Metabolomes are an emergent property of marine microbial systems

Joshua S. Sacks

A dissertation
Submitted in partial fulfillment of the
Requirements for the degree of

Doctor of Philosophy

University of Washington
2025

Reading Committee:
Anitra E. Ingalls, Chair
E. Virginia Armbrust
Randelle M. Bundy

Program Authorized to Offer Degree:
Oceanography

University of Washington

Abstract

Metabolomes are an emergent property of marine microbial systems

Joshua S. Sacks

Chair of the Supervisory Committee:
Anitra E. Ingalls
Oceanography

Microbial communities underpin ocean biogeochemical cycles and form the basis of marine ecosystems. Microbes use and maintain pools of small, organic molecules called metabolites as the core intermediates and building blocks of their metabolisms and as currencies in the exchanges of nutrients, energy, and information. Microbes also use metabolites to adapt to environmental conditions such as osmotic pressure and oxidative stress. Metabolomics, or the measurement of all metabolites in a system (the metabolome), is a powerful tool for characterizing microbial systems in the marine environment. Metabolomics studies have revealed the role of taxonomy and physiochemical factors in controlling marine microbial metabolite pools. However, microbial ecosystems are also shaped by the myriad of biological and chemical interactions that occur among their component parts, leading to the emergence of new dynamics at the community level not predictable from studying a single organism. In this dissertation, I employ metabolomics to characterize the metabolite pools of marine microbial communities at an unprecedented scale and investigate the roles of a diverse suite of interactions in shaping metabolic processes. In Chapter 2, I developed a novel extraction approach for separating dissolved metabolites from seawater, enabling the first measurement of many dissolved compounds in marine samples. In Chapter 3, I used metabolomics and transcriptomics to characterize viral infection of the globally important cyanobacterium *Prochlorococcus*, revealing extensive metabolic remodeling and demonstrating the importance of viral infection in shaping marine metabolite dynamics. In Chapter 4, I used uptake experiments to quantify the cycling of two metabolites, glycine betaine and homarine, and show that the uptake rates and standing stocks of these compounds is dependent on the concentrations of other metabolites present in the system. Finally, in Chapter 5, I measured pools of osmolytes inside and outside of cells at the basin scale in the North and Equatorial Pacific Ocean and the Salish Sea. I found that the composition of osmolyte pools was largely conserved at the community level despite vast changes in microbial biomass, community composition, nutrient status, and productivity. Taken together, these projects demonstrate that the composition and fluxes of marine metabolite pools are an emergent property that stem from a wide range of interactions across scales in microbial ecosystems.

Table of Contents

Acknowledgements 1

Chapter 1: Introduction 3

1.1 A microbial ocean 3

1.2 Metabolomics 4

1.3 Emergent Properties and Interactions in Microbial Systems 7

1.4 Metabolomics in the search for life on other worlds 9

Chapter 2: Quantification of Dissolved Metabolites in Environmental Samples through Cation-Exchange Solid Phase Extraction (CX-SPE) paired with Liquid Chromatography-Mass Spectrometry 10

2.1 Abstract:...11

2.2 Introduction:..12

2.3 Methods:..15

2.4 Assessment: 24

2.5 Discussion:...36

2.6 Conclusion:..41

2.7 Acknowledgements:...41

2.8 Figures:..42

2.9 Tables:..48

2.10 Supplemental Text:..53

2.11 Supplemental Figures: 54

2.12 Supplemental Tables: 63

Chapter 3: Remodeling of *Prochlorococcus* Metabolism During Viral Infection 66

3.1 Abstract:...66

3.2 Introduction:..67

3.3 Methods:..68

3.4 Results and Discussion:..72

3.5 Conclusions:..85

3.6 Acknowledgements:..86

3.7 Figures: 87

3.8 Supplemental Text: 94

3.9 Supplemental Figures: 108

3.10 Supplemental Tables: 113

Chapter 4: The cycling of glycine betaine and homarine in marine microbial communities: quantitative flux measurements and the role of competitive uptake inhibition 115

4.1 Abstract: 115

4.2 Scientific significance statement: 116

4.3 Introduction: 116

4.4 Methods: 118

4.5 Results and Discussion: 121

4.6 Conclusion: 127

4.7 Acknowledgements: 127

4.8 Figures: 128

4.9 Tables: 131

4.10 Supplemental Text: 133

4.11 Supplemental Figures: 142

4.12 Supplemental Tables: 149

Chapter 5: A conserved osmolome in the North Pacific Ocean 150

5.1 Abstract: 150

5.2 Introduction: 151

5.3 Results and Discussion: 152

5.4 Methods: 166

5.5 Acknowledgements: 174

5.6 Figures: 175

5.7 Supplemental Figures: 181

5.8 Supplemental Tables: 190

Chapter 6: Conclusion: 192

Bibliography: 196

Acknowledgements

First and foremost, thank you to my advisor, Anitra Ingalls. For the past six years you have been an incredible champion, mentor, and friend. You pushed me to constantly think deeper and more critically and gave me a safe space to fail, grow, and learn to lead. The Ingalls Lab has been an amazing place to train as a scientist. Thank you to Katherine Heal and Angie Boysen for teaching me the foundations of metabolomics and oceanography, serving as role models of how to perform cutting edge and inclusive science, and continuing to act as professional and scientific mentors to this day. Thank you to our lab manager Laura Carlson. Without your remarkable analytical and laboratory skills, this dissertation would not have been possible. Will Kumler and I have moved through every stage of graduate school together. Thank you for being a great friend and often being the only person in the room who actually understands what I am talking about. Thank you to the other current and former Ingalls Lab members: Frank Xavier Ferrer-González, Susan Garcia, Natalie Kellog, Hannah Dawson, Regina Lionheart, Caludia Luthy, Raisha Rahman for creating an exciting and intellectually stimulating environment and for extensive field, lab, and moral support. An additional thank you to the undergraduate and post-baccalaureate students in the Ingalls Lab who have contributed to collecting and analyzing the samples in this dissertation: Alex Meyers, Leland Wood, Everetta Rasyid, Anna Finch, Kate Faber, Amy Wang, Delilah Kaufman, Natalie Kledzik, Andrew Margolis.

Thank you to my committee: Ginger Armbrust, Randie Bundy, Drew Groman-Lewis, and David Butman for your personal and professional guidance and thank you to the School of Oceanography and the UW Astrobiology Program for providing a supportive community and numerous opportunities to expand my intellectual and professional horizons.

My PhD research has taken place in the larger context of the SCOPE (Simmons Collaboration on Ocean Processes and Ecology) and DiNiMITE (Dissolved Nitrogenous Metabolites: Investigations into Transformations and Exchanges) Programs. Through fieldwork, meetings, and shared projects, these collaborations have connected me to a broad network of mentors and peers who have taught me so much and become wonderful friends. Thank you to Debbie Lindell, Angel White, Dave Karl, Oscar Sosa, Daniel Muratore, Emily Seelen, Shiri Graff van Creveld, Francois Ribalet, Lexi Jones-Kellett, Stephen Blaskowski, Zinka Bartolek, Elaina Thomas, Kathy Qi, Travis Mellett, Patrick Monreal, Jennifer Beatty, and Brandon Brenes, among many others, for keeping me sane at sea and enabling my research to expand beyond the bounds of metabolomics into a truly interdisciplinary endeavor.

Seattle and my community here have provided an incredible backdrop to my graduate school experience. Thank you to Mary Margaret Stoll and Zoe Krauss for always being there through the highs and lows of graduate school. Thank you to Ryan Newell, Koll Roberts, and all of Seattle Frontrunners for constant companionship over hundreds of miles of long runs and countless laps around the track. Thank you to Adam Sokol and Kara Dastrup for being the being the best Craigslist roommates one could ask for. Finally, thank you to my partner, Matt Simon, for your constant love and support and for always being excited to look at my latest figure.

My parents, Paul Sacks and Linda Walters, nurtured in me a love for the natural world and a hunger for discovery and adventure. Thank you for your encouragement, for your guidance, and for giving me the space to find my own way in the world of science. I love you both and I am honored to be joining your ranks with a PhD of my own.

Chapter 1: Introduction

1.1 A microbial ocean

Marine microbes drive biogeochemical cycles in the ocean[1]. Phytoplankton use photosynthesis to fix carbon (convert inorganic carbon dioxide into organic matter). The magnitude of oceanic net carbon fixation is staggering at approximately 50 gigatons of carbon per year, roughly equivalent to the amount of carbon fixation performed by all terrestrial systems combined[2]. This fixed organic matter is the basis for all marine ecosystems and fuels the export of carbon to the deep ocean, partially regulating Earth's climate[3,4]. The amount of carbon fixation that occurs and its fate within an ecosystem are controlled by a complex and interconnected array of physical (ex. light, temperature), chemical (ex. pH, nutrients), and biological (ex. predation, symbioses) factors[5–8]. Understanding the controls on marine microbial communities and the fate of the organic matter they produce is critical to understanding the Earth system and predicting how it will change with climate change.

A critical component of organic matter in microbial systems is small, microbially produced, organic molecules called metabolites. Inside cells, metabolites serve as the building blocks of macromolecules, the intermediates of metabolic processes, and important functions such as resource storage[9,10] and stress response[11–13]. Outside cells, metabolites mediate microbial interactions (ex. cross feeding, signaling)[14–16] and serve as an important control on microbial community composition[17,18]. Metabolites are also thought to serve as key currencies in global elemental cycles of carbon, nitrogen, and sulfur[3,19–21]. In this dissertation, I use metabolomics, or the direct measurements of all metabolites in a system, to characterize microbial interactions, quantify the role of metabolites in marine biogeochemical cycles, and identify basin scale patterns in metabolites across diverse microbial communities in the North Pacific Ocean.

1.2 Metabolomics

Metabolomics offers a window into microbial systems by providing a snapshot of the standing stocks of metabolites inside and outside of cells[22]. Metabolomics can be combined with experimental treatments or environmental gradients to examine how compounds change across time, space, or condition, providing insights into changes in microbial metabolism across these factors. Marine microbial metabolomics is a young and rapidly advancing field, and much work remains to be done in developing analytical techniques, interpretive frameworks, and datasets to realize the promise of this powerful approach.

Methodology

The principal tool in metabolomics is liquid chromatography-mass spectrometry. Liquid chromatography separates compounds based on chemical properties such as polarity while mass spectrometry measures the mass and abundance of the compound. Metabolites in marine samples can be measured both inside of cells (particulate) or outside of cells (dissolved), each with its own unique challenges in sampling and analysis[23]. Particulate metabolites are captured on a filter, then concentrated for later extraction. However, these samples tend to display substantial matrix effects from salts during LC-MS analysis and therefore require complex normalization techniques to account for this measurement variability[24]. Measurements of dissolved metabolites have lagged behind particulate metabolites due to the additional challenge of extracting and concentration small polar compounds from a matrix containing inorganic salts that are roughly nine orders of magnitude more concentrated[19]. However, recent advances in extraction and derivatization offer new analytical windows into this pool[25,26]. In Chapter 2, I developed a new extraction approach, cation-exchange solid phase extraction (CX-SPE) that enabled the extraction and measurement of many dissolved metabolites that had not been measured before[25].

Conceptual Frameworks

As metabolomics methods improve and the first generation of large-scale environmental metabolomics datasets comes online, the community must now grapple with how to interpret the complex data of metabolomics. To help glean meaning from these datasets, a variety of conceptual frameworks are used or have been proposed, each with its own set of advantages and disadvantages.

For the study of individual organisms, metabolomics can be used to directly assess changes to cellular metabolism in response to a range of conditions. By directly measuring changes in the concentrations of key substrates, intermediates, and cofactors of cellular metabolisms, alterations to metabolic pathways can be inferred. This approach has led to insights into metabolic changes that occur in a range of phytoplankton species in response to nutrient and vitamin limitation[12,13,27,28]. In chapter 3, I apply this framework to study how viral infection alters the metabolism of the marine cyanobacterium *Prochlorococcus*. In these laboratory cultures dominated by a single organism, we could ascribe changes in the relative cellular abundances of metabolites to a single process (in this case viral infection), allowing us to identify signals of carbon and energy limitation and oxidative stress. However, this approach is almost impossible to apply in natural communities as environmental samples contain billions of cells of diverse lineages, all using similar compounds and pathways, imprinted on top of each other into a whole ecosystem measurement.

Other conceptual approaches have adapted concepts from microbiology and ecology to classify metabolites into interpretable groups. Moran et al. classified metabolites based on their function in microbial systems as substrates (fuel for microbial growth), facilitators (reusable enablers of biochemical reactions), and signals (communication chemicals)[19]. Durham et al. borrow terms from macroecology to designate metabolites as dominant (highly abundant), keystone (important with low functional redundancy), or indicator (used to assess ecosystem function or stress)[29]. Many compounds may bridge these divides such as osmolytes, a set of compounds produced by organism primarily in response to osmotic stress. Osmolytes are highly abundant sources of carbon, nitrogen, sulfur and energy to marine communities (substrate, dominant), preserve protein structure facilitating biochemical reactions (facilitator), and can mediate organism interactions by driving microbial chemotaxis or growth strategy shifts (signal)[10,11,30–35]. In chapter 4, I employ these frameworks to test the hypothesis that two abundant osmolytes, glycine betaine and homarine, are important components of marine biogeochemical cycles. I found that the fluxes of these compounds were equivalent to up to 1.2% of net community production, suggesting they can be classified as both dominant and substrate metabolites. While these classifications are incomplete, they provide a common language and intellectual framework for studying these compounds.

A third approach is to study metabolites as functional groups. Metabolite functional groups are a set of compounds with a shared biochemical function that may be differentially produced by different organisms or the same organism under different conditions. Therefore, changes in the relative or absolute composition of a metabolite functional group can provide insights into community function. A classic example of this are membrane lipids, where cyanobacteria and eukaryotes substitute sulfur and nitrogen containing lipids for phosphorus containing lipids under phosphate limitation[36,37]. In Chapter 5, I apply the metabolite functional group framework to osmolytes. I measure community osmolomes, or the sum of all osmolytes in a system, across the diverse biomes in the North and Equatorial Pacific Ocean and the Salish Sea. Individual organisms produce different suites of osmolytes to adapt to nutrient stress and salinity[11,12,32,38]. However, the high functional redundancy in osmolyte production across diverse taxonomic groups results in largely stable osmolomes across our transects with nutrient supply driven changes in community composition primarily responsible for regional differences.

1.3 Emergent Properties and Interactions in Microbial Systems

Emergent properties are defined as aspects of a system that cannot be linearly predicted from the sum of its component pieces, or put another way, when communities are more than the sum of their parts[39–42]. Emergent properties include phenomenon such as self-organization, community scale elemental cycling, and ecosystem resilience[39,40,43]. Life itself can be considered an emergent property from the combination of chemical structures and processes that compose a cell[42].

In marine microbial systems, emergent phenomena often arise from the interaction of microbes with one another. The interactions range from predatory to symbiotic[7,14,44]. Phytoplankton and bacteria will often form mutualisms with the phytoplankton providing fixed carbon substrates to the bacteria and the bacteria in turn providing essential vitamins and nitrogen sources to the phytoplankton[14,20,45]. These interactions can increase the growth of both the phytoplankton and the bacteria. In contrast, predatory interactions such as viral infection and zooplankton grazing both cause mortality for phytoplankton, and in doing so alter microbial populations and biogeochemical cycling[46–50]. Interactions can also occur at the chemical level where enzymes or reactions can be inhibited or facilitated by the presence of certain molecules, altering the chemical composition and metabolism of the system. Collectively, these biological and chemical interactions combine to structure the ecosystems and community level metabolic processes we observe in the ocean.

The central argument of this dissertation is that the pools of metabolites we observe (collectively referred to as the metabolome) are an emergent property of the interactions that occur within these systems. In chapter 3, I show that viral infection dramatically alters phytoplankton metabolism, changing the abundance of approximately 25% of phytoplankton metabolites (a biological interaction). In chapter 4, I demonstrate that the dissolved concentrations and uptake rates of metabolites in natural microbial communities are highly dependent on the other compounds present due to a process known as uptake competition (an interaction among metabolites). Finally, in chapter 5, I show that the composition of osmolomes of marine microbial communities from diverse environments is remarkably consistent. While individual organisms have highly distinct osmolyte profiles, a stable community osmolome composition emerges from the collective pressures and interactions that shape mixed communities. Taken together, these results highlight the importance of biological interactions in oceanic biogeochemical cycles and provide some hope that community metabolic function can be preserved in the face of future ecosystem change.

1.4 Metabolomics in the search for life on other worlds

Metabolomics is also a promising tool in the search for life on other worlds[51]. Mass spectrometry-based metabolite analyses are key features of both in-situ analysis and sample return missions with metabolites such as amino acids and nucleic acids identified as key targets to measure for presence, relative abundance patterns, isotope ratios, and chirality[51,52]. Additionally, new work on agnostic biosignatures suggests that metabolomics analyses can provide measures of chemical complexity of metabolites that can serve as an indicator of life[53]. However, in order to use mass spectrometry to search for life in the oceans of other worlds, we must first understand what life looks like through a mass spectrometer here on Earth.

Metabolomics analysis of microbial communities in Earth's oceans will allow us to identify chemical analytical targets and distributional patterns that are strong indicators of biosynthesis. Many metabolites also have abiotic synthesis pathways and have been documented in extraterrestrial samples such as asteroids and Martian meteorites[54,55]. Therefore, our ability to use small, polar metabolites as biomarkers requires a thorough understanding of their biological controls and relationship with other environmental parameters to interpret their presence and distributions in future life detection missions[56]. Beyond life detection, metabolomics can provide valuable insights into the function of microbial communities and their adaptations to the environment. Metabolomics is therefore a critical tool in Astrobiologically relevant microbiology research to understand how organisms adapt to low energy, oligotrophic, and extreme systems, a key goal of NASA's Astrobiology strategy[56]. In my dissertation I develop new methods for measuring metabolites in the marine environment (Chapter 2), I apply these methods to characterize microbial metabolite pools in Earth's oceans to identify new biosignature targets

(Chapters 4 and 5), and I constrain the variability and influences on these metabolites across a wide range of physical, chemical, and biological gradients (Chapters 3, 4, and 5).

Chapter 2: Quantification of Dissolved Metabolites in Environmental Samples through Cation-Exchange Solid Phase Extraction (CX-SPE) paired with Liquid Chromatography-Mass Spectrometry

2.1 Abstract[1]:

Small, biologically produced, organic molecules called metabolites play key roles in microbial systems where they directly mediate exchanges of nutrients, energy, and information. However, the study of dissolved polar metabolites in seawater and other environmental matrices has been hampered by analytical challenges including high inorganic ion concentrations, low analyte concentrations, and high chemical diversity. Here we show that a cation-exchange solid phase extraction (CX-SPE) sample preparation approach separates positively charged and zwitterionic metabolites from seawater and freshwater samples, allowing their analysis by liquid chromatography-mass spectrometry (LC-MS). We successfully extracted 69 known compounds from an in-house compound collection and evaluated the performance of the method by establishing extraction efficiencies and limits of detection (pM to low nM range) for these compounds. CX-SPE extracted a range of compounds including amino acids and compatible solutes, resulted in very low matrix effects, and performed robustly across large variations in salinity and dissolved organic matter (DOM) concentration. We compared CX-SPE to an established solid phase extraction procedure (PPL-SPE) and demonstrate that these two methods extract fundamentally different fractions of the dissolved metabolite pool with CX-SPE extracting compounds that are on average smaller and more polar. We use CX-SPE to analyze four environmental samples from distinct aquatic biomes, producing some of the first CX-SPE dissolved metabolomes. Quantified compounds ranged in concentration from 0.0093 nM to 49

---

[1] This chapter was published as: Sacks, J.S., Heal, K.R., Boysen, A.K., Carlson, L.T. and Ingalls, A.E., 2022. Quantification of dissolved metabolites in environmental samples through cation-exchange solid-phase extraction paired with liquid chromatography–mass spectrometry. *Limnology and Oceanography: Methods*, *20*(11): 683-700. 2022. DOI: 10.1002/lom3.10513

nM and were composed primarily of amino acids (0.15 – 16 nM) and compatible solutes such as TMAO (0.89 – 49 nM) and glycine betaine (2.8 – 5.2 nM).

2.2 Introduction:

Metabolomics is commonly used to measure intracellular pools of small molecules where it is a powerful tool in characterizing cellular phenotype and biochemical pathways[22]. In aquatic environments, dissolved metabolites (operationally defined as passing through a 0.2 μm filter) mediate microbial nutrient and energy exchanges and can serve as critical controls on microbial community composition and activity[12,14,19,57]. However, our ability to measure extracellular metabolomes in environmental samples is hampered by low metabolite concentrations, high inorganic ion concentrations, and the complex matrix of dissolved organic matter (DOM)[19]. This problem is particularly acute in seawater since salt (~0.6 M) and background DOM (~40-80 μM) concentrations exceeds those of individual metabolites (low pM to low nM) by roughly 9 and 4 orders of magnitude, respectively[19,58]. In liquid chromatography mass spectrometry (LC-MS) analyses, salts and background DOM can cause severe ion suppression, preventing the detection and quantification of compounds of interest[24,59]. Furthermore, metabolites are extremely chemically diverse, spanning a range of formal charges (positive to negative), polarities (nonpolar to polar), functional classifications, and masses (10s – 1000s of Daltons). This diversity makes measuring all metabolites with a singular analytical approach nearly impossible as the techniques used to separate and concentrate analytes from the matrix rely on differences in charge, polarity, size, and functional groups. Therefore, specific extraction techniques are needed to target the polar, low-molecular weight fraction of metabolites and remove them from the salts and background DOM present in aquatic samples to enable their analysis via LC-MS.

In recognition of this need, a variety of methods targeting low molecular weight, polar compounds dissolved in seawater are now available[26,59–62], but critical gaps remain in the analytical windows of these techniques[19]. The commercially available modified styrene divinyl benzene polymer solid phase extraction Bond Elute PPL column (PPL-SPE) (Agilent Technologies, Santa Clara, CA USA) is widely used in low molecular weight DOM analysis, retains 43–62% of DOM from marine samples, and has been used widely for dissolved metabolite studies[34,59,63–66]. The PPL extracted DOM, while containing some metabolites, is comprised primarily of non-polar molecules that are representative of the large, recalcitrant ocean DOM pool that can cause substantial matrix effects, limiting the ionization of polar metabolites of interest during electrospray ionization mass spectrometry[59,63]. Benzyl chloride (BC) derivatization of primary amine, secondary amine, and alcohol containing functional groups before extraction via PPL-SPE captures a range of metabolites not extracted using just PPL-SPE[26]. However, polar compounds without those functional groups, such as glycine betaine, are not derivatized with BC, cannot be extracted with PPL[59], and often co-elute with other compounds with a reversed phase LC-MS run[10,32]. In addition to the metabolites of interest, the BC-derivatization/PPL approach also has the potential to extract a large number of unidentified compounds present in DOM including hydrophobic DOM that binds to the PPL column without derivatization as well as unknown molecules that bind to the PPL column post derivatization. These molecules may introduce matrix effects, particularly in high DOM samples such as those from coastal environments or laboratory cultures. Derivatization paired with gas chromatography[60] and with liquid/liquid separations[61] as well as salt tolerant reversed phase columns[62] have also been used, but without sufficient sensitivities to measure compounds at the low nanomolar to picomolar concentrations expected in seawater.

A critical gap in understanding dissolved organic compounds is in the ability to measure polar zwitterionic compounds containing positively charged quaternary amine or sulfonium groups including dimethylsulfoniopropionate (DMSP), gonyol, dimethylsulfonioacetate (DMS-Ac), glycine betaine, homarine, trigonelline, and trimethylamine N-oxide (TMAO). These compounds have been identified as quantitatively important components of marine particulate metabolite pools[10,31,32], are often accumulated as osmolytes to help microbes respond to osmotic stress[11,13,67,68], mediate microbial interactions such as phytoplankton-bacteria symbioses[34], and play important roles in the cycling of carbon and other elements[11,69,70]. Many polar charged and zwitterionic compounds are largely unextractable with existing dissolved metabolomics extraction approaches and invisible to some LC-MS analytical methods, requiring the development of new metabolomics methods to study the distribution and roles of these compounds in marine systems alongside other key dissolved metabolite groups such as amino acids.

Cation-exchange solid phase extraction (CX-SPE) has the potential to broaden our view of the polar components in DOM by targeting compounds with a permanent positive charge or that can be protonated under acidic conditions. CX-SPE has been successfully used for the extraction of small organoarsenic compounds in seawater[71], therapeutic and illicit drugs from aqueous sewage samples[72], and drugs and metabolites from urine samples[73]. In CX-SPE, protonated analytes and cationic salts in an acidified sample bind to negatively charged binding sites on the stationary phase resin through electrostatic attractions. The analytes are then eluted through the addition of a basic solution which deprotonates the analytes or alters their interactions with the stationary phase while the salts remain attached to the resin, separating the analytes from the salts. The eluted analytes can then be concentrated, separated, and analyzed using LC-MS. The stationary phase is regenerated by flushing the columns with a concentrated strong acid solution to displace the salts on the column.

Here we present a CX-SPE approach for extraction of dissolved metabolites from seawater and pair it with established liquid chromatography-mass spectrometry methods optimized for the separation and quantification of small, polar molecules[24]. We identify compounds that are extracted using this approach, characterize their extraction efficiencies, and examine the robustness of this new method across a range of salinities and dissolved organic matter concentrations. We also compare this new approach to the common PPL-SPE approach for extracting DOM from seawater. Finally, we demonstrate the utility of the method by quantifying dissolved metabolite pools in a variety of environmental samples.

2.3 Methods:

Materials:

Most metabolite standards were purchased from Sigma Aldrich, Santa Cruz Biotechnology, Toronto Research Chemicals, Cambridge Chemicals, Spectrum Chemical, and Cambridge Isotope Laboratories. The sulfonate compound 2,3-dihydroxypropane-1-sulfonate (DHPS) was provided by A. Bourdon and S. Champagna (University of Tennessee, Knoxville, USA). Dimethylsulfonioacetate (DMS-Ac) and Gonyol were provided by Dr. G. Pohnert, (Friedrich Schiller University Jena, Germany). N-acetyltaurine was provided by A. Cook and K. Denger

(University of Konstanz, Germany). 2-(3,5-Dichlorophenylcarbamoyl)-1,2-dimethylcyclopropane-1-carboxylic acid (cinnamoyl-HSL) was provided by C. Harwood (University of Washington, Seattle, USA). TMAP (β-alanine betaine) was synthesized in house using the procedure detailed in[67]. Full supplier information is provided in Table 2.3

Sample Collection:

Natural water samples for method assessment and environmental analysis were collected from four different locations, Station ALOHA in the North Pacific Subtropical Gyre (ALOHA, 22º 45' N, 158º W), the North Pacific Transition Zone (NP, 41º 24' N, 158º W), Puget Sound (PS, 47º 41' N, 122º 25' W), and Lake Washington (LW, 47º 38' N, 122º 18' W). ALOHA and NP samples were collected at a depth of 15 m on the research cruise KOK1606 aboard the R.V. Ka'imikai-O-Kanoloa between 20 April and 2 May 2016 from Niskin bottles. The PS sample was collected on 7 May 2017 at a depth of 8 m from aboard the R.V. Rachel Carson using Niskin bottles. The ALOHA, PS, and NP samples were collected by filtering seawater through 142-mm, 0.2 µm polytetrafluoroethylene (PTFE) filters (Omnipore Membrane Filters, Merck Millipore Ltd., Tullagreen, Ireland) using a peristaltic pump into 2 L polycarbonate bottles and stored at -20˚C until analysis. The LW sample was collected at a depth of 1 m from the University of Washington Waterfront Activities Center Dock on 20 June 2020. It was filtered through a 47-mm 0.2 µm PTFE filter (Omnipore Membrane Filters, Merck Millipore Ltd., Tullagreen, Ireland) using a glass vacuum filtration setup before being stored in 45 mL polypropylene falcon tubes at -20˚C until analysis. Samples were stored at -20˚C for a duration of several weeks to 4 years. We cannot rule out the possibility that storage influenced the measured metabolite concentrations in natural samples. The main objectives of our method development and assessment were to determine the extraction efficiency and detection limits of metabolites within samples with a variety of salt and dissolved organic matter matrices. We expect the sample matrices (inorganic ions and background DOM) to be unaffected by storage time. GO-SHIP protocols allow for multi-year storage of samples prior to analysis of bulk DOM, suggesting that while there could be some structural alteration, there is no detectable net loss of the DOM matrix over multiple years of storage[74].To remove trace organic contaminants, all plastic used in collection and sample storage were soaked in a 10% HCl acid bath for 24 hours and rinsed three times with MilliQ H_2O and all glass was combusted to 450 ˚C for 5 hours before use.

Sample Preparation:

Cation-Exchange Solid Phase Extraction:

The cation exchange solid phase extraction (CX-SPE) approach to measuring trace organic molecules requires that the water used in all steps of the process (column equilibration, extraction, elution, regeneration, and blank preparation) is free of analytes. In addition to the purification provided by a Milli-Q system (Sigma-Aldrich, Vienna, Austria), further purification was required to ensure the complete removal of all analytes of interest (glycine betaine in particular) (Figure 2.6). To prepare analyte-free water, Milli-Q water was rinsed through CX-SPE columns before being used, producing what we term “Cation Exchange Clean Water” or “CXC H_2O”. The washing columns were set up and equilibrated as follows. First, 35 g of strong cation-exchange resin (Dowex 50WX8; H+ form, 100-200 mesh, Sigma-Aldrich, Vienna, Austria) was placed in a glass chromatography column with a fritted disk and a PTFE stopcock (1 in ID x 12 in, Chemglass Life Sciences, Vineland, NJ USA). The resin was then equilibrated for use by washing with 50 mL of H_2O, 100 mL of 1M NH_3, 50 mL of H_2O, 100 mL of 3M HNO_3, and 50 mL of H_2O using gravity flow. CXC H_2O was then generated by flowing Milli-Q H_2O through the washing columns and storing in combusted 2 L glass bottles with polypropylene caps and kept in the dark until use. CXC H_2O was then used to prepare CXC 1 M NH_3 and 3 M HNO_3 solutions. The CXC H_2O and solutions were then used to prepare a second set of CXC CX-SPE columns for sample analysis in the same manner as the washing columns. The CX-SPE columns were prepared in-house rather than using pre-packed columns to enable thorough cleaning for trace organic compounds, the reuse of columns, and the flexibility to control column design to meet our needs.

Sample processing (outlined in Figure 2.1) was performed by acidifying 40 mL of sample to pH 2 using CXC 3 M HNO_3. This sample volume was chosen for convenience during sample collection, storage, and transport, efficient thawing during analysis, and to minimize the chemical waste generated during the extraction procedure. The sample was added to the CXC columns along with a 20 µL spike of our "Extraction" isotopically labeled internal standards dissolved in HPLC water and allowed to stand for 5 minutes (internal standard information detailed in Table 2.4). Our internal standard concentrations were chosen to be within or slightly above the range of expected natural values for dissolved metabolites (Table 2.1). The sample was then drained from the column and the column was rinsed with 50 mL CXC H_2O by trickling along the side with a combusted glass Pasteur pipette so as not to disturb the resin. NH_3 (1 M) was added to the column (once again trickled down the side using a glass pipette to avoid disturbing the resin) and eluted in 10 mL fractions into combusted 20 mL glass vials. The pH of each fraction was measured by dabbing a small amount of sample onto a pH strip (Panpeha pH indicator strips, Sigma-Aldrich) using a glass Pasteur pipette. pH strips were chosen over pH probes to minimize cross-contamination among samples. The pH of the fractions was measured until the alkaline front was reached (pH change from 2–4 to 9–11) and the alkaline front fraction, the fraction before, and the two fractions after were collected and combined (40 mL total collected of approximately 150 mL total NH_3 used). These fractions were selected for as they contained, on average, >90% of the total recoverable signal for compounds in our standards library from the CX-SPE extraction process (Supplemental Text, Table 2.5, Figure 2.7). The combined fractions were dried down in the dark in a TurboVap Evaporator (Biotage) under N_2 gas (flow rate of 1.8 L/min) in a water bath at 32 °C. The dried fractions were stored at -20 ˚C until reconstitution. The column was regenerated through sequential washing with 50 mL 1 M CXC NH_3, 50 mL CXC H_2O, 100 mL 3 M CXC HNO_3, and 50 mL of CXC H_2O. The Teflon coated stopcock pieces were removed and solvent rinsed sequentially with 10% formic acid, methanol, and dichloromethane to remove any residual trace organics. When in storage, columns were filled with 0.01 M CXC HNO_3.

PPL Extraction:

PPL-SPE extraction was performed following established literature protocols[59,63]. Agilent Bond Elut PPL cartridges (1 g bed mass, 6 mL volume) were rinsed with 1 cartridge volume of 0.01 M HCl. Forty mL of sample was acidified to pH 2 with 3 M HCl, spiked with 20 µL "Extraction" internal standards, and pushed through the column using a peristaltic pump (Flow rate: 10 mL/min) followed by 2 cartridge volumes 0.01 M HCl. Finally, the sample was eluted with 1 cartridge volume of methanol, dried down under N_2 gas, and stored at -20 ˚C until reconstitution.

Reconstitution:

All samples were reconstituted in 380 µL of HPLC grade H_2O and 20 µL of our “Injection” isotopically labeled internal standards mix dissolved in HPLC grade H_2O for a total volume of 400 µL H_2O (Table 2.4). As a standard precaution to prevent clogging of the liquid chromatography or mass spectrometry systems, each sample was then syringe filtered through a 13 mm syringe filter with 0.22 µm PTFE membranes to remove any potential particulates, and stored at -80 ˚C until analysis. There were no visible particulates in the vials after reconstitution of dry residues. The removal of salts during the CX-SPE extraction was verified by measuring the salinity of reconstituted samples of ALOHA seawater which had a starting salinity of 36 PPT and a post-reconstitution salinity of 0 PPT (Portable Refractometer, Agriculture Solutions, Kingfield, ME, USA). A similar CX-SPE approach using a smaller resin bed volume for an equal sample volume also found that <0.5% of total dissolved solids remained following solid phase extraction[71].

Chromatography and Mass Spectrometry:

In this study, two complementary types of liquid chromatography were used to enable the separation and detection of a greater range of compound polarities. Reversed-phase (RP) chromatography uses a non-polar stationary phase and polar mobile phase, resulting in more polar compounds eluting early on while less polar compounds elute later in the run, making it ideal for separating more non-polar compounds for analysis. Hydrophilic Interaction Liquid Chromatography (HILIC) is a type of normal-phase approach that uses a polar stationary phase and a less polar mobile phase. In HILIC, less polar compounds elute early in the run while more polar compounds are better retained on the column, enabling their separation. Each technique has benefits and drawbacks. HILIC chromatography tends to be more variable, with changes in retention time for compounds over the course of a run. However, RP chromatography is more vulnerable to ion suppression, particularly for more polar compounds that elute at or near the dead volume and does not achieve separation of important polar metabolites that are structural isomers (same *m/z*) such as homarine and trigonelline, two compounds that have been highlighted as components of particulate polar metabolite pools[24,32]. The chromatography and mass spectrometry methods detailed here are adapted from Boysen et al. 2018 with small changes in the instrument scan range and the HILIC column solvent timings.

HILIC Analysis:

HILIC chromatography was performed using a SeQuant ZIC-pHILIC column (5 µm particle size, 2.1 mm x 150 mm, from Millipore) with 10mM ammonium carbonate in 85:15 water to acetonitrile (Solvent A) and 10mM ammonium carbonate in 85:15 acetonitrile to water (Solvent B) at a flow rate of 0.15 mL/min. The column was held at 100% B for 2 minutes, ramped to 64% A over 18 minutes, ramped up to 100% A over 1 minute, held at 100% A for 7 minutes, and equilibrated at 100% B for 22 minutes (total time is 50 minutes). The column was maintained at 30°C throughout the analysis. The injection volume was 2 µL.

Reversed-Phase Analysis:

Reversed-phase chromatography was performed using a Waters Acquity UPLC HSS Cyano column (1.8 µm particle size, 2.1 mm x 100 mm) equipped with an Acquity UPLC HSS Cyano guard column (1.8 μm particle size, 2.1 mm x 5 mm) with 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B) at a flow rate of 0.4 mL/min. The column was held at 5% B for 2 minutes, ramped to 100% B over 18 minutes, held at 100% B for 2 minutes, and equilibrated at 5% B for 5 minutes (total run time is 25 minutes). The column was maintained at 35°C throughout the analysis. The injection volume was 15 μL.

Mass Spectrometry:

Targeted and untargeted mass spectrometry was performed using a Thermo Orbitrap Q-Exactive HF Mass Spectrometer (QE) using the settings detailed in Boysen et al.[24]. For HILIC analysis, a capillary temperature of 320ºC, an H-ESI spray voltage of 3.5 kV, an auxiliary gas heater temperature of 90 ºC, an S-lens RF level of 65, a sheath gas flow rate of 16 L/h, an auxiliary gas flow rate of 3 L/h, and a sweep gas flow rate of 1 L/h were used. Polarity switching was employed with a scan range of 60–900 *m/z* and a resolution of 60,000. For RP analysis, a capillary temperature of 320 ºC, an H-ESI spray voltage of 3.8 kV, an auxiliary gas heater temperature of 90 ºC, an S-lens RF level of 65, a sheath gas flow rate of 40 L/h, an auxiliary gas flow rate of 10 L/h, and a sweep gas flow rate of 1 L/h were used. A full scan method was used in positive mode with a scan range of 60–900 *m/z* and a resolution of 120,000. As in Boysen et al.[24], polarity switching was used with the HILIC analysis as many compounds in our targeted compound lists ionized well in either positive or negative ionization modes, allowing us to successfully measure these compounds. However, for RP analysis, most compounds only ionized well in positive ionization mode so a single ionization mode was chosen to allow for higher resolution data collection.

Data Processing and Quality Control:

Targeted:

MS data from the QE were converted to .mzML files using MSConvert[75]. All peaks of targeted compounds and internal standards were manually integrated using Skyline, with peak identification performed by comparison to authentic standards in matrix for both retention time and peak shape[76]. Quality control was then performed to remove low quality peaks from the data by requiring that peaks met requirements for minimum area (40,000 for HILIC, 5,000 for RP) and exact mass (<6 ppm). Best-matched internal standard normalization (B-MIS) was applied to account for changes in instrument response throughout the run[24]. All data analysis and visualization was performed in the R statistical environment (version 4.2.0) using R Studio (version 2022.02.0) with the the "readr," "dyplr," "stringr," "ggplot2," and "RaMS" packages.

Untargeted:

To expand the analytical window and further compare CXC and PPL extraction approaches beyond our standards library, an untargeted metabolomics approach was employed. Untargeted metabolomics decreases the analytical bias inherent in targeted metabolomics approaches by allowing for the comparison of a greater number of mass features within each sample. ALOHA and PS samples prepared using both CX-SPE and PPL-SPE and analyzed using both HILIC and RP chromatography were selected for this comparison. Mass features were then detected and integrated using MSDial (see Table 2.6 for parameters) separately for HILIC positive, HILIC negative, and RP positive modes, normalized using B-MIS normalization and filtered using the quality control filters detailed below to curate a final list of high-quality mass features[77]. All samples were aligned to pooled samples made up of both the ALOHA and PS samples analyzed with both the CX-SPE and PPL-SPE enabling the detection of the same mass features across samples and methods. Mass features were required to have a signal to noise ratio of 5, a minimum normalized peak area of 5000, a relative standard deviation within the pooled samples of < 0.3, and a peak area at least 3 times greater than the average value for that mass feature in the corresponding methodological blank. HILIC mass features with retention times less than 2 minutes and greater than 20 minutes and RP mass features with retention times less than 0.5 minutes and greater than 17.8 minutes were discarded as these regions demonstrate high levels of ion suppression[24]. These parameters were applied to each analytical triplicate independently and mass features not passing all quality control parameters in each individual analytical triplicate were discarded.

2.4 Assessment:

Identification of Extracted Compounds, Extraction Efficiencies, and Linearity:

We identify compounds extracted by our CX-SPE method by measuring extraction efficiencies (EEs), relative standard deviations (RSD_{EE}) of the extraction efficiencies, and the linearity of method and instrument response to standard spikes at different environmentally relevant concentrations (R^2) of our in-house library of authentic standards. Compounds included in the compound library are molecules involved in fundamental biological processes (amino acids, intermediates in central carbon metabolism, nucleobases), signaling, cross feeding, or bioactivity (e.g., B vitamins, homoserine lactones)[78,79]. Compounds involved in microbial adaptation to salt or other stressors (ex. compatibles solutes such as glycine betaine or DMSP) that enable organisms to handle osmotic stress and have been documented to have high concentrations in marine phytoplankton and bacteria[11] are also a part of our compound library. While not exhaustive, the compound library employed in this study includes many of the most abundant compounds documented in marine particulate community metabolomes[10,31,32] as are compounds that are suggested by previous work to be actively cycled and potentially abundant in marine dissolved metabolomes[26,64,66,70,80–86]. Our in-house library consists of 179 standards separated and analyzed using HILIC chromatography (HILIC Standards) and 70 standards separated and analyzed using RP chromatography (RP Standards) (Table 2.3).

The extraction efficiency of a compound is the percentage of a compound recovered at the end of sample preparation compared to the initial amount present in the sample. Using HILIC chromatography, leucine and isoleucine do not achieve chromatographic separation and therefore are integrated together and referred to collectively as (iso)leucine. The median spike concentration for HILIC Standards was 25 nM and the median spike concentration for RP Standards was 2.5 nM (spike values for each compound are presented in Table 2.3). For method evaluation, standard spike concentrations were required to be at least as concentrated as the most abundant environmental concentrations of common metabolites (~1-10 nM, Table 2.1) and high enough such that a high-quality signal on the instrument was achieved after reconstitution, as determined previously for this combination of chromatography and instrument[24].

To determine extraction efficiencies, 3 sets of samples were run on the columns in triplicate: a “spike-before” sample where the standard spike was added to the natural seawater or lake water before processing the sample, a “spike-after” sample where the standard spike was added during reconstitution, and a “no-spike” sample to which no standard spike was added.

Standards were dissolved in 1 mL HPLC H_2O with a concentration of each compound of approximately 1000 nM for most HILIC standards and 100 nM for most RP standards. These standard spikes result in the final concentrations reported in Table 2.3 when combined with 40 mLs of sample. We are not aware of any potential interferences in our standards mixes due to cross reactions. Extraction efficiencies were then determined using the following equation.

$$Extraction\ Efficiency\ (EE) = \frac{Area_{Spike\ Before} - Area_{No\ Spike}}{Area_{Spike\ After} - Area_{No\ Spike}} \times\ 100\% \qquad (1)$$

Subtracting the "No Spike" sample from both the "Spike Before" and "Spike After" samples removes background noise as well as the signal contributed by the natural abundance of the compound from the EE calculation. EEs were determined in triplicate for all four environmental sample matrices (ALOHA, NP, PS, LW) for the HILIC standards and in ALOHA and PS sample matrices for the RP standards. An EE value was calculated separately for each triplicate in each sample and the RSD_{EE} for each compound was calculated from the standard deviation of all EE values (n = 12 for HILIC standards, n = 6 for RP standards), divided by the mean EE value, and multiplied by 100 to attain a percentage value.

The linearity of the extraction efficiency of the compounds in ALOHA seawater was evaluated by spiking in our standards mix at 100%, 50%, or 10% of our original spike concentration (ex. 25 nM, 12.5 nM, 2.5 nM) in triplicate to build a calibration curve for each compound. The goal of this test was to determine if there was any variability in extraction efficiency in response to changes in concentrations. The dynamic ranges of these compounds on this instrument were established in Boysen et al. 2018. The R^2 value was calculated using a linear model to assess the linearity of the relationship between spike concentration and peak area. A strongly linear relationship between these two values indicates that EE is not changing in response to changes in analyte concentration. This range of concentrations was chosen as the lowest values still allowed for a spike signal to be detected despite the natural abundance of many compounds. We did not test higher concentrations as we did not expect many environmental metabolite concentrations to be above 25 nM.

Compounds were considered successfully extracted by CX-SPE when they met three extraction quality control thresholds indicating that the method could be used to generate reproducible metabolomics data in environmental samples. Compounds were required to have EEs greater than 1% and below 150%, RSD_{EE} values of less than 0.5, and R^2 values greater than 0.7. These values were selected to remove compounds that were not extracted through CX-SPE (EEs < 1%), demonstrated very high levels of variability in calculated extraction efficiencies (RSD_{EE} values > 0.5 or $R^2 < 0.7$), or demonstrated increases in concentration potentially related to the breakdown of other standards into these molecules (ex. acetylcarnitine into carnitine) during the CX-SPE procedure (EE > 150%). However, we also attempted to avoid over penalizing compounds that may have suffered from higher or lower levels of ion-suppression in the “spike after” or “spike before” samples due to the unnaturally high total analyte concentrations in these samples introduced by the standard spike. The final selection of compounds meeting the extraction quality control thresholds were designated as “reproducibly extracted” compounds and were used for further analysis to compare method performance across different sample matrices. A total of 69 targeted compounds (Table 2.2) were deemed reproducibly extracted using CX-SPE after passing quality control thresholds for EE, RSD_{EE}, and R^2 values. 70 of the compounds were HILIC standards (6 negative mode, 58 positive mode) while 5 were RP standards. For compounds passing quality control thresholds, EEs range from 2.24% for glycerophosphocholine to 131% for L-glutamine (Table 2.2, Figure 2.2). RSD_{EE} values ranged from 12.2% for nicotinic acid acid to 47.8% for glycerophosphocholine. R^2 values ranged from 0.731 for hordenine to 0.996 for betonicine. Example chromatograms of several compounds in ALOHA seawater at natural abundance and with a range of standard spike concentrations are shown in Figure 2.3.

Method Performance Across Environmental Samples:

To assess the robustness of the method to differences in environmental sample matrix, the EEs and response factors (RFs) of the reproducibly extracted compounds in the four different sample types were compared. Response factors are the amount of signal detected on the instrument relative to concentration of that compound present in the sample and can vary over orders of magnitude in ESI-MS. Four samples (ALOHA, NP, PS, LW) representing a range of DOM concentrations (oligotrophic (low DOM)–eutrophic (high DOM)) and salinities (freshwater to open ocean seawater) were chosen. The EEs and RFs of the HILIC standards were compared for all 4 environmental samples in triplicate while the EEs and RFs of the RP standards were compared for just ALOHA and Puget Sound samples in triplicate (Table 2.7). RF values were calculated using equation 2.

$$RF = \frac{Area_{Spike\ After} - Area_{No\ Spike}}{Spike\ Concentration} \qquad (2)$$

The EE and RF values for each sample matrix were then compared by one-way ANOVA and the p-values were then adjusted to control the false discovery rate (FDR) through the Benjamini-Hochberg Procedure[87] (Table 2.7). Differences in EE and RF were considered significant if the FDR adjusted p-value was below 0.05. No compounds displayed significant differences in EE, suggesting that CX-SPE extraction efficiencies are robust to changes in salinity and DOM concentration. Only two compounds, adenosine monophosphate and L-glutamic acid, had significant differences in RF values across environmental matrices, suggesting that CX-SPE helps minimize differences in instrument response caused by sample variation for the majority of compounds.

Matrix Effects

Matrix effects in LC-MS refer to the impact of the sample matrix (solvent, salts, and coeluting organic compounds present in the solution) on the ionization efficiency of target analytes. Ionization efficiencies are affected by the presence of matrix components that influence the extent of ionization of target analytes in the electrospray source. One goal of SPE is to reduce matrix effects by separating analytes from the sample matrix. To assess the ability of CX-SPE to remove matrix effects, a response factor ratio (RF ratio) for each compound was calculated. RF ratios equal to 1 are indicative of no matrix effects, those below 1 indicate ion suppression caused by the matrix, and those above 1 indicate ionization enhancement. Our HILIC and RP standard mixes were spiked at equal concentrations into a pooled sample matrix of our 4 environmental samples (matrix) and into HPLC water (water). Background concentrations of the analytes in the matrix were assessed by spiking CXC-H_2O water into the matrix at an equal dilution factor to that of the authentic standards. Response factors were calculated for each compound in the pooled matrix and in water and the ratio of these two response factors was determined.

$$RF\ ratio = \frac{RF_{matrix}}{RF_{water}} \tag{3}$$

The RF ratio of 68 of the 69 successfully extracted compounds was between 0.89 and 1.16 with one RP standard, nicotinic acid, having a much higher RF ratio (1.56) (Figure 2.8). The RF ratios determined here are much closer to 1 than for particulate metabolite analyses using an identical LC-MS analysis where many compounds exhibited substantial ion suppression (RF ratio < 0.5), indicating that CX-SPE excels at removing matrix effects, particularly for compounds analyzed with HILIC chromatography[24].

Limits of Detection:

We define the limit of detection (LOD) as the concentration at which the chance of a false positive is 5%[88]. We determined the LODs for each compound by calculating the 95% confidence interval of signal in the CXC-H_2O methodological blanks using the mean and standard deviation of the integrated peak area of the *m/z* and retention time window of each compound. LODs can be difficult to assess for ESI-MS measurements because of the difficulty in attaining a matrix matched blank while also being analyte free. However, CX-SPE demonstrated low matrix-effects (Figure 2.8), suggesting that CXC-H_2O serves as an appropriate sample for determining the "true" blank value of the compound [71]. Furthermore, we observed no indication of significant changes in EE or RF under different salinities or DOM concentrations for most metabolites, suggesting that these factors do not impact CX-SPE analysis. CXC-H_2O blanks were run throughout the analysis of the environmental samples for a total of 15 blanks. LODs were calculated for each compound using the following equation where B is the average value of the blank, S_B is the standard deviation of the blank, 1.761 is the Student's T-value for a 95% confidence interval, and n is the number of blanks. The response factor, RF ratio, and EE are used to convert the LOD from a signal value to an environmental concentration. For two compounds, cytidine and inosine, no signal was observed in the blanks, initially resulting in a calculated LOD of 0 nM which is not a true representation of method performance. To control for this, the signal blank value (B) was set at half the quality control threshold for a real peak (20,000) to enable a LOD calculation for these compounds. This adjustment resulted in LOD values for these three compounds that fell well within the range of the LOD values calculated for the other compounds.

$$LOD = \left(B + \frac{1.761(s_B)}{\sqrt{n}}\right) \times \frac{1}{Response\ Factor} \times \frac{1}{RF\ ratio} \times \frac{1}{EE/100} \tag{4}$$

Limits of detection ranged from the low-pM range for dimethylsulfoniopropionate (DMSP) and DMS-Ac (0.000103 nM and 0.000131 nM, respectively) to the low of nM for L-Serine and L-Ornithine (4.04 nM and 6.84 nM, respectively) (Table 2.2). The large variation in these values is due to the large range in instrument response factors and background ion concentrations in the blanks.

Comparison of CX-SPE and PPL-SPE

Targeted

To compare the performance and analytical windows of CX-SPE and the commonly used PPL-SPE approach, we compared the compounds extracted from our compound-library using each approach in ALOHA and PS seawater. PPL-SPE compounds were considered successfully extracted if they met the same quality control thresholds of EE ($1\% < \mathrm{EE} < 150\%$) and $\mathrm{RSD}_{\mathrm{EE}}$ ($\mathrm{RSD} < 50\%$) as in the CX-SPE analysis. PPL-SPE successfully extracted 71 compounds, 25 in the HILIC positive fraction, 18 in HILIC negative, and 28 in RP (Table 2.7). EEs ranged from 1.8% to 136.2% with $\mathrm{RSD}_{\mathrm{EE}}$ values between 4.4% and 49.1% (Figure 2.2, Table 2.7).

Between CX-SPE and PPL-SPE, a total of 116 compounds were extracted but only 24 compounds (21%) were extracted by both methods, indicating that the two methods are suited to extracting unique types of compounds (Table 2.3). We compared CX-SPE and PPL-SPE compounds using the cheminformatics classification system Chem-Ont[89] (Figure 2.9) and elemental composition using H/C and O/C ratios (Figure 2.10) to attempt to distinguish between these two groups but did not identify clear differences between the methods, potentially as a result of the limited coverage and potential bias of our finite compound library.

Untargeted:

The curated lists of untargeted mass features for CX-SPE and PPL-SPE were compared by retention time (RT) and mass to charge ratio (*m/z*) to characterize and distinguish the fractions of the untargeted, dissolved metabolome extracted by each SPE approach. It is worth noting that while each unique *m/z* and RT identifier likely represents a unique compound, some isomers may not be fully separated using our chromatography, some compounds may be duplicated in both HILIC and RP chromatography or in both positive and negative ion modes, and some compounds may result in multiple mass spectral peaks due to having multiple ionization states or adducts. However, we expect that the highly conservative quality control thresholds we applied removed the vast majority of duplicated peaks for a single compound. RT serves as a proxy for polarity where higher HILIC RTs and lower RP RTs indicate greater polarity. The *m/z* of a mass feature is primarily a measure of molecular weight, particularly with singly charged ions. Comparisons were made using the non-parametric Kruskal-Wallis approach for samples that do not meet assumptions of normality or homogeneity of variances. Additionally, the sample/method pairs (ex. PS/CX-SPE) were compared by the presence and absence of mass features determined using each type of chromatography to identify the degree of overlap between the two SPE approaches and to determine if the different samples contained unique metabolites.

The untargeted analysis of both CX-SPE and PPL-SPE resulted in a total of 244, 91, and 1458 mass features (MFs) for the HILIC Positive, HILIC Negative, and RP fractions respectively for a total of 1793 total mass features that passed all quality control thresholds (Table 2.8, Table 2.9). This set of untargeted MFs includes compounds characterized using the targeted approach. HILIC Positive and HILIC Negative features were combined (hereafter HILIC features) for further comparisons (resulting in a total of 335 HILIC features) while acknowledging that there may be some compounds that are represented in both HILIC positive and HILIC negative fractions. We also tentatively identified 31 mass features (1 HILIC negative feature, 25 HILIC positive features, and 5 RP features) as compounds within our standards library through comparisons to m/z and retention time (Table 2.9).

PPL-SPE resulted in a greater total number of MFs compared to CX-SPE for both forms of chromatography and in both samples. For the RP features, 1,036 (71.1%) were uniquely present in PPL-SPE samples, 143 (9.8%) were uniquely present in CX-SPE samples, and 279 (19.1%) were present in both PPL-SPE and CX-SPE samples. For the HILIC features, 219 (65.3%) were uniquely present in PPL-SPE samples, 87 (26%) were uniquely present in CX-SPE samples, and 29 (8.7%) were present in samples from both methods. The two samples had similar numbers of total MFs (PPL-SPE and CX-SPE combined) with the ALOHA and PS samples having 1324 and 1419 MFs respectively (Figure 2.11). ALOHA and PS HILIC MFs were not significantly different in terms of average retention time (RT) or average *m/z* for both CX-SPE (Kruskal-Wallis, $p = 0.34$; $p = 0.85$, respectively) and PPL-SPE (Kruskal-Wallis, $p = 0.46$; $p = 0.19$, respectively) and so ALOHA and PS HILIC MFs were grouped together for further comparisons of the CX-SPE and PPL-SPE methods. ALOHA and PS RP MFs were significantly different in terms of average RT and average *m/z* for both CX-SPE (Kruskal-Wallis, $p = 0.031$; $p = 0.000081$, respectively) and PPL-SPE (Kruskal-Wallis, $p = 0.0026$; $p = 0.00014$, respectively) and so the two samples were compared separately.

CX-SPE MFs were more polar than PPL-SPE MFs. CX-SPE MFs had a significantly higher average retention time in HILIC chromatography (Kruskal-Wallis, $p = <2.2\times10^{-16}$) and a significantly lower average retention time in RP chromatography in both the ALOHA and PS samples (Kruskal-Wallis, $p = 0.00033$; $p = 4.24\times10^{-15}$, respectively) (Figure 2.4). Both types of chromatography suggest that CX-SPE extracts a more polar set of compounds than PPL-SPE. CX-SPE MFs also had significantly lower average *m/z* values than PPL-SPE MFs, suggesting that CX-SPE extracts a set of compounds with a lower average molecular weight than PPL-SPE (Figure 2.4). The mean *m/z* values of CX-SPE MFs were significantly lower for both HILIC chromatography (Kruskal-Wallis, $p = <2.2\times10^{-16}$) and in RP chromatography for both ALOHA and PS samples (Kruskal-Wallis, $p = 0.036$; $p = 0.00074$, respectively). These results show that CX-SPE and PPL-SPE extract different subsets of the low molecular weight dissolved organic matter pool.

