Untargeted Metabolomic Analysis of Oregon Coast Sediment Pore Water using Ultrahigh Resolution Mass Spectrometry

by
James Merlo

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science Chemical Engineering
(Honors Scholar)

Presented June 4, 2020
Commencement June 2020

AN ABSTRACT OF THE THESIS OF

James Merlo for the degree of Honors Baccalaureate of Science in Chemical Engineering presented on June 4, 2020. Title: Untargeted Metabolomic Analysis of Oregon Coast Sediment Pore Water using Ultrahigh Resolution Mass Spectrometry.

Abstract approved:_____

Rene Boiteau

One of the principal drivers of climate change is the concentration of greenhouse gases such as CO_2 in the atmosphere. A large portion of this CO_2 ends up in the waters off the continental coasts where it transforms into biomass. The major sink for this matter is the ocean sediments on continental margins where organic matter can be stored for thousands of years in extremely complex forms that are largely uncharacterized. Some of compounds remineralize into carbon dioxide and eventually re-enter the atmosphere, but others get locked into recalcitrant forms that are stored for millennia. Bacteria play a role in this cycle by transforming carbon from reactive forms to unreactive forms, but the mechanisms that this occurs by remains largely unclear. Progress in this field has largely been held back by the lack of available analytical tools to study the environment at such a small scale. New technologies such as Ultra High resolution mass spectrometry coupled with Liquid Chromatography gives us the ability to investigate these systems. Here, we investigated using 21T Fourier Transform Inductively Coupled Resonance Mass Spectrometry to conduct a metabolomic analysis of Oregon Coast sediment pore waters. This preliminary study has successfully predicted molecular formulas and classified compounds into groups of lipids, proteins, phytochemicals, carbohydrates, aminosugars, and nucleotides based on their stoichiometries. The method has allowed for elucidation of key differences between two distinct sediment regions (top of the sediments, and 4-8 cm into the sediments) efficiently using data plotting techniques such as van Krevelen diagrams and volcano plots. The study has captured the expected pattern of molecular weight distribution of DOM in the sediment pore waters, while also finding differences in oxidation states and compound class distributions that lead to additional research questions.

Key Words: Metabolomics, Dissolved Organic Matter, FT-ICRMS, Sediment Pore Waters

Corresponding e-mail address: jmerlo1110@sbcglobal.net

Untargeted Metabolomic Analysis of Oregon Coast Sediment Pore Water using Ultrahigh Resolution Mass Spectrometry

by
James Merlo

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Chemical Engineering
(Honors Scholar)

Presented June 4, 2020
Commencement June 2020

Honors Baccalaureate of Science in Chemical Engineering project of James Merlo presented on June 4, 2020.

APPROVED:

_____

Rene Boiteau, Mentor, representing CEOAS

_____

Miguel Goñi, Committee Member, representing CEOAS

_____

Clare Reimers, Committee Member, representing CEOAS

_____

Peter Chace, Committee Member, representing CEOAS

_____

Toni Doolen, Dean, Oregon State University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, Honors College. My signature below authorizes release of my project to any reader upon request.

_____

James Merlo, Author

Table of Contents

Background

Global Impact

The ocean and its underlying sediments are the largest sinks of carbon dioxide on the planet and are able to respond to changes in atmospheric carbon dioxide on anthropogenic and geological time scales. Understanding their dynamics is critical to making confident projections of global climate change in response to anthropogenic disturbances such as CO_2 emissions and natural disturbances. The interplay of biological, geochemical, and physical processes that transfer carbon from the surface ocean where it is fixed by primary producers to the underlying sediments where it is consumed or buried is not entirely understood.

Dissolved Organic Matter

Dissolved organic matter (DOM), contains both dissolved organic carbon (DOC) and dissolved organic nitrogen (DON). The ocean's DOC reservoir contains approximately 662 ± 32Pg (1015 g) C, exceeding the inventory of organic carbon in particles in the oceans by 200-fold and making it one of the largest bioreactive pools of carbon in the ocean, second only to dissolved inorganic carbon (DIC) at 38,100 Pg C (Hansell and Carlson 1988, Sarmiento and Gruber, 2006) The ocean inventory of DOC is comparable to the mass of inorganic C in the atmosphere (Eppley et al., 1987). Net oceanic uptake of CO_2 is ~1.9 Pg C $year^{-1}$ (Sarmiento and Gruber, 2006), so small perturbations in the processes regulating DOC production and removal affect the ocean and atmosphere CO_2 balance and organic carbon export. The large number of processes affecting DOM greatly complicates progress in understanding its cycling, processes including photoxidation, microbial chemoautotrophy, chemoheterotrophy and decomposition (Carlson and Hansell, 2015).

DOM can be categorized as labile (reactive), semi-labile, and refractory (unreactive), and understanding the chemical compositions of these categories is fundamental to understanding their cycling (Ridgwell and Arndt, 2015). Labile (or reactive) DOM consists of easily characterized low-molecular-weight biochemicals (simple sugars and carbohydrates, amino acids, proteins, vitamins, etc.). Labile DOM is released by metabolic processes of phytoplankton, zooplankton, and bacteria or photooxidation of upwelled DOM. Labile DOM may also be produced via deep-water chemoautotrophy. The biologically labile fraction, perhaps < 1% of the DOM inventory, may have reactivity lifetimes of minutes whereas biologically resistant DOM can be sequestered in the deep ocean for millennia (Jiao et al., 2014). At a fundamental level, it is recognized that certain biopolymers such as proteins and nucleic acids are particularly labile under a wide range of conditions due to relatively weak bonds between monomers (peptide bonds) or the nutrient requirements (phosphates and nitrogen) that such compounds provide. At the opposite end of the degradability spectrum are refractory molecules typically comprised of aliphatic compounds and cross-linked by relatively non-reactive ether bonds. Microbial activity transfers organic carbon from low concentrations of reactive carbon to progressively higher concentrations of refractory carbon (Ridgwell and Arndt, 2015).

