Composition of dissolved organic matter in groundwater

Krista Longnecker* and Elizabeth B. Kujawinski

Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Revised version submitted to: *Geochimica et Cosmochimica Acta*

Date revision submitted: *15 February 2011*

*Corresponding author. Mailing address: WHOI MS#4, Woods Hole, MA 02543. Phone: (508) 289-2824. Fax: (508) 457-2164. E-mail: klongnecker@whoi.edu

ABSTRACT

Groundwater constitutes a globally important source of freshwater for drinking water and other agricultural and industrial purposes, and is a prominent source of freshwater flowing into the coastal ocean. Therefore, understanding the chemical components of groundwater is relevant to both coastal and inland communities. We used electrospray ionization coupled with Fourier-transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS) to examine dissolved organic compounds in groundwater prior to and after passage through a sediment-filled column containing microorganisms. The data revealed that an unexpectedly high proportion of organic compounds contained nitrogen and sulfur, possibly due to transport of surface waters from septic systems and rain events. We matched 292 chemical features, based on measured mass:charge (*m/z*) values, to compounds stored in the Kyoto Encyclopedia of Genes and Genomes (KEGG). A subset of these compounds (88) had only one structural isomer in KEGG, thus supporting tentative identification. Most identified elemental formulas were linked with metabolic pathways that produce polyketides or with secondary metabolites produced by plants. The presence of polyketides in groundwater is notable because of their anti-bacterial and anti-cancer properties. However, their relative abundance must be quantified with appropriate analyses to assess any implications for public health.

1. INTRODUCTION

Groundwater is a globally important source of freshwater and as such, its composition is critically important for several reasons. First, groundwater is used as drinking water and for a variety of agricultural and industrial uses (e.g., MARGAT, 1994; ZEKTSER and EVERETT, 2004). Groundwater can also be an abundant source of freshwater entering the marine environment (MULLIGAN and CHARETTE, 2006; MOORE, 2010), and it can carry high levels of inorganic nutrients and other elements (SLOMP and VAN CAPPELLEN, 2004; SANTOS et al., 2009). These inorganic nutrients can cause regional increases in primary production, decreases in the size of seagrass beds, and ultimately play a role in the extent of coastal hypoxia (VALIELA et al., 1990; BOWEN et al., 2007; MOORE, 2010).

While many investigations have studied the transport and reactivity of inorganic ions within groundwater, considerably less is known about the sources and concentrations of organic matter in groundwater. Groundwater has variable concentrations of dissolved organic carbon (SAÑUDO-WILHELMY et al., 2002; GOÑI and GARDNER, 2003), which fluctuate with changes in flow rate (MONTLUÇON and SAÑUDO-WILHELMY, 2001), with seawater mixing in the subsurface (BECK et al., 2007), and with microbial activity (PABICH et al., 2001). The $\delta^{13}C$ and $\Delta^{14}C$ values of bulk organic carbon in groundwater are similar to the values for surface soils, and this suggests that organic matter percolates through the soils to the subsurface (MURPHY et al., 1989; WASSENAAR et al., 1990). However, the proportion of organic matter from surface soils varies regionally (ROUTH et al., 2001; LAPWORTH et al., 2008). Organic matter can also be released from the sediments through subsurface microbial activity (MURPHY et al., 1989; ARAVENA and WASSENAAR, 1993; BUCKAU et al., 2000).

There is limited information available on the composition of dissolved organic matter in groundwater. Groundwater is known to contain both humic and fulvic acids which appear to be released from sedimentary organic carbon found in the soil/subsurface matrix (WASSENAAR et al., 1990; ARTINGER et al., 2000). Furthermore, free and combined biological monomers such as neutral sugars and amino acids have been identified in groundwater, although they represent less than 10% of the total organic carbon (ROUTH et al., 2001; CHAPELLE et al., 2009). Finally, the presence of microorganisms has been linked to changes in the fluorescence characteristics of organic matter in the subsurface (CHAPELLE et al., 2009). However, there has been no molecular level assessment of the composition of organic matter in groundwater and this hinders our ability to characterize its fate and reactivity.

The molecular level composition of dissolved organic matter can be assessed using ultrahigh resolution mass spectrometry. In particular, electrospray ionization coupled with Fourier-transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS) has proven useful in characterizing organic matter composition in a wide array of environmental samples (e.g., SLEIGHTER and HATCHER, 2008; SCHMIDT et al., 2009; BHATIA et al., 2010). This technique preferentially detects compounds that are ionic in aqueous mixtures and often provides sufficient mass accuracy and resolution to assign elemental formulas based on the mass measurement alone (KUJAWINSKI and BEHN, 2006; KOCH et al., 2007). In the present project, we provide information on the molecular level composition of organic matter in groundwater, and link this composition with the metabolic pathways responsible for the production of specific organic compounds.

We established an experimental setup designed to examine the impact of protozoan grazers on the composition of dissolved organic matter in groundwater. Protozoan grazers affect bulk

carbon cycling in the subsurface (MADSEN et al., 1991; KINNER et al., 2002; TSO and TAGHON, 2006; CUNNINGHAM et al., 2009), although their role in altering the presence or absence of specific organic compounds remains unknown. Laboratory experiments have identified organic compounds that appear or disappear when specific protozoan grazers are cultured with model bacterial organisms (KUJAWINSKI et al., 2004; GRUBER et al., 2006). However, assessing the impact of protozoan grazers on the composition of dissolved organic matter in natural ecosystems is more complex and requires establishing experimental conditions with and without protozoan grazers. Therefore, our experiment was designed to allow us to test the hypothesis that protozoan grazers have a significant impact on the composition of dissolved organic matter in groundwater.

2. METHODS

2.1. Sampling site and experimental setup

Groundwater was sampled from the freshwater zone of the aquifer at the Waquoit Bay National Estuarine Research Reserve. The groundwater was continuously pumped from 2.4 m below the surface and then through cylinders 25 cm high and 7 cm wide which were filled with autoclaved aquifer sediments to mimic in situ conditions. The flow rate of groundwater through the cylinders was 30 ml hr^{-1} which resulted in an 8 hour residence time within the cylinders. In half of the cylinders, groundwater was filtered with a 1 μm filter to remove protozoan grazers; the other half of the cylinders received unfiltered ('whole') groundwater. Groundwater was allowed to flow through the sediment-filled cylinders for one month prior to the onset of the experiment. The first day following this pre-conditioning period is designated as day zero, and all sample collection during the present project starts on day zero. From day zero to day eleven

of the project, the cylinders received uniformly-labeled ^{13}C-acetate (99% $^{13}CH_3$-$^{13}COOH$, Cambridge Isotope Laboratories, Andover MA) such that the concentration of acetate in the groundwater was 200 μM. Groundwater exiting the cylinders was collected in acid-washed low-density polyethylene cubitainers which were kept cold in the dark by a recirculating water bath. The cubitainers were returned to the lab every three to six days.

