

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/26316343>

# Natural Organic Matter and the Event Horizon of Mass Spectrometry

ARTICLE in ANALYTICAL CHEMISTRY · JANUARY 2009

Impact Factor: 5.64 · DOI: 10.1021/ac800464g · Source: PubMed

CITATIONS

120

READS

95

6 AUTHORS, INCLUDING:



Norbert Hertkorn

Helmholtz Zentrum München

132 PUBLICATIONS 2,860 CITATIONS

SEE PROFILE



Matthias Witt

Bruker Corporation

106 PUBLICATIONS 2,698 CITATIONS

SEE PROFILE



Boris Peter Koch

Alfred Wegener Institute Helmholtz Centre ...

72 PUBLICATIONS 1,777 CITATIONS

SEE PROFILE



Ph. Schmitt-Kopplin

Helmholtz Zentrum München

348 PUBLICATIONS 6,609 CITATIONS

SEE PROFILE

Article

## Natural Organic Matter and the Event Horizon of Mass Spectrometry

N. Hertkorn, M. Frommberger, M. Witt, B. P. Koch, Ph. Schmitt-Kopplin, and E. M. Perdue

*Anal. Chem.*, **2008**, 80 (23), 8908-8919 • Publication Date (Web): 28 October 2008

Downloaded from <http://pubs.acs.org> on December 15, 2008

### More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)



ACS Publications  
High quality. High impact.

# Natural Organic Matter and the Event Horizon of Mass Spectrometry

N. Hertkorn,<sup>\*,†</sup> M. Frommberger,<sup>†</sup> M. Witt,<sup>‡</sup> B. P. Koch,<sup>§</sup> Ph. Schmitt-Kopplin,<sup>†</sup> and E. M. Perdue<sup>||</sup>

Institute of Ecological Chemistry, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany, Bruker Daltonics, Fahrenheitstrasse 4, D-28359 Bremen, Germany, Alfred-Wegener-Institut für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany, and School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332

Soils, sediments, freshwaters, and marine waters contain natural organic matter (NOM), an exceedingly complex mixture of organic compounds that collectively exhibit a nearly continuous range of properties (size-reactivity continuum). NOM is composed mainly of carbon, hydrogen, and oxygen, with minor contributions from heteroatoms such as nitrogen, sulfur, and phosphorus. Suwannee River fulvic acid (SuwFA) is a fraction of NOM that is relatively depleted in heteroatoms. Ultrahigh resolution Fourier transform ion cyclotron (FTICR) mass spectra of SuwFA reveal several thousand molecular formulas, corresponding in turn to several hundred thousand distinct chemical environments of carbon even without accountability of isomers. The mass difference  $\Delta m$  among adjoining C,H,O-molecules between and within clusters of nominal mass is inversely related to molecular dissimilarity: any decrease of  $\Delta m$  imposes an ever growing mandatory difference in molecular composition. Molecular formulas that are expected for likely biochemical precursor molecules are notably absent from these spectra, indicating that SuwFA is the product of diagenetic reactions that have altered the major components of biomass beyond the point of recognition. The degree of complexity of SuwFA can be brought into sharp focus through comparison with the theoretical limits of chemical complexity, as constrained and quantized by the fundamentals of chemical binding. The theoretical C,H,O-compositional space denotes the isomer-filtered complement of the entire, very vast space of molecular structures composed solely of carbon, hydrogen, and oxygen. The molecular formulas within SuwFA occupy a sizable proportion of the theoretical C,H,O-compositional space. A 100 percent coverage of the theoretically feasible C,H,O-compositional space by SuwFA molecules is attained throughout a sizable range of mass and H/C and O/C elemental ratios. The substantial differences between (and complementarity of) the SuwFA molecular formulas that are observed using six different modes of ionization (APCI, APPI, and ESI in positive and negative modus) imply considerable selectivity of the ionization process and suggest that the observed mass spectra represent simplified projections of still more complex mixtures.

Navigation through the extremely vast total space of molecular structures, for which count estimates of chemically relevant constitutional isomers<sup>1</sup> range from  $10^{60}$  to  $10^{200}$ ,<sup>2</sup> and the assessment of chemical complexity and molecular diversity are of ever increasing importance in modern chemistry, biology, and medicine.<sup>3–20</sup> A useful novel concept proposed in this article is termed the compositional space of molecules and represents the isomer-filtered complement of the entire space of molecular structures. The transformation of the total structural space of molecules into its compositional space is equivalent to its projection on a highly resolved molecular weight axis. The conversion of the entire

\* To whom correspondence should be addressed. E-mail: hertkorn@helmholtz-muenchen.de. Phone: +4989-31872834. Fax: +4989-31872705.

<sup>†</sup> Helmholtz Zentrum Muenchen.

<sup>‡</sup> Bruker Daltonics.

<sup>§</sup> Alfred-Wegener-Institut für Polar- und Meeresforschung.

<sup>||</sup> Georgia Institute of Technology.

- (1) Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P. *Anal. Bioanal. Chem.* **2007**, *389*, 1311–1327.
- (2) Petit-Zeman, S. Exploring Biological Space. Horizon Symposium, Black Point Inn, Maine, Charting Chemical Space: Finding New Tools to Explore Biology. Nature Publishing Group, 2004; 1–3.
- (3) Bohacek, R. S.; McMartin, C.; Guida, W. C. *Med. Res. Rev.* **1996**, *16*, 3–50.
- (4) Burke, M. D.; Schreiber, S. L. *Angew. Chem., Int. Ed.* **2004**, *43*, 46–58.
- (5) Dobson, C. M. *Nature* **2004**, *432*, 824–828.
- (6) Ertl, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374–380.
- (7) Fink, T.; Bruggesser, H.; Reymond, J. L. *Angew. Chem., Int. Ed.* **2005**, *44*, 1504–1508.
- (8) Fink, T.; Reymond, J. L. *J. Chem. Inf. Model.* **2007**, *47*, 342–353.
- (9) Koch, B. P.; Witt, M. R.; Engbrodt, R.; Dittmar, T.; Kattner, G. *Geochim. Cosmochim. Acta* **2005**, *69*, 3299–3308.
- (10) Nielsen, T. E.; Schreiber, S. L. *Angew. Chem., Int. Ed.* **2008**, *47*, 48–56.
- (11) Paolini, G. V.; Shapland, R. H. B.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. *Nat. Biotechnol.* **2006**, *24*, 805–815.
- (12) Paton, R. S.; Goodman, J. M. *J. Chem. Inf. Model.* **2007**, *47*, 2124–2132.
- (13) Schreiber, S. L. *Science* **2000**, *287*, 1964–1969.
- (14) Wang, M. L.; Hu, X. Q.; Beratan, D. N.; Yang, W. T. *J. Am. Chem. Soc.* **2006**, *128*, 3228–3232.
- (15) Larsson, J.; Gottfries, J.; Muresan, S.; Backlund, A. *J. Nat. Prod.* **2007**, *70*, 789–794.
- (16) Schuffenhauer, A.; Brown, N.; Selzer, P.; Ertl, P.; Jacoby, E. *J. Chem. Inf. Model.* **2006**, *46*, 525–535.
- (17) Schurer, S. C.; Tyagi, P.; Muskall, S. A. *J. Chem. Inf. Model.* **2005**, *45*, 239–248.
- (18) Sutherland, J. J.; Higgs, R. E.; Watson, I.; Vieth, M. *J. Med. Chem.* **2008**, *51*, 2689–2700.
- (19) Schuffenhauer, A.; Brown, N.; Ertl, P.; Jenkins, J. L.; Selzer, P.; Hamon, J. *J. Chem. Inf. Model.* **2007**, *47*, 325–336.
- (20) Lipkus, A. H.; Yuan, Q.; Lucas, K. A.; Funk, S. A.; Bartelt, W. F.; Schenck, R. J.; Trippe, A. J. *J. Org. Chem.* **2008**, *73*, 4443–4451.

structural space into a mass spectrum amounts to an extreme data reduction, while crucial molecular-level information is retained.

From another perspective, the mass-limited vision from the outside observer onto the molecular compositional space can be regarded as the event horizon of mass spectrometry, because structural information will remain elusive when mass spectrometry is used without additional techniques capable of resolving isomers from peaks of identical mass.<sup>21–23</sup> Here, the ultimate in information obtainable from mass spectrometry will remain confined to the depiction of the mathematically possible, and chemically relevant, molecular compositions. The compositional space is quantized as defined by the laws of chemical binding, and it will increase with molecular size and number of contributing elements recognized. The quantization of the compositional space implies fundamental restrictions in the occurrence and positioning of mass peaks, which can be probed within the scope of current ultrahigh-resolution FTICR mass spectrometry. By direct analysis of mass spectra and by the use of mathematical data analysis and data sorting schemes,<sup>24</sup> such as Kendrick mass analysis<sup>25,26</sup> and van Krevelen diagrams,<sup>27–29</sup> both the intrinsic order of the entire, quantized compositional space and the substance-specific informative order that is obtained from FTICR mass spectra of complex materials can be readily compared and contrasted. By these means, the concept of the compositional space will help establish novel and significant assignment and calibration criteria for line-rich mass spectra of complex materials.<sup>1,30</sup>

An excellent example of naturally occurring very extensive intricacy of composition is that of natural organic matter (NOM). The formation of NOM on earth preceded the evolution of life, and from early stages in the earth's history, coevolution occurred among prebiotic/abiotic molecules, NOM, and primitive and higher forms of life.<sup>1</sup> NOM does not originate from a genetic code but is rather generated through combined action of biotic and abiotic pathways, governed by common laws of thermodynamics and kinetics.<sup>1</sup> The resulting near continuum of reactivities toward inorganic ions and organic molecules acts to buffer against any environmental extremes in the geo- and biospheres that might damage life because of potent reactivity.<sup>31</sup> This primary function of NOM in the ecosphere will be most appropriately satisfied by means of a superlative array of diversity in binding sites and chemical environments. Although much experimental evidence has been consistent with the perception of NOM as a highly complex mixture, direct evidence of its prolific complexity has

only emerged with the advent of ultrahigh resolution FTICR mass spectroscopy.<sup>32–34</sup>