Organic matter oxidation is coupled to the sequential utilization of terminal electron acceptors (TEAs), typically in the order of O_2, NO_3^-, Mn(VI), Fe(III) and SO_4^{2-} followed by methanogenesis and/or fermentation. Depending on the degradation pathway, organic matter is directly oxidized to CO_2, partly oxidized to intermediate compounds or reduced to CH_4. The

availability of oxygen plays a central role in the discussion of changes in organic carbon consumption rates on climate-relevant time scales. The preservation of sedimentary organic carbon tends to increase under anoxic conditions (Canfield et al., 1993, Hartnett et al., 1998). This increase could be the result of a thermodynamically limited degradation of refractory organic compounds in the absence of the powerful electron acceptor oxygen, the inability of anaerobic organisms to directly and completely degrade organic matter to CO_2, a decreased enzymatic activity, and/or a decreased availability of DOC adsorbed to mineral surfaces under anoxic conditions (Arndt et al., 2013). LaRowe and van Cappellen have developed a metric to help link the Gibbs Energy of the oxidation reaction to its nominal oxidation state of carbon (NOSC), providing a simple way to estimate the thermodynamic driving force for the mineralization of an organic compound with a given NOSC when coupled to the predominant electron acceptors in the system.

$$NOSC = -\left(\frac{-Z + 4a + b - 3c - 2d + 5e - 2f}{a}\right) + 4 \quad (1)$$

Z corresponds to the net charge of the organic compound, and the coefficients a, b, c, d, e, and f refer to the stoichiometry of C, H, N, O, P, and S. We can thus predict ΔG_{rxn} (change in Gibbs energy) for the oxidation of an organic compound with half reactions containing O_2 or with anaerobic electron acceptors such as $Fe(OH)_3$ simply through its composition. Low NOSC values under anaerobic oxidation reactions mechanisms may indicate that the compound is thermodynamically limited from remineralizing due to a low thermodynamic driving force, whereas compounds with high NOSC values correspond to a higher thermodynamic driving force. The microbial mineralization of abundant compound classes such as lipids and fatty acids may be thermodynamically limited if not completely inhibited under anaerobic conditions (Keiluweit et al., 2017).

Sediment Pore Waters

Dissolved organic matter (DOM) in marine sediment pore waters plays an important role in sediment carbon and nutrient (N and P) cycling. In many sediments, pore water concentrations of DOM are elevated by up to an order of magnitude over bottom water values (Burdige and Zeng, 1998). This implies that there is net production of DOM in sediments as a result of organic matter degradation processes. Some of the classic anaerobic processes are illustrated in *Figure 1* as sediment POM is transformed by microbial activity into inorganic end products or recalcitrant DOM. DOM accumulates with depth in most sediments as rates of particulate organic matter (POM) remineralization and the reactivity of

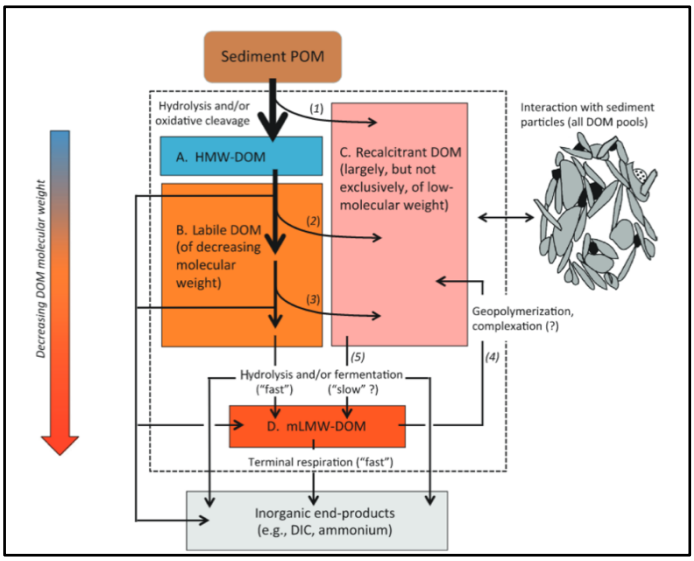

Figure 1: A conceptual model for DOM cycling in sediments based on the classic anaerobic food chain. Taken from Burdige and Komada, 2015. Sediment POM may end up in inorganic end products or recalcitrant DOM pools along its path.

sediment POM both decrease. (Burdige and Komada, 2015). The initial breakdown of POM in sediments occurs by microbial exoenzymes as well as by metazoan feeding. Viral lysis of living bacterial cells is also important in adding DOM compounds to sediment pore waters (Rowe and Deming, 2011).

As with other OM reservoirs, the redox conditions in sediments are a major driver of DOM accumulation, with low oxygen environments driving carbon storage for otherwise labile DOM. Several previous studies have attempted to characterize this DOM's composition. Deaminated peptides may represent ~25–45% of DOC accumulating in sediment pore waters as refractory molecules, as shown in studies by Abdulla et al., 2018 on sediment pore waters in the Santa Barbara Basin using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS). Others have reported high abundance of DON formulas having O/C and H/C ratios that overlap with those of carboxyl-rich alicyclic (CRAM) molecules (Schmidt et al., 2011). Schmidt et al., 2011 has further shown that some of these DON formulas could result from the deamination of peptides. Recent studies using this technique from sediment pore waters collected in the Santa Barbara basin have elucidated that abiotic nucleophilic addition reactions involving bisulfide (HS^-) and polysulfide (HSx^-) are the major sulfurization pathways that form dissolved organic sulfate (Gan et al., 2020). Some DOM in sediment pore waters also complexes trace metals and affects dissolved metal concentrations and fluxes from sediments (Burdige and Komada, 2015). Thus, high resolution mass spectrometry appears to be a promising method for determining the chemical mechanisms of DOM production, transformation, and fate.

