©Copyright 2018
Katherine Rose Heal

The Power and Promise of Direct Measurements of Metabolites in Marine Systems

Katherine Rose Heal

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Anitra E. Ingalls, Chair

E. Virginia Armbrust

Allan H. Devol

Program Authorized to Offer Degree:
Oceanography

University of Washington

Abstract

The Power and Promise of Direct Measurements of Metabolites in Marine Systems

Katherine Rose Heal

Chair of the Supervisory Committee:
Professor Anitra E. Ingalls
Oceanography

Organic molecules play a myriad of roles in the surface ocean. They are the building blocks of proteins, the alphabet of genetic information, and the language in which microorganisms communicate. The small biomolecules that are the readout of cellular activity, or metabolites, or are of particular interest. Some metabolites like compatible solutes fuel the microbial loop, while others like vitamins act as micronutrients to algae. In this dissertation, I aim to use liquid-chromatography mass-spectrometry (LC-MS) to observe pools of metabolites to answer long-standing questions in microbial ecology and organic matter cycling in the surface ocean, specifically tackling gaps in our knowledge regarding cobalamin (vitamin B_{12}) and microbial metabolomics. In Chapters 2 and 3, I develop and utilize LC-MS techniques to directly measure cobalamin in dissolved and particulate organic matter. These chapters reveal an underappreciated chemical diversity of cobalamin and cobalamin-like compounds in marine systems, revealing that cyanobacteria produce a cobalamin-like compound, pseudocobalamin. I tie this chemical diversity to the biochemistry, ecology, and evolution of marine cobalamin producers in Chapter 3. In Chapters 4 and 5, I use an analytical technique to perform targeted and untargeted metabolomics. These techniques aim to identify and quantify the entire metabolome of an organism (as in Chapter 4) or community (as in Chapter 5). In a controlled laboratory setting, I harness metabolomics to understand how two species of diatoms experience cobalamin limitation; this work is presented in Chapter 4.

In Chapter 5, I explore the whole community metabolome across a natural oceanographic gradient in temperature, salinity, and nutrients in the North Pacific ocean. Here I offer some of the first observations of several metabolites in natural marine systems and hypothesize on their roles in microbial processing of organic matter.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation could not have been possible without a whole team of people supporting me — I have had tailwinds for most of the last seven years.

First and foremost, I want to acknowledge all the female scientists before me who have paved the way for women in science. Much of the work in this dissertation relies on the groundbreaking work done from the 1940s by Dorthy Crowfoot Hodgkin, a chemist in a man's world who awed the scientific world and didn't fit into social conventions. There is much work to be done to make the culture of science more inclusive, but I wholeheartedly appreciate the giants on whose shoulders I stand. Thank you for allowing us to be women, weirdos, *and* scientists.

I would like to thank my advisor, Anitra Ingalls, for her unwavering support throughout graduate school. It has been a true pleasure to work in Anitra's lab, where she fosters creative scientific thinking, independence, and collaboration. Thank you for leading me when I teetered and allowing me to lead when I was capable. I'm looking forward to many more years of collaboration.

Others in Anitra's lab also are deserving of the highest praise. Laura Truxal Carlson has served as my constant companion in the lab and at sea; I often say that working alongside Laura is like having four hands, we make a great team and it has been an immense pleasure to work with such a capable technician and scientist throughout my tenure at UW. I have had the privilege of mentoring and working with several incredible undergraduates including Davey French, Willow Coyotes-Maestas, Alexa Weid, Regina Lionheart, and Natalie Kellog. There is no scenario in which the data presented in this dissertation could have been collected, processed, and analyzed without this team of helpers — thanks guys. Labmates Rachel

Lundeen, Laura Hmelo, Jamie Collins, and Wei Qin have been great sounding boards for my science. My fellow graduate student, Angie Boysen, has taught me heaps about how to be a better coder, analytical chemist, and writer — collaborating closely with her has been a tremendously fruitful and fun experience and that I hope to continue.

Outside of my lab, I have also benefitted from scientific input from my committee, Gabrielle Rocap, Mike Brett, Al Devol, and especially Ginger Armbrust. Throughout my time in graduate school, I have been a part of close collaborations with Ginger's lab (specifically with François Ribalet, Bryndan Durham, and Shady Amin) and Dave Stahl's lab (specifically with Wei Qin and Tony Bertangnolli). Both of these collaborations have taught me a tremendous amount about microbiology, microbial ecology, and biological oceanography in general.

I have spent a fair amount of time at sea collecting the samples presented in my dissertation; a heartfelt thank you to captains, crew members, and scientific teams aboard KM1314, TN280, MGL1707, KOK1606, and several Barnes cruises.