Unit-resolution mass spectra of NOM samples are commonly characterized by a rather continuous envelope, circumfencing many hundreds of signals (with a maximum intensity around 350–700 Da), and by the occurrence of several regular mass spacing patterns.<sup>34–36</sup> Mass spacings based on methylene ( $\Delta m = \pm 14$  Da) and double-bond equivalents (DBE,  $\Delta m = \pm 2$  Da) dominate,<sup>36</sup> simply because they refer to these most fundamental building blocks of organic chemistry. Unfortunately, linear combinations of methylene and DBE cannot be distinguished from oxygen-containing fragments such as H<sub>2</sub>O ( $\Delta m = \pm 18$  Da), CO ( $\Delta m = \pm 28$  Da), and CO<sub>2</sub> ( $\Delta m = \pm 44$  Da) at this level of resolution, leaving descriptive mass listing and mass comparison as one of the few remaining options for mass analysis.<sup>37,38</sup> In case of MS<sup>n</sup> experiments, mass loss of 18 Da (H<sub>2</sub>O) and 44 Da (CO<sub>2</sub>) dominate the mass spectra of NOM, providing limited assistance in molecular-level assignment. Most of the informative, i.e., substance-specific, molecular-level order in mass spectra of NOM resides in the patterns with (sub)millimass spacing (mDa), which are accessible only via FTICR mass spectrometry at high-field. Good-quality FTICR mass spectra of NOM show several thousands to ten-thousands of resolved signals, for which molecular formulas have been computed across a sizable mass range.<sup>34,39,40</sup> The conventional analysis of these line-rich mass spectra of NOM is incapable of furnishing useful information about molecular composition and structure.<sup>24</sup> The mass peaks representing pairs of molecules of similar structure and function, like CH<sub>2</sub> homologues, are widely separated on the mass axis. In general, any pair of mass peaks separated by less than 2.1057 Da ( $\pm$ DBE), irrespective of its position on the mass axis and independent of its composition (C<sub>n</sub>H<sub>m</sub>O<sub>q</sub>Z<sub>Σ</sub>; Z, any combination of elements) necessarily represents considerably dissimilar chemical structures. This key relationship in general refers to any pair of mass peaks within any unit mass cluster, irrespective of elemental composition. Accordingly, inconspicuous and seemingly minuscule intensity and mass shifts between FTICR mass spectra of complex systems like NOM, not easily recognized in standard display and interpretation,<sup>24</sup> could in fact represent major and even dramatic alterations in the composition and molecular-level structure.

## EXPERIMENTAL SECTION

**FTICR Mass Spectrometry.** Ultrahigh resolution mass spectra were acquired with a Bruker APEX Qe Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) equipped with a 9.4 T superconducting magnet and an Apollo II electrospray

- (21) Barnes, C. A. S.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *Anal. Chem.* **2002**, *74*, 26–36.
- (22) Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. *Anal. Chem.* **2006**, *78*, 4161–4174.
- (23) Trimpin, S.; Plasencia, M.; Isailovic, D.; Clemmer, D. E. *Anal. Chem.* **2007**, *79*, 7965–7974.
- (24) Meija, J. *Anal. Bioanal. Chem.* **2006**, *385*, 486–499.
- (25) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. N. *Anal. Chem.* **2001**, *73*, 4676–4681.
- (26) Kendrick, E. *Anal. Chem.* **1963**, *35*, 2146–2154.
- (27) Kim, S.; Kramer, R. W.; Hatcher, P. G. *Anal. Chem.* **2003**, *75*, 5336–5344.
- (28) Visser, S. A. *Environ. Sci. Technol.* **1983**, *17*, 412–417.
- (29) Wu, Z. G.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2004**, *76*, 2511–2516.
- (30) Hertkorn, N.; Frommberger, M.; Perdue, E. M.; Thaller, S.; Holzmann, E.; Schmitt-Kopplin, P.; Humic Substances-Linking Structure to Functions, Schriftenreihe des Lehrstuhls für Wasserwirtschaft und der DVGW-Forschungsstelle am Engler-Bunte-Institut der Universität Karlsruhe (TH); **2006**; 313–316.
- (31) MacCarthy, P. *Soil Sci.* **2001**, *166*, 738–751.

- (32) Koch, M. A.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17272–17277.
- (33) Kujawinski, E. B. *Environ. Forensics* **2002**, *3*, 207–216.
- (34) Stenson, A. C.; Marshall, A. G.; Cooper, W. T. *Anal. Chem.* **2003**, *75*, 1275–1284.
- (35) Kujawinski, E. B.; Freitas, M. A.; Zang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. *Org. Geochem.* **2002**, *33*, 171–180.
- (36) Plancque, G.; Amekraz, B.; Moulin, V.; Toulhoat, P.; Moulin, C. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 827–835.
- (37) Reemtsma, T.; These, A. *Environ. Sci. Technol.* **2005**, *39*, 3507–3512.
- (38) These, A.; Reemtsma, T. *Anal. Chem.* **2003**, *75*, 6275–6281.
- (39) Koch, B. P.; Dittmar, T.; Witt, M.; Kattner, G. *Anal. Chem.* **2007**, *79*, 1758–1763.
- (40) Mopper, K.; Stubbins, A.; Ritchie, J. D.; Bialk, H. M.; Hatcher, P. G. *Chem. Rev.* **2007**, *107*, 419–442.

source with ion funnel technology. Spectra were acquired in positive and negative ion mode using electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photo ionization (APPI). An amount of 0.5 mg of Suwannee river fulvic acid (SuwFA; International Humic Substances Society, IHSS, 1S101F, Suwannee River Fulvic Acid Standard I) was dissolved in 1 mL of 50% aqueous methanol (2  $\mu\text{L}$  of formic acid was added for positive ion measurements) and was introduced into the ion source by syringe infusion with a flow rate of 120  $\mu\text{L h}^{-1}$  in the ESI mode and 12 mL  $\text{h}^{-1}$  in the APCI and APPI modes. The drying gas temperature was set to 250  $^{\circ}\text{C}$  in all ionization modes. The nebulizer gas temperature was set to 400  $^{\circ}\text{C}$  in the APCI and APPI experiments. The spectra were externally calibrated on arginine clusters with a 10  $\mu\text{g mL}^{-1}$  solution in 50% methanol using a linear calibration and then internally recalibrated with fatty acids in the negative ion mode and siloxanes in the positive ion mode. This results in a mass accuracy better than 1 ppm with an average error of all detected  $[\text{C}_n\text{H}_{m-1}\text{O}_q]^{-}$  peaks in negative and  $[\text{C}_n\text{H}_{m+1}\text{O}_q]^{+}$  peaks in positive ion mode of only 0.29 ppm. The spectra were acquired with 2 M data points with a mass range of 145–2000  $m/z$  in the negative ion mode (transient of 1.05 s) and 174–2000  $m/z$  in the positive ion mode (transient of 1.26 s) resulting in a resolving power of 250 000 at  $m/z$  300 (150 000 at  $m/z$  500) in the negative ion mode and 300 000 at  $m/z$  300 (180 000 at  $m/z$  500) in the positive ion mode, respectively. A single sine apodization was performed before Fourier transformation of the time-domain signal. The ion accumulation time in the ion source was set to 0.5–2 s in the positive ion mode and 0.2–0.5 s in the negative ion mode for each scan depending on the signal-to-noise in a single scan. A total of 120–500 scans were added for each mass spectrum.

**Validity Criteria To Determine Elemental Compositions from Mass Spectra.** Van Krevelen plot analysis was performed on SuwFA molecular compositions  $\text{C}_n\text{H}_m\text{O}_q$  derived from negative ions  $[\text{C}_n\text{H}_{m-1}\text{O}_q]^{-}$  and positive ions  $[\text{C}_n\text{H}_{m+1}\text{O}_q]^{+}$  with a signal-to-noise ratio of 3 or greater for both  $^{12}\text{C}_n\text{H}_m\text{O}_q$  and  $^{12}\text{C}_{n-1}^{13}\text{CH}_m\text{O}_q$ . Formulas were considered valid only if the following criteria were met (1)  $M_{\text{ion}} \leq 1000$  Da; (2) error  $\leq 0.5$  ppm; (3)  $^{13}\text{C}$  peak (defined as a  $M_{\text{ion}} + 1.003\,354$  Da) detected at subppm accuracy; (4)  $m$  even (nitrogen rule in absence of nitrogen, applied to the neutral molecule); (5)  $n \geq 1$  (must contain C); (6)  $2 \leq m \leq 2n + 2$  ( $\text{CH}_2$  to saturated linear alkanes, applied to the neutral molecule); (7)  $0 \leq q \leq n + 2$  (formally including  $\text{H}_2\text{CO}_3$ ); (8) unsaturation (DBE)  $\leq 40$ .

**Computation of the C,H,O-Compositional Space.** An algorithm to compute all chemically relevant C,H,O- (and C,H,N-, Figure 1 in the Supporting Information) molecular compositions within the mass range from 0–1000 Da was designed and exercised with Matlab routines. It was based upon the valences of carbon, oxygen, nitrogen, and hydrogen and resulted in uncharged C,H,O- and C,H,N-molecules, respectively. The observance of the nitrogen rule resulted in C,H,O-molecules of even nominal mass and C,H,N-molecular compositions of even nominal mass in the case of an even count of nitrogen atoms and an uneven mass in the case of uneven counts of nitrogen atoms. No other radicals or ions were recognized. A further restriction imposed referred to the elemental ratios ( $\text{H/C ratio} \leq 2$ ,  $\text{O(N)/C ratio} \leq 1$ ), which is typical of molecules detected by mass spectrometry in

SuwFA and natural organic matter in general. The obtained mass lists were transformed into Excel format for sorting and preparation of the figures shown.

**Venn Diagrams.** The Venn diagrams provided in Figure 2 were computed with Venn Diagram Plotter, version 1.3.3103.24313 (June 30, 2008; Kyle LITTLEFIELD, PNNL, Richland, WA) from data equivalent to Table 1 (or Figure 2 in the Supporting Information). These three-circle Venn diagrams represent computer optimized representations that are not exactly to scale except coincidental in rare circumstances.