The sediment pore waters collected in this investigation originated from the continental margins off the Oregon Coast. Continental margins are the transition zones between the coastal (water depth < 200 m) and the deep ocean, and they are comprised of the continental shelf (water depth < 200 m), the continental slope and the continental rise. They are important, but extremely complex sites for transformations affecting the global carbon cycle that are influenced by a variety of organic matter sources, by intensive lateral transport and large river plumes and by abrupt relocation of sediment through successive erosion/deposition cycles. They also reveal a strong temporal variability and a marked seasonality in energy, organic matter fluxes and redox conditions (Arndt et al., 2013.) The vast majority (over 90%) of carbon preservation and remineralization in all marine sediments occurs in estuarine and continental margin sediments (Hedges and Keil, 1995). This burial is intrinsically linked to the cycling of N, P, S, Fe and Mn and controls the oxygen content of the atmosphere on geological timescales (Hedges and Keil, 1995).

The connection between sediment DOM inputs and outputs including all of the biological, geochemical, and physical processes is poorly understood, but has major consequences on ocean and atmospheric carbon contents. The rates of DOM formation and decomposition, and the stoichiometry between DOC and other elements may help to understand these processes. Combining DOM chemical characterization approaches with "omic" technologies such as metabolomics that reveal microbial community structure and metabolic strategies is an approach that promises linking specific microbial groups with specific DOM compounds in sediment pore waters (Carlson and Hansell, 2015).

Untargeted Metabolomic Analysis

Metabolomics is the study of the low-molecular weight metabolites within a biological system. Untargeted metabolomics seeks to detect and describe as much of the metabolome (complete set of small-molecule chemicals found within a biological sample) as possible without bias (Kido Soule et al., 2015). The purpose of environmental metabolomics is the characterization of organisms' metabolic response to natural or anthropogenic stressors in the environment. The metabolic pathways of microbial organisms shape the marine carbon cycle through processes of carbon fixation and remineralization, and metabolic intermediates pass through DOM. Environmental metabolomics can be used to identify these compounds and gain insight into microbe-DOM interactions within the marine ecosystem (Kido Soule et al., 2015). Two mass spectrometry based methods for molecular analysis of DOM have been used traditionally, Liquid chromatography mass spectrometry, and direct infusion Fourier transform ion cyclotron resonoance mass spectrometry.

Using liquid chromatography-mass spectrometry (LC-MS) metabolomic methods, thousands of peaks can be detected from biological samples. Each of these peaks is referred to as a metabolite feature and corresponds to a detected ion with a unique mass-to-charge ratio and a unique retention time (Patti et al. 2012). Statistical evaluation of dense data matrices and chemically characterizing features that are different between samples is a bottleneck of metabolomics. The goal is to identify the metabolites that are significantly changing between classes of biological samples. T-tests or ANOVA together with Fold Change analysis serve for creation of volcano plots, are useful approaches for extraction and visualization of differentially accumulated metabolites in two dimensional space of statistical importance (p-value) and biological importance (fold change) (Piasecka et al., 2019). Inductively coupled mass spectrometry is unique because it enables assignment of molecular formulas, which is first step of identifying a compound and can be used to predict a compound class for the feature, nominal oxidation state, elemental stoichiometry, and molecule size, which all provide information relevant to carbon and nutrient cycling in sediments. Using ultrahigh resolution and high mass accuracy mass spectrometer techniques such as Fourier Transform Ion Cyclotron Resonance Mass Spectrometry enables assignment of molecular formulas that can be attributed to biomolecules (e.g. proteins and peptides) and any altered by-products without destroying their chemical structures. (Abdulla et al., 2018). FT-ICRMS is traditionally used with direct infusion analysis of samples without chromatography, because the long transient times require to collect mass spectra are too slow to resolve chromatographic peaks that elute over in narrow time windows. However, the benefit is that it provides mass accuracy needed to assign molecular formulas.

Elemental compositions can be visualized in van Krevelen plots of H/C and O/C ratios where they may be compared with the same ratios in possible biochemical precursors (lipids, proteins, carbohydrates), between samples and across various spatial scales. H/C and O/C ratios that diverge from precursors indicate DOM that has been degraded or transformed (Wakeham and Lee, 2019). Newer methods for determining compound classes that have proved more accurate than van Krevelen plots specifically for metabolites use C/H/O/N/P stoichiometric ratios of over 130,000 elemental formulas of compounds classified in 6 main categories: lipids, peptides, amino sugars, carbohydrates, nucleotides, and phytochemical compounds (oxy-aromatic compounds) (Rivas-Ubach et al., 2018). Multidimensional stoichiometric compound

classification (MSCC) constraints have shown categorization of elemental formulas to the main compound categories in biological samples with over 98% accuracy, representing a substantial improvement over any classification based on the classic H/C and O/C criteria alone (Rivas-Ubach et al., 2018).

The aforementioned graphical techniques and MSCC constraints are used here to conduct a pairwise comparison of sedimentary DOM with LC-FTICR-MS, comparing DOM found in a surface layer to a deeper sediment layer >4 cm. This represents an analytical advancement over previous FT-ICR-MS methods because chromatography provides (1) a 'retention' time and the ability to collect MS/MS fragmentation spectra, which enable the identification of particular metabolites, and (2) reduced ion suppression which enables relative quantitation of metabolites across different samples, and thus the implementation of the statistical data analysis workflows commonly used in metabolomics. The goals of this study were to develop a workflow that can combine the traditional LCMS data processing (feature detection) with the traditional FT-ICR MS data processing (formula assignment). A second goal was to evaluate the workflow by determining what proportion of features can be assigned to compound classes and what types. The third goal was to develop ways to analyze the new data comprised of 'metabolite features' that are binned into compound classes that describe their chemistry.