The present study considers five samples: one sample from the groundwater entering the sediment-filled cylinders on day 30 of the experiment and four samples from groundwater exiting the sediment-filled cylinders. Two of the samples are from the treatment with grazers (day 0 and day 30), and two of the samples are from the treatment where grazers were removed by filtration (also from day 0 and day 30). The samples collected on day 0 were collected before acetate was added to the sediment-filled cylinders. In addition, data from the present project was compared to Suwannee River fulvic acid (SRFA) which had been previously analyzed by our laboratory in both positive and negative ion modes (KIDO SOULE et al., 2010).

2.2. Analysis of discrete groundwater samples

Unfiltered groundwater in the cubitainers was used to obtain the abundance of protozoan grazers and for δ^{13}C measurements. The abundance of protozoan grazers was obtained using epifluorescence microscopy. Cells were first preserved with 0.05% (final concentration) alkaline Lugol's solution, followed by 0.1% (final concentration) sodium thiosulfate, and finally 2% (final concentration) of borate-buffered formalin. Samples were incubated at 4 ºC for 24 hours, stained with DAPI (25 μg ml^{-1} final concentration) for 10 min, and then filtered onto black 0.8 μm polycarbonate filters (SHERR et al., 1993). Concentrations and carbon stable-isotopic ratios of total organic carbon (TOC) and dissolved inorganic carbon (DIC) in the groundwater exiting the cylinders were obtained with an O.I.-analytical 1010 TOC/TIC analyzer in series with a Europa

20-20 mass spectrometer. The coefficient of variability between duplicate injections averaged <1%. $\delta^{13}C$ values were reported relative to PeeDee belemnite using standard notation: $\delta^{13}C$ (‰)= (R_{sample} / $R_{standard}$ - 1)*1000, where R is the ratio of the heavy to light element. The $\delta^{13}C$ values were converted to atom $\%^{13}C$ for ease of presentation.

2.3. Extraction and analysis of dissolved organic matter (DOM)

Prior to analysis of the concentration and composition of DOM, the bacterial community in the groundwater was removed by filtration of groundwater through combusted 0.2 µm Anodisk filters (Whatman). The concentration of dissolved organic carbon (DOC) was measured with a Shimadzu TOC-V_{CSH} total organic carbon analyzer. The coefficient of variability between replicate injections was <1%. Comparisons to standards provided by Prof. D. Hansell (University of Miami) were made daily.

DOM in groundwater is too dilute to directly analyze using ESI FT-ICR MS and we used a solid phase extraction method to concentrate the DOM. The 0.2-µm filtered water was acidified and extracted using stacked C_{18}/SDB resin disks, and eluted off the disks with 70% methanol as previously described (KIM et al., 2003; KUJAWINSKI et al., 2009). Different extraction methods may alter the measured chemical characteristics of DOM (KAISER et al., 2003; SCHWEDE-THOMAS et al., 2005); a single extraction method was used throughout the present project to minimize this issue. The combination of C_{18} and SDB should result in a higher extraction efficiency than just C_{18}. While we did not measure extraction efficiencies for the present project, we estimate the DOM extraction efficiency is at least 40% based on previous research with similar extraction resins (TREMBLAY et al., 2007; DITTMAR et al., 2008). We cannot speculate about the composition of the dissolved organic matter we were not able to extract using the

C_{18}/SDB resin disks. A blank, consisting of acidified Milli-Q water, was processed and analyzed along with the five samples described above.

Samples were analyzed in both positive and negative ion modes on a 9.4 T ESI FT-ICR mass spectrometer at the National ICR Users' Facility at the National High Magnetic Field Laboratory at Florida State University in Tallahassee FL as previously described (KUJAWINSKI et al., 2009). Positive ion mode will preferentially ionize compounds with amines, which are common in proteins. Negative ion mode will preferentially ionize carboxylic acids which are common in lignins, humic acids, and some lipids. One hundred scans were co-added, Hanning apodized, zero-filled once, and fast Fourier-transformed (SENKO et al., 1996a; SENKO et al., 1996b). Spectra were internally calibrated with a series of compounds present in all spectra and mass accuracy errors were approximately 0.5 ppm after internal calibration. The noise level was individually determined for each sample, and only peaks with a signal at least three times the noise level were analyzed further. Each peak is a mass:charge (*m/z*) value which is the measured mass of the observed ion divided by its charge. Spectra were aligned (MANTINI et al., 2007) to generate a master list of *m/z* values present in all spectra. Any *m/z* values found in the blank were removed from the rest of the dataset. Data were converted to neutral masses assuming a loss of one proton (H^+) in negative ion mode and addition of one sodium ion (Na^+) in positive ion mode.

Elemental formulas were assigned using the Compound Identification Algorithm (CIA: KUJAWINSKI and BEHN, 2006; KUJAWINSKI et al., 2009) using a formula error of 1 ppm, and a relationship error of 20 ppm. The mass limit above which elemental formulas were assigned only by functional group relationships was 500 Da. Elements considered in CIA are C, H, O, N, S, and P. Isotopomers with a ^{13}C atom were identified and elemental formulas were corrected to reflect ^{13}C content. ESI FT-ICR MS is not quantitative, and peak heights are affected by

differences in ionization efficiency among compounds. Therefore we only consider the presence or absence of a compound rather than relative peak heights. We could not do MS^n to confirm the structure/identity of individual *m/z* values due to the high number of peaks observed at each nominal mass.

2.4. Computational and statistical analysis

Three datasets were downloaded in February 2010 from the Kyoto Encyclopedia of Genes and Genomes (KEGG, KANEHISA et al., 2008): biochemical compounds, biochemical reactions, and a comprehensive list of metabolic pathways. This combination of KEGG datasets allows us to identify metabolic pathways which are involved in the production or alteration of specific biochemical compounds. The lists were imported into MATLAB, and the neutral masses for each of the 16,143 biochemical compounds were recalculated using exact elemental masses. Duplicate neutral masses were possible in this dataset due to the presence of structural isomers and thus we compared the list of *m/z* values from the present project to KEGG in two ways. First, we looked for any compound listed in KEGG within 1 ppm of our neutral masses. Second, we culled this list to consider only compounds found once in KEGG, i.e., those without structural isomers. Figures of the metabolic pathways in KEGG with biochemical compounds found in the present project were generated using the KEGG application programming interface via the SOAP/WSDL web service from within MATLAB. Select metabolic pathways are provided in the Electronic Annex as EA Figures 2 to 10. We recognize that our measured *m/z* values cannot be linked to KEGG compounds with absolute certainty due to the possibility of structural isomers that are not included in KEGG. This caveat should be considered when we use the word 'compound' in the discussion of our data and KEGG.