Field Application of CX-SPE for Targeted Dissolved Metabolomics

To assess the utility of CX-SPE for characterizing and quantifying dissolved metabolomes, we characterized 4 distinct environmental samples (ALOHA, PS, NP, LW) using HILIC chromatography and 2 samples using RP chromatography (ALOHA, PS). Calculations were performed using equation 4 where A_{norm} is normalized peak area.

$$Concentration = A_{norm} \times \frac{1}{Response\ Factor} \times \frac{1}{RF\ ratio} \times \frac{Vol_{reconstituted}}{Vol_{sample}} \times \frac{1}{EE/100} \tag{4}$$

Using our CX-SPE approach we quantified 21, 34, 36, and 41 dissolved metabolites in the NP, LW, PS, and ALOHA samples respectively that were above LODs in our four environmental samples, ranging in concentration from 0.0093 nM desthiobiotin in the ALOHA sample to 48.5 nM for and trimethylamine n-oxide (TMAO) in the ALOHA sample (Figure 2.5, Figure 2.12, Table 2.10). Amino acids such as glutamic acid and proline as well as other osmolytes such as the nitrogen-containing compounds glycine betaine and TMAO and the sulfoniums DMSP, DMS-Ac, and gonyol were among the most abundant compounds measured (Figure 2.5, Table 2.10). The freshwater LW sample was noticeably different than the marine samples as many of the major osmolytes that contributed significantly to the marine samples were below detection limits and LW contained higher relative percentages of many amino acids (Figure 2.5). We also characterized two environmental samples using PPL-SPE using the same quantification approach as was used with CX-SPE. While we did not attempt to calculate a true LOD, we determined a conservative "applied LOD threshold" using equation 4 and just 3 methodological blanks of CXC-H_2O analyzed with PPL-SPE (Figure 2.13, Figure 2.14, Table 2.11). The PPL-SPE approach resulted in the quantification of 20 and 21 targeted compounds above the applied LOD threshold in ALOHA and PS seawater respectively. Concentrations ranged from 0.00017 nM methylthioadenosine in the ALOHA sample to 9.31 nM tryptophan, also in the ALOHA sample. It should be noted that the targeted compound library used in this study is focused on small, polar metabolites and likely lead to more targeted compounds being successfully measured with CX-SPE compared with PPL-SPE.

2.5 Discussion:

In the ocean, the dissolved polar metabolite pool is a small but rapidly cycled pool of carbon and other essential elements through which roughly half of net primary production flows[19].The magnitude of this flux makes measuring and understanding the cycling of these highly labile molecules essential for understanding marine biogeochemical cycles and marine microbiology. The CX-SPE method presented here provides an effective way to extract small, polar, organic molecules from seawater that either have a positive charge or can be protonated by acidifying the sample. This sample preparation approach paired with a combination of HILIC and RP liquid chromatography and high-resolution mass spectrometry enables the measurement of dissolved metabolites at environmentally relevant concentrations. Key aspects that led to the success of the method presented here include the trace-organics cleaning protocols (the use of CXC H_2O for all reagents and blanks), the near complete removal of salts and background organic matter from the analytes of interest resulting in very low matrix effects, and the use of HILIC chromatography which enables the separation of very polar metabolites such as betaines (ex. glycine betaine and homarine), amino acids (ex. glutamic acid and proline), and sulfonium ions (ex. DMSP, DMS-Ac, and gonyol) that are highly-abundant components of marine particulate metabolite pools[10,31,32].

We show in both the targeted and untargeted comparisons that CX-SPE and PPL-SPE extract fundamentally different portions of dissolved metabolite pools, suggesting that each approach provides a complimentary window into marine DOM. In the targeted study, among the 116 extracted compounds, only 24 compounds were successfully extracted by both methods (Table 2.3). PPL-SPE resulted in a higher total number of mass features in both HILIC and RP LC-MS analyses which fit our expectations given that PPL columns are favored for DOM analysis, where they retain the greatest total percentage of DOM on solid phase extraction resins that have been tested[63]. PPL-SPE also resulted in a set of mass features that had a higher average molecular weight and lower polarity than CX-SPE (Figure 2.4), suggesting that these features are more representative of the much higher concentration, less-polar, refractory DOM pool[3]. Compared to PPL-SPE, CX-SPE extracted a smaller number of mass features that had a lower molecular weight and were more polar, making these features likely more representative of the dissolved polar metabolite pool rather than the background DOM pool (Figure 2.4). The targeted features uniquely extracted by CX-SPE include highly abundant known groups of marine particulate metabolites (free amino acids, betaines, sulfoniums) that are not successfully extracted using PPL-SPE[10,31,32]. CX-SPE and PPL-SPE are complimentary analyses, which, when combined, provide an analytical window into metabolites with a range of polarities.

CX-SPE compared favorably in method performance to other dissolved metabolomics approaches, with each approach enabling access to a unique set of compounds. CX-SPE LODs ranged from 0.00010 nM – 6.8 nM, which is similar to the derivatization/PPL method (0.010 – 150 nM) with both methods enabling measurements of compounds in the pM to low nM range expected of dissolved metabolites in seawater[26]. CX-SPE LODs were lower than the derivatization/GC-MS method (60–1000 nM)[60]. CX-SPE did not achieve comparable LODs for amino acids (0.025– 6.8 nM) compared with focused dissolved free amino acid (DFAA) methods (0.009–0.163 nM)[83], but CX-SPE did still enable the quantification of many DFAAs in environmental samples. The CX-SPE LODs for DFAAs and other compounds can potentially be improved in the future through the use of larger sample volumes and micro or nanoflow HPLC systems.

Different dissolved metabolomics approaches provide different analytical windows, enabling measurements of unique sets of compounds. For example, of the 69 compounds successfully extracted and measured by CX-SPE, 15 (22%) do not contain primary amine, secondary amine, or alcohol functional groups and therefore would not be derivatized by the BC reagent used in the BC derivatization/PPL method[26]. Compounds that would not be derivatized using the BC approach include many of the most abundant metabolites measured such as glycine betaine and TMAO (Figure 2.5). CX-SPE was developed in part to focus on positively charged and polar, zwitterionic metabolites and successfully extracted most positively charged compounds from our standards library (betaines, sulfonium ions). However, some positively charged compounds were not extracted reproducibly with high confidence because other compounds decomposed into them during analysis (carnitine) or poor performance on the instrument. While a good number of zwitterionic compounds were successfully extracted (amino acids), some were not, either through extraction efficiencies below threshold values or poor reproducibility (such as the amino acid glycine). CX-SPE also failed to extract a large number of other compounds within our in-house standards library including sugars such as trehalose and sucrose, some B vitamins, compounds with negative functional charges such as DHPS. These molecules often do not develop strong positive charges in acidic solutions, preventing their extraction via cation exchange; have highly variable response factors, preventing confident quantification; or have very low ionization efficiencies, preventing their detection via LC-MS at environmentally relevant concentrations. Derivatization type approaches are likely the best option for studies focusing on these compounds[26,60].

The concentrations of many of the compounds measured in environmental samples by CX-SPE analysis are within or close to the environmental concentration ranges reported by other studies of dissolved metabolites such as amino acids[83,90,91], TMAO[80], creatine[92], DMSP[81], and arsenobetaine[71] (Table 2.1). Furthermore, the concentrations of some compounds measured for the first time using CX-SPE such as glycine betaine were of roughly the same order of magnitude (low nM) as those predicted by uptake kinetics studies[70,81,93]. The overall composition of the compounds in the environmental samples fit our expectations based on observed particulate metabolite pools, with the most abundant particulate metabolites also contributing significantly to the quantifiable dissolved metabolite pool. We cannot exclude the possibility that the dissolved concentrations of compounds measured are impacted by cell breakage during filtration, resulting in metabolites moving from the particulate into the dissolved phase, impacting our dissolved phase measurements and contributing to these similarities. We were surprised by the high concentration of many metabolites at station ALOHA relative to other stations. We predicted the concentrations of dissolved metabolites would be the lowest in this oligotrophic environment. The response factors and extraction efficiencies of CX-SPE are robust to changes in environmental parameters, suggesting that these values are accurate. However, it is important to note that each sample presents a single snapshot of a dissolved metabolome that is likely constantly and rapidly remodeled by a range of dynamic uptake and release processes such as viral lysis and sloppy feeding by grazers[94–96]. More samples from a variety of locations and timepoints will be required to accurately constrain the geographic, temporal, and environmental variability of dissolved metabolomes.

The results from the analysis of these four environmental samples underscores the importance of CX-SPE as a tool for enhancing our understanding of microbial communities and the cycling of DOM in marine and aquatic environments. For example, glycine betaine and TMAO are key molecules in marine ecosystems that serve as osmolytes for marine microbes and animals[11,97], fixed C and N sources to marine microbes[11,70,98], or precursors of methane and atmospheric aerosols[99,100]. Until now, studies of these molecules have been restricted to the particulate phase[32], molecular biology techniques[101], indirect approximations through uptake kinetics studies[70,102], or a single, very targeted set of measurements[80]. CX-SPE enables some of the first measurements of these molecules, paving the way for future studies in understanding the distribution and cycling of these compounds. Finally, CX-SPE can be paired with untargeted metabolomics approaches to discover previously unidentified dissolved metabolites that may play key roles in shaping microbial systems.

2.6 Conclusion:

The CX-SPE method presented and evaluated in this study is a valuable addition to the suite of sample preparation approaches aimed at the challenging problem of measuring polar metabolites that are dissolved in complex matrices. Paired with LC-MS, this extraction technique can be utilized in a wide range of marine and freshwater environments to improve our understanding of microbial community interactions and labile DOM cycling. We compare CX-SPE to the popular PPL-SPE and demonstrate that the two approaches capture fundamentally different components of DOM, with CX-SPE excelling at extracting small, positively charged and zwitterionic molecules that often contain nitrogen or sulfur. Finally, we present the first CX-SPE dissolved environmental metabolomes, including some of the first measurements of several compounds with documented importance in aquatic microbial systems such as homarine, glycine betaine and TMAO, providing baseline measurements to facilitate future research.

2.7 Acknowledgements:

We thank Alec Meyers, Everetta Rasyid, and Leland Wood for assistance with laboratory analyses and preliminary data processing; William Kumler, Natalie Kellogg, and Regina Lionheart for feedback and assistance with data analysis and interpretation; and the science party and crew of KOK1606 and the R.V. Rachel Carson for support during sample collection. We thank G. Pohnert, C. Harwood, A. Bourdon, S. Champagna, A. Cook, and K. Denger for generously providing metabolite standards. This work was supported by grants from the Simons Foundation (LS award ID: 385428, A.E.I.; SCOPE award ID 329108, A.E.I.; award ID 598819, K.R.H.; award ID 731716, A.K.B.) and the National Science Foundation (OCE-2125886 to A.E.I.).

2.8 Figures:

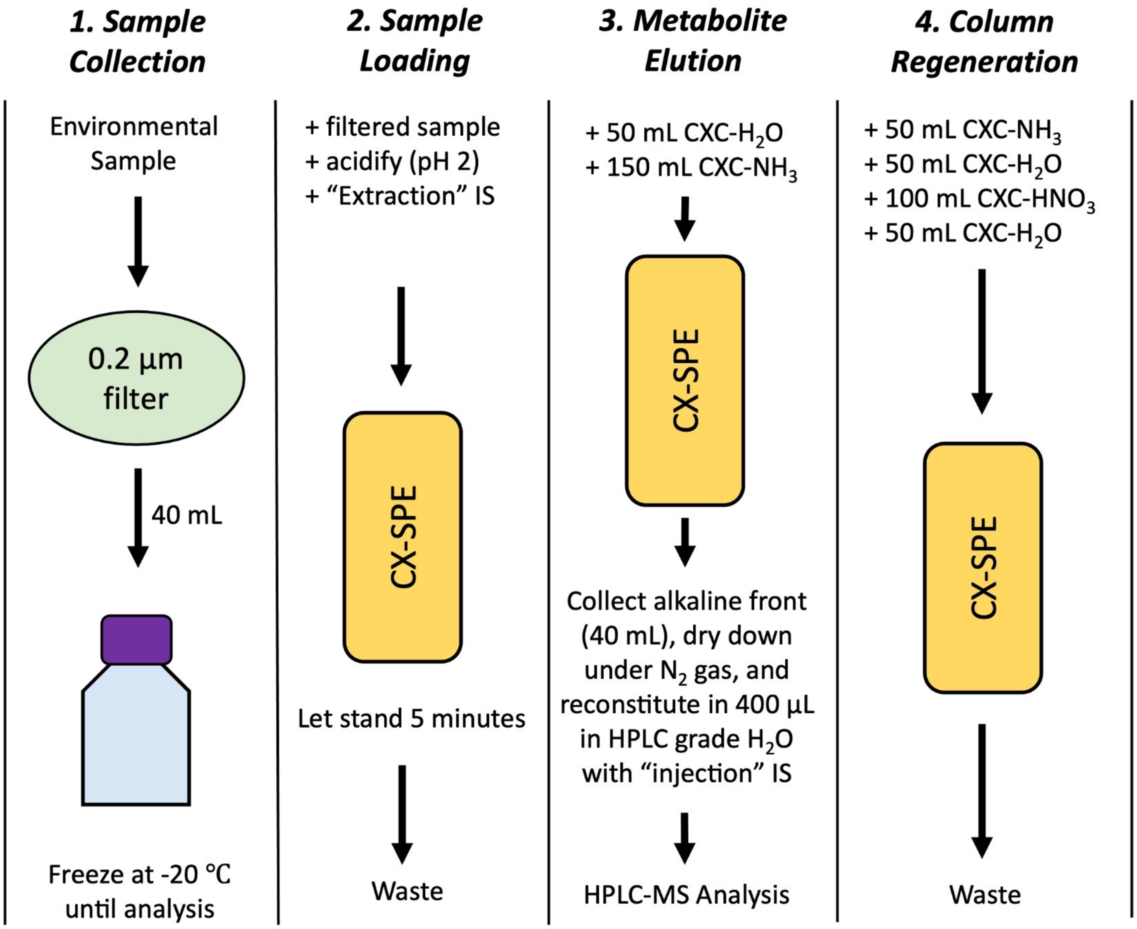

Figure 2.1: Visual outline of the CX-SPE workflow for dissolved metabolomics. IS is internal standards.

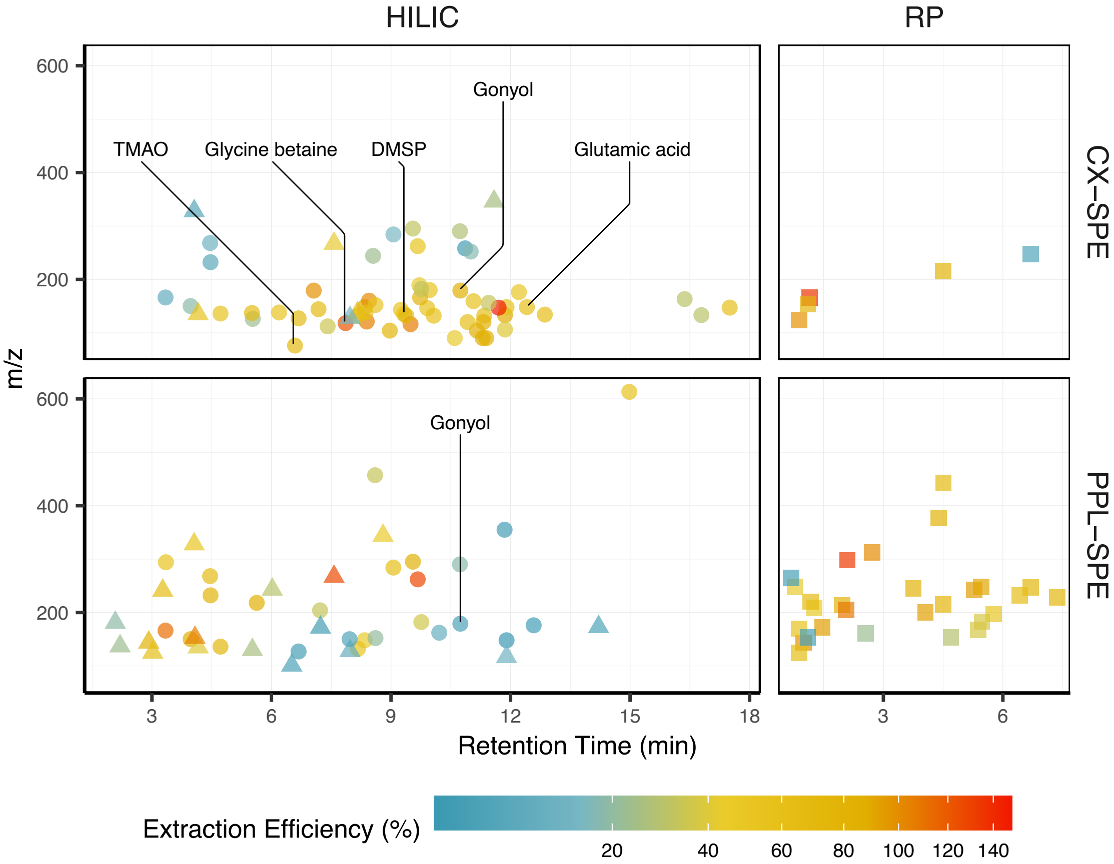

Figure 2.2: Extraction efficiency of targeted compounds extracted by CX-SPE and PPL-SPE shown by the compounds' *m/z* and retention time for both types of chromatography. Different analytical fractions are denoted by shape (HILIC Positive, circle; HILIC Negative, triangle; RP, square). Several compounds representative of various function classes of metabolites extracted by CX-SPE (TMAO (trimethylamine N-oxide), amine oxides; homarine and glycine betaine, betaines; DMSP (dimethylsulfoniopropionate) and gonyol, sulfoniums; Glutamic acid, amino acids) are annotated. Note that the color bar has a square root transformation.

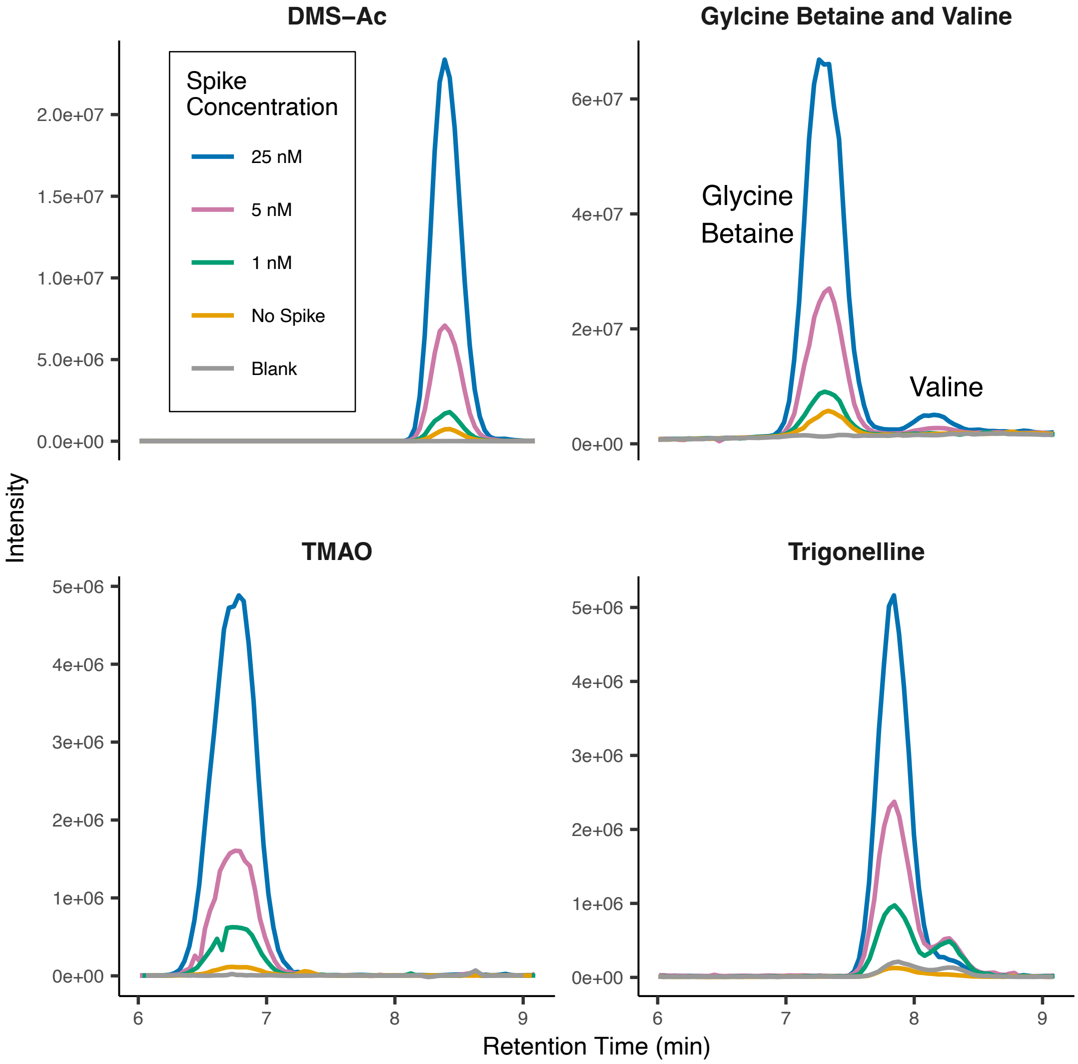

Figure 2.3: Extracted ion chromatograms (EIC) of selected compounds at different standard spike concentrations in ALOHA seawater processed using CX-SPE. “Blank” is a CXC-H_2O blank. “No Spike” is an unamended sample of ALOHA seawater. “1 nM,” “5 nM,” and “25 nM” are samples where the depicted standards were spiked into a sample of ALOHA seawater at 1, 5, and 25 nM concentrations before processing with CX-SPE. DMS-Ac is dimethylsulfonioacetate and TMAO is trimethylamine N-oxide. These chromatograms were not used in any of the method assessment presented in this study.

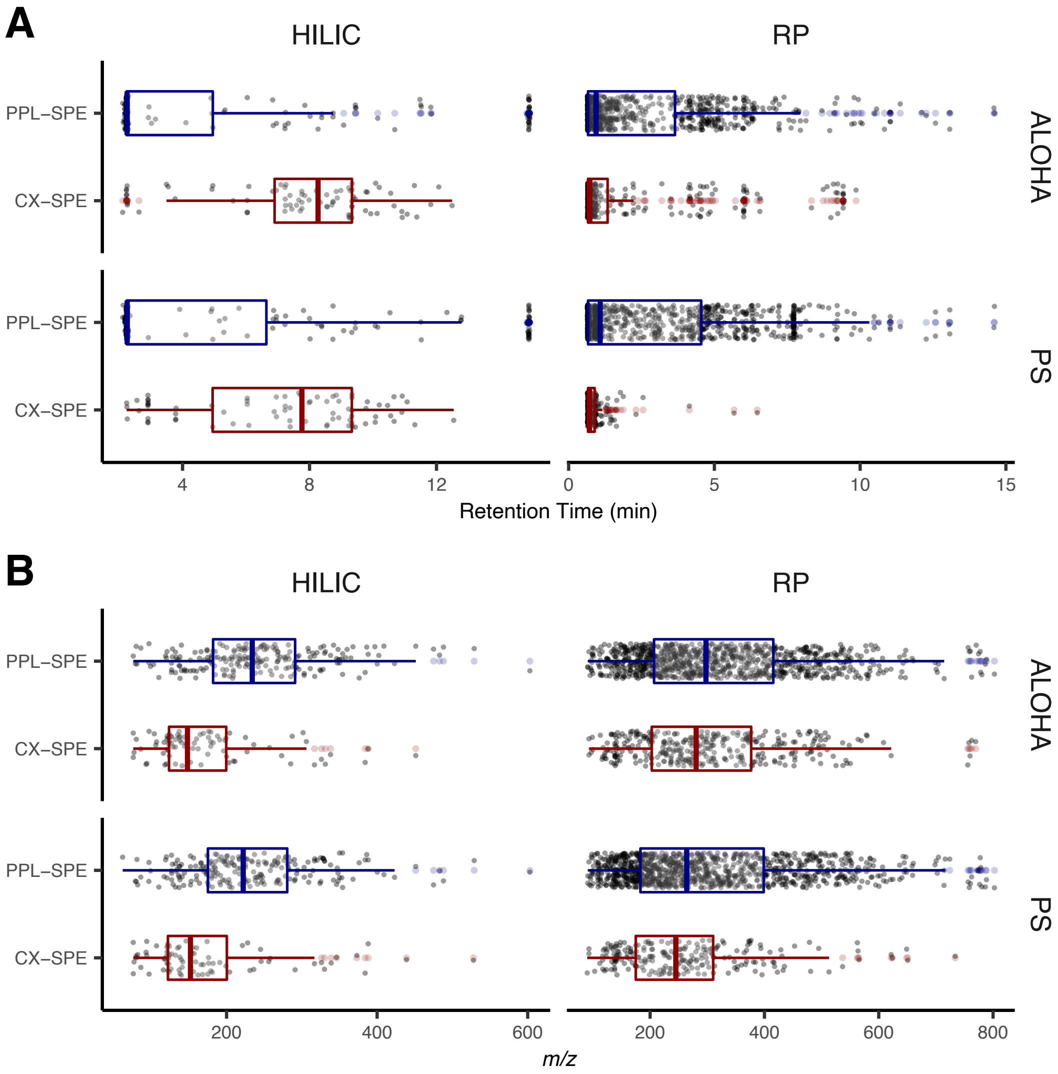

Figure 2.4: scatter, and box plots showing the distribution of retention times (A) and *m/z* values (B) of untargeted, high-quality mass features generated by CX-SPE and PPL-SPE using HILIC and RP chromatography from ALOHA and PS samples.

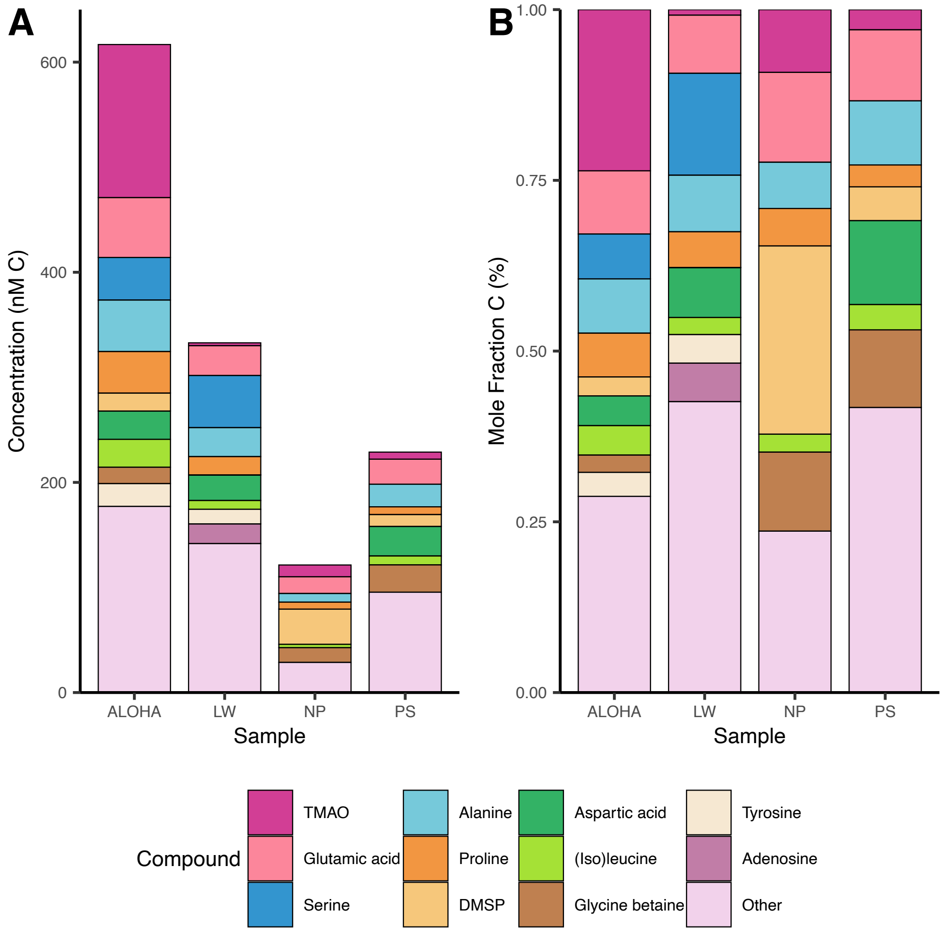

Figure 2.5: Average concentration in nM C (A) and average relative concentration in mole fraction C (B) of compounds quantified by CX-SPE in nM in four discrete environmental samples. The 11 compounds that contribute the most carbon are shown individually while all other compounds are grouped together.

2.9 Tables:

Compound or Compound Class	Concentration measured with CX-SPE (nM)	Locations measured with CX-SPE	Literature Concentrations (nM)	Literature Locations	Reference
Individual dissolved free amino acids	0.15 - 17	ALOHA, PS, NP, LW	<0.01-62.5	South Atlantic Transect	Sabadel et al. 2017[a]
			0.06 - 3.0	Bermuda Atlantic Time Series	Widner et al. 2021[b]
			0-16.9	Jardines de la Reina, Cuba Reef System	Weber et al. 2020[c]
			0 – 20	Equatorial Pacific	Lee and Bada, 1975[a]
			1.0-15	New York Bight	Fuhrman and Ferguson, 1986[a]
TMAO	0.89 - 49	ALOHA, PS, NP, LW	<1.65 - 76.9	Antarctic Coastal Waters	Gibb and Hatton 2004[d]
DMSP	0.013 - 6.7	ALOHA, PS, NP, LW	1.1-2.8	Sargasso Sea	Kiene et al. 2006[d]
Creatine	0.20 - 0.90	ALOHA, PS, NP, LW	6.3-57	North Pacific Subtropical Gyre	Wawrik et al. 2017[e]
Arsenobetaine	0.015	ALOHA	0.0067-0.13	North Atlantic	Glabonjat et al. 2018[f]
Xanthine	0.073 - 0.092	ALOHA, PS	0-1.06	Jardines de la Reina, Cuba Reef System	Weber et al. 2020[c]
Ectoine	0.046 - 0.054	ALOHA, NP	0.10-0.37	Bermuda Atlantic Time Series	Widner et al. 2021[b]
Citrulline	2.0 - 2.6	ALOHA, LW	0.28-0.31	Bermuda Atlantic Time Series	Widner et al. 2021[b]

Ornithine	8.2 - 8.3	ALOHA, LW	0.42-0.54	Bermuda Atlantic Time Series	Widner et al. 2021[b]
Sarcosine	0.54 - 1.2	ALOHA, NP, PS	0.09-0.22	Bermuda Atlantic Time Series	Widner et al. 2021[b]

Table 2.1: Concentrations of dissolved metabolites in environmental samples measured in this study with CX-SPE and in other studies. Superscript denotes extraction approach ("a", DFAA specific approach; “b”, BC derivatization/PPL-SPE; “c”, PPL-SPE, “d”, chemical or enzymatic degradation to volatile analyte; “e”, Select SCX-SPE; “f”, CX-SPE.)

Compound	**Fraction**	**EE (%)**	**RSD of EE (%)**	**R^2**	**LOD (nM)**
(3-Carboxypropyl)trimethylammonium	HILIC Pos	46.8	27.6	0.964	0.012
(Iso)leucine	HILIC Pos	59.7	17.6	0.949	0.23
3',5'-Cyclic Adenosine Monophosphate	HILIC Neg	5.92	39.7	0.886	0.0071
4-Aminobutyric acid	HILIC Pos	89.2	16.9	0.934	0.67
5-Hydroxyectoine	HILIC Pos	47.2	16	0.994	0.0074
5-Methylcytosine	HILIC Pos	21	15.1	0.962	0.042
5-Oxoproline	HILIC Neg	18.1	38.5	0.912	17
Abscisic acid	RP	4.77	38.1	0.925	0.62
Adenine	HILIC Pos	49.1	18.4	0.954	0.23
Adenosine	HILIC Pos	10.6	20.9	0.953	0.12
Adenosine monophosphate	HILIC Neg	23	39.3	0.961	0.027
Allopurinol	HILIC Neg	46	38.6	0.859	1.2
Arsenobetaine	HILIC Pos	88.9	14.1	0.992	0.0013
beta-Alanine	HILIC Pos	82.1	15.4	0.977	1.3
beta-Alaninebetaine	HILIC Pos	66.2	15.1	0.991	0.014
beta-Glutamic acid	HILIC Pos	52.9	12.7	0.843	1.0
Betonicine	HILIC Pos	89.2	14.1	0.996	0.0059
Butyrylcarnitine	HILIC Pos	5.56	39.2	0.93	0.009
Citrulline	HILIC Pos	40.2	27.7	0.857	1.5
Creatine	HILIC Pos	50.4	19.5	0.98	0.047
Cytidine	HILIC Pos	19.9	21.3	0.97	0.038
Cytosine	HILIC Pos	30.8	15.2	0.969	0.05
Desthiobiotin	RP	66.1	38.3	0.784	0.0042
Dimethylglycine	HILIC Pos	65.1	14.9	0.993	0.7

Dimethylsulfonioacetate	HILIC Pos	94.8	12.6	0.996	0.00013
Dimethylsulfoniopropionate	HILIC Pos	55.5	15.5	0.973	0.0001
Ectoine	HILIC Pos	47.6	12.5	0.99	0.025
Glucosamine	HILIC Pos	49.4	15.9	0.988	0.21
Glutamylphenylalanine	HILIC Pos	26	23.4	0.923	0.012
Glycerophosphocholine	HILIC Pos	2.24	47.8	0.931	0.25
Glycine betaine	HILIC Pos	100	20.6	0.969	0.38
Gonyol	HILIC Pos	71	26.2	0.974	0.17
Guanine	HILIC Pos	47.1	26	0.955	0.35
Guanosine	HILIC Pos	13.2	30	0.952	0.17
Homarine	HILIC Pos	50	15.2	0.995	0.15
Hordenine	HILIC Pos	7.29	41.9	0.732	0.06
Hydroxyisoleucine	HILIC Pos	63.2	39.1	0.916	0.25
Hydroxyproline	HILIC Pos	54.5	16.1	0.965	0.041
Hypoxanthine	HILIC Pos	59.6	18.7	0.978	0.39
Inosine	HILIC Neg	36.7	28.2	0.95	0.023
L-Alanine	HILIC Pos	106	30.3	0.974	1.8
L-Asparagine	HILIC Pos	82.7	14.4	0.916	0.49
L-Aspartic acid	HILIC Pos	67.2	18.5	0.963	3.4
L-Glutamic acid	HILIC Pos	53.7	22.4	0.85	2.0
L-Glutamine	HILIC Pos	131	16.8	0.988	0.024
L-Histidine	HILIC Pos	33	36.8	0.969	0.66
L-Homoserine	HILIC Pos	68.8	18.1	0.986	0.18
L-Hydroxylysine	HILIC Pos	38.2	23.4	0.732	0.0026
L-Lysine	HILIC Pos	56.2	39.1	0.888	1.5
L-Methionine S-oxide	HILIC Pos	72.6	18.6	0.973	0.16

L-Ornithine	HILIC Pos	47.2	30.7	0.907	6.8
L-Phenylalanine	RP	128	21	0.914	0.2
L-Proline	HILIC Pos	104	16.4	0.984	0.51
L-Serine	HILIC Pos	75.3	38.4	0.919	4.0
L-Threonine	HILIC Pos	62.2	17.5	0.942	0.21
L-Tyrosine	HILIC Pos	28.6	26.7	0.938	0.88
Melamine	HILIC Pos	60.2	18.2	0.975	0.36
Muramic acid	HILIC Pos	14.6	25.5	0.968	0.046
N6-Acetyl-L-lysine	HILIC Pos	37.8	26.4	0.984	0.015
N6-Methyladenine	HILIC Pos	15.8	29.2	0.917	0.02
Nicotinic acid	RP	102	12.2	0.894	1.1
O-Methylmalonyl-L-carnitine	HILIC Pos	47.4	20.7	0.989	0.0064
Ophthalmic acid	HILIC Pos	23	27.8	0.959	0.02
Proline betaine	HILIC Pos	58.2	17.3	0.985	0.022
Sarcosine	HILIC Pos	41	14.4	0.968	0.32
Trigonelline	HILIC Pos	53.3	15.7	0.987	0.057
Trimethylamine N-oxide	HILIC Pos	54.7	33	0.946	0.13
Urocanic acid	HILIC Neg	65.6	17.6	0.939	0.73
Xanthine	RP	69.2	22.7	0.965	0.055

Table 2.2: Analysis fraction, extraction efficiency (EE), relative standard deviation of extraction efficiency (RSD_{EE}), R^2 values, and limits of detection (LODs) for compounds meeting quality control thresholds using CX-SPE. Isoleucine and leucine are not fully separated with our chromatography and so our treated together as (iso)leucine.

2.10 Supplemental Text:

Fraction Collection Optimization Test:

To determine the optimal number of fractions to collect, we conducted an experiment where our HILIC positive standards mix was spiked into a sample of ALOHA seawater and the CX-SPE procedure was performed as stated in the methods. Only the HILIC positive ion mode standards were evaluated as those compounds were the primary targets of the CX-SPE approach and we attempted to optimize our procedure for their successful extraction. Each 10 mL fraction during elution was collected and dried down separately. In this experiment, the alkaline front was fraction 10. The percent of total peak area (proportional to concentration) in each fraction was then compared (Supplemental Table S3, Supplemental Figure S3). The alkaline front (fraction 10), the fraction before (fraction 9), and the two fractions after (fractions 11 + 12) contained, on average, 91.4% of the total signal area for each compound (Supplemental Table S3). This trend was shown to be consistent across almost every compound measured in the HILIC positive standards mix (Supplemental Figure S3). Therefore, the four fractions selected for analysis and further method evaluation going forward were the alkaline front, the fraction before, and the two fractions after as these represented over 90% of the total signal and could be dried down for reconstitution in a reasonable time frame.

2.11 Supplemental Figures:

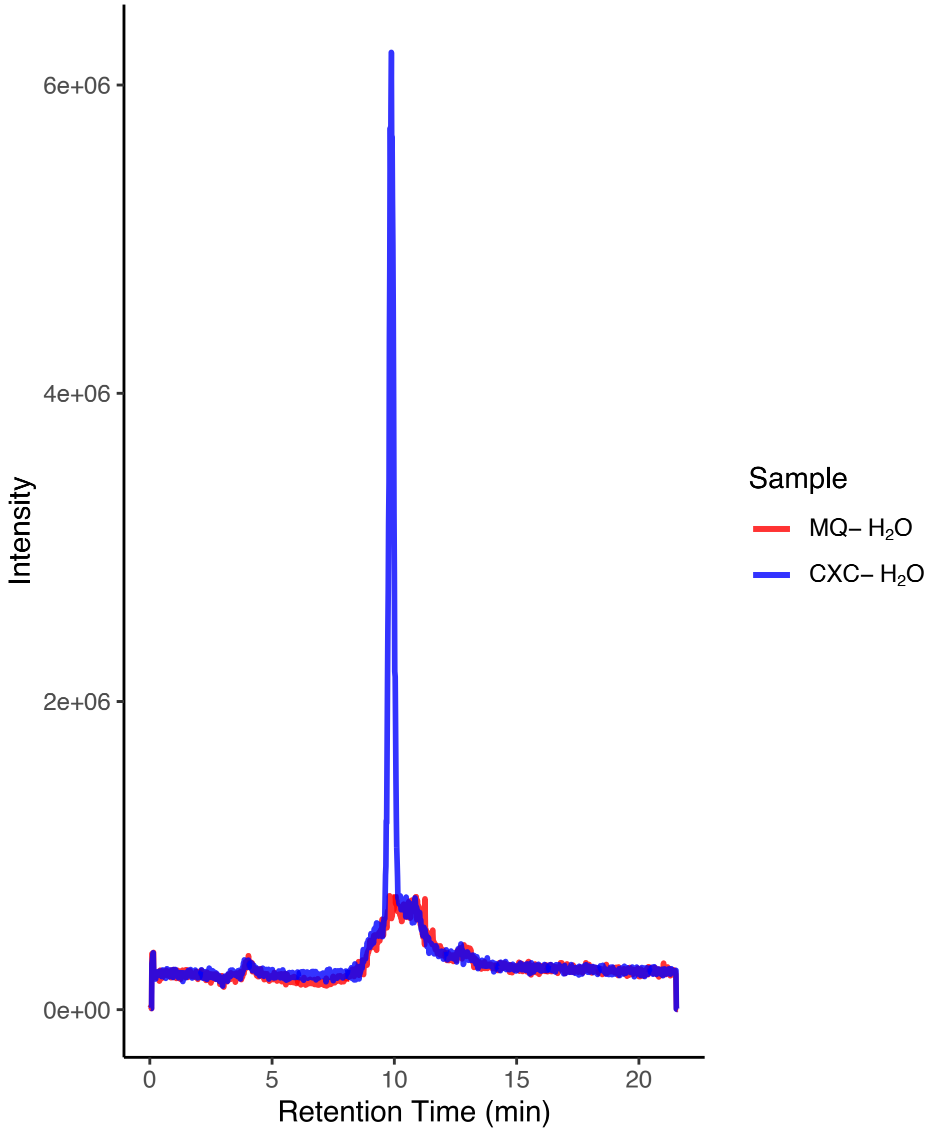

Figure 2.6: Extracted ion chromatograms of m/z 118.0868 (the exact mass of glycine betaine) in MilliQ H_2O before (MQ-H_2O) and after (CXC-H_2O) running on CX-SPE cleaning column to create the CXC-H_2O. The peak in the MQ-H_2O EIC at approximately 10 minutes is glycine betaine.

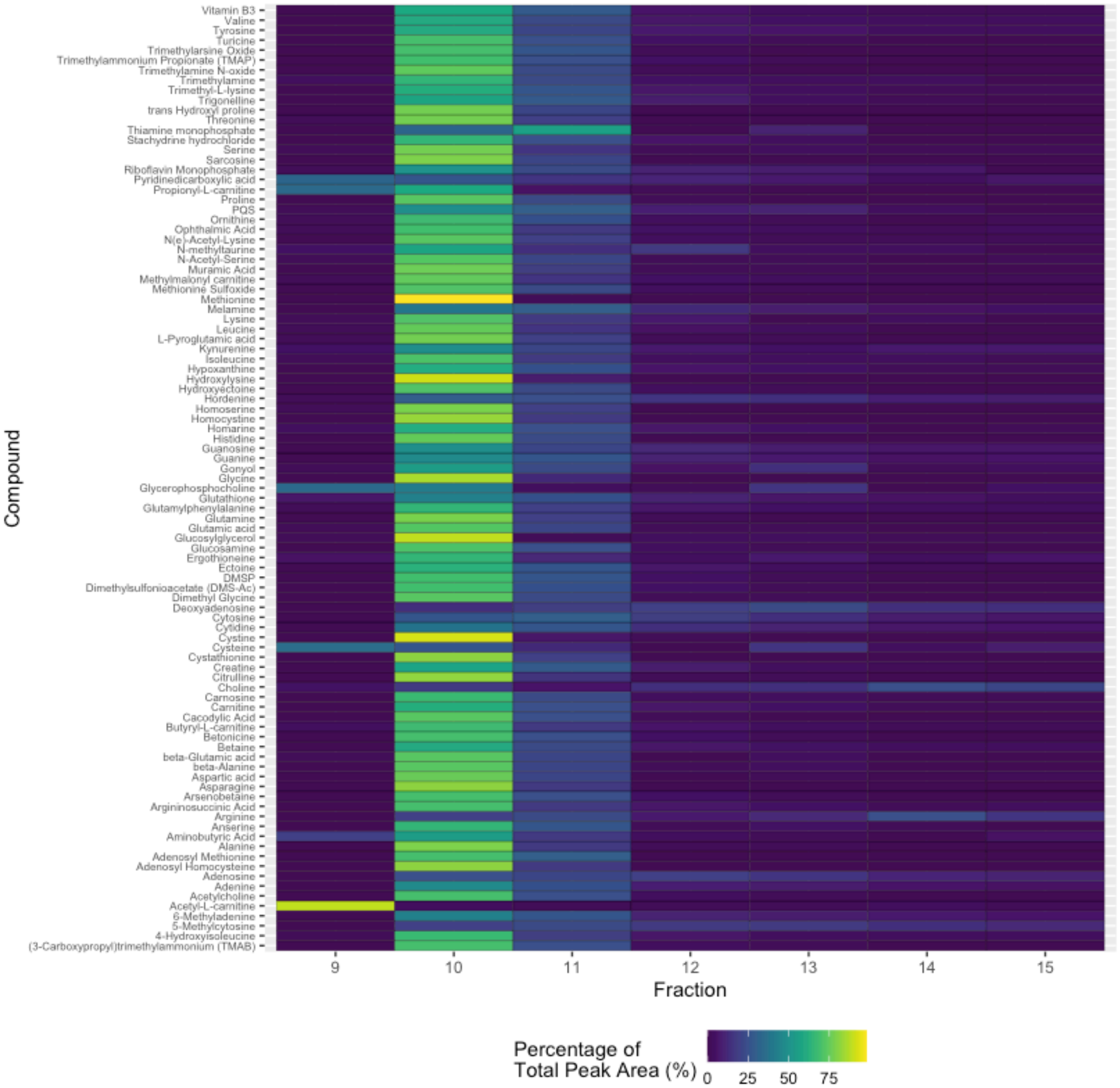

Figure 2.7: The heatmap shows the percentage of total peak area in each elution fraction for each compound in the HILIC standard mix measured in positive ion mode. Brighter colors indicate higher relative percentage of peak area. In this test, fraction 10 was the alkaline front.

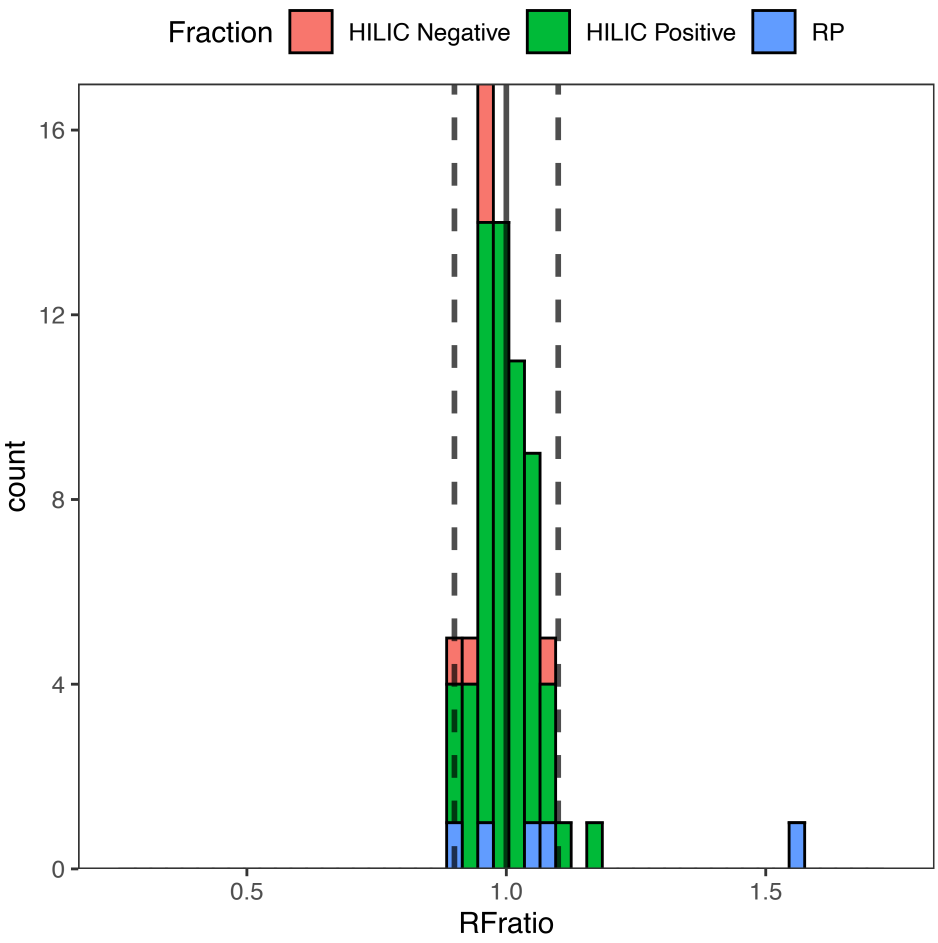

Figure 2.8: Histogram of the RF ratios of compounds reproducibly extracted using CX-SPE. An RF ratio of 1 indicates low matrix effects. an RF ratio greater than 1 indicates ion enhancement, and RF ratio less than 1 indicates ion suppression. Compounds are colored by their analysis fraction. The solid vertical line is at an RF ratio value of 1 and the two dotted lines show RF ratio values of 0.9 (left) and 1.1 (right).

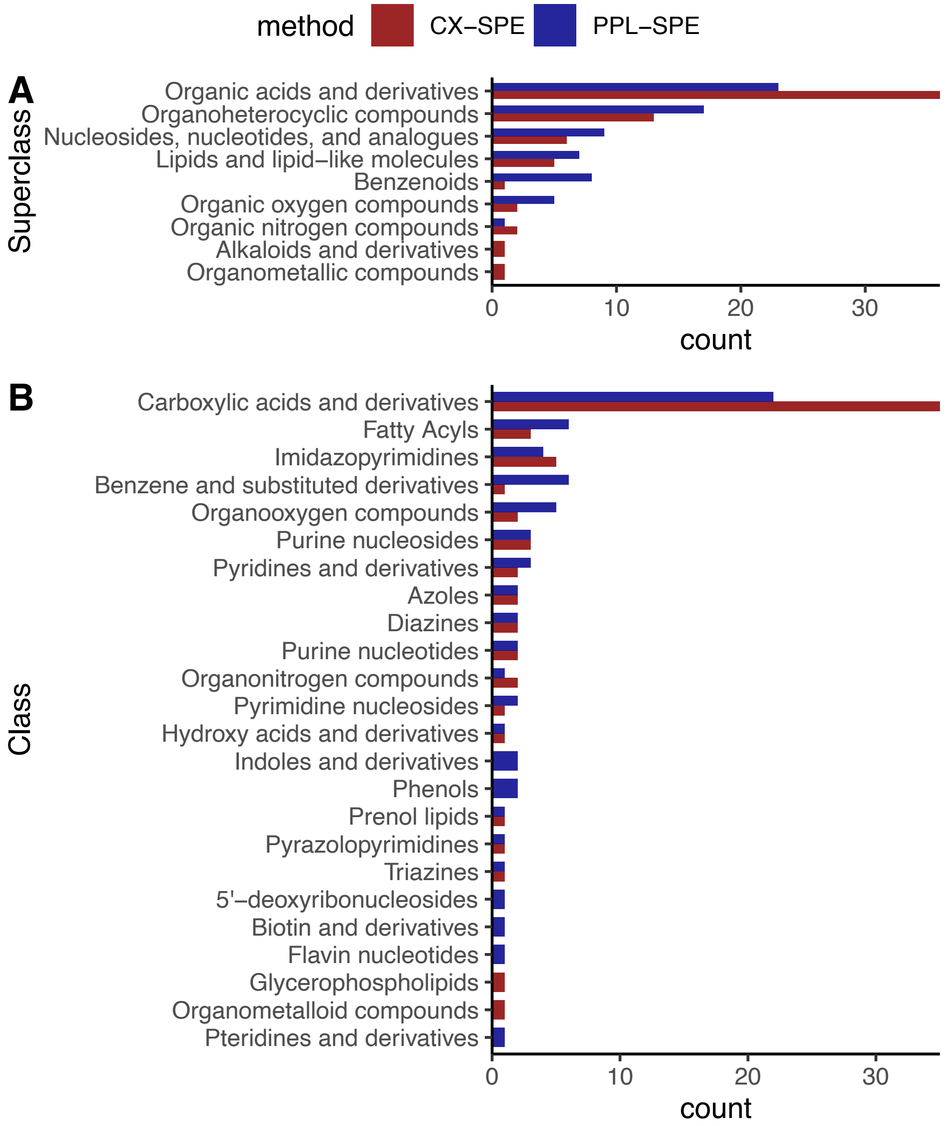

Figure 2.9: Bar charts showing the number of targeted compounds extracted by CX-SPE and PPL-SPE in each superclass (A) and class (B) of the Chem-Ont cheminformatics classification system.

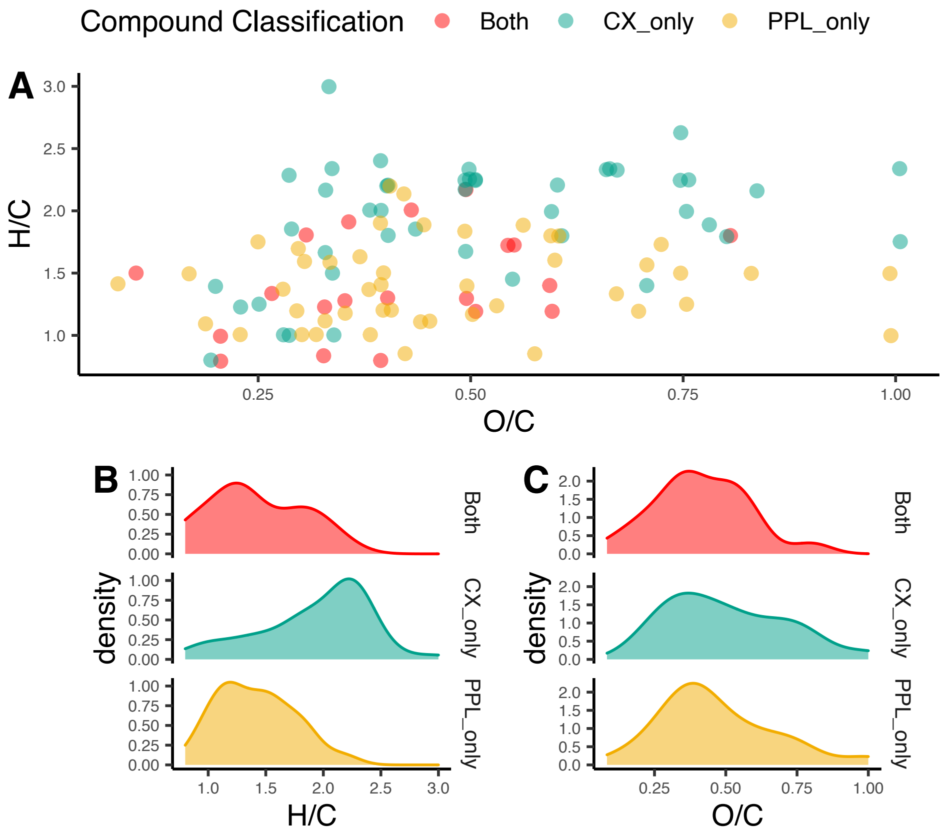

Figure 2.10: Van Krevelen plot (A) of targeted compounds showing the C/O ratio and C/H ratio of metabolites reproducibly extracted by CX-SPE (green), PPL-SPE (yellow), and both CX-SPE and PPL-SPE (red). A small amount of jitter was added to the points to minimize overplotting of points. B and C show density distributions of C/H ratio and C/O ratio, respectively, separated by analytical category (Both, CX-SPE only, and PPL-SPE only).