## RESULTS AND DISCUSSION

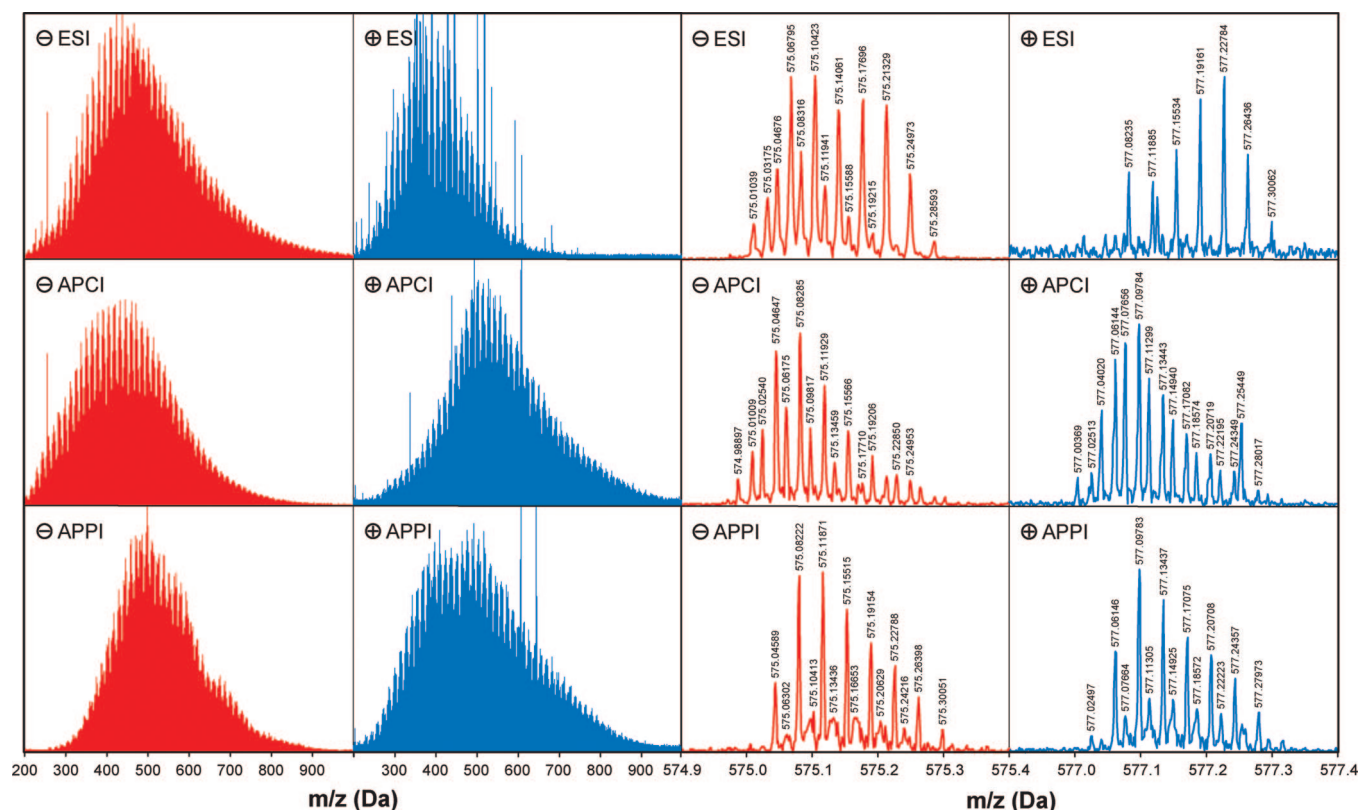
Suwannee river fulvic acid (SuwFA) is an operationally defined fraction of Suwannee river natural organic matter (SuwNOM) and comprises its water-soluble complement at any pH value. On a mass basis, SuwFA contains 52.34% C, 4.36% H, 42.98% O, 0.67% N, 0.46% S, and 0.004% P. Heteroatoms (N, S, and P) account collectively for only about 1% of its mass; so, to a very good approximation, SuwFA can be considered as a mixture of molecules, each of which has a molecular formula of  $\text{C}_n\text{H}_m\text{O}_q$ . We report here a comprehensive mass spectral analysis of SuwFA, in which electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photo ionization (APPI), each in positive and negative ion modes, have yielded six complementary data sets of ions (Figure 1).

In a rather conservative analysis of these six data sets, only signals having a signal-to-noise ratio of 3 or greater for both the ( $^{12}\text{C}_n\text{H}_m\text{O}_q$ ) and the ( $^{12}\text{C}_{n-1}^{13}\text{CH}_m\text{O}_q$ ) peaks were processed further. The subset of “valid” peaks was further restricted to those that met the strict exclusion criteria,<sup>39,41</sup> cf. Experimental Section. The three complementary negative mode FTICR mass spectra of SuwFA produced a total of 16 147 different resolved ion signals in the mass range from 200 to 700 Da, of which 3678 ions (22.8%) could be assigned as members of the C,H,O-compositional space, i.e., the complement of all feasible C,H,O molecular compositions (Table 1 and Figures 1 and 2 of the Supporting Information). Only 611  $\text{C}_n\text{H}_m\text{O}_q$  ions (3.8% of detected  $\text{C}_n\text{H}_m\text{O}_q$  compositions) were observed in all three methods of negative ionization. Less conservative exclusion criteria and recognition of peaks with lesser signal-to-noise ratio increase both “valid” peak counts and the proportion of  $\text{C}_n\text{H}_m\text{O}_q$  molecules (data not shown).

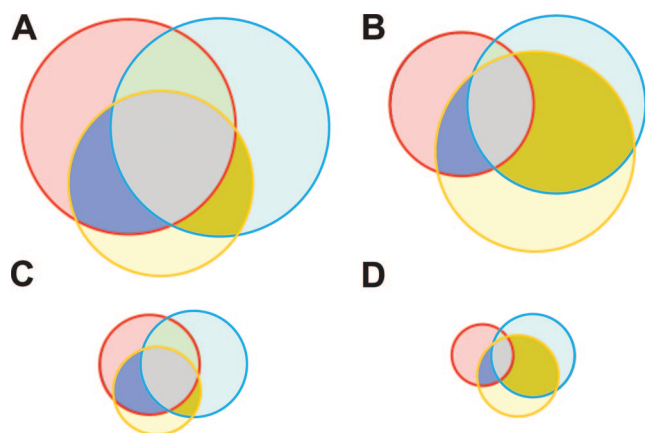
Collectively, the three modes of positive ionization yielded 11 820 positive ions (about  $2/3$  of the count of negative ions), of which 2555 (21.6%) could be assigned unique  $\text{C}_n\text{H}_m\text{O}_q$  molecular compositions. Only 204 (1.7%) positive ions derived from  $\text{C}_n\text{H}_m\text{O}_q$  compositions were common to all three ionization modes. The total numbers of SuwFA ions and  $\text{C}_n\text{H}_m\text{O}_q$  molecules, as well as the count of ions unique to any single method, decrease in the order  $\text{APCI} > \text{ESI} > \text{APPI}$  for negative ions and  $\text{APPI} > \text{APCI} > \text{ESI}$  for positive ions (Figures 2 and 3, Table 1, and Figure 2 in the Supporting Information). Within both the positive and negative ionization modes, the comparison of total ion counts and those representing solely C,H,O-molecular compositions indicate very similar ionization selectivity for these two groups of SuwFA molecules. This situation is very different when positive and negative ionization of SuwFA is compared: here, a very profound variance in selectivity is observed for both total ion counts and

(41) Kujawinski, E. B.; Behn, M. D. *Anal. Chem.* **2006**, *78*, 4363–4373.





**Figure 1.** 9.4 T FTICR mass spectra of SuwFA in ESI (top), APCI (middle), APPI modes (bottom) in negative (left) and positive (right) ionization mode for the mass range 200–1000 Da (left two columns) and the nominal mass cluster” at 576 Da (right two columns; cf. parts A and B of Figure 3 in the Supporting Information).



**Figure 2.** Venn diagrams, scaled with respect to (A, B) total counts of SuwFA-derived FTICR mass spectral peaks and (C, D) identified C,H,O-molecular compositions, depending on ionization modus: ESI (red), APCI (blue), APPI (yellow). (A) All negative ions, (B) all positive ions, (C) all negative ions with C,H,O-compositions, and (D) all positive ions with C,H,O-compositions.

those representing C,H,O-molecular compositions (Figure 2). Similar considerations apply for the comparative analysis of the van Krevelen plots of molecular formulas that are derived from negative (or positive) ions which also reveal high compositional selectivity for the three modes of ionization (Figure 3). This outcome is consistent with numerous studies of factors that affect

ionization efficiencies. In general, ESI ionizes a wide range of analytes; however, APCI and APPI tend to ionize less polar and smaller molecules.<sup>42–44</sup> The extent of complementarity between positive and negative SuwFA ions and between the methods is striking: APCI shares more than two-thirds of its total number of positive ions with those of APPI and less than a fifth with those of ESI; in contrast, APCI shares about half of its negative ions with those of both ESI and APPI (Figures 2 and 3; Figure 2 in the Supporting Information).

APPI-derived van Krevelen diagrams (Figure 3) and CH<sub>2</sub>-based Kendrick mass analysis (Figure 4) demonstrate the most well-defined discrimination of positive and negative ions according to the oxygenation of the analytes: across the entire bandwidth of relative unsaturation (i.e., H/C ratios of C<sub>n</sub>H<sub>m</sub>O<sub>q</sub> molecules), more oxygenated molecules are preferentially ionized in negative APPI mode and vice versa. In APCI-derived van Krevelen diagrams, positive and common (to positive and negative) ions represent restricted compositional subspaces of areas fully covered by negative ions. Similarly, ESI common ions occupy a subspace half the area of ESI negative ions. Kendrick mass defect analysis (Figure 4) reveals a unique preferential ionization of small, oxygen-deficient molecules in the negative APCI mode; both ESI and APPI preferentially ionize oxygen-rich molecules in the negative mode.

The patterns observed in the van Krevelen diagrams shown here (Figure 3) closely resembled the FTICR mass spectra-derived van Krevelen diagrams which were obtained previously from

(42) Cai, S. S.; Syage, J. A. *J. Chromatogr., A* **2006**, *1110*, 15–26.