The community of graduate students in oceanography at UW is remarkable. Ashley Maloney has been a stalwart companion in this often challenging journey: remarkably generous with her time, chocolate, and foam roller. Leah Johnson has similarly been generous — ready to ride a bike, grab a cocktail, or hit the spa in moments of need. Nancy Williams will continue to awe everyone, but I've known she was impressive from day one — we made a good little cohort. There are too many other graduate students who have provided varied support to name by name, but here I go and do it anyway; thank you to Rick Berg, Elizabeth Brasseale, Seth Bushinsky, Sarah Dewey, Ryan Grossman, Megan Duffy, Nemiah Ladd, Shirley Leung, Rachelle Lim, Erin Martin, Anna McLasky, Susanna Michaels, Dan Nelson, Hilary Palevsky, Noel Pelland, Justin Penn, Diane Rico, Jake Steinberg, Marta Wolfshorndl and I'm sure others that I'll regret not naming later.

Throughout my education, my loving family has attempted to keep me grounded by

reminding me that there is life outside of this ivory tower: that it still snows in Minnesota and children still grow up too fast.

My non-academic community in Seattle is amazing; the friends I've made here are some of the best people I've ever had the pleasure of meeting. Martina Brimmer is a remarkable friend and woman; I've learned a lot about what makes a good person from her and the community she introduced me to. My house has been a consistent source of love, thank you to all past and current 321 occupants for making our home so special. Finally, infinite thanks to the continued patience, support, and cheerleading from Gabe Ehlert — let's go ride bikes!

DEDICATION

To my mother, for allowing me to chase dreams different than hers.

Chapter 1
INTRODUCTION

1.1 Marine organic carbon

Marine organic carbon is a large pool of actively cycling carbon in the earth system. This pool originates primarily through the fixation of CO_2 by phytoplankton in the surface ocean, where about half of it is readily respired back to CO_2 by marine bacteria [1]. The balance between carbon fixation and respiration is responsible for forming the base of marine food webs and sequestering carbon into the deep ocean through the biological pump. After its initial formation, organic matter is transformed by a number of biotic and abiotic factors, resulting in a chemically complex pool of both dissolved and particulate organic matter.

Previous work has used bulk organic matter measurements to try to discern the cycling of this pool of carbon. Analyzing atomic quotients (i.e. O:C and H:C) yields a rough perspective of the contributions of classes of biomolecules to a bulk organic matter pool, and one can use these bulk measurements to watch how these pools change over time and space. The isotopic composition of bulk organic matter can also yield insight into the origins of different classes of biomolecules by harnessing the average 6‰ difference between land plants and marine plankton in the δ^{13}C. These techniques give us a broad understanding of the source and processing of marine organic matter, though many factors alter the atomic quotients and δ^{13}C of organic matter that cannot be investigated when taking this bulk approach.

Another avenue of research in marine organic geochemistry has focussed on specific classes of molecules that can yield insight into specific processes in the ocean. For instance, fatty acids have been used as biomarkers to assess trophic transfer and food quality [2], pigments yield information on phytoplankton community composition [3], and ^{14}C dating of archaeal

lipids revealed that most archaea in the deep ocean are primarily autotrophic [4]. Several researchers put forth a major effort to characterize particulate marine material into classes of macromolecules to learn about selective remineralization and marine organic matter composition (i.e. [5, 6, 7]). These studies are founded on analytical methods specific to certain analytes that have allowed researchers to isolate and detect compounds of interest. Overall, a quantitative investigation of the small biomolecules, or metabolites, in organic matter has remained analytically unreasonable due to a lack of technique that allows for the observation and quantification of the diverse chemical compound classes that metabolites occupy.

1.1.1 Liquid chromatography-mass spectrometry

In recent decades, liquid chromatography-mass spectrometry (LC-MS) has improved to the point that it now can simultaneously measure thousands of different molecules of many diverse compound classes. LC-MS operates on a separation technique (liquid chromatography) that is remarkably adaptable, capable of performing separation on biopolymers like peptides and intact lipids as well as small organic molecules like osmolytes and amino acids. Unlike gas chromatography, LC-MS does not rely on the chemical alteration of the analyte for detection (derivatization). The mass spectrometry detection technique (electrospray ionization) in LC-MS is also capable of detecting a similar wide array of molecules over a large analytical range of concentrations and detecting most of these compounds in their native forms.

This dissertation employs a variety of LC-MS techniques to measure small organic molecules in the dissolved and particulate pools of marine organic matter to tackle longstanding questions in marine microbial ecology and chemical oceanography. Specially, I investigate distributions, sources, and metabolic consequences of cobalamin in both natural and controlled conditions and explore community metabolomics in the North Pacific ocean.

1.1.2 Power and promise observing metabolites in the ocean

The small organic molecules that are the focus of this dissertation all fall into the larger category of metabolites, or organic molecules that are the byproduct of cellular processing.

