysis of original HA (RES0) was typical for a terrestrial HA (Table 1).[^3] The TIC HPSEC weight of fractionated material was unaccounted for (Table 1), whereas AQU3 and ORG4 fractions yielded too small amounts of ORG1, ORG2, ORG3, AQU2, and RES4 fractions, among hydroxyunsaturated acids, the order of abundance was: mono-ω-C16 (0.39 mg/g TOM), tri-ω-C18 (0.25 mg/g TOM), and mono-ω-C18 (0.24 mg/g TOM). Among hydroxy-unsaturated acids, the order of abundance was C18 (0.42 mg/g TOM), C16 and C18 (Table S-1 of the Supporting Information). As for saturated dioic acids (Table 2), C16 (1.2 mg/g TOM) was the most abundant (Table S-1 of the Supporting Information), whereas C24 (0.22 mg/g TOM) revealed the largest amount among hydroxy-diacids. Other negatively charged compounds were also detected, for which empirical formulas were determined, but no structure could be assigned. Among these, C16H26O4 (0.44 mg/g TOM) was the most important.

The 13C–CPMAS–NMR spectrum (Figure 3B) of RES0 showed signals for carboxylic (175 ppm), sp^2 aromatic/double bond (110–150 ppm), C–O, C–N (65–105 ppm), and unsubstituted alkyl (0–65 ppm) carbons. The large intensity of the latter NMR signal (Figure 3B) suggests that HPSEC–ESI–MS may overestimate cyclic over linear structures (Table 2).

Unbound Fraction (ORG1). The weight yield of ORG1 accounted for 13% of RESO with smaller C/H and larger C/O ratios than RESO, thus indicating a greater content of aliphatic and saturated compounds (Table 1). Both 13C–CPMAS–NMR and 1H NMR spectra confirmed a large alkyl content by C9 was barely detectable. The C15 (0.78 mg/gTOM) and C17 (0.48 mg/gTOM) were the most abundant odd-numbered acids, whereas oleic acid (1.39 mg/gTOM) was the largest among unsaturated alkanolic acids.

Moreover, HPSEC–ESI–MS allowed detection of hydroxylated C14–C24 acids, di- and trihydroxylated C18 acids, mono-unsaturated C14, C16, and C18 acids, hydroxy-unsaturated C6–C20 acids, C4–C24 diocic, C16–C24 hydroxy-dioic, cyclic acids, and several other unassigned acidic structures (Table S-1 of the Supporting Information). The most important hydroxy-saturated acids were: mono–OH-C16 (0.39 mg/gTOM), tri–OH-C18 (0.25 mg/gTOM), and mono–OH-C24 (0.24 mg/gTOM). Among hydroxy-unsaturated acids, the order of abundance was C18 (0.42 mg/gTOM), C16 and C18 (Table S-1 of the Supporting Information).
showing intense aliphatic signals in the $^1\text{H} = 0$ to 2.30 ppm and $^{13}\text{C} = 0$ to 65 ppm ranges (Figure 3). The ORG1 spectra revealed a distribution of different signals: C–O, C–N, or both in the $^1\text{H} = 2.30$ to 4.50 ppm and $^{13}\text{C} = 65$ to 95 ppm ranges, unsaturated compounds in the $^1\text{H} = 4.50$ to 9.0 ppm and $^{13}\text{C} = 105$ to 150 ppm ranges, and carbonyl structures ($^{13}\text{C} > 150$ ppm).

GC-MS results (Figure 4A, Table S-2 of the Supporting Information) were consistent with NMR information. In fact, ORG1 contained C$_{12-32}$ saturated acids, C$_{16-18}$ mono- and di-unsaturated acids, C$_{14-18}$ mono- and polyhydroxylated acids in $\alpha$, $\beta$, and $\omega$ positions, C$_{15-18}$ branched alkanolic acids, C$_{9-26}$ $\alpha$, $\omega$-dioic acids, aromatic acids, and C$_{14-28}$ n-alkanols. Even-numbered linear monoacids were more abundant than odd-numbered ones (Table 2). Despite differences in MS techniques, the pool of n-alkanoic acids in ORG1 was qualitatively similar to RES0 as for chain length (Table S-2 of the Supporting Information). Mono- and dicarboxylic acids and some iso- and anteiso-branched alkanolic acids were the most important components in ORG1, as opposed to RES0 where cyclic acids were predominant (Table 2, Table S-2 of the Supporting Information). C$_{18}$ and C$_{16}$ linear acids were the most abundant among short-chain compounds in ORG1, whereas C$_{26}$ and C$_{22}$ acids were the largest among long-chain homologues (Table S-2 of the Supporting Information). C$_9$, C$_{16}$, and C$_{22}$ dioic acids were most significant (Table S-2 of the Supporting Information), with amounts similar to RES0. Hydroxyacids as $\omega$-C$_{16-18}$, $\beta$-C$_{14}$, and $\alpha$-C$_{16}$ were particularly abundant in ORG1.