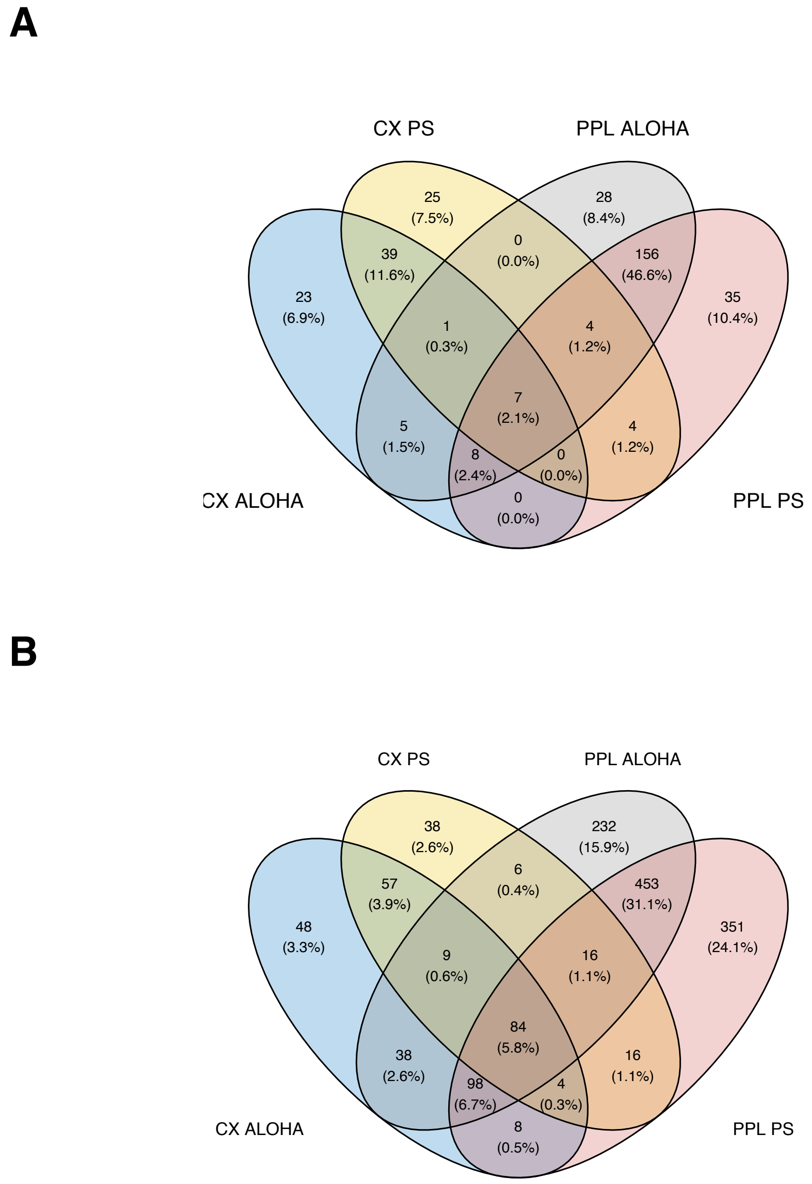

Figure 2.11: Venn diagrams showing the presence/absence of mass features that passed quality control filters for HILIC chromatography (A) and reversed phase chromatography (B). Samples/method pairs are identified by the station the sample was collected from (ALOHA or PS) and the SPE method by which they were processed (CX or PPL). There were 335 and 1458 HILIC and RP features, respectively.

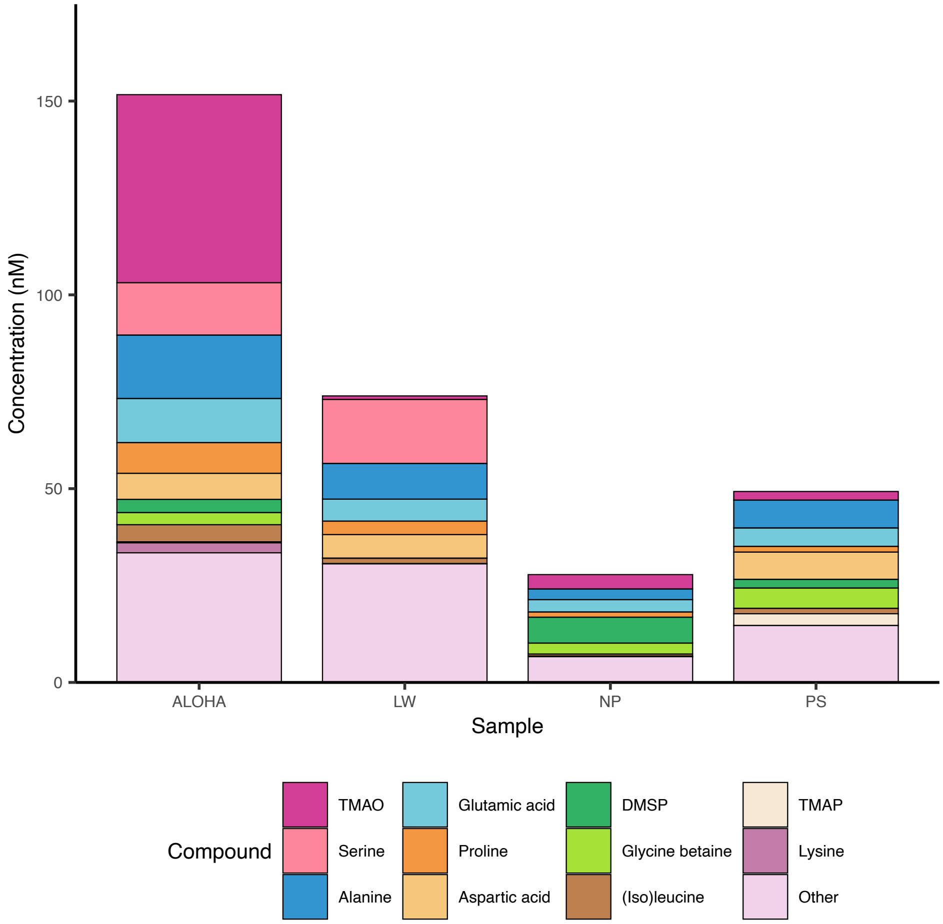

Figure 2.12: Average concentration in nM compound of compounds quantified by CX-SPE in four discrete environmental samples. The 11 most abundant compounds are shown individually while all other compounds are grouped together.

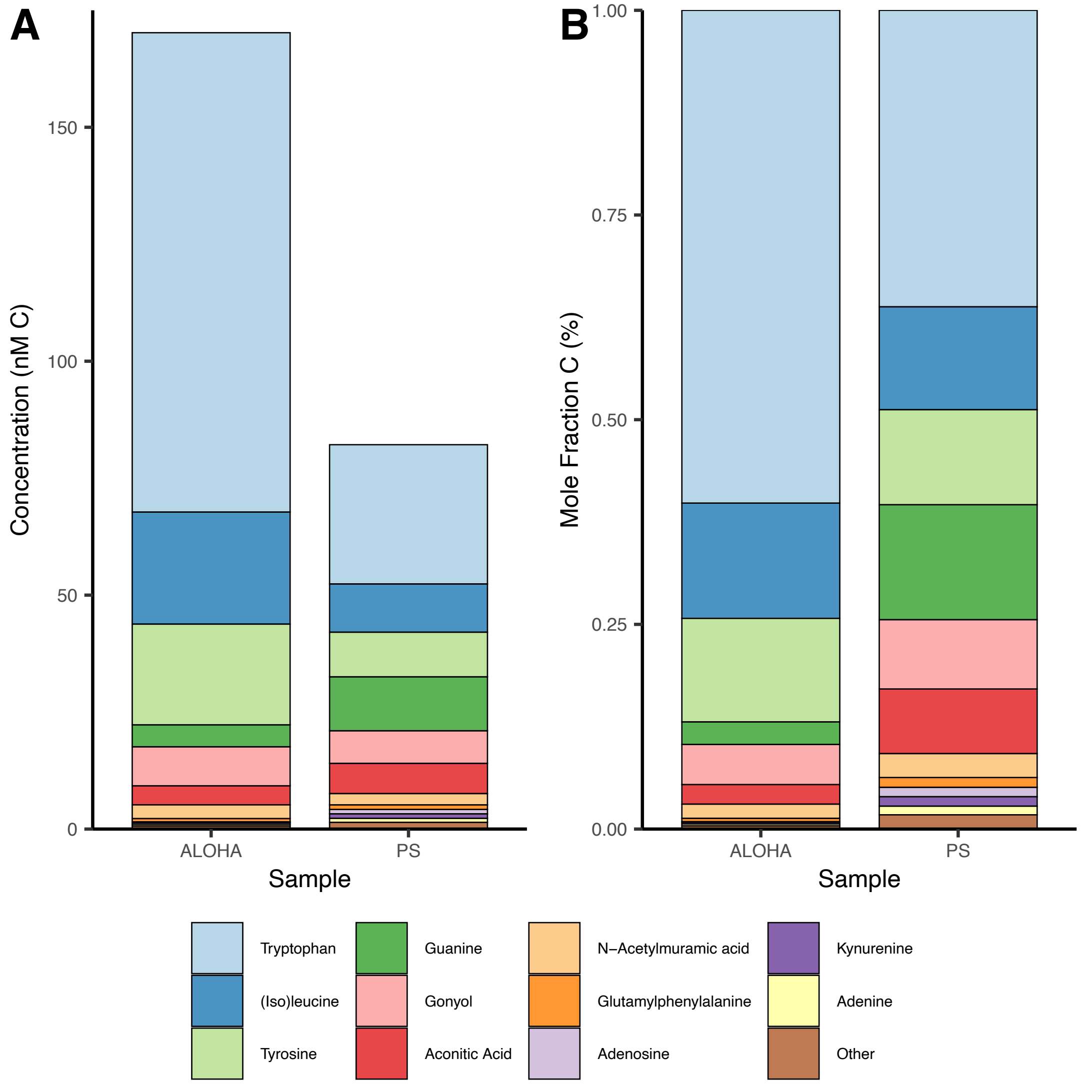

Figure 2.13: Average concentration in nM C (A) and average relative concentration (mole fraction C) of compounds quantified by PPL-SPE in ALOHA and PS samples. The 11 compounds that contribute the most carbon are shown individually while all other compounds are grouped together.

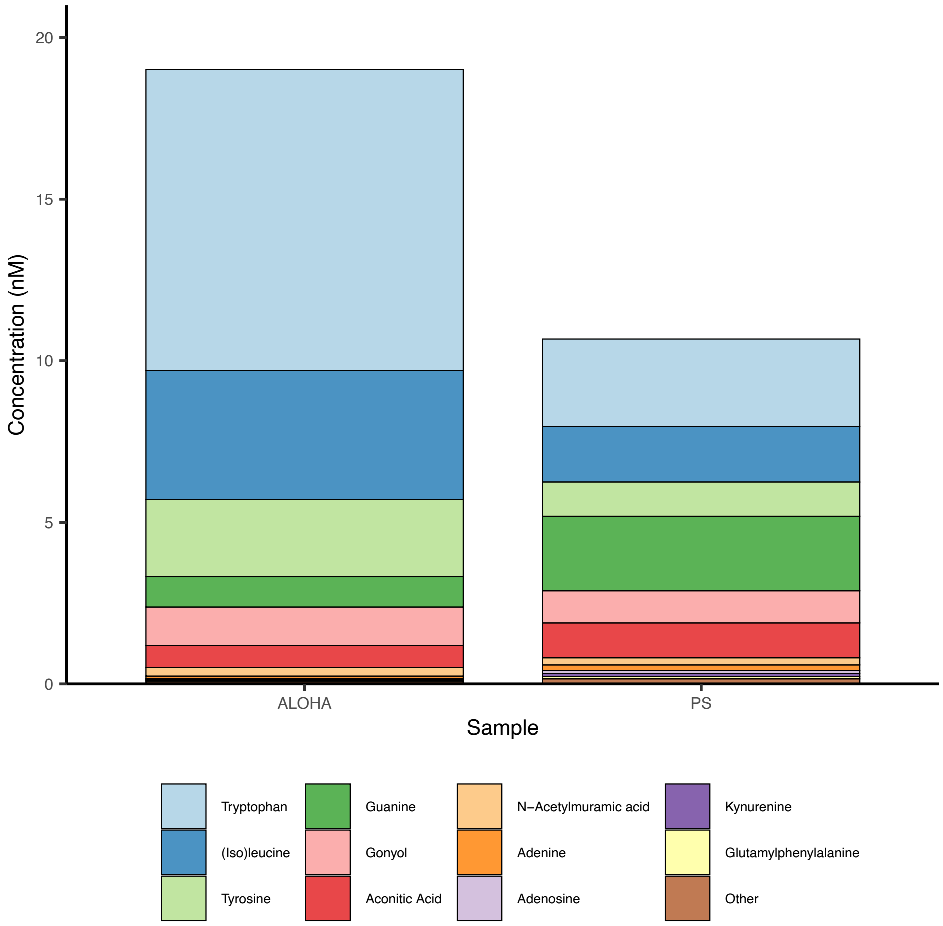

Figure 2.14: Average concentration in nM compound of compounds quantified by PPL-SPE in ALOHA and PS samples. The 11 most abundant compounds are shown individually while all other compounds are grouped together.

2.12 Supplemental Tables:

Table 2.3: Identities, mass spectrometry and chromatography details, tested spike concentrations, and supplier information of all standards used in this study. This table also includes a “Status” column that summarizes whether the compound could be successfully and reliably extracted by CX-SPE, PPL-SPE, both, or neither method. This table is provided in a separate file.

Table 2.4: Identities and concentrations of internal standards used in “Extraction” and “Injection” internal standards mixes. This table is provided in a separate file.

Elution Fraction	**Average Percentage of Total Peak Area (%)**	**Cumulative Average Percentage of Total Peak Area (%)**
9	3.5	3.5
10 (Alkaline Front)	62.5	66
11	20.5	86.5
12	4.9	91.4
13	4.0	95.4
14	2.4	97.8
15	2.2	100

Table 2.5: Results of preliminary optimization test. This table shows the average and cumulative average percentage of peak area (proportional to concentration) in each elution fraction of the total peak area observed in all elution fractions.

Table 2.6: MS-DIAL parameters used in untargeted analysis. This table is provided in a separate file.

Table 2.7: CX-SPE sample specific extraction efficiencies (EE) and EE standard deviations for compounds passing quality thresholds for in each of the environmental sample (four for HILIC compounds, two for RP compounds) matrices tested. This table is provided in a separate file.

Table 2.8: PPL-SPE sample specific extraction efficiencies (EE) and EE standard deviations for compounds passing quality thresholds in both environmental sample matrices tested. This table is provided in a separate file.

Table 2.9: Final list of mass features (MFs) identified in untargeted analysis that passed quality control thresholds. MFs are identified by analysis fraction, m/z, and retention time. If the MF passed quality control thresholds and was considered present in a method-sample pair (ex. CX-SPE – PS), the average peak area of the three analytical replicates is reported. If the MF did not pass quality control thresholds it is marked as not detected (ND). This table is provided in a separate file.

Table 2.10: Environmental concentrations and standard deviations in nM compound, nM C, and nM N for compounds quantified using CX-SPE. This table is provided in a separate file.

Table 2.11: Environmental concentrations and standard deviations in nM compound, nM C, and nM N for compounds quantified using PPL-SPE. This table also contains the calculated LOD thresholds (nM) applied to the PPL environmental dataset. This table is provided in a separate file.

Chapter 3: Remodeling of *Prochlorococcus* Metabolism During Viral Infection

3.1 Abstract[2]:

The marine cyanobacterium *Prochlorococcus* is the most abundant photoautotroph in the world and a major contributor to oceanic primary productivity. Viruses are important controls on *Prochlorococcus* populations with infection rates reaching up to 10% of *Prochlorococcus* populations. During infection, viruses remodel their host's metabolic machinery, creating metabolically distinct cells, termed virocells. However, the specific consequences of viral infection on *Prochloroccus* metabolism remain poorly understood. Here, we characterize the infection of non-axenic cultures of *Prochlorococcus* MED4 by the T7-like virus P-SSP7 using a combination of metabolomics, transcriptomics, and population modeling approaches. P-SSP7 infection dramatically altered the culture metabolome with 25% of metabolites showing differential abundance. Infected cells exhibited decreased carbon fixation and the draw down of intracellular stores of carbon structures and energy such as glycogen and the osmolytes sucrose and aspartic acid. In contrast, another osmolyte, glucosylglycerol, was accumulated in high concentrations and came to dominate the virocell metabolome. Infected cells also experienced pseudocobalamin (pB12) stress, as evidenced by reduced pB12 concentrations, increased expression of genes to synthesize pB12, and depletion of metabolites whose synthesis relies on

---

[2] This chapter is current in preparation for submission as: Sacks, J.S**.**, Ribalet, F., Linney, M., Kearney, S., Bjorkman, K., Carlson, L.T., Haley, S., Karl, D., Dyhrman, S., Armbrust, E.V., Lindell, D., Ingalls, A.E. “Remodeling of *Prochlorococcus* Metabolism During Viral Infection.”

pB12 including S-adnosylmethionine (SAM) and the antioxidant glutathione. Collectively, our results suggest that the observed metabolic remodeling is driven by the demand for carbon and energy for virion production and infection-induced oxidative stress. Viral infection changes the substrate and vitamin pools provided by *Prochlorococcus* to the microbial community, potentially altering the speciation and flux of organic matter in marine systems and acting as a selective force on microbial community composition and function.

3.2 Introduction:

The picocyanobacterium *Prochlorococcus* is the most abundant primary producer in the surface ocean where it is estimated to perform 8.5% of marine primary production globally[103]. *Prochlorococcus* forms the base of oligotrophic marine ecosystems, supplying growth substrates, biosynthetic precursors, and vitamins to biological communities[27,104,105]. *Prochlorococcus* abundances and ecosystem function are in part controlled by mortality agents such as viruses and grazers[47,48,104]. Through infection and lysis, viruses contribute to regulating the abundance, diversity, and biogeography of *Prochlorococcus*. Viral lysis of *Prochlorococcus* also releases cellular metabolites and macromolecules into the environment, providing an additional source of bioavailable organic matter to the microbial community[50,106].

During infection, viruses transform cells into metabolically and phenotypically distinct forms, termed virocells, to facilitate virion production[107]. Virion production is a resource and energy intensive process that requires carbon, nitrogen, and phosphorus in the form of biosynthetic precursors as well as reducing power to fuel the production of amino acids and nucleotides[108,109]. To meet these demands, viruses alter their host's metabolism through a variety of mechanisms including changes to host gene expression and the use of auxiliary metabolic genes (AMGs) which encode phage metabolic proteins[108,110,111]. High rates of cyanobacterial viral infection have been measured in the ocean, reaching up to 10% of the *Prochlorococcus* population in some areas[47,112]. In those regions, *Prochlorococcus* virocells likely represent a significant fraction of the community and play an important role in ecosystem function and biogeochemical cycling.

Most of these changes to cellular metabolism are not predicted by a virus's AMGs and therefore direct observation of infected cells is required to characterize virocell metabolism[113]. Metabolomics is a powerful tool for directly characterizing virocell phenotype and can provide insights not predicted by molecular biology methods[22,113,114]. While previous work suggests that virally driven alterations to *Prochlorococcus* metabolism include alterations to photosynthesis, the pentose phosphate pathway, nucleotide synthesis, and nutrient acquisition[108,109,115–119], current understanding of *Prochlorococcus* virocell metabolism is limited by a lack of direct metabolic observations. Thus, the role of metabolic processes in structuring viral infection dynamics remain unknown and the consequences of viral infection on community level processes remain unconstrained. Here, we investigated the infection of *Prochlorococcus* MED4 by the T7-like cyanophage P-SSP7. Using a combination of metabolomics, transcriptomics, and population modeling approaches, we characterized the remodeling of *Prochlorococcus* metabolism during P-SSP7 infection.

3.3 Methods:

Experimental Design, Abundance, and ATP measurements:

Details on experimental design, abundance, and ATP measurements are presented in Lindell et al. (in prep) and in the supplemental methods. In brief, we examined differences between uninfected and P-SSP7 infected MED4 cells, by incubating three treatments consisting of a control (C, no virus), a low virus treatment (LV, 0.001 initial phage to *Prochlorococcus* ratio), and a high virus treatment (HV, 0.7 initial phage to *Prochlorococcus* ratio) (Figure 3.1A). All treatments were non-axenic. The experiment was run for a total of 48 hours. Samples collection frequencies were as follows: *Prochlorococcus* abundance: 2 h; heterotrophic bacteria and phage abundance: 6 h; ATP, transcriptomics, and metabolomics: 12 h (6 h for metabolomics control treatment samples) (Figure 3.1B). *Prochlorococcus* and heterotrophic bacteria abundances were measured using flow cytometry and phage abundance was measured using qPCR. ATP was measured using the firefly bioluminescence assay[120].

Biomass measurements and size-structured matrix population modeling:

Flow cytometry-based estimates of the cell size of *Prochlorococcus* and two classes of heterotrophic bacteria ("small" and "large") were measured using light scatter and fluorescence to discriminate between populations. The equivalent spherical diameter of each cell was estimated using Mie light scatter theory and converted to a carbon quota for each cell using volume-to-carbon relationships as described in Ribalet et al.[121]. We then applied a size-structured matrix population model to the *Prochlorococcus* size-distribution dataset following the approach described in Mattern et al.[122]. This model quantifies the rates of cell division, carbon fixation, and carbon loss (respiration+exudation) from the changes in size classes observed during the day/night cycle.

Metabolomics Analysis:

Particulate metabolites were sampled using vacuum filtration, flash frozen in liquid nitrogen, and stored at -80 ˚C until analysis. Metabolites were extracted using a modified Bligh and Dyer two-phase extraction as described in Boysen et al.[24]. Metabolite MS1 and DDA MS2 data were collected using established liquid chromatography mass spectrometry (LC-MS) methods for targeted and untargeted metabolomics and targeted vitamin analyses[24,25]. Metabolite data was processed using a hybrid targeted-untargeted approach combining MSDIAL and Skyline[76,77]. Vitamin data were manually processed using Skyline. Metabolite data were normalized using Best Matched Internal Standard Normalization[24] and quality controlled to remove low quality molecular features, adducts, isotopes, and duplicated compounds between positive and negative ionization modes using in-house scripts. Annotation of untargeted features was performed using SIRIUS 5 (V5.5.7)[123].

Targeted metabolites were quantified using authentic standards or isotopically labeled internal standards. Vitamins were quantified using a standard curve. Glutathione is not stable under our extraction procedure, with a least a portion of glutathione oxidizing to glutathione disulfide. We therefore interpret our glutathione disulfide measurements to reflect the sum of glutathione and glutathione disulfide concentrations and refer to it henceforth as “glutathione (disulfide)”. One sample (C_LV_T48) was removed from the analysis as it was compromised during sample processing.

We identified metabolites that were differentially abundant under viral infection by comparing three treatment/timepoint combinations that had high levels of viral infection (immediately preceeding significant Prochlorococcus mortality) and high relative biomass of Prochlorococcus to heterotrophic bacteria (Prochlorococcus > 80% of microbial carbon) (HV-T12, HV-T24, and LV-T36). We classified these samples as high viral infection (HVI) samples. We used a fold change analysis on biomass normalized peak areas to identify differentially abundant compounds in the HVI samples relative to the control (t-test and a Wilcoxon signed-rank test). We removed all metabolites significantly correlated ($p < 0.05$) with bacterial biomass to minimize the impact of heterotrophic bacterial growth on our interpretation.

Transcriptomic Analysis:

Samples were collected for C and HV treatments using vacuum filtration, flash frozen in liquid nitrogen and stored at -80 ˚C until extraction. Total RNA was extracted from individual filters using a Qiagen RNeasy Mini Kit with a modified lysis step and the addition of synthetic standards. Samples were then processed after rRNA depletion to enrich for prokaryotic mRNA. Library preparation and sequencing (NovaSeq6000) of Illumina®TruSeq libraries was performed at the JP Sulzberger Columbia Genome Center (New York, New York) following Center protocols. RNA was sequenced to produce 40 million 100-bp, paired-end reads. Reads were quality controlled, mapped to the MED4 reference genome, counted, and normalized to both spike-in standards and a *Prochlorococcus* housekeeping gene (*rnpB*)[117]. Differential expression analysis was performed using the DESeq2 R package (V1.36.0) between pairwise HV and C samples.

For details on sampling, flow cytometry and population modeling, metabolite data acquisition and processing, transcriptomics sequencing and processing, and statistical analyses, see supplemental methods. All statistical analyses for metabolomics and transcriptomics were performed in R (V4.4.1).

3.4 Results and Discussion:

Intracellular metabolomes differed significantly between infected and uninfected treatments

P-SSP7 addition led to substantial *Prochlorococcus* mortality and increases in the abundances of both P-SSP7 and heterotrophic bacterial over the course of the experiment, suggesting high levels of viral infection and lysis (Figure 3.1C, 3.1D, 3.1E) (Lindell et al. in prep) (Table 3.1). Untargeted intracellular metabolomes differed significantly over time and with treatment (Figure 3.8, Table 3.2). NMDS ordination of untargeted molecular features (i.e., distinct ions detected in the untargeted analysis, n = 623) separated samples based on treatment and time, with infected and control samples diverging at later timepoints (Figure 3.2A). A perMANOVA analysis confirmed that treatment and time were both significant contributors to variability in the dataset, explaining 17 % (p = 0.001) and 22 % (p = 0.001) of the variance, respectively.

The relative abundance of quantified known particulate metabolites (n = 79) also showed clear differences with time and treatment (Figure 3.2B, Table 3.3). Metabolites of the C treatment and pre-HVI timepoints for the LV and HV (T0, T12, and T24 for LV and T0 for HV) treatments were dominated by the known *Prochlorococcus* MED4 metabolites sucrose, glutamic acid, aspartic acid, glutathione (disulfide), and glutamine[32,38]. At the HVI timepoints and later, the LV and HV treatments contain increasing relative amounts of other compounds including glucosylglycerol, ectoine, and hypotaurine. Ecotine and hypotaurine are not detected in axenic *Prochlorococcus* MED4 cultures and likely originate from heterotrophic bacteria[32,38]. In contrast, glucosylglycerol abundance peaked specifically during periods of high viral infection, suggesting a link to the infection process. The observed metabolite changes reflect both viral infection of *Prochlorococcus* and the changing contribution of bacteria to community metabolite pools.

The total carbon contained in quantified metabolites (metabolite carbon) exhibited similar trends to total microbial biomass estimated using both ATP and flow cytometry approaches (microbial carbon), suggesting that the fraction of living carbon present in metabolites is relatively constant (Figure 3.9, Tables 3.1, 3.4). Metabolite carbon represented 4.0-10.5% (mean of 7.0%) of microbial carbon (Figure 3.2C), roughly three-fold greater than has benn estimated for metabolite carbon as a percentage of bulk marine particulate organic matter (POM) (1.7-2.5 %)[10,32]. This difference likely reflects the larger percentage of living biomass in this laboratory system relative to the oligotrophic marine environment, where living biomass is estimated to be just 30% of POM[124]. These findings suggest that metabolites represent a large fraction of the living carbon pool and thus viral remodeling of metabolomes can substantially alter carbon speciation and cycling in marine ecosystems.

We used a fold-change analysis to identify metabolites with statistically significant differences in abundance between samples with high viral infection (HVI) levels and control samples (Figure 3.3). We identified 157 (25%) out of 623 molecular features that were significantly different (false discovery rate-adjusted p-value < 0.05, $|\log2fc| > 0.5$) under HVI conditions, with 94 decreased (15% of total, 9 quantified, 24 tentatively annotated) and 50 increased (8% of total, 5 quantified, 6 tentatively annotated) (Figure 3.3, Figure 3.10, Table 3.5).

This change is similar to the infection of *P. aeruginosa* by phage YuA, where approximately 15% of untargeted metabolites were decreased and 5% were increased[113]. In contrast, phage infection of *P. aeruginosa*, *Sulfitobacter,* and *Synechococcus* WH8102 by a variety of other phages altered 0–71% of measured metabolites and primarily showed increased concentrations[113,114,118]. The diversity of metabolite responses across host-virus systems supports the idea that metabolic alterations under infection are variable and virus-specific[113,125,126].

Some metabolites altered in our study were also altered in other virocell metabolomics studies including the uridine diphosphate (UDP), the activated sugar UDP-N-acetylglucosamine, and the antioxidant glutathione (disulfide)[113,114]. Of unknown molecular features decreased under viral infection, 8 were tentatively identified as pyrimidine and pyrimidine derivatives, which may indicate altered nucleotide metabolism during infection, as seen in other studies[113,114,127] (Table 3.6). In addition to metabolites, P-SSP7 infection altered MED4 gene expression (29% decreased, 25% increased at timepoint 24), similar to previous studies[117] (Table 3.7). We used the metabolites identified in our fold change analysis, transcripts for genes related to these compounds, and the outputs of the matrix population model to develop a conceptual model of metabolic stressors and alterations experienced by *Prochlorococcus* during viral infection.

P-SSP7 infection results in reduced carbon fixation and drawdown of host carbon stores

Viral replication requires significant amounts of energy and carbon skeletons for nucleotide synthesis and the production of new virions. However, the sources of these resources (new photosynthate or use of cellular reserves) during P-SSP7 infection are not fully understood. The results of the matrix population model suggest that carbon fixation from photosynthesis was reduced under viral infection (two-way ANOVA, *p*-treatment < 0.001, *p*-time < 0.001, *p*-interaction = 0.002). The HV treatment displayed lower modeled population level carbon fixation rates during the 0–24-hour time window, and the low virus treatment displayed lower modeled carbon fixation rates during the 24–48-hour time window (Figure 3.4A, Table 3.8). Our model results are consistent with direct observations of decreased carbon fixation in the marine cyanobacterium *Synechococcus* during infection by the cyanophages S-RSM4 and S-PM2d, suggesting that decreased carbon fixation is a widespread phenomenon during marine cyanobacterial infection[128]. Exposure of *Prochlorococcus* MIT9313 to virally produced dissolved organic matter decreases carbon fixation, suggesting that the predicted reduction in carbon fixation may in part be due to uninfected cells responding to the lysate of other cells[129]. The matrix population model also suggested reduced population-level division rates (two-way ANOVA, *p*-treatment = 0.001, *p*-time = 0.035, *p*-interaction = 0.36) and population-level carbon loss rates (a combination of respiration and carbon release) (two-way ANOVA, *p*-treatment < 0.001, *p*-time = 0.001, *p*-value interaction < 0.001) during infection (Figure 3.4B, 3.4C, Table 3.8). Overall, our population model suggests that *Prochlorococcus* infected with P-SSP7 reduced their carbon fixation, growth, and exudation rates, implying that photosynthetic carbon fixation is not the primary source of energy and carbon structures for virion production.

Despite the modeled decrease in carbon fixation, ATP levels were not significantly different during viral infection (log2fc = -0.05, *p* = 0.53), suggesting that cellular energy levels were maintained. We therefore hypothesized that internal carbon stores were drawn down to fuel viral replication. In *Prochlorococcus*, fixed carbon and energy is stored primarily as carbohydrates such as the glucose polymer glycogen[130,131]. We observed increased expression of the glycogen degradation genes *glgP* and *glgX* (log2fc = 1.5, *p* < 0.001; log2fc = 2.0, *p* < 0.001

at 24 h, respectively) and a decrease in expression of genes related to glycogen production such as *glgA*, *glgB*, and *glgC* (log2fc = -1.5, $p < 0.001$; log2fc = -0.7, $p = 0.048$; log2fc = -1.4, $p < 0.001$ at 24 h, respectively) consistent with use of carbon stores (Figure 3.5). Expression of putative sigma factors from cyanophages of *Synechococcus elongatus* PCC 7942 has been shown to upregulate the expression of *glgP* and *glgX*, suggesting that use of host glycogen stores can fuel viral replication in a range of host-virus systems[116].

In addition to glycogen, *Prochlorococcus* accumulates the compatible solute sucrose at high concentrations as an additional energy and fixed carbon storage compound (Figure 3.2B)[10]. Sucrose could therefore be an important source of these resources during viral replication[10]. Indeed, we found that sucrose concentrations were significantly reduced in HVI samples relative to the control (log2fc = -1.15, $p < 0.001$). Despite this reduction, the abundance of transcripts encoding sucrose phosphate synthase (*spsA*: log2fc = 0.7, p = 0.033 at 24 h) increased. This suggests an overall increase in carbon flux through sucrose or increased demand for this compound as a source of energy and fixed carbon during infection[109]. Collectively, the use of glycogen and sucrose stores is consistent with a “leaching” type model for P-SSP7 infection for Prochlorococcus, where host metabolism is heavily modified and cellular resources are drawn down to fuel replication[125].

P-SSP7 infection results in the remodeling of Prochlorococcus compatible solute pools

Marine microbes accumulate organic compounds known as compatible solutes at high concentrations to preserve osmotic balance under saline conditions and fulfill other metabolic functions[11,30]. Prochlorococcus MED4 cells accumulate sucrose, glucosylglycerol, glutamic acid, aspartic acid, and glutamine as compatible solutes[32]. In addition to sucrose, we identified changes to aspartic acid and glucosylglycerol under viral infection. Aspartic acid concentrations were significantly decreased in our HVI samples relative to the control (log2fc = -0.64, p = 0.0015) (Figure 3.6). We identified changes in the expression of genes related to aspartic acid production at 24 h (*aspC*: log2fc = -1.2, $p < 0.001$; *ansA*: log2fc = 0.9, $p = 0.018$) and consumption (*nadB*: log2fc = 1.5, $p < 0.001$; *purA*: log2fc = -1.4, $p < 0.001$). Collectively, these results suggest decreased production of aspartic acid from oxaloacetate, increased production from asparagine, increased use for NAD synthesis, and decreased use for AMP synthesis. NADPH, produced from NAD, is understood to provide reductant for deoxynucleotide synthesis for phage genome replication[109]. Our results suggest that the drawdown of aspartic acid may be used to fuel this process[109]. AMP concentrations were also significantly reduced under viral infection (log2fc = -2.38, $p = 0.005$), potentially due to decreased synthesis from aspartic acid.

The drawdown of sucrose and aspartic acid under viral infection could lead to virocells being vulnerable to osmotic stress. However, glucosylglycerol, another compatible solute, increased significantly in concentration during viral infection (log2fc = 1.74, $p < 0.001$) (Figure 3.5)[132]. Glucosylgylcerol was initially present at very low concentrations but came to dominate the *Prochlorococcus* metabolome during periods of high viral infection (Figure 3.2B). Gene expression for the synthesis of glucosylglycerol from glycerol-3-phosphate and ADP-glucose (*ggpP*: log2fc = 0.6, $p = 0.011$) was increased at 36 h in the high virus treatments, suggesting that glucosylglycerol may be produced via this pathway. We hypothesize that glucosylglycerol is synthesized to compensate for the loss of osmotic stress protection from sucrose and aspartic acid as it has a lower per-molecule carbon and energy content than sucrose.

The combined drawdown of sucrose and aspartic acid, paired with the increase in glucosylglycerol, represents a major remodeling of virocell compatible solute pools. Compatible solute remodeling in phytoplankton has been observed in response to carbon, nitrogen, and vitamin B12 stress, where alterations to compatible solute pools facilitate adaptation to the stressor (e.g., compatible solutes with reduced C, N or B12 requirements)[11,12]. Our results suggest that P-SSP7 infection alters the composition of Prochlorococcus compatible solute pools in response to the demands for fixed carbon and energy required for virion production.

P-SSP7 infection leads to changes in the interconnected pB12, SAM, and antioxidant metabolisms

Vitamin B12 is an essential cofactor for most cells where it is involved in several critical reactions, including methionine synthesis, deoxyribonucleotide synthesis, and production of succinyl-CoA during the breakdown of fatty acids and amino acids[133]. Concentrations of pB12, the form of vitamin B12 used by *Prochlorococcus*, were significantly decreased under viral infection (log2fc = -0.82, $p = 0.012$, Figure 3.10). In contrast, we observed a significant increase in the expression of genes encoding enzymes in the pB12 synthesis pathway (*cobO*, *cobJ*, *cobI*, *cobH*, *cobB*, and *cobN*; log2fc > 0.5 and $p < 0.05$ at 12 and/or 24 h) in the HV treatment at 12 and/or 24 hours (Fig 3.7, Figure 3.12). Decreased cellular pB12 quotas alongside increased expression of genes to produce pB12 suggests that infected *Prochlorococcus* cells did not have sufficient pB12 to perform pB12 dependent reactions and therefore were experiencing pB12 stress.

Consistent with pB12 stress, S-adenosylmethionine (SAM), was decreased under viral infection (log2fc = -0.73, $p = 0.007$), implying infection induced changes to the methionine cycle (Figure 3.7). In agreement with our metabolite measurement, transcripts for the use of SAM increased (*speH*: log2fc = 1.4, $p = 0.003$ at 24 h) and transcripts related to the synthesis of SAM decreased (*metK*: log2fc = 1.9, $p < 0.001$ at 24 h) suggesting both decreased SAM production and increased SAM consumption. Transcripts for methionine synthase were also increased (*metH*: log2fc = 1.3, $p < 0.001$ at 24 h), suggesting further disruption of the methionine cycle. The decrease in SAM concentrations suggests a reduction in the cell's capacity to carry out methylation reactions[134]. One important pathway that requires SAM as a methyl donor is biotin (vitamin B7) synthesis. While we did not measure biotin in our analysis, the precursor desthiobiotin was depleted (log2fc = -1.89, $p = 0.0036$) under infection, suggesting biotin production may be reduced as a consequence of decreased methylation capacity (Figure 3.3, Figure 3.10).

The methionine cycle is involved in the production of the antioxidant metabolite glutathione through its role in the synthesis of cysteine (Figure 3.7). We found that glutathione (disulfide) concentrations were decreased in our HVI samples (log2fc = -1.11, $p < 0.001$). The expression of genes related to the synthesis of glutathione (*gshB*: log2fc = -1.5, $p < 0.001$ at 24 h) and the reduction of glutathione disulfide back to glutathione post ROS scavenging (*gorA*: log2fc = -0.8, $p = 0.025$ at 24 h) were significantly decreased, providing further evidence for a decrease in antioxidant metabolism activity (Figure 3.7). Another antioxidant, ophthalmic acid, was also significantly decreased (log2fc = -0.75, $p = 0.0032$). Ophthalmic acid is a structurally similar tripeptide to glutathione and is synthesized by the same enzyme (GshB), suggesting its synthesis was also decreased[135]. Collectively, these results suggest that during PSSP-7 infection of *Prochlorococcus,* there is an overall decrease in the production and pools of antioxidant metabolites, reducing the virocell's ability to combat oxidative stress.

The biochemical links between pB12, the methionine cycle, and antioxidant metabolism suggest a connection between the observed changes (Figure 3.7). pB12 availability is directly related to the production of SAM through serving as a cofactor in the production of methionine, and SAM is directly related to the production of glutathione (disulfide) through the transulfuration pathway (Figure 3.7). Therefore, a decrease in pB12 could lead to a decrease in SAM, which in turn leads to a decrease in glutathione (disulfide). A culture study examining the metabolic consequences of vitamin B12 limitation on the diatom *Thalassiosira pseudonana* found decreases in both SAM and glutathione (disulfide), supporting this interpretation [12]. The widespread alterations to gene expression related to all three metabolic processes also likely contribute to the observed changes.

The lower observed pB12 concentrations alongside increased expression of pB12 synthesis genes suggests that pB12 degradation was occurring during viral infection. While pB12 and other forms of vitamin B12 are light sensitive, light levels were the same across treatments, suggesting photodegradation is not the primary cause (although viral infection may alter photoprotective mechanisms)[136]. Reactive nitrogen species (RNS) and reactive oxygen species (ROS) can also damage cobalamins[137,138]. We hypothesize that the unifying driver behind the metabolic changes identified here is increased oxidative stress caused by an imbalance in ROS production and reduction during infection. Populations of ammonia-oxidizing archaea experiencing reactive nitrogen stress also display both increased transcripts for B12 synthesis and decreased pools of bioavailable B12, supporting our interpretation[138,139]. We examined the expression of genes involved in antioxidant defense to look for further evidence of oxidative stress. We found that expression of the antioxidant gene *sodN* was also decreased under viral infection (log2fc = -1.7, $p < 0.001$), suggesting further decreases in antioxidant capacity (Table 3.7)

ROS production and the induction of oxidative stress are common under viral infection across diverse virus/host systems, including eukaryotic phytoplankton, plants, and humans[140–142]. Interestingly, the virion production process of P-SSP7 is sensitive to external sources of ROS, with increased oxidative stress causing increases in both empty capsid production and DNA mispackaging[143]. This sensitivity of P-SSP7 to ROS aligns with our metabolomic and transcriptomic data, which suggest a decrease in antioxidant capacity during infection. While direct measurements of ROS are required to confirm our hypothesis, we find widespread evidence to support our interpretation that oxidative stress contributes to, and is potentially a result of, the observed changes to pB12, SAM, and antioxidant metabolisms. Our results also suggest that a positive feedback cycle may emerge where oxidative stress-driven decreases in pB12 lead to decreases in cellular antioxidants, resulting in further increases in oxidative stress.

An important consideration for understanding the role of ROS and antioxidant metabolism in *Prochlorococcus* is that *Prochlorococcus* strain MED4 lacks catalase and relies instead on “helper” bacteria to reduce hydrogen peroxide[144]. Therefore, P-SSP7 virocells may also be reliant on "helper” bacteria to compensate for the infection induced decreases in antioxidant metabolism[143]. However, heterotrophic bacteria were present in our experiment and signs of oxidative stress were still observed, suggesting that “helper” bacteria are unable to fully counteract the infection induced oxidative stress.

Insights into the roles of AMGs during P-SSP7 infection:

Our observations of metabolic remodeling provide insights into the role of AMGs in P-SSP7 infection of *Prochlorococcus*. P-SSP7 encodes AMGs for a photosystem II protein (PsbA), a high-light inducible protein, a transaldolase involved in the pentose phosphate pathway (TalC), and a class I ribonucleotide reductase (nrd)[117,145,146]. These AMGs are transcribed at the beginning of infection, suggesting they are important for successful viral replication, but their exact roles and selective advantages are unclear[117]. The photosystem II gene *psbA* results in the production of these proteins during infection, suggesting increased photosynthesis rates[115]. However, our matrix population model predicts that carbon fixation was reduced during infection, suggesting that increasing carbon fixation is not the primary role for this AMG (Figure 3.4). Therefore, we hypothesize that *psbA* primarily increases photosynthesis for the purpose of producing energy and reductant for nucleotide synthesis. The presence of *talC* as an AMG supports this interpretation, as it catalyzes conversion of ribose 5-phosphate to glucose 6-phosphate as part of the pentose phosphate pathway, directing ribose 5-phosphate away from the Calvin cycle and carbon fixation[109].

Our results also provide insights into the role of the class I ribonucleotide reductase. Ribonucleotide reduction is a key step in the synthesis of deoxyribonucleotides, an essential process for successful viral replication. Unlike the class II ribonucleotide reductases encoded by *Prochlorococcus*, class I ribonucleotide reductases do not require pB12 or B12[146]. During infection, we identified metabolite and transcriptional changes that are consistent with pB12 stress, suggesting that pB12 could limit deoxyribonucleotide synthesis with class II ribonucleotide reductases (Figure 3.7, Figure 3.12). Class I ribonucleotide reductases would not be impacted. The ribonucleotide reductase encoded by P-SSP7 also lacks a common tyrosine residue that makes other ribonucleotide reductases vulnerable to inactivation by nitric oxide (a RNS), suggesting it is also resistant to both RNS and ROS[146]. Therefore, the class I ribonucleotide reductase encoded by P-SSP7 appears adapted to both the observed pB12 stress and implied oxidative stress present during infection, facilitating viral DNA synthesis and providing a clear benefit to P-SSP7. The majority of sequenced cyanophage encode class I ribonucleotide reductases, suggesting that many cyanophages may experience similar metabolic conditions (low pB12 availability and high levels of oxidative stress) during infection[146].

Implications for Environmental Infection Dynamics

In the oligotrophic gyres where *Prochlorococcus* dominates, diel processes play a key role in structuring organism metabolisms and community level processes, driven by the strong day-night cycle of light availability and its impact on primary productivity[147–149]. Many metabolites identified as being modified by viral infection in this study, including sucrose, aspartic acid, glutathione (disulfide), and SAM were all identified as displaying diel periodicity, peaking in the evening or at night in a study at Station ALOHA in the North Pacific Subtropical Gyre[10]. Viral infection of *Prochlorococcus*, both from direct measurements and inferred from transcriptional activity, also displayed diel periodicity during this same sampling campaign, with infection peaking in the evening and at night[96,150].

This diel infection pattern is hypothesized to be explained by viruses preferentially infecting when host genome replication is happening prior to cell division, leading to more resources for phage genome replication[96,150]. Our results suggest that this late-evening infection is also connected to the availability of host metabolic resources and the diel variation in cellular stressors. In the evening, when cellular carbon and energy stores (ex. glycogen, sucrose) are at their peak[10,151], may represent an optimal time for infection. Additionally, oxidative stress is likely lower at night when light stress and photosynthesis are not occurring, reducing cellular demand for antioxidant type molecules[152]. Previous studies have noted striking similarities between the metabolic effects of cyanophage infection and the dark-phase metabolism of uninfected cyanobacteria, particularly regarding decreased carbon fixation and increased use of carbon stores[110]. Our results provide further evidence in *Prochlorococcus* to support these observations and expand its scope to incorporate compatible solutes, antioxidants, and vitamins. However, in culture, P-SSP7 does have lower replication rates in the dark than in the light, suggesting daytime infection is also advantageous[115,153].

Implications for Microbial Communities and Biogeochemical Cycles

Through both the release of metabolites from infected cells (e.g., via exudation or leakage) and from cell lysis, viruses contribute to labile dissolved organic matter (DOM) and dissolved vitamin pools that support marine microbial communities[147,154]. Our results show that the composition of metabolites is significantly different in virally infected cells than in healthy cells, with infected cells showing depletion of sucrose, aspartic acid, and antioxidants, and a striking increase in glucosylglycerol (Figure 3.2B). Therefore, we hypothesize that the metabolite composition of the lysate and exudate from P-SSP7 infected MED4 cells will be compositionally distinct from healthy cells, as has been found in other cyanobacterial lysate studies[106,155]. These altered metabolite pools may select for different heterotrophic bacterial communities and may drive chemotaxis towards infected cells, leading to a distinct sub-community supported by viral lysis products[17,86,106,156]. The matrix population model also suggested reduced carbon fixation and carbon loss (respiration + exudation) rates during infection, suggesting that pre-lysis, *Prochlorococcus* virocells may contribute less to both labile POM and DOM pools than their uninfected counterparts.

Changes to B vitamin pools and production during viral infection may also influence microbial communities. Auxotrophies for B vitamins such as vitamin B12 (approximately 50% of phytoplankton and some bacteria) and the biotin precursor desthiobiotin (8.5-13.1% of the prokaryotic community) are prevalent in the ocean[133,157–160]. While only B12 auxotrophs capable of remodeling pB12 into B12 would be affected (approximately 11% of eukaryotic phytoplankton strains), genes involved in pB12 remodeling are present in alpha- and gammaproteobacteria in the Northwest Atlantic and are actively transcribed by eukaryotes in the NPSG[147,161,162]. This suggests that infection induced reductions in pB12 and desthiobiotin pools would impact community level vitamin availability in *Prochlorococcus*-dominated regions. Therefore, vitamin metabolism remodeling is another way in which viral infection may influence microbial community composition beyond being a direct top-down pressure.

3.5 Conclusions:

This study provides a comprehensive view of the metabolic remodeling that occurs during viral infection of the globally important marine cyanobacterium *Prochlorococcus* MED4 by the cyanophage P-SSP7. Using a combination of metabolomics, transcriptomics, and population modeling, we identify widespread and often interconnected changes, including a decrease in carbon fixation, depletion of carbon and compatible solute reserves, disruption of vitamin B12 and methionine metabolism, and a compromised antioxidant defense system. Based on these findings, we hypothesize that carbon and energy limitation and oxidative stress are major stressors driving the observed metabolic reprogramming during infection. The extent of these metabolic changes is not directly predicted by P-SSP7's AMGs, suggesting that AMGs alone are insufficient to capture the full range of viral effects on host metabolism[113,117,145]. While our study provides valuable insights, it is important to acknowledge that we cannot definitively distinguish between host defense responses and metabolic alterations caused by the virus. Nonetheless, the metabolic remodeling observed in this study has significant implications for our understanding of the roles of AMGs in metabolic remodeling, the dynamics of viral infection in the environment, and the role of viruses in shaping microbial community composition and biogeochemical cycles.

3.6 Acknowledgements:

We thank Tara Clemente, Dror Shitrit, Michael Carlson, and Julia Weissenbach for assistance with sample collection, Everetta Rasyid for assistance with metabolomics laboratory analysis, and Nhi Vo and Konnor Von Emster for assistance with RNASeq data processing. We thank Will Kumler, Frank Xavier Ferrer-González, and Max Jahns for valuable conversations throughout data analysis and interpretation. This work was supported by grants from the Simons Foundation (LS award ID 385428 to AEI; SCOPE award ID 3721225 to STD; SCOPE award ID 721252 to DMK; LS award ID 721254 to DL; SCOPE award ID 329108 to DL; and award ID 649394 to SMK).

3.7 Figures:

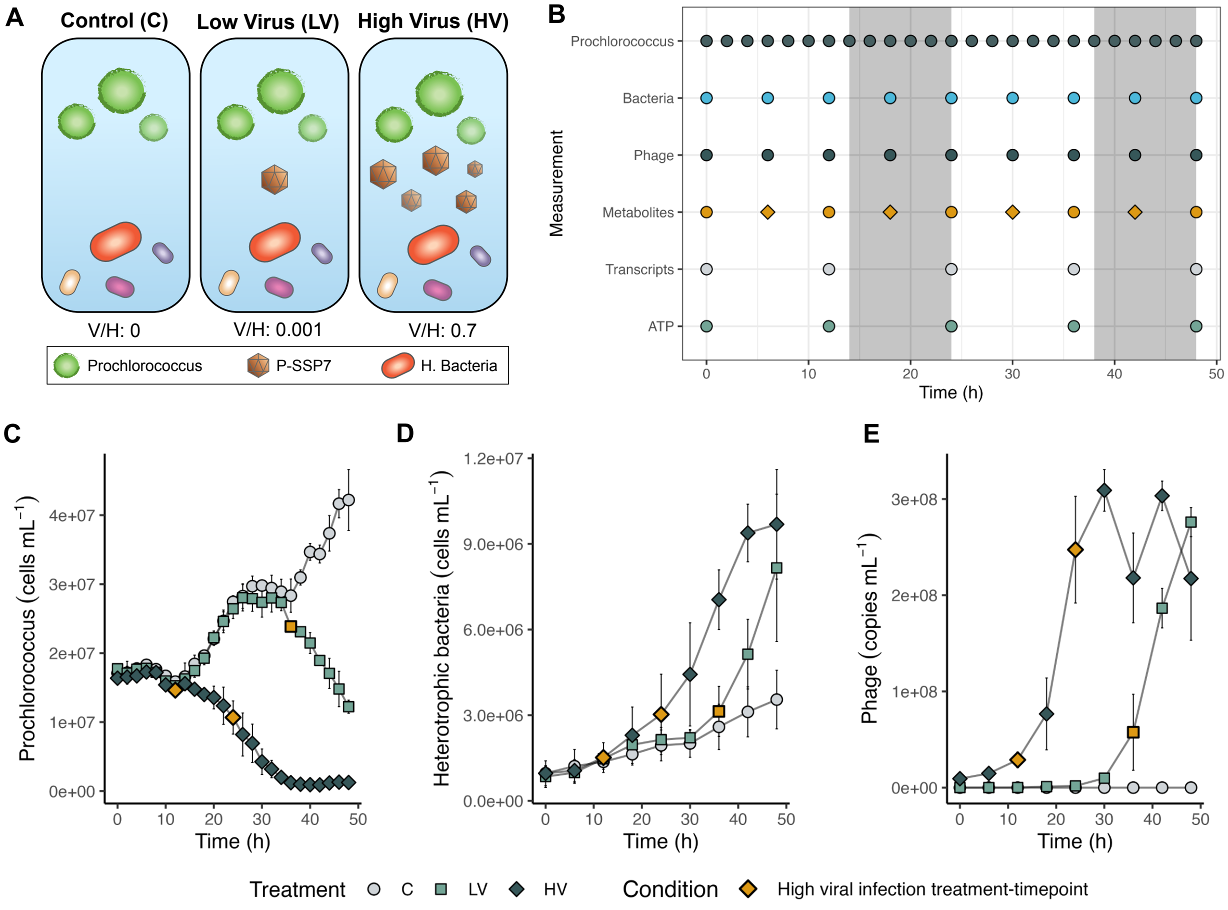

Figure 3.1: Visual overview of experimental design showing treatments (A) and sampling timepoints for all data included in this study (B). Light and dark shading indicates day and night. The abundance of Prochlorococcus (C), heterotrophic bacteria (D), and phage P-SSP7 (E) estimated through flow cytometry (C, D) and qPCR (E). Points show mean values of three biological replicates and error bars shows standard deviations of the mean. Color and shape indicate treatment. Points highlighted in yellow are the points classified as having high levels of viral infection and low levels of bacterial biomass that are used in the metabolite fold change analysis. In panel A, H. Bacteria is heterotrophic bacteria and bacteria of different colors are shown to emphasize that a variety of bacterial groups were present.

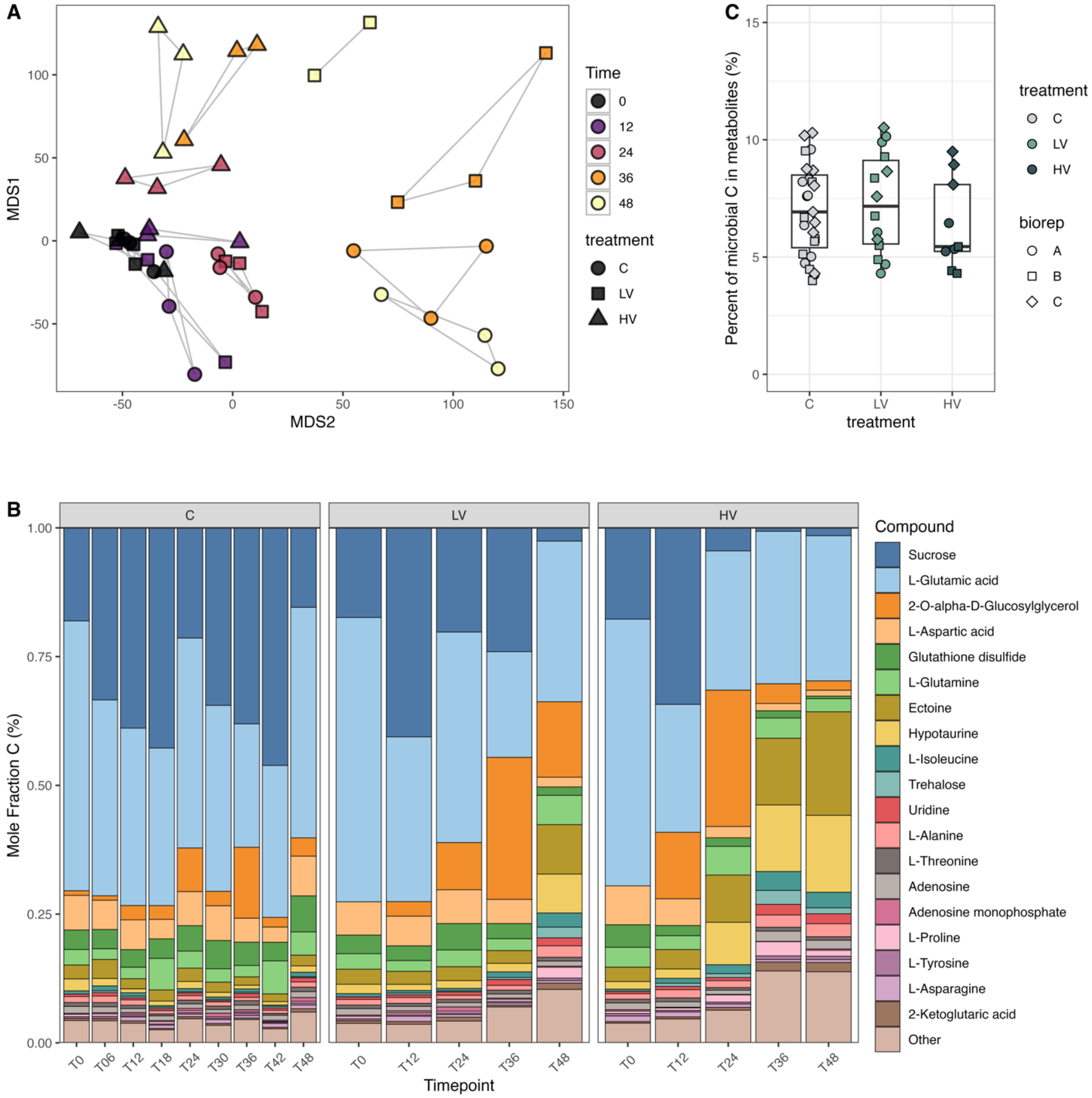

Figure 3.2: NMDS plot of untargeted particulate metabolomes (A) with color denoting time and shape denoting treatment. Biological triplicates are linked by hulls. NMDS plot has a stress value of 0.1 and the projection was found to be significant through a Monte Carlo analysis (p = 0.01). Stacked bar charts show the mean relative abundance of quantified targeted metabolites in mole fraction carbon space (B. The percent of flow cytometry estimated microbial carbon (in Prochlorococcus and heterotrophic bacteria) in quantified metabolites for individual samples is shown as points and for each treatment as boxplots (C).