Studying the abundance and identity of the small molecules in microbial systems yields a unique perspective that is complementary to genomic and transcriptomic approaches. Beyond looking at the genomic potential or transcriptomic signal in a system, I can see which molecules an organism or community has produced and explore how these molecules change in concentration over environmental parameters of interest. In this way, I can explore the composition of marine organic matter and the metabolic processes that shape the organic matter pool in the ocean.

However, measuring metabolites in the marine environment presents a unique set of analytical challenges. A diverse set of analytical techniques are required to effectively measure such chemically diverse elements, requiring different LC-MS techniques. Dissolved organic matter is notoriously difficult to isolate from the complex salt matrix that dominates the marine environment, making sample preparation difficult. This sea salt causes matrix-induced ion suppression during LC-MS analysis, which must be accounted for and analytical drift is common using LC-MS. Some of these challenges have been recently tackled in our group [8], though much analytical method development work remains.

1.2 Cobalamin in marine systems

One area of research that has been hindered as a result of analytical challenges is the study of cobalamin (vitamin B_{12}) in the ocean. Cobalamin is an organic micronutrient that is produced by a select cohort of bacteria and archaea and is essential to organisms in all domains of life. In particular, over half of eukaryotic phytoplankton require exogenous cobalamin to grow [9], and areas of the ocean have been shown to be limited or co-limited by cobalamin [10, 11]. Measurements of the molecule in seawater, however, have been scarce in part due to low concentration (often < 1 pM) and in part due to the chemical diversity of the pool of molecules that make up the bioavailable pool of cobalamin.

Chapter 2 develops and tests a method for directly measuring cobalamins and other B vitamins dissolved in seawater. In that work, I found that the major form of cobalamin in seawater is a form formerly under-recognized in the field of microbial ecology, hydroxocobal-

amin.

Before the ability to make direct measurements of the compounds, our understanding of the source of cobalamin in the oceans has been limited to genomic inference. Chapter 3 bridges this gap in knowledge and shows that different microbes in the ocean produce cobalamin and cobalamin-like compounds that are not equally bioavailable to the community at large. I combine field and laboratory work to show that this phenomenon is common in natural systems.

1.3 Metabolomics as a window into phytoplankton physiology

Phytoplankton are responsible for fixing organic carbon in the marine environment, producing about half of the worlds oxygen, and forming the base of marine food webs. In many parts of the ocean, their growth is limited by the availability of nutrients, including cobalamin. Phytoplankton have adapted to use cobalamin at low concentrations [12, 13], but the metabolic consequences of cobalamin limitation are still largely unknown. Chapter 4 investigates how metabolite pools in two diatoms change under cobalamin limitation to understand how these globally important organisms cope with cobalamin limitation. For this, I used a metabolomics approach which attempts to identify and quantify the full suite of biomolecules that result from cellular processing, using a method developed by our lab [8]. Our work revealed that diatoms experience an imbalance in the methionine cycle, rearrangement of the osmolyte suite, and change in transsulfuration pathway. I also present strong evidence that diatoms show a decrease in the efficiency of a cobalamin-dependent enzyme, methylmalonyl CoA mutase, whose role is still mostly unknown in diatom biochemistry.

1.4 Community metabolomics

Organisms do not live in isolation, and performing metabolomics of natural microbial communities can offer a unique perspective beyond controlled laboratory organism studies. Community metabolomics has been used in microbiome research to understand how environmental pressures affect the microbiome as a whole. This approach has been rarely applied to marine

systems, despite its promise to inform us about the composition of the marine organic matter and give us insight into microbial metabolism *in situ*. In Chapter 5, I explore metabolite pools over a natural gradient using community metabolomics — specifically the *transitional zone* in the North Pacific Ocean. I identify compounds at high abundance that are previously underappreciated, explore differences in metabolite pools in different oceanographic regimes, and offer some of the first measurements of a suite of biomolecules in marine particulate matter.

Chapter 2

DETERMINATION OF FOUR FORMS OF VITAMIN B_{12} AND OTHER B VITAMINS IN SEAWATER BY LIQUID CHROMATOGRAPHY TANDEM MASS SPECTROMETRY

2.1 Abstract[1]

2.1.1 Rational

Vitamin B_{12} is an essential nutrient for more than half of surveyed marine algae species, but methods for directly measuring this important cofactor in seawater are limited. Current mass spectrometry methods do not quantify all forms of B_{12}, potentially missing a significant portion of the B_{12} pool.

2.1.2 Methods

We present a method to measure vitamins B_1, B_2, B_6, B_7 and four forms of B_{12} dissolved in seawater. The method entails solid-phase extraction, separation by ultra performance liquid chromatography, and detection by triple-quadrupole tandem mass spectrometry using stable isotope-labeled internal standards. We demonstrated the use of this method in the environment by analyzing B_{12} concentrations at different depths in Hood Canal, part of the Puget Sound estuarine system in Washington State.