Two-dimensional spectra (Figure S-1A of the Supporting Information) provide detailed information of how functional groups are bound to one another. Cross peaks therein suggested spin systems such as alkyl-alkenes (Figure S-1A of the Supporting Information, peaks a–c), vicinal hydroxyls (Figure S-1A of the Supporting Information, peak f), alkyl-hydroxyl (Figure S-1A of the Supporting Information, peak h), carboxyl-alkyl (Figure S-1A of the Supporting Information, peak m), and some substituted aromatic patterns (Figure S-1A of the Supporting Information, peaks q–w).

**Organosoluble Weakly Bound Fraction (ORG2).** The weight yield of organosoluble ORG2 was 11% of RES0 (Table 1). ORG2 had C/H and C/O ratios of 0.93 and 1.82 (Table 1), respectively, indicating an unsaturation more similar to RES0 than to ORG1. Similarities to ORG1 were found in both $^{13}\text{C}$–CPMAS and $^1\text{H}$ spectra, except for the intense $^{13}\text{C}$ peak around 55 ppm due to methylation of carboxyl groups induced by methanolic transesterification (Figure 3B). In particular, $^1\text{H}$ NMR showed signals for CH-C-X (X: electron-withdrawing group) systems in the 1.9–3.0 ppm interval and CH-X (X: O/N) groups in the 3.0–4.5 ppm range (Figure 3A).

GC-MS chromatograms of derivatized components in both neutral and acidic subfractions of ORG2 (Figure 4B–C) revealed C$_{9-29}$ linear and branched alkanolic acids, $\alpha/$$\beta$ C$_{9-26}$ mono- and dihydroxyacids, the C$_{18}$ unsaturated acid, C$_{9-30}$ $\alpha$, $\omega$-dioic acids, sterols, C$_{12-28}$ n-alkanols, and substituted benzoic and cinnamic acid derivatives (Table S-3 of the Supporting Information). These compounds were similar to those found in RES0, except for the lack of cyclic structures in ORG2 (Table 2). The presence of steroids and aromatic compounds should be due to their liberation from ester bonds cleaved during BF$_3$-methanolic ORG2 extraction (Table 2).

Quantitative differences from ORG1 were shown by GC-MS results (Table 2, Table S-3 of the Supporting Information). Unlike ORG1, hydroxyacids were most abundant in ORG2 (Table 2), whereas alkanoic and unsaturated acids were found in lesser amount than for ORG1 (Table S-3 of the Supporting Information). The distribution of monohydroxyacids also differed from ORG1 because di- and tri-OH chains were more abundant (10.8 mg/g TOM) than $\alpha$ and $\omega$-OH homologues (Table S-3 of the Supporting Information). Whereas diaicids were found in lesser amounts (2.4 mg/g TOM) than for ORG1, other compounds such as sterols, linear alcohols, and aromatic compounds (mostly 3- and 3,4-substituted benzoic and cinnamic acids) were largely more abundant in ORG2 (Table S-3 of the Supporting Information).

Despite such compositional differences, 2D NMR spectra of ORG2 showed similar signals as ORG1. Both homo- and heterocorrelated spectra revealed cross peaks for internal hydroxyl and end-chain carboxyl groups (Figures S-2A and S2-B of the Supporting Information, peaks b, d, and j) as well as different spin correlations.
systems, such as methyl ester signals (Figure S-2D of the Supporting Information, peak j), and primary alcohols (Figures S-2A and S-2B of the Supporting Information, peaks e and peaks g and h, respectively). Signals in the aromatic region suggested the presence of 3-monosubstituted and 3,4-disubstituted ring patterns (Figure S-2A of the Supporting Information, peaks l–p), typical of benzene rings with hydroxyl or methoxyl substituents.