Cluster analysis was used to analyze variability in the ESI FT-ICR MS data; separate analyses were conducted for positive and negative ion modes. Distances between samples were calculated with the Bray-Curtis distance measure using the Fathom toolbox (David Jones, University of Miami) and cluster analysis was performed using Ward's linkage method (McCune and Grace, 2002).

3. RESULTS AND DISCUSSION

Our experimental design was successful in reducing the number of protozoan grazers in the 1 μm-filtered treatments. By day 30 of the experiment, the groundwater exiting the cylinders for the whole treatments (with grazers) had 5880 ± 610 protozoan grazers ml^{-1}. The groundwater exiting the 1 μm-filtered treatments (grazers removed by filtration) had 1140 ± 380 protozoan grazers ml^{-1}.

The DOC concentrations in the groundwater prior to entering the sediment-filled cylinders averaged 75.1 μM (66.3 to 83.9 μM, 95% confidence interval, n = 4). The concentration of DOC exiting the cylinders increased to an average of 110.8 μM (75.9 to 145.6 μM, 95% confidence interval, n = 27) in the cylinders receiving 1 μm-filtered groundwater and 146.8 μM (95.4 to 198.2 μM, 95% confidence interval, n = 21) in the cylinders receiving whole groundwater. The increase in dissolved organic carbon concentrations in the cylinders which did not receive added carbon was likely due to carbon leaching off the sediment within each cylinder, although we cannot discount the contribution of DOC exuded by microbial cells within the cylinders.

The atom % of ^{13}C of TOC in the groundwater increased rapidly after the addition of ^{13}C-acetate, and then declined after the ^{13}C-acetate addition was terminated on day 11 of the experiment (Fig. 1). However, ^{13}C-TOC includes both ^{13}C-acetate added as dissolved organic

carbon and any of the ^{13}C-acetate assimilated into bacterial biomass. Our dissolved inorganic carbon (DIC) data revealed increases in atom % ^{13}C in DIC above the natural abundance of ^{13}C in both the whole and the 1 μm-filtered treatments (Fig. 1). The increase in DIC was measured at the first sampling point following the addition of the ^{13}C-labeled carbon and was sustained throughout the duration of the experiment. The loss of carbon as DIC ranged from less than 1% to a maximum of 7% of all calculated losses. Therefore, the ^{13}C-DIC data indicate the microbial community within the sediment-filled cylinders was capable of utilizing the added organic material.

We analyzed the groundwater exiting the sediment-filled cylinders using ultrahigh resolution mass spectrometry (ESI FT-ICR MS) which provided insight into the molecular-level composition of the groundwater in our experiment. ESI FT-ICR MS yielded between 3200 and 8300 *m/z* values, with slightly more *m/z* values observed in negative ion mode (Fig. 2, Table 1). Although we observed organic compounds with ^{13}C replacing ^{12}C in the elemental formulas, there was no increase in the proportion of ^{13}C-compounds following the incubation of the groundwater with ^{13}C-acetate. Therefore, ^{13}C-acetate was apparently not incorporated into new molecules as groundwater was transported through the columns. There are multiple possibilities for this observation. The most likely possibilities are that the ^{13}C-labeled organic matter was washed out of the columns between day 0 and day 30 of the experiment, or the bacterial community respired all of the ^{13}C-acetate. Methodological issues could have also limited our ability to detect the ^{13}C compounds if they were present in concentrations below our detection limit, or if the ^{13}C was present in molecules whose molecular mass was less than *m/z* 300.

Our hypothesis was that the presence of protozoan grazers would have a significant impact on the composition of dissolved organic matter. We used both positive and negative ion modes to

address this question because the different modes will preferentially ionize different compounds. However, 45% of *m/z* values in positive ion mode and 58% of *m/z* values in negative ion mode were found in all five samples. The overlap between samples is also evident in the Van Krevelen diagrams used to visualize the O:C and H:C molar ratios of each elemental formula (EA Figure 1). Furthermore, in positive ion mode the difference between the whole and filtered treatments decreased during the incubation (Fig. 3). In negative ion mode, there were fewer *m/z* values in the treatment without grazers by the conclusion of the experiment (Table 1). The cluster analysis also revealed that the compounds were different from those observed at the onset of the experiment (Fig. 3). However, the level of differences between samples in negative ion mode was quite small, with Bray-Curtis differences between samples ranging from 0.15 to 0.32 in negative ion mode. Based on laboratory experiments, Gruber et al. (2006) conclude that the bacterial community has a greater impact on the composition of dissolved organic matter than do bacterial predators. While we do not have a bacteria-free treatment for comparison, our data do not support our hypothesis that protozoan grazers affect the composition of dissolved organic matter in groundwater on a 30-day timescale. Additional work will be needed to assess if grazers were able to impact the abundance of specific compounds in groundwater; however, ESI FT-ICR MS is not considered quantitative and for the remaining discussion we will consider the composition of organic matter in the full data set rather than focusing on differences between individual samples.

While compounds only containing C, H, and O were the most prevalent elemental formulas assigned here, a larger than expected proportion of the *m/z* values were assigned formulas containing CHON. The percent of CHON formulas ranged from 15 to 38% in our samples. This range is higher than that observed in marine and riverine samples collected off the

eastern United States (KUJAWINSKI et al., 2009) and in freshwater samples from inland rivers (SLEIGHTER et al., 2009). However, higher proportions of organic nitrogen compounds have been observed in pore waters within offshore marine sediments (SCHMIDT et al., 2009) and from glaciated surfaces in Greenland (BHATIA et al., 2010). The samples for the present project were collected in an area where dissolved organic nitrogen is predominantly from septic systems (COLE et al., 2006; KROEGER et al., 2006). Furthermore, the transformation of nitrogen from organic to inorganic forms was correlated to the distance groundwater travels in the subsurface (KROEGER et al., 2006). Therefore we posit that the high proportion of CHON compounds in our dataset is due to limited microbial remineralization of dissolved organic nitrogen prior to the groundwater reaching our study site. However, deciphering which biotic and abiotic processes in these disparate environments may contribute to high proportions of CHON compounds will be an exciting avenue for future research.