---

[1] An edited version of this chapter was published as K. R. Heal, L. T. Carlson, A. H. Devol, E. V. Armbrust, J. W. Moffett, D. A. Stahl, and A. I. Ingalls, "Determination of four forms of vitamin B_{12} and other B vitamins in seawater by liquid chromatography/tandem mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 28, no. 22, pp. 2398–2404, 2014

2.1.3 Results

Recovery of vitamin B_{12} forms during the preconcentration steps was $> 71\%$ and the limits of detection were < 0.275 pM in seawater. Standard addition calibration curves in three different seawater matrices were used to determine analytical response and to quantify samples from the environment. Hydroxocobalamin was the main form of B_{12} in seawater at our field site.

2.1.4 Conclusions

We developed a method for quantifying four forms of B_{12} in seawater by liquid chromatography-mass spectrometry with the option of simultaneous analysis of vitamins B_1, B_2, B_6, and B_7. We validated the method and demonstrated its application in the field.

2.2 Introduction

Phytoplankton play a pivotal role in global carbon and nitrogen cycles as the base of the marine food web and as a major vehicle for the uptake of carbon dioxide from the atmosphere. Recent work suggests that auxotrophy for specific B vitamins can control primary productivity and phytoplankton diversity in certain ecosystems [14, 11]. Croft and coworkers [15] found that of 300 surveyed eukaryotic phytoplankton cultures, over 50% were auxotrophic for thiamine (B_1), biotin (B_7), or cobalamin (B_{12}). Community composition and activity of vitamin producers could influence marine eukaryotic primary productivity or diversity through these important biomolecules. For instance, B_{12} is only produced only by select bacteria and archaea but is also required by many marine eukaryotes [15, 16, 17]. While these B vitamins appear to be important and potentially limiting nutrients for some phytoplankton, our understanding of B vitamin cycling in the ocean is largely limited by unvetted and still evolving methods. Vitamin B_{12} is a suite of compounds that contain a cobalamin with different chemical groups attached to the cobalt atom (Figure 2.1). Four main types have been well studied in the medical field but have been largely ignored in environmental set-

tings: cyanocobalamin (CN-Cbl), methylcobalamin (Me-Cbl), hydroxocobalamin (OH-Cbl), and deoxyadenosylcobalamine (Ado-Cbl). Aquacobalamin is another potentially important form, but under acidic conditions is converted to its conjugate base, OH-Cbl [18]. The cobalamins have different bio- and photo- chemistries, making it important to differentiate them in studies of marine nutrient cycling. Of the four forms of cobalamin, Me-Cbl and Ado-Cbl are actively used as cofactors, while OH-Cbl and CN-Cbl are not used directly and are first converted into an active form of B_{12} [19]. All forms of cobalamins are light sensitive, with Me-Cbl and Ado-Cbl identified as extremely labile, photodegrading into OH-Cbl in a matter of seconds after light exposure [20].

Early analyses quantified B_{12} in seawater by monitoring the growth of B_{12}-requiring diatoms [21, 22, 17]. New approaches allow the simultaneous measurement of several B vitamins (B_1, B_2, B_6, B_7) in seawater using solid-phase extraction (SPE) followed by high pressure liquid chromatography (HPLC) coupled to triple quadrupole mass spectrometry (MS), but identify only one form of B_{12}, CN-Cbl [23, 24] The most recently published MS method has a detection limit of 0.18 pM for cyanocobalamin concentrated from 2 L of seawater [24]. All four forms of B_{12} have been quantified in seawater using a diode array detector (DAD) method [25], but this required 4 L of seawater to achieve detection limits of 0.37 pM for OH- Me-, 0.148 pM for CN-, and 0.044 pM for Ado-Cbl. With processing rates of 1 mL min-1, SPE of this large volume is time consuming, subject to breakthrough and unsuitable for unstable analytes.

Measured B_{12} half-saturation constants for cultured algae range from > 0.02 to 13 pM [12] and ambient concentrations of this nutrient are often below existing detection limits [24], demonstrating the need for a highly sensitive analysis. Published extraction methods in the food and medical fields for measuring different B vitamins do not transfer well to the marine environment due to the low analyte concentrations and high salt content of seawater [26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. Here we present an improved method to measure dissolved vitamins in seawater, focusing on the different forms of vitamin B_{12}.

2.3 Methods

2.3.1 Sample collection

Environmental samples were taken near Hoodsport in Hood Canal (HC), Washington, USA at 47° 25.309 N, 123° 6.755 W on 31 May 2013 at approximately 1 pm. Water used for standard curves was collected from the surface at HC, along historical Line P at station P4 (48° 39.0 N, 126° 40.0 W, collected 21 May 2012) and at station ALOHA (22° 45.62 N, 158° 0.03 W, collected 4 September, 2013). These three stations represent very different oceanic regimes (HC, coastal; P4, transitioning between coastal and north pacific gyre; ALOHA, oligotrophic subtropical gyre) and provide three distinct matrices. Seawater was collected in Niskin bottles using a conductivity-temperature-depth (CTD) rosette in a single cast. Temperature and chlorophyll fluorescence measurements were obtained with a Seabird CTD sensor package. Samples were collected into opaque, acid-cleaned, high-density polyethlyene bottles after rinsing three times with the collection water. Seawater was 0.2 μm filtered immediately with a peristaltic pump. The filtrate was collected and separated into two 500 mL acid cleaned opaque bottles. These samples were frozen at -20°C until analyzed.