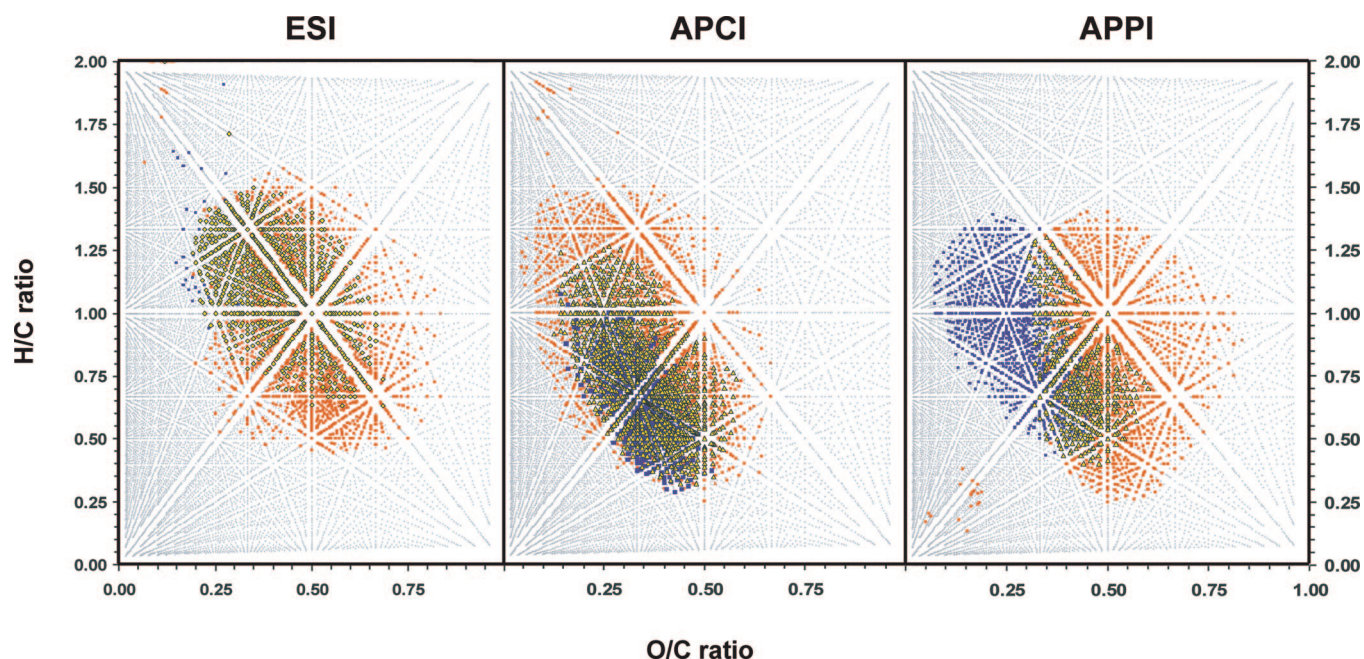
(43) Hanold, K. A.; Fischer, S. M.; Cormia, P. H.; Miller, C. E.; Syage, J. A. *Anal. Chem.* **2004**, *76*, 2842–2851.

(44) Keski-Hyynilä, H.; Kurkela, M.; Elovaara, E.; Antonio, L.; Magdalou, J.; Luukkainen, L.; Taskinen, J.; Kostianen, R. *Anal. Chem.* **2002**, *74*, 3449–3457.

**Table 1. Counts of SuwFA-Derived  $C_nH_mO_n$  Ions and Number of FTICR Mass Spectral Peaks<sup>a</sup>**

ionization method	positive ions			negative ions		
	number of peaks	identified C,H,O-compositions	percent	number of peaks	identified C,H,O-compositions	percent
all ions	11 820	2555	21.6	16 147	3678	22.8
APCI total	6 504	1542	23.7	10 011	2393	23.9
APPI total	8 172	1445	17.7	7 103	1644	23.1
ESI total	4 220	840	19.9	9 600	2123	22.1
APCI only	1 664	618	37.1	4 315	1083	25.1
APPI only	2 655	392	14.8	1 237	241	19.5
ESI only	1 915	475	24.8	3 207	482	15.0
APCI and APPI	4 771	908	19.0	4 193	842	20.1
APCI and ESI	1 554	220	14.2	4 701	1079	23.0
APPI and ESI	2 236	349	15.6	4 890	1172	24.0
APCI and APPI and ESI	1 490	204	13.7	3 198	611	19.1

<sup>a</sup> Counts and percentages of positive (left) and negative (right) ions, sorted according to total number of peaks, peaks representing  $C_nH_mO_q$  formulas and percentage of valid  $C_nH_mO_q$  compositions related to total number of peaks, depending on ionization method and charge.



**Figure 3.** Selectivity of SuwFA ionization according to ionization method (ESI, APCI, and APPI, respectively) as displayed in van Krevelen diagrams of C,H,O-molecular formulas  $C_nH_mO_q$  derived from positive (blue), negative (red), and common to positive and negative (yellow) ions of SuwFA; gray dots represent the feasible C,H,O-compositional space of 18 141 elemental compositions  $C_nH_mO_q$  within the mass range from 200 to 700 Da and within the H/C and O/C ratios given.

natural organic matter samples isolated from soil-, freshwater-, and marine ecosystems,<sup>9,27,40,45–48</sup> although higher S/N ratio available in this study enabled an improved coverage of molecular compositions.

This conspicuous accordance indicated the existence of rules governing the appearance of these recurring periodical patterns. These were found to reflect the fundamental relationships how molecules, composed solely of carbon, hydrogen, and oxygen ( $C_nH_mO_q$ ) are being assembled as imposed by the laws of chemical binding. These laws refer to the mass and valences of the atoms and the entire mass range allowed. They define all feasible linear

combinations of atoms in accordance with the nitrogen rule: i.e., only molecules of even nominal mass and an even count of hydrogen are allowed for any combination of C, H, and O atoms to result in the C,H,O-compositional space of chemically relevant molecules. If this C,H,O-compositional space is projected onto a highly resolved mass axis, a mass spectrum with ultimate capacity to depict molecular dissimilarity is obtained, based on exact mass and molecular composition alone; it will, however, remain incapable of resolving isomers. The intrinsic order of that compositional space results in regular patterns in standard van Krevelen diagrams (Figure 1 in the Supporting Information) which are two-dimensional projections of a three-dimensional, mass-edited van Krevelen diagram<sup>29</sup> (cf. Figure 7). The regularity of these standard patterns is defined by the valences of the contributing atoms (e. g., hydrogen = 1, carbon = 4, nitrogen = 3, oxygen = 2), and the coverage is defined by the mass range applied (Figure 1 in the Supporting Information).

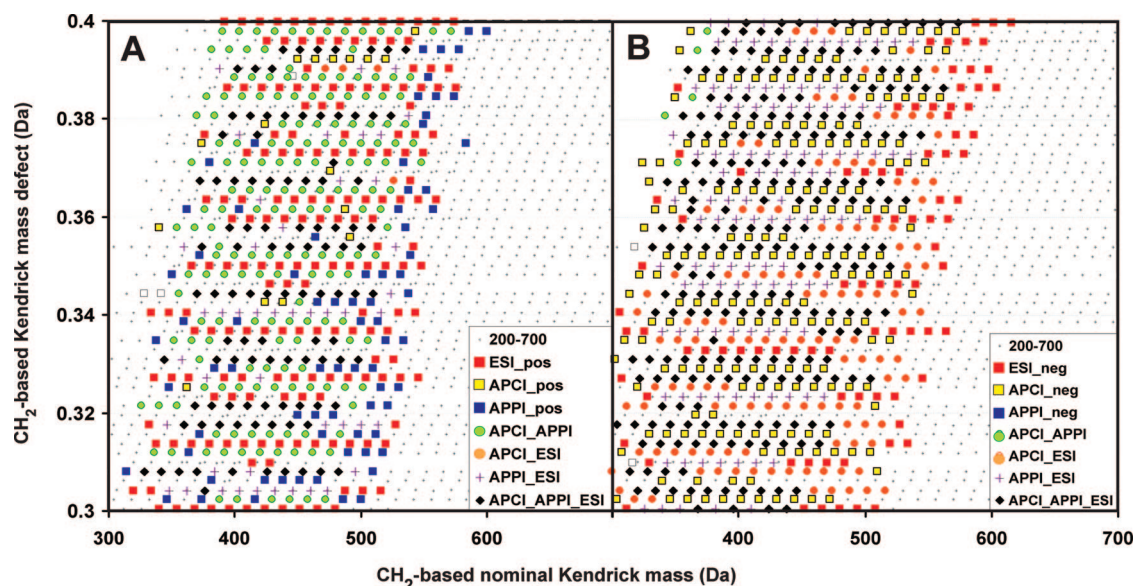
(45) Kramer, R. W.; Kujawinski, E. B.; Hatcher, P. G. *Environ. Sci. Technol.* **2004**, *38*, 3387–3395.

(46) Sleighter, R. L.; Hatcher, P. G. *J. Mass Spectrom.* **2007**, *42*, 559–574.

(47) Dittmar, T.; Whitehead, K.; Minor, E. C.; Koch, B. P. *Mar. Chem.* **2007**, *107*, 378–387.

(48) Einsiedl, F.; Hertkorn, N.; Wolf, M.; Frommberger, M.; Schmitt-Kopplin, P.; Koch, B. P. *Geochim. Cosmochim. Acta* **2007**, *71*, 5474–5482.





**Figure 4.** Kendrick mass analysis of mass spectra of SuwFA. Within both positive (A) and negative (B) ionization, a  $\text{CH}_2$ -based Kendrick mass defect analysis arranges the seven groups of ionization preferences of nominally identical C,H,O-molecular compositions in a nonstatistical, orderly manner. The selectivity of positive and negative ionization of SuwFA varies profoundly. Preferences of SuwFA ionization relate significantly to the values of Kendrick mass defect: compounds of identical class are preferentially ionized by a single one or only a few methods. Within a given Kendrick mass defect class, mass can be further decisive in directing the selectivity of ionization. Positions of the C,H,O-compositional space not occupied by mass peaks are denoted by the tiny gray dots.

Up to a nominal mass of 1550 Da, the compositional space of chemically relevant C,H,O-molecules projects into clearly separated “nominal-mass” clusters (Figure 5 and Figures 3B and 5A in the Supporting Information). Addition and removal of any chemically relevant molecular fragments, like double-bond-equivalents (DBE;  $\Delta m$ ,  $\pm 2.0157$  Da) and methylene ( $\text{CH}_2$ ;  $\Delta m$ ,  $\pm 14.0156$  Da), is associated with mass shifts across the boundaries of these “nominal-mass” clusters. Conversely, any pairs of chemical structures derived from any pairs of peaks within any nominal mass cluster are necessarily considerably dissimilar and disagree at least to the extent of significant functional group alteration (Table 2, Figure 4 in the Supporting Information) and bond displacement.

Within any nominal mass cluster of  $\text{C}_n\text{H}_m\text{O}_q$  composition, four fundamental mathematical combinations of element-shifts characterize the C,H,O-compositional space:  $+4\text{C}/-3\text{O}$  ( $\Delta m_1 = +15.26$  mDa),  $+\text{CH}_4/-\text{O}$  ( $\Delta m_2 = +36.39$  mDa),  $+12\text{H}/-\text{C}$  ( $\Delta m_3 = +93.90$  mDa), and  $+16\text{H}/-\text{O}$  ( $\Delta m_4 = +130.29$  mDa; Figure 3A in the Supporting Information and Table 2). The minimum chemical dissimilarity imposed by variance in molecular composition corresponds to the formal exchange of  $\text{CH}_4$  against oxygen, which can be visualized by a nominal exchange of a  $\text{HC}-\text{CH}_3$  unit against a carbonyl group. The exchange of one carbon atom against 12 hydrogen atoms (as well as the other alterations described above; cf. parts A and B of Figure 5 in the Supporting Information and Table 2) would impose much more severe functional group alterations and bond displacements. Hence, any pairs of molecular compositions of unequal mass [ $0 < \Delta m < 2.0157$  Da (DBE)] within C,H,O-nominal mass clusters will exhibit a significant mandatory chemical dissimilarity with respect to composition, structure, and reactivity, irrespective of chemical environment.