Sulfur-containing elemental formulas were also prevalent in the original groundwater sample as CHONS in positive ion mode and in negative ion mode as CHOS (Table 1). Sulfur-containing organic matter can be a dominant component of rainwater (ALTIERI et al., 2009) which could percolate to the subsurface. In contrast, organic sulfur compounds are only a small component of subsurface sedimentary organic matter (RYU et al., 2006). Although we cannot definitively identify the source of the organic sulfur compounds in the present project, the decrease in the proportion of these compounds in groundwater collected from the cylinder terminus was striking. This suggests that organic sulfur compounds were consumed or remineralized during transit through the cylinders. Organic sulfur compounds have been recognized as both carbon and sulfur sources for marine microorganisms (SIMÓ et al., 2002; SIEVERT et al., 2007) and sulfate-reducing bacteria have been found in subsurface microbial

communities (VAN BEEK and VAN DER KOOIJ, 1982; CHANG et al., 2001). While sulfate-reducing bacteria require an organic carbon source to reduce sulfate to sulfide, to our knowledge there has been no investigation of the consumption or alteration of organic sulfur compounds by subsurface microorganisms.

To characterize potential sources of organic matter, we compared our groundwater data with metabolic pathways collated in KEGG. For this analysis, we considered *m/z* values found in any of the five groundwater samples, and therefore do not consider observations about individual samples. As a control exercise, we compared the *m/z* values in Suwannee River Fulvic Acid (SRFA) to compounds in KEGG. We recognize that only a subset of organic compounds are listed in KEGG, but putatively linking compounds in groundwater with metabolic pathways in KEGG is one step towards linking organic compounds with biological sources. When the measured *m/z* values were converted to neutral masses, there were 292 compounds in the KEGG database within 1 ppm of an *m/z* value found within groundwater; 88 of those compounds did not have any structural isomers in KEGG (EA Table 1). While this is a small number of compounds relative to the *m/z* values observed in all the samples, this project provides an important step towards linking organic compounds detected *in situ* with biochemical pathways. The organic compounds with no structural isomers were contained within 25 different metabolic pathways (Table 2). For most of the metabolic pathways, additional metabolic intermediates were tentatively identified in our dataset but these compounds contain structural isomers in KEGG and so the identifications cannot be fully constrained. However, the possible presence of these additional intermediates provides confidence that the metabolic pathway is active in our system (SUHRE and SCHMITT-KOPPLIN, 2008). In some cases, up to half of the biochemical compounds within a metabolic pathway were putatively identified in groundwater. The compounds within

the biochemical pathways which were not identified within our samples are either not present or are not amenable to measurement using ESI FT-ICR MS. Additional work will be needed to determine the concentration of the compounds we observed in groundwater, and in the following discussion we only consider the presence or absence of compounds. Two broad groups of metabolic pathways (Table 2) were the majority of the *m/z* values matched to KEGG: (1) microbial pathways involved in the production of polyketides or other pharmaceutically interesting compounds (biosynthesis of polyketides and terpenoids), and (2) pathways related to plant-based metabolisms (biosynthesis of other secondary metabolites). The remaining pathways (carbohydrate metabolism, lipid metabolism, amino acid metabolism, metabolism of cofactors and vitamins, and xenobiotics biodegradation and metabolism) had no more than four compounds without structural isomers and represent diverse classes of metabolic pathways.

Identification of polyketides in groundwater is an important finding because these compounds affect human metabolism. Polyketides are a class of compounds used as anti-bacterial and anti-cancer drugs, and are generally classified into three groups based on their structure and/or their biosynthetic pathways (RAWLINGS, 2001; SHEN, 2003). In the present project (Table 2), we found the strongest evidence for the biochemical pathways involved in the biosynthesis of type II polyketides (EA Figure 2) and the biosynthesis of 12-, 14-, and 16-membered macrolides, which are type I polyketides with macrocyclic lactone rings (EA Figure 3). Furthermore, while not quite half of the compounds in the pathway for the biosynthesis of type II polyketides were also found in SRFA, none of the compounds in the pathway for the biosynthesis of 12-, 14-, and 16-membered macrolides, tetracycline, or ansamycins were found in SRFA. Fungi and bacteria, especially Actinomycetes, are the main producers of type II polyketides (HUTCHINSON and FUJII, 1995), and type II polyketide synthase genes have

previously been observed in soils (SEOW et al., 1997; WAWRIK et al., 2005). For some of the polyketides, we observed elemental formulas consistent with intermediate metabolites in the majority of the chemical steps needed to produce a polyketide (e.g., doxorubicin, auramycinone, urdamycin A, urdamycin B; EA Figure 2). Identification of these biosynthetic intermediates suggests that polyketides may be actively produced in the subsurface. However, we cannot eliminate the possibility that polyketides are also being transported through the subsurface from septic systems in the region.

Polyketides may be degraded by biota in the subsurface. In laboratory cultures, the fungus *Alternaria alternata* has been shown to degrade the polyketides it produces (JONSSON et al., 1987). Bacterial degradation of fungal polyketides has also been demonstrated for the polyketide cercosporin (MITCHELL et al., 2002). For example, cercosporin is degraded by microorganisms into xanosporic acid, through two intermediate steps (MITCHELL et al., 2003; TAYLOR et al., 2006). Cercosporin, xanosporic acid, and one of the two intermediates were observed in all of our groundwater samples, suggesting that this degradation pathway (or a similar one) is present and active in our system.

The second major category of compounds identified in the present study is secondary metabolites produced by plants (Table 2, EA Figures 4 to 10). In most of the biochemical pathways, the compounds identified in KEGG without structural isomers were also identified within SRFA. Quantification of fatty acids has revealed that plant-based organic matter can be a dominant component of sedimentary organic matter in aquifers (HARTOG et al., 2004). Our observation of specific compounds produced by plant-based metabolisms and the high degree of overlap with compounds in SRFA highlights the impact of surface-based processes on the subsurface organic matter. However, we cannot determine whether these plant metabolites are

available to subsurface microbial communities and thus whether they are degraded there. While the presence of plant-derived compounds in groundwater is not surprising, the observation emphasizes that both micro- and macro-fauna have the potential to impact the composition of dissolved organic matter in the subsurface.