2.3.2 Preconcentration

Filtered, frozen sea water samples were thawed prior to vitamin preconcentration using a slight modification of previously published methods [25, 23]. Briefly, a C18 SPE column (Waters, 35 mL capacity, 10 g resin) was first conditioned by passing 20 mL of Optima grade (Fisher) methanol followed by at least 20 mL of Milli-Q water through the column. The 500 mL seawater samples were adjusted to pH 5.5-6.5 with HCl and loaded onto the SPE column at 1 mL min^{-1} using a peristaltic pump with tygon tubing. This low flow rate and pH range has been shown [23, 36] to be important for high retention of CN-Cbl and B_1 on C18 columns, and our preliminary tests supported this finding. SPE columns were kept wet during the entire process in a closed system to avoid needing to rewet the columns. The column was then washed with 20 mL Milli-Q water (adjusted to pH 6.5 with NaOH). Samples were eluted

from the column with 40 mL of methanol, which was dried down under clean N_2 gas at low heat (solution temperature was kept below 35°C to avoid thermal degradation of B_1 [37]). Finally, samples were reconstituted in 480 μL of solvent A of the ultra performance liquid chromatography (UPLC) method (see below) and 20 μL of an internal standard mix. To minimize photodegradation, samples were collected in opaque bottles and preconcentration was performed in a dark, windowless room using opaque tubing and foil-covered vials.

2.3.3 Recovery

To determine the percent recovery of each vitamin during SPE, the 0.2 μm filtered seawater samples were spiked with known amounts of the 7 analytes in triplicate either before or after preconcentration on SPE. All samples were analyzed in the same fashion. These values of recovery were used in calculations of vitamins concentrations in seawater to correct for SPE recovery.

2.3.4 UPLC/MS Conditions

A UPLC-electrospray ionization (ESI) MS method was adapted from previously published methods [25, 24, 38]. A Waters Acuity UPLC system coupled to a triple quadrupole (Waters Xevo TQ-S) was used to analyze all samples and standards. A UPLC HSS Cyano column (Waters Acuity, 1.8 μm, 2.1 x 100 mm) was eluted with 20 mM ammonium formate with 0.1% formic acid in water (Solvent A) and acetonitrile (Solvent B). All solvents used for UPLC were Optima LC/MS grade from Fisher. A linear gradient was employed for 5.3 minutes (from 98:2 to 70:30 Solvent A: B), followed by 1 min at 5:95, and re-equilibrated for 1.7 minutes. The flow rate was 0.6 mL/min and the column temperature was 35 °C. All analyses were done in positive ion mode using selective reaction monitoring (SRM). Waters UPLC ESI source temperature was 130°C, with a desolvation temperature of 600°C. Collision energies, cone voltages, and retention times for each analyte are listed in Table 2.1. Peak areas of the listed transitions (in Table 2.1) were used for identification and quantification with internal standards (see below). With the exception of B_7, the SRM method proved very

selective, generating a single peak (per transition) within the expected retention window. To identify B_7, we used the corresponding retention time of the internal standard and the ratio of the two transitions. An example chromatogram (with extracted ions) is shown in Figure 2.2.

2.3.5 Quantification

Preparation and storage of standards

Standards of B_1 (thiamine hydrochloride), B_2 (riboflavin), B_6 (pyridoxine), B_7 (biotin), and CN-Cbl were obtained from AccStandard (water soluble Vitamin Kit); Ado-Cbl, Me-Cbl, OH-Cbl, and ^{13}C labeled B_1, B_2, B_7 (for internal standards) were purchased from Sigma Aldrich. All stock solutions were prepared by dissolving 1 mg of standard in 1 mL of Optima LC/MS grade water (except B_2, which was prepared in 5 mL water). Stock solutions were stored at -20°C in the dark. Working stocks (1 ng/μL) B_1, B_2, B_6, B_7, CN-Cbl and OH-Cbl were prepared fresh for standard curves; no appreciable degradation was noted over a 6 month period when stored at 4°C in the dark. Working stocks of Me-Cbl and Ado-Cbl (1 ng/μL) were prepared daily from frozen stock solutions due to rapid degradation of these compounds. Internal standard stock solutions were kept at -20°C in the dark and did not exhibit degradation over a 6 month period.