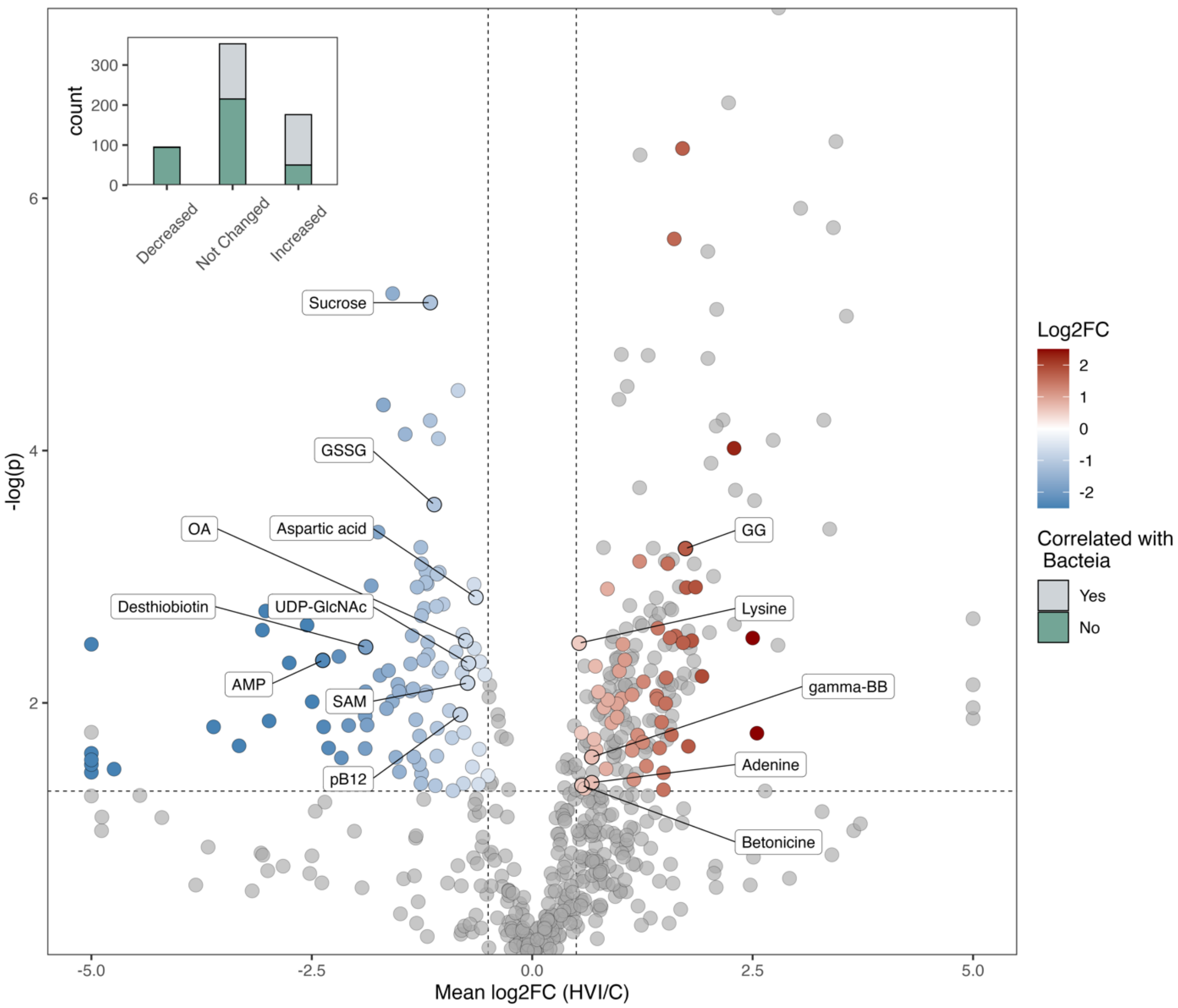

Figure 3.3: Mean log2 fold change (log2FC) of molecular features in “high viral infection” (HVI) samples relative to their replicate-matched controls (C). fdr-adjusted p-values are from a one-tailed t-test. Gray circles either were not significantly different or were positively correlated with heterotrophic bacterial abundance. Colored points show metabolites significantly altered under viral infection with red indicating increases (log2fc > 0) and blue indicating decreases (log2fc < 0). Known quantifed metabolites are that were significantly different are annotated with labels. Dotted lines indicate thresholds for significance ($\log2FC > 0.5$, $p < 0.05$). Abbreviations are as follows: GSSG: glutathione disulfide; OA: ophthalmic acid; UDP-GlcNAc: UDP-N-acetylglucosamine; AMP: adenosine monophosphate; SAM: S-adenosylmethionine; pB12: pseudocobalamin; GG: glucosylglycerol; gamma-BB: (3-Carboxypropyl)trimethylammonium (gamma-butyrobetaine). Inset shows the count of molecular features that increased, decreased, or unchanged under high viral infection conditions with color indicting whether features were significantly positively correlated with heterotrophic bacteria (gray) or not (green).

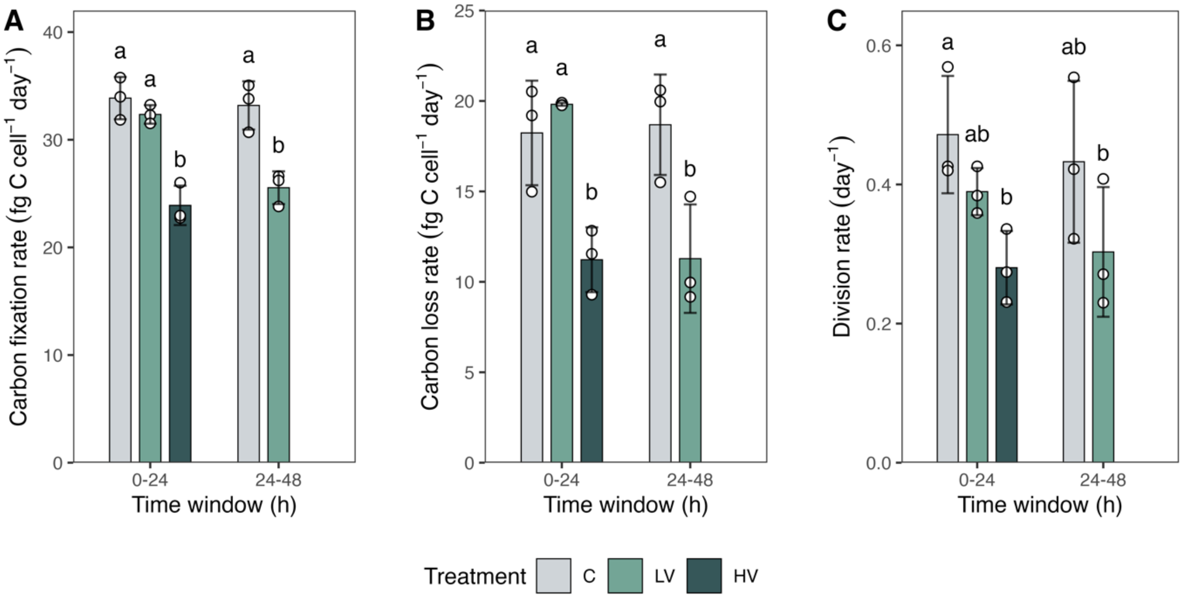

Figure 3.4: Estimated (matrix population model) rates of carbon fixation (A), carbon loss (B), and cell division (C). Error bars represent one standard deviation of the mean of estimated rates from biological triplicates, with individual estimates indicated with white circles. Color indicates treatment. Letters indicate significance groupings from a two-way ANOVA where samples labeled with different letters are significantly different from each other ($p < 0.01$).

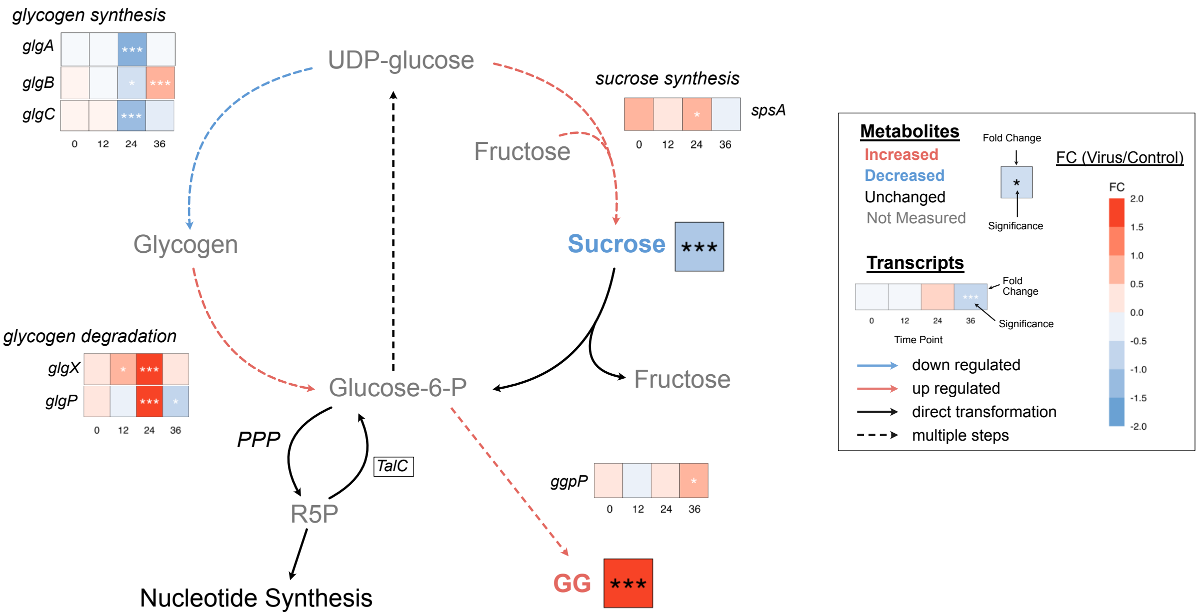

Figure 3.5: Metabolic map depicting changes to glycogen, sucrose, and glucosylglycerol (GG) metabolism. Heatmaps show changes in gene expression in the HV treatment relative to the control. Boxes besides metabolites show fold change of metabolites in “high viral infection” treatments relative to the control. Color of arrows and metabolites indicates change under viral infection. The box around TalC indicates that it is an auxillary metabolic gene encoded for by PSSP-7. Asterisks indicate significance (“*”: 0.05 > p > 0.01; “**”: 0.01 > p > 0.005; “***”: p < 0.005).

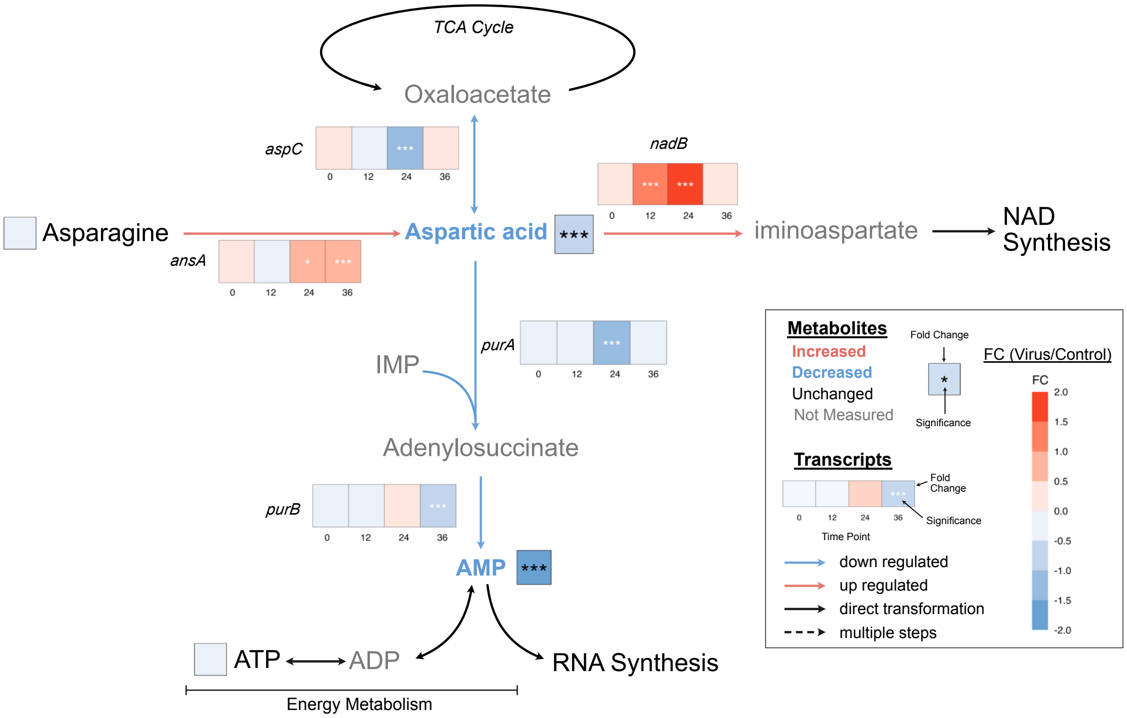

Figure 3.6: Metabolic map depicting changes to aspartic acid and adenosine monophosphate (AMP) metabolism. Heatmaps show changes in gene expression in the HV treatment relative to the control. Boxes besides metabolites show fold change of metabolites in “high viral infection” treatments relative to the control. Color of arrows and metabolites indicates change under viral infection. Asterisks indicate significance (“*”: 0.05 > p > 0.01; “**”: 0.01 > p > 0.005; “***”: p < 0.005).

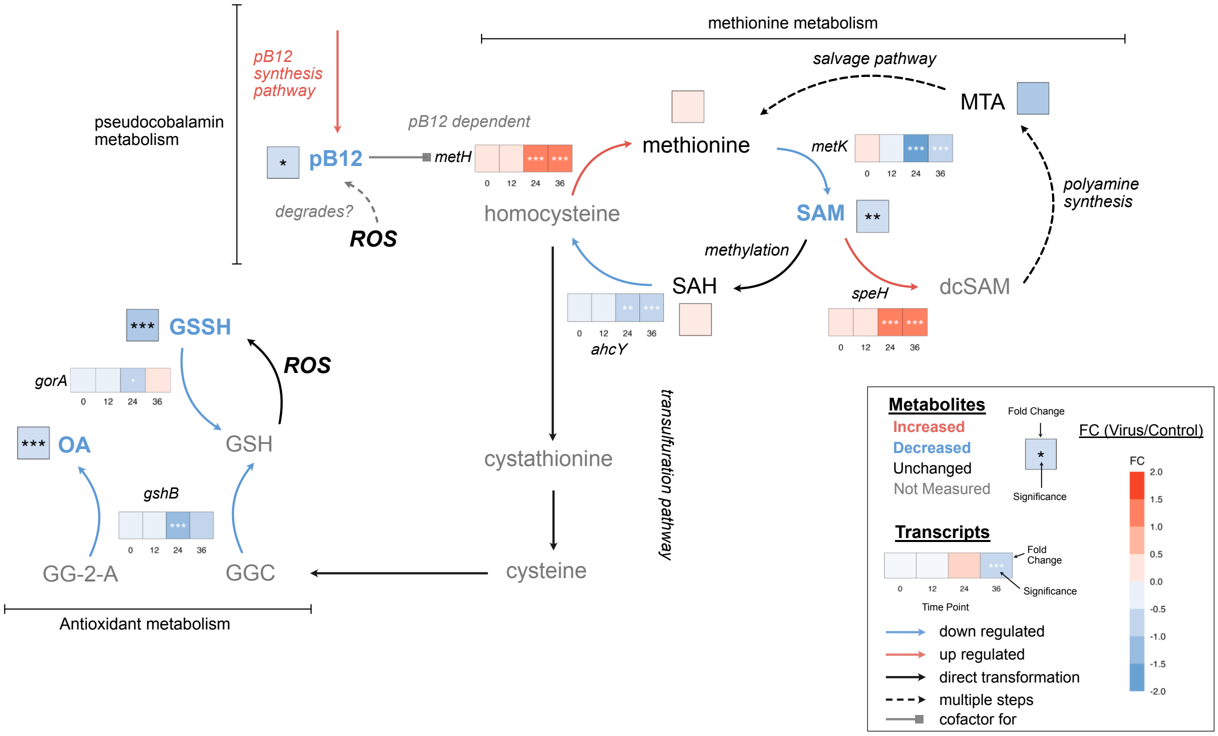

Figure 3.7: Metabolic map depicting changes to pseudocobalamin metabolism (pB12), the methionine cycle and S-Adenosylmethionine (SAM) metabolism, and antioxidant metabolism (glutathione (GSG) and ophthalmic acid (OA)). Heatmaps show changes in gene expression in the HV treatment relative to the control. Boxes besides metabolites show fold change of metabolites in “high viral infection” treatments relative to the control. Color of arrows and metabolites indicates change under viral infection. Asterisks indicate significance (“*”: $0.05 > p > 0.01$; “**”: $0.01 > p > 0.005$; “***”: $p < 0.005$).

3.8 Supplemental Text:

Supplemental Methods:

Materials:

Metabolite Standards were acquired from Sigma Aldrich, Santa Cruz Biotechnology, Toronto Research Chemicals, Cambridge Chemicals, Spectrum Chemical, Cambridge Isotope Laboratories, or were provided by other labs. Details of standards and origins are provided in [25].

Culturing:

Prochlorococcus MED4 was grown in Pro99 medium with 12 mM bicarbonate. Cultures were incubated at 21°C with slow stirring and a 14:10 light:dark cycle with a cool white light source with 50 $\mu mol \cdot photon \cdot m^{-2} \cdot s^{-1}$. Three independent cultures for each biological replicate were maintained for multiple generations prior to the experiment. MED4 cultures were not axenic and contained 4-7 % heterotrophic bacterial cells at the start of the experiment. Cyanophage P-SSP7 was propagated on exponentially growing cells of *Prochlorococcus* MED4 under the same conditions it was grown at for the experiment.

Biomass Estimation for Prochlorococcus and Heterotrophic Bacteria:

Flow cytometry details for Prochlorococcus and two populations of heterotrophic bacteria (“small” and “large”) are reported in Lindel et al. (in prep). Cell size was estimated from individual particle using forward light scatter (FSC) measurements using Mie light scattering theory applied to a simplified optical model, as described in[121]. This model was calibrated using polystyrene beads of known refractive index and diameter, demonstrating strong agreement with Mie-predicted particle diameters. Assuming spherical cell morphology, cell volumes were calculated and converted to carbon quotas using the allometric equation: pgC $\text{cell}^{-1} = 0.261 \times \text{Volume}^{0.860}$. While not explicitly validated for heterotrophic bacteria, this approach yields carbon quota estimates consistent with those derived from cyanobacterial cultures (see [121] for details).

<u>Matix population modeling:</u>

Model Parameters:

We applied a size-structured matrix population model following the approach described in [122] to estimate phytoplankton division rates. This category of model has demonstrated its suitability for estimating phytoplankton division rates [163–168]. It mechanistically simulates the evolution of the size distribution over a day/night cycle. The model discretizes cell counts across size classes and time (using the parameter Δv) and time (using the parameter Δt). The model quantifies the rates of the three key biological processes (cell division, carbon fixation, and carbon loss), which are represented by the parameter vector $\theta = (\theta\delta, \theta\gamma, \theta\rho, \omega 0)$. These parameters, along with a concentration parameter σ (which allows for overdispersion in the data), determine the dynamics of the model. Below is a breakdown of the modeled processes:

1. Cell Division (δ): Division rates are size-dependent, reflecting the probability of cells transitioning to subsequent life-cycle stages. The model assumes a monotonically increasing division rate as a function of cell size.

2. Carbon Fixation (γ): Carbon fixation rates are determined by light availability (PAR measurements) and cellular carbon content. This captures diel cycle dynamics and phytoplankton growth rates. The model incorporates a power-law relationship between carbon fixation and cell size.

3. Carbon Loss (ρ): The model accounts for carbon loss through respiration and exudation, estimating rates separately for each size class. This reflects cell shrinkage due to metabolic processes other than division.

Prochlorococcus size distribution was discretized into 30 size classes based on the log of their carbon quotas, ranging from 80 to 260 fgC cell^{-1}. This range was selected to fully encompass the variation in *Prochlorococcus* cell size observed in the dataset. We chose the volume bins and a time step of 10 minutes to ensure cells would likely only undergo a single transition at a time within the model. Model source code and workflow used in this study are publicly available on Github (https://github.com/seaflow-uw/mpm-workflow).

Data Analysis:

We used Bayesian inference, implemented through the Stan software package and Hamiltonian Monte Carlo (HMC) algorithms, to process the data. Six HMC chains were run for 2000 Markov Chain Month Carlo iterations. This approach strikes a balance between computational efficiency and the complexity of our model. A list of model parameters and priors is available in [122]. We implemented rigorous quality control measures to ensure accuracy and reliability of our model estimates, only including good model convergence (R-hat <1.1) to guarantee reliability of model-derived estimates.

Metabolomics:

Metabolite Sampling:

Particulate metabolites were sampled every 12 hours using gentle vacuum filtration onto 47 mm 0.2-mm polytetrafluoroethylene (PTFE) omnipore filters (Omnipore Membrane Filters, Merck Millipore Ltd). Glass and polysulfone filtration rigs were used. Glass rigs were combusted at 450 ˚C for 4 hours before and between experiments. The polysulfone filtration setups were soaked in 10% HCl for 24 hours and triple rinsed with MiliQ water between experiments. In between sampling different treatments, timepoints, and biological replicates, the filtration setups were triple rinsed with MiliQ water, rinsed with 10% HCl, and then triple rinsed again with MiliQ. After filtration, samples were wrapped in combusted foil and flash frozen in liquid nitrogen before storage at -80˚C until analysis. Sample volumes ranged from 50 – 300 mL to account for variations in biomass throughout the experiment and filtration times ranged from 5–15 minutes. Blanks were collected by re-filtering the filtrate from T0 and T48 samples from the HV treatments as well as the uninoculated culture media, using the same cleaning and filtering approaches. These blanks are intended to serve as a control against metabolite signal introduced by metabolites and inorganic ions contained in any residual media that remains on the filters following filtration or that adsorbs onto the filters.

Metabolite Extraction:

Particulate metabolites were extracted using a modified Bligh and Dyer two-phase extraction as previously described in Boysen et al. [24]. Briefly, filters were extracted using bead-beating in 1:1:2 methanol:water:dichloromethane at -20 ˚C resulting in the separation of more-polar metabolites into the aqueous (water-methanol) fraction and more-nonpolar metabolites into the organic fraction (dichloromethane). After extraction and separation, metabolites were dried down under N_2 gas. Aqueous fractions were resuspended in 400 mL water and organic fractions were resuspended in methanol. Only the aqueous fractions were analyzed for this study. Isotopically labeled internal standards (detailed in Supplemental Table 9) were added during extraction (extraction standards) and during reconstitution (injection standards) for monitoring extraction, injection, and instrument performance, to normalization, and quantification [24]. After reconstitution samples were kept at -80˚C until analysis.

Metabolite data acquisition:

Metabolite data were collected using established liquid chromatography mass spectrometry (LC-MS) methods for targeted and untargeted metabolomics and targeted vitamin analysis in seawater and culture sample [24,25]. Three LC-MS approaches were employed to maximize metabolite coverage: 1) normal-phase hydrophilic interaction liquid chromatography (HILIC) paired with orbitrap mass spectrometry to measure more-polar compounds (in polarity switching mode resulting in positive (Pos) and negative (Neg) data outputs), 2) reversed-phase cyano chromatography (RP) paired with orbitrap mass spectrometry to measure less polar compounds (with just positive ionization), and 3) reversed-phase chromatography with a cyano column paired with a triple quadrupole (TQS) mass spectrometer for targeted vitamin analysis (Vitamin). Targeted and untargeted data for the HILIC and RP chromatography fractions was obtained using a Thermo Q-Exactive (QE) mass spectrometer using the methods detailed in Sacks et al. [25]. For untargeted metabolite annotation, fragmentation spectra were collected on pooled samples using a data dependent acquisition approach. Data dependent MS/MS scans were collected on pooled samples with separate injections performed for HILIC positive, HILIC negative, and RP positive fractions at three fragment collision energies (20, 35, and 50 eV). Identical chromatography and mass spectrometry conditions were used as for MS1 analysis. The top five most abundant ions in each cycle were fragmented with a resolution of 30,000 with a dynamic exclusion time of 20 seconds. Vitamin data were acquired using a triple quadrupole (TQS) mass spectrometer using selected reaction monitoring (SRM) as detailed in Heal et al. [169,170].

Metabolite Data Processing:

Data processing was performed using a hybrid targeted/untargeted approach. The targeted approach allowed us to accurately identify and quantify metabolites for which we have standards. The untargeted approach generates a less biased and more comprehensive list of molecular features for exploration but is not quantitative, nor are the features confidently identified. Details on compounds and standards incorporated into the targeted analysis are available in Sacks et al. [25].

Hybrid targeted and untargeted metabolite data processing:

Data files from the QE were converted to .mzMLs using MSConvert [75]. Targeted compounds were manually integrated in pooled samples and isotopically labeled internal standards were manually integrated in all samples using Skyline [76]. These results were then exported and used to generate a batch specific exact mass and retention time (RT) file for all targeted metabolites in our in-house compound library. MS-DIAL (Version 4.90) was then used for untargeted peak-picking and integration using the settings detailed in Supplemental Table 10 [77]. The batch-specific exact mass and RT values were used to identify targeted compounds in the MS-DIAL analysis with a higher degree of accuracy compared with non-batch specific values as they allow for much narrower RT and mz ppm flexibility and account for the batch-specific RT variability that occurs when using HILIC chromatography. Targeted compounds that were either manually identified as present but were not detected by MS-DIAL or targeted compounds identified in MS-DIAL but poorly integrated (greater than 10% difference in mean peak area between the manually integrated pooled samples and the MS-DIAL integrated pooled samples) were removed from the MS-DIAL data. These compounds were then manually integrated across all samples in Skyline and combined with the MS-DIAL dataset to form the full peak list. HILIC positive ion mode, HILIC negative ion mode, and RP-cyano fractions were all analyzed through the hybrid targeted/untargeted pipeline separately. Vitamin data were manually integrated in Skyline.

Metabolomics Data Quality Control, Normalization, and Adduct Detection:

Metabolite data was subjected to quality control to remove low quality molecular features (MFs) and normalization to address variability due to matrix and instrument variability. First, MFs not meeting minimum standards for peak area (40,000 for HILIC, RP, and Vit) in at least 12 samples were removed. Best-Matched Internal Standard Normalization was then applied to all features in the HILIC and RP datasets [24]. Vitamin data was normalized to a matched internal standard for Vitamins B1 and B2. If a matched internal standard was not available for a vitamin feature, no normalization was performed. A final MFs feature list was then generated by imposing thresholds for minimum relative standard deviation (0.3) in the pooled samples (to remove features displaying high variability in repeat injections of the same sample) and a minimum number of triplicates (present in all three replicates of 7 sets of triplicates). MFs were defined as present in a sample if they had a minimum peak area greater than 40,000 and the peak area was more than 3 times higher than the average value for the molecular feature in the blanks.

Untargeted mass spectrometry datasets contain large numbers of MFs that can be attributed to multiple isotopologues or adducts of a single compound. Additionally, when working with both positive and negative mode data, MFs can be duplicated between these datasets. Isotopologues, adducts, and shared positive/negative mode signals were identified in our curated final peak list by identifying MFs with a shared retention time (+/- 3 s relative to MF), m/z values corresponding to a predicted m/z values (+/- 5 ppm) for adducts and isotopologues, and peak area correlations across samples of 95% or greater. We decided between MFs duplicated in HILIC Pos and HILIC Neg mode based on which MF had the lower relative standandard deviation in the pooled sample. We removed duplicated MFs and 13C adducts but left other adducts in so as not to remove any actual features potentially misannotated as adducts. This resulted in a final metabolomics dataset containing 253 HILIC Positive features (42 known, 211 unknown), 136 HILIC Negative features (17 known, 119 unknown), 228 RP features (14 known, 214 unknown), and 6 Vitamin features (all known).

Untargeted Metabolite Annotation:

Annotation of untargeted MFs was performed using SIRIUS 5 (version 5.5.7) [123,171,172]. HILIC Pos, HILIC Neg, and RP DDA pooled samples were analyzed separately using default settings. Annotations were matched to MFs by exact mass (+/- 5 ppm) and retention time (+/- 20 s). Only annotations related to the curated MFs that passed quality control thresholds were kept resulting in annotations for 35 HILIC Pos MFs, 15 HILIC Neg MFs, and 35 RP MFs. These annotations are shown in Supplemental Table 6.

Quantification of Targeted HILIC and RP Metabolites:

As in Heal et al. [32], targeted metabolites were quantified by comparing their peak areas to the peak areas of authentic standards at a known concentration in water and in the sample matrix. The following equation was used to calculate the absolute concentration in the experimental samples:

$$\mathrm{Concentration_{samp}} = \mathrm{PA_{samp}} \times \frac{1}{\mathrm{RF}} \times \frac{1}{\mathrm{RFratio}} \times \mathrm{Vol_{vial}} \times \frac{1}{\mathrm{Vol_{samp}}}$$

where PA_{samp} is the normalized peak area, RF is the response factor of the standard in water, RFratio is the ratio of the response factor of the standard in water to the response factor of the standard in the sample matrix, Vol_{vial} is the volume of the extracted sample after reconstitution, and Vol_{samp} is the volume of the sample collected during the experiment.

For samples with matched internal standards included in the analysis, quantification was performed using these internal standards using the following equation:

$$\mathrm{Concentration}_{\mathrm{samp}} = \mathrm{PA}_{\mathrm{samp}} \times \frac{\mathrm{Concentration}_{\mathrm{IS}}}{\mathrm{PA}_{\mathrm{IS}}} \times \mathrm{Vol}_{\mathrm{vial}} \times \frac{1}{\mathrm{Vol}_{\mathrm{samp}}}$$

where $Concentration_{IS}$ is the concentration of the matched internal standard and PA_{IS} is the peak area of the internal standard in that sample.

Quantification of Vitamins

Cobalamins were quantified using a calibration curve. Standards for pseudocoblamains were not analyzed for this study but have been analyzed on our LC-MS system in the past [169]. We calculated the ratio of the response factor of OH-pseudocobalamin and OH-cobalamin in our historic data and applied that ratio to the cobalamin standard for this study (1.675 OH-cobalamin:OH-pseudocobalamin). Cobalamin and pseudocobalamin are notoriously light sensitive compounds and the active forms of cobalamin (Me- and Ado-) have been shown to degrade into OH-cobalamin on the order of minutes [136]. OH-cobalamin is then stable on the timescales of days. Sampling and extraction for this experiment did not take place in a dark environment and therefore we expect all bioavailable forms of cobalamin to have converted into OH-cobalamin. We assume that similar photodegradation processes apply to pseudocobalamin. We therefore interpret our OH-psuedocobalamin results to represent the summation of all pseudocobalamin within the system. However, we acknowledge that we cannot fully discount any losses of our total cobalamin measurement from photodegradation or that some of the ME- and Ado-pseudocobalamins did not convert to OH-pseudocobalamin, both resulting in underestimations of the total pseudocobalamin pools.

Identification of differentially abundant metabolites

To understand how viral infection alters the metabolism of *Prochlorococcus*, we sought to identify metabolites that were differentially abundant when viral infection was high and bacterial abundance was low. To do this, we identified three treatment/timepoint combinations (HV-T12, HV-T24, and LV-T36) that had both high levels of viral infection (timepoints immediately preceding high levels of *Prochlorococcus* mortality) and high amounts of *Prochlorococcus* biomass relative to heterotrophic bacterial biomass (>85% Pro C/total C) (Supplemental Table 3). We classified these samples as high viral infection (HVI) samples. The peak area of each metabolite was then normalized to the amount of microbial carbon estimated by flow cytometry and the log2 fold change (log2fc) was calculated for each HVI sample compared to the biological replicate-matched control. We used both an fdr-adjusted, 1-tailed t-test and a one-sided Wilcoxon signed-rank test to determine if the log2fc values of metabolites in HVI were significantly different from 0 (i.e. the control) ($p < 0.05$). The mean log2fc absolute difference was required to be at least 0.5.

To further account for the potential of bacterial growth to influence our results (metabolites appearing to be enriched during viral infection due to the growth of heterotrophic bacteria), we calculated the Pearson correlation coefficients for all metabolites compared to bacterial biomass across the HV and LV treatments. We assessed the significance of these correlations using fdr-adjusted linear models. Compounds that showed significant positive correlations with heterotrophic bacteria ($p<0.05$) were removed from our interpretation.

Metabolite Multivariate Statistical Analyses:

We performed multivariate analyses on our samples using the vegan package (V2.6-6.1) in R (V4.4.1). B-MIS normalized peak areas for all MFs were normalized to volume filtered and standardized across the entire sample set (value between 1 and 0). Difference between samples, timepoints, and treatments were visualized using a nonmetric dimensional scaling (NMDS) plot with a Manhattan distance matrix using the metaMDS function. Using a scree plot we selected 2 dimensions as optimal. The significance of the projection was defined using a Monte Carlo permutation. We performed an analysis of similarities using the adonis2 function on the same Manhattan distance matrix to identify the significance of differences due to treatment, time, and biological replicate across our samples.

Transcriptomics:

Sampling and RNA Extraction:

Samples were collected using vacuum filtration to filter 50-200 mL of culture per sample (depending on predicted culture biomass) onto 47 mm, 0.2 µm polycarbonate filters. Filtration times ranged from 10–30 minutes. Filters were transferred to cryovials, flash-frozen in liquid nitrogen, and stored at -80 ˚C until extraction. Filtration funnels were bleached for 20 minutes and washed with MilliQ water between time points. Total RNA was extracted from individual filters using a Qiagen RNeasy Mini Kit with a modified lysis step. Briefly, lysis buffer and Biospec zirconia/silica beads (0.5 mm) beads were added to each sample for each treatment (n=3) and vortexed for 1 min, placed on ice for 30 s, and vortexed again for 1 min. Lysate from each sample was removed with a pipette and transferred into clean 1.5 mL microcentrifuge tubes. Synthetic standards (Invitrogen by ThermoFisher ERCC RNA Spike-in mix) were then added to each sample based upon estimated total RNA (derived using experimental cell numbers for *Prochlorococcus* sp. and a published estimate of RNA per cell [173]). The remainder of the Qiagen RNeasy Mini Kit protocol was then followed according to the manufacturer's instructions, incorporating the on-column DNase digestion step using a Qiagen RNase-free DNase kit. Resulting total RNA was eluted with RNase-free water. Quantity and quality of extracted total RNA was assessed on an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA).

Sequencing:

Extracted samples were sequenced after rRNA depletion (Illumina®TruSeq Stranded Total RNA with Ribo-Zero Plus library preparation kit) to enrich for prokaryotic mRNA. Library preparation and sequencing (NovaSeq6000) of Illumina®TruSeq libraries was performed at the JP Sulzberger Columbia Genome Center (New York, New York) following Center protocols. RNA was sequenced to produce 40 million 100-bp, paired-end reads.

Quality Control, Mapping, and Normalization:

Reads were trimmed with BBMap (https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh), quality controlled with fastqc (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and aligned to the MED4 reference genome with bowtie2 [174]. The resultant sam files were sorted and converted to bam files with samtools [175]. Reads were counted in the aligned bam with htseq-count to produce sample by gene count tables [176]. The gene count table served as input along with the MED4 gff annotation file to create the annotated count tables. To account for biological and sequencing variability, counts were normalized using both the spike-in ERCC standard and a *Prochlorococcus* housekeeping gene (RnpB) [117].

Differential Expression Analysis:

The DESeq2 R package (V1.36.0) was employed to identify differentially expressed transcripts [177]. Standard functions and default workflows were applied. Tests for differential expression were conducted between each pairwise infection and control sample at the same time point for all time points (0, 12, 24, and 36 hours) using a negative binomial generalized linear model. The genes and resulting log fold change and Benjamini-Hochberg adjusted *p*-values were matched with KEGG (Kyoto Encyclopedia of Genes and Genomes) annotations [178]. Only differentially expressed genes (fdr-adjusted *p*-value < 0.05) with KEGG annotations were used for interpretation.

3.9 Supplemental Figures:

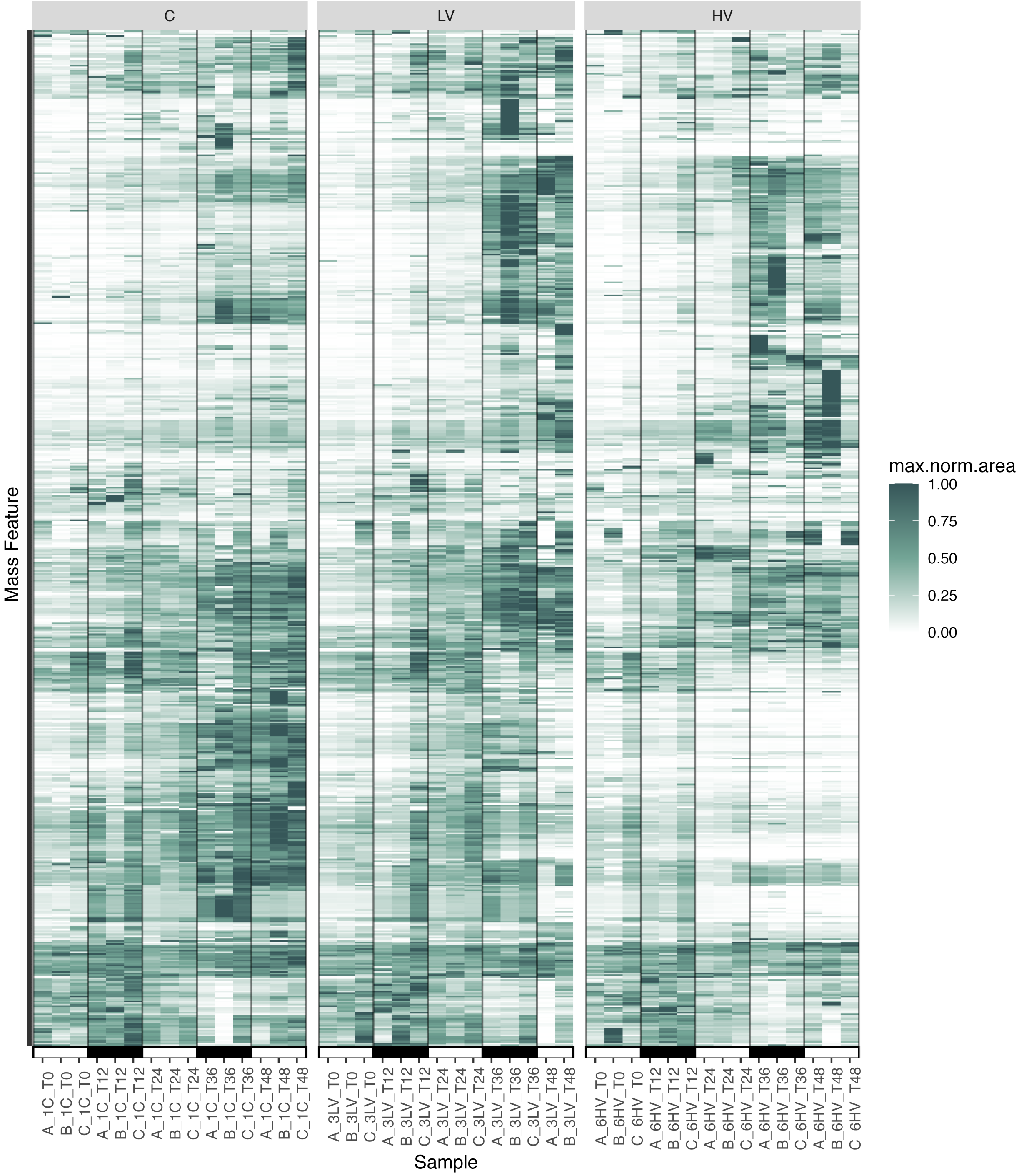

Figure 3.8: Heatmap of untargeted mass features showing max normalized area. Mass features are clustered hierarchically. Triplicate samples are grouped by treatment (top panel) and time point with morning samples identified by white bars and evening samples identified by black bars.

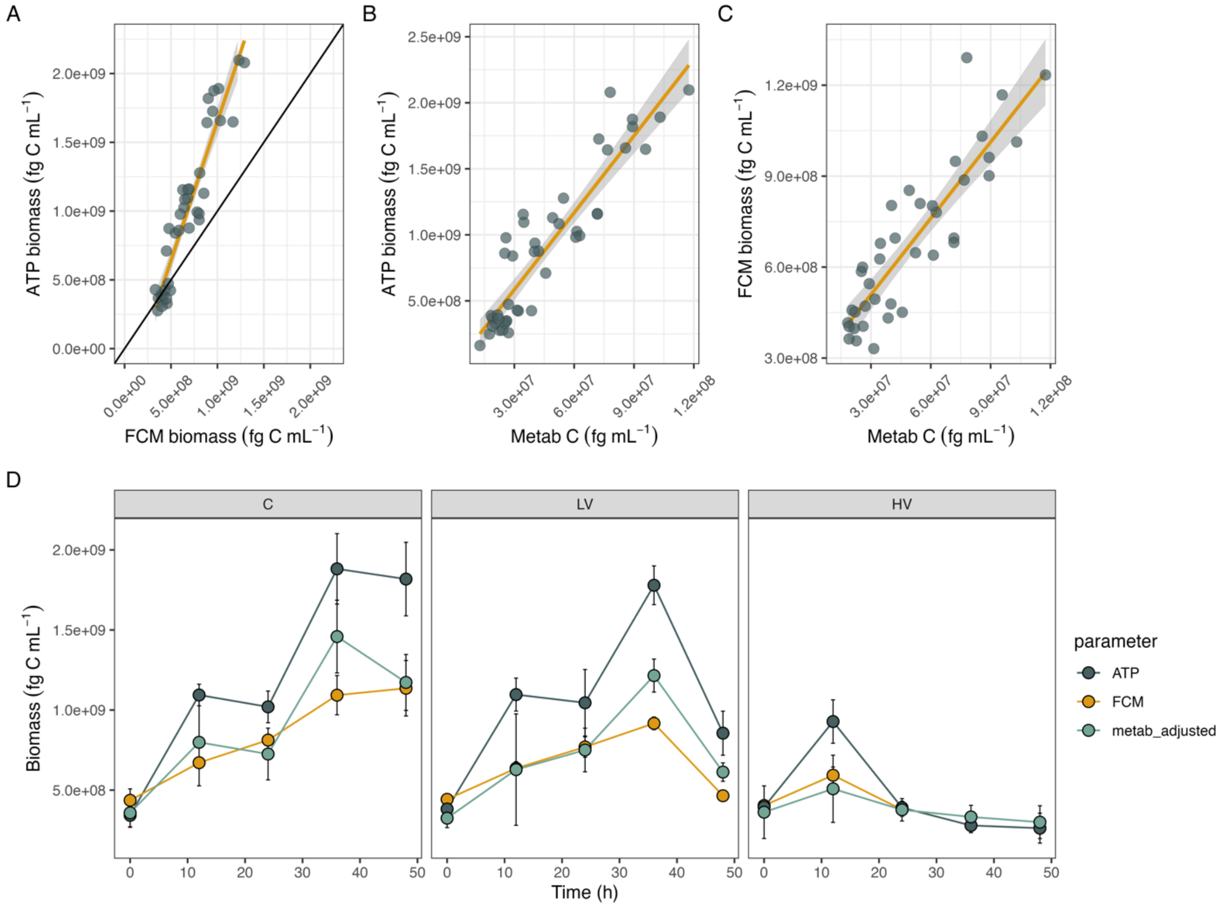

Figure 3.9: Relationship between flow cytometry (FCM) estimated carbon biomass and ATP estimated carbon biomass (A). The black line represents the 1:1 line between these two variables and the yellow line represents the best linear fit. Relationship between total metabolite carbon and ATP estimated carbon (B) and FCM estimated carbon (C). The yellow line represents the best linear fit for the data. The R^2 values for A, B, and C are 0.89, 0.84, and 0.78, respectively. Mean ATP estimated carbon, FCM estimated carbon, and adjusted metabolite carbon (metabolite carbon divided by 0.07 to reflect the percentage of microbial carbon in metabolites (7%)) over the time course of experiment. Error bars represent one standard deviation of the mean of three biological triplicates.

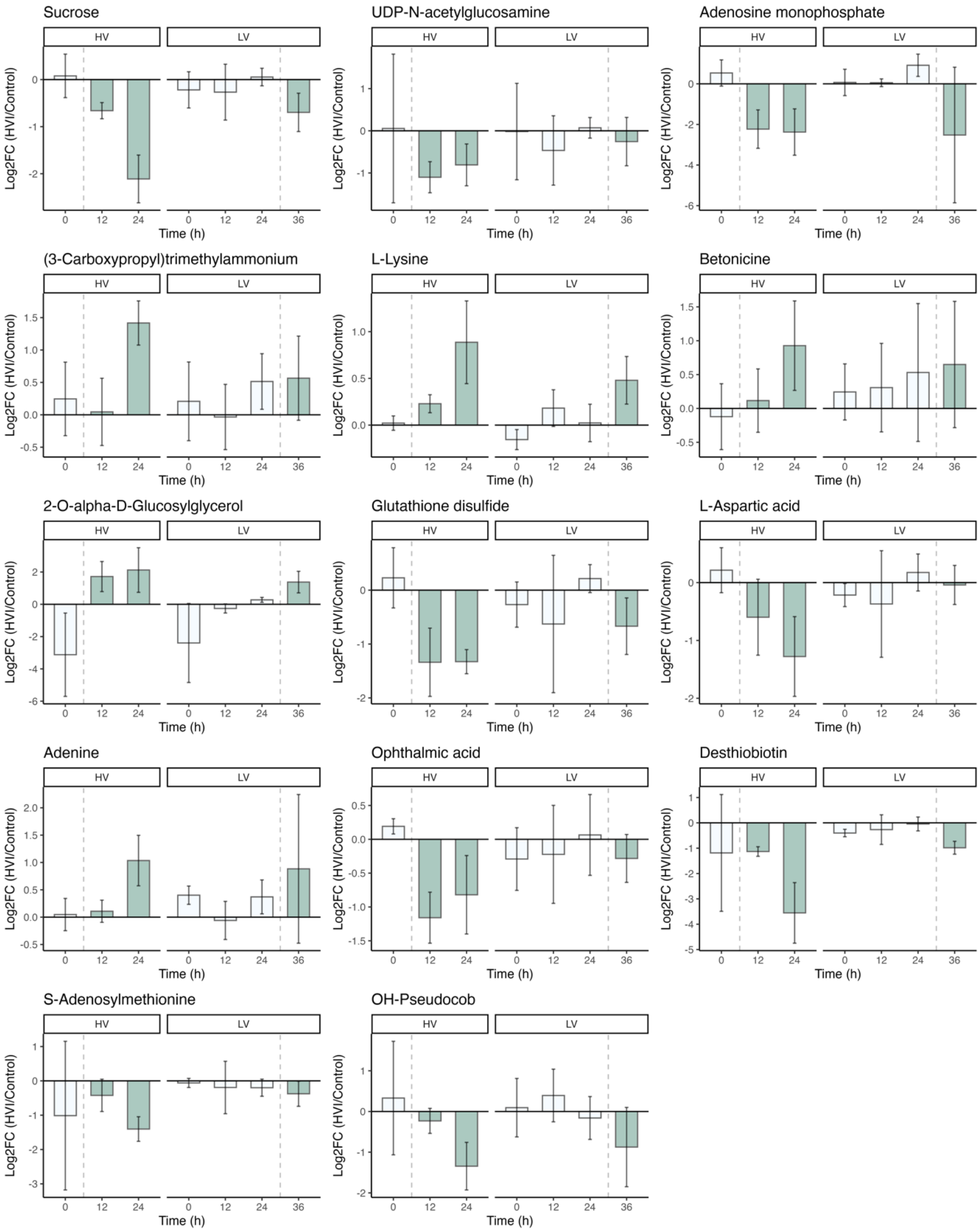

Figure 3.10: Fold change of known metabolites identified as significantly different under viral infection over time. Timepoints designated as "high viral infection (HVI)" are colored green and to the right of the dotted lines (HV 12 and 24 h and LV 36 h) and represent the timepoints used for characterizing the impact of viral infection on cellular metabolism. Bars display the mean of three biological replicates and error bars show the standard deviation of the mean.

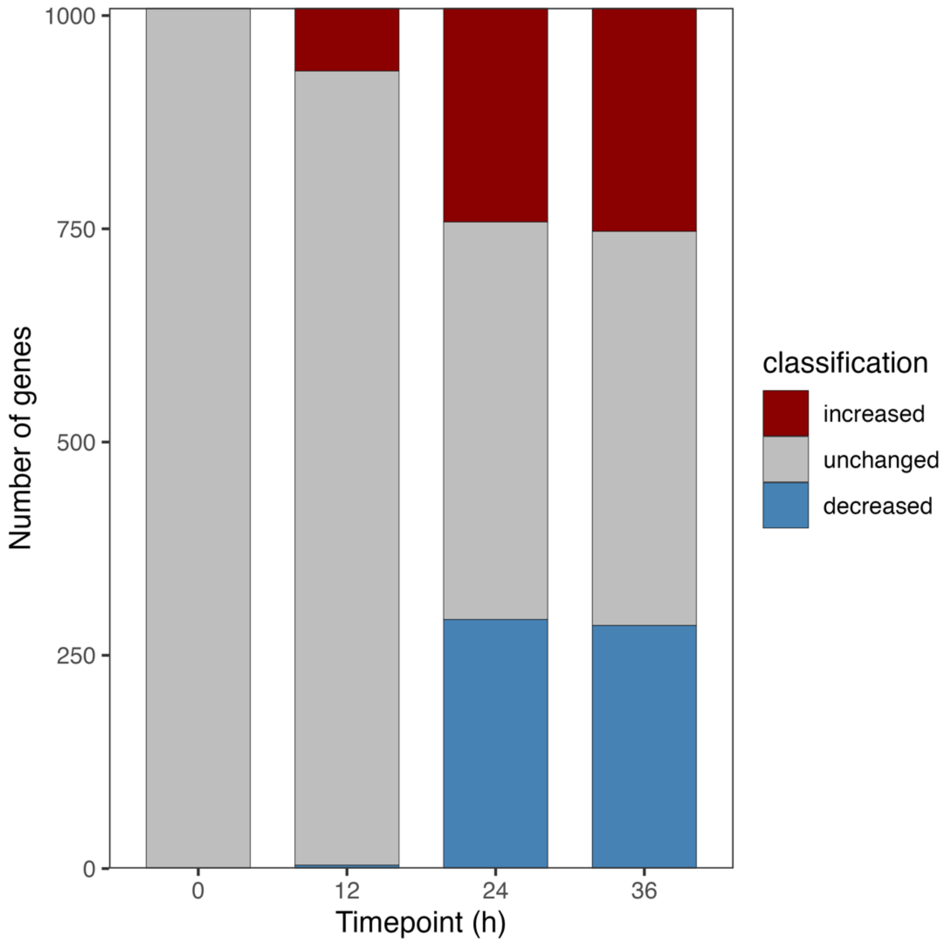

Figure 3.11: Summary of changes to the expression of *Prochlorococcus* genes over the course of the experiment. Genes were classified as having significantly (DESeq2, $p < 0.05$) increased expression (blue) or decreased expression (red) or unchanged (gray) in the HV treatment relative to the control.

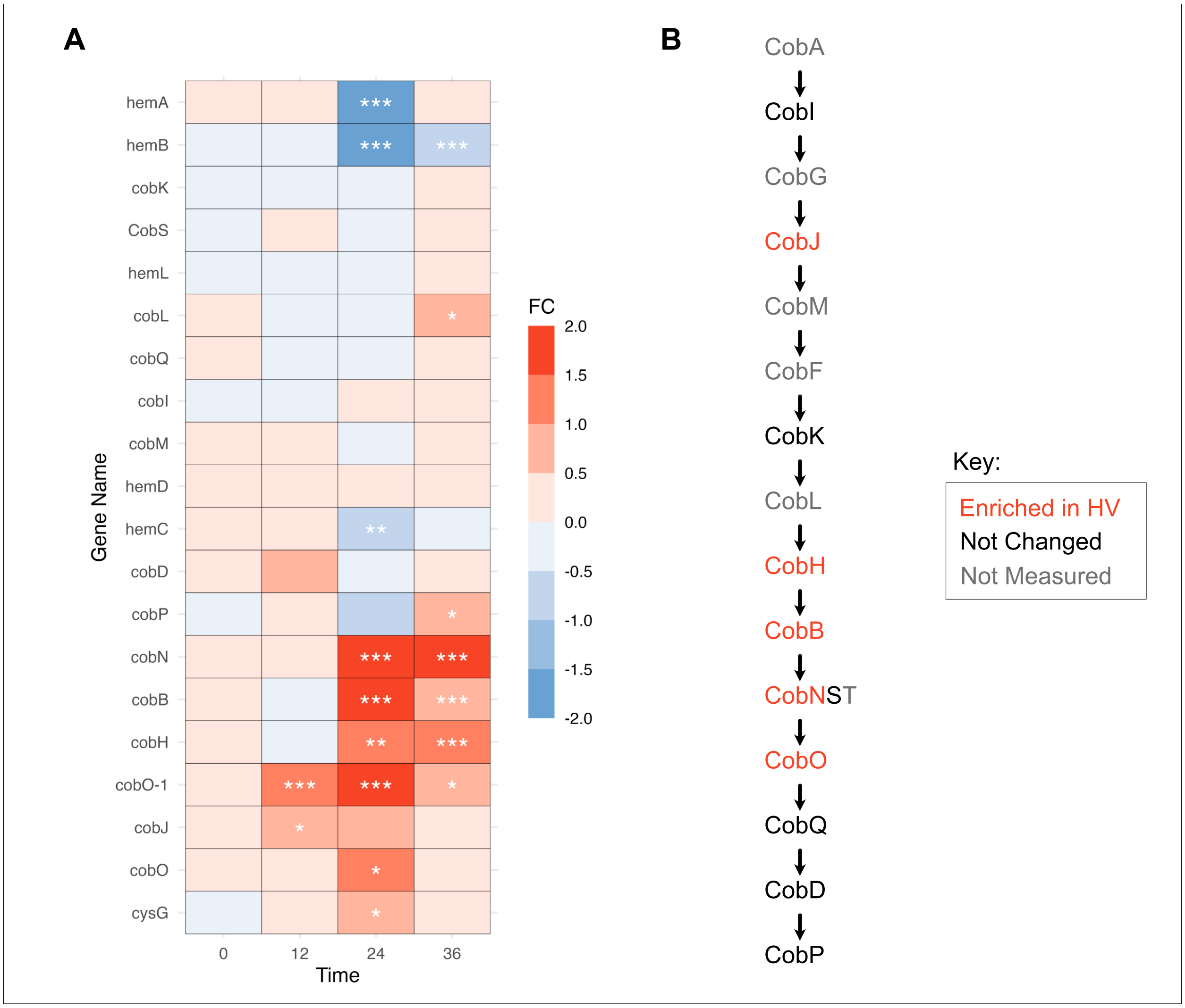

Figure 3.12: Hierarchically clustered heatmap showing the relative transcriptional behavior of genes involved in pseudocobalamin synthesis in Prochlorococcus (A). Color indicates fold change and asterisks indicate significance (“*”: $0.05 > p > 0.01$; “**”: $0.01 > p > 0.005$; “***”: $p < 0.005$). Simplified metabolic map of genes involved in psuedocobalamin synthesis showing genes that were upregulated under viral infection, unchanged under viral infection, or not measured in this study (B).

3.10 Supplemental Tables:

Table 3.1: Abundances and biomass of *Prochlorococcus* (Pro), small bacteria (BacSml), large bacteria (BacLrg), total bacteria (BacTot) and all microbes (TotMicrobial) estimated by flow cytometry. Abundances (count) are in cells mL^{-1}, carbon quotas (Qc) are in fg C $cell^{-1}$, and biomass estimates (C) are in fg C mL^{-1}. The fraction of total biomass in *Prochlorococcus* is calculated in Pro.C.Fraction. Abundances are reproduced from Lindell et al. in prep. This table is provided as a separate file.