The knowledge of the molecular masses comprising the C,H,O-compositional space enabled to unambiguously differentiate mo-

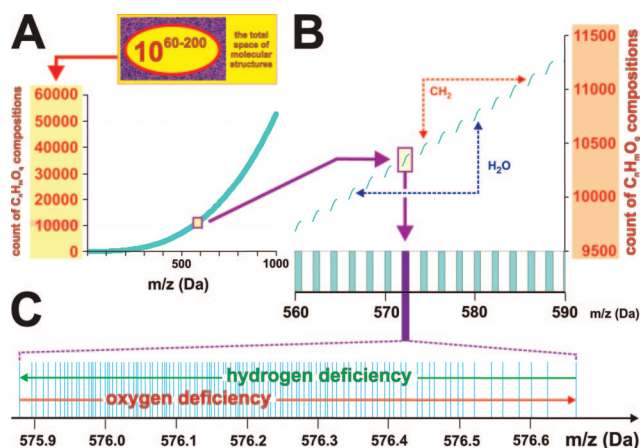
lecular compositions composed solely of C, H, and O from any other heteroatom-containing ( $\text{C}_n\text{H}_m\text{O}_q\text{Z}_\Sigma$ ;  $\text{Z}_\Sigma$ , any combinations of heteroatoms) molecular compositions in the acquired mass spectra.

The key information which enabled the positive identification of molecular formulas composed solely of C, H, and O, based on accurate mass determination, resulted from the finding that any mass spacing  $\Delta m$  (the mass difference between neighboring C,H,O-molecules if sorted according to mass) of any C,H,O-molecules in the mass range from 0–1000 Da was large compared with standard high-field FTICR mass resolution. The minimum mass spacing  $\Delta m$  between neighboring masses of any C,H,O-molecules between 0 and 1000 Da is 1.1474 mDa and corresponds to an alteration of molecular composition of  $\Delta = \text{C}_{43}-\text{H}_{20}\text{O}_{31}$ . This (mathematical) combination becomes available for C,H,O-molecular formulas exceeding mass 908.1554 Da (Table 2). The mass spacing  $\Delta m$  between any contiguous C,H,O-molecule will become even larger in the case of a smaller mass range considered. For example, in the mass range below 908.1554 Da, the minimum mass difference  $\Delta m$  among any adjoining C,H,O-molecules was 2.36 mDa (Table 2).

All mathematically possible mass differences  $\Delta m$  between adjoining mass peaks within C,H,O-nominal mass clusters of the entire C,H,O-compositional space range below 1 Da ( $\Delta m \leq 36.3854$  mDa) and are arranged into eight distinct groups, each corresponding to a specific variance in atom counts (Table 2).

Computation of all mass differences among adjoining mass peaks leads to continually decreasing distances in-between the nominal mass clusters of C,H,O-composition with increasing mass (Figure 5A in the Supporting Information). Within the nominal C,H,O-mass clusters, eight groups of mass differences  $\Delta m$ , the mass difference between adjoining mass peaks, define the entire C,H,O-compositional space. Each of these  $\Delta m$  represents an unique combination of molecular fragments (Figures 4 and 5 in





**Figure 5.** Fundamental characteristics of the C,H,O-compositional space. (A) The transformation of the vast total space of chemically relevant constitutional isomers into its isotope-filtered complement, the compositional space, is equivalent to its projection on a high-resolution mass axis and amounts to a tremendous data reduction. (B) Up to a nominal mass of 1550 Da, the compositional space of C,H,O-molecules arranges into clearly separated “nominal-mass” clusters. Addition and removal of any chemically relevant subunits, like double-bond-equivalents (DBE;  $\Delta m = 2.0157$  Da) and methylene ( $\Delta m = 14.0156$  Da) is associated with mass shifts across the boundaries of these “nominal-mass” clusters. Conversely, within any nominal mass cluster, a mandatory, extensive variance in chemical composition and structure is imposed upon any pairs of  $C_nH_mO_q$  molecular compositions with  $0 < \Delta m < 2.0157$  Da. This general relationship even extends to molecular compositions with any combination of heteroatoms  $C_nH_mO_qZ_\Sigma$  ( $Z_\Sigma$ , any combination of heteroatoms). (C) Oxygen-rich molecules dominate the lower mass range (cumulative negative mass defect of  $^{16}\text{O}$ , 15.9949 Da), while hydrogen-rich molecules (cumulative positive mass defect of  $^1\text{H}$ , 1.0078) preponderate the higher mass range of each  $C_nH_mO_q$  nominal mass cluster. Here, the cluster of  $C_nH_mO_q$  compositions with nominal mass 576 is selected; its mass sweep extends to 749 mDa, ranging from  $\text{C}_{24}\text{O}_{18}$  (575.908 Da) up to  $\text{C}_{41}\text{H}_{84}$  (576.657 Da); see also parts A and B of Figure 3 in the Supporting Information.

the Supporting Information and Table 2). The already mentioned variance of formal exchange of  $\text{CH}_4$  against oxygen ( $\Delta m = 36.39$  mDa) represents the least drastic molecular variance among adjoining pairs of C,H,O-molecular compositions. Further decrease of the mass differences  $\Delta m$  among adjoining mass peaks leads to ever more pronounced molecular disparity. Because of the larger size of molecular fragments contributing to  $\Delta m$ , specific minimum values of C,H,O-molecular weight must be exceeded before certain mass differences  $\Delta m$  can appear in the mass spectra. For instance, no  $\Delta m$  lesser than 2.3628 mDa (1.1474 mDa) can occur for any combination of  $C_nH_mO_q$ -molecules in the mass range up to 548.026 25 (908.155 40) Da (Table 2). Any of these mass spacings  $\Delta m$  can be easily resolved within the scope of current FTICR mass spectrometry.

In conclusion, any mass spacing  $\Delta m$  between pairs of molecular composition which amounts to less than  $\Delta m = 2.1057$  Da ( $\pm\text{DBE}$ ) necessarily implies considerable molecular dissimilarity. Further decrease of  $\Delta m$  imposes an ever increasing mandatory molecular disparity (valid for any  $\Delta m \leq 36.39$  mDa). This relationship operates within the entire C,H,O-compositional space, i.e., in-between as well as within nominal mass clusters (Figure 5 in the Supporting Information) and imposes an enforced growth of dissimilarity in composition and, necessarily, chemical structure.

While direct comparison of mass spectra is viable, van Krevelen diagrams, in particular, efficiently expose compositional characteristics of complex organic materials and allow both the intrinsic order of the quantized compositional space and the informative order of the sample composition to be projected onto a common area (Figures 3 and 5).<sup>1,30</sup> It has to be noted that very closely related concepts as developed here for the van Krevelen diagrams also apply for Kendrick mass defect analyses (cf. Figure 4).

It was noted earlier that ESI, APCI, and APPI yield a higher number of resolved signals and unique molecular formulas in the negative ion mode than in the positive ion mode. It was unclear at that point whether this disparity is attributable to greater compositional diversity, greater mass range, or both. Molecular formulas that arise collectively from the negative ion mode also provide more extensive coverage of theoretically possible H/C and O/C atomic ratios in van Krevelen plots (and thus greater compositional diversity) than molecular formulas obtained from the positive ion mode (Figure 6). It is difficult to ascertain from the raw spectra, however, whether the population of negative ions also have collectively a greater mass range (Figure 1).

Another view of the compositional diversity of SuwFA is provided by consolidating all resolved peaks and  $C_nH_mO_q$  molecular formulas that are obtained from positive ions and negative ions into two separate groups. It is then possible to isolate the resolved peaks and  $C_nH_mO_q$  molecular formulas that are found only in positive ion modes (3526, 584), in negative ion modes (7853, 1715), and both modes (8294, 1971). The considerable extent of noncongruence of positive and negative ions (Figure 7) indicates the occurrence of a sizable proportion of SuwFA molecules that apparently do not contain functional groups which would be easily ionized in both the positive and negative ionization modes under the conditions applied.

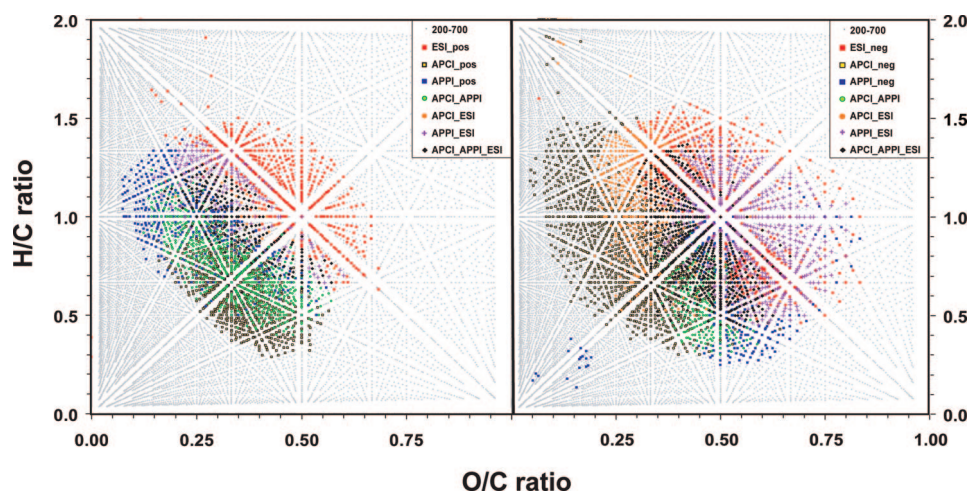
With the extremely conservative assumption that there are no structural isomers and that nominally identical peak positions and  $C_nH_mO_q$  formulas derived from positive and negative ions necessarily represent identical molecules, the minimum possible number of unique C,H,O-molecules in SuwFA is calculated to be 4270. With the use of the known  $C_nH_mO_q$  molecular formulas of the 4270 unique molecules, the corresponding minimum numbers of unique chemical environments for C, H, and O are readily calculated to be  $1.19 \times 10^5$ ,  $1.01 \times 10^5$ , and  $0.46 \times 10^5$ , respectively (Figure 7 and Table 3). Such diversity of chemical environments for carbon and hydrogen is manifested in considerable overlap of NMR resonance frequencies and thus accounts for the low resolution signatures of  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectra of SuwFA specifically and of NOM and related materials in general (data not shown).<sup>48</sup>