Evaluation of matrix effect

A representative set of environmental matrices of inorganic and organic molecules consisted of 0.2 μm-filtered seawater collected from three locations (Hood Canal, Line P station P4 and station ALOHA) into clear polycarbonate bottles and maintained in a full spectrum light incubator for 2 days to degrade the endogenous vitamins. Although this significantly lowered the matrix blanks, some vitamins (B_1, B_6, B_7, CN-Cbl, OH-Cbl) were not completely removed. Therefore, standard addition curves with additions of 0.05 to 5 pg (for B_1, B_2, B_6, and all the forms of B_{12}) or from 0.25 to 25 pg (for B_7, which had a more extreme matrix

effect) were made in 500 mL of each matrix. These standards were then preconcentrated and analyzed as described above. Internal standards were added after preconcentration and drying to account for different matrix effects between samples at final concentrations of 2 (B_2) or 5 pg/μL (B_1, B_7).

Limit of detection, precision, carry over

Limits of detection and quantification were defined as three and ten times, respectively, the standard deviation of the most dilute concentration in the calibration curve over several analyses, using the more intense transition (Table 2.1), though the 2nd transition was always present at LOQ [39]. Intra- and inter-day precision was defined as standard deviation of 5 injections. MQ blanks were run after standard curve and between each sample to avoid any carry-over effect, although carry over was always $< 1\%$ of the sample peak in these blanks.

2.4 Results and Discussion

Here we present a new SPE and LC/MS-based method for the analysis of B_1, B_2, B_6, B_7 and the four major forms of cobalamin (B_{12}) in seawater. This method reduces sample processing time, explicitly accounts for matrix effects by using isotope labeled internal standards, allows chromatographic separation of the four forms of B_{12}, and improves detection limits beyond previous methods [24, 25].

The published stable pH range for all analyzed cobalamins is pH 4-7 [40], with lower pH's resulting in lower retention on a C18 SPE and acid-induced heterolysis of Ado-Cbl [23, 41]. Me-Cbl and Ado-Cbl are particularly labile, so we focused on decreasing processing time (from 66 to 8 hours for the preconcentration step) and eliminating the 2nd SPE (pH 2) step required in previously published methods. We found poorer recoveries of B_1 (37%) and B_7 (48%) when the pH 6.5 SPE was followed by a pH 2 SPE on a single C18 resin. We therefore chose to eliminate the pH 2 SPE step, which halved the processing time and reduced the potential for methodological artifacts, especially for Me-Cbl and Ado-Cbl. Our method reduces the required volume 4-8 fold and involves a single SPE step, saving over 58

hours of time in the preconcentration step per sample [25, 24] .

2.4.1 Method validation

Recoveries for all analytes during the SPE step were high (> 70%) at a flow rate of 1 ml min^{-1} and pH 5.5-6.5, except B_1 and B_7, which were never higher than 55% and 67% respectively (Table 2.2). The poorer recovery for B_1 is likely due to breakthrough on the C18 SPE column, as the chromatography indicates that this very polar molecule with a quaternary amine elutes with the solvent front at 0.61 min (Figure 2.1). The method was specifically optimized for the B_{12} forms.

We used ^{13}C isotope-labeled internal standards of vitamins B_1, B_2, and B_7 to account for matrix effects. A standard curve was generated for each analyte in the three different seawater matrices using three different isotope-labeled internal standards (B_1, B_2, and B_7). The standard curves were evaluated based on variation in the response factors in the matrix and the R_2 of the resulting linear regressions. We normalized the peak areas of each analyte to each isotopically-labeled internal standard. The standard deviation of the response factor of each internal standard against each analyte (Figure 2.3a) represents the degree of variation observed in samples analyzed in each of the matrices. Both the standard deviation of the response factors and the R^2 of the linear regression produced by internal standard normalized curves (Figure 2.3b) were used to select the best internal standard for each compound (Table 2.2). We found that in all cases except B_6, using a standard curve normalized to an internal standard corrected the data for variations in the response factor in different matrices (Figure 2.4).

Standard addition calibration curves for all analytes gave high R^2 values (> 0.97 with internal standards), and low limits of detection (< 0.3 pM in seawater) with the exception of B_7, which had a much greater matrix effect (Table 2.2). The low response factor of B_7 in matrix is likely due to coelution of other organic molecules, which introduced large errors in peak integrations (Figure 2.1). Our detection limit was lower than those reported using existing methods for B_1, B_6, B_2, and CN-Cbl, despite using only 25% of the original sample

size, but was higher for B_7 due to a large matrix effect. Measurements of vitamins B_1 and B_7 were associated with large errors and poor recoveries, which may explain our inability to reproduce previously reported recoveries [24]. Intra- and inter-day precision emphasized the importance of running a matrix curve and samples on the same day in a single batch (Table 2.2).