Materials and Methods

Pore water collection

Sediment pore water samples were obtained from cruise OC1904A on the R/V OCEANUS on April 22-26, 2019. The samples were collected from a depth of 142 m at site NH20 (Newport Hydrographic line 44.651˚N/124.528˚W). Sediment cores were obtained using two methods. The first method was using a hydraulically dampened gravity corer described in Reimers et al., 2012. This sediment core (length ~ 30 cm) was moved in its tube to a thermally controlled bath, maintained at 7.4˚C. The second method involved using a box-corer where sediment was brought onboard and 12-15 cm-long sub-cores were obtained. For each extraction method, Rhizon suction samplers (0.1 µm porous polymer, Rhizosphere Research, Wageningen, the Netherlands) were inserted vertically into the top of the core samples for 60-120 minutes. There were two different Rhizon sampler collections, one set that reached pore waters between 0-4 cm and another that reached 5-8 cm into the column. It was noted that no nitrogen purging was conducted during the pore water extraction which may have resulted in a white precipitate indicative of sulfidic compounds appearing in the samples. The samples were transferred to acid washed pre-combusted borosilicate vials and stored at -20˚C.

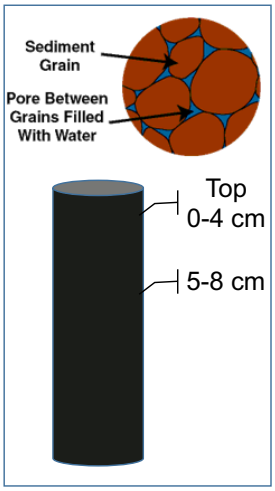

Figure 2: Locations of pore water samples obtained within a sediment core

Extraction

Each porewater sample (10mL) was acidified with the addition of 100µL LCMS grade formic acid. The organic matter was preconcentrated and desalted by loading onto a primed and rinsed solid phase extraction (SPE) column (agilent PPL, 100mg), rinsed with 2mL of 1% formic acid in qH_2O, and eluted with 1mL of MeOH. The eluted sample was concentrated in a vacuum

centrifuge and brought up in a final volume of 500mL qH_2O. As an internal standard, 25µL of 50µM cyanocobalamin was added to each sample.

LC-MS analysis

The concentrated porewater extracts were separated on a bio-compatible high-pressure liquid chromatography (HPLC) system (Dionex Ultimate 3000RSLC) using a C18 column (2.1 x 100 mm, 3 µm particle size; Hamilton). Samples were injected in a 50µL volume and separated with a gradient from 95% solvent A (5 mM aqueous ammonium formate) and 5% solvent B (5 mM ammonium formate in distilled MeOH) to 95% B over 20 minutes, followed by isocratic elution at 95% B for 10 min at a flow rate of 0.2 mL/min. The flow from the LC was coupled to the Ultra-high resolution 21 Tesla Fourier Transform – Ion Cyclotron Resonance Mass Spectrometer (FT-ICRMS) using a heated ESI source set to a capillary voltage of 3500 V; sheath, auxiliary, and sweep gas flow rates of 12, 6, and 2 (arbitrary units); and ion transfer tube and vaporizer temperatures of 300 and 75 °C. Mass spectra were collected with 1 second transients in the ICR cell, yielding 600K resolution at 400 *m/z* and sub ppm mass accuracy. MS/MS fragmentation spectra were collected by collision induced dissociation (CID) of the major features with a collision energy of 40.

Data processing

LC-MS data were converted to opensource MzML file formats using MsConvert (proteowizard). Features were detected and peaks aligned in MSDial with a minimum peak height setting of 3000, a mass slice width of 0.01 Da, and including ion adducts of [M+H]+, [M+NH4]+, and [M+Na]+ (Tsugawa et al., 2014). MS/MS identification was referenced using the publicly available library MSP Spectral Database. The *m/z* values were calibrated and formulas were assigned using custom scripts in R based on the Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry methods (Kind and Fiehn, 2007). Compound motifs including lipids, proteins, amino-sugars, nucleotides, carbohydrates, and phytochemicals were assigned based on stoichiometric constraints of hydrogen, oxygen, nitrogen, and carbon outlined in Rivas-Ubach et al., 2018. Features that matched a molecular formula and *m/z* ratio with compounds found in the Metlin database were recorded.

Results and Discussion:

Out of 11 pore water samples collected, 8 originated from the top 0-4 cm into the sediment column (referred to as "top") and 3 originated from the deeper layer 4-8 cm into the sediment column. MS Dial aligned 14,576 features that were detected in the samples from the LCMS. Of these features, 6,214 (43%) were assigned molecular formulas. Some of these molecular formula assignments had masses that deviated from their expected mass values, and after filtering the data by including only features with an absolute mass error < 0.5 ppm, there were 4,055 features remaining. These features are illustrated in a volcano plot as shown in *Figure 3*. Of these features, 816 (20%) showed significantly different mean intensities from the top of the sediments to the deeper layer ($p<0.05$). 798 features were significantly more prevalent in the top of the sediments, and 18 were more prevalent in the bottom of the sediments. The breakdown of compound classes in each sediment group is shown in *Figure 4*.