Conclusions

The present project assessed the composition of dissolved organic matter in groundwater and identified possible sources of a subset of the organic matter. We observed a higher proportion of organic nitrogen and sulfur compounds compared to organic matter characterized from other environments, possibly due to inputs from septic systems and rain events. Furthermore, the appearance of degradation products of one polyketide, cercosporin, and ^{13}C-labeled dissolved inorganic carbon indicates the bacterial community is capable of utilizing organic compounds in groundwater. We also found evidence of microbial production of organic compounds in the subsurface because a large proportion of intermediates within polyketide biosynthetic pathways were present in groundwater. These latter compounds are of particular relevance to communities that rely on groundwater as drinking water because the presence of pharmaceutically-interesting compounds has unknown consequences. Our data do not include the concentration of each compound, therefore we cannot assess the relevance to human health. Additional research with compound-specific methods is needed in order to definitively identify these compounds, and quantify any spatial and temporal variability in their concentrations.

Acknowledgments. We thank Meagan Eagle Gonneea for help installing the groundwater well, Ann Mulligan for discussions about the project and the research site, Dave Harris at the University of California, Davis for the analysis of the carbon stable isotopes, and Alan Marshall,

369 Jerry Purcell and Tanner Schaub at the National High Magnetic Field Laboratory for access to
370 and assistance with the ultrahigh resolution mass spectrometer used in the present study. The
371 comments of three anonymous reviewers were helpful in improving the manuscript. Finally, we
372 extend our special gratitude to the Waquoit Bay National Estuarine Research Reserve for
373 allowing us to install the well and conduct this experiment in their boathouse. Funding was
374 provided by NSF grants EAR-0525166 and OCE-0751897 to EBK.
375

Figure Legends

Fig. 1. Atom %^{13}C of total organic carbon (TOC, squares) and dissolved inorganic carbon (DIC, diamonds) in groundwater exiting the sediment-filled cylinders exposed to (A) whole groundwater or (B) 1 μm-filtered groundwater. Dashed line indicates the natural abundance of ^{13}C. Data are plotted on a $\log_{10}$ scale.

Fig. 2. Negative ion mode mass spectra from whole groundwater sampled at day 30. The y-axis is relative peak height. The dotted line is the peak detection threshold for this sample. Elemental formulas are given above select peaks. The compound number(s), beginning with 'C', are given for compounds found in KEGG. Compounds with structural isomers in KEGG (marked with a triangle) and without structural isomers (marked with a star) were both identified in the groundwater samples.

Fig. 3. Cluster analysis based on Bray-Curtis distance measures calculated for the ultrahigh resolution mass spectrometry data collected in (A) positive ion mode and (B) negative ion mode. Note the different x-axis scale for the two figures.

392 Table 1. Samples were analyzed in positive and negative ion mode using ESI FT-ICR MS.
393 'Whole' and 'Filtered' in the sample column refer to whole groundwater and 1 μm-filtered
394 groundwater exiting the sediment-filled cylinders. The table summarizes the number of *m/z*
395 values found in each sample, the percent of *m/z* values assigned elemental formulas, and the
396 percent of formulas containing only CHO, CHON, CHOS, or CHONS in the elemental formula.
397

Sample	Total # *m/z* values	% *m/z* with formulas	%CHO	%CHON	%CHOS	%CHONS
Positive ion mode						
Whole, day 0	3941	97	41	38	2	3
Filtered, day 0	3790	99	55	28	0	5
Whole, day 30	4759	97	63	21	0	3
Filtered, day 30	3260	96	66	23	0	1
Groundwater, day 30	3368	98	35	15	2	20
Negative ion mode						
Whole, day 0	8329	94	48	31	3	2
Filtered, day 0	6035	91	59	22	4	0
Whole, day 30	4282	98	44	24	5	5
Filtered, day 30	3376	99	64	26	0	0
Groundwater, day 30	5915	94	45	24	10	7

Table 2. Number of *m/z* values matching compounds in the KEGG database, including compounds with and without structural isomers. Matches to compounds in KEGG are shown for the data from the present project ('Waquoit Bay') and Suwannee River Fulvic Acid ('SRFA'). The total number of compounds is the entire set of compounds contained in the metabolic pathway at KEGG.

		# compounds with no isomers		# compounds, including isomers	
	Total # compounds	Waquoit Bay	SRFA	Waquoit Bay	SRFA
Biosynthesis of polyketides and terpenoids					
Biosynthesis of 12-, 14- and 16-membered macrolides	85	16	0	27	0
Biosynthesis of type II polyketide products	103	29	12	54	7
Biosynthesis of ansamycins	30	1	0	1	0
Tetracycline biosynthesis	28	1	0	2	0
Diterpenoid biosynthesis	83	3	3	29	21
Carotenoid biosynthesis	109	1	1	2	2
Biosynthesis of other secondary metabolites					
Phenylpropanoid biosynthesis	72	2	2	9	7
Stilbenoid, diarylheptanoid and gingerol biosynthesis	38	2	1	5	2
Flavonoid biosynthesis	86	1	1	14	11
Flavone and flavonol biosynthesis	50	4	4	13	9
Isoflavonoid biosynthesis	78	3	3	17	13
Novobiocin biosynthesis	55	2	0	4	0
Carbohydrate metabolism					
Ascorbate and aldarate metabolism	67	1	0	1	0
Glyoxylate and dicarboxylate metabolism	72	1	0	1	0
Lipid metabolism					
Steroid hormone biosynthesis	115	3	2	20	6
Arachidonic acid metabolism	75	1	0	21	0
Amino acid metabolism					
Cysteine and methionine metabolism	100	3	0	3	0
Glutathione metabolism	71	1	0	1	0
Metabolism of cofactors and vitamins					
Retinol metabolism	38	1	1	1	1
Porphyrin and chlorophyll metabolism	135	4	3	10	5
One carbon pool by folate	46	1	1	1	1
Xenobiotics biodegradation and metabolism					
1,4-Dichlorobenzene degradation	72	1	0	1	0
Drug metabolism - cytochrome P450	103	1	0	3	0
gamma-Hexachlorocyclohexane degradation	55	1	0	1	0
Metabolism of xenobiotics by cytochrome P450	88	2	0	4	0

References

Altieri K. E., Turpin B. J., and Seitzinger S. P., (2009) Oligomers, organosulfates, and nitrooxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry. *Atmos. Chem. Phys.* **9**, 2533-2542.

Aravena R. and Wassenaar L. I., (1993) Dissolved organic carbon and methane in a regional confined aquifer, southern Ontario, Canada: Carbon isotope evidence for associated subsurface sources. *Appl. Geochem.* **8**, 483-493.

Artinger R., Buckau G., Geyer S., Fritz P., Wolf M., and Kim J. I., (2000) Characterization of groundwater humic substances: influence of sedimentary organic carbon. *Appl. Geochem.* **15**, 97-116.