Hydrosoluble Weakly Bound Fraction (AQU2). The weight yield of hydrosoluble AQU2 was ∼14% of RES0, and both C/H and C/O ratios were 0.80 (Table 1), thus indicating a lower unsaturation and larger oxygen substitution than that for RES0 and previous organosoluble fractions. The large N content in AQU2 implies a greater affinity of humic nitrogenous components to the aqueous phase, as opposed to the lower N content found in ORG2. These indications were supported by the $^{13}$C-CPMAS spectrum (Figure 3B), showing prominent signals for CH–N, CH–O, and O–CH–O groups at 50–60, 60–80, and 90–105 ppm intervals, respectively, and for carboxyl groups (180 ppm). Moreover, the $^1$H spectrum revealed an increase in the 2.2–4.5 ppm interval, attributable to CH–X (X: OH/NH$_2$) groups (Figure 3A).

HPSEC–ESI–MS provided further molecular characterization to support NMR data (Table S-4 of the Supporting Information). It showed content of odd- and even-numbered C$_{10–18}$ saturated acids, C$_{15–18}$ unsaturated acids, C$_{15–18}$ hydroxy-unsaturated acids, C$_{5–9}$ diacids as well as C$_{7–18}$ negatively charged compounds containing one to five nitrogen atoms and cyclic acids as proxy of linear substituted aliphatic acids and C$_{3}$H$_{2}$O$_{5}$ saccharide structures, respectively. The HPSEC–ESI–MS chromatogram indicated that linear compounds in AQU2 had the same elution times as those in RES0, whereas short chain lengths and less hydrophobic homologues had different size distribution. The TIC chromatogram (Figure 5) showed three major signals at around 26, 31, and 38 min with mass distribution related to different elution of analytes. Most substances eluted at 26 min (Figure 5A), except for the majority of nitrogenated compounds and two of the abundant cyclic acids, which eluted between 31 and 38 min (Figures 5B–C).

The odd- and even-numbered saturated acids were abundant (>21 μg/g$_{\text{TOM}}$), particularly the C$_{15–16}$ homologues (Table S-4 of the Supporting Information), as observed in ORG2 and RES0. Similar abundance was shown by unsaturated and hydroxylated acids, with the C$_{16}$ homologue being the largest one, whereas saturated alkanoic acids were found only in traces (Table S-4 of the Supporting Information). Unidentified nitrogenated compounds were mostly represented (42 μg/g$_{\text{TOM}}$) by the C$_6$H$_{10}$O$_2$N$_3$ empirical formula (Table S-4 of the Supporting Information). The large carboxyl CPMAS signal may be assigned to such structures (Figure 3B) that contribute to diversify AQU2.

Figure 4. GC-MS (TIC) chromatograms of (A) ORG1, (B) ORG2 neutral subfraction, (C) ORG2 acidic subfraction, and (D) ORG3. Peak labeling identifies compounds classes. Structural details of labeled peaks are reported in Tables S-2, S-3, and S-5 of the Supporting Information.
from both ORG1 and ORG2 and, to a lesser extent, RES0. This suggests that small and hydrophilic compounds are favorably extracted in aqueous media. Other nitrogenated compounds (2–10 μg/gTOM, Table S-4 of the Supporting Information) were
identified by the MS-Orbitrap software with most probable empirical formulas: C_{16}H_{29}O_{6}N_{3}, C_{10}H_{29}O_{7}N_{5}, C_{7}H_{13}O_{6}N_{4}, and C_{12}H_{33}O_{6}N_{5}. These compounds are likely to be responsible for the multiplicity of CH-OH proton signals in the 'H NMR spectrum of AQU2 (Figure 3A). Although unambiguous assignment of these structures was not possible, they largely characterized AQU2 because all other fractions showed poor evidence of aminoglycosides. However, the sensitivity of LC-MS was limited, probably because of less ionizable weakly acidic hydroxyl groups prevailing over more acidic carboxyl groups.