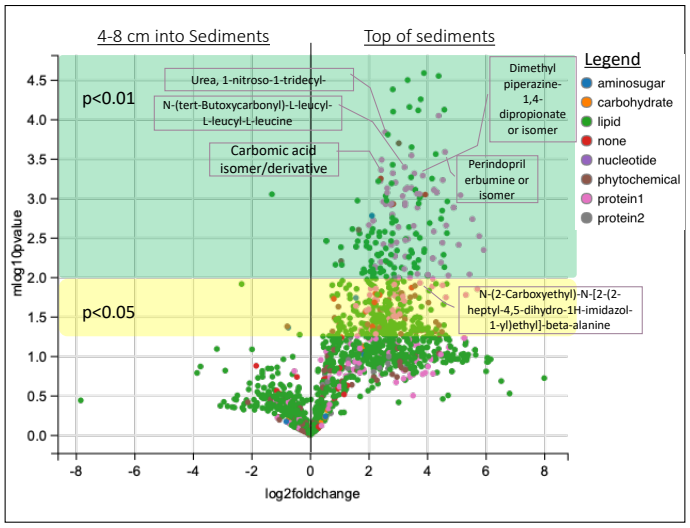

Figure 3: Volcano plot showing the distribution of the 2000 most intense features with predicted molecular formulas by its log fold change. X axis show the $\log_2$ of the fold change where the fold change is defined as the mean intensity of a feature found in the top sediment section divided by the mean intensity of a feature found in the lower sediment section. The y-axis shows the -log(pvalue), illustrating which features have mean intensities that are statistically significant between the two regions for comparison. Features in the positive region had higher intensity values detected in the top of the sediments compared to the bottom and vice versa. Significant features which matched M/Z and molecular formulas to known compounds in METLIN Library are labeled. This visualization helps determine which features may be significantly different in two distinct sample groups.

Lipids were by far the most abundant compound class found in both samples, some being more prevalent in the uppermost sediments, while others being more prevalent in the lower sediments. To understand the reason for this distribution, further investigation into the chemical makeup of the lipids in these two sections would need to be undertaken. Part of the explanation is that the SPE using a C18 column and the chromatography system is optimal for these compound classes compared to sugars and other polar metabolites because the polar metabolites pass through the extraction column without being retained. It is possible that the older, deeper sediments have bacteria that are producing these compounds in situ, or they are oxidized forms of lipids that have been degrading in the sediments over time. The average molecular weights of the lipids that were found in significantly higher intensities in the top of the sediments, were larger than those found in the lower sediment section (*Figure 5)*. This trend follows the prediction of DOM cycling described in *Figure 1*, and supports the idea that these lipids may be remnants of anaerobic degradation of sedimentary particulate organic matter. The average NOSC of the lipids found in these compounds (N=10) was -1.243 in the top of the sediments and -0.999 in the bottom of the sediments (*Figure 4)*. The more oxidized lipids in the lower sediments suggests more 'energy yielding' lipids make up a larger proportion of the DOM in lower sediments compared to top sediments. This conflicts with the explanations of previous studies including Keiluweit et al. 2017 that argue the thermodynamics of preservation explain the DOM composition in soils. At the individual compound class level, other controls over composition might be more important than the NOSC

in these sediments such as the biomass precursors in the sediments or the specific microbial mechanisms in each layer.

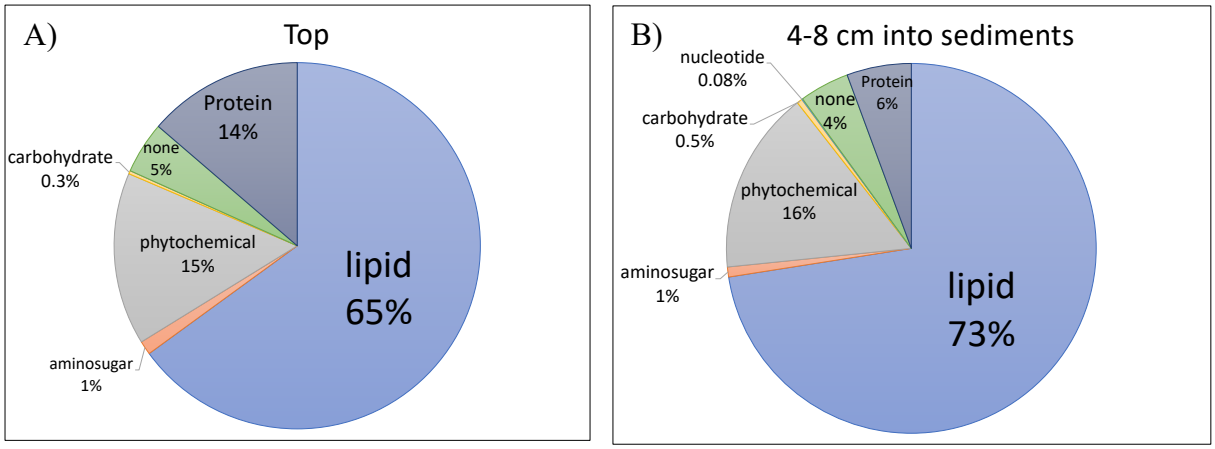

Figure 4: A) Breakdown of class features found in the Top sediments N = 2796. B) Breakdown of class features found in the samples 4-8 cm into the sediments (N=1259). There is an overlap in features found in both sediments. Proteins make up a smaller portion of the features in the lower sediment group.

Phytochemical-like compounds were found in similar ratios in both sediment sections, with a small portion being unique to only the lower sediments. Similar to lipids, the molecular weights decreased significantly from the top to the lower sediments (*Figure 5*) suggesting they may undergo similar transformation mechanisms during DOM cycling. Phytochemicals differ from lipids when comparing NOSC, as phytochemicals did not show any significant difference in mean NOSC values in the two sediment pore water layers. The NOSC for the phytochemical-like compounds was larger than that of lipids, consistent with the trend observed in Keiluweit et al., 2017 that was attributed to thermodynamics.