Beck A. J., Tsukamoto Y., Tovar-Sanchez A., Huerta-Diaz M., Bokuniewicz H. J., and Sañudo-Wilhelmy S. A., (2007) Importance of geochemical transformations in determining submarine groundwater discharge-derived trace metal and nutrient fluxes. *Appl. Geochem.* **22**, 477-490.

Bhatia M. P., Das S. B., Longnecker K., Charette M. A., and Kujawinski E. B., (2010) Molecular characterization of dissolved organic matter associated with the Greenland ice sheet *Geochim. Cosmochim. Acta* **74**, 3768-3784.

Bowen J. L., Kroeger K. D., Tomasky G., Pabich W. J., Cole M. L., Carmichael R. H., and Valiela I., (2007) A review of land-sea coupling by groundwater discharge of nitrogen to New England estuaries: Mechanisms and effects. *Appl. Geochem.* **22**, 175-191.

Buckau G., Artinger R., Geyer S., Wolf M., Fritz P., and Kim J. I., (2000) Groundwater in-situ generation of aquatic humic and fulvic acids and the mineralization of sedimentary organic carbon. *Appl. Geochem.* **15**, 819-832.

Chang Y.-J., Peacock A. D., Long P. E., Stephen J. R., McKinley J. P., Macnaughton S. J., Hussain A. K. M. A., Saxton A. M., and White D. C., (2001) Diversity and characterization of sulfate-reducing bacteria in groundwater at a uranium mill tailings site. *Appl. Environ. Microbiol.* **67**, 3149-3160.

Chapelle F. H., Bradley P. M., Goode D. J., Tiedeman C., Lacombe P. J., Kaiser K., and Benner R., (2009) Biochemical indicators for the bioavailability of organic carbon in ground water. *Ground Water* **47**, 108-121.

Cole M. L., Kroeger K. D., McClelland J. W., and Valiela I., (2006) Effects of watershed land use on nitrogen concentrations and $\delta^{15}N$ in groundwater. *Biogeochemistry* **77**, 199-215.

Cunningham J. J., III, Kinner N. E., and Lewis M., (2009) Protistan predation affects trichloroethene biodegradation in a bedrock aquifer. *Appl. Environ. Microbiol.* **75**, 7588-7593.

Dittmar T., Koch B., Hertkorn N., and Kattner G., (2008) A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Meth.* **6**, 230-235.

Goñi M. A. and Gardner I. R., (2003) Seasonal dynamics in dissolved organic carbon concentrations in a coastal water-table aquifer at the forest-marsh interface. *Aquat. Geochem.* **9**, 209-232.

Gruber D. F., Simjouw J. P., Seitzinger S. P., and Taghon G. L., (2006) Dynamics and characterization of refractory dissolved organic matter produced by a pure bacterial culture in an experimental predator-prey system. *Appl. Environ. Microbiol.* **72**, 4184-4191.

Hartog N., van Bergen P. F., de Leeuw J. W., and Griffioen J., (2004) Reactivity of organic matter in aquifer sediments: geological and geochemical controls. *Geochim. Cosmochim. Acta* **68**, 1281-1292.

Hutchinson C. R. and Fujii I., (1995) Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* **49**, 201-238.

Jonsson L. M. V., Häggblom P., and Söderhäll K., (1987) Turnover of polyketides in the fungus *Alternaria alternata* with a possible involvement of peroxidase. *Exp. Mycol.* **11**, 223-230.

Kaiser E., Simpson A. J., Dria K. J., Sulzberger B., and Hatcher P. G., (2003) Solid-state and multidimensional solution-state NMR of solid phase extracted and ultrafiltered riverine dissolved organic matter. *Environ. Sci. Technol.* **37**, 2929-2935.

Kanehisa M., Araki M., Goto S., Hattori M., Hirakawa M., Itoh M., Katayama T., Kawashima S., Okuda S., Tokimatsu T., and Yamanishi Y., (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480-484.

Kido Soule M. C., Longnecker K., Giovannoni S. J., and Kujawinski E. B., (2010) Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter. *Org. Geochem.* **41**, 725-733.

Kim S., Simpson A. J., Kujawinski E. B., Freitas M. A., and Hatcher P. G., (2003) High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C_{18} solid phase disk. *Org. Geochem.* **34**, 1325-1335.

Kinner N. E., Harvey R. W., Shay D. M., Metge D. W., and Warren A., (2002) Field evidence for a protistan role in an organically-contaminated aquifer. *Environ. Sci. Technol.* **36**, 4312-4318.

Koch B. P., Dittmar T., Witt M., and Kattner G., (2007) Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **79**, 1758-1763.

Kroeger K. D., Cole M. L., and Valiela I., (2006) Groundwater-transported dissolved organic nitrogen exports from coastal watersheds. *Limnol. Oceanogr.* **51**, 2248-2261.

Kujawinski E. B. and Behn M. D., (2006) Automated analysis of electrospray ionization Fourier-transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **78**, 4363-4373.

Kujawinski E. B., Del Vecchio R., Blough N. V., Klein G. C., and Marshall A. G., (2004) Probing molecular-level transformations of dissolved organic matter: insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar. Chem.* **92**, 23-37.

Kujawinski E. B., Longnecker K., Blough N. V., Del Vecchio R., Finlay L., Kitner J. B., and Giovannoni S. J., (2009) Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim. Cosmochim. Acta* **73**, 4384-4399.

Lapworth D. J., Gooddy D. C., Butcher A. S., and Morris B. L., (2008) Tracing groundwater flow and sources of organic carbon in sandstone aquifers using fluorescence properties of dissolved organic matter (DOM). *Appl. Geochem.* **23**, 3384-3390.

Madsen E. L., Sinclair J. L., and Ghiorse W. C., (1991) In situ biodegradation: microbiological patterns in a contaminated aquifer. *Science* **252**, 830-833.

Mantini D., Petrucci F., Pieragostino D., Del Boccio P., Di Nicola M., Di Ilio C., Federici G., Sacchetta P., Comani S., and Urbani A., (2007) LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* **8**, 101.

Margat J., 1994. Groundwater operations and management. In: Gibert, J., Danielopol, D. L., and Stanford, J. A. Eds.), *Groundwater Ecology*. Academic Press.

McCune B. M. and Grace J. B., (2002) *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, Oregon.

Mitchell T. K., Alejos-Gonzalez F., Gracz H. S., Danehower D. A., Daub M. E., and Chilton W. S., (2003) Xanosporic acid, an intermediate in bacterial degradation of the fungal phototoxin cercosporin. *Phytochemistry* **62**, 723-732.

Mitchell T. K., Chilton W. S., and Daub M. E., (2002) Biodegradation of the polyketide toxin cercosporin. *Appl. Environ. Microbiol.* **68**, 4173-4181.