Table 3.2: Volume normalized peak areas of all mass features in all samples. Mass features are uniquely identified (MF) with known compounds annotated in by Name. The final best-matched internal standard used for normalization is indicated by FinalBMIS. The analytical fraction, retention time (RT), mass-charge ratio (mz), and predicted adduct (Predicted_Adduct_Ion) for all features is provided. This table is provided as a separate file.

Table 3.3: Quantified metabolite concentrations in each sample and the percent of total microbial carbon represented by each metabolite. Metabolite concentrations are presented as concentrations of compound (Metab_nM_in_sample), concentration of carbon in that metabolite (Metab_nM_C_in_sample), and as the concentration of nitrogen in that metabolite (Metab_nM_N_in_sample). The percentage of carbon represented by that metabolite (Percent_metab_C) of the total microbial carbon estimated by flow cytometry (Total_FCM_biomass_C_in_sample_nM) is also shown. This table is provided as a separate file.

Table 3.4: Particulate ATP data and biomass estimates. The mean and standard deviation of particulate ATP cocnentrations are reported in nmol L^{-1} and living biomass estimates are reported fg C mL^{-1}. Particulate ATP data reproduced from Lindell et al. in prep. This table is provided as a separate file.

Table 3.5: Results of fold change analysis for each MF. Mass features are identified by MF and Name. The average log2 fold change of high viral infection samples compared to the control (mean_log2_fc) and significance of this fold change determined by both a t-test (t_test_p) and Wilcoxon Ranked Sum Test (wilcox_p) are reported. The correlation of the MF with heterotrophic bacterial biomass (bact_cor_pearson), significance of this correlation (bact_cor_lm_p), and final determination as to whether or not the MF is bacteria associated are shown (bact_cor_determination). This table is provided as a separate file.

Table 3.6: Annotations for untargeted mass features from SIRIUS. MFs with predicted identities are identified by MF, known name if applicable (Name), mass-charge ratio (mz), retention time (RT), and analytical fraction (Fraction). The predicted adduct and molecular formula are provided in SIRIUS_adduct and SIRIUS_molecular_formula, respectively. All Classyfire predictions and the probability estimate of the most specific classyfire class are provided. Finally, the predicted SIRIUS annotation for the MF as well as the confidence score are shown in SIRIUS_predicted_identity and SIRIUS_identity_confidence_score, respectively. This table is provided as a separate file.

Table 3.7: Results of fold change analysis for transcripts. Transcripts are identified by name, KEGG ID, KEGG pathway ID, and KEGG pathway description. The log2 fold change (HV/C) and fdr adjusted p-values are reported for each timepoint. This table is provided as a separate file.

Table 3.8: Results of matrix population model containing mean and standard deviation values for cell division rates, carbon fixation rates, and carbon loss rates. This table is provided as a separate file.

Table 3.9: All internal standards used in for best-matched internal standard (B-MIS) normalization and quantification for each fraction. This table is provided as a separate file.

Table 3.10: MS-DIAL parameters used for untarageted metabolomics peak picking and integration. This table is provided as a separate file.

Chapter 4: The cycling of glycine betaine and homarine in marine microbial communities: quantitative flux measurements and the role of competitive uptake inhibition

4.1 Abstract[3]:

The flux of carbon through the labile dissolved organic matter (DOM) pool supports marine microbial communities and represents the fate of approximately half of marine net primary production (NPP). However, the behavior of individual chemical structures that make up labile DOM remain largely unknown. We performed twelve uptake kinetics and two uptake competition experiments on the abundant betaine osmolytes glycine betaine (GBT) and homarine. Combining uptake kinetics with dissolved metabolite measurements, we quantified fluxes through the DOM pool. Fluxes were correlated with particulate concentrations and ranged from 0.53–41 and 0.003–0.54 nmol L^{-1} day^{-1} for GBT and homarine, respectively, equivalent to up to 1.2% of NPP. Turnover times of dissolved GBT and homarine ranged from 1–57 days. Betaines and sulfoniums such as dimethylsulfoniopropionate (DMSP) competitively inhibited homarine uptake. Our results quantify GBT and homarine cycling and suggest an important role for uptake competition in regulating dissolved metabolite concentrations.

---

[3] This paper is currently in review at Limnology and Oceanography Letters as: Sacks, J.S., Carlson, L.T., Sosa, O., Boysen, A.K., Heal, K.R., Finch, A.H., Ferrer-Gonzalez, F.X., Karl, D., White, A., Ingalls, A.E. "The cycling of glycine betaine and homarine through marine microbial communities: quantitative flux estimates and the role of competitive uptake inhibition" In review at *Limnology and Oceanography: Letters.*

4.2 Scientific significance statement:

Dissolved metabolites are the currency for exchanges of carbon, nutrients, chemical structures, and energy among members of marine microbial communities. However, the qualitative and quantitative importance of specific metabolites within these communities is largely unknown, limiting our understanding of the role these compounds play in microbial interactions and biogeochemical cycles. We combine dissolved metabolite concentration measurements with uptake kinetics experiments to quantify the fluxes of two compounds, glycine betaine and homarine, through the dissolved pool across diverse marine environments. We also demonstrate that the presence of structurally similar compounds decreases homarine uptake and decouples transporter kinetics and concentrations, indicating that uptake competition may be an important factor regulating dissolved metabolite concentrations and cycling in the environment.

4.3 Introduction:

As a part of the DOM pool, dissolved metabolites collectively mediate the exchange of nutrients, energy, and chemical structures in marine microbial communities [19]. The most abundant metabolites measured in marine microbial communities are osmolytes that organisms synthesize in response to osmotic stress[10,11,31,32]. Among these abundant osmolytes are the betaines GBT and homarine, which often are among the ten most abundant characterized metabolites in marine particulate (GBT: 0.39–73 nmol L^{-1}, homarine: 0.2–2.7 nmol L^{-1}) and dissolved (GBT: 2.8–5.2 nmol L^{-1}, homarine: 0.62–1.7 nmol L^{-1}) metabolomes[10,25,31,32,179–182]. In addition to serving as osmolytes and growth substrates for marine microbes such as SAR11 and roseobacters[33,183], GBT and homarine can act as sources of nitrogen, be building blocks for other metabolites[70], serve as methyl donors[70,184], facilitate organismal interactions as bioactive molecules[185,186], and serve as precursors to climate active gasses such as methane[100,187]. Their high concentrations suggest that these compounds play important roles in marine biogeochemical cycles, but quantitative estimates of their fluxes do not exist. Furthermore, the biological and chemical controls on their uptake rates, kinetic profiles, and concentrations are largely unknown.

Uptake studies of GBT and DMSP have provided insights into the rates and controls of the cycling of betaines and their sulfur analogs, sulfoniums. Marine bacterial communities display high affinities for GBT and DMSP with low half-saturation constants (K_t) ranging from approximately 1–79 nmol L^{-1}[70,102,188], similar to measured dissolved concentrations in the low nmol L^{-1} range[25,81]. Based on the correspondence between K_t values and dissolved concentrations, it is hypothesized that the concentration of a compound is controlled by the affinity of the oligotrophic bacterial community's transporters for that compound[19]. However, this interpretation is complicated by the fact that GBT and DMSP are taken up by a wide range of marine microbes including *Prochlorococcus*, *Synechococcus*, picoeukaryotes, and diatoms, which often have higher K_t values than oligotrophic heterotrophic bacteria[189–191]. Additionally, GBT, DMSP, and other betaines and sulfoniums act as competitive inhibitors to each other's uptake, suggesting that a compound's cycling is partially controlled by the concentrations of other metabolites present[188,189,192,193]. Homarine was only recently recognized as an important component of marine metabolomes and to our knowledge, no studies exist describing its transporters or cycling[32]. Here we characterize community level uptake kinetics of GBT and homarine, identify competitive inhibitors to homarine uptake in marine microbial communities, and quantify the fluxes of these compounds through the DOM pool.

4.4 Methods:

Study Sites and Environment Characterization

Experiments were performed on five cruises: KM1906 (Spring 2019), TN397 (Fall 2021), and TN412 (Winter 2023) in the North Pacific spanning the equatorial upwelling (Eq), North Pacific Subtropical Gyre (NPSG), and North Pacific Transition Zone (NPTZ) and RC078 (Summer 2022) and RC104 (Summer 2023) in Puget Sound (PS) (Figure 4.1). Samples were collected for particulate organic carbon (POC), particulate nitrogen (PN), chlorophyll *a* (chl), nitrate and nitrite (N+N), and primary productivity. Details of locations, dates, and analyses are provided in the supplemental methods and Table 4.3.

Uptake Kinetics and Competition Experiments

For TN397, TN412, RC078, and RC104, seawater was collected through the ship underway flowthrough systems and prefiltered through 100 µm mesh. Triplicate samples were spiked with seven different concentrations of isotopically labeled substrate ($^{2}H_{3}$-homarine, $^{13}C_{5}$,$^{15}N_{1}$-GBT) ranging from 0 nmol L^{-1} to 1000–5000 nmol L^{-1}, depending on expected *in situ* concentrations, and incubated for 30 minutes in temperature and light-controlled incubators. Samples were collected using peristaltic pumps onto PVDF membrane filters, flash frozen in liquid nitrogen, and stored at -80 °C. Blanks were collected for each concentration by refiltering the filtrate onto a new filter. For dissolved metabolite analysis, 40 mL of filtrate from the 0 nmol L^{-1} treatments was collected in acid washed 50 mL polypropylene falcon tubes and frozen at -20 °C. The uptake kinetic experiment on KM1906 was similar, with details in Boysen et al.[70]. Uptake competition experiments (TN397_UCH1, RC104_UCH1) were identical to uptake kinetics experiments except that for each treatment the spike consisted of 50 nmol L^{-1} of $^{2}H_{3}$-homarine and 100 nmol L^{-1} of unlabeled competitor (GBT, DMSP, TMAO, trigonelline, or glucose). Additional details are in the supplemental methods.

Metabolite Extraction, Data Acquisition, and Quantification:

Extraction, analysis, and quantification of KM1906 samples are detailed in Boysen et al. (2022). For all other particulate samples, extractions were performed with 40:40:20:0.1 methanol:acetonitrile:water:formic acid solution as the extraction solvent[194]. Dissolved metabolite samples were extracted using cation-exchange solid phase extraction and quantified as detailed in Sacks et al.[25]. Dissolved and particulate metabolites were measured using liquid chromatography-mass spectrometry (full LC-MS methods detailed in supplemental methods). Standard curves were used to quantify labeled GBT and homarine. Blanks showed a linear relationship with spike concentration across all experiments (Figure 4.4). We used this relationship to calculate the expected concentration of labeled compound on the filter not due to biological uptake and subtracted it from our measured concentrations. Unlabeled particulate metabolites were quantified using internal standards added during reconstitution.

Uptake Kinetics, Flux, and Competition Calculations

Uptake kinetics parameters were calculated using the Michaelis–Menten (MM) equation (Equation 1),

Equation 1:

$$V_A = \frac{V_{max}(A)}{K_t + S_n + A}$$

Where V_A is the uptake rate of the labeled compound, V_{max} is the maximum uptake rate, K_t is the half saturation constant, S_n is the in-situ substrate concentration, and A is the concentration of added labeled compound. These models were fit using a non-linear least-squares (NLS) approach. Assuming steady state conditions, fluxes through the dissolved phase were calculated as the *in situ* uptake rate using the MM equation and measured *in situ* concentrations. Turnover times (TT) were estimated by dividing the dissolved *in situ* concentration by the flux. V_{max}, K_t+S_n, and TT were also estimated using Wright-Hobbie (WH) plots[195]. NLS and WH derived MM parameters were tightly correlated and fell largely on the 1:1 line, suggesting our results are robust to changes in model-fitting approach (Figure 4.5). We report the NLS results in the main text. As in Boysen et al.[70], we estimated errors for K_t, V_{max}, and TT using Monte Carlo simulations that sampled the analytical error around each datapoint and calculated the mean and standard deviation of all non-outlier models. Uptake inhibition was assessed by comparing the measured uptake rate of $^2H_3$-homarine in the presence of either no competitor or glucose, a "negative control" that we predicted would not impact homarine uptake based on experiments with GBT[188,193,196].

4.5 Results and Discussion:

Environmental Conditions:

Experiments spanned a wide range of marine biogeochemical gradients (Table 4.3). The oligotrophic NPSG was defined by low concentrations of chlorophyll *a* (0.073–0.097 mg m^{-3} chl), low nutrient concentrations (0.017–0.058 µmol L^{-1} N+N), and low productivity (307–456 nmol C L^{-1} day^{-1}). The Eq and NPTZ had intermediate levels of biomass (0.14–0.20 and 0.85 mg m^{-3} chl, respectively), nutrients (2.9–6.6 and 6.4 µmol L^{-1} N+N, respectively), and productivity (1320–1630 and 675 nmol C L^{-1} day^{-1}, respectively). The coastal/estuarine PS had high but variable biomass (0.73–1.6 mg m^{-3} chl) and nutrients (1.2–22 µmol L^{-1} N+N), and high productivity (median value of 16700 nmol C L^{-1} day^{-1})[197].

Environmental Metabolite Concentrations:

Particulate concentrations of GBT and homarine were 0.32±0.05–8.9±1.6 nmol L^{-1} and 0.014±0.003–2.4±0.2 nmol L^{-1}, respectively, with the lowest concentrations in the oligotrophic NPSG and the highest concentrations in coastal PS (Table 4.1, Figure 4.6). Dissolved concentrations of GBT and homarine were higher than the particulate concentrations and ranged from 0.47±0.14–56±9 nmol L^{-1} and 0.06±0.03–13±2 nmol L^{-1} L, respectively (Table 4.1). Total dissolved concentrations of quantified zwitterionic metabolites (betaines, sulfoniums, and trimethylamine N-oxide (TMAO)) ranged from 4.5±1.1 nmol L^{-1} in the NPSG and up to 150±20 nmol L^{-1} in PS and were dominated by DMSP, gonyol, GBT, TMAO, beta-alanine betaine, and homarine (Figure 4.7, Table 4.4).

Uptake Competition:

Our uptake competition experiments showed that DMSP, GBT, unlabeled homarine (positive control), and trigonelline significantly inhibited the uptake of $^2H_3$-homarine relative to a glucose control ($p < 0.05$, Dunnett's test) (Figure 4.2A). Glucose was not significantly different from a "no addition" control in the RC104 experiment suggesting it serves as an appropriate control. Results from the TMAO incubations disagreed between experiments and in the TN397 experiment, the inhibition was significant but small compared to the other compounds. Therefore, we did not classify TMAO as a competitive inhibitor of homarine. We combined our results with similar uptake experiments from the literature examining competitive inhibition of GBT, DMSP, and choline in natural communities and in cultures of SAR11 (Figure 4.8, Table 4.5). We found that the compounds that inhibited homarine uptake also inhibited the uptake of GBT, DMSP, and choline[188,189,193,196]. We hypothesize that other inhibitory compounds from these studies, including dimethylsulfonioacetate (DMSA), beta-alanine betaine, and proline betaine also inhibit the uptake of homarine and vice versa. To understand the uptake dynamics of any one of these compounds, the concentration of the whole pool of betaines and sulfoniums must be considered.

Uptake Kinetics:

Community uptake of both homarine and GBT followed MM uptake kinetics. V_{max} values ranged from 0.28±0.02–3.7±0.2 nmol L^{-1} hr^{-1} for GBT and 0.052±0.005–0.67±0.08 nmol L^{-1} hr^{-1} for homarine (Table 4.1, Figure 4.2B, Figure 4.8). K_t values ranged from 5.5±1.9–72±22 nmol L^{-1} for GBT and 35±9.4–491±170 nmol L^{-1} for homarine (Table 4.1). Communities had lower K_t values for GBT than for homarine in the NPSG (5.5 vs. 35–38 nmol L^{-1}) and in PS (55–65 vs. 87–491 nmol L^{-1}), suggesting consistently higher affinity for GBT than for homarine across diverse marine ecosystems.

For both GBT and homarine, the oligotrophic NPSG stations had roughly an order of magnitude lower K_t values and lower V_{max} values than the NPTZ, Eq, or PS stations (Table 4.1). The NPSG stations had K_t values closer to the K_t value of SAR11 for GBT (~1 nmol L^{-1}), an oligotrophic bacterium, while the other regions had K_t values that were more similar to those found for diatoms for GBT (189–315 nmol L^{-1}) or DMSP (632 nmol L^{-1})[190,191,193]. We conclude from these results that the uptake of betaines in oligotrophic environments is dominated by free living oligotrophs, while in eutrophic environments such as PS, phytoplankton and copiotrophic bacteria control the kinetics and cycling of these compounds. This conclusion agrees with results from experiments in cultures and natural systems that show substantial uptake of DMSP and GBT by diatoms and other phytoplankton[189–191,198,199].

High affinity (low K_t) transporters are thought to maintain low concentrations of dissolved metabolites in the ocean[19]. Therefore, we predicted that lower K_t values would correspond to lower *in situ* dissolved concentrations. We did not find a significant correlation between K_t and dissolved GBT ($p = 0.12$, $R^2 = 0.65$) or homarine ($p = 0.20$, $R^2 = 0.14$). Additionally, homarine generally had higher K_t values and lower concentrations than GBT, suggesting that community K_t alone does not control dissolved metabolite concentrations (Figure

4.3A, Figure 4.3B). A subset of extremely high-affinity transporters could draw down concentrations below community K_t, explaining these discrepancies. However, GBT and homarine measurements in the NPSG were well below SAR11's 1 nmol L^{-1} K_t for GBT, suggesting high-affinity transporters are insufficient to explain measured dissolved concentrations.

We hypothesized that this disconnect between K_t and *in situ* concentrations may be influenced by competitive inhibition and that the magnitude of this effect would be related to the compound's relative abundance in the total pool of competitors (lower relative abundance leads to a larger difference between K_t and S_n). We quantified this effect as the ratio of K_t to S_n (K_t/S_n) where higher values of K_t/S_n correspond to larger deviations of K_t above *in situ* concentrations (Table 4.6). We compared K_t/S_n to the relative concentration of GBT or homarine in the total pool of potential competitors (all measured betaines and sulfoniums) and found a significant power law relationship ($p < 0.0001$, $R^2 = 0.85$), suggesting that competitive uptake inhibition decouples K_t and dissolved concentrations (Figure 4.3C, Figure 4.10). We hypothesize that the concentration of the total pool of betaines and sulfoniums is controlled by the collective K_t for all transporters of these compounds in the community, but the concentrations of individual compounds are maintained below that concentration. Our results suggest that competitive inhibition is an important process in regulating dissolved metabolite concentrations, especially given the broad specificity of many transporters in marine bacteria[183].

Turnover Times and Fluxes:

GBT TTs ranged from 21.4±12–220±72 h (0.9–9.2 days) while homarine TTs ranged from 390±120–1400±860 h (16–57 days) (Table 4.2). GBT TTs were roughly an order of magnitude shorter than homarine TTs in both the NPSG and PS, suggesting GBT is consistently more rapidly cycled than homarine across diverse environments. Oligotrophic NPSG stations typically had shorter turnover times than NPTZ, Equator, or PS stations for both compounds. These TTs place GBT and homarine in the labile (TTs of hours to days) and semilabile (TTs of weeks to months) DOM categories and suggest that these categorizations are environment dependent.

Fluxes of GBT and homarine through the dissolved pool ranged from 0.53±0.26–41±16 nmol L^{-1} day^{-1} and 0.003±0.001–0.54±0.24 nmol L^{-1} day^{-1}, respectively (Table 4.2). These daily values correspond to turnover rates of 156–459% of the particulate and 10.9–112% of the dissolved GBT pools and 3.3–47% of the particulate and 1.8–6.2% of the dissolved homarine pools (Table 4.7). The entire particulate GBT pool and up to half of the particulate homarine pool is cycled every day. The highest fluxes were observed in PS, reaching daily carbon fluxes of up to 200±79 nmol C L^{-1} day^{-1} for GBT and 3.8±1.6 nmol C L^{-1} day^{-1} for homarine (Table 4.2). The fluxes of GBT and homarine represent 0.05–1.2% and 0.0049–0.023% of estimated daily primary production, respectively (Table 4.7). The flux of nitrogen in GBT at our NPSG station is equivalent in magnitude to 29% of mean nitrogen fixation at Station ALOHA (1.84±1.09 nmol N L^{-1} day^{-1}), suggesting that betaines could be a quantitatively important component of the marine nitrogen cycle[200].

Fluxes of GBT through the dissolved pool were similar to published values for other abundant osmolytes such as taurine (0.2–6.0 nmol L^{-1} day^{-1}) and DMSP (1.7–58.5 nmol L^{-1} day^{-1})[201–205]. The fluxes of homarine were typically one to three orders of magnitude lower, commensurate with homarine's lower concentrations (Table 4.2, Table. 4.8). Fluxes of metabolites from the dissolved pool into the particulate pool should not be directly interpreted as catabolism. Significant portions (79–95%) of GBT remain unaltered in the particulate pool for days following uptake, likely reflecting the use of GBT as an osmolyte rather than as a growth substrate[70,206].

Determining flux rates is time and resource intensive. For this reason, we explored if fluxes can be predicted from other parameters. We identified a strong positive linear relationship between particulate concentration and flux for both compounds (Figure 4.3D, $p = 0.0003$, $R^2 = 0.72$) as well as each compound independently (GBT: $p = 0.0005$, $R^2 = 0.99$; homarine: $p = 0.004$, $R^2 = 0.73$). For these compounds, particulate concentration is a strong predictor of flux.

Comparisons to other studies:

Compared to other studies of GBT uptake, our K_t values were slightly higher than reported K_t+S_n values (5.5–72 nmol L^{-1} vs. 1.2–49 nmol L^{-1}), our V_{max} values were slightly lower (0.28–3.7 nmol L^{-1} hr^{-1} vs. 0.39–44 nmol L^{-1} hr^{-1}), and our TTs were much longer (21–220 hr vs. 0.2–11 hr) (Table 4.8)[102,188]. We attribute these differences to differences in experimental design. In previous studies, samples were prefiltered (1.2 µm) to remove phytoplankton and preincubated for 24 hours in the dark to allow for bacterial drawdown of dissolved metabolites prior to the start of the experiment. In contrast, we used 100 µm filtered seawater and performed our experiments immediately, minimizing potential alterations to bacterial GBT transporter expression and preserving *in situ* dissolved metabolite conditions[84]. Previous studies characterized the high affinity transporters of free-living bacteria without competitive uptake inhibition. However, they likely do not reflect *in situ* cycling rates or the whole community uptake profile. Our results suggest that the cycling rate of GBT may be lower than previously reported and that phytoplankton uptake and competitive inhibition should be considered when measuring uptake kinetics.

4.6 Conclusion:

We quantified the fluxes of two important osmolytes, GBT and homarine, in diverse marine ecosystems, facilitating their incorporation into biogeochemical budgets and informing our understanding of their role in microbial interactions. This work also highlights the importance of competitive inhibition in controlling dissolved metabolite concentrations, kinetics, and cycling in the marine environment.

4.7 Acknowledgements:

We thank the captains, crews, and science parties of the *R/V Kilo Moana* on cruise KM1906, of the *R/V Thomas G. Thompson* on cruises TN397 and TN412, and *R/V Rachel Carson* on cruises RC078, and RC104. We thank Kate Faber, Natalie Kledzik, and Susan Garcia for assistance with laboratory analyses, Will Kumler for valuable conversations and assistance with data analysis, and E. Virginia Armbrust for serving as chief scientist of KM1906, TN397, and TN412. This work was supported by grants from the National Science Foundation (2125886 to AEI and KRH and 2124712 to OAS) and grants from the Simons Foundation (LS award ID 385428, AEI; SCOPE Award ID 329108 to AEI, AEW, and DMK; and SF award ID 598819, KRH).

4.8 Figures:

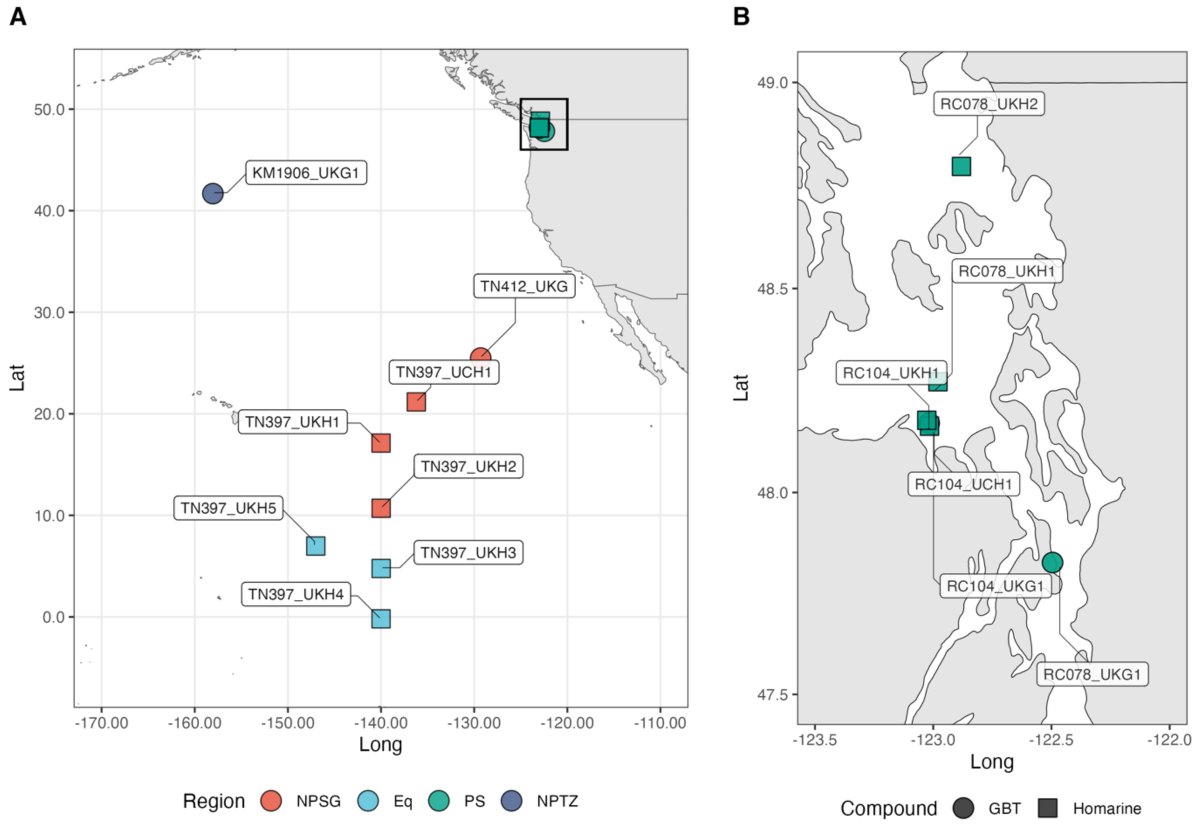

Figure 4.1: Map of experiment locations in the North and Equatorial Pacific (A) with inset focused on Puget Sound (B). The shape of the symbol indicates the compound investigated in the uptake experiment while the color indicates the ocean region (North Pacific Subtropical Gyre (NPSG); equatorial upwelling (Eq); Puget Sound (PS); North Pacific Transition Zone (NPTZ)) where the experiment was performed.

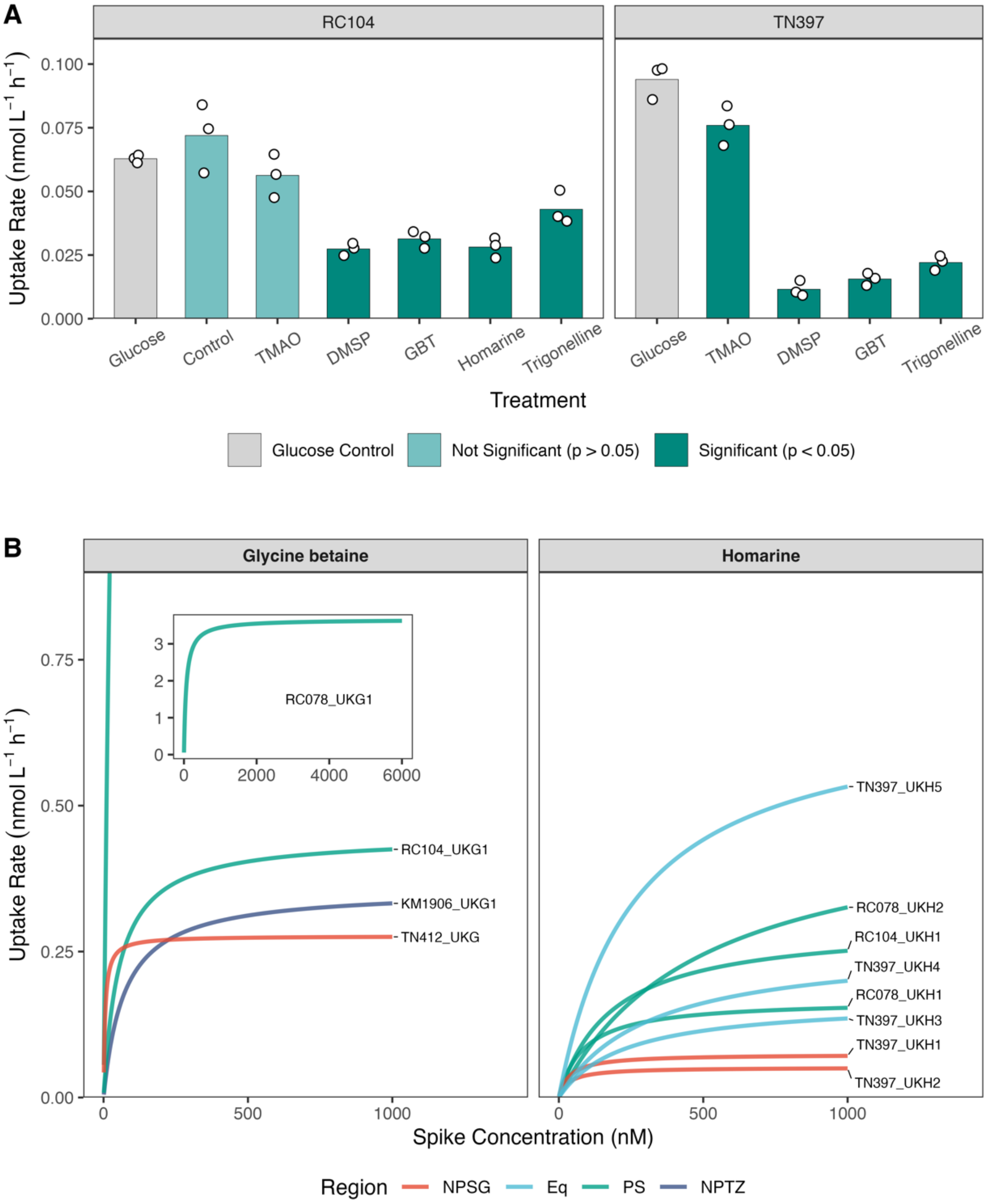

Figure 4.2: Results of uptake competition experiment where bars represent the mean uptake rate of a 50 nmol L^{-1} spike of $^{2}H_3$-homarine and the open circles show the uptake rate of each of the three triplicates (A). Treatments (a 100 nmol L^{-1} spike of competitor) were considered significantly different from the "negative control" of glucose based on Dunnett's Test (p-value <0.05). "Control" refers to a no competitor control and "Homarine" refers to unlabeled homarine. Michaelis–Menten uptake kinetic profiles for the uptake of labeled GBT and homarine

(B). Michaelis–Menten parameters for each experiment are detailed in Table 1. Inset shows the full uptake profile for RC078-UKG1. Color indicates experiment region.

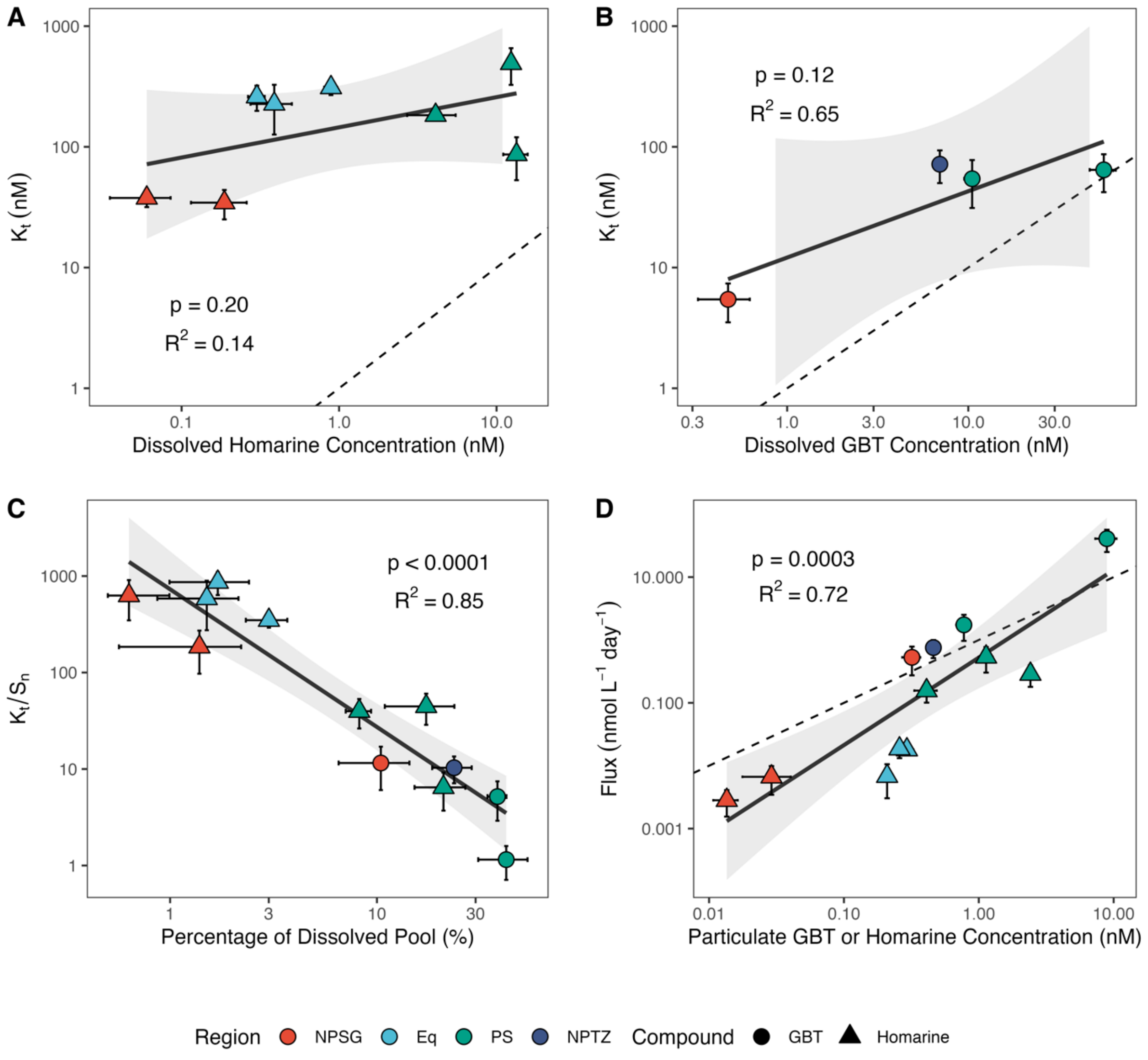

Figure 4.3: Relationship between *in situ* dissolved concentration and K_t for homarine (A) and GBT (B). Points show mean values for individual experiments. Relationship between K_t/S_n and the mole percent of GBT or homarine as a component of the dissolved pool of total quantified betaines and sulfoniums (C). Relationship between *in situ* particulate concentration of either GBT or homarine and the daily flux of these compounds through the dissolved pool (D). For all panels, solid black lines correspond to linear models and gray shading represents 95% confidence intervals. Symbol color shows region and symbol shape shows compound. Error bars represent one standard deviation. Text shows p-values for linear models. Dashed lines are 1:1 lines. Note log10 scaling for both axes for all panels.

4.9 Tables:

Region	Cruise	Experiment	Compound	Particulate Concentration (nmol L^{-1}, mean ± standard deviation)	Dissolved Concentration (nmol L^{-1}, mean ± standard deviation)	Kt (nmol L^{-1}, mean ± standard deviation)	Vmax (nmol L^{-1} h^{-1}, mean ± standard deviation)
NPTZ	KM1906	UKG1	GBT	0.46±0.05	7.0±0.5	72± 22	0.36±0.03
PS	RC078	UKG1	GBT	8.9±1.6	56±9.0	65± 22	3.7±0.2
PS	RC104	UKG1	GBT	0.78±0.10	11± 1.0	54± 23	0.45±0.03
NPSG	TN412	UKG	GBT	0.32±0.05	0.47±0.15	5.5±1.9	0.28±0.02
PS	RC078	UKH1	Homarine	1.10±0.14	13±2	87± 34	0.18±0.02
PS	RC078	UKH2	Homarine	2.40±0.23	12±0.4	490±170	0.49±0.08
PS	RC104	UKH1	Homarine	0.41±0.08	4.1±1.4	180± 19	0.30±0.01
NPSG	TN397	UKH1	Homarine	0.014±0.003	0.06±0.03	38±6.1	0.074±0.004
NPSG	TN397	UKH2	Homarine	0.029±0.011	0.19±0.07	35±9.4	0.052±0.005
Eq	TN397	UKH3	Homarine	0.21±0.02	0.39±0.11	230±100	0.17±0.03
Eq	TN397	UKH4	Homarine	0.29±0.04	0.89±0.08	310± 41	0.26±0.01
Eq	TN397	UKH5	Homarine	0.26±0.02	0.30±0.04	260± 61	0.67±0.08

Table 4.1: GBT and homarine environmental concentrations and uptake kinetic parameters.

Region	Cruise	Experiment	Compound	Turnover Time (h, mean ± standard deviation)	Metabolite Flux (nmol compound L^{-1} day^{-1}, mean ± standard deviation)	Carbon Flux (nmol C L^{-1} day^{-1}, mean ± standard deviation)
NPTZ	KM1906	UKG1	GBT	220± 72	0.76±0.24	3.8±1.2
PS	RC078	UKG1	GBT	33± 14	41± 16	200± 79
PS	RC104	UKG1	GBT	150± 66	1.7±0.8	8.7±3.9
NPSG	TN412	UKG	GBT	21± 12	0.53±0.26	2.7±1.3
PS	RC078	UKH1	Homarine	600±280	0.54±0.24	3.80±1.6
PS	RC078	UKH2	Homarine	1000±390	0.29±0.11	2.0±0.8
PS	RC104	UKH1	Homarine	630±310	0.15±0.06	1.1±0.4
NPSG	TN397	UKH1	Homarine	510±310	0.003±0.001	0.02±0.009
NPSG	TN397	UKH2	Homarine	670±420	0.007±0.003	0.047±0.022
Eq	TN397	UKH3	Homarine	1400±860	0.007±0.004	0.048±0.026
Eq	TN397	UKH4	Homarine	1200±230	0.018±0.003	0.13±0.02
Eq	TN397	UKH5	Homarine	390±120	0.019±0.005	0.13±0.04

Table 4.2: GBT and homarine turnover times and daily fluxes in nmol L^{-1} and nmol carbon L^{-1}.

4.10 Supplemental Text:

Supplemental Methods

Sample Collection for POC, PON, Chl *a*, nutrients, and productivity

Particulate carbon and particulate nitrogen were collected from the underway system (TN397, TN412) or 5-15m CTD casts (KM1906, RC078) and analyzed according to HOT protocols (https://hahana.soest.hawaii.edu/hot/protocols/) for cruises KM1906, TN397, and TN412 and according to University of Washington Marine Chemistry Lab protocols (https://www.ocean.washington.edu/file/Sampling_Procedures_2023) for RC078. Samples for inorganic nutrients in Puget Sound were collected in high density polyethylene (HDPE) bottles and analyzed according to UW Marine Chemistry Lab Protocols (https://www.ocean.washington.edu/file/Sampling_Procedures_2023). High sensitivity nutrient samples were collected in HDPE bottles and analyzed according to HOT protocols (magnesium induced co-precipitation for low level P and chemiluminescence for low level nitrate+nitrite; https://hahana.soest.hawaii.edu/hot/protocols) for TN397, TN412, and KM1906. Temperature and salinity measurements were taken using sensors on either the CTD (KM1906) or the flow through underway system (TN397, TN412, RC078, RC104). Chlorophyll *a* concentrations were determined using an underway flowthrough fluorometer (ECO, Seabird Scientific) for TN397 and TN412 or taken from the CTD for KM1906 and RC078 using factory calibration settings. For TN397, KM1906, and TN412, primary production values were determined using deck board ^{14}C tracer incubations according to standard HOT protocols (https://hahana.soest.hawaii.edu/hot/protocols/) or deck board ^{13}C tracer incubations as per[207].

Isotopic abundances of particles were analyzed at the Biogeochemical Stable Isotope Facility at University of Hawaii (https://www.soest.hawaii.edu/GG/isotope_biogeochem/index_files/Page532.htm).

For RC078 and RC104, productivity was not measured during the cruise, but we used Puget Sound summertime (June-Sept to match cruises) ^{14}C derived primary production values compiled from [197]. Values displayed extremely high variability (standard deviation approximately equal to mean) so the overall median was chosen to provide an estimate for comparison. When both ^{14}C and ^{13}C primary production estimates were available, we compared the fluxes of GBT and homarine to whichever value was higher.

For TN397 and TN412, samples for PC, PN, Chlorophyll-a, and nutrients were not sampled exactly when the experiments were performed so a Stineman interpolation approach was used to co-locate data for these variables with the kinetics experiments using the R package stinepack. Values from KM1906 are from Boysen et al. (2022). Details of experiment locations and environmental measurements are provided in Supplemental Table 1.

Experimental Sample Collection and Design

Seawater for uptake kinetics experiments was collected into acid-washed, 2L polycarbonate bottles from either the Niskin™ bottle rosette on the CTD at a depth of 15 m (KM1906) or underway flowthrough systems (TN397, TN412, RC078, RC104) at depths of 8 m (TN397, TN412) or 1 m (RC078, RC104). For pelagic cruises (TN397, TN412, KM1906), a 100 µm prefiltration step was applied to remove large particulate matter (aggregates and large phytoplankton or zooplankton). Particulate metabolite and uptake rate data from KM1906 experiments are from Boysen et al. (2022) but new dissolved data is reported. Samples in 2-L bottles were stored in flow through (RC078, RC104) or temperature-controlled (KM1906, TN397, TN412) incubators matching the *in situ* temperatures with either blue (KM1906, TN397, TN412) or mesh (RC078, RC104) shading to keep light levels at approximately mixed layer light levels (approximately 30% shading). Samples were then spiked in triplicate with 7 different concentrations of isotopically labeled substrate ($^{2}H_{3}$-Homarine, $^{13}C5$,$^{15}N1$-GBT) ranging from 0 nmol L^{-1} as the lowest concentration and 1000-5000 nmol L^{-1} as the highest concentration, depending on expected *in situ* concentrations (higher spike concentrations were chosen for high expected concentrations). For $^{2}H_{3}$-Homarine, just the methyl group is labeled. The highest spike concentration was 100 to 10,000 times greater than *in situ* concentrations. Samples were incubated for 30 minutes before gentle filtration using peristaltic pumps on 47 mm diameter PVDF Durapore filters. Only six samples could be filtered at a time, so isotope labeled compounds at two concentrations were spiked into triplicate bottles at a time, incubated and filtered immediately. Filtration times were less than 15 minutes and were used with incubation times to calculate uptake rates. The entire experiment took place over the course of several (typically 3) hours. Upon completion of filtration, filters were folded, wrapped in combusted foil, placed in a cryotube, and flash frozen in liquid nitrogen. Samples were kept at -80 ºC until analyzed. For dissolved metabolite analysis, 40 mL of filtrate from the 0 nmol L^{-1} treatments was collected in acid washed 50 mL polypropylene falcon tubes and frozen at -20 ºC until analyzed. Blanks were collected for each concentration be refiltering the filtrate onto a new filter to estimate the amount of signal coming simply from seawater trapped on the filters rather than in particles. Uptake competition experiments were performed identically to uptake kinetics experiments except for each treatment our spike consisted of 50 nmol L^{-1} of $^2H_3$-homarine and 100 nmol L^{-1} of unlabeled competitor.

A major concern during dissolved metabolite sampling is cellular breakage leading to elevated concentrations in the dissolved phase[81]. We took care during dissolved metabolite sampling to minimize filtration pressure and never expose the filtered microbial community to air. Our values for DMSP (2.40-10.9 nmol L^{-1}, mean 4.63 nmol L^{-1}) are within reported literature concentrations[81,201–203,208–210] made using small volume gravity filtration, a method demonstrated to minimize cell breakage (0.5-67.6 nmol L^{-1}, Supplemental Table 7), including for our main study regions of the North Pacific and Salish Sea (1.3-8.8 nmol L^{-1})[208]. The agreement between our values and literature values suggest that cell breakage was minimized during our sampling.

Particulate Metabolite Extraction

Particulate metabolites were extracted two ways. For the KM1906 samples, a modified Bligh and Dyer approach was used as detailed in Boysen et al. (2018). Briefly filters were extracted in Teflon™ tubes with a two-phase extraction with 50:50 methanol:water as the aqueous solvent and dichloromethane as the organic solvent. For the TN397, TN412, RC078, and RC104 samples, a one phase extraction was performed with 40:40:20:0.01 methanol:acetonitrile:water:formic acid solution as the extraction solvent [194]. Briefly, filters were placed in 15 mL Teflon™ tubes with solvent, bead beaten with glass beads, and centrifuged. The solvent was then collected, and the procedure was repeated three times while keeping samples at -20 ºC through the process. In-house tests showed no difference in extraction performance between the two methods for our metabolites of interest. Following all three extraction approaches, samples were dried down under N_2 gas, reconstituted in 400 µL of H2O, filtered through a 0.2 µm syringe filter to remove any particulate matter, and stored at -80 ºC until data acquisition.

Metabolite Data Acquisition

Labeled GBT and homarine were measured in particulate samples by hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray ionization liquid chromatography-mass spectrometry (ESI-LC-MS) on a Waters Xevo TQ-S triple quadrupole mass spectrometer (TQ-S) in selected reaction monitoring mode (mass spectrometry and chromatography conditions detailed below). The ions used for quantification are detailed in Table 4.10. For dissolved metabolites and unlabeled particulate metabolites, ESI-LC-MS was performed with identical LC conditions but a Q-Exactive HF Orbitrap Mass Spectrometer (QE) in full scan mode was used for data collection. Exact chromatography and mass spectrometry conditions are detailed below. Metabolite data from both the TQ-S and QE were converted from .raw files to .mzML files using MSConvert and peaks were integrated using Skyline[75,76]. All raw peak areas from skyline integrations are available on Zenodo at https://zenodo.org/records/14606227.

HILIC Chromatography Conditions

The HILIC conditions used were the same as in[25] and we provide details here for ease of reference. HILIC chromatography was performed using a SeQuant ZIC-pHILIC column (5 µm particle size, 2.1 mm × 150 mm, from Millipore) with 10 mM ammonium carbonate in 85:15 water to acetonitrile (Solvent A) and 10 mM ammonium carbonate in 85:15 acetonitrile to water

(Solvent B) at a flow rate of 0.15 mL min^{-1}. The column was held at 100% B for 2 min, ramped to 64% A over 18 min, ramped up to 100% A over 1 min, held at 100% A for 7 min, and equilibrated at 100% B for 22 min (total time is 50 min). The column was maintained at 30°C throughout the analysis. The injection volume was 2 μL.

TQ-S Mass Spectrometry Conditions

Metabolites for the uptake kinetics and competition experiments were measured on a Waters Acquity UPLC system coupled to a Waters TQ-S triple quadrupole mass spectrometer equipped with electrospray ionization (ESI) in positive ion mode. Selected reaction monitoring (SRM) conditions for each compound (collision energy, cone voltage, parent and daughter ions) were optimized by infusion of each metabolite standard (Table 4.9). For all metabolites, a quantification and confirmation SRM transition (parent and product ion pair) were selected based on maximum peak areas. A labeled internal standard was used for identification when possible and only clear peaks that matched a standard retention time were considered. The SRM transitions were monitored over a 5-10-minute window around its retention time for the HILIC method to ensure at least 12 data points per peak. MS parameters for the HILIC TQ-S method were as follows: capillary voltage of 0.5 kV, source temperature of 130 °C, desolvation temperature of 500 °C, cone gas flow at 150 L h^{-1} and desolvation gas flow at 1000 L h^{-1} in positive mode.

QE Mass Spectrometry Conditions

The QE mass spectrometry conditions used were the same as in [25] and we provide details here for ease of reference. Instrument parameters were as follows: a capillary temperature of 320 °C, an heated-electrospray ionization probe (H-ESI) spray voltage of 3.5 kV, an auxiliary gas heater temperature of 90 °C, an S-lens RF level of 65, a sheath gas flow rate of 16 L h^{-1}, an auxiliary gas flow rate of 3 L h^{-1}, and a sweep gas flow rate of 1 L h^{-1} were used. Polarity switching was employed with a scan range of 60–900 m/z and a resolution of 60,000.

TQ-S Particulate Metabolite Quantification

Quantification of the KM1906 samples is detailed in Boysen et al. (2022). For the TN397, TN412, RC078, and RC104 samples, standard curves were used to quantify labeled GBT and homarine. Separate standard curves were created for each experiment using the 0 nmol L^{-1} spike samples with concentrations ranging from 0-200 nmol L^{-1}. Blanks showed a highly linear relationship (GBT: $p = 0.003$, $R^2 = 0.99$; homarine: $p < 0.0001$, $R^2 = 0.95$) with spike concentration across all experiments (Figure 4.3). We used this relationship to calculate the expected concentration of labeled compound on the filter that was not due to biological uptake and subtracted it from our measured concentrations.

QE Dissolved and Particulate Metabolite Quantification

Dissolved metabolites and *in situ* particulate metabolites were quantified as reported previously[25] and Boysen et al. (2018). Briefly, raw peak areas were normalized using best-matched internal standard normalization and quantified by comparing to an authentic standard in H_2O and in matrix within each batch on the instrument or using a matched internal standard added before extraction. Concentrations were adjusted for extraction efficiencies (detailed in Sacks et al. (2022)) and batch specific limits of detection (LOD) were calculated for each compound using CX-SPE blanks (Table 4.5). The average blank concentration was subtracted from both our dissolved measurements and our reported limits of detection to accurately reflect environmental dissolved metabolite concentrations our methods ability to measure them with CX-SPE and LC-MS. Unlabeled particulate metabolites were quantified using internal standards added during reconstitution.

Calculations

Wright-Hobbie Transformations

K_t+S_n, V_{max} and turnover time (TT) were estimated using Wright-Hobbie (WH) plots. These plots show the linear relationship between treatment concentration on the x-axis and uptake time divided by the fraction of labeled compound taken up on the y-axis. K_t+S_n is the x-intercept, V_{max} is the slope, and turnover time is the y-intercept[195]. Values for all experiments are reported in the Supplemental Table 8. WH and NLS kinetics parameters and turnover times showed strong linear relationships (K_t+S_n: $p < 0.0001$, $R^2 = 0.91$; V_{max}: $p < 0.0001$, $R^2 = 0.99$; TT: $p < 0.0001$, $R^2 = 0.92$) with most values falling on the 1:1 line, suggesting that these two approaches are largely interchangeable (Figure 4.5).

Monte Carlo Simulations

As in Boysen et al. (2022) we estimated errors for K_t, V_{max}, and TT using Monte Carlo simulations where we sampled the error around each datapoint and calculated the mean and standard deviation of all non-outlier models. Outliers were defined as having adjusted R^2 values two standard deviations outside of the mean or as models that failed to converge. We performed 1000 iterations of each Monte Carlo simulation.

Statistical Analysis

Statistics were performed using the R statistical environment (version 4.4.1). Data manipulations and statistical analyses were performed using the tidyverse, lubridate, broom, and rstatix packages. Data visualization was performed with the tidyverse, scales, cowplot, ggthemes, ggsci, ggpubr, network, GGally, and ggrepel packages.

Literature Synthesis

We compiled literature on uptake competition experiments studying the related compounds DMSP, GBT, and choline and curated the data by requiring compounds to significantly inhibit uptake according to reported statistical analysis. When no statistical analysis was reported, we included experiments that showed uptake inhibition by at least 30% relative to the control[188,193,196].

4.11 Supplemental Figures:

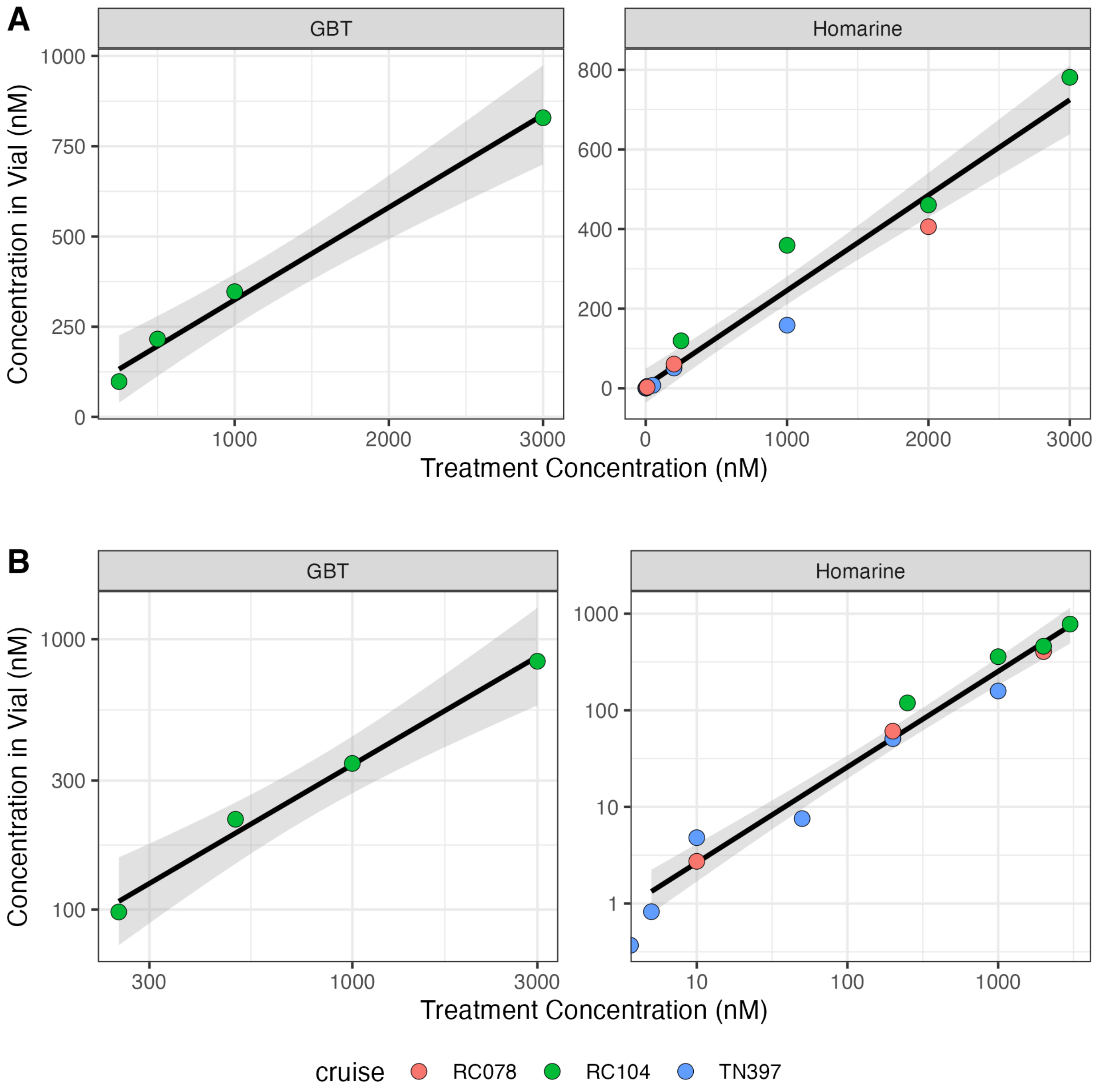

Figure 4.4: Relationship between the measured vial concentration and treatment concentration for blanks for GBT (left) and homarine (right) with linear (A) and log10 scaling (B). Symbols show individual blank samples. Symbol color represents cruise. The black line represents the fit linear model and the gray shaded area represents the 95% confidence interval.