Proteins accounted for a larger proportion of the metabolite features found in the top of the sediments than in the lower sediment sections (*Figure 4)*. There were no protein features that were found significantly more abundant in the lower sediment section than the top section. To identify their origin or understand their cycling, further experiments would have to be conducted testing the different peptide degradation methods as shown in *Figure 6*. Most of the masses observed do not match the masses of canonical peptides, so there is a likelihood that the dissolved features have undergone structural changes. Proteinaceous material has been evidenced to be transformed in soils by mechanisms such as: hydrolysis and deamination, both reducing the molecular size and nitrogen content of the products and intermediates; oxidation and hydration of the intermediates; and methylation and dehydration (Schmidt et al., 2011).

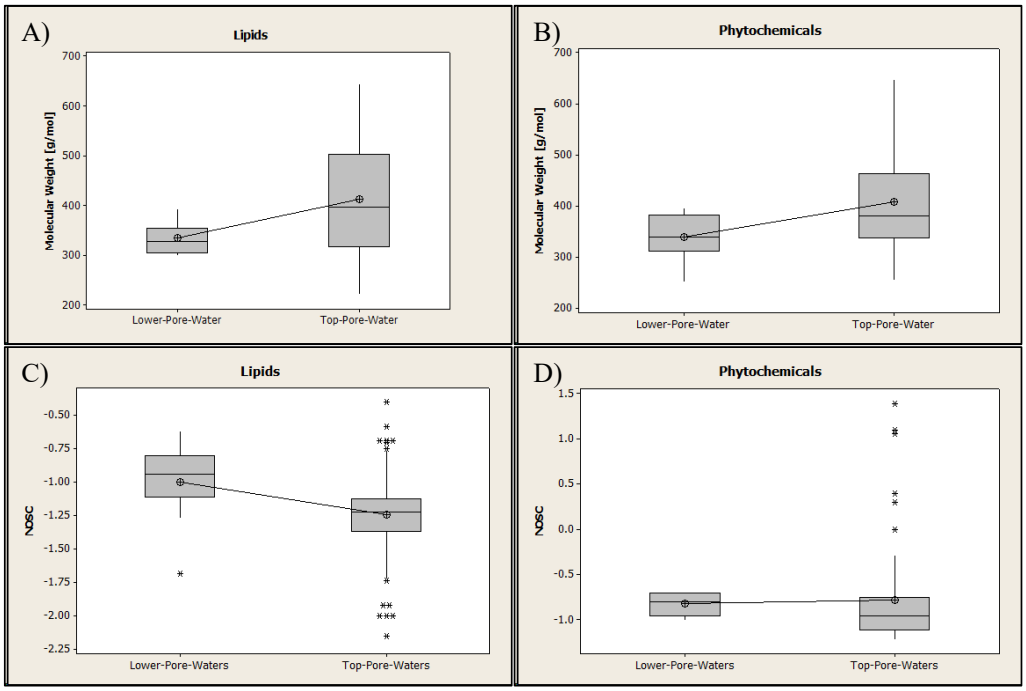

Figure 5: Molecular weight boxplot of A) Lipids found with significantly higher abundances in the lower sample (N=10, $\bar{x}$ = 334.7 g mol^{-1}) and with significantly higher abundances in the top water (N = 514, $\bar{x}$ = 339.9 g mol^{-1}), (p<0.001), B) Phytochemicals found with significantly higher abundances in the lower sample (N=7, $\bar{x}$ = 334.7 g mol^{-1}) and with significantly higher abundances in the top water (N = 73, $\bar{x}$ = 408 g mol^{-1}), (p=0.01) . The decrease in molecular weight can be attributed to the oxidative and fermentative processes that transform DOM as they cycle through sediments. Boxplot of NOSC for C) Lipids found with significantly higher abundances in the lower sample (N=10, $\bar{x}$ = -0.999) and with significantly higher abundances in the top water (N = 514, $\bar{x}$ = -1.243), (p=0.029), and D) Phytochemicals found with significantly higher abundances in the lower sample (N=7, $\bar{x}$ = -0.822) and with significantly higher abundances in the top water (N = 73, $\bar{x}$ =-0.785), (p=0.633). Unlike lipids, the means of the NOSC for phytochemicals are not statistically different.

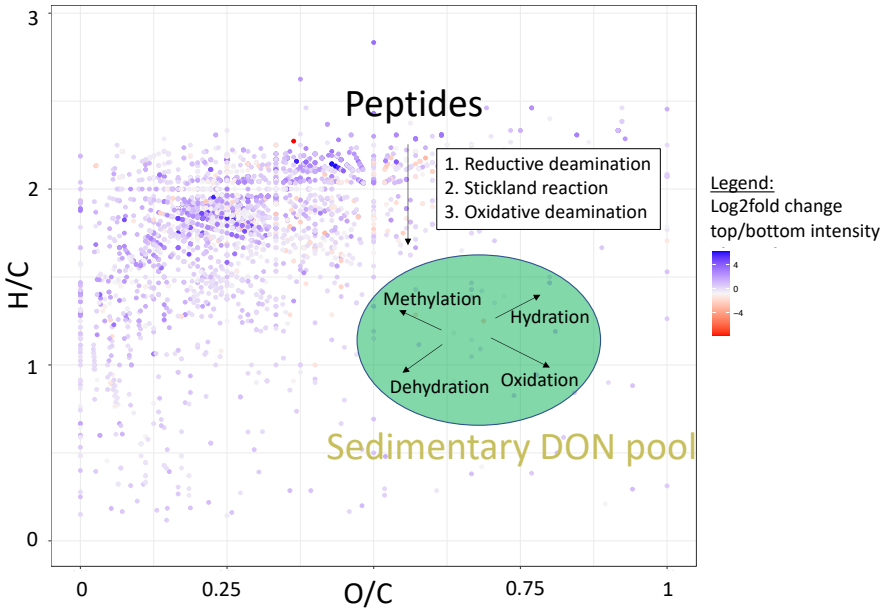

Figure 6: van Krevelen Diagram of all the features detected with nitrogen bearing molecular formulas. A mathematical model described in Schmidt et al., 2011 shows that the assemblages of nitrogen-bearing molecular formulas are potential products of proteinaceous material that have been transformed by the following reactions: hydrolysis and deamination, both reducing the molecular size and nitrogen content of the products and intermediates; oxidation and hydration of the intermediates; and methylation and dehydration.