Montluçon D. and Sañudo-Wilhelmy S. A., (2001) Influence of net groundwater discharge on the chemical composition of a coastal environment: Flanders Bay, Long Island, New York. *Environ. Sci. Technol.* **35**, 480-486.

Moore W. S., (2010) The effect of submarine groundwater discharge on the ocean. *Annual Review of Marine Science* **2**, 59-88.

Mulligan A. E. and Charette M. A., (2006) Intercomparison of submarine groundwater discharge estimates from a sandy unconfined aquifer. *J. Hydrol.* **327**, 411-425.

Murphy E. M., Davis S. N., Long A., Donahue D., and Jull A. J. T., (1989) Characterization and isotopic composition of organic and inorganic carbon in the Milk River aquifer. *Water. Resour. Res.* **25**, 1893-1905.

Pabich W., Valiela I., and Hemond H., (2001) Relationship between DOC concentration and vadose zone thickness and depth below water table in groundwater of Cape Cod, U.S.A. *Biogeochemistry* **55**, 247-268.

Rawlings B. J., (2001) Type I polyketide biosynthesis in bacteria (Part A--erythromycin biosynthesis). *Nat. Prod. Rep.* **18**, 190-227.

Routh J., Grossman E. L., Murphy E. M., and Benner R., (2001) Characterization and origin of dissolved organic carbon in Yegua ground water in Brazos County, Texas. *Ground Water* **39**, 760-767.

Ryu J.-h., Zierenberg R. A., Dahlgren R. A., and Gao S., (2006) Sulfur biogeochemistry and isotopic fractionation in shallow groundwater and sediments of Owens Dry Lake, California. *Chem. Geol.* **229**, 257-272.

Santos I. R., Burnett W. C., Dittmar T., Suryaputra I. G. N. A., and Chanton J., (2009) Tidal pumping drives nutrient and dissolved organic matter dynamics in a Gulf of Mexico subterranean estuary. *Geochim. Cosmochim. Acta* **73**, 1325-1339.

Sañudo-Wilhelmy S. A., Rossi F. K., Bokuniewicz H., and Paulsen R. J., (2002) Trace metal levels in uncontaminated groundwater of a coastal watershed: importance of colloidal forms. *Environ. Sci. Technol.* **36**, 1435-1441.

Schmidt F., Elvert M., Koch B. P., Witt M., and Hinrichs K.-U., (2009) Molecular characterization of dissolved organic matter in pore water of continental shelf sediments. *Geochim. Cosmochim. Acta* **73**, 3337-3358.

Schwede-Thomas S. B., Chin Y. P., Dria K. J., Hatcher P., Kaiser E., and Sulzberger B., (2005) Characterizing the properties of dissolved organic matter isolated by XAD and C-18 solid phase extraction and ultrafiltration. *Aquatic Sciences* **67**, 61-71.

Senko M. W., Canterbury J. D., Guan S., and Marshall A. G., (1996a) A high-performance modular data system for fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **10**, 1839-1844.

Senko M. W., Hendrickson C. L., Paša-Tolić L., Marto J. A., White F. M., Guan S., and Marshall A. G., (1996b) Electrospray ionization fourier transform ion cyclotron resonance at 9.4 T. *Rapid Commun. Mass Spectrom.* **10**, 1824-1828.

Seow K. T., Meurer G., Gerlitz M., Wendt-Pienkowski E., Hutchinson C. R., and Davies J., (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned

546 from soil DNA: a means to access and use genes from uncultured microorganisms. *J.*
547 *Bacteriol.* **179**, 7360-7368.

548 Shen B., (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase
549 paradigms. *Curr. Opin. Chem. Biol.* **7**, 285-295.

550 Sherr E. B., Caron D. A., and Sherr B. F., 1993. Staining of heterotrophic protists for
551 visualization via epifluorescence microscopy. In: Kemp, P. F., Sherr, B. F., Sherr, E. B.,
552 and Cole, J. J. Eds.), *Handbook of Methods in Aquatic Microbial Ecology*. Lewis
553 Publishers, Boca Raton.

554 Sievert S. M., Kiene R. P., and Schulz-Vogt H. N., (2007) The sulfur cycle. *Oceanography* **20**,
555 84-90.

556 Simó R., Archer S. D., Pedrós-Alió C., Gilpin L., and Stelfox-Widdicombe C. E., (2002)
557 Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the
558 microbial food web in surface waters of the North Atlantic. *Limnol. Oceanogr.* **47**, 53-61.

559 Sleighter R. L. and Hatcher P. G., (2008) Molecular characterization of dissolved organic matter
560 (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh
561 resolution electrospray ionization Fourier transform ion cyclotron resonance mass
562 spectrometry. *Mar. Chem.* **110**, 140-152.

563 Sleighter R. L., McKee G. A., and Hatcher P. G., (2009) Direct Fourier transform mass spectral
564 analysis of natural waters with low dissolved organic matter. *Org. Geochem.* **40**, 119-
565 125.

566 Slomp C. P. and Van Cappellen P., (2004) Nutrient inputs to the coastal ocean through
567 submarine groundwater discharge: controls and potential impact. *J. Hydrol.* **295**, 64-86.

568 Suhre K. and Schmitt-Kopplin P., (2008) MassTRIX: mass translator into pathways. *Nucleic*
569 *Acids Res.* **36**, W481-484.

570 Taylor T. V., Mitchell T. K., and Daub M. E., (2006) An oxidoreductase is involved in
571 cercosporin degradation by the bacterium *Xanthomonas campestris* pv. zinniae. *Appl.*
572 *Environ. Microbiol.* **72**, 6070-6078.

573 Tremblay L. B., Dittmar T., Marshall A. G., Cooper W. J., and Cooper W. T., (2007) Molecular
574 characterization of dissolved organic matter in a North Brazilian mangrove porewater and
575 mangrove-fringed estuaries by ultrahigh resolution Fourier Transform-Ion Cyclotron
576 Resonance mass spectrometry and excitation/emission spectroscopy. *Mar. Chem.* **105**,
577 15-29.

578 Tso S.-F. and Taghon G., (2006) Protozoan grazing increases mineralization of naphthalene in
579 marine sediment. *Microb. Ecol.* **51**, 460-469.

Valiela I., Costa J., Foreman K., Teal J. M., Howes B., and Aubrey D., (1990) Transport of groundwater-borne nutrients from watersheds and their effects on coastal waters. *Biogeochemistry* **10**, 177-197.

van Beek C. G. E. M. and van der Kooij D., (1982) Sulfate-reducing bacteria in ground water from clogging and nonclogging shallow wells in the Netherlands river region. *Ground Water* **20**, 298-302.