The informative order of the consolidated data set and the intrinsic order of C,H,O-compositional space were compared within an arbitrarily selected, mass-restricted C,H,O-compositional space having atomic H/C ratios of 0.3–1.5, O/C ratios of 0.1–0.7, and molecular masses of 200.011 Da ( $\text{C}_{11}\text{H}_4\text{O}_4$ ) to 700.507 Da ( $\text{C}_{46}\text{H}_{68}\text{O}_5$ ). The intent of this comparison was to estimate the degree of coverage of possible molecular formulas by those formulas found in the SuwFA sample. The intrinsic order of this C,H,O-compositional space consists of a total of 6503 unique  $C_nH_mO_q$  molecular formulas (Figure 7). In this limited sector, 1715 molecular formulas were detected only as negative ions, 584

**Table 2. Unique Nominal Mass Differences  $\Delta m$  Which Define Any Spacing between Adjoining C,H,O-Molecules within Nominal Mass Clusters of the Entire C,H,O-Compositional Space in the Mass Range from 0–1000 Da<sup>a</sup>**

index	nominal variance of fragments defining $\Delta m$	mass difference among adjoining mass peaks $\Delta m$ (mDa)	minimum mass of first occurrence of $\Delta m$ in the C,H,O-compositional space (Da)	formula with minimum mass for $\Delta m$ to occur (C,H,O) <sub>x</sub>	formula of next higher mass (C,H,O) <sub>x+1</sub>	mass of molecule (C,H,O) <sub>x+1</sub> (Da)
A	C <sub>43</sub> –H <sub>20</sub> O <sub>31</sub>	1.1474	908.155 40	C <sub>31</sub> H <sub>40</sub> O <sub>31</sub>	C <sub>74</sub> H <sub>20</sub>	908.156 49
B	H <sub>12</sub> O <sub>18</sub> –C <sub>25</sub>	2.3628	548.026 25	C <sub>44</sub> H <sub>4</sub> O <sub>1</sub>	C <sub>19</sub> H <sub>16</sub> O <sub>19</sub>	548.028 56
C	C <sub>18</sub> –H <sub>8</sub> O <sub>13</sub>	3.5102	382.074 74	C <sub>13</sub> H <sub>18</sub> O <sub>13</sub>	C <sub>31</sub> H <sub>10</sub>	382.078 25
D	H <sub>4</sub> O <sub>5</sub> –C <sub>7</sub>	5.8730	164.062 61	C <sub>13</sub> H <sub>8</sub>	C <sub>6</sub> H <sub>12</sub> O <sub>5</sub>	164.068 48
E	C <sub>11</sub> –H <sub>4</sub> O <sub>8</sub>	9.3832	230.006 27	C <sub>8</sub> H <sub>6</sub> O <sub>8</sub>	C <sub>19</sub> H <sub>2</sub>	230.015 66
F, $m_1$	C <sub>4</sub> –O <sub>3</sub>	15.2562	86.000 40	C <sub>3</sub> H <sub>2</sub> O <sub>3</sub>	C <sub>7</sub> H <sub>2</sub>	86.015 65
G	H <sub>4</sub> O <sub>2</sub> –C <sub>3</sub>	21.1292	60.000 00	C <sub>5</sub>	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	60.021 13
H, $m_2$	CH <sub>4</sub> –O	36.3854	42.010 57	C <sub>2</sub> H <sub>2</sub> O	C <sub>3</sub> H <sub>6</sub>	42.046 95
$m_3$	H <sub>12</sub> –C	93.9000	84.000 00	C <sub>7</sub>	C <sub>6</sub> H <sub>12</sub>	84.093 90
$m_4$	H <sub>16</sub> –O	130.2854	111.994 92	C <sub>8</sub> O	C <sub>8</sub> H <sub>16</sub>	112.125 20

<sup>a</sup> The letters indicated in column 1 also refer to  $\Delta m$  given in Figures 4 and 5B in the Supporting Information: A–H represent  $\Delta m$  observed in mass spectra;  $m_1$ – $m_4$  represent fundamental mathematical relationships among molecular compositions (Figure 3A in the Supporting Information). Any of these  $\Delta m$  (column 3) is associated with a specific alteration of molecular compositions (column 2) and require a certain minimum mass to first appear within the C,H,O-compositional space (column 4); cf. text.



**Figure 6.** Selectivity of ionization of SuwFA. Occurrence of C,H,O-elemental ratios in van Krevelen diagrams derived from SuwFA positive (left) and negative (right) ions in the mass range from 200–700 Da, color-coded according to occurrence in the different ionization modes, provided with the entire feasible compositional space of C<sub>n</sub>H<sub>m</sub>O<sub>q</sub> molecules (small gray dots; ne = 18 414) within the given limits of H/C and O/C ratios.

molecular formulas were detected only as positive ions, and 1971 molecular formulas were detected both as positive and negative ions. Collectively, these 4270 molecular formulas occupy 65.7% of possible C<sub>n</sub>H<sub>m</sub>O<sub>q</sub> molecular formulas in the selected C,H,O-compositional space. Similar extents of coverage have been observed with respect to the total number of diverse chemical environments for the elements carbon (60.4%), hydrogen (58.7%), and oxygen (78.1%; Figure 7 and Table 3). This noticeable discrimination implies a noticeable preference of oxygenated molecules to become ionized under these experimental conditions.

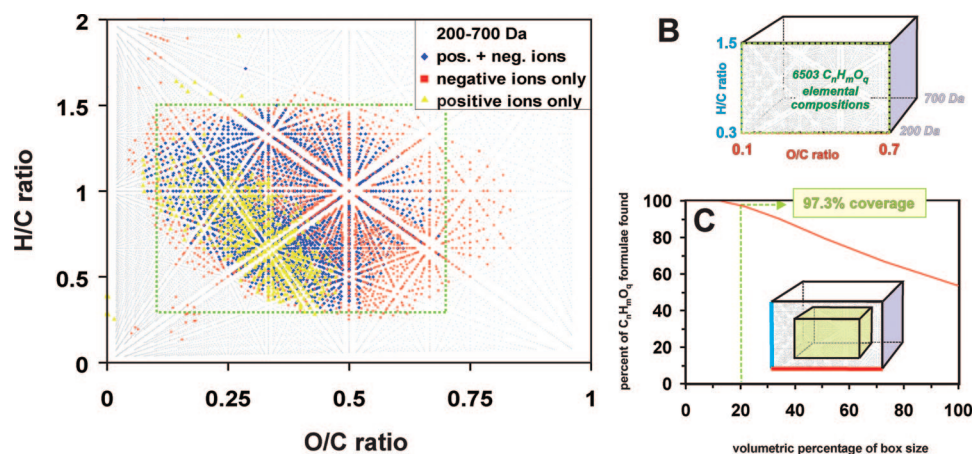
The C<sub>n</sub>H<sub>m</sub>O<sub>q</sub> compositions found in mass spectra of SuwFA occupy 100 percent of the feasible C,H,O-compositional space across a fairly wide range of mass and H/C and O/C elemental ratios (Figure 7C). This NOM compositional and structural variance supports the concept of a size–reactivity continuum of NOM, in which the variability of compositions ranges near the theoretically feasible capacity of molecular arrangements.

When the consolidated C<sub>n</sub>H<sub>m</sub>O<sub>q</sub> molecular compositions of SuwFA derived from positive and negative ions are depicted as

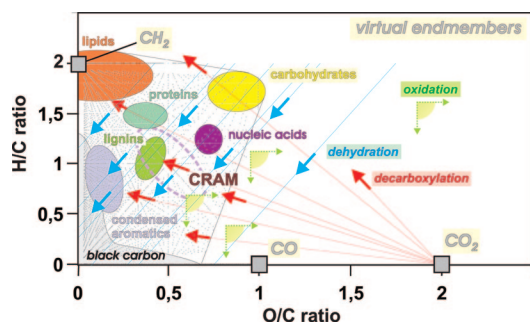
van Krevelen diagrams (Figure 7),<sup>49</sup> very few compositions are observed in the specific sectors of the diagram that are anticipated for precursor biomolecules (Figure 8).<sup>27,40,50</sup> This significant finding implies a very far-reaching processing of biomolecules in SuwFA to the extent of nonrecognition, in line with previous findings.<sup>1,51–56</sup>

- (49) Koch, B. P.; Dittmar, T. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 926–932.
- (50) Hockaday, W. C.; Grannas, A. M.; Kim, S.; Hatcher, P. G. *Geochim. Cosmochim. Acta* **2007**, *71*, 3432–3445.
- (51) Hertkorn, N.; Permin, A.; Perminova, I.; Kovalevskii, D.; Yudov, M.; Petrosyan, V.; Kettrup, A. *J. Environ. Qual.* **2002**, *31*, 375–387.
- (52) Reemtsma, T.; These, A.; Springer, A.; Linscheid, M. *Environ. Sci. Technol.* **2006**, *40*, 5839–5845.
- (53) Ritchie, J. D.; Perdue, E. M. *Geochim. Cosmochim. Acta* **2003**, *67*, 85–96.
- (54) Cabaniss, S. E.; Madey, G.; Leff, L.; Maurice, P. A.; Wetzel, R. *Biogeochemistry* **2005**, *76*, 319–347.
- (55) Cabaniss, S. E.; Madey, G.; Leff, L.; Maurice, P. A.; Wetzel, R. *Biogeochemistry* **2007**, *86*, 269–286.
- (56) Dittmar, T.; Kattner, G. *Mar. Chem.* **2003**, *82*, 115–123.
- (57) Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. *Geochim. Cosmochim. Acta* **2006**, *70*, 2990–3010.





**Figure 7.** Coverage of the C,H,O-compositional space by SuwFA ions: (A) overlay of the consolidated positive (blue), negative (red), and common ions (yellow) observed in both the positive and negative APCI, APPI, and ESI ionization modes of SuwFA. The broken green line denotes the projection of the panel B C,H,O-compositional space onto a van Krevelen diagram. (B) Volume of the C,H,O-compositional space selected to assess the extent of coverage by SuwFA molecules. Within the limits specified (200–700 Da; H/C, 0.3–1.5; and O/C, 0.1–0.7), full coverage is equivalent to 6503 distinct  $C_nH_mO_q$  molecular compositions. (C) Relative coverage of the panel (B) C,H,O-compositional space by  $C_nH_mO_q$  molecules of SuwFA in dependence of arbitrarily selected volumes of the compositional C,H,O-space, centered at 450 Da; H/C, 0.9; and O/C, 0.4. At a volume of 20% (as identified by the green box), the coverage of SuwFA molecules still exceeds 97% of the maximum feasible options for C,H,O-compositions.