Based on results of this untargeted analysis, more targeted approaches should be conducted to further understand the role of microbial activity on DOM degradation in these Oregon Coast sediments. Potential future work may include observing the transformations of proteins, lipids, and phytochemicals in centimeter intervals down the sediment core in order to better understand the types of processes that are occurring. Obtainment of DOC concentrations and nutrient concentrations of the sediment water and the overlying water may aid in elucidating the flux of DOM between the two locations. Incubating sediments under different redox conditions and observing the changes using this method may also help explain the evidence for metabolite features at different sediment depths. Furthermore, detection and characterization of recalcitrant DOM or of carboxyl-rich alicyclic molecules (CRAM) could be a useful tool in these types of untargeted studies, but additional ways to analyze the data for their detection may need to be considered to obtain this capability. Now that there is a developed framework for observing differences in chemical formulas and metabolite classes with high mass accuracy the research questions are nearly limitless.

Significant progress was made in each of three goals in the study: (1) developing a workflow that can combine the traditional LCMS data processing (feature detection) with the traditional FT-ICR MS data processing (formula assignment), (2) evaluating of the workflow (what proportion of features can be assigned to compound classes and what types), and (3) developing ways to analyze the new data comprised of 'features' that are binned into compound classes that describe their chemistry. Even though this preliminary study was successful in each of these three goals, there were a number of challenges that were faced when attempting to analyze this dataset that may need further development before the methodology can reach its full potential. Some challenges included efficiently referencing the features to libraries of compounds. MSDial and Metlin were both used as databases, with only Metlin providing lists of compounds that matched both m/z ratios and molecular formulas. MSDial has the capability to match mass spectra with their reference databases, but no clear matches were assigned in this dataset. Furthermore, a few contaminants from sample collection and processing methods were detected in the system. Compounds that are known plasticizers as well as a constituent with the molecular formula of nylon was detected. Although these features can be discounted based on process blanks, future studies should aim to minimize the presence of such contaminants as their presence may hinder the ability to detect metabolites of interest. Future experimental designs should aim to keep the sample size of every comparison group large enough to ensure that features are observed in multiple samples, raising confidence that they are environmentally significant.

Conclusion

This pilot study for using LC-FTICR-MS to study sediment pore waters has been successful at obtaining a wide variety of molecular formulas and classifying compounds based on elemental stoichiometries. The method has allowed for elucidation of key differences between two distinct sediment regions in an efficient way using data plotting techniques such as van Krevelen diagrams and volcano plots. The study has captured the expected pattern of molecular weight distribution of DOM in the sediment pore waters, while also finding differences in oxidation states and distributions that challenge the currently prevailing dogma of organic matter accumulation under reducing conditions. This study detected thousands of compounds that were not available in the current libraries of compounds, demonstrating that the vast majority of metabolites comprising marine DOM are currently missing from compound libraries. The

approach, algorithms, and workflows to process metabolomic data developed from this investigation can be utilized to start answering questions about microbial activity that has never before been attainable, giving scientists and climate modelers insight into key microbial processes that can impact the global carbon cycle.

References

Abdulla, H.A.; Burdige, D.J; Komada, T., Abiotic formation of dissolved organic sulfur in anoxic sediments of Santa Barbara Basin, Organic Geochemistry, **2020,** Volume 139, 103879, ISSN 0146-6380, https://doi.org/10.1016/j.orggeochem.2019.05.009.

Abdulla, H.A.; Burdige, D.J; Komada, T., Accumulation of deaminated peptides in anoxic sediments of Santa Barbara Basin, Geochimica et Cosmochimica Acta, **2018**, Volume 223, Pages 245-258, ISSN 0016-7037, https://doi.org/10.1016/j.gca.2017.11.021.

Arndt, S.; Jørgensen, B.B.; LaRowe, D.E.; Middelburg, J.J.; Pancost, R.D.; Regnier, P., Quantifying the degradation of organic matter in marine sediments: a review and synthesis. Earth-Sci. **2013,** Rev. 123, **2013,** 53–86.

Azam, F.; Godson R. E.; Size distribution and activity of marine microheterotrophs, Limnology and Oceanography, **1977,** 3, doi: 10.4319/lo.1977.22.3.0492.

Burdige, D.J; Komada, T., Chapter 12 - Sediment Pore Waters, Editor(s): Dennis A. Hansell, Craig A. Carlson, Biogeochemistry of Marine Dissolved Organic Matter (Second Edition), Academic Press, **2015,** Pages 535-577, ISBN 9780124059405, https://doi.org/10.1016/B978-0-12-405940-5.00012-1.

Burdige, D.J.; Zheng, S., The biogeochemical cycling of dissolved organic nitrogen in estuarine sediments. Limnol. Oceanogr. **1998,** 43, 1796–1813.

Canfield, D.; Thamdrup, B.; Hansen, J.W.; The anaerobic degradation of organic matter in Danish coastal sediments. Geochim. Cosmochim. **1993,** Acta 57, 3867–3883.