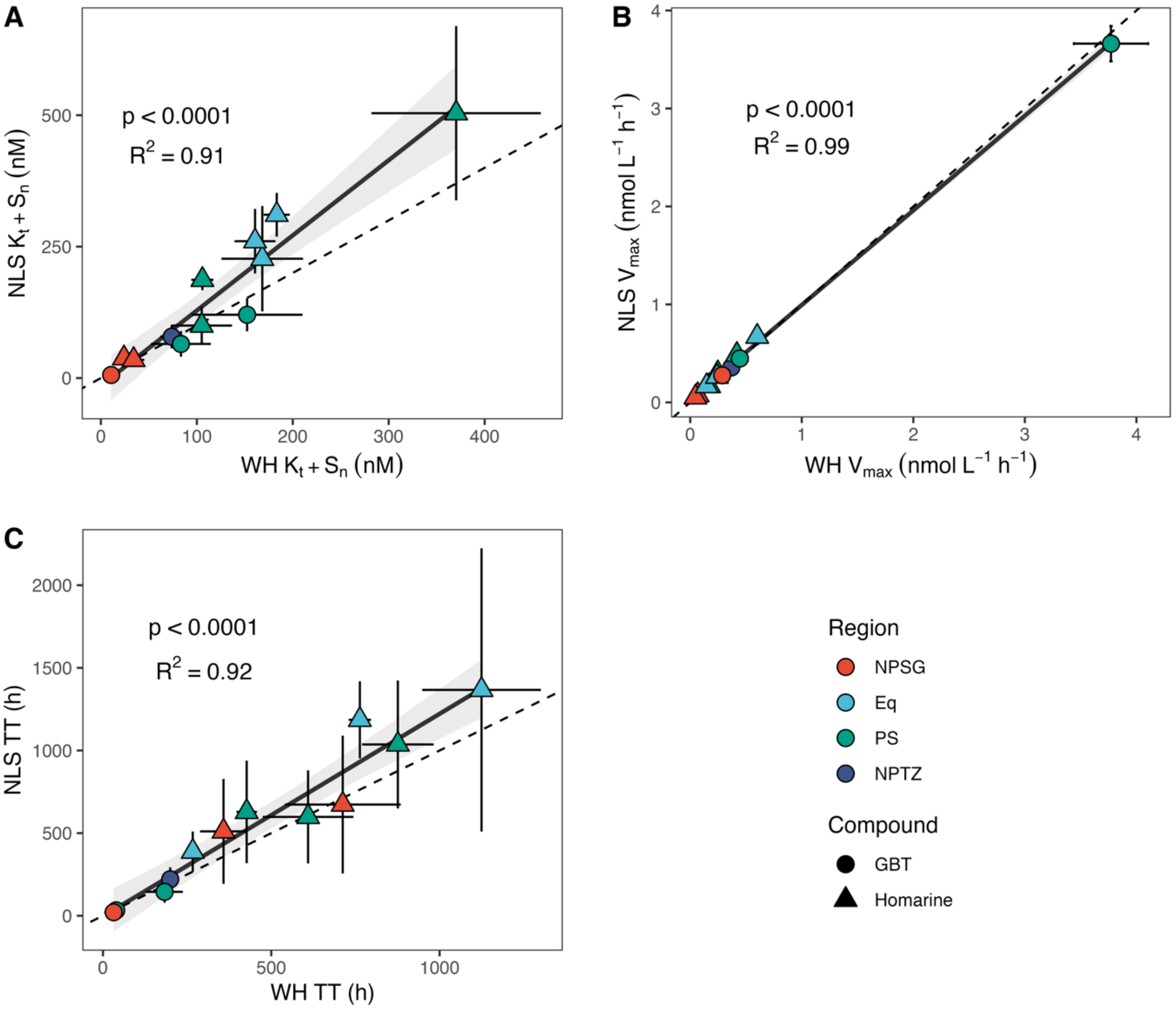

Figure 4.5: Comparison of K_t+S_n (A), V_{max} (B), and TT (C) values determined with the nonlinear least squares (NLS) and the Wright-Hobbie (WH) transformations for all experiments. Experiment region is shown by symbol color and the compound tested is identified by shape. Error bars represent one standard deviation of all Monte Carlo models. The linear line of best fit is shown as a solid black line and the gray region represents the 95% confidence interval. The dashed line is the 1:1 line. The *p* and R^2 values for a linear model are shown on each plot.

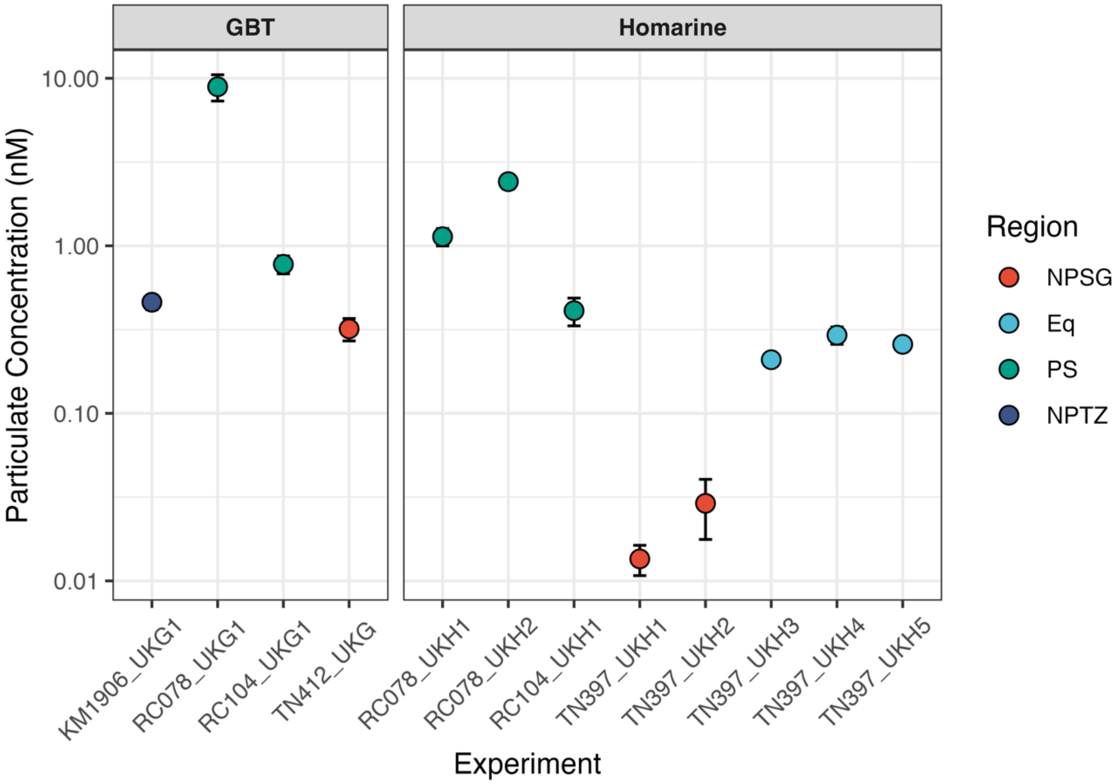

Figure 4.6: Particulate concentrations of GBT and homarine at experimental locations. Error bars represent one standard deviation. Colors represent different regions. Note log10 scaling of y-axis.

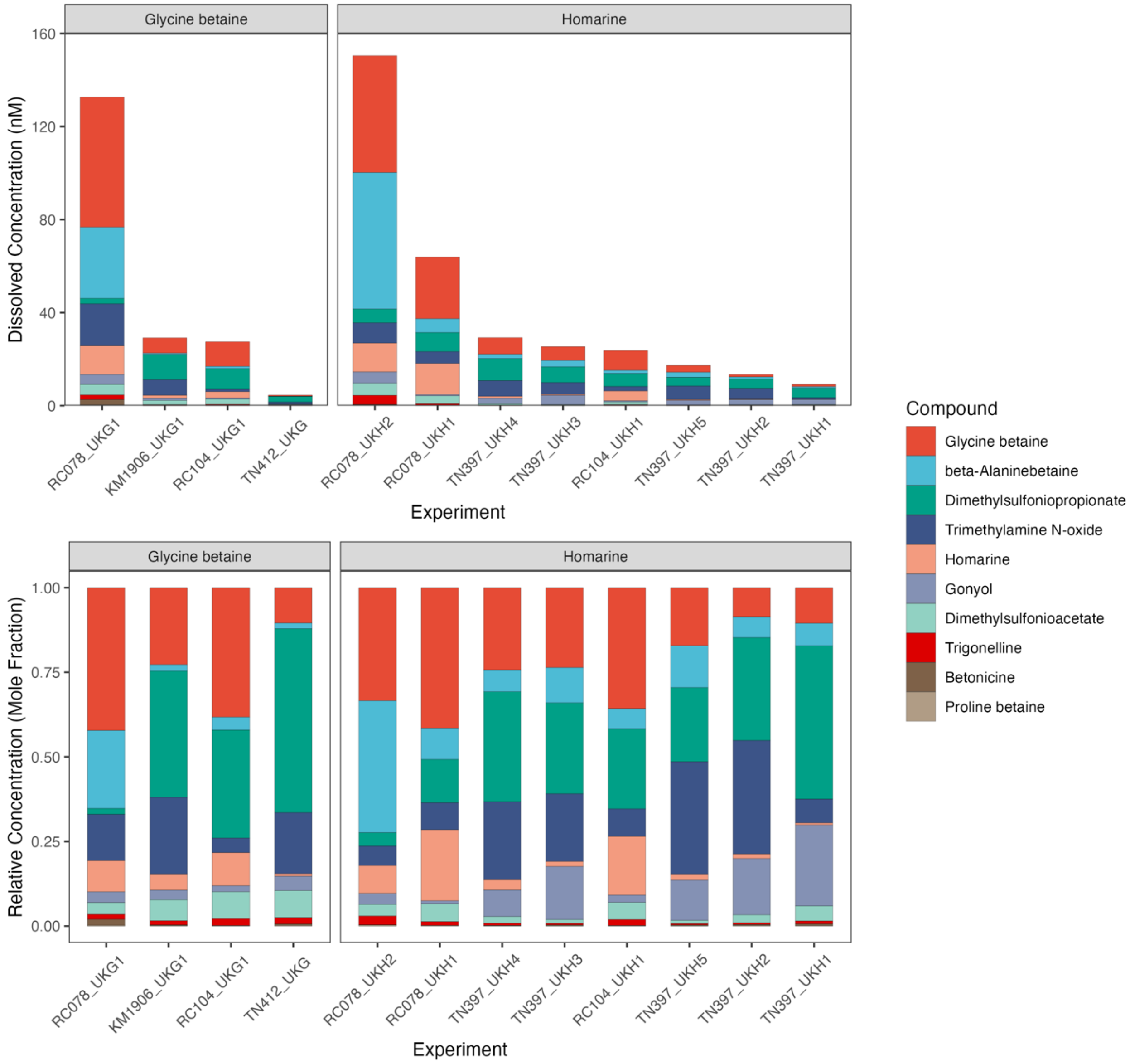

Figure 4.7: Absolute (A) and relative (B) environmental concentrations of dissolved zwitterionic metabolites (sulfoniums, betaines, TMAO) at locations selected for experiments investigating the uptake kinetics of GBT (left) or homarine (right).

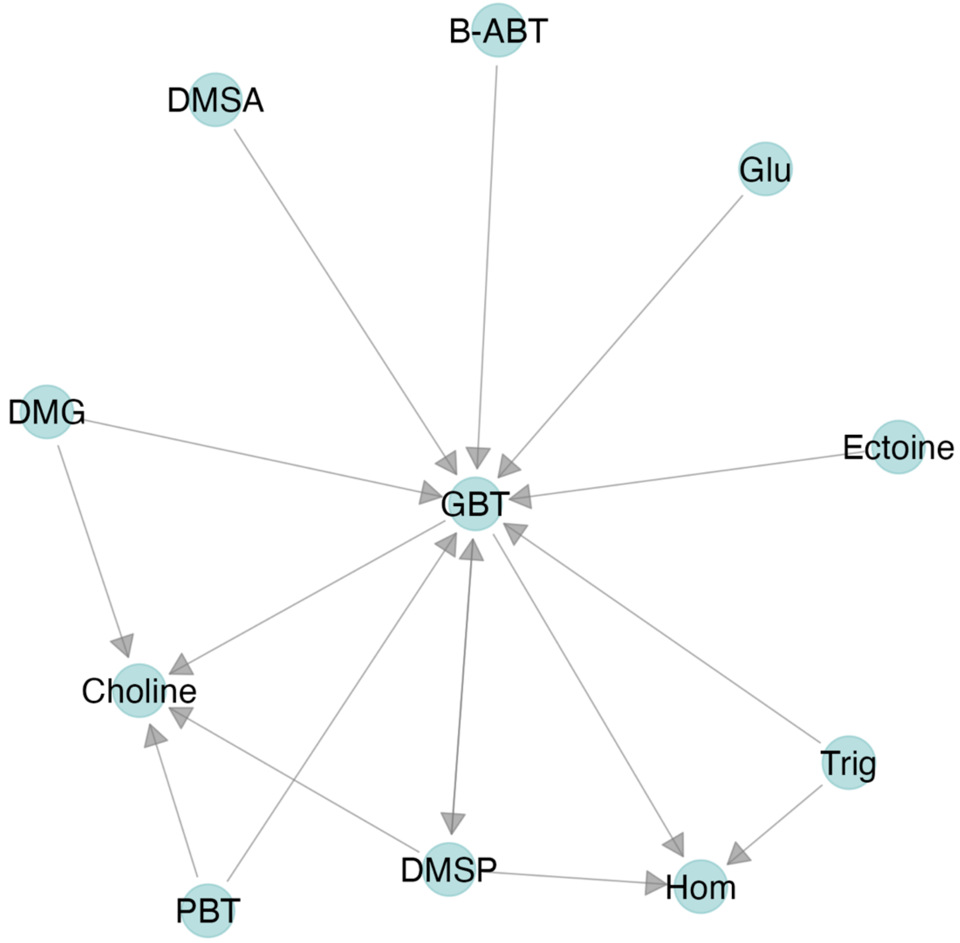

Figure 4.8: The uptake-inhibition network created by synthesizing the results from this study with literature values. Arrows point from the inhibitor toward the inhibited compound. Abbreviations are as follows: GBT: glycine betaine, Glu: glutamic acid, Hom: homarine, Trig: trigonelline, DMSA: dimethylsulfonioacetate, DMG: dimethylglycine, PBT: proline betaine, B-ABT: beta-alanine betaine.

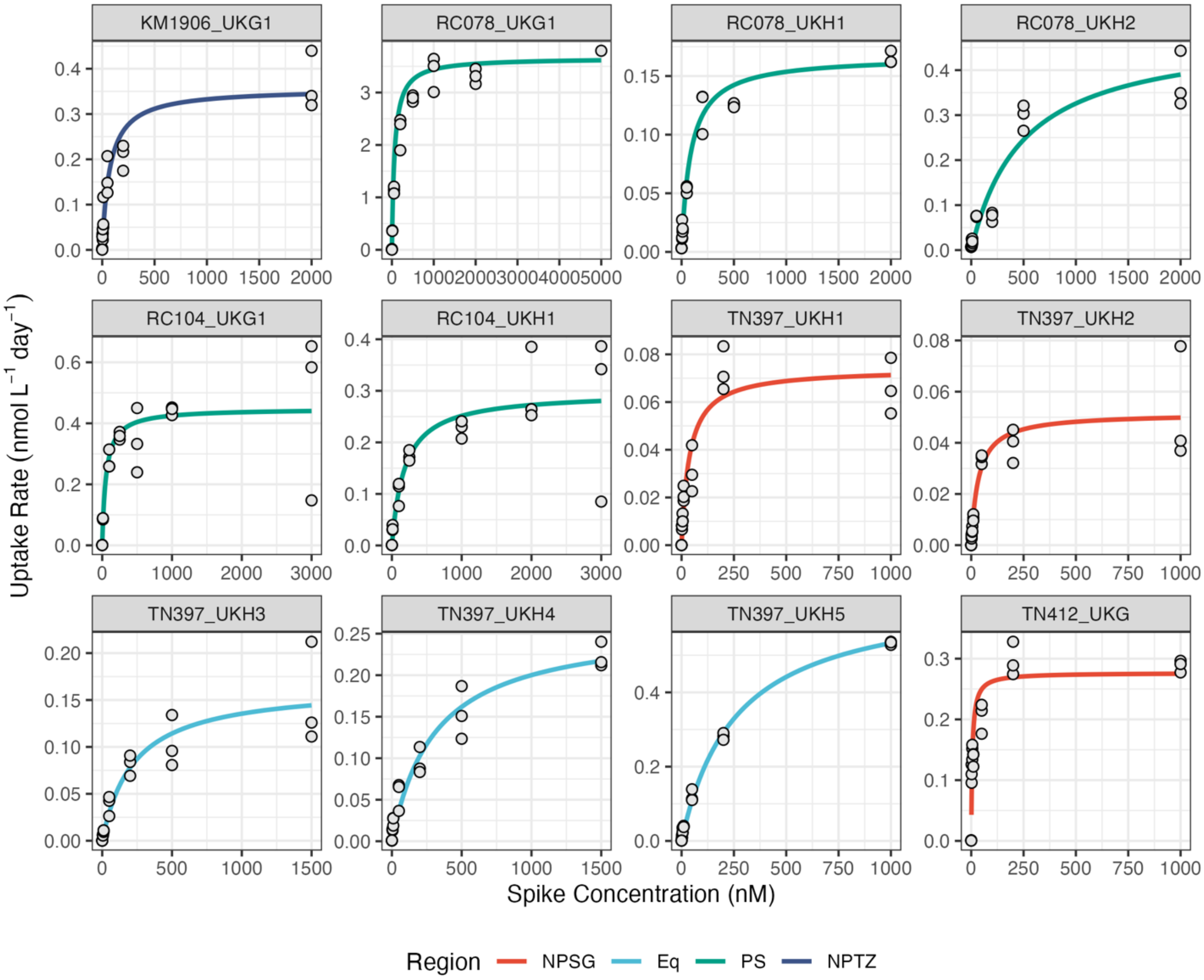

Figure 4.9: Results of uptake kinetics experiments where x-axis represents the spike concentration of the isotopically labeled compound (glycine betaine: UKG, homarine: UKH) and the y-axis represents the uptake rate of the labeled compound. Individual results from each bottle are shown as white circles and the fit Michalis-Menten kinetics equation (Table 1) is shown as the colored solid line. Color indicates the region the experiment was performed in.

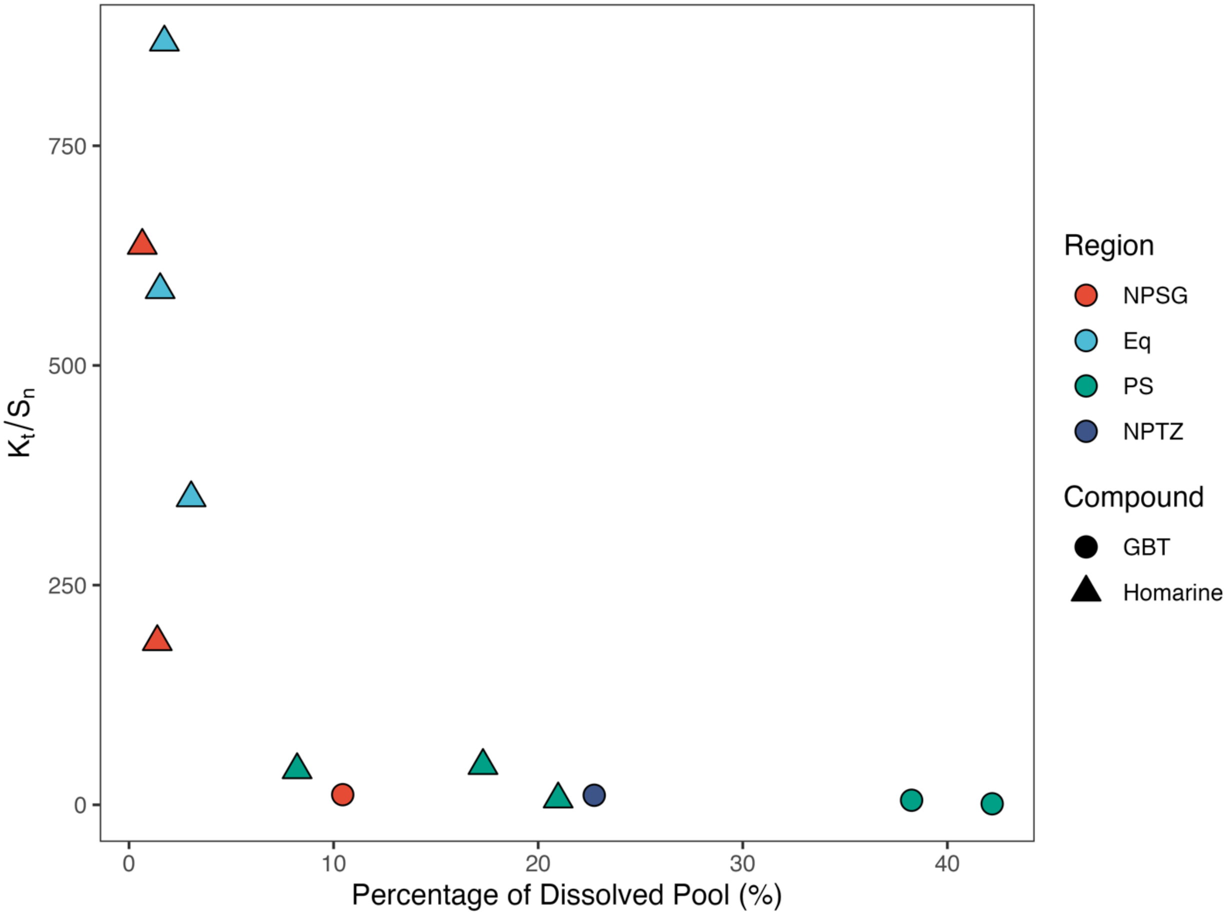

Figure 4.10: Comparison of K_t/S_n and the relative abundance of the compound investigated in the total dissolved pool of betaines and sulfoniums expressed as a percentage. Experiment region is shown by color and the compound tested is identified by shape.

4.12 Supplemental Tables:

Table 4.3: Experimental location and time details and environmental results of temperature, salinity, chlorophyll *a*, particulate carbon, particulate nitrogen, DIN, and primary productivity. This table is provided as a separate file.

Table 4.4: Environmental dissolved concentrations and batch specific limits of detection (LODs) for the betaines, sulfoniums, and TMAO measured in this study. LODs are presented with and without the blank subtraction that was applied to the environmental values. LODs with blank subtraction were used for this study to accurately compare to the blank subtracted environmental values. This table is provided as a separate file.

Table 4.5: Compiled uptake competition experiment results from the literature of GBT, DMSP, and choline. This table is provided as a separate file.

Table 4.6: Calculated K_t/S_n values and uncertainties and the relative abundance of the compound of interest (GBT or homarine) as a percentage of the total pool of betaines and sulfoniums (competitive inhibitors) for each experiment. This table is provided as a separate file.

Table 4.7: GBT and homarine fluxes presented as a percentage of the particulate and dissolved pools and as a percentage of daily primary production. This table is provided as a separate file.

Table 4.8: Compiled literature values on the dissolved concentrations, kinetics, turnover times, and fluxes for osmolytes including GBT, homarine, taurine, dissolved free amino acids (DFAAs), choline, and DMSP. This table is provided as a separate file.

Table 4.9: TQ-S settings, fragmentation, and quantification ion information for compounds measured in this study. This table is provided as a separate file.

Table 4.10: Wright-Hobbie plot derived K_t/S_n, Vmax, turnover times, and Monte Carlo derived uncertainties for all experiments. This table is provided as a separate file.

Chapter 5: A conserved osmolome in the North Pacific Ocean

5.1 Abstract[4]:

Osmotic stress is a key challenge for life. To overcome this challenge, organisms across all domains of life accumulate a small set of metabolites, termed osmolytes, in high concentrations to preserve cellular integrity and macromolecular structure, among other stress response and metabolic functions. Osmolytes are the most abundant small molecules measured in microbial systems and are hypothesized to be important currencies in microbial interactions and global biogeochemical cycles. However, our understanding of the concentrations, compositions, and controls on marine microbial community osmolomes, the sum of all osmolytes in a system, is limited. Here, we show that despite wide variation across taxonomic groups, intracellular and extracellular osmolome composition in surface communities is largely conserved across the North Pacific Ocean. Almost all osmolytes were highly correlated across our dataset and the relative abundance of each osmolyte was mostly consistent, despite total osmolyte concentrations varying by two orders of magnitude. Size fractionated osmolomes were distinct in composition and variable across regions, suggesting that that the producers of specific osmolytes change across space. While largely consistent, we identified increases in the relative abundance of sulfur containing osmolytes under high nitrogen conditions which we attribute to changes in microbial community composition. Osmolytes represent up to 4.2%, 2.6%, and 23%

---

[4] This paper is currently in preparation for submission as: Sacks, J.S., Carlson, L.T., Sosa, O., Boysen, A.K., Heal, K.R., Ribalet, F., John, S., Seleen, E., White, A., Armbrust, E.V., Ingalls, A.E. “A Convergent Osmolome in the North Pacific.” In preparation.

of particulate organic carbon, nitrogen, and sulfur pools, respectively, supporting the idea that these compounds are important components of marine elemental cycles. Our findings suggest that osmolyte composition will be robust to future changes in community composition and nutrient conditions, acting to stabilize marine communities and ecosystem level metabolic function in a changing ocean.

5.2 Introduction:

For life in salty environments, osmoregulation is a fundamental challenge. A cell must maintain osmotic balance while minimizing the deleterious effects of inorganic ions. To overcome this challenge, cells of all domains of life accumulate high concentrations of a small set of low molecular weight metabolites, called osmolytes or compatible solutes, that have similar osmotic properties to salt but preserve macromolecular structure and function[30,211]. Osmolytes also function to protect cells against fluctuations in pressure, temperature, hydration, and reactive oxygen species, and can serve as intracellular stores of fixed carbon and nitrogen, metabolic intermediates, and energy stores [11]. Marine microbes synthesize and accumulate osmolytes including free amino acids, betaines, sulfonates and sugars to extremely high intracellular concentrations (mM range), making them the most abundant metabolites measured in the ocean[10,16,25,31,32,38,182,211,212]. Due to their high concentrations, these compounds play a major role in structuring marine microbial interactions, acting as key growth substrates and signaling molecules, and are important pathways in global biogeochemical cycles of carbon nitrogen, and sulfur[11,19].

Controls on the concentration and composition of the osmolome (the sum of all osmolytes in a system) at the cellular level in individual organisms are well characterized. Total osmolyte concentrations are regulated primarily by salinity while composition is controlled by a mix of taxonomy, nutrient status, salinity, and biotic interactions[11–13,38,67,132,191,213]. However, controls on the concentration and composition of intracellular osmolomes at the community level are not well understood. Results from sparse environmental samples highlight both similarities in the abundant molecules detected and differences in the abundances of specific osmolytes across different communities, times of day, and sinkings vs. suspended particles[10,31,32,214]. The concentration and composition of extracellular osmolomes remains largely uncharacterized, but laboratory studies suggest that the composition of extracellular metabolite pools may diverge from intracellular pools[27,64,215]. Recent advances in separation science and liquid chromatography-mass spectrometry now the measurement of intra and extracellular osmolytes in environmental samples at scale[24,25]. In this study, we use these methods to map oceanic osmolomes at a basin scale across the North Pacific and ask: what controls the abundance and composition of intracellular and extracellular osmolytes in the ocean and how do microbial communities alter their osmoadaptation strategies across different nutrient regimes?

5.3 Results and Discussion:

We measured intracellular (particulate) and extracellular (dissolved) osmolytes in surface microbial communities on three cruises in the North Pacific (KM1906 and TN397) and Puget Sound (RC078) (Figure 5.1A, 5.1B, Table 5.1), cultures of dominant marine eukaryotic and prokaryotic microbial groups (Table 5.2), and a mesocosm nutrient amendment incubation experiment. The North Pacific cruises spanned four distinct biogeochemical regions: the low biomass, nitrogen-limited North Pacific Subtropical Gyre (NPSG), the iron-limited Equatorial Pacific upwelling region (Equator), the high biomass and productivity North Pacific Transition Zone (NPTZ) between the NPSG and the subpolar gyre, and the high nutrient California Current system (CC). On RC078 we sampled throughout the Puget Sound and Salish Sea estuarine system (PS) that represents a highly productive but variable coastal system. Our samples span a range of temperatures, salinities, community compositions, and nutrient regimes (Figure 5.1C)[7].

Osmolyte Concentrations Reflect Microbial Biomass

Summed particulate metabolite concentrations were primarily controlled by the total microbial biomass present (Table 5.4). In the open ocean systems, summed particulate osmolyte concentrations ranged from 2.0 nM in the NPSG to 46.4 nM in the NPTZ (Mean values: NPSG = 5.0±2.1 nM, Equator = 10.3±2.2 nM, CC = 11.3±5.1 nM, NPTZ = 19.6±11.0 nM) (Figure 5.1D). PS had high particulate osmolyte concentrations but high variability (mean = 192.7±217.2 nM). Stations 6 and 7 had high concentrations of particulate osmolytes (up to 702.5 nM, 15 times higher than the highest open ocean sample) but stations 2 and 3 having concentrations down to 13.3 nM, equal to or lower than many open ocean values (Figure 5.1E). Patterns of summed particulate osmolyte concentrations largely matched patterns of particulate organic carbon (POC, correlated with microbial biomass in marine systems), with POC explaining 83% of variability in osmolyte concentrations (linear model, $p < 0.001$) (Figure 5.1D, 5.1E, Figure 5.7). While organisms in culture increase their osmolome pools in response to increases in salinity, salinity showed a negative correlation with summed particulate osmolyte concentrations ($R^2 = 0.74$, $p < 0.001$), suggesting salinity was not the primary driver of differences in particulate osmolyte concentrations across our study area (Figure 5.7).

Summed dissolved osmolyte and summed total osmolyte concentrations were also closely correlated with POC. Although dissolved osmolomes did not include sugars or sulfonates, summed dissolved osmolyte concentrations (Range: 6.2–586.0 nM; Mean values: NPSG = 18.6±10.5 nM, Equator = 36.0±16.6 nM, CC = 42.7±23.4 nM, NPTZ = 62.9±33.6 nM, PS = 269.3±151.4 nM) were typically higher than particulate concentrations in the open ocean. In contrast, particulate osmolyte concentrations were often higher than dissolved concentrations in PS. For some compounds, measured dissolved osmolyte concentrations may be higher due to possible leakage from cells during filtering. Summed total osmolyte concentrations were typically only slightly lower than dissolved concentrations (Range: 9.0–810.6 nM; Mean values: NPSG = 22.0±10.7 nM, Equator = 42.5±17.3 nM, CC = 51.0±27.0 nM, NPTZ = 73.3±35.8 nM, PS = 405.4.3±219.0 nM). Summed dissolved and total osmolyte concentrations were closely correlated in log-log space with POC (dissolved: $R^2 = 0.76$, $p < 0.001$; total: $R^2 = 0.82$, $p < 0.001$) and summed particulate osmolyte concentrations (dissolved: $R^2 = 0.82$, $p < 0.001$; total: $R^2 = 0.89$, $p < 0.001$) suggesting that these two osmolyte pools both closely follow POC concentrations as well (Figure 5.7). Salinity also displayed a negative relationship with summed dissolved and total osmolyte concentrations (dissolved: $R^2 = 0.68$, $p < 0.001$; total: $R^2 = 0.74$, $p < 0.001$), further demonstrating that microbial biomass, not salinity, controls bulk osmolyte concentrations in the ocean (Figure 5.7).

Osmolytes are Important Components of Particulate Organic Carbon, Nitrogen, and Sulfur Pools

Osmolytes were an important component of marine POC and particulate nitrogen (PN) pools. In the open ocean, particulate osmolytes comprised 0.4-2.9% of POC (NPTZ: mean = 1.4±0.8 %; NPSG: mean = 1.5±0.5 %; Equator: mean = 2.1±0.5 %) and 0.1-2.6% of PN (NPTZ: mean = 1.0±0.6 %; NPSG: mean = 1.4±0.5 %; Equator: mean = 1.9±0.4 %) in our open ocean regions (Supplemental Figure 2). ATP based estimates of the contribution of living biomass to the total POC pool suggest ~30% of POC is in living biomass in the NPSG[124,216]. Assuming particulate osmolytes are only in living organisms, they may represent closer to 3–7 % of living carbon biomass in the NPSG. We observed latitudinal shifts in the percent of osmolytes in POC, with osmolytes representing a higher portion of POC and PN at the Equator and the northernmost section of the transition zone (Figure 5.8). In PS, particulate osmolytes represented 0.5–4.3% of POC (mean = 1.8±1.0 %) and 0.06–1.65% of PN (mean = 0.6±0.4 %), with the high values reflecting high amounts of living phytoplankton biomass relative to detritus (stations 6, 7) and the lower values (stations 3, 5) likely reflecting high detrital organic matter, likely from both marine and terrestrially sources.

Particulate osmolytes potentially represent a major component of the marine particulate organic sulfur (POS) pools. The average ratio of C:S and N:S in particulate osmolytes (22±8.2 and 3.2±1.5, respectively) is far lower than the ratio of POC:POS and PON:POS in phytoplankton biomass (95 and 12.3, respectively)[217–219]. Therefore, we expect osmolytes to contribute disproportionately to marine sulfur pools and cycling. POS was not measured here, but assuming a fixed C:S of 95:1 for organic matter, we estimate that in the open ocean 3.2–42.4 % of POS is in particulate sulfur containing osmolytes with regional variability (NPTZ: mean = 15.3±10.1 %; NPSG: mean = 5.8±1.5 %; Equator: mean = 8.7±1.7 %) (Supplemental Figure 2). In our coastal samples, particulate sulfur osmolytes represented a high but variable amount of

POS 1.3–48.7 % (mean = 12.6±13.0 %). Our values potentially represent an underestimate of sulfur osmolyte concentrations as the abundant sulfur metabolite DMSP may leak out of cells during sampling but may also be an under- or overestimate if the POC/POS ratio is not fixed[81]. Our estimates of the percentage of POS in osmolytes are far higher than our estimates of the percentage of POC and PN in osmolytes in open ocean samples (0.4-2.9% % and 0.1-2.6%, respectively). The global inventory of POS in phytoplankton biomass in the surface ocean is estimated to be 0.028 Pg S with an annual assimilation rate of 1.36 Pg S per year. Therefore, our data suggest that the sulfonates taurine, isethionate, and DHPS and the sulfoniums gonyol, DMSP, and DMSA have combined marine inventories of 0.001–0.012 Pg S and annual fluxes of 0.049-0.57 Pg S per year, similar in magnitude to estimates of terrestrial emissions of sulfur to the atmosphere (0.067 – 0.132 Pg S per year)[218]. We conclude that these sulfonates and sulfonium compounds and the metabolisms that produce and consume them are major components of the global sulfur cycle.

Osmolome Composition is Largely Conserved Across Diverse Marine Biomes

We detected most osmolytes in all of our samples, suggesting that these compounds are ubiquitous in the marine environment. Of the compounds we were able to measure, we detected 22/33 compounds in every particulate sample and 29/33 in over 90% of our particulate samples. In the dissolved samples, we detected 11/26 compounds in every sample and 17/26 in over 90% of our samples. The compounds not detected in some samples include the extremely low abundance compounds such as arsenobetaine and compounds taxonomically restricted to a small set of organisms such as sucrose which is almost exclusively produced in high abundance by *Prochlorococcus*[32,38]. The most abundant compounds include those identified as important osmolytes in other environmental and culture studies including the amino acids glutamic acid, hydroxyisoleucine, and (iso)leucine, the betaine glycine betaine, the sugar glucosylglycerol, the sulfoniums gonyol and DMSP, and the sulfonates isethionate and taurine (Figure 5.2A-F)[10,16,31,32,182,212,214].

Across all studied regions, particulate osmolomes were composed of 49±13 % amino acids (mean ± standard deviation), 15±11 % sulfonates, 14±4 % betaines, 13±6 % sulfoniums, and 4±4 % sugars, and 2±1 % of ectoines, arsenobetaine, and TMAO (Figure 5.2A, 5.2B). Dissolved osmolomes were composed of 32±16 % amino acids, 25±10 % betaines, 25±14 % sulfoniums, 17±11 % TMAO, and less than 1% ectoines and arsenobetaine (Figure 5.2C, 5.2D). We observed a high level of congruence in size and composition between the particulate and dissolved pools for compounds measured in both phases, contrasting with culture studies [27,64,215]. A mantel test suggested that dissolved osmolomes were closely related to particulate osmolomes (multivariate $R^2 = 0.85$, $p < 0.001$). This result suggests that in natural systems, the amount and composition of intracellular osmoloytes produced by the microbial community acts as a primary control on the dissolved osmolome.

We observed a consistent osmolome composition across regions in the particulate, dissolved, and total fractions (Figure 5.2). Our sample set crosses large gradients in community composition and nutrient status due to changes in nutrient supply rates, temperature, and biotic interactions[7,220]. Therefore, we predicted there would be dramatic shifts in osmolome composition across these gradients, reflecting community taxonomic and physiological shifts. Instead, we observed that osmolome composition was largely conserved across our sampling locations with some regional differences, particularly in the NPTZ and PS samples (Figure 5.2). A principal component analysis (PCA) on the log-transformed osmolyte concentrations revealed that a single principal component explained over 70%, 52%, and 63% of the variability in particulate, dissolved, and total osmolyte pools (Figure 5.3A–C). PC1 was tightly correlated with the log10 transformed POC (particulate: $R^2 = 0.89$, $p < 0.001$; dissolved: $R^2 = 0.81$, $p < 0.001$; total: $R^2 = 0.87$, $p < 0.001$), further supporting the argument that osmolomes are primarily controlled by biomass with a relatively constant composition (Figure 5.3D–F). A single principal component explaining most of the variability in a sample set suggests that most variables in the dataset are correlated. Indeed, we observed that in our particulate, dissolved, and total osmolyte datasets, 29/32, 15/23, and 19/23 osmolytes were correlated ($r > 0.5$) with the most abundant particulate amino acid osmolyte (glutamic acid), respectively (Figure 3G–I; Figures 5.9, 5.10, 5.11). Compounds not correlated with glutamic acid were often unique cases such as arsenobetaine, a degradation product of arsenosugars and arsenolipids that is most abundant in the mesopelagic, and DMSP, which is known to leak out of cells during sample collection[32,71,221]. Our findings are similar to observations in other marine targeted metabolomics datasets that show largely consistent patterns of subsets of the metabolites measured in this study across the North and South Atlantic and along the same 158°W transect in the North Pacific as our KM1906 samples[31,32,214].

The conservation of osmolomes across diverse oceanic regions is striking in comparison to the highly heterogenous osmolomes of individual cultures of marine phytoplankton, bacteria, and archaea (Figure 5.4A, Figure 5.12, Table 5.5) [32,38]. Most groups of organisms only produce a few osmolytes in high abundance and often in different ratios than they are present in the environment (Figures 5.2A, 5.2B, Figure 5.12). However, many osmolytes are produced by multiple organisms from diverse lineages. For example, glycine betaine accounts for at least 10% of the average osmolome of *Synechoccocus*, Prasinophytes, Diatoms, and Haptophytes (Figure

5.4A). Therefore, we interpreted the convergence observed in marine osomolomes to represent widespread conservation of function in osmolyte production at the community level, despite changes in taxonomy.

We tested our hypotheses that the community level production of an osmolyte is conserved despite taxonomic shifts using size fractionated samples collected along a transect from the NPSG to the NPTZ in 2017 along the same transect as KM1906. We found that the fraction of a compound in our small size fraction (< 3 μm) compared to our large size fraction (> 3 μm) varied substantially across the transect for many compounds, suggesting changes in producers of these compounds in the environment (Figure 5.4B, Table 5.6). For example, *Prochlorococcus*, Prasinophytes, and Diatoms all accumulate aspartic acid to high intracellular concentration. Along the transect, the percentage of aspartic acid in the small size fraction shifts from 48% in the oligotrophic NPSG (latitiude 28.46) to less than 20% in the highly productive NPTZ (latitude 39.75), suggesting a shift in the dominant producers from *Prochlorococcus* to diatoms and other eukaryotes (Figure 5.4B). Aspartic acid changes by just 0.5% in relative abundance (0.3–0.8%) across the same samples. We found substantial changes (>10%) in the percentage of a compound in the less than 3 μm size class across the transect for 30/31 compounds with 16/31 changing more than 20%, suggesting substantial changes in producers for almost all osmolytes. These estimates for changes in producers represent lower bounds as large changes in the producers of a compound can occur within these size classes as well. While largely consistent in our overall dataset, we observe some variability in the relative abundances of compounds in our size fractionated samples (Figure 5.4B). For example, homarine ranges in relative concentration from 0.6 % in the NPSG to 14 % in the NPTZ, likely related to the high abundances of *Synechococcus* in this region [47]. However, most compounds have greater variability in their size fraction partitioning than their overall relative abundance (Figure 5.13). Our results demonstrate that at the community level, osmolome composition converges due to the production of a conserved suite of osmolytes by taxonomically distant organisms.

Convergence in the relative composition of organic matter in the ocean across ecosystems is well documented. The Redfield ratio describes the relatively consistent ratio of carbon, nitrogen, and phosphorus (canonically 106:16:1) in the ocean[222]. The relative abundance of amino acids in proteins, the relative abundance of functional groups in organic matter, and the overall macromolecular composition of freshly produced organic matter also remain largely constant across marine ecosystems[223–227]. Our results document another aspect of marine organic matter in the surface ocean that is conserved across disparate biogeochemical conditions. This largely consistent composition hints at an evolutionary benefit for mixed communities to contain a range of osmolytes in these relative proportions. Many osmolytes play slightly different roles in stabilizing macromolecules due to different chemical interactions with water, proteins, and other osmolytes, suggesting that a mixture of osmolytes may be advantageous[30,228,229]. While each organism makes only a subset of these compounds, most marine phytoplankton and bacteria contain transporters allowing for the uptake of many of these compounds[189,191,213,230]. Therefore, osmolyte pools may partially serve as a community resource with total pool size and composition partially regulated by osmolyte exchanges rather than simply production at the cellular level.

A “Redfield Ratio” for Betaines

We also observed the conservation of osmolome composition within specific compound classes. The betaines had consistent relative abundances across all samples (Figure 5.14). In particulate, dissolved, and total osmolomes, glycine betaine (GBT) is almost always the most abundant compound, followed by beta-alanine betaine, and homarine, with the other measured betaines at consistently low relative abundances (Figure 5.5A–C). We were therefore able to calculate a "Redfield Ratio" for betaines across our dataset, standardizing our results to proline betaine, consistently the least abundant compound. The "Betaine Redfield Ratio" is 48 : 26 : 14 : 3.4 : 3.4 : 1.5 : 1.3 : 1 (glycine betaine : beta-alaninebetaine : homarine : trigonelline : TMAB : carnitine : betonincine : proline betaine) for particulate osmolomes, 125 : 41 : 16 : 4.9 : 2.7 : 1.3 : 1.8 : 1 for dissolved osmolomes; and 100 : 36 : 14 : 4.1 : 3 : 1.2 : 1.5 : 1 for total osmolomes. While not identical, all three datasets show a clear power law type distribution, where a few compounds (GBT, beta-alanine betaine, and homarine) are typically much more abundant than every other compound.

This non-random relationship suggests a benefit, and therefore evolutionary pressure, for communities to produce these structurally and functionally similar compounds in this order. We explored aspects of the chemical structure of these compounds and found a relationship between molecular weight (shown as mass to charge ratio, *m/z*) and average rank. Larger, heavier compounds were less abundant (higher rank) than smaller, lighter compounds in particulate ($R^2 = 0.71$, $p < 0.001$), dissolved ($R^2 = 0.69$, $p = 0.007$), and total ($R^2 = 0.68$, $p = 0.007$) (Figure 5D–F). This result suggests that at the ecosystem level, organisms are preferentially making smaller compounds than larger ones. These smaller compounds require fewer resources and likely require less molecular machinery to synthesize, making them cheaper and more efficient to produce. This observation of greater production of smaller compounds provides at least a partial answer to the longstanding question of why organisms produce some osmolytes over others.

Particulate osmolomes display regional variability in stoichiometry due to community composition shifts related to nitrogen availability

While approximately 70% of variance in osmolomes can be explained by a single principal component, differences in microbial community composition and physiology in response to different environmental conditions likely explain the remaining 30%. In our PCA analyses, samples cluster by region (Figure 5.3). We suspected some of these differences may be attributable to the different nutrient supply regimes across these regions. In response to nutrient stress, microbes remodel their cellular composition to reduce demand for limiting nutrients and maintain growth rates. For example, in low phosphorus regions of the ocean, cyanobacteria switch to using sulfur-containing lipids rather than phosphorus-containing lipids to lower their phosphorous demand[36,37]. Microbes have been shown to remodel their osmolomes to switch from N-containing osmolytes to N-free osmolytes in N-stressed cultures[11]. We therefore hypothesized that in response to N-stress, we would observe a community level shift in osmolome stoichiometry to have higher C/N ratios and lower N/S ratios.

We tested this hypothesis in a month long mesocosm experiment where natural mixed environmental communities taken from Station ALOHA in the NPSG were exposed to additions of N, P, and Fe alone and in combination[231]. Contrary to our expectations, the osmolome of communities receiving N additions had higher C/N ratios (mean = 15.2 ± 5.5, Three-Way ANOVA, $p < 0.001$) and lower N/S ratios (mean = 0.5 ± 0.3, Three-Way ANOVA, $p < 0.001$) compared with communities not receiving N additions (C/N: mean = 8.5 ± 1.4; N/S: mean = 2.6 ± 1.3) (Figure 5.6A, 5.6B). No other nutrient amendments resulted in changes in osmolome stoichiometry, and no interaction effects among nutrients in combination were observed (Three-Way ANOVA, $p > 0.05$). The lower N/S ratio was driven by increases in the abundance of sulfur-containing sulfonates, particularly isethionate (Figure 5.15). Isethionate is a two-carbon sulfonate produced by prasinophytes and diatoms, likely from taurine[16]. Community composition analysis showed dramatic increases in diatom abundance under N-addition[231]. These results suggest that changes in osmolome stoichiometry related to ecosystem nutrient status are driven primarily by changes in community composition rather than by physiological adaptation.

We then looked for evidence of osmolome stoichiometry shifts in the environment and found that in our open ocean systems (NPSG, NPTZ, CC, Equator), osmolome N/S ratios were significantly different ($p < 0.001$, ANOVA). In these systems, we found that regions with higher nutrient supply rates and biomass (ex. NPTZ, CC) had lower osmolome N/S ratios (NPTZ: mean = 1.3 ± 0.5 , Tukey HSD $p < 0.001$; CC: mean = 2.7 ± 0.9, Tukey HSD $p = 0.017$) compared with the lowest nutrient supply region (NPSG: mean = 3.5 ± 1.4) (Figure 5.6C). These changes also appeared to be primarily driven by increases in the concentrations of isethionate and other sulfonates (Figure 5.2A). These low N/S regions also display higher relative diatom biomass than the lowest nutrient supply regions[7,220]. In a 2015 study following the same transect from the NPSG to NPTZ, our group observed similar latitudinal patterns of the S-containing osmolytes taurine, isethionic acid, and DHPS, suggesting interannual reproducibility in our observed stoichiometric patterns[32]. Our results show that community level osmoadaptation responds to nutrient supply due to changes in microbial community composition. Counterintuitively, microbial communities in N-replete environments rely more on sulfur to fulfill their osmoregulation requirements than communities in N-deplete environments. It is worth noting that N-supply likely does not entirely explain the osmolome N/S ratio as both the Equator and NPTZ have high nitrogen supply rates, but the Equator has more similar N/S ratio to the NPSG than the NPTZ, potentially due to differences in other nutrients such as iron. However, our results still demonstrate a connection between inorganic nitrogen supply, the growth of eukaryotic phytoplankton, and the accumulation of sulfonates, showing that sulfur osmolytes link the marine nitrogen and sulfur cycles.

Why some eukaryotes favor S-containing osmolytes is not yet understood. S-containing osmolytes may save N for other uses, allowing N to be directed towards growth and cellular replication in environments where fast growth is required to remain competitive. Sulfonates are also accumulated on diel cycles, suggesting roles in redox management and energy storage[16]. The production of sulfonates may also facilitate adaptation to Fe or vitamin B12 stress. Nitrate reduction to ammonia is an iron-intensive process and iron-limited diatoms are less able to use nitrate to grow[232–234]. The use of sulfur-containing osmolytes may also reduce nitrogen-demand through their relatively short biosynthetic pathways, and as a result, iron-demand, facilitating growth in low-iron environments. Many compatible solutes including betaines and sulfoniums require methylation through vitamin B12 dependent pathways. Some of these compounds including DMSP and GBT are decreased under vitamin B12 limiting conditions[12,235]. Sulfonate synthesis does not require SAM or vitamin B12, suggesting that these compounds could facilitate adaptation to lower B12 environments.

Community level functional redundancy in osmotic stress response may stabilize microbial communities in the face of global change

Functional redundancy, or the conservation of function at the ecosystem level despite taxonomic shift, is well documented in marine microbial communities and includes a range of ecosystem properties relate to biogeochemical cycling (ex. nitrogen fixation) and organic matter decomposition (ex. polysaccharide degradation)[236,237]. Our results suggest that the relative abundance of functional groups of osmolyte producers (all microbes that produce a specific osmolyte) are largely conserved across most marine ecosystems with differences primarily relating to eukaryote-specific sulfur osmolytes. Open questions remain on how and why osmolome composition remains consistent across communities, but we hypothesize that the widespread potential for osmolyte uptake and exchange may play a role in facilitating community level osmoregulation.

The conservation of particulate and dissolved osmolome composition may act as a stabilizing force on marine heterotrophic communities. Osmolytes are important growth substrates for dominant marine bacterial groups including SAR11 and roseobacters which each make up significant fractions of marine heterotrophic bacterial communities in the surface ocean . Laboratory experiments show that a mixed environmental community of bacteria grown on different osmolyte pools selects for different community compositions[17]. Therefore, our observation of a conserved osmolome suggests that similar bacterial functional groups may be selected for across different regions as similar growth substrates will be present in similar ratios. However, different total osmolyte concentrations across regions may still select for oligotrophs and copiotrophs in different environments, depending on transporter affinity and growth rate[239]. Additionally, while the most abundant osmolytes may serve as the primary growth substrates for the community, the less abundant osmolytes such as trigonelline, proline betaine, and DMSA may support low level resource partitioning that can help maintain bacterial functional and taxonomic, diversity even when a dominant growth substrate is shared[240]. Additionally, heterogeneous microenvironments such as phycospheres may provide osmolytes in non-standard ratios and amounts, supporting additional diversity in osmolyte consumers[15].

Functional redundancy is typically associated with resilience in microbial systems as the loss of a single member from a functional guild can be compensated for by others[236,237]. The ecosystem level functional redundancy in osmolyte pools therefore suggests that in relative composition it will be robust to future changes as the global ocean becomes warmer and more stratified[241,242]. This in turn will act to help stabilize the composition and phenotypes of marine heterotrophic bacterial communities that rely on osmolytes as substrates. However, reduced N-supply to the surface ocean will likely reduce the overall concentrations of marine osmolyte pools as biomass decreases. The shift to lower N-environments will also shift osmolomes to have less S-containing osmolytes, decreasing organic sulfur pools and cycling in the ocean.

5.4 Methods:

A note on environmental metabolite measurements:

Metabolite measurements in marine microbial communities are difficult to make due to a wide array of analytical challenges associated with sampling, extraction, and measurement[19,24,81,194]. Our particulate methods capture the majority of osmolytes documented in the literature (Table 5.3), but we are missing several important compounds including the sugar alcohols glycerol and inositol[33]. Our approach for extracting compounds from seawater, cation-exchange solid phase extraction, allows us to measure zwitterionic compounds that can be protonated (amino acids, betaines, TMAO, and sulfoniums) but is unable to extract sugars and sulfonates[25]. Additionally, some osmolytes such as DMSP may leak out of cells during sampling, potentially decreasing the measured particulate concentrations and increasing the measured dissolved concentrations[81]. We therefore present our results as particulate, dissolved, and total pools and acknowledge the uncertainties present in these values. Our results are similar to observed concentrations of osmolytes in other particulate and dissolved metabolomics studies on marine microbial communities, suggesting that for the compounds measured in this study, analytical artifacts have been minimized and are consistent with past measurements[10,32,80,83,180,181,212,214].

Environmental Sampling:

Samples were collected for particulate and dissolved metabolites on four cruises: MGL1704 (Spring 2017), KM1906 (Spring 2019), TN397 (Winter 2021), and RC078 (Summer 2022). These cruises sampled across the North Pacific (the North Pacific Transition Zone, the North Pacific Subtropical Gyre, and the Equatorial Pacific) and the Salish Sea and Puget Sound. Samples for particulate metabolites were collected as previously described[32]. Briefly, seawater was collected in triplicate samples from underway seawater intake systems (MGL1704, KM1906, TN397) and from Niskin bottles attached to a CTD rosette from the top 10 meters of the water column (RC078). Water was filtered using peristaltic pumps onto 0.2 um 147 mm Omnipore PTFE membrane filters (MGL1704, KM1906) or 0.2 um 47 mm Durapore PVDF membrane filters (TN397, RC078), flash frozen in liquid nitrogen, and stored at -80 °C until extraction. Samples for MGL1704, KM1906 and TN397 were prefiltered through 100 um nylon mesh to remove large, rare organisms but prefiltration was not performed on RC078 samples as a larger fraction of coastal biomass is >100 μm. For small size fraction (< 3 μm) samples on MGL1704, an inline 3 μm PES Supor filter was used as an additional prefilter. Dissolved metabolite samples were collected on KM1906, TN397, and RC078 by collecting the filtrate from the particulate samples in acid washed 50 mL polypropylene falcon tubes (KM1906, TN397) and 60 mL polycarbonate bottles (RC078). Dissolved metabolites samples were then frozen at -20 °C until extraction. Blanks for particulate samples were collected by refiltering the filtrate onto new filters and handled identically to environmental samples.

Mesocosm Incubations:

Particulate metabolite samples were collected from PERI-FIX, a 29-day mesocosm study testing the effects of nutrient amendment on microbial communities North Pacific Subtropical Gyre[243]. Briefly, water was collected from 21.0892 N, -158.0743 W and incubated in PERIcosm incubators at the University of Hawai'I Marine Center in Honolulu, HI, USA in August–September, 2021[231]. Triplicate tanks were either unamended (control, C) or spiked with N (75 nmol of ammonia and 75 nmol nitrate per L of seawater), P (9.4 nmol phosphate per L of seawater), and F (1 nmol iron chloride per L of seater), individually or in combination (NP, NF, PF, NPF). Tanks were incubated on 12 hour day/night cycles at 25 °C. Particulate metabolite samples were collected at the beginning of the experiment (Tote samples) and the end of the experiment in the same manner as the environmental particulate samples.

Culture Samples

Culture samples were reanalyzed from previous studies using updated liquid chromatography-mass spectrometry approaches to enable more consistent comparisons to environmental samples. Organism details are listed in Supplemental Table 2 and culturing details are in Durham et al. and Heal et al.[32,38].

Auxiliary Data:

Particulate organic carbon (POC) and particulate nitrogen (PN) samples were collected from the underway seawater intake systems (KM1906, TN397) or Niskin bottles on the CTD rosette (RC078). POC samples were analyzed according to HOT protocols (KM1906, TN397) or University of Washington Marine Chemistry Lab protocols (RC078). Temperature and salinity measurements were taken using sensors on underway seawater intake systems (KM1906, TN397) or CTD (RC078). For KM1906 and TN397, samples were collected on the same transect but were not directly paired with metabolite samples. We therefore performed a Stineman interpolation (stinepack package, R) to estimate POC and PN values that correspond to our metabolite measurements. Particulate organic sulfur was estimated from POC measurements by assuming a fixed ratio of 95:1 POC:POS.

Metabolomics:

Identification of Osmolytes:

We identified osmolytes using a literature search of primary and secondary studies showing accumulation of the compound in response to salt stress (detailed in Table 5.3)[30,33,67,221,230,244–253]. The compounds identified were amino acids, betaines, sugars, sulfoniums, sulfonates, ectoines, trimethylamine N-oxide, and arsenobetaine.