**Figure 8.** Integrated view of the SuwFA compositional space. Van Krevelen plots of consolidated SuwFA molecular compositions indicate loss of signature from the distinct classes of biomolecular precursors [lipids (orange), peptides (blue), carbohydrates (yellow), nucleic acids (purple), lignins (green), polycyclic aromatics (light violet)], oxidation products (carboxyl-rich alicyclic molecules, CRAM)<sup>57</sup> and hydrogen-deficient geogenic materials (e.g., black carbon).<sup>45,50,58,59</sup> The near complete coverage of the quantized C,H,O-compositional space of SuwFA molecules near average H/C and O/C values implies a very far-reaching processing of biomolecules in SuwFA, affected by covalent binding of precursors in conjunction with a succession of small scale oxidative and hydrolytic degradation reactions.

Any covalent bond formation between different classes of SuwFA precursors<sup>60,61</sup> will initially produce novel compounds that are (within van Krevelen diagrams) positioned close to the graphs connecting the precursor molecules. Hence, covalent binding between different classes of biochemical precursor molecules will preferentially generate molecular compositions with near average H/C and O/C ratios. This also represents the area within the C,H,O-compositional space for which the largest number of chemically relevant isomers is expected.<sup>1</sup> Humification is governed

by the general constraints of thermodynamics and kinetics, and a growing diversity of molecular compositions can be judged as a credible contribution to an increase of entropy  $S$ , which is defined as  $S = k \ln N$  [ $k$ , Boltzmann constant  $1.38 \times 10^{-23} \text{ J K}^{-1}$ ;  $N$ , number of microstates (partition function; here, count of diverse molecular compositions) consistent with the given macrostates].

Carboxyl-rich aliphatic acids (CRAM)<sup>57,62</sup> which occupy a central position within the van Krevelen plot, and black carbon (BC), which covers a rather extended area of unsaturation and oxidation within the C,H,O-compositional space,<sup>45,58,59</sup> represent SuwFA constituents of biogeochemical origin. Covalent binding of CRAM and BC with other SuwFA constituent classes of ultimate biogenic origin will most effectively fill the gaps in-between the initially fragmented biochemical compositional space of SuwFA precursors. A succession of small-scale chemical reactions, like dehydration, dehydrogenation, decarbonylation, and decarboxylation will continuously create novel compounds, which further close remaining gaps in the SuwFA compositional space. Oxidation will increase O/C (or decrease H/C) ratios, while loss of oxygenated molecules like  $H_2O$  (blue arrows) and  $CO_2$  (red arrows) because of, e.g., mineralization will decrease O/C and decrease ( $-H_2O$ ) and increase ( $-CO_2$ ) H/C ratios (Figure 8).

Up to this point, the assessment of molecular complexity of SuwFA has relied solely upon assigned C,H,O-molecular formulas, which collectively account for 22% of observed peaks and approximately 61% of total observed ion intensity, and has ignored the existence of structural isomers. A more realistic assessment of molecular complexity should include an estimate of the additional number of molecular formulas in the unidentified peaks and should consider the possibility of structural isomers. Taking a stepwise approach, it might be assumed, for example, that the remaining 78% of observed peaks is attributable to a proportionate number of around 15 100 unidentified C,H,O-molecular formulas, thus bringing the total number of observed molecules to around

(58) Dickens, A. F.; Gelinas, Y.; Masiello, C. A.; Wakeham, S.; Hedges, J. I. *Nature* **2004**, *427*, 336–339.

(59) Knicker, H.; Totsche, K. U.; Almendros, G.; Gonzalez-Vila, F. J. *Org. Geochem.* **2005**, *36*, 1359–1377.

(60) Hsu, P. H.; Hatcher, P. G. *Geochim. Cosmochim. Acta* **2005**, *69*, 4521–4533.

(61) Hsu, P. H.; Hatcher, P. G. *Org. Geochem.* **2006**, *37*, 1694–1704.

(62) Lam, B.; Baer, A.; Alaei, M.; Lefebvre, B.; Moser, A.; Williams, A.; Simpson, A. J. *Environ. Sci. Technol.* **2007**, *41*, 8240–8247.



**Table 3. Minimum Numbers of Total Ions, C,H,O-Compositions and Minimum Counts of Different Chemical Environments of C, H, and O in SuwFA<sup>a</sup>**

	consolidated ions of SuwFa from six ionization methods	identified C,H, O-compositions	percent coverage of C,H,O-compositional space within limits specified	total count of carbon from identified C,H,O-compositions	total count of hydrogen from identified C,H,O-compositions	total count of oxygen from identified C,H,O-compositions
negative ions	7 853	1715	26.4	45 557	39 748	21 064
positive ions	3 526	584	9.0	21 776	16 380	6 725
minimum common ions	8 294	1971	30.3	51 180	44 554	18 369
minimum total ions	19 403	4270	65.7	118 513	100 682	46 158
C,H,O-compositional space		6503	100.0	196 146	171 494	59 102
percent coverage				60.4	58.7	78.1

<sup>a</sup> Obtained from consolidated sets of positive and negative SuwFA ions in relation to the respective counts of the C,H,O compositional space within the specified limits, 200–700 Da; H/C, 0.3–1.5; and O/C, 0.1–0.7 (Figure 7).

19 400. A proportionate scaling of the number of unique chemical environments of C, H, and O yields total estimates of  $4.22 \times 10^5$ ,  $3.58 \times 10^5$ , and  $1.63 \times 10^5$  for C, H, and O, respectively.

The level of complexity of SuwFA can reasonably be scaled further by considering the presence of heteroatoms, which have thus far been neglected entirely. Using the bulk elemental analysis of SuwFA, the corresponding atomic ratios are H/C = 0.993, O/C = 0.616, N/C = 0.0110, S/C = 0.0033, and P/C = 0.000030. In the absence of heteroatoms, the resolving power of high-field FTICR mass spectrometers is sufficient to completely resolve  $C_nH_mO_q$  ions having masses of 1000 Da or greater (Table 2; Figure 1; Figures 4 and 5 in the Supporting Information). If this were the situation for SuwFA, then its level of chemical complexity would be well represented by the peak counts in Table 3. Heteroatoms are not quantitatively abundant in SuwFA; however, it is still quite reasonable to expect a substantial fraction of molecules to contain N and, to a lesser extent, S. As a simple example, let the bulk average chemical composition of SuwFA be approximated as  $C_{25}H_{24}O_{15}$ . The bulk N/C ratio of SuwFA can be accounted for, if only one average molecule in four contains a single nitrogen atom. This line of reasoning indicates that up to 25% of all peaks could contain heteroatoms. It does not follow, however, that all such peaks would be fully resolved, because overlap of very closely spaced peaks in mass spectra, which may result in envelopes of unresolved peaks, is mathematically much more likely to occur in the case of molecules containing heteroatoms ( $C_nH_mO_qZ_\Sigma$ , where  $Z$  is any combination of heteroatoms). As an example, the compositional subunits  $N_4O$  and  $S_2H_8$  differ in mass by only 0.45 mDa.<sup>63,64</sup> If it is assumed conservatively that up to 10% of  $C_nH_mO_q$  peaks might have been compromised by overlap with peaks from  $C_nH_mO_qZ_\Sigma$  molecules, and thus be included in the count of resolved peaks, then the number of unique chemical environments of C, H, and O increase further to  $4.64 \times 10^5$ ,  $3.94 \times 10^5$ , and  $1.79 \times 10^5$  for C, H, and O, respectively.

An alternative arrangement of the C,H,O-compositional space, introduced by Reemtsma et al. to depict perceived order in natural organic matter, is shown in Figure 9.<sup>38,52,65</sup> The expansion of the compositional space in this display is governed by the number of

carbon atoms in molecules  $C_nH_mO_q$  (upper limit of carbon count) and by the allowed range of O/C ratios (lower limit of carbon count; Figure 9A). Sorting of this C,H,O-compositional space according to molecular formulas  $C_nH_mO_q$  with constant numbers  $Z$  ( $Z = n + q$ ) of (C + O) atoms produces trapezoidal areas with limits defined by the fundamentals of chemical binding (parts A and B of Figure 9). In any of these trapezoids, which are defined by a constant count of (C + O; e.g.,  $Z = n + q = 60$ ; Figure 9B), a fullerene-type endmember  $C_n$  is located at its upper left corner. Progressive hydrogenation eventually yields a maximum count of hydrogen at full saturation  $C_nH_{2n+2}$  [horizontal line at constant count of carbon (Figure 9B)]. The exchange of carbon by oxygen decreases the carbon count by 1, eventually down to the specified maximum ratio of O/C (here  $(O/C)_{\max} = 1$ ; Figure 9B). This alteration initially produces  $C_{n-1}$ -series of molecules (ranging from  $C_{n-1}O$  to  $C_{n-1}H_{2n}O$ ) and displaces the minimum mass of the molecule by approximately 4 Da [ $\Delta m(+O-C) = 3.994\,915$  Da] and the maximum mass by approximately 2 Da [ $\Delta m(+O-CH_2) = 1.979\,265\,2$  Da].

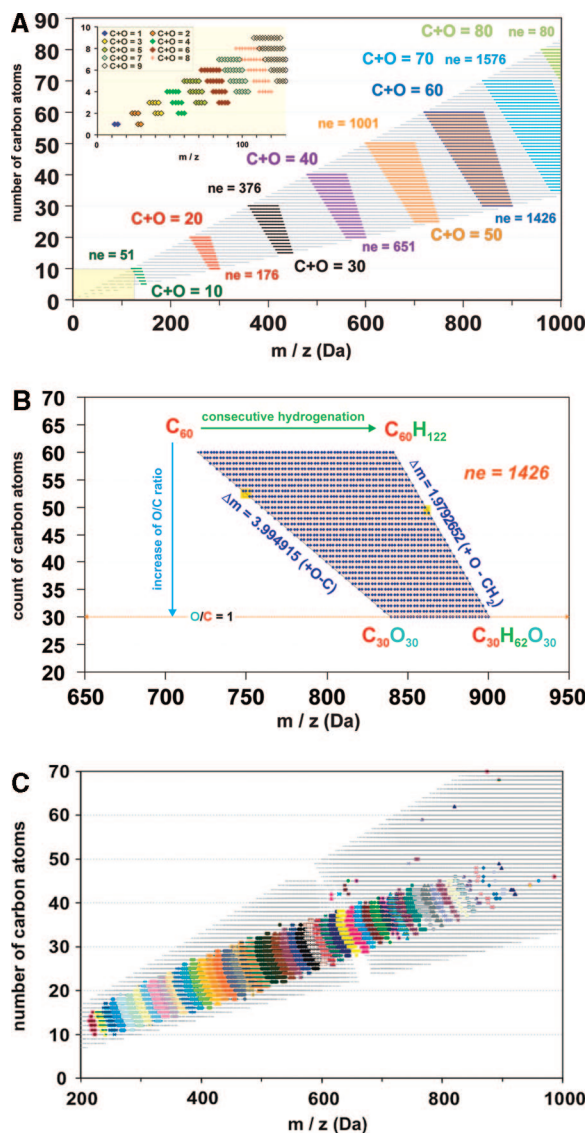
While in the total accessible C,H,O-compositional space, the areas representing  $(C + O) = Z$  are separated for  $Z = 1-7$  only (forming isolated “island” patterns; Figure 9A insert), FTICR mass spectra of NOM, which depict only the most intense signals, produce islandlike patterns also at higher  $Z$  (e.g.,  $Z$ , 24–28).<sup>38,52,65</sup> In the data set of SuwFA ions presented in Figure 9C, in which all valid C,H,O-compositions of SuwFA are consolidated irrespective of charge (see also Figure 7A), the so-called (C + O)-island patterns merge for  $Z > 10$  because of the higher S/N ratio of the FTICR mass spectra of SuwFA used here. This continuous pattern observed here applies even with the rather conservative exclusion criteria used for this analysis of FTICR mass spectra. The positioning of the identified  $C_nH_mO_q$ -ions of SuwFA in this (as well as in the van Krevelen) diagram is most likely determined by feasible counts of chemically relevant isomers, which reach maximum numbers at intermediate H/C and O/C ratios.<sup>1</sup> Here, average unsaturation and oxygenation in  $C_nH_mO_q$  molecules refer to near center positions within the trapezoids, which are defined by  $Z$  ( $Z = n + q$ ).