The 2D COSY spectrum indicated correlations for CH-CH spin system (Figure S-3A of the Supporting Information, peak g), whereas TOCSY also showed long-range CHX-CHX (Figure S-3B of the Supporting Information, peaks d-f) and CH-CHX
(Figure S-3B of the Supporting Information, peaks a–c) interactions, probably belonging to alkyl structures bearing hydroxyl or amino groups (Figure S-3B of the Supporting Information). The same TOCSY regions in ORG1 (Figure S-1B of the Supporting Information) were scarcely populated, thereby suggesting isolated hydroxyl moieties, as opposed to possible glycosidic or polyolic structures in AQU2. Aromatic COSY (Figure S-3A of the Supporting Information, peak i) and TOCSY (Figure S-3B of the Supporting Information, peaks o–q) correlations in AQU2 appeared in more deshielded intervals than those for ORG2 (Figures S-2A and S-2B of the Supporting Information, peaks l–q and peaks s–v, respectively), and their larger chemical shift values (8.5 to 8.7 ppm) suggest that they possibly derive from double-condensed rings. No evidence of these structures was found by LC-MS, possibly because of lack of ionizable groups.

**Organosoluble StronglyBound Fraction (ORG3).** ORG3 gave a weight yield of only 0.9% of RES0 (Table 2) and had a thick glue-like consistency that prevented meaningful elemental analysis. The $^1$H NMR spectrum (Figure 3A) showed large aliphatic (0–2.30 ppm) and CH-X (X: O/N, 2.30–4.50 ppm) signals due to hydroxylated acids. Contrary to ORG1 or ORG2 spectra, a number of low intensity signals for unsubstituted sp$^2$ systems (4.50–9.00 ppm) were visible for ORG3, possibly responsible for the observed deliquescence.

These indications were supported by GC-MS (Figure 4D), which revealed the presence in ORG3 of C$_{6}$–C$_{12}$ saturated and unsaturated n-alkanoic acids, C$_{6}$–C$_{8}$ α,ω-diacids, C$_{16,22,24}$ ω-monohydroxyacids, C$_{22}$–C$_{26}$ β-monohydroxyacids, C$_{18}$ di- and trihydroxyacids, C$_{12}$–C$_{18}$ n-alkanols, phenolic acids, and steroids (Table S-5 of the Supporting Information). The most abundant saturated acid was C$_{16}$, whereas C$_{18,1}$ was the largest among unsaturated acids, thus showing consistency with $^1$H NMR observations. The great degree of hydroxylation in ORG3 was confirmed by the abundant poli-hydroxylated C$_{18}$ and monohydroxylated C$_{16}$ acids. Also, diacids and n-alkanoic acids showed quantitative importance, although less than that for ORG1 or ORG2, the largest amount being measured for C$_{16}$ (302 mg/gTOM) and C$_{12}$ (11.8 mg/gTOM) homologues, respectively (Table S-5 of the Supporting Information).

Two-dimensional NMR spectra of ORG3 suggested spin systems (Figure S-4 of the Supporting Information) similar to those of ORG1 and ORG2. Homocorrelated COSY and TOCSY spectra showed many cross peaks in the aromatic (Figure S-4A and S-4B of the Supporting Information, peaks m–r and peaks h and i, respectively) and CH-X (Figure S-4A of the Supporting Information, peaks b, f, i, h) regions, which were confirmed by heterocorrelated HSQC spectra (Figure S-4C of the Supporting Information, peaks r–z inclusive) and CH-X (Figure S-4B of the Supporting Information, peaks m–s) regions. A smaller amount of compounds for all classes was found in ORG3 than in ORG1 and ORG2 (Table 2), thus indicating that previous transesterification had already broken most of the existing ester bonds.

**Solid Residue after Ethers Cleavage (RES4).** The residue after cleavage of ether bonds accounted for 20% of RES0 weight and provided a C/H and C/O ratio of 1.19 and 1.07, respectively, as an indication of larger unsaturation and oxidation than that for RES0 (Table 1). As compared with RES0 and intermediate fractions, RES4 was depleted of linear structures and completely deprived of alkanoic acids (Table 2).

This was confirmed by VSL NMR experiments (Table S-6 of the Supporting Information), which provided proton relaxation times in the solid state ($T_1$,H) as indicators of molecular rigidity.$^{36}$ The $T_1$,H values were larger for RES4 than for RES0 in all spectral regions and especially for sp$^2$ carbons (95–150 ppm), thereby indicating a poorer molecular flexibility and tighter packing for the final RES4 residue. $^{13}$C–CPMAS–NMR and DD (Figure S-5 of the Supporting Information) spectra were used, respectively, to characterize total and quaternary carbons. In fact, a large content of sp$^3$ carbons in $^{13}$C–CPMAS–NMR resonated at 95–150 ppm (Figure S3B), but only 66% of which were also found in DD spectra and hence identified as quaternary carbons (Figure S-5 of the Supporting Information). This suggests either totally substituted or condensed aromatic carbons in RES4. Also, the $^1$H solution spectrum of RES4 (Figure 3A) indicated a broad band for sp$^2$ signals (5.5–9.00 ppm), thereby supporting the unsaturated/oxidized nature of RES4.