Data Acquisition:

Particulate metabolites were extracted as described in Boysen et al.[24]. Briefly, we used a modified Bligh and Dyer extraction which separates compounds into a polar phase (methanol and water) and a nonpolar phase (dichloromethane). Only the polar fraction was analyzed for this study. Dissolved metabolites were extracted as described in Sacks et al.[25]. Briefly, we used cation-exchange solid phase extraction to separate positively charged and zwitterionic metabolites from salts and other dissolved organic matter. Following both extraction procedures, compounds were dried down under N_2 gas, reconstituted in water with internal standards, filtered through a 0.2 μm syringe filter to remove any particulate matter, and stored at -80 ºC until data acquisition. The internal standards used in this study are detailed in Table 5.7. Metabolomics data were collected using liquid chromatography mass spectrometry with the settings detailed in Sacks et al.[25]. Briefly, we used hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray ionization liquid chromatography-mass spectrometry source with a Q-Exactive HF Orbitrap Mass Spectrometer. Data were collected in full scan mode with polarity switching to measure compounds in both positive and negative ion modes.

Data Processing, Quantification, and Quality Control:

Metabolite data were converted from .raw files to .mzML files using MS Convert and chromatograms for osmolytes were integrated in skyline[75,76]. We identified osmolytes based on exact mass and comparison in retention time space to an authentic internal standard. Particulate metabolites were normalized and quantified as in Heal et al. and dissolved metabolites were quantified as in Sacks et al.[25,32]. Briefly, raw peak areas were normalized using best-matched internal standard normalization and quantified by comparing to an authentic standard in water and a pooled sample (all environmental samples combined) within each batch on the instrument or using a matched internal standard added before extraction. The unlabeled authentic standards enable the calculation of a response factor (RF: the ratio of peak area to concentration for a given compound) and a response factor ratio (RFratio: the ratio of the RF in water to the RFin a pooled sample).

In some instances, RFs and RFratios may be difficult to calculate for a certain sample set if a standard was not run in that analytical batch or there are high natural concentrations of a compound in the pooled sample. If RFs or RFratios could not be calculated or were far outside the normal range for our measurements (RFratio < 0.3 or > 3), we took the mean of all RFs or RFratios across the batches analyzed in this project to calculate a consensus RF and RFratio. We then used these consensus values in quantification calculations for the compound of interest.

For dissolved measurements, we adjusted concentrations for extraction efficiencies and calculated batch specific limits of detection as detailed in Sacks et al.[25]. We performed blank subtraction on both the dissolved measurements and limits of detection by subtracting the mean concentration in the blank from both values. We applied quality control to our samples by requiring measurements in particulate samples to be 3 times greater than the average blank concentration for all analytical batches, and measurements in dissolved samples to be above batch-specific limits of detection.

Estimating dissolved GBT concentrations in KM1906 Samples:

Dissolved samples for KM1906 were analyzed without “Cation-Exchange Clean” (CXC) procedures. These procedures are required to remove any remaining zwitterionic metabolites from Milli-Q cleaned water before analysis. Glycine betaine contamination is observed in our Milli-Q system and measurements made without CXC procedures are therefore not trustworthy[25]. We therefore removed measurements of glycine betaine from this sample set. However, glycine betaine displays high correlations with other betaines in the environment, particularly homarine ($R^2 = 0.64$). We used this relationship between homarine and glycine betaine to predict dissolved glycine betaine concentrations in the KM1906 sample set to facilitate comparisons of absolute and relative abundances with our other cruises.

Size Fractionation Estimates

We estimated the relative partitioning of osmolytes between size fractions for the MGL1704 samples. We calculated the average total concentration of each osmolyte in the full-size fraction samples by calculating the mean concentration in each triplicate. We then divided the concentration in the small size fraction sample by the total to calculate the fraction of the osmolyte measured in the small size fraction (fraction less than 3 µm).

Osmolome Particulate Carbon, Nitrogen, and Sulfur Estimates

We calculated the concentration of carbon, nitrogen, and sulfur in osmolytes in each sample by multiplying the number of carbon, nitrogen, and sulfur atoms in each osmolyte by the concentration of that compound. We then summed the total concentration of carbon, nitrogen, and sulfur in osmolytes in each sample. We compared this value to measurements of POC and PON as well as estimates of POS to estimate the percentage of POC, PON, and POS in osmolytes. We also used the estimates of total carbon, nitrogen, and sulfur in osmolytes to calculate the overall osmolome stoichiometry (C/N ratio, C/S ratio, N/S ratio).

Statistical Approaches

We used a principal component analysis (PCA) to constrain the sources of variability in osmolome composition and concentration. To perform the PCA, we applied a log10 transformation to all measurements and range scaled the values of each osmolyte so that all values for each compound were between 0 and 1. We then used the “rda()” function in in the vegan package (version 2.6-6.1). To assess the relationship between particulate and dissolved osmolomes, we performed a mantel test on range-standardized, Euclidean distance matrices of particulate and dissolved metabolites using the “mantel()” function in the vegan package. To most accurately reflect the data, we did not remove, replace, or interpolate concentrations of compounds not passing blank or limit of detection thresholds in our multivariate analyses. For all other statistical tests we used the “aov()” function for ANOVAs, the “TukeyHSD()” function for post-hoc tests, and the “lm()” function for linear models. All analysis were performed in R version 4.4.1.

5.5 Acknowledgements:

We thank the captains, crews, and science parties of the *R/V Kilo Moana* on cruise KM1906, the *R/V Thomas G. Thompson* on cruise TN397, the *R/V Marcus G. Langseth* on cruise MGL1704, and *R/V Rachel Carson* on cruise RC078. We thank Kate Faber, Natalie Kledzik, Susan Garcia, Claudia Luthy, and Anna Finch for assistance with fieldwork and laboratory analyses, Will Kumler, Frank X. Ferrer-González for valuable conversations and assistance with data analysis. This work was supported by grants from the National Science Foundation (2125886 to AEI and KRH and 2124712 to OAS) and grants from the Simons Foundation (LS award ID 385428, AEI; SCOPE Award ID 329108 to AEI; and SF award ID 598819, KRH).

5.6 Figures:

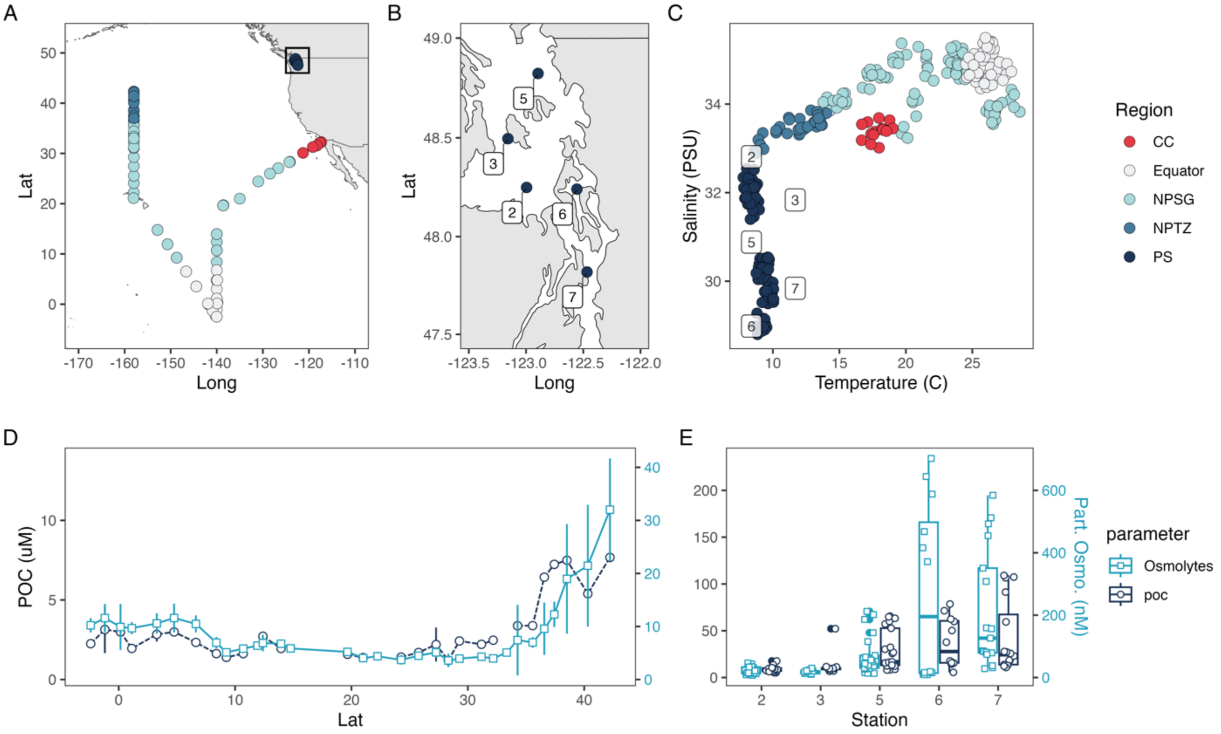

Figure 5.1: Locations and environmental context of sampling locations in the North and Equatorial Pacific Ocean (A) and Puget Sound and the Salish Sea (B). Temperature and salinity plots for all samples (C). In A–C, color indicates region (CC: California Current, NPSG: North Pacific Subtropical Gyre, NPTZ: North Pacific Transition Zone, PS: Puget Sound and the Salish Sea). Number labels in B and C indicate station numbers for PS samples. Total particulate osmolyte concentrations (blue) and particulate organic carbon (POC) measurements (black) along the open ocean transects (NPSG, NPTZ, and Equator) (D) and at each station in PS (E). In D, POC and osmolyte measurements are grouped into 1 degree latitude bins and the average (points) and standard deviation (error bars) are shown. In E, individual samples are shown as points with boxplots showing the median and quartiles for each station.

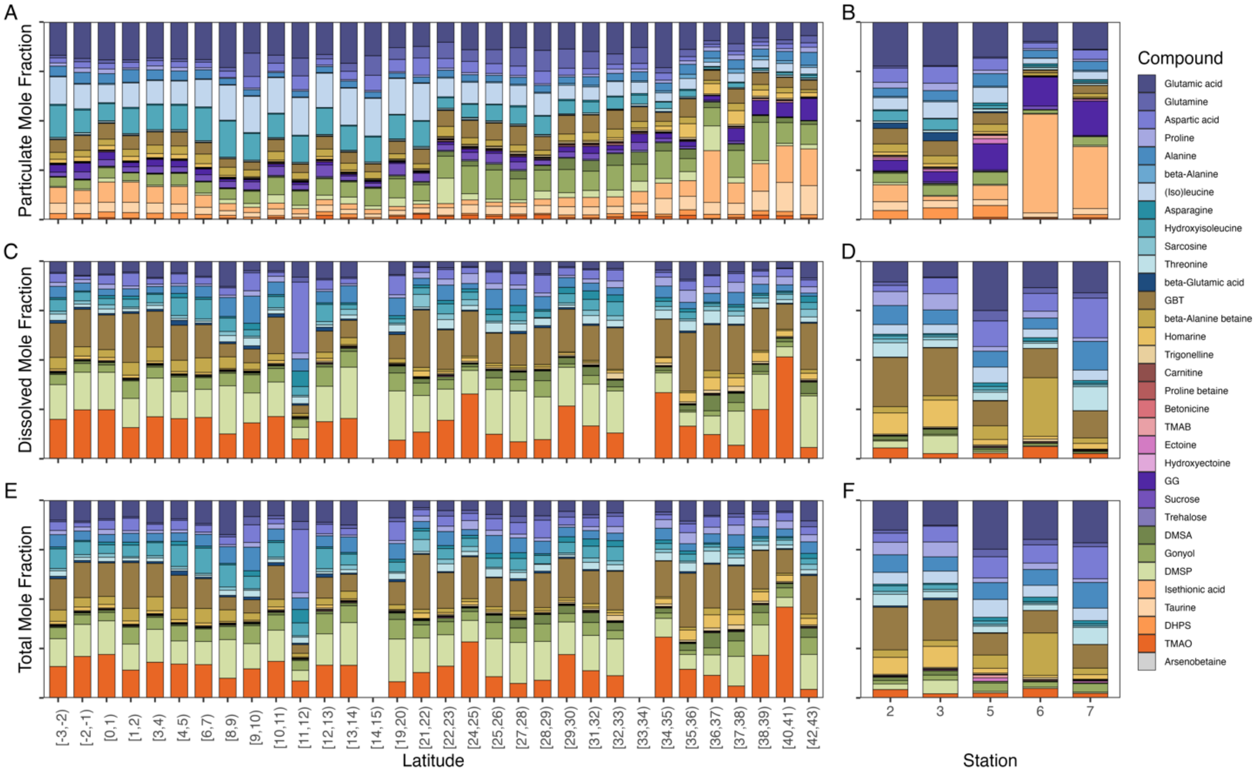

Figure 5.2: Relative concentrations of osmolytes in particulate (A, B), dissolved (C, D), and total (E, F) osmolomes for the combined open ocean latitudinal transects of the Equator, NPSG, and NPTZ (A, C, E) and for the estuarine samples from PS (B, D, F). For A, C, and E, osmolyte measurements are grouped into 1 degree latitude bins and the average relative concentration for each compound is shown. For B, D, and F, the average osmolyte concentration for samples collected from surface (top 10 m) are shown. Total osmolyte measurements only include compounds measured in both particulate and dissolved phases. Compound identity is shown by color with different color groups indicating compound class (AA: blues, betaines: yellow and reds; ectoines: pinks; sugars: purples; sulfoniums: greens; sulfonates: oranges).

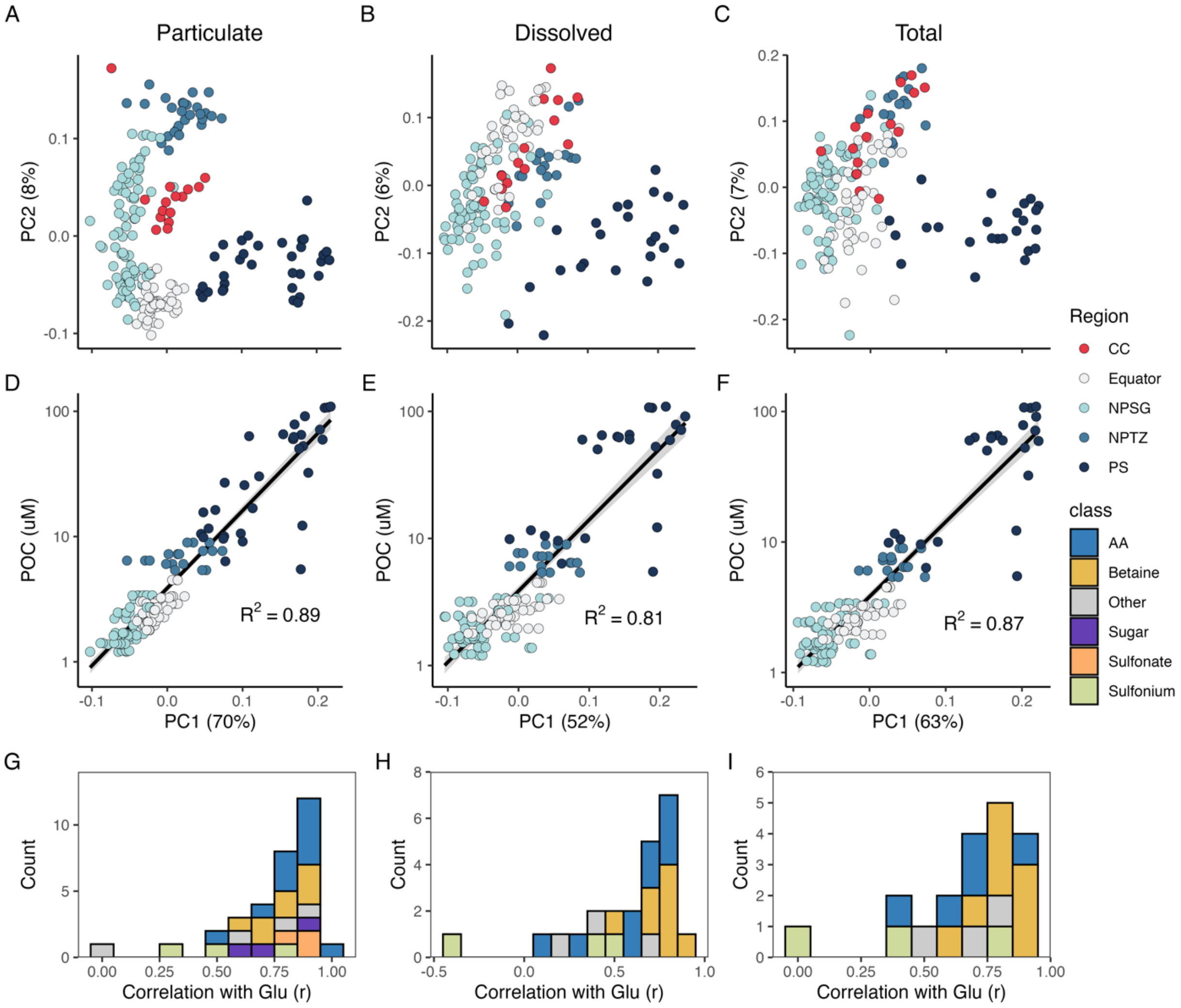

Figure 5.3: Plots of the first two principal components (PC1, PC2) from principle component analyses (PCA) of log10 normalized concentrations of particulate (A), dissolved (B), and total osmolomes (C). Scatter plots showing the relationship between PC1 and particulate organic carbon (POC) for particulate (D), dissolved (E), and total (F) osmolomes. The black line indicates the linear line of best fit and the gray shading indicates the 95% confidence interval. Histograms of the correlation between glutamic acid (Glu) and all other osmolytes. Color indicates osmolyte class (AA: amino acids).

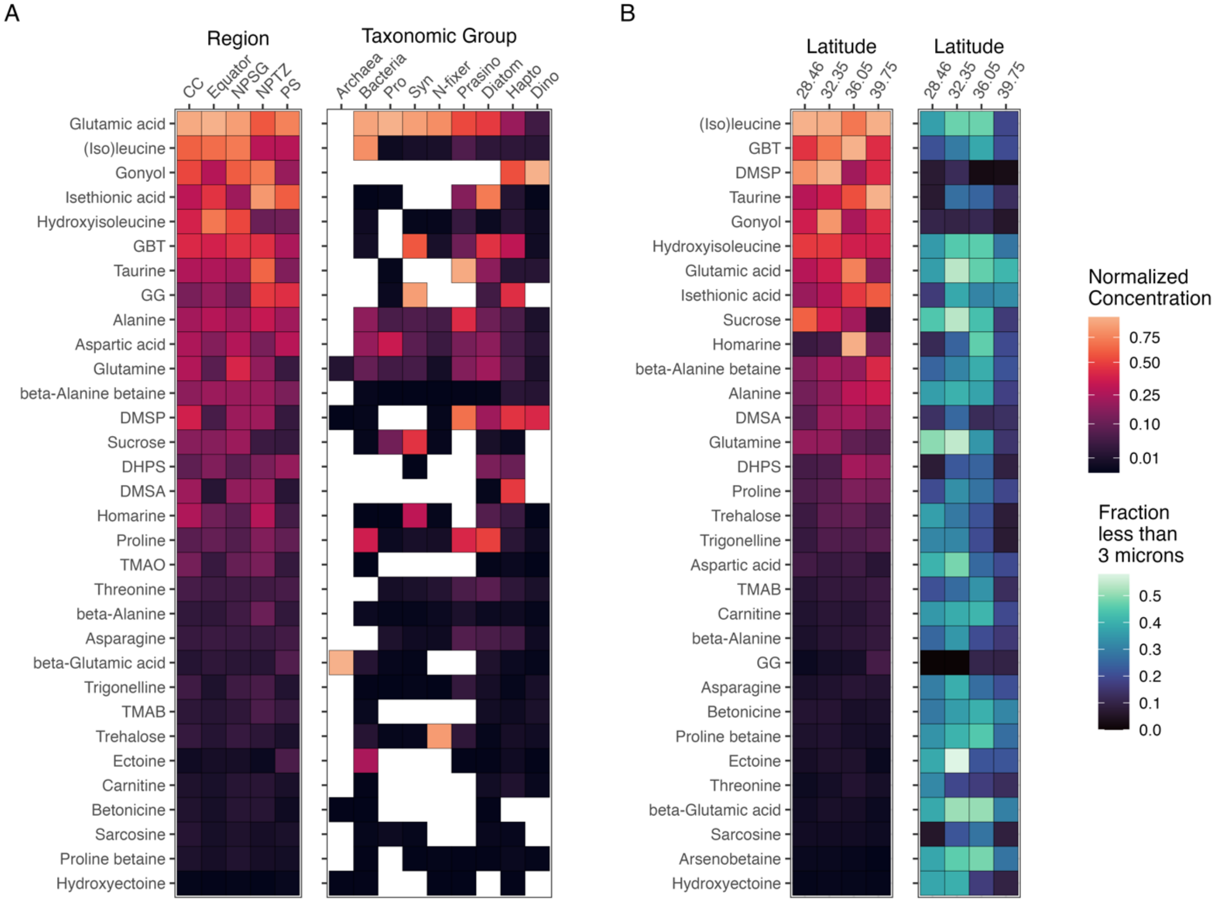

Figure 5.4: Heatmaps showing the average max normalized concentration (concentration normalized to highest concentration compound in the sample) of osmolytes in environmental samples from 5 distinct regions (left) and cultures of distinct taxonomic groups (right) (A). White squares indicate where a compound was not detected in a particular taxonomic group. Note square root scaling of color bar. Heatmaps showing the relative concentration of osmolytes in 4 size fractionated osmolome samples collected along a latitudinal transect from the NPSG to NPTZ along 158° W (left) and the relative fraction of each osmolyte in the small size fraction (less than 3 µm). Compounds in both sample sets are ordered from highest to lowest max normalized concentration.

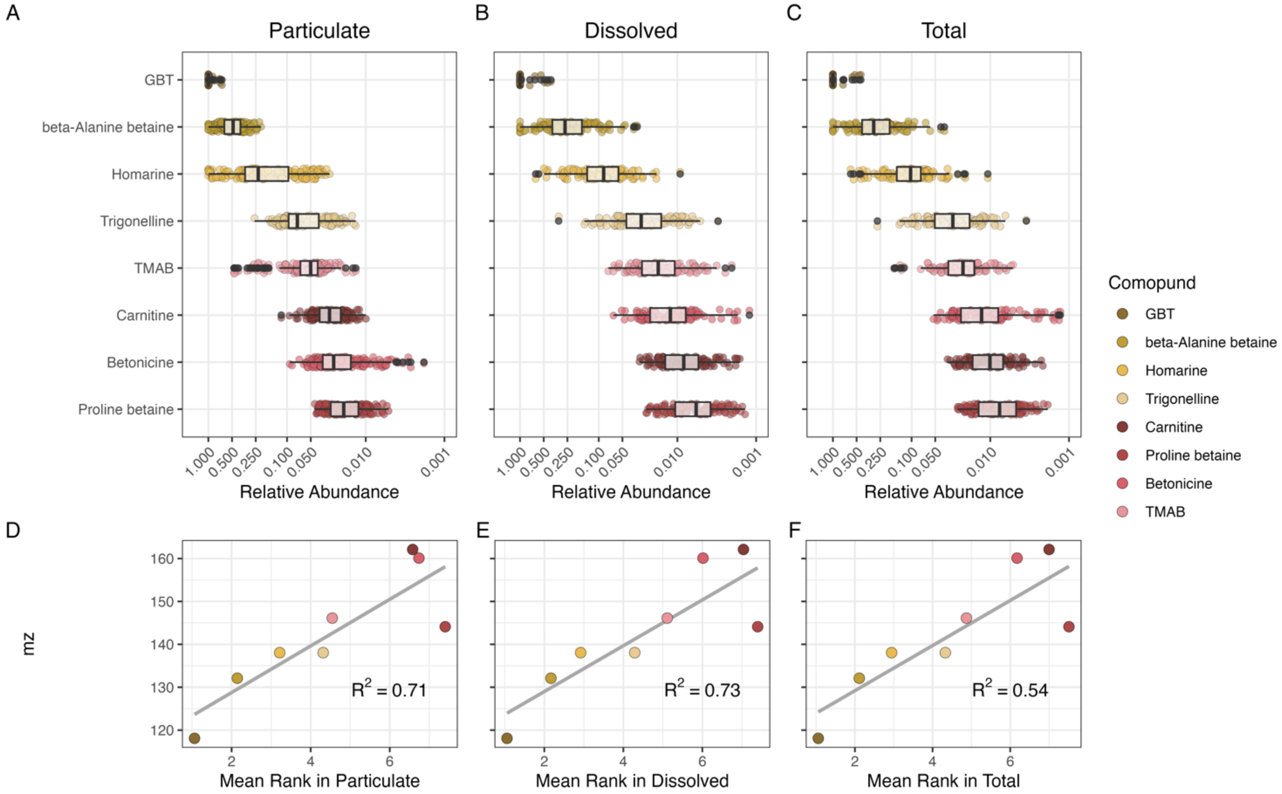

Figure 5.5: Relative abundance of betaine osmolytes normalized to the most betaine in each sample for particulate (A), dissolved (B), and total (C) osmolomes. Points indicate individual samples and boxplots show the median and quartiles for each compound. Relative abundance is log10 scaled and compounds are ordered from overall most (top) to least abundant (bottom). Relationship between mass to charge ratio (m/z) and mean betaine rank in particulate (D), dissolved (E), and total (C) osmolomes. Points indicate average values for each compound and the line indicates the line of best fit between these two parameters. TMAB is trimethylammonium butyrate or gamma-butyrobetaine.

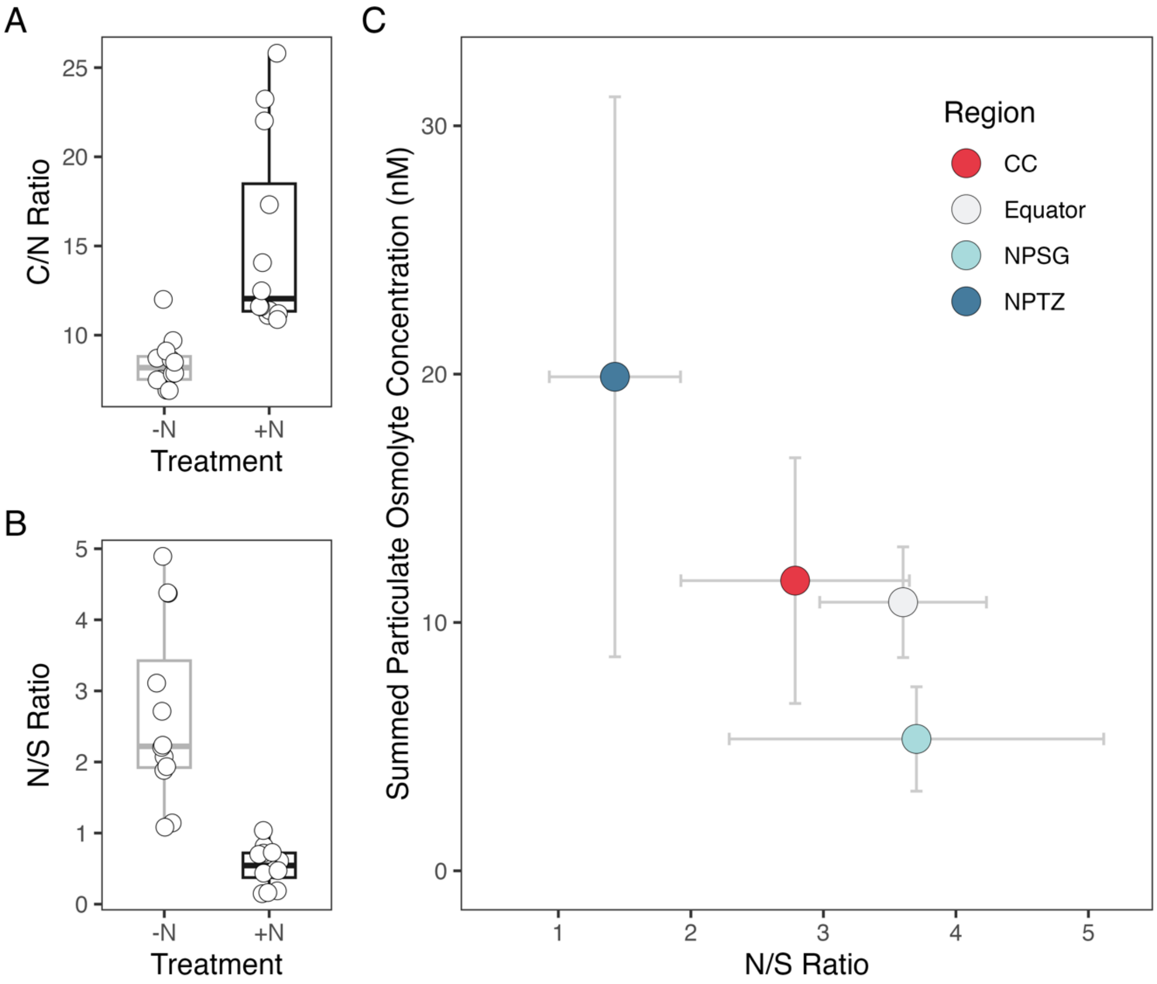

Figure 5.6: Osmolome stoichiometric differences for C/N ratio (A) and N/S ratio (B) for microbial communities from the NPSG after one-month mesocosm incubations under nitrogen amendment. Individual biological replicates are shown as points and boxplots show medians and quartiles. Relationship between summed particulate osmolyte concentration and particulate osmolome N/S ratio for four open ocean regions. Points indicate regional averages and error bars represent one standard deviation. Color indicates region.

5.7 Supplemental Figures:

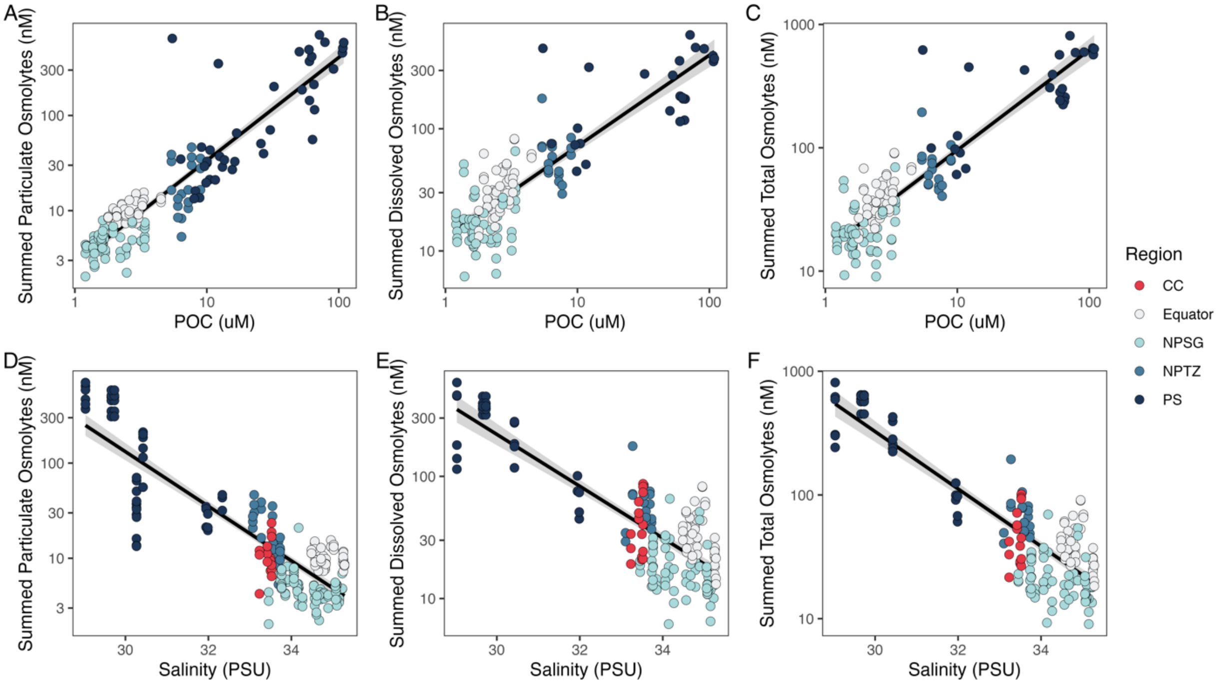

Figure 5.7: Scatter plots showing the relationship between particulate organic carbon (POC) and summed particulate (A), dissolved (B), and total (C) osmolyte concentrations. Scatter plots showing the relationship between salinity and summed particulate (D), dissolved (E), and total (F) osmolyte concentrations. Point color indicates sample region and black lines indicate the linear best fit line. Gray shading around the best fit line indicates the 95% confidence interval.

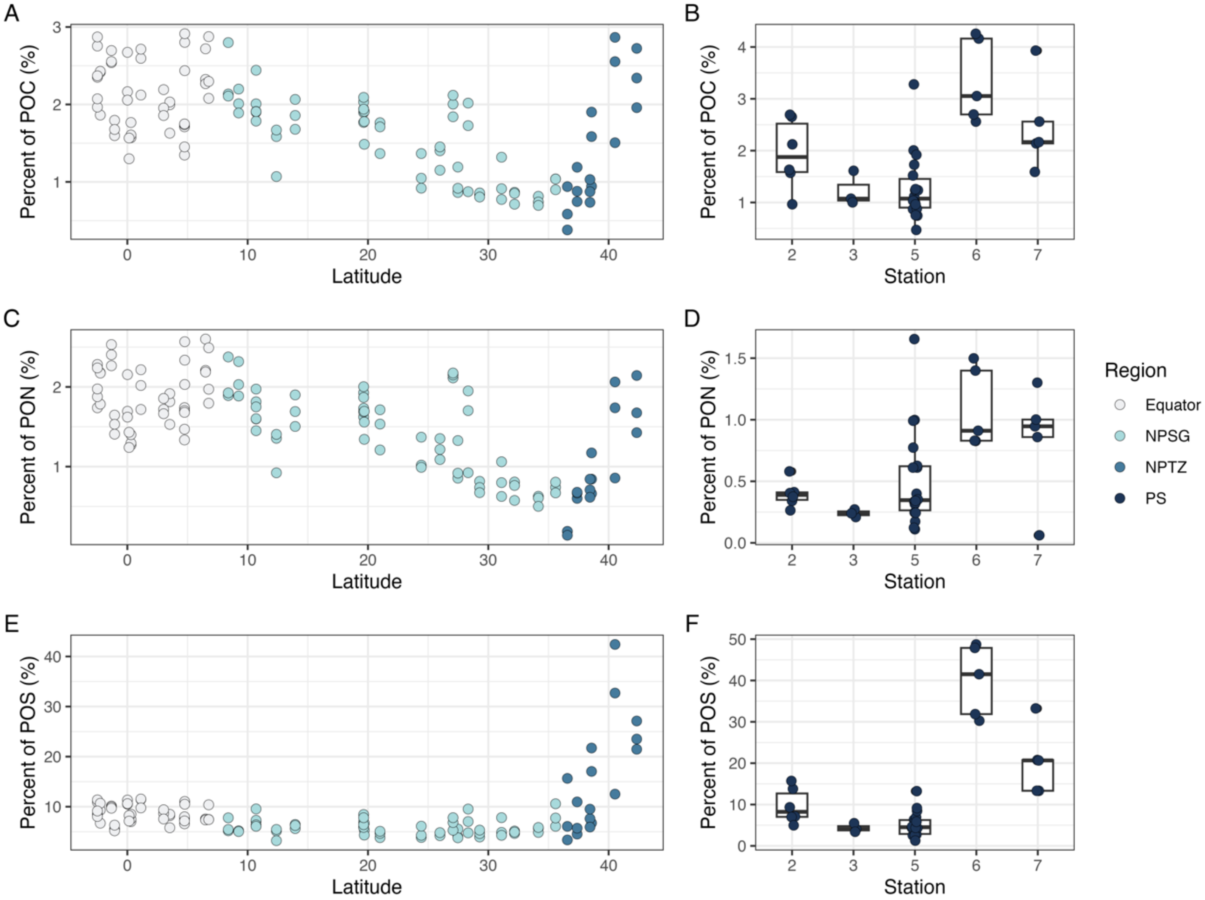

Figure 5.8: The percentage of particulate organic carbon (A, B), particulate organic nitrogen (C, D), and estimated particulate organic sulfur (E, F) in quantified particulate osmolytes for the combined open ocean latitudinal transect (Equator, NPSG, NPTZ) (A, C, E) and PS (B, D, F). Color indicates region and points represent individual measurements. In B, D, and F, boxplot show median and quartiles for each station.

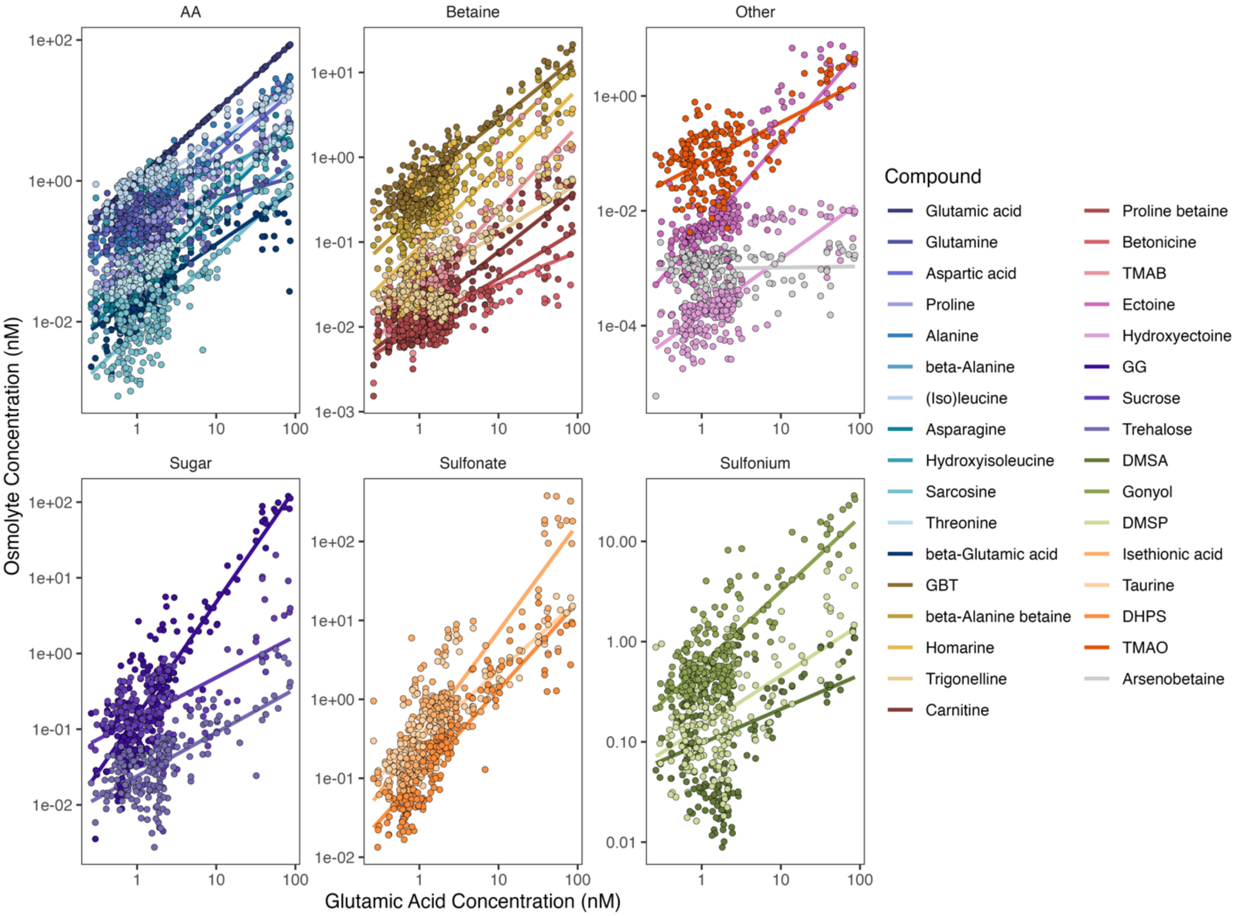

Figure 5.9: Scatter plots and linear lines of best fit showing the correlation between particulate glutamic acid concentrations and the particulate concentrations of all other measured osmolytes. Compounds indicated by color and are grouped by class (AA: amino acids). Points represent measurements in individual samples.

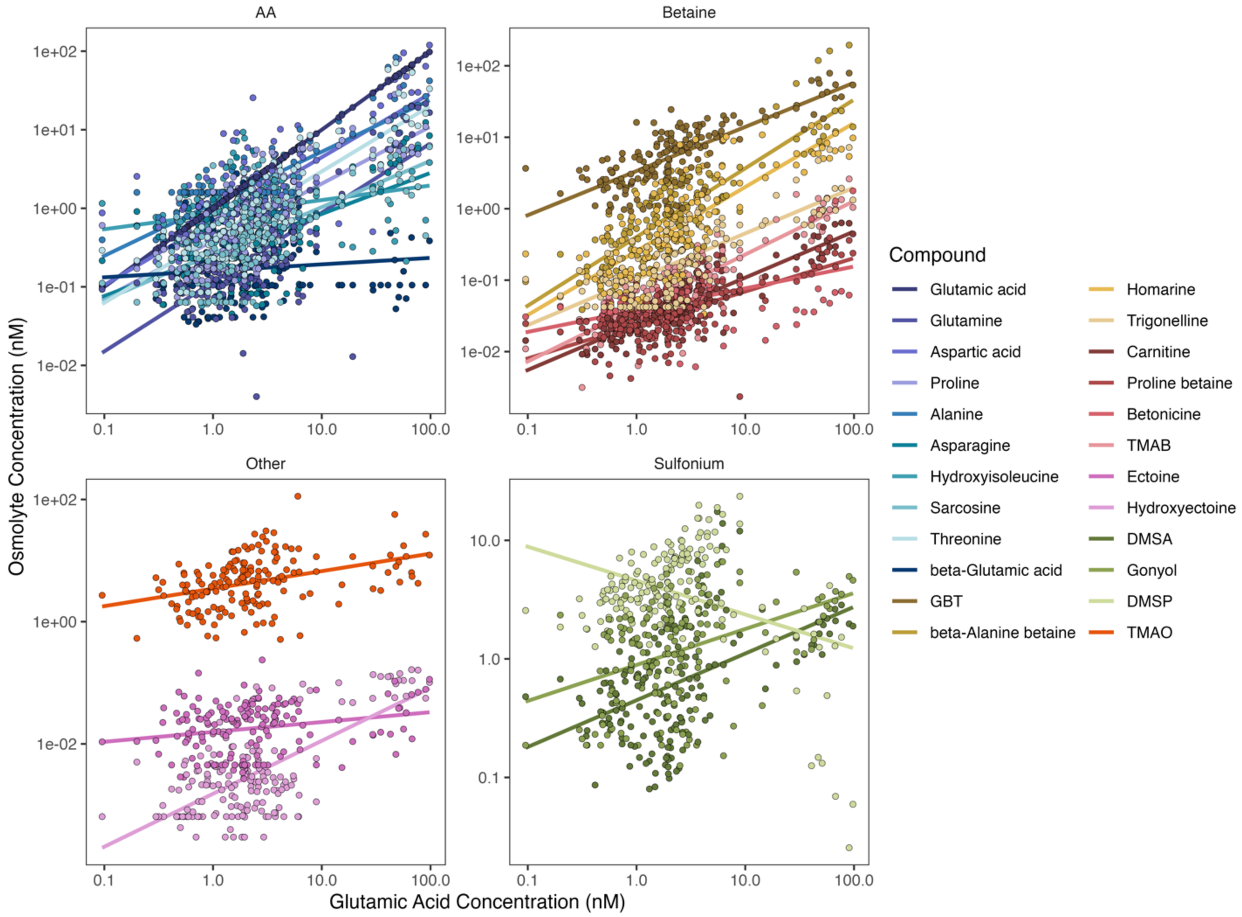

Figure 5.10: Scatter plots and linear lines of best fit showing the correlation between dissolved glutamic acid concentrations and the dissolved concentrations of all other measured osmolytes. Compounds indicated by color and are grouped by class (AA: amino acids). Points represent measurements in individual samples.

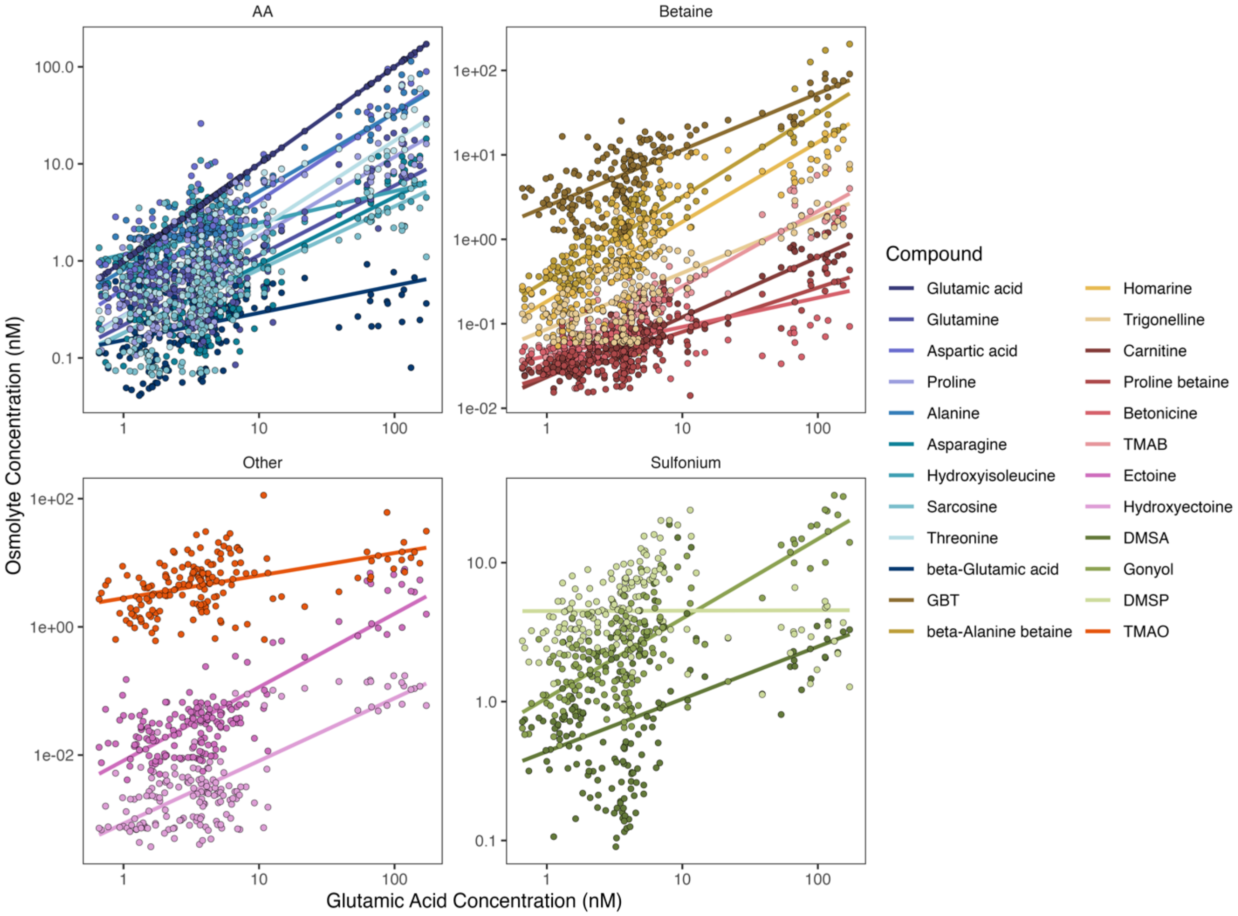

Figure 5.11: Scatter plots and linear lines of best fit showing the correlation between total glutamic acid concentrations and the total concentrations of all other measured osmolytes. Compounds indicated by color and are grouped by class (AA: amino acids). Points represent measurements in individual samples.

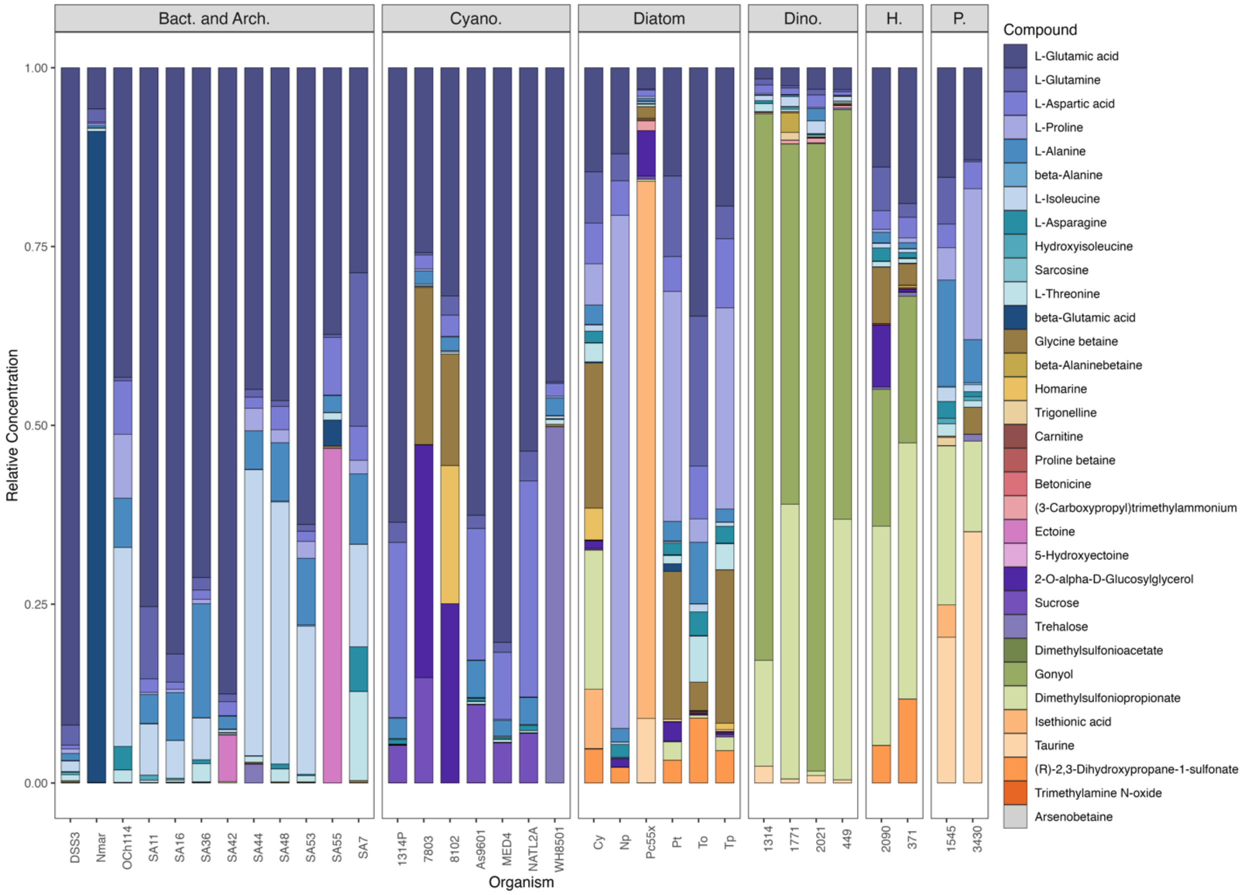

Figure 5.12: Relative concentrations of osmolytes in individual cultures of phytoplankton, bacteria, and archaea. Bars represent the mean of 2–3 individual cultures and color indicates compound. Organisms are grouped by taxonomy (Bact. And Arch.: bacteria and archaea; Cyano.: cyanobacteria; Dino.: dinoflagellates; H.: haptophytes; P.: prasinophytes).

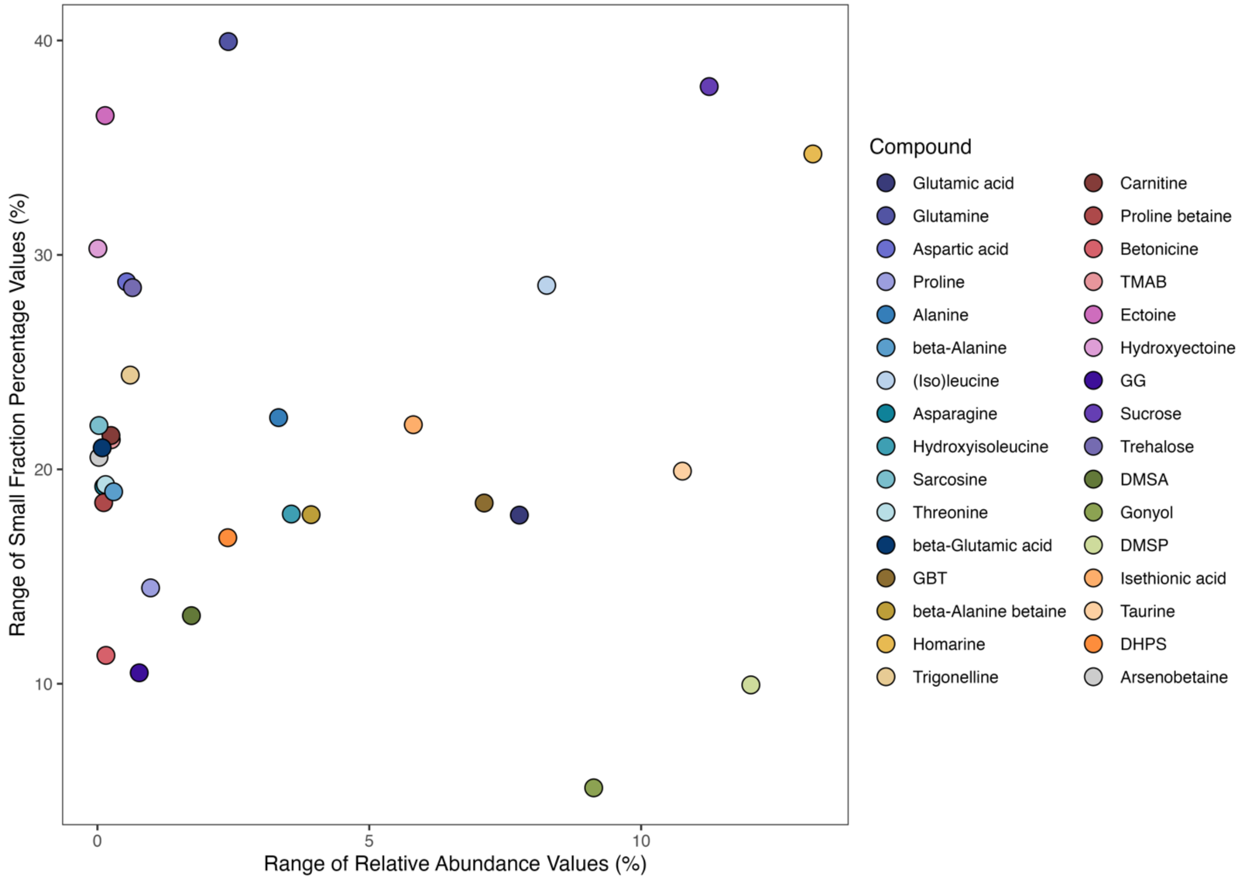

Figure 5.13: Range (maximum - minimum) of relative abundance values (concentration of the compound as a percentage of the total osmolyte concentration in the sample) values plotted against the range of small fraction percentage values for each osmolyte from size fractionated osmolome samples collected along a transect from the NPSG to NPTZ (MGL1704).