The estimates of complexity described so far have ignored the existence of structural (constitutional and stereo-) isomers, a highly unlikely proposition, considering the size and composition of these molecules, which allow for arranging oxygen and unsaturation (DBE) in the form of double bonds and (ali)cyclic

(63) He, F.; Emmett, M. R.; Hakansson, K.; Hendrickson, C. L.; Marshall, A. G. *J. Proteome Res.* **2004**, *3*, 61–67.

(64) Kim, S.; Rodgers, R. P.; Marshall, A. G. *Int. J. Mass Spectrom.* **2006**, *251*, 260–265.

(65) Reemtsma, T.; These, A. V.; Linscheid, M.; Leenheer, J.; Spitz, A. *Environ. Sci. Technol.* **2008**, *42*, 1430–1437.



**Figure 9.** (A) C,H,O-compositional space arranged according to count of carbon atoms in  $C_nH_mO_q$  formulas versus mass (Dalton). The trapezoidal patterns highlighted are defined by constant numbers  $Z$  ( $Z = n + q$ ), equivalent to the count of (C + O) atoms, for  $Z = 10$ –80 in units of 10; ne denotes counts of molecular formulas for the  $Z$ -values provided within the shown limits of mass ( $200 < m < 1000$  Da). (B) Trapezoid for  $Z = 60$  (1426 different  $C_nH_mO_q$  molecular formulas) with endmembers  $C_{60}$ ,  $C_{60}H_{122}$ ,  $C_{30}O_{30}$ ,  $C_{30}H_{62}O_{30}$  given; horizontal displacement indicates (de)hydrogenation and vertical displacement variance in O/C ratio; the lower count of carbon is defined by the proposed maximum O/C ratio (here,  $(O/C)_{\max} = 1$ ). The mass difference denotes exchange of oxygen versus carbon (left margin,  $\Delta m = 3.994\,915$  Da) and oxygen versus  $CH_2$  (right margin,  $\Delta m = 1.979\,265\,2$  Da). (C) The C,H,O-compositional space as defined in part A together with the entire set of consolidated SuwFA ions irrespective of charge, grouped, and color-coded according to  $Z$  (otherwise equivalent to ions provided in Table 3 and Figure 7).

structures in a huge number of isomers, easily ranging into the millions even in the low mass range near 200 Da.<sup>1,6,8,12,66,67</sup> The occurrence of isomers in SuwFA is indirectly confirmed by the observation of unbroken mass traces that occur across a considerable bandwidth of retention times (LC) and electrophoretic

mobilities (CE), when mass spectrometry of SuwFA is performed in conjunction with these separation techniques (data not shown). In fact, SuwFA molecular ions as derived from FTICR mass spectra appear to occupy the compositional space with the maximum feasible counts of chemically relevant isomeric structures, which is centered at intermediate H/C and O/C ratios.<sup>1</sup> In comparison with carboxyl-rich alicyclic molecules CRAM, another recently identified main constituent in marine organic matter likely present also in SuwFA,<sup>57</sup> which are mainly devoid of aromatics and olefinic unsaturation, SuwFA features sizable amounts of aromatic structures (23% as derived from  $^{13}C$  NMR spectra), which open the compositional space of chemically relevant species toward lower H/C ratios.<sup>1</sup>

For any ionization technique used (Figure 3), the loci of H/C and O/C ratios for molecular formulas  $C_nH_mO_q$  that can be assigned for SuwFA are significantly displaced toward lower H/C and O/C ratios compared with those derived from elemental analysis of the bulk material.<sup>68</sup> The inclusion of molecular ions for which molecular formulas are not yet assigned because they contain heteroatoms ( $C_nH_mO_qZ_z$ ) would move these formulas-derived loci toward higher H/C ratios (because incorporation of N necessarily increases H) and toward higher O/C ratios because a significant fraction of S and P in SuwFA occur as sulfates and phosphates. The low abundance of heteroatoms in SuwFA limits the magnitude of these effects, so the underestimation of H/C and O/C from assigned molecular formulas is most probably attributable to lower ionization efficiencies of oxygen-rich and more saturated molecules such as carbohydrate-like moieties, which, according to  $^{13}C$  NMR data, contribute about one-fifth of carbon present in SuwFA. In fact, consolidated FTICR mass spectra of SuwFA (free-flow electrophoretic) fractions cover a substantially larger area of van Krevelen diagrams than FTICR mass spectra of nonfractionated SuwFA (data not shown). Mass spectra of complex mixtures are decisively affected by competition for ionization. This is in stark contrast to the mass spectroscopic characterization of compounds that have been separated beforehand, as commonly realized in the hyphenation of separation and mass spectrometry (e.g., GC/MS, LC/MS, CE/MS).

All these findings clearly suggest that the minimum reasonable count of distinctly different chemical environments of carbon in SuwFA numbers in the millions. The inclusion of isomers would immediately enlarge these numbers by several orders of magnitude, resulting in billions of distinct chemical environments for carbon in SuwFA. SuwFA is an operationally defined subfraction of SuwNOM, and the preparation of SuwFA from SuwNOM is analogous to a reduction of its compositional and structural variance.<sup>69</sup> So, the huge variety of binding sites established here for SuwFA will be even larger in the case of SuwNOM.

## CONCLUSIONS

The ecological significance of this quintessential molecular heterogeneity of NOM<sup>31</sup> in the range of the theoretical limits defined by the laws of chemical binding is to procure a crucial

(66) Davies, R. E.; Freyd, P. J. *J. Chem. Educ.* **1989**, *66*, 278–281.

(67) Wieland, T.; Kerber, A.; Laue, R. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 413–419.

(68) Witt, M.; Schmitt-Kopplin, P.; Frommberger, M.; Hertkorn, N. *Humic Substances-Linking Structure to Functions*; Schriftenreihe des Lehrstuhls für Wasserwirtschaft und der DVGW-Forschungsstelle am Engler-Bunte-Institut der Universität Karlsruhe (TH); 2006; 297–300.

(69) Perdue, E. M.; Ritchie, J. D. *Dissolved Organic Matter in Fresh Water*; Elsevier: Amsterdam, The Netherlands, 2003.

life-sustaining force with a well balanced persistence and reactivity to enable and sustain microbial life in aquatic, marine, and terrestrial ecosystems and to maintain plant growth and soil quality, indispensable prerequisites for at least any higher terrestrial life. SuwFA as an example represents on the one hand remnants of life but on the other hand (together with its degradation products) it typifies direct molecular sources for microbial life as well as indirect life support because of the capacity to mitigate reactive chemical entities which could harm life functions because of potent reactivity.

Novel and useful criteria of signal assignment and assessment of complexity emerge, if the substance-specific informative order of mass spectra is projected upon the intrinsic order of the quantized compositional space of molecules.<sup>1,30</sup> If the compositional variability of NOM materials (such as SuwFA) approaches the theoretically feasible limits defined by the compositional space, then mass spectrometry will operate close to its own event horizon, and the peak patterns of high-quality mass spectra will become looking alike irrespective of materials investigated. Hence, accurate quantification of mass spectral peaks and faithful recognition of heteroatom-containing molecules will become crucial in the analysis of NOM (and of any other exceedingly complex material as well).

FTICR mass spectrometry alone will not be capable to resolve the chemical complexity of SuwFA and alike materials to its full extent because of extensive signal overlap. The meaningful exploration of the huge structural space of NOM will necessitate

integrated studies based upon a mathematical analysis of fractionation and high-resolution organic structural spectroscopy.<sup>1,48,57,70,71</sup> These means of integrated data acquisition and analysis will provide the necessary extensive synergy required to further advance the structural analysis of SuwFA and other intricate natural materials which are poorly amenable to target analysis.<sup>1,72–74</sup>

## ACKNOWLEDGMENT

The authors gratefully acknowledge a GSF research fellowship for E.M.P. (FE-75184) and a DAAD exchange fellowship (DAAD/NSF PPP exchange program USA–Germany, 2002–2003) and valuable discussions with E. V. Kunenkov (Moscow State University). We thank E. Holzmann and S. Thaller (HMGU) for skillful technical assistance during measurements and preparation of the manuscript.

## NOTE ADDED AFTER ASAP PUBLICATION

This article was released ASAP on October 28, 2008. The paper was reposted on November 1, 2008 with revised figures added.

## SUPPORTING INFORMATION AVAILABLE

Additional information as noted in the text and high-resolution Figures 1–9. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review March 5, 2008. Accepted September 12, 2008.

AC800464G

- (70) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. *Anal. Chem.* **2006**, *78*, 363–371.
- (71) Smith, L. M.; Maher, A. D.; Cloarec, O.; Rantalainen, M.; Tang, H.; Elliott, P.; Stamler, J.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Anal. Chem.* **2007**, *79*, 5682–5689.

- (72) Benner, R. *Chemical Composition and Reactivity in Biogeochemistry of Marine Dissolved Organic Matter*; Academic Press: Amsterdam, The Netherlands, 2002.
- (73) Hedges, J. I.; Eglinton, G.; Hatcher, P. G.; Kirchman, D. L.; Arnosti, C.; Derenne, S.; Evershed, R. P.; Kogel-Knabner, I.; de Leeuw, J. W.; Littke, R.; Michaelis, W.; Rullkotter, J. *Org. Geochem.* **2000**, *31*, 945–958.
- (74) Hedges, J. I.; Oades, J. M. *Org. Geochem.* **1997**, *27*, 319–361.