$^{13}$C–CPMAS and $^1$H NMR spectra of RES4 (Figure 3) showed other compounds with carbonyl (150–185 ppm) and aliphatic (15–40 ppm) C and aliphatic H (0.5–2.2 ppm) signals, respectively. These findings were supported by the TIC chromatogram of HPSEC–ESI–MS for RES4 (Figure 6). Its two main peaks contained masses corresponding to odd- and even-numbered C$_{10}$–C$_{18}$ saturated acids, C$_{14}$–C$_{18}$ unsaturated acids, C$_{6,8}$ hydroxy-unsaturated acids, cyclic acids, and other C$_{15–17}$ unidentified acidic structures comprising four to six oxygen atoms, which were by far the most abundant components (Table S-7 of the Supporting Information). Except for four compounds, most masses in the first TIC peak were not attributable to known chemical structures (Figure 6A), whereas most recognizable structures were comprised in the second TIC signal (Figure 6B). Moreover, three cyclic acids were eluted later in the chromatogram (>40 min).

RES4 large insaturation suggests that some of these unidentified acidic compounds must be cyclic to remain chemically stable (Table S-7 of the Supporting Information). Other unidentified structures, probably linear because their empirical formula resembled that of n-alkanoic acids, were also detected in considerable amount (Table S-7 of the Supporting Information). These polar substances were also found in RES0 but not in AQU2 and must hence be bound tightly enough to humic matrices to resist previous extractions.

The 2D-COSY spectra showed correlations between CH-X hydroxylalkyl and C–H alkyl protons (Figure S-6A of the Supporting Information, peaks j and l), the latter presumably attributable to methyl groups. This correlation may be assigned to CH$_3$–CHOH–R structures, which were not observed in other extracts. Cross peaks for vicinal hydroxyl protons (Figure S-6A of the Supporting Information, peak k) and in the aromatic region (Figure S-6A of the Supporting Information, peaks m and n) may be attributed to 4-hydroxy substituted rings. The double peaks may be explained with the concomitant presence of phenol and phenol forms at the alkaline $p$H used for NMR analysis (Figure S-6 of the Supporting Information). The heterocorrelated HSQC spectrum showed few signals, mainly due to alkyl groups in fatty acids (Figure S-6C of the Supporting Information, peaks g–k). The high-field chemical shift for both $^1$H and $^{13}$C nuclei of these signals indicates the proximity of electron-withdrawing nuclei such as oxygen, thereby suggesting probable assignment to methylene adjacent to multihydroxylated structures.
**DISCUSSION**

The “Humeomics” approach adopted here involved a mild chemical fractionation to separate humic molecules from their complex suprastucture and identification of their molecular structure by advanced analytical techniques. Humic molecules solvated in organic solvents in each fractionation step (ORG 1–3) were further simplified by an SPE separation and then structurally analyzed by GC-MS after derivatization (Tables S-2, S-3, and S-5 of the Supporting Information). Conversely, humic molecules separated in aqueous phases (AQU2) and humic constituents in the original (RES0) and final (RES4) solid humic materials were characterized by ESI-MS following a HPSEC separation (Tables S-1, S-4, and S-7 of the Supporting Information). The high-resolution Orbitrap-MS detection provided reliable empirical formulas for HPSEC-separated analytes. For most of linear acids, this was sufficient to identify their structures, with the exception of structural isomerism.

Because of appropriate standards calibration, a quantitative comparison was possible between ORG fractions (measured by GC-MS) and AQU and RES (measured by LC-MS) materials (Table 2). Although the gravimetric yields of separated fractions indicated a partial weight loss as compared with RES0 (Table 1), the sum of compounds identified by GC-MS in all fractions was larger than that for bulk HA (RES0) (Table 2). This may be hardly attributed to molecular change introduced by derivatization because it similarly affected GC-MS detectability of calibration standards used to quantify unknown fractionation compounds. Therefore, our mild fractionation was successful not only in preserving the structural features of progressively removed humic molecules but also in enhancing their structural identification as specific compounds classes.