2.4.2 Application

We applied our new method to a depth profile from the Hood Canal (Puget Sound, Washington) field site (Figure 2.5). Vitamin concentrations varied with depth and were generally higher in surface waters, lower at mid-depth and then increased in our deepest samples. B_1 concentrations ranged from 0.58 to 1.5, B_2 from 45 to 128, B_6 from 1.3 to 5.7, and B_7 from < 0.9 to 16.9 pM. Throughout the water column, B_{12} concentrations were dominated by OH-Cbl (1.56 to 5.8 pM) with Ado-Cbl contributing a significant percentage of the B_{12} pool, especially at the surface (< 0.14 to 1.2 pM). By contrast, Me-Cbl and traditionally measured CN-Cbl were much lower, never contributing more than 2% (Me-Cbl) or 21% (CN-Cbl) to total B_{12} inventories at any given depth.

The most extensive published data sets of direct measurements of B_{12} in the ocean measure the CN-Cbl form. In the Hood Canal system, the dominant cobalamin was OH-Cbl, a degraded but still bioavailable product of other B_{12} forms, with CN-Cbl contributing little to the pool of B_{12} at most depths (> 0.7 pM, Figure 5) [24, 17]. Previously published MS methods [24] would suggest a B_{12} concentration an order of magnitude lower than the entire cobalamin pool in Hood Canal by not including OH-Cbl in the quantification of dissolved B_{12}. Suarez-Suarez *et. al.* also found the B_{12} pool distributed between the four forms in surface waters of Palma de Mallorca Bay, Spain [25], corroborating that measurement of the different forms of B_{12} is paramount when attempting to evaluate the availability of dissolved B_{12} to marine communities.

Our data and those of Sañudo-Wilhelmy and coworkers [24] are the only published B_2, B_6, and B_7 concentrations in seawater. Our B_6 and B_7 values are of the same order of

magnitude as those previously reported, while our values for vitamin B_2 are consistently two orders of magnitude higher than those previously reported in seawater (< 0.7 pM to 7 pM) [24]. Our B_1 values are consistently lower than previously reported [24], which may reflect a systematic underestimation due to poor recovery, despite the values being corrected for our estimated SPE recovery efficiencies. Concentrations of B vitamins can vary dramatically over space and depth and often vary independently of one another [24, 17]. The use of consensus reference materials would bolster the ability to compare datasets between labs. We are just beginning to understand the spatial and temporal variability of these metabolites, but accurate measurements of labile B_{12} forms in the smaller sample volumes processed here should allow us to better constrain the variability through time series and rate experiments. There are no published methods for direct measurement of vitamins in marine particulates, which may be a large pool of vitamins in the ocean.

It is unclear whether measurements of the thiamine or pyridoxine (the B_1 and B_6 forms measured in this study respectively) are representative of the suite of molecules that make up the total pool of those vitamins. Some forms of B_1 and B_6 are extremely polar, which would not be efficiently retained on the SPE column (like the form of B_1 analyzed in this study). The ecological importance of the varied forms and moieties of B_1 is only beginning to be understood [42, 17], and a second methodology utilizing a concentrating mechanism that incorporates these polar molecules would further improve our ability to study these compounds.

2.5 Conclusions

In summary, we have improved upon existing methods to simultaneously quantify four forms of vitamin B_{12} as well as B_1, B_2, B_6, and B_7. The method greatly decreases sample size, processing time, and pH manipulations. We quantify our analytes using an isotope-labeled internal standard, giving us high confidence that in a complex mixture of organic molecules a change in the analytical response is due to a change in concentration rather than a difference in matrix effects. We applied this technique by quantifying the eight analytes throughout

a water column in Hood Canal, Washington. This work demonstrates the importance of measuring all forms of B_{12} in order to understand the cycling of this important nutrient in the marine environment.

2.6 Acknowledgements

We would like to acknowledge Willow Coyote-Maestas for assistance in processing samples, and the crew of R/V Barnes for help in the field. This research was supported by the National Science Foundation's Graduate Research Fellowship Program, the University of Washington's Program on Climate Change Graduate Student Fellowship, and NSF MRI award OCE-1228770 to AEI and DAS and Dimensions of Biodiversity award OCE-1046017 to AEI, EVA, DAS, AHD, and JWM. We would also like to acknowledge two anonymous reviewers who greatly improved this manuscript.