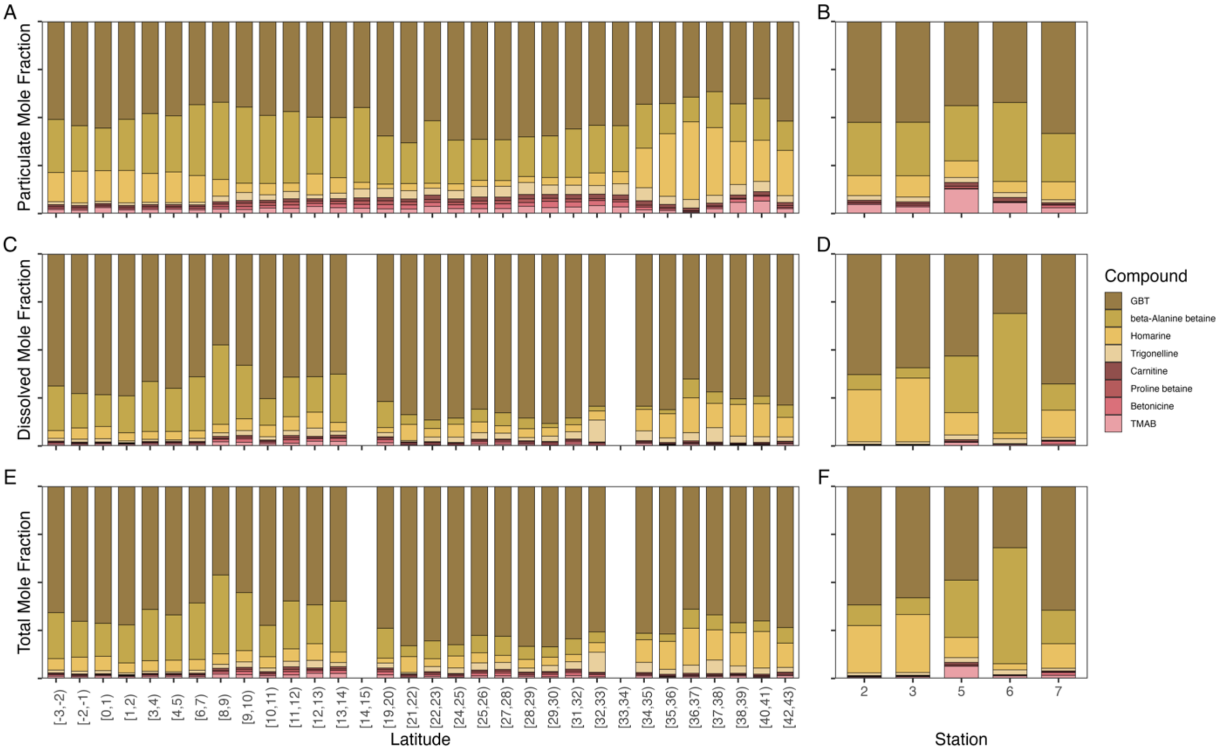

Figure 5.14: Relative concentrations of betaines in particulate (A, B), dissolved (C, D), and total (E, F) osmolomes for the combined open ocean latitudinal transects of the Equator, NPSG, and NPTZ (A, C, E) and for the estuarine samples from PS (B, D, F). For A, C, and E, osmolyte measurements are grouped into 1 degree latitude bins and the average relative concentration for each compound is shown. For B, D, and F, the average osmolyte concentration for samples collected from surface (top 10 m) are shown.

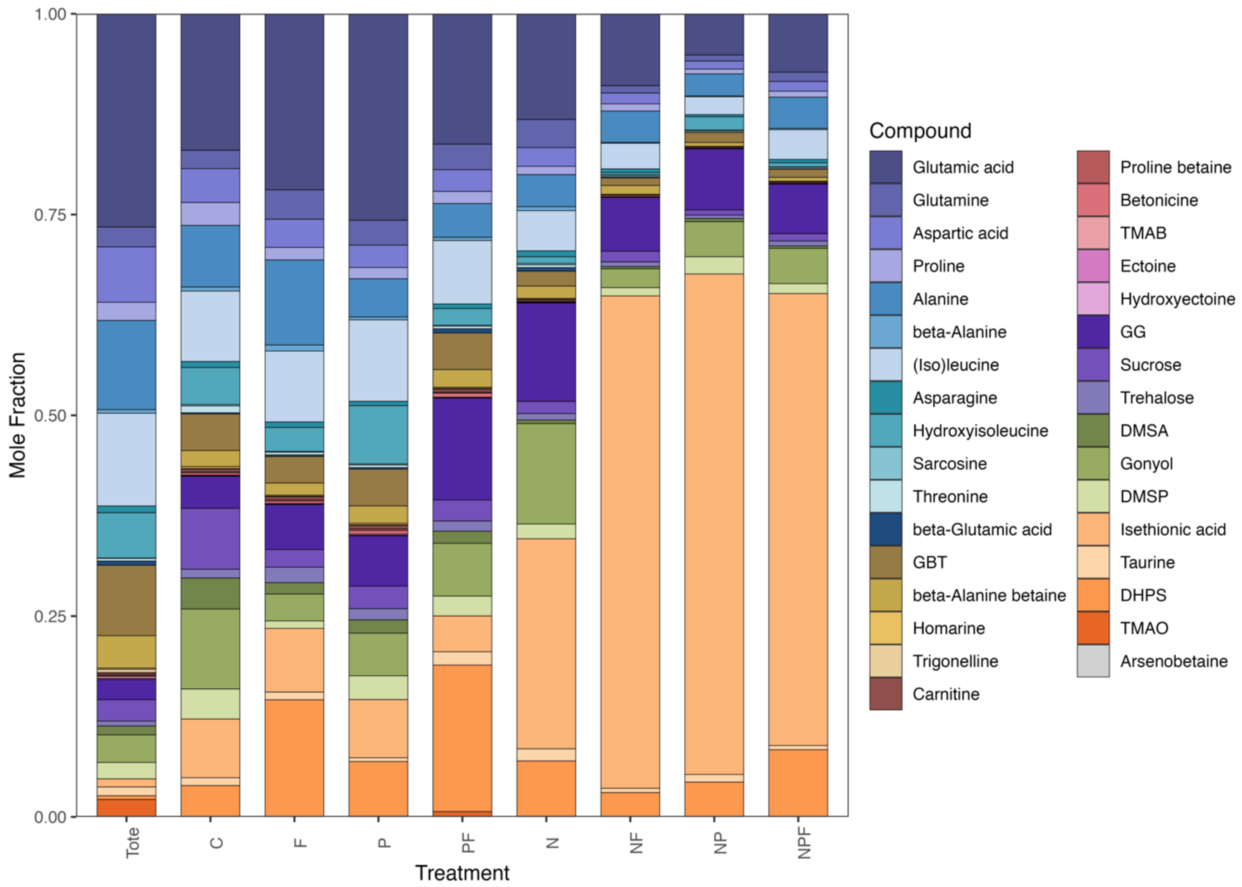

Figure 5.15: Relative concentrations of osmolytes in microbial communities from the NPSG after one-month mesocosm incubations. Color indicates compound and bars represent the mean of three biological replicates (four for “Tote” samples). Treatment indicates nutrient amendment (C: control, F: iron, P: phosphorus; N: nitrogen) and multiple letters indicates addition of multiple nutrients. “Tote” indicates samples of the natural community from the beginning of the experiment.

5.8 Supplemental Tables:

Table 5.1: Environmental and mesocosm sample details. Samples are defined by Part.SampID and Diss.SampID (particulate and dissolved sample names), Cruise, Region (NPSG: North Pacific Subtropical Gyre; NPTZ: North Pacific Transition Zone; PS: Puget Sound and the Salish Sea), Latitude (Lat), Longitude (Long), depth (in meters), station, treatment (for PERIFIX mesocosm incubations), temperature (sst, °C), salinity (sss, PSU), particulate organic carbon (poc, µmol/L), particulate nitrogen (pn, µmol/L). This table is provided as a separate file.

Table 5.2: Information about culture samples from Durham et al. 2019. Cultures samples are indicated by sample ID, organism group, organism culture ID, volume filtered (in mL), the total cells filtered, and the total cell volume filtered (in µL). This table is provided as a separate file.

Table 5.3: Results of literature search to identify osmolytes used in microbial systems. For each compound, we list the compound name, whether or not it is considered an osmolyte, our source justifying our categorization, and whether or not we are able to measure the compound with our particulate metabolomics methods. Compounds are grouped by osmolyte class (AA: amino acid). This table is provided as a separate file.

Table 5.4: Environmental and mesocosm sample osmolyte particulate and dissolved osmolyte concentrations. Samples are defined by Part.SampID and Diss.SampID (particulate and dissolved sample names), Region (NPSG: North Pacific Subtropical Gyre; NPTZ: North Pacific Transition Zone; PS: Puget Sound and the Salish Sea), compound, shortened compound name (compound.name.fig), particulate concentration in nM, dissolved concentration in nM, total concentration in nM, and quality control flags for particulate, dissolved, and total measurements ("Flag" indicates a measurement is below quality control thresholds of 3x greater than the blank for particulate or above calculated limit of detection for dissolved measurements). This table is provided as a separate file.

Table 5.5: Particulate osmolyte concentrations in culture samples. Samples are defined by sample ID, organism group, organism culture ID and compound measurements are reported in µM in sample vial and then µM in sample (normalized to volume filtered). This table is provided as a separate file.

Table 5.6: The fraction of each compound in the small (less than 3 μm) fraction at the four size fraction stations from MGL1704. This table is provided as a separate file.

Table 5.7: Internal standards and concentrations used in metabolomics data processing for normalization and quantification. This table is provided as a separate file.

Chapter 6: Conclusion:

Overview

This body of work contributes to the field of marine metabolomics through advances in methodology and the production of first of their kind datasets. These datasets enable both the testing of existing hypotheses about controls on marine metabolomes and the development of new hypotheses inspired by these observations. This body of work also contributes to the field of marine microbial ecology by enhancing our understanding of how biotic interactions and ecosystem scale emergent properties shape marine microbial ecosystems and metabolomes.

Marine metabolomics:

Marine metabolomics is a young but rapidly advancing field[19,22]. One of the core challenges at this its current stage is the lack of metabolomics measurements, particularly of compounds in the dissolved phase. The lack of dissolved measurements is due to the substantial analytical challenges associated with extracting and measuring dissolved metabolites from seawater. In the last decade, an array of new methods were developed to separate these samples from seawater[25,26,60,62]. In Chapter 2, I developed and evaluated one of these new approaches (CX-SPE) to extract polar, zwitterionic compounds from seawater. CX-SPE uniquely enables the measurement of many of the most abundant particulate and dissolved metabolites in marine microbial systems. In Chapters 3 and 5, I apply CX-SPE to measure dissolved metabolites at the basin scale in North Pacific, providing the first high resolution measurements of many of these compounds in the surface ocean, providing a set of dissolved measurements at an unprecedented scale and resolution for most compounds measured. Additionally, in Chapter 3 I combine metabolomics measurements with uptake kinetics studies to measure the fluxes of glycine betaine and homarine through the dissolved pool and demonstrate the role of uptake competition in regulating dissolved metabolite concentrations and cycling. Many microbial transporters are promiscuous beyond the transporters of betaines and sulfoniums, including amino acids, sugars, and sulfonates[183]. The impact of uptake competition on these compounds classes is currently unknown but should be targeted for future investigation as it may substantially alter the flow of carbon, nitrogen, and sulfur through microbial communities. Additionally, while the forementioned advances in methodology and the creation of the first large-scale dissolved metabolomics datasets have advanced the field substantially, large challenges remain. Measurements made in different labs using different sampling and analytical approaches are not necessarily comparable and often display large variations in measured concentrations[25,26]. The dissolved metabolomics community must now enter the phase of intercomparison and standardization in order to validate these methods and enable robust comparisons across datasets and labs. Finally, most mass features in untargeted metabolomics studies remain unannotated. A central focus moving forward must be to better annotate mass features to better understand their controls and function in marine systems.

Microbial ecology.

Microbial interactions are known to be important in marine systems where they shape community composition by altering the realized niches of marine organisms, alter biogeography, and ultimately act as an important control on the fate of organic matter in marine systems[3,7,44,47,49]. Metabolites are often key components of these interactions where they facilitate communication, act as toxins, and serve as the currencies in cross feeding relationships[19,29].

While the importance of these interactions is widely appreciated, the specifics of the mechanisms and consequences of these interactions remain largely uncharacterized. In chapter 3, I investigated the interaction of a host and virus and demonstrate dramatic changes in the metabolism of infected cells. These changes then can have downstream impacts for the substrate and vitamin pools available to the community. In chapter 5 I showed that diverse communities lead to a convergent osmolome composition, suggesting that diversity acts as a stabilizing force on marine metabolite-based networks. Additionally, I show that community shift, rather than physiological acclimation, likely drives shifts in the osmolyte pools of marine communities.

These projects identify important themes for understanding microbial interactions through metabolomics going forward, but expansions in scale are required to achieve robust generalizability for these compounds. First, virus-host systems can have highly varied impacts on host metabolism, even for different viruses infecting the same host[113,125]. Therefore, while some aspects of cyanobacterial infection documented in Chapter 2 are likely universal, more studies on a diverse set of host-phage systems are required to isolate the commonalities among these systems. Identifying these common features will allow for the potential robust incorporation of viruses into model to determine their impact on marine microbial systems and biogeochemical cycles. Finally, in Chapter 5, I show that osmolome composition is largely conserved at the basin scale in the North Pacific, likely due to functional redundancy in osmolyte producers across marine organisms. The next challenge in determining the levels of conservation in osmolome composition is exploring new oceanic regions. Samples of the Atlantic, Indian, and Southern Oceans will be required to characterize surface ocean osmolomes at a global scale and samples from deeper depths, polar latitudes, more extreme environments (ex. Oxygen Minimum Zones, Hydrothermal Vents), and freshwater environments to characterize the full range of marine and aquatic microbial osmolomes.

Metabolomes and Microbial Communities in a Changing Ocean

A mechanistic understanding of how marine communities and biogeochemical cycles function is critical for predicting how they ocean will change in the future. Metabolomics is a powerful tool for exploring both of these features of the ocean, providing insights into microbial phenotype and composition as well as the cycling and fate of organic molecules at the compound-level. The work presented in this dissertation shows that metabolomes are an emergent property of marine systems due to the role of a wide variety of interactions in structuring their composition and cycling. One component of this is that disparate communities produce similar osmolomes through functional redundancy. Functional redundancy is a characteristic associated with resiliency and therefore suggests that some marine community metabolic networks may be maintained in the face of warming or stratification driven shifts in microbial communities.

Bibliography:

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**, 1034–1039 (2008).
2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**, 237–240 (1998).
3. Moran, M. A. *et al.* Deciphering ocean carbon in a changing world. *PNAS* **113**, 3143–3151 (2016).
4. Siegel, D. A., DeVries, T., Cetinić, I. & Bisson, K. M. Quantifying the Ocean's Biological Pump and Its Carbon Cycle Impacts on Global Scales. *Annual Review of Marine Science* **15**, 329–356 (2023).
5. Chavez, F. P., Messié, M. & Pennington, J. T. Marine Primary Production in Relation to Climate Variability and Change. *Annual Review of Marine Science* **3**, 227–260 (2011).
6. Moore, C. M. *et al.* Processes and patterns of oceanic nutrient limitation. *Nature Geosci* **6**, 701–710 (2013).
7. Dutkiewicz, S. *et al.* Multiple biotic interactions establish phytoplankton community structure across environmental gradients. *Limnology and Oceanography* **69**, 1086–1100 (2024).
8. Behrenfeld, M. J. & Boss, E. S. Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms. *Annual Review of Marine Science* **6**, 167–194 (2014).
9. Mojzeš, P. *et al.* Guanine, a high-capacity and rapid-turnover nitrogen reserve in microalgal cells. *Proceedings of the National Academy of Sciences* **117**, 32722–32730 (2020).

10. Boysen, A. K. *et al.* Particulate metabolites and transcripts reflect diel oscillations of microbial activity in the surface ocean. *mSystems* **6**, e00896-20 (2021).

11. Welsh, D. T. Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS Microbiol. Rev.* **24**, 263–290 (2000).

12. Heal, K. R., Kellogg, N. A., Carlson, L. T., Lionheart, R. M. & Ingalls, A. E. Metabolic consequences of cobalamin scarcity in the diatom <i>Thalassiosira pseudonana<i> as revealed through metabolomics. *Protist* **170**, 328–348 (2019).

13. Dawson, H. M. *et al.* Potential of temperature- and salinity-driven shifts in diatom compatible solute concentrations to impact biogeochemical cycling within sea ice. *Elementa: Science of the Anthropocene* **8**, (2020).

14. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101 (2015).

15. Seymour, J. R., Amin, S. A., Raina, J.-B. & Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* **2**, 1–12 (2017).

16. Durham, B. P. *et al.* Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat Microbiol* **4**, 1706–1715 (2019).

17. Fu, H., Uchimiya, M., Gore, J. & Moran, M. A. Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 3656–3662 (2020).

18. Kieft, B. *et al.* Phytoplankton exudates and lysates support distinct microbial consortia with specialized metabolic and ecophysiological traits. *Proceedings of the National Academy of Sciences* **118**, e2101178118 (2021).

19. Moran, M. A. *et al.* Microbial metabolites in the marine carbon cycle. *Nat Microbiol* **7**, 508–523 (2022).

20. Durham, B. P. *et al.* Cryptic carbon and sulfur cycling between surface ocean plankton. *Proceedings of the National Academy of Sciences* **112**, 453–457 (2015).

21. Moran, M. A. *et al.* The Ocean's labile DOC supply chain. *Limnology and Oceanography* **67**, 1007–1021 (2022).

22. Bundy, J. G., Davey, M. P. & Viant, M. R. Environmental metabolomics: a critical review and future perspectives. *Metabolomics* **5**, 3 (2008).

23. Kido Soule, M. C., Longnecker, K., Johnson, W. M. & Kujawinski, E. B. Environmental metabolomics: Analytical strategies. *Marine Chemistry* **177**, 374–387 (2015).

24. Boysen, A. K., Heal, K. R., Carlson, L. T. & Ingalls, A. E. Best-matched internal standard normalization in liquid chromatography–mass spectrometry metabolomics applied to environmental samples. *Anal. Chem.* **90**, 1363–1369 (2018).

25. Sacks, J. S., Heal, K. R., Boysen, A. K., Carlson, L. T. & Ingalls, A. E. Quantification of dissolved metabolites in environmental samples through cation-exchange solid-phase extraction paired with liquid chromatography–mass spectrometry. *Limnol. Oceanogr.: Methods* **20**, 683–700 (2022).

26. Widner, B., Kido Soule, M. C., Ferrer-González, F. X., Moran, M. A. & Kujawinski, E. B. Quantification of Amine- and Alcohol-Containing Metabolites in Saline Samples Using Pre-extraction Benzoyl Chloride Derivatization and Ultrahigh Performance Liquid Chromatography Tandem Mass Spectrometry (UHPLC MS/MS). *Anal. Chem.* **93**, 4809–4817 (2021).

27. Kujawinski, E. B. *et al.* Metabolite diversity among representatives of divergent Prochlorococcus ecotypes. *mSystems* **8**, e01261-22 (2023).
28. Szul, M. J., Dearth, S. P., Campagna, S. R. & Zinser, E. R. Carbon fate and flux in prochlorococcus under nitrogen limitation. *mSystems* **4**, 10.1128/msystems.00254-18 (2019).
29. Durham, B. P. *et al.* An ecological framework for microbial metabolites in the ocean ecosystem. *L&O Letters* (in revision).
30. Yancey, P., Clark, M. E., Hand, S., Bowlus, R. D. & Somero, G. Living with water stress: evolution of osmolyte systems. *Science* **217 4566**, 1214–22 (1982).
31. Johnson, W. M. *et al.* Metabolite composition of sinking particles differs from surface suspended particles across a latitudinal transect in the South Atlantic. *Limnology and Oceanography* **65**, 111–127 (2020).
32. Heal, K. R. *et al.* Marine community metabolomes carry fingerprints of phytoplankton community composition. *mSystems* **6**, e01334-20 (2021).
33. McParland, E. L., Alexander, H. & Johnson, W. M. The osmolyte ties that bind: genomic insights into synthesis and breakdown of organic osmolytes in marine microbes. *Front. Mar. Sci.* **8**, (2021).
34. Johnson, W. M., Kido Soule, M. C. & Kujawinski, E. B. Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. *ISME J* **10**, 2304–2316 (2016).
35. Raina, J.-B. *et al.* Chemotaxis shapes the microscale organization of the ocean's microbiome. *Nature* **605**, 132–138 (2022).

36. Van Mooy, B. A. S. *et al.* Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458**, 69–72 (2009).

37. Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T. & Devol, A. H. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences* **103**, 8607–8612 (2006).

38. Durham, B. P. *et al.* Chemotaxonomic patterns in intracellular metabolites of marine microbial plankton. *Frontiers in Marine Science* **9**, (2022).

39. Ponge, J.-F. Emergent properties from organisms to ecosystems: towards a realistic approach. *Biological Reviews* **80**, 403–411 (2005).

40. Staudinger, M. D. *et al.* How Does Climate Change Affect Emergent Properties of Aquatic Ecosystems? *Fisheries* **46**, 423–441 (2021).

41. van den Berg, N. I. *et al.* Ecological modelling approaches for predicting emergent properties in microbial communities. *Nat Ecol Evol* **6**, 855–865 (2022).

42. Novikoff, A. B. The Concept of Integrative Levels and Biology. *Science* **101**, 209–215 (1945).

43. Wintermute, E. H. & Silver, P. A. Emergent cooperation in microbial metabolism. *Molecular Systems Biology* **6**, 407 (2010).

44. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews* **76**, 667–684 (2012).

45. Durham, B. P. *et al.* Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. *Environmental Microbiology* **19**, 3500–3513 (2017).

46. Weinbauer, M. G. *et al.* Synergistic and antagonistic effects of viral lysis and protistan grazing on bacterial biomass, production and diversity. *Environmental Microbiology* **9**, 777–788 (2007).

47. Carlson, M. C. G. *et al.* Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nat Microbiol* **7**, 570–580 (2022).

48. Beckett, S. J. *et al.* Disentangling top-down drivers of mortality underlying diel population dynamics of Prochlorococcus in the North Pacific Subtropical Gyre. *Nat Commun* **15**, 2105 (2024).

49. Follett, C. L. *et al.* Trophic interactions with heterotrophic bacteria limit the range of Prochlorococcus. *Proceedings of the National Academy of Sciences* **119**, e2110993118 (2022).

50. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801–812 (2007).

51. Chou, L. *et al.* Planetary Mass Spectrometry for Agnostic Life Detection in the Solar System. *Frontiers in Astronomy and Space Sciences* **8**, (2021).

52. Seyler, L. *et al.* Metabolomics as an Emerging Tool in the Search for Astrobiologically Relevant Biomarkers. *Astrobiology* **20**, 1251–1261 (2020).

53. Marshall, S. M. *et al.* Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nat Commun* **12**, 3033 (2021).

54. Glavin, D. P., Bada, J. L., Brinton, K. L. F. & McDonald, G. D. Amino acids in the Martian meteorite Nakhla. *Proceedings of the National Academy of Sciences* **96**, 8835–8838 (1999).

55. Naraoka, H. *et al.* Soluble organic molecules in samples of the carbonaceous asteroid (162173) Ryugu. *Science* **379**, eabn9033 (2023).

56. National Academies of Sciences, E., and Medicine. *An Astrobiology Strategy for the Search for Life in the Universe*. (The National Academies Press, Washington, DC, 2019). doi:10.17226/25252.

57. Shibl, A. A. *et al.* Diatom modulation of select bacteria through use of two unique secondary metabolites. *PNAS* **117**, 27445–27455 (2020).

58. Emerson, S. & Hedges, J. *Chemical Oceanography and the Marine Carbon Cycle*. (Cambridge University Press, Cambridge, 2008). doi:10.1017/CBO9780511793202.

59. Johnson, W. M., Soule, M. C. K. & Kujawinski, E. B. Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. *Limnology and Oceanography: Methods* **15**, 417–428 (2017).

60. Sogin, E. M., Puskás, E., Dubilier, N. & Liebeke, M. Marine Metabolomics: a Method for Nontargeted Measurement of Metabolites in Seawater by Gas Chromatography–Mass Spectrometry. *mSystems* **4**, e00638-19 (2019).

61. Xu, C. *et al.* MetFish: a Metabolomics Pipeline for Studying Microbial Communities in Chemically Extreme Environments. *mSystems* **6**, e01058-20 (2021).

62. Pontrelli, S. & Sauer, U. Salt-Tolerant Metabolomics for Exometabolomic Measurements of Marine Bacterial Isolates. *Anal. Chem.* **93**, 7164–7171 (2021).

63. Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* **6**, 230–235 (2008).

64. Fiore, C. L., Longnecker, K., Kido Soule, M. C. & Kujawinski, E. B. Release of ecologically relevant metabolites by the cyanobacterium Synechococcus elongatus CCMP 1631. *Environmental Microbiology* **17**, 3949–3963 (2015).

65. Petras, D. *et al.* High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science* **4**, (2017).

66. Weber, L. *et al.* Extracellular Reef Metabolites Across the Protected Jardines de la Reina, Cuba Reef System. *Frontiers in Marine Science* **7**, (2020).

67. Gebser, B. & Pohnert, G. Synchronized Regulation of Different Zwitterionic Metabolites in the Osmoadaption of Phytoplankton. *Marine Drugs* **11**, 2168–2182 (2013).

68. Dawson, H. M. *et al.* Large Diversity in Nitrogen- and Sulfur-Containing Compatible Solute Profiles in Polar and Temperate Diatoms. *Integrative and Comparative Biology* **60**, 1401–1413 (2020).

69. Yoch, D. C. Dimethylsulfoniopropionate: Its Sources, Role in the Marine Food Web, and Biological Degradation to Dimethylsulfide. *Applied and Environmental Microbiology* **68**, 5804–5815 (2002).

70. Boysen, A. K. *et al.* Glycine betaine uptake and metabolism in marine microbial communities. *Environmental Microbiology* **24**, 2380–2403 (2022).

71. Glabonjat, R. A., Raber, G., Van Mooy, B. A. S. & Francesconi, K. A. Arsenobetaine in Seawater: Depth Profiles from Selected Sites in the North Atlantic. *Environ. Sci. Technol.* **52**, 522–530 (2018).

72. Fontanals, N. *et al.* Evaluation of strong cation-exchange polymers for the determination of drugs by solid-phase extraction–liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A* **1343**, 55–62 (2014).

73. Logan, B. K., Stafford, D. T., Tebbett, I. R. & Moore, C. M. Rapid screening for 100 basic drugs and metabolites in urine using cation exchange solid-phase extraction and high-

performance liquid chromatography with diode array detection. *J Anal Toxicol* **14**, 154–159 (1990).

74. Halewood, E. *et al.* GO-SHIP Repeat Hydrography: Determination of dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) in seawater using High Temperature Combustion Analysis. (2022) doi:10.25607/OBP-1745.
75. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918–920 (2012).
76. Adams, K. J. *et al.* Skyline for small molecules: a unifying software package for quantitative metabolomics. *J. Proteome Res.* **19**, 1447–1458 (2020).
77. Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, 523–526 (2015).
78. Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C. & Webb, E. A. The role of B vitamins in marine biogeochemistry. *Annu. Rev. Mar. Sci.* **6**, 339–367 (2014).
79. Hmelo, L. R. Quorum Sensing in Marine Microbial Environments. *Annual Review of Marine Science* **9**, 257–281 (2017).
80. Gibb, S. W. & Hatton, A. D. The occurrence and distribution of trimethylamine-N-oxide in Antarctic coastal waters. *Marine Chemistry* **91**, 65–75 (2004).
81. Kiene, R. P. & Slezak, D. Low dissolved DMSP concentrations in seawater revealed by small-volume gravity filtration and dialysis sampling. *Limnology and Oceanography: Methods* **4**, 80–95 (2006).
82. Poretsky, R. S., Sun, S., Mou, X. & Moran, M. A. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology* **12**, 616–627 (2010).

83. Sabadel, A. J. M. *et al.* Determination of picomolar dissolved free amino acids along a South Atlantic transect using reversed-phase high-performance liquid chromatography. *Marine Chemistry* **196**, 173–180 (2017).

84. Vorobev, A. *et al.* Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals. *Environmental Microbiology* **20**, 3012–3030 (2018).

85. Mayali, X. & Weber, P. K. Quantitative isotope incorporation reveals substrate partitioning in a coastal microbial community. *FEMS Microbiology Ecology* **94**, fiy047 (2018).

86. Ferrer-González, F. X. *et al.* Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. *ISME J* **15**, 762–773 (2021).

87. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).

88. Currie, L. A. Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal. Chem.* **40**, 586–593 (1968).

89. Djoumbou Feunang, Y. *et al.* ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* **8**, 61 (2016).

90. Lee, C. & Bada, J. L. Amino acids in equatorial Pacific Ocean water. *Earth and Planetary Science Letters* **26**, 61–68 (1975).

91. Fuhrman, J. & Ferguson, R. Nanomolar concentrations and rapid turnover of dissolved free amino acids in seawater: agreement between chemical and microbiological measurements. (1986) doi:10.3354/MEPS033237.

92. Wawrik, B. *et al.* Bacterial utilization of creatine in seawater. *Aquatic Microbial Ecology* **80**, (2017).

93. Mausz, M. A. *et al.* Microbial uptake dynamics of choline and glycine betaine in coastal seawater. *Limnology and Oceanography* **n/a**, (2022).

94. Connell, P., Ribalet, F., Armbrust, E., White, A. & Caron, D. Diel oscillations in the feeding activity of heterotrophic and mixotrophic nanoplankton in the North Pacific Subtropical Gyre. *Aquat. Microb. Ecol.* **85**, 167–181 (2020).

95. Bandara, K., Varpe, Ø., Wijewardene, L., Tverberg, V. & Eiane, K. Two hundred years of zooplankton vertical migration research. *Biological Reviews* **96**, 1547–1589 (2021).

96. Mruwat, N. *et al.* A single-cell polony method reveals low levels of infected <i>Prochlorococcus<i> in oligotrophic waters despite high cyanophage abundances. *ISME J* **15**, 41–54 (2021).

97. Yancey, P. H., Gerringer, M. E., Drazen, J. C., Rowden, A. A. & Jamieson, A. Marine fish may be biochemically constrained from inhabiting the deepest ocean depths. *PNAS* **111**, 4461–4465 (2014).

98. Lidbury, I., Murrell, J. C. & Chen, Y. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *PNAS* **111**, 2710–2715 (2014).

99. Lidbury, I. D., Murrell, J. C. & Chen, Y. Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: implications for marine carbon and nitrogen cycling. *ISME J* **9**, 760–769 (2015).

100. Jones, H. J. *et al.* A new family of uncultivated bacteria involved in methanogenesis from the ubiquitous osmolyte glycine betaine in coastal saltmarsh sediments. *Microbiome* **7**, 120 (2019).

101. Ngugi, D. K., Ziegler, M., Duarte, C. M. & Voolstra, C. R. Genomic Blueprint of Glycine Betaine Metabolism in Coral Metaorganisms and Their Contribution to Reef Nitrogen Budgets. *iScience* **23**, (2020).

102. Mausz, M. A. *et al.* Microbial uptake dynamics of choline and glycine betaine in coastal seawater. *Limnology and Oceanography* **67**, 1052–1064 (2022).

103. Flombaum, P. *et al.* Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9824–9829 (2013).

104. Biller, S. J., Berube, P. M., Lindell, D. & Chisholm, S. W. Prochlorococcus: the structure and function of collective diversity. *Nat Rev Microbiol* **13**, 13–27 (2015).

105. Braakman, R. *et al.* Global niche partitioning of purine and pyrimidine cross-feeding among ocean microbes. *Science Advances* **11**, eadp1949 (2025).

106. Xiao, X. *et al.* Viral lysis alters the optical properties and biological availability of dissolved organic matter derived from prochlorococcus picocyanobacteria. *Appl Environ Microbiol* **87**, e02271-20 (2021).

107. Forterre, P. The virocell concept and environmental microbiology. *ISME J* **7**, 233–236 (2013).

108. Jacobson, T. B., Callaghan, M. M. & Amador-Noguez, D. Hostile takeover: how viruses reprogram prokaryotic metabolism. *Annu. Rev. Microbiol.* **75**, 515–539 (2021).

109. Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E757–E764 (2011).

110. Warwick-Dugdale, J., Buchholz, H. H., Allen, M. J. & Temperton, B. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virol. J.* **16**, 15 (2019).

111. Breitbart, M., Bonnain, C., Malki, K. & Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nat Microbiol* **3**, 754–766 (2018).

112. Shopen Gochev, C. *et al.* Cold surface waters of the sub-antarctic pacific ocean support high cyanophage abundances and infection levels. *Environ. Microbiol.* **27**, e70031 (2025).

113. De Smet, J. *et al.* High coverage metabolomics analysis reveals phage-specific alterations to Pseudomonas aeruginosa physiology during infection. *ISME J* **10**, 1823–1835 (2016).

114. Ankrah, N. Y. D. *et al.* Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* **8**, 1089–1100 (2014).

115. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).

116. Sawa, N. *et al.* Modification of carbon metabolism in *Synechococcus elongatus* PCC 7942 by cyanophage-derived sigma factors for bioproduction improvement. *J. Biosci. Bioeng.* **127**, 256–264 (2019).

117. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83–86 (2007).

118. Howard-Varona, C. *et al.* Protist impacts on marine cyanovirocell metabolism. *ISME COMMUN.* **2**, 1–14 (2022).

119. Zimmerman, A. E. *et al.* Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* **18**, 21–34 (2020).

120. Karl, D. M. & Holm-Hansen, O. Methodology and measurement of adenylate energy charge ratios in environmental samples. *Mar. Biol.* **48**, 185–197 (1978).

121. Ribalet, F. *et al.* SeaFlow data v1, high-resolution abundance, size and biomass of small phytoplankton in the North Pacific. *Sci Data* **6**, 277 (2019).

122. Mattern, J. P. *et al.* A Bayesian approach to modeling phytoplankton population dynamics from size distribution time series. *PLOS Computational Biology* **18**, e1009733 (2022).

123. Dührkop, K. *et al.* SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods* **16**, 299–302 (2019).

124. Karl, D. M. *et al.* Temporal dynamics of total microbial biomass and particulate detritus at Station ALOHA. *Progress in Oceanography* **205**, 102803 (2022).

125. Howard-Varona, C. *et al.* Phage-specific metabolic reprogramming of virocells. *ISME J* **14**, 881–895 (2020).

126. Howard-Varona, C. *et al.* Environment-specific virocell metabolic reprogramming. *The ISME Journal* **18**, wrae055 (2024).

127. Chevallereau, A. *et al.* Next-generation "-omics" approaches reveal a massive alteration of host rna metabolism during bacteriophage infection of <i>Pseudomonas aeruginosa<i>. *PLOS Genetics* **12**, e1006134 (2016).

128. Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Viruses inhibit co2 fixation in the most abundant phototrophs on earth. *Curr. Biol.* **26**, 1585–1589 (2016).

129. Fang, X. *et al.* Transcriptomic responses of the marine cyanobacterium <i>Prochlorococcus<i> to viral lysis products. *Environ. Microbiol.* **21**, 2015–2028 (2019).

130. Ofaim, S., Sulheim, S., Almaas, E., Sher, D. & Segrè, D. Dynamic allocation of carbon storage and nutrient-dependent exudation in a revised genome-scale model of <i>Prochlorococcus<i>. *Front. Genet.* **12**, (2021).

131. Luan, G., Zhang, S., Wang, M. & Lu, X. Progress and perspective on cyanobacterial glycogen metabolism engineering. *Biotechnol. Adv.* **37**, 771–786 (2019).

132. Kirsch, F., Klähn, S. & Hagemann, M. Salt-regulated accumulation of the compatible solutes sucrose and glucosylglycerol in cyanobacteria and its biotechnological potential. *Front. Microbiol.* **10**, (2019).

133. Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C. & Webb, E. A. The role of B vitamins in marine biogeochemistry. *Annu. Rev. Mar. Sci.* **6**, 339–367 (2014).

134. Lu, S. C. *S*-Adenosylmethionine. *Int. J. Biochem. Cell Biol.* **32**, 391–395 (2000).

135. Narainsamy, K. *et al.* Oxidative-stress detoxification and signalling in cyanobacteria: the crucial glutathione synthesis pathway supports the production of ergothioneine and ophthalmate. *Mol. Microbiol.* **100**, 15–24 (2016).

136. Bannon, C. C., Mudge, E. M. & Bertrand, E. M. Shedding light on cobalamin photodegradation in the ocean. *Limnol. Oceanogr. Lett.* **9**, 135–144 (2024).

137. Nazhat, N. B., Golding, B. T., Johnson, G. R. A. & Jones, P. Destruction of vitamin B12 by reaction with ascorbate: The role of hydrogen peroxide and the oxidation state of cobalt. *J. Inorg. Biochem.* **36**, 75–81 (1989).

138. Heal, K. R. *et al.* Accumulation of NO2-cobalamin in nutrient-stressed ammonia-oxidizing archaea and in the oxygen deficient zone of the eastern tropical North Pacific. *Environ. Microbiol. Rep.* **10**, 453–457 (2018).

139. Qin, W. *et al.* Stress response of a marine ammonia-oxidizing archaeon informs physiological status of environmental populations. *ISME J* **12**, 508–519 (2018).

140. Sheyn, U., Rosenwasser, S., Ben-Dor, S., Porat, Z. & Vardi, A. Modulation of host ROS metabolism is essential for viral infection of a bloom-forming coccolithophore in the ocean. *ISME J* **10**, 1742–1754 (2016).

141. Bottero, V., Chakraborty, S. & Chandran, B. Reactive oxygen species are induced by Kaposi's sarcoma-associated herpesvirus early during primary infection of endothelial cells to promote virus entry. *J Virol* **87**, 1733–1749 (2013).
142. Hernández, J. A. *et al.* Oxidative stress and antioxidative responses in plant–virus interactions. *Physiol. Mol. Plant Pathol.* **94**, 134–148 (2016).
143. Laurenceau, R., Raho, N., Forget, M., Arellano, A. A. & Chisholm, S. W. Frequency of mispackaging of <i>Prochlorococcus<i> DNA by cyanophage. *ISME J* **15**, 129–140 (2021).
144. Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M. & Zinser, E. R. Dependence of the cyanobacterium <i>Prochlorococcus<i> on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLOS ONE* **6**, e16805 (2011).
145. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three <i>Prochlorococcus<i> cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3**, e144 (2005).
146. Harrison, A. O., Moore, R. M., Polson, S. W. & Wommack, K. E. Reannotation of the ribonucleotide reductase in a cyanophage reveals life history strategies within the virioplankton. *Front. Microbiol.* **10**, (2019).
147. Muratore, D. *et al.* Complex marine microbial communities partition metabolism of scarce resources over the diel cycle. *Nat Ecol Evol* **6**, 218–229 (2022).
148. Barthelmeß, T. *et al.* Pronounced diel cycling of dissolved carbohydrates and amino acids in the surface ocean and across diverse regimes. *Environ. Sci. Technol.* **59**, 419–429 (2025).
149. Ottesen, E. A. *et al.* Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**, 207–212 (2014).

150. Aylward, F. O. *et al.* Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11446–11451 (2017).

151. Zinser, E. R. *et al.* Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, <i>Prochlorococcus<i>. *PLOS ONE* **4**, e5135 (2009).

152. Welkie, D. G. *et al.* A hard day's night: cyanobacteria in diel cycles. *Trends Microbiol.* **27**, 231–242 (2019).

153. Liu, R., Liu, Y., Chen, Y., Zhan, Y. & Zeng, Q. Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14077–14082 (2019).

154. Bratbak, G., Thingstad, F. & Heldal, M. Viruses and the microbial loop. *Microb. Ecol.* **28**, 209–221 (1994).

155. Ma, X., Coleman, M. L. & Waldbauer, J. R. Distinct molecular signatures in dissolved organic matter produced by viral lysis of marine cyanobacteria. *Environ. Microbiol.* **20**, 3001–3011 (2018).

156. Henshaw, R. J. *et al.* Metabolites from intact phage-infected <i>Synechococcus<i> chemotactically attract heterotrophic marine bacteria. *Nat Microbiol* **9**, 3184–3195 (2024).

157. Wienhausen, G. *et al.* The overlooked role of a biotin precursor for marine bacteria - desthiobiotin as an escape route for biotin auxotrophy. *ISME J* **16**, 2599–2609 (2022).

158. Johnson, W. M. *et al.* Auxotrophic interactions: a stabilizing attribute of aquatic microbial communities? *FEMS Microbiol Ecol* **96**, fiaa115 (2020).

159. Tang, Y. Z., Koch, F. & Gobler, C. J. Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20756–20761 (2010).

160. Feng, X. *et al.* Mechanisms driving genome reduction of a novel Roseobacter lineage. *ISME J* **15**, 3576–3586 (2021).

161. Helliwell, K. E. *et al.* Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12. *Current Biology* **26**, 999–1008 (2016).

162. Soto, M. A., Desai, D., Bannon, C., LaRoche, J. & Bertrand, E. M. Cobalamin producers and prokaryotic consumers in the Northwest Atlantic. *Environ. Microbiol.* **25**, 1300–1313 (2023).

163. Ribalet, F. *et al.* Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. *Proceedings of the National Academy of Sciences* **112**, 8008–8012 (2015).

164. Sosik, H. M., Olson, R. J., Neubert, M. G., Shalapyonok, A. & Solow, A. R. Growth Rates of Coastal Phytoplankton from Time-Series Measurements with a Submersible Flow Cytometer. *Limnology and Oceanography* **48**, 1756–1765 (2003).

165. Hunter-Cevera, K. R. *et al.* Diel size distributions reveal seasonal growth dynamics of a coastal phytoplankter. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9852–9857 (2014).

166. Hunter-Cevera, K. R. *et al.* Physiological and ecological drivers of early spring blooms of a coastal phytoplankter. *Science* **354**, 326–329 (2016).

167. Fowler, B. L. *et al.* Dynamics and functional diversity of the smallest phytoplankton on the Northeast US Shelf. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 12215–12221 (2020).

168. Hamilton, M. *et al.* Dynamics of Teleaulax -like cryptophytes during the decline of a red water bloom in the Columbia River Estuary. *Journal of Plankton Research* **39**, (2017).

169. Heal, K. R. *et al.* Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *PNAS* **114**, 364–369 (2017).

170. Heal, K. R. *et al.* Determination of four forms of vitamin B12 and other B vitamins in seawater by liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **28**, 2398–2404 (2014).

171. Ludwig, M. *et al.* Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat Mach Intell* **2**, 629–641 (2020).

172. Dührkop, K. *et al.* Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat Biotechnol* **39**, 462–471 (2021).

173. Roth-Rosenberg, D., Aharonovich, D., Omta, A. W., Follows, M. J. & Sher, D. Dynamic macromolecular composition and high exudation rates in Prochlorococcus. *Limnology and Oceanography* **66**, 1759–1773 (2021).

174. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

175. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

176. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

177. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

178. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30 (2000).

179. Keller, M. D., Matrai, P. A., Kiene, R. P. & Bellows, W. K. Responses of coastal phytoplankton populations to nitrogen additions: dynamics of cell-associated

dimethylsulfoniopropionate (DMSP), glycine betaine (GBT), and homarine. *Can. J. Fish. Aquat. Sci.* **61**, 685–699 (2004).

180. Beale, R. & Airs, R. Quantification of glycine betaine, choline and trimethylamine N-oxide in seawater particulates: Minimization of seawater associated ion suppression. *Analytica Chimica Acta* **938**, 114–122 (2016).

181. Airs, R. L. *et al.* Seasonal measurements of the nitrogenous osmolyte glycine betaine in marine temperate coastal waters. *Biogeochemistry* **162**, 309–323 (2023).

182. Dawson, H. M. *et al.* Microbial metabolomic responses to changes in temperature and salinity along the western Antarctic Peninsula. *The ISME Journal* **17**, 2035–2046 (2023).

183. Clifton, B. E., Alcolombri, U., Uechi, G.-I., Jackson, C. J. & Laurino, P. The ultra-high affinity transport proteins of ubiquitous marine bacteria. *Nature* **634**, 721–728 (2024).

184. Sperfeld, M. *et al.* Algal methylated compounds shorten the lag phase of Phaeobacter inhibens bacteria. *Nat Microbiol* **9**, 2006–2021 (2024).

185. Kokoeva, M. V., Storch, K.-F., Klein, C. & Oesterhelt, D. A novel mode of sensory transduction in archaea: binding protein-mediated chemotaxis towards osmoprotectants and amino acids. *EMBO J* **21**, 2312–2322 (2002).

186. Poulin, R. X. *et al.* Chemical encoding of risk perception and predator detection among estuarine invertebrates. *Proceedings of the National Academy of Sciences* **115**, 662–667 (2018).

187. Li, L. *et al.* Bacteria and archaea synergistically convert glycine betaine to biogenic methane in the formosa cold seep of the South China Sea. *mSystems* **6**, e0070321 (2021).

188. Kiene, R., Hoffmann Williams, L. & Walker, J. Seawater microorganisms have a high affinity glycine betaine uptake system which also recognizes dimethylsulfoniopropionate. *Aquat. Microb. Ecol.* **15**, 39–51 (1998).

189. Vila-Costa, M. *et al.* Dimethylsulfoniopropionate uptake by marine phytoplankton. *Science* **314**, 652–654 (2006).

190. Petrou, K. & Nielsen, D. A. Uptake of dimethylsulphoniopropionate (DMSP) by the diatom Thalassiosira weissflogii: a model to investigate the cellular function of DMSP. *Biogeochemistry* **141**, 265–271 (2018).

191. Torstensson, A., Young, J. N., Carlson, L. T., Ingalls, A. E. & Deming, J. W. Use of exogenous glycine betaine and its precursor choline as osmoprotectants in Antarctic sea-ice diatoms. *Journal of Phycology* **55**, 663–675 (2019).

192. Kiene, R. P. & Gerard, G. Evaluation of glycine betaine as an inhibitor of dissolved dimethylsulfoniopropionate degradation in coastal waters. *Marine Ecology Progress Series* **128**, 121–131 (1995).

193. Noell, S. E. & Giovannoni, S. J. SAR11 bacteria have a high affinity and multifunctional glycine betaine transporter. *Environmental Microbiology* **21**, 2559–2575 (2019).

194. Canelas, A. B. *et al.* Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal. Chem.* **81**, 7379–7389 (2009).

195. Wright, R. R. & Hobbie, J. E. Use of glucose and acetate by bacteria and algae in aquatic ecosystems. *Ecology* **47**, 447–464 (1966).

196. Kiene, R. P. Uptake of choline and its conversion to glycine betaine by bacteria in estuarine waters. *Applied and Environmental Microbiology* **64**, 1045–1051 (1998).

197. Newton, J. & Van Voorhis, K. *Seasonal Patterns and Controlling Factors of Primary Production in Puget Sound's Central Basin and Possession Sound*. https://apps.ecology.wa.gov/publications/SummaryPages/0203059.html (2002).

198. Fernandez, E., Ostrowski, M., Siboni, N., Seymour, J. R. & Petrou, K. Uptake of dimethylsulfoniopropionate (dmsp) by natural microbial communities of the Great Barrier Reef (GBR), Australia. *Microorganisms* **9**, 1891 (2021).

199. Meyer, N., Rydzyk, A. & Pohnert, G. Pronounced uptake and metabolism of organic substrates by diatoms revealed by pulse-labeling metabolomics. *Front. Mar. Sci.* **9**, (2022).

200. Böttjer, D. *et al.* Temporal variability of nitrogen fixation and particulate nitrogen export at Station ALOHA. *Limnology and Oceanography* **62**, 200–216 (2017).

201. Kiene, R. P. & Linn, L. J. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. *Limnology and Oceanography* **45**, 849–861 (2000).

202. del Valle, D. A., Kiene, R. P. & Karl, D. M. Effect of visible light on dimethylsulfoniopropionate assimilation and conversion to dimethylsulfide in the North Pacific Subtropical Gyre. *Aquatic Microbial Ecology* **66**, 47–62 (2012).

203. del Valle, D. A., Martínez-García, S., Sañudo-Wilhelmy, S., Kiene, R. & Karl, D. Methionine and dimethylsulfoniopropionate as sources of sulfur to the microbial community of the North Pacific Subtropical Gyre. *Aquatic Microbial Ecology* **75**, (2015).

204. Clifford, E. L. *et al.* Taurine is a major carbon and energy source for marine prokaryotes in the north atlantic ocean off the iberian peninsula. *Microb Ecol* **78**, 299–312 (2019).

205. Clifford, E. L. *et al.* Mesozooplankton taurine production and prokaryotic uptake in the northern Adriatic Sea. *Limnology and Oceanography* **65**, 2730–2747 (2020).

206. Kiene, R. P. & Williams, L. P. H. Glycine betaine uptake, retention, and degradation by microorganisms in seawater. *Limnology and Oceanography* **43**, 1592–1603 (1998).

207. Gradoville, M. R. *et al.* Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography* **65**, 1858–1875 (2020).

208. Asher, E., Dacey, J. W., Ianson, D., Peña, A. & Tortell, P. D. Concentrations and cycling of DMS, DMSP, and DMSO in coastal and offshore waters of the Subarctic Pacific during summer, 2010-2011. *Journal of Geophysical Research: Oceans* **122**, 3269–3286 (2017).

209. Bertrand, E. M. *et al.* Vitamin B12 and iron colimitation of phytoplankton growth in the Ross Sea. *Limnology and Oceanography* **52**, 1079–1093 (2007).

210. O'Brien, J. *et al.* Microbial dimethylsulfoniopropionate (DMSP) cycling in the ultraoligotrophic eastern Indian Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography* **206**, 105195 (2022).

211. Kinne, R. The role of organic osmolytes in osmoregulation: from bacteria to mammals. *The Journal of experimental zoology* **265 4**, 346–55 (1993).

212. Kumler, W. *et al.* Metabolites reflect variability introduced by mesoscale eddies in the North Pacific Subtropical Gyre. *Front. Mar. Sci.* **11**, 1481409 (2024).

213. Hagemann, M. Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiology Reviews* **35**, 87–123 (2011).

214. Johnson, W. M. *et al.* Particulate and dissolved metabolite distributions along a latitudinal transect of the western Atlantic Ocean. *Limnology and Oceanography* **68**, 377–393 (2023).

215. Brisson, V. *et al.* Identification of Effector Metabolites Using Exometabolite Profiling of Diverse Microalgae. *mSystems* **6**, e00835-21 (2021).

216. Henderikx-Freitas, F., Karl, D. M., Björkman, K. M. & White, A. E. Constraining growth rates and the ratio of living to nonliving particulate carbon using beam attenuation and adenosine-5′-triphosphate at Station ALOHA. *Limnology and Oceanography Letters* **6**, 243–252 (2021).

217. Ho, T.-Y. *et al.* The Elemental Composition of Some Marine Phytoplankton. *Journal of Phycology* **39**, 1145–1159 (2003).

218. Ksionzek, K. B. *et al.* Dissolved organic sulfur in the ocean: Biogeochemistry of a petagram inventory. *Science* **354**, 456–459 (2016).

219. Moran, M. A. & Durham, B. P. Sulfur metabolites in the pelagic ocean. *Nat Rev Microbiol* **17**, 665–678 (2019).

220. Juranek, L. W. *et al.* The Importance of the Phytoplankton "Middle Class" to Ocean Net Community Production. *Global Biogeochemical Cycles* **34**, e2020GB006702 (2020).

221. Popowich, A., Zhang, Q. & Le, X. C. Arsenobetaine: the ongoing mystery. *National Science Review* **3**, 451–458 (2016).

222. Martiny, A. C. *et al.* Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. *Nature Geosci* **6**, 279–283 (2013).

223. Lee, C., Wakeham, S. G. & I. Hedges, J. Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep Sea Research Part I: Oceanographic Research Papers* **47**, 1535–1568 (2000).

224. Hedges, J. I. *et al.* Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* **409**, 801–804 (2001).

225. Hedges, J. I. *et al.* The biochemical and elemental compositions of marine plankton: A NMR perspective. *Marine Chemistry* **78**, 47–63 (2002).

226. Ingalls, A. E., Lee, C., Wakeham, S. G. & Hedges, J. I. The role of biominerals in the sinking flux and preservation of amino acids in the Southern Ocean along 170°W. *Deep Sea Research Part II: Topical Studies in Oceanography* **50**, 713–738 (2003).

227. Liefer, J. D. *et al.* Latitudinal patterns in ocean C:N:P reflect phytoplankton acclimation and macromolecular composition. *Proceedings of the National Academy of Sciences* **121**, e2404460121 (2024).

228. Yancey, P. & Burg, M. Counteracting effects of urea and betaine in mammalian cells in culture. *The American journal of physiology* **258 1 Pt 2**, R198-204 (1990).

229. Yancey, P. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology* **208**, 2819–2830 (2005).

230. Burg, M. B. & Ferraris, J. D. Intracellular Organic Osmolytes: Function and Regulation. *J Biol Chem* **283**, 7309–7313 (2008).

231. Seelen, E. A. *et al.* Pelagic ecosystem research incubators (PERIcosms): optimized incubation tanks to investigate natural communities under long term, low nutrient, and low metal conditions. *Limnology and Oceanography: Methods* **22**, 548–571 (2024).

232. Cohen, N. R. *et al.* Diatom Transcriptional and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Front. Mar. Sci.* **4**, (2017).

233. Cohen, N. R. *et al.* Transcriptomic and proteomic responses of the oceanic diatom Pseudo-nitzschia granii to iron limitation. *Environmental Microbiology* **20**, 3109–3126 (2018).

234. Kolody, B. C. *et al.* Nitrogen and Iron Availability Drive Metabolic Remodeling and Natural Selection of Diverse Phytoplankton during Experimental Upwelling. *mSystems* **7**, e00729-22 (2022).

235. Nef, C. *et al.* Cobalamin Scarcity Modifies Carbon Allocation and Impairs DMSP Production Through Methionine Metabolism in the Haptophyte Microalgae Tisochrysis lutea. *Front. Mar. Sci.* **7**, (2020).

236. Louca, S. *et al.* Function and functional redundancy in microbial systems. *Nat Ecol Evol* **2**, 936–943 (2018).

237. Ramond, P., Galand, P. E. & Logares, R. Microbial functional diversity and redundancy: moving forward. *FEMS Microbiology Reviews* **49**, fuae031 (2025).

238. Giovannoni, S. J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science* **9**, 231–255 (2017).

239. Norris, N., Levine, N. M., Fernandez, V. I. & Stocker, R. Mechanistic model of nutrient uptake explains dichotomy between marine oligotrophic and copiotrophic bacteria. *PLOS Computational Biology* **17**, e1009023 (2021).

240. Yu, X. A. *et al.* Low-level resource partitioning supports coexistence among functionally redundant bacteria during successional dynamics. *The ISME Journal* **18**, wrad013 (2024).

241. Henson, S. A., Cael, B. B., Allen, S. R. & Dutkiewicz, S. Future phytoplankton diversity in a changing climate. *Nat Commun* **12**, 5372 (2021).

242. Zhang, Z. *et al.* Global biogeography of microbes driving ocean ecological status under climate change. *Nat Commun* **15**, 4657 (2024).

243. Seelen, E. *et al.* A tale of two nutrients: how nitrogen and phosphorus differentially control marine biomass production and stoichiometry. Preprint at https://doi.org/10.21203/rs.3.rs-4203527/v1 (2024).

244. Robertson, D. E., Lesage, S. & Roberts, M. F. β-aminoglutaric acid is a major soluble component of *Methanococcus thermolithotrophicus*. *Biochimica et Biophysica Acta (BBA) - General Subjects* **992**, 320–326 (1989).

245. Martin, D. D., Ciulla, R. A. & Roberts, M. F. Osmoadaptation in Archaea. *Applied and Environmental Microbiology* **65**, 1815–1825 (1999).

246. Peluso, G. *et al.* Carnitine: An osmolyte that plays a metabolic role. *Journal of Cellular Biochemistry* **80**, 1–10 (2001).

247. Sleator, R. D. & Hill, C. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiology Reviews* **26**, 49–71 (2002).

248. Parida, A. K. & Das, A. B. Salt tolerance and salinity effects on plants: a review. *Ecotoxicology and Environmental Safety* **60**, 324–349 (2005).

249. Bashir, A. *et al.* Plant-derived compatible solutes proline betaine and betonicine confer enhanced osmotic and temperature stress tolerance to Bacillus subtilis. *Microbiology* **160**, 2283–2294 (2014).

250. Slama, I., Abdelly, C., Bouchereau, A., Flowers, T. & Savouré, A. Diversity, distribution and roles of osmoprotective compounds accumulated in halophytes under abiotic stress. *Annals of Botany* **115**, 433–447 (2015).

251. Durham, B. P. *et al.* Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat Microbiol* **4**, 1706–1715 (2019).

252. Podbielski, I., Schmittmann, L., Sanders, T. & Melzner, F. Acclimation of marine invertebrate osmolyte systems to low salinity: A systematic review & meta-analysis. *Front. Mar. Sci.* **9**, (2022).

253. Guo, K., Glatter, T., Paczia, N. & Liesack, W. Asparagine Uptake: a Cellular Strategy of Methylocystis to Combat Severe Salt Stress. *Applied and Environmental Microbiology* **89**, e00113-23 (2023).