Both NMR and MS investigation showed that saturated and unsaturated n-alkanoic acids were largely present in all fractions (Table 2; Tables S-1–S-5 and S-7 of the Supporting Information). Being abundant in living cells and capable of forming hydrophobic associations during cell decay, they become a considerable part of humified suprastructure. Moreover, their presence in all separated fractions underlines their role in forming and stabilizing humic supramolecular associations. Whereas part of saturated acid is loosely bound in the humic matrix, as shown by their qualitative similarity in both RES0 and ORG1, most n-alkanoic acids are involved in ester bonds and separated only after hydrolysis steps (ORG2–3 and AQU2). Unsaturated alkanic acids were also recovered in hydrolysable AQU2 to an extent greater than their limited water solubility would allow, suggesting that aggregation with hydrophilic substances increased their repartition in aqueous media. However, a large portion of unsaturated acids was found in final RES4 residue (Table 2). Whereas RES4 molecular rigidity was generally larger than that of RES0, the T1 values of aliphatic sp-carbons were similar to those of sp-carbons (Table S-6 of the Supporting Information).

This indicates that flexible unsaturated acids were still kept incorporated in the tight RES4 residue, possibly with a bridging role between compact sp-carbons. Alternatively, new insaturation may be formed from dehydration of hydroxyl groups as a result of HI treatment.

Hydroxylated acids and n-alkanols originate from decay of plant biopolysters such as cutin and suberin or from bacterial metabolism. Various amounts of these classes were found in all separated humic fractions (Tables S-1–S-5 and S-7 of the Supporting Information). Besides characterization by MS techniques, 2D NMR showed how hydroxyl groups were bound to carbon chains (Figures S-1–S-4 and S-6 of the Supporting Information), thereby indicating their molecular architectures.

The largest abundance of hydroxy compounds was found in ORG1 and ORG2, whereas di- and trihydroxy substituted acids were prominent in both ORG2 and ORG3 fractions (Table 2). The latter compounds were thus more involved in multiple ester bonds and had a greater role in stabilizing humic suprastructure.

In fact, the sterical hindrance conferred by di- and trihydroxy-lated components to humic associations through cross-linking with multihydroxylated linear acids required the smaller alkaline nucleophile (KOH) to release them in ORG3. The abundance of multifunctional hydroxyacids might also explain the reported nonsolvent consistency of ORG3. These observations rank this compounds abundance in the order: unbound > linear esters > cross-linked esters.

Alkandioic acids produced from either α- or γ-oxidation of hydroxy acids or direct decay of plant material. Both GC-MS and LC-MS detected these compounds, thereby showing that the “Humeomics” approach enhanced understanding of the structural role of diacids. In fact, the HPSEC–ESI–MS chromatogram of RES0 (Figure 2) and its mass analysis (Table S-1 of the Supporting Information) showed that diacid eluted earlier (Figure 2B) than monocarboxyl homologues (Figure 2C). This suggests a correlation between diacid content and hydrodynamic radius of humic associates because the diacids dipolar nature allowed different interactions with extracting solutions from that of monocarboxylic acids. Moreover, quantitative results (Table 2) indicated that diacids were extensively isolated as unbound material in ORG1, whereas a lesser amount was cleaved from esters and was absent in the final RES4 residue (Table 2). In fact, HPSEC–ESI–MS showed that the shoulder corresponding to diacids in RES0 (Figure 2) disappeared in RES4 (Figure 6). These findings suggest that alkandioic acids have a less relevant role in stabilizing humic superstructures than their monoketoic acid equivalents. An explanation may reside on the diacids polar terminals that prevent their stable association with other apolar alkyl chains in humic matrices. Their greater hydrophilicity may also justify the traces of hydroxylated diacids found in AQU2 (Table 2). It is noticeable that a partial application of this fractionation procedure on a compost HA showed a more abundant content of diacids and alkanols in ORG 3 than for this soil HA, The difference in molecular content observed between the two HA suggests that the humeomic fractionation proposed here enables distinction among humic matters of different origin and formation.