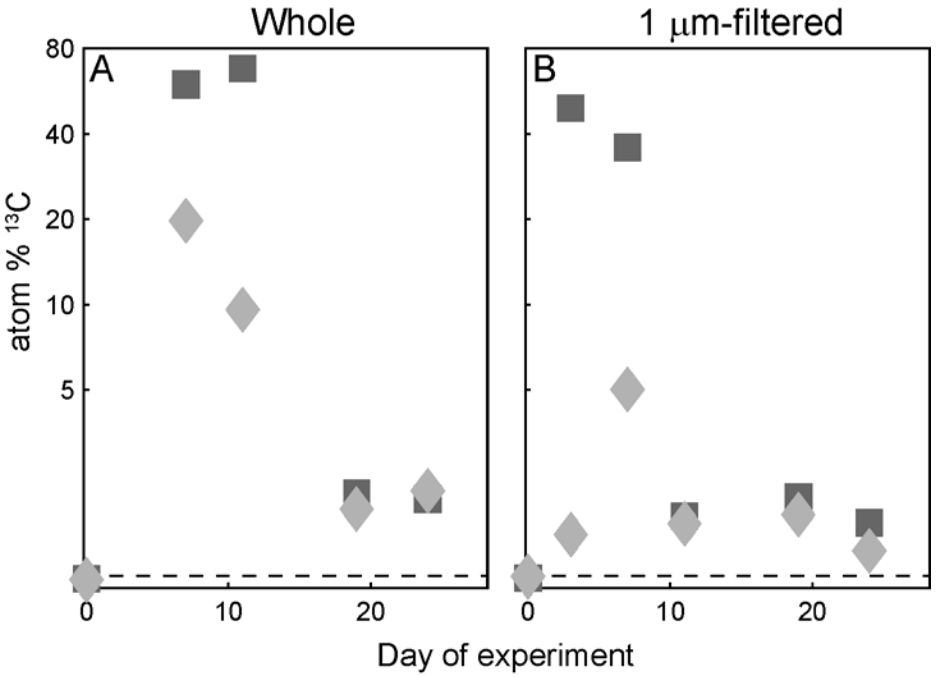
Wassenaar L., Aravena R., Fritz P., and Barker J., (1990) Isotopic composition (^{13}C, ^{14}C, ^{2}H) and geochemistry of aquatic humic substances from groundwater. *Org. Geochem.* **15**, 383-396.

Wawrik B., Kerkhof L., Zylstra G. J., and Kukor J. J., (2005) Identification of unique type II polyketide synthase genes in soil. *Appl. Environ. Microbiol.* **71**, 2232-2238.

Zektser I. S. and Everett L. G., (2004) *Groundwater resources of the world and their use*. United Nations Educational, Scientific and Cultural Organization, IHP-VI, Series on Groundwater No. 6.

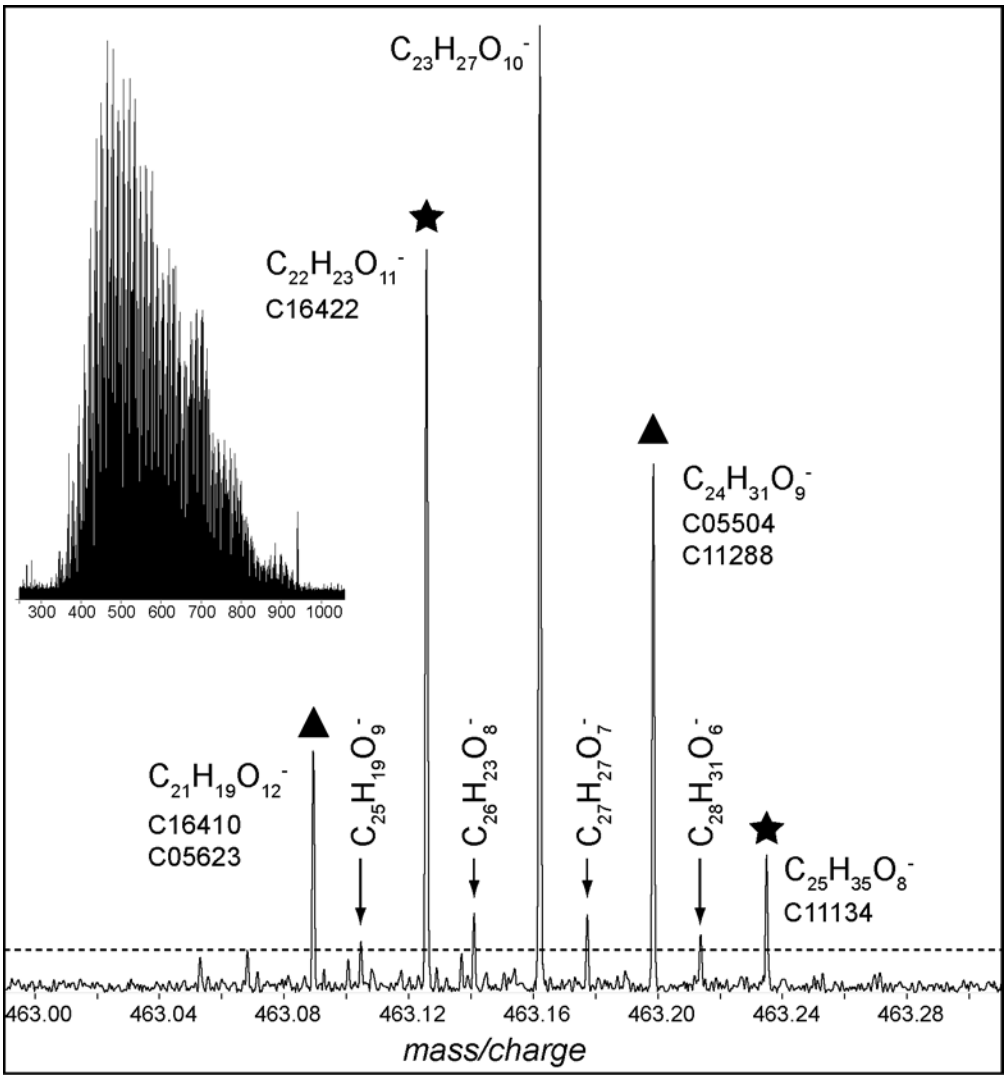
596 Longnecker and Kujawinski

597 Figure 1



599 Figure 2

600



601 Longnecker and Kujawinski

602 Figure 3