Carlson, C.A.; Hansell, D.A.; DOM sources, sinks, reactivity, and budgets. In: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter, 2nd ed., **2015,** pp. 65–126. https://doi.org/10.1016/B978-0-12-405940-5.00003-0. Academic press

Eppley, R.W.; Stewart, E.; Abbott, M.R.; Owen, R.W., Estimating ocean production from satellited-derived chlorophyll: insights from the Eastropac data set. Oceanol. Acta SP, **1987,** 109–113.

Gan, S.; Schmidt, F.; Heuer, V.; Goldhammer, T.; Witt, M.; Hinrichs, K.; Impacts of redox conditions on dissolved organic matter (DOM) quality in marine sediments off the River Rhône, Western Mediterranean Sea, Geochimica et Cosmochimica Acta, **2020,** Volume 276, Pages 151-169, ISSN 0016-7037, https://doi.org/10.1016/j.gca.2020.02.001.

Hartnett, H.; Keil, R.; Hedges, J., Influence of oxygen exposure time on organic carbon preservation in continental margin sediments. *Nature,* **1998,** 391, 572–575 , https://doi.org/10.1038/3535.1

Hansell, D.A.; Carlson, C.A., Deep ocean gradients in dissolved organic carbon concentrations. Nature **1998,** 395, 263–266.

Hedges, J.I. and Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. Mar. Chem. **1995,** 49, 81–115.

Jiao, N.; Herndl, G.J.; Hansell, D.A.; Benner, R.; Kattner, G.; Wilhelm, S.W.; Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. Nat. Rev. Microbiol. **2010,** 8, 593–599.

Jiao, N.; Robinson, C.; Azam, F.; Thomas, H.; Baltar, F.; Dang, H.; Hardman-Mountford, N.J.; Johnson, M.; Kirchman, D.L.; Koch, B.P.; Legendre, L.; Li, C.; Liu, J.; Luo, T.; Luo, Y.-W.; Mitra, A.; Romanou, A.; Tang, K.; Wang, X.; Zhang, C.; Zhang, R., Mechanisms of microbial carbon sequestration in the ocean – future research directions. Biogeosciences **2014,** 11, 5285–5306. https://doi.org/10.5194/bg-11-5285-2014.

Keiluweit, M.; Wanzek, T.; Kleber, M. Anaerobic microsites have an unaccounted role in soil carbon stabilization. *Nat Commun* **2017**, 8, 1771. https://doi.org/10.1038/s41467-017-01406-6

Kind T.; Fiehn O., Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics,* **2007,** 8-105. doi:10.1186/1471-2105-8-105

LaRowe, D. E.; and Van Cappellen, P., Degradation of natural organic matter: a thermodynamic analysis. Geochim. Cosmochim. Acta 75, **2011,** 2030–2042.

Kido Soule, C. M.; Longnecker, K.; Johnson, W. M.; Kujawinski, E. B., Environmental metabolomics: Analytical strategies, Marine Chemistry, **2015,** Volume 177, Part 2, Pages 374-387, ISSN 0304-4203, https://doi.org/10.1016/j.marchem.2015.06.029.

Patti, G.; Yanes, O.; and Siuzdak, G., Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **2012,** 13, 263–269, https://doi.org/10.1038/nrm3314

Piasecka, A.; Kachlicki, P.; and Stobiecki, M.; Analytical Methods for Detection of Plant Metabolomes Changes in Response to Biotic and Abiotic Stresses. *International journal of molecular sciences*, **2019,** *20*(2), 379. https://doi.org/10.3390/ijms20020379

Reimers, C. E.; Özkan-Haller, H. T.; Berg, P., Devol, A.; McCann-Grosvenor, K.; and Sanders, R. D., Benthic oxygen consumption rates during hypoxic conditions on the Oregon continental shelf: Evaluation of the eddy correlation method, *J. Geophys. Res.*, **2012,** 117, C02021, doi:10.1029/2011JC007564.

Ridgwell, A. and Arndt, S., Chapter 1 - Why Dissolved Organics Matter: DOC in Ancient Oceans and Past Climate Change, Editor(s): Dennis A. Hansell, Craig A. Carlson,

Biogeochemistry of Marine Dissolved Organic Matter (Second Edition), Academic Press, **2015,** Pages 1-20, ISBN 9780124059405, https://doi.org/10.1016/B978-0-12-405940-5.00001-7.

Rowe, G.T.; Deming, J.W., An alternative view of the role of heterotrophic microbes in the cycling of organic matter in deep-sea sediments. Mar. Biol. Res. **2011,** 7, 629–636.

Sarmiento, J.L.; Gruber, N., Ocean Biogeochemical Dynamics. Princeton University Press, **2006,** Princeton, NJ.

Schmidt, F.; Koch, B.P.; Elvert, M.; Schmidt, G.; Witt, M.; Hinrichs, K.U.; Diagenetic transformation of dissolved organic nitrogen compounds under contrasting sedimentary redox conditions in the Black Sea. Environ. Sci. Technol. **2011,** 45, 5223–5229.

Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; and Siuzdak, G., METLIN: a metabolite mass spectral database. *Therapeutic drug monitoring*, **2005,** *27*(6), 747–751. https://doi.org/10.1097/01.ftd.0000179845.53213.39.

Wakeham, S.G.; Lee, C., Limits of our knowledge, part 2: Selected frontiers in marine organic biogeochemistry, Marine Chemistry, **2019,** Volume 212, Pages 16-46, ISSN 0304-4203, https://doi.org/10.1016/j.marchem.2019.02.005.

Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; and Arita, M., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods*, **2015,** *12*(6), 523–526. https://doi.org/10.1038/nmeth.3393