Polyhydroxylated compounds such as carbohydrates, amino sugars, or their derivatives are commonly associated with humic matter. The decay of plant cellulose or fungal chitin may yield these hydrophilic substances, which become protected from further biodegradation in humic hydrophobic domains. The high hygroscopicity of hydrolysable AQU2 was probably due to large content of these substances. In fact, solution-state NMR spectra for AQU2 revealed densely populated signals in spectral regions attributed to hydroxalkyl and aminoalkyl residues (Figure S-3C of the Supporting Information, peaks a–h). Although polyhydroxylated species in AQU2 could not be structurally identified by LC-MS because of poor sensitivity of ESI sources for non acidic structures, empirical formulas inferred their presence as nitrogen-containing compounds (Table S-4 of the Supporting Information). However, they were detected in different TIC regions (Figure 5A–C), thereby indicating
association of nitrogenated polyhydroxylated structures with other compounds and different aggregation dynamics. Because empirical formulas are consistent with aminoglycosidic rather than heteroaromatic nitrogen (Table S-4 of the Supporting Information), “Humeomics” show potential to contribute to characterize unknown humic nitrogen and solve a highly controversial aspect in humic chemistry.65

Aromatic structures are found in HS.27 They were mainly detected in ORG2 and ORG3 fractions by GC-MS (Table 2), whereas other non aromatic cyclic structures were shown in RES0, RES4, and AQU2 by LC-MS (Tables S-1, S-4, and S-7 of the Supporting Information). Empirical formulas of latter compounds showed large C/H and small C/O ratios, thereby implying multiple unsaturations and few unexchangeable protons. The HPSEC distribution of non aromatic cyclic structures (Figures 2, 5, 6) differed from that of linear compounds, thereby implying a mechanism for specific supramolecular associations rather than their random inclusion in macromolecular network. Because these compounds, to our knowledge, had not been previously reported in literature, our “humeomic” approach appears to be versatile enough to reveal both aromatic and non aromatic structures. Nevertheless, an unambiguous characterization of non aromatic substances would require extensive use of MS−MS analysis, which is beyond the scope of this work.

It is interesting to note that whereas the $^{13}$C−CPMAS spectrum of RES4 (Figure 3B) showed a considerable signal around 130 ppm and the DD spectrum (Figure S-5 of the Supporting Information) revealed that ~66% of such sp$^2$ carbon pool is quaternary, the corresponding sp$^3$-rich structures were not detected by either mass spectrometry or solution $^1$H NMR. This suggests the presence of either π−π stacked fully substituted29 or highly condensed aromatic structures in residual RES4.47

Few linear hydrocarbons and steroids as a result of breakdown of plant roots cells28,48 were easily detected by GC-MS in organosoluble fractions (Table 2). Hydrocarbons were found mainly in ORG1 as unbound species, whereas steroids appeared in ORG2 and ORG3, indicating that they were tightly bound to the humic matrix and liberated only after ester hydrolyses (Table 2). The hydrophobicity of these classes well integrate around 130 ppm and the DD spectrum (Figure S-5 of the Supporting Information). Empirical formulas of latter compounds showed large C/H and small C/O ratios, thereby implying multiple unsaturations and few unexchangeable protons. Most of these compounds showed signals in aromatic C1 positions (Table S-5) which are further stabilized when these compounds are in ester forms.

### CONCLUSIONS

This “humeomic” approach promises to reach an exhaustive molecular characterization of HA and natural organic matter (NOM) in general. “Humeomic” is nondestructive of carbon−carbon bonds but isolates humic molecules with intact carbon structure, which bind to humic matrix only by weak dispersive forces or ester and ether linkages. The progressive separation of undisturbed single humic components allows their structural identification by modern chromatographic and physical−chemical techniques, thereby advancing knowledge on humus molecular characteristics as related to its origin and formation. This stepwise “Humeomics” may be generally applied to all kinds of NOM and, because of its systematic reproducibility, may lead to develop real models of humic conformational architecture. An advanced structural and stereochemical comprehension of humic suprastructures is required to relate their molecular structure to biological activities and functions in environmental ecosystems.

### ASSOCIATED CONTENT

#### Supporting Information. Chromatographic data and spectral images. This material is available free of charge via the Internet at http://pubs.acs.org.

### AUTHOR INFORMATION

**Corresponding Author**

Phone: +39 081 253 91 60. Fax: +39 081 253 91 86. E-mail: alessandro.piccolo@unina.it.

### ACKNOWLEDGMENT

A.N. gratefully acknowledges the use of the ESI-MS-Orbitrap at the Institut für Umweltforschung, Universität Dortmund, Germany.

### REFERENCES

(47) Knicker, H. Biogeochemistry 2007, 85, 91–118.