2.7 Figures

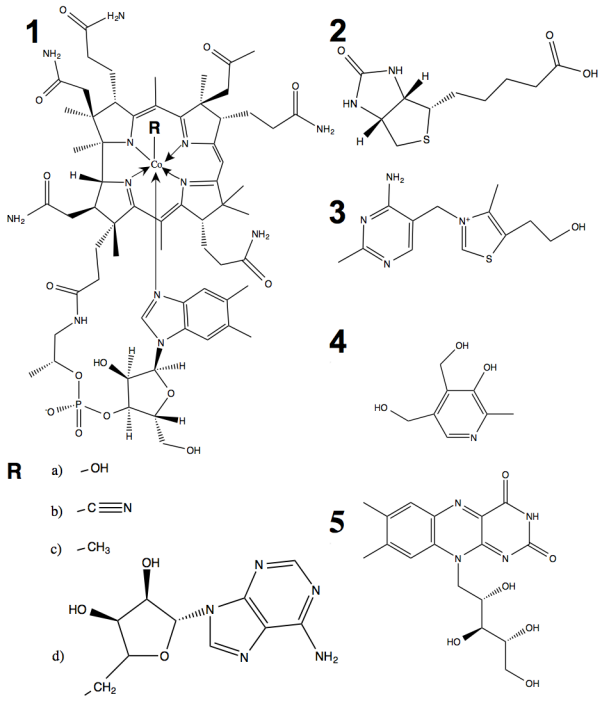

Figure 2.1: Structures of analytes in this study. (1) Cobalamin (B_{12}), with four functional forms: (a) Hydroxocobalamin (OH-Cbl), (b) Cyanocobalamin (CN-Cbl), (c) Methylcobalamin (Me-Cbl), and (d) 5'-deoxy-5'-adenosylcobalamin (Ado-Cbl); (2) biotin (B_7); (3) thiamine (B_1); (4) pyridoxine (B_6); (5) riboflavin (B_2)

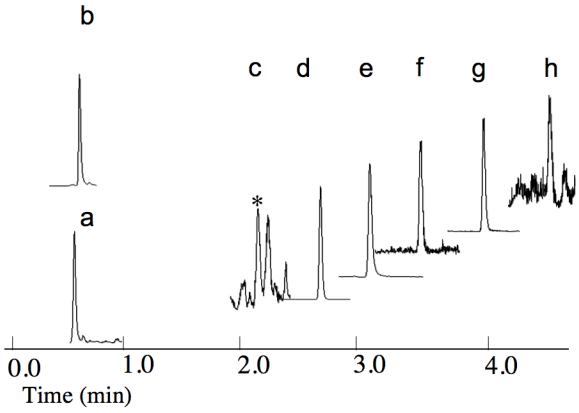

Figure 2.2: Example chromatogrms of extracted ion pairs from a processed sample (surface sample at Hood Canal): a) B_1, b) B_6, c) B_7 (* marking peak), d) B_2, e) OH-Cbl, f) CN-Cbl, g) Ado-Cbl, h) Me-Cbl. Y axis is intensity of individual transitions, each chromatogram set to 100% relative intensity for visualization since the instrumental response can vary over three orders of magnitude between analytes. For biotin, the peak is identified by both IS and secondary transition. See Table 2.1 for extracted ion pairs

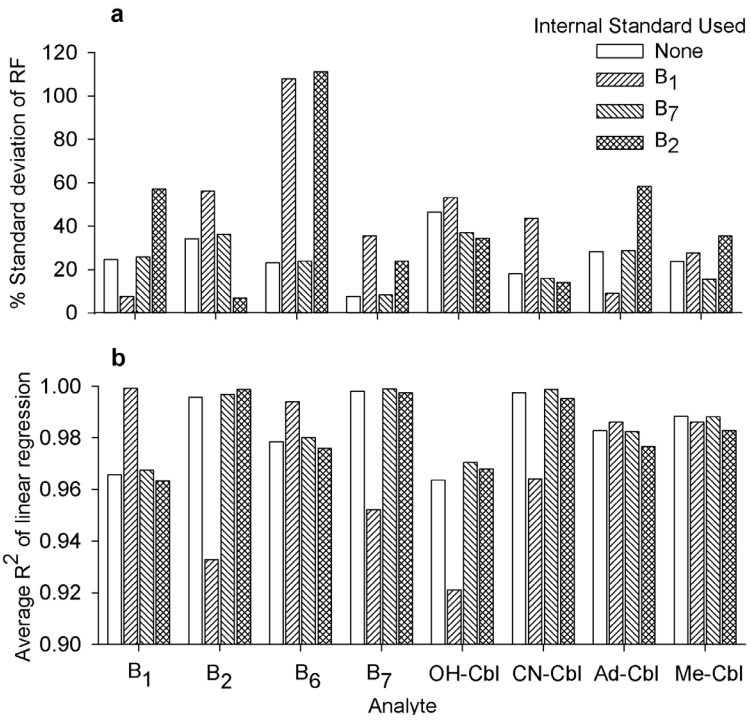

Figure 2.3: Identification of appropriate standard for analyte quantification. For each analyte, a linear regression was performed on calibration curves prepared in three different matrices using different isotope-labeled internal standards (B_1, B_2, B_7). a) Standard deviation of the response factor (RF) for each analyte-internal standard pair; b) average R^2 of the linear regressions in three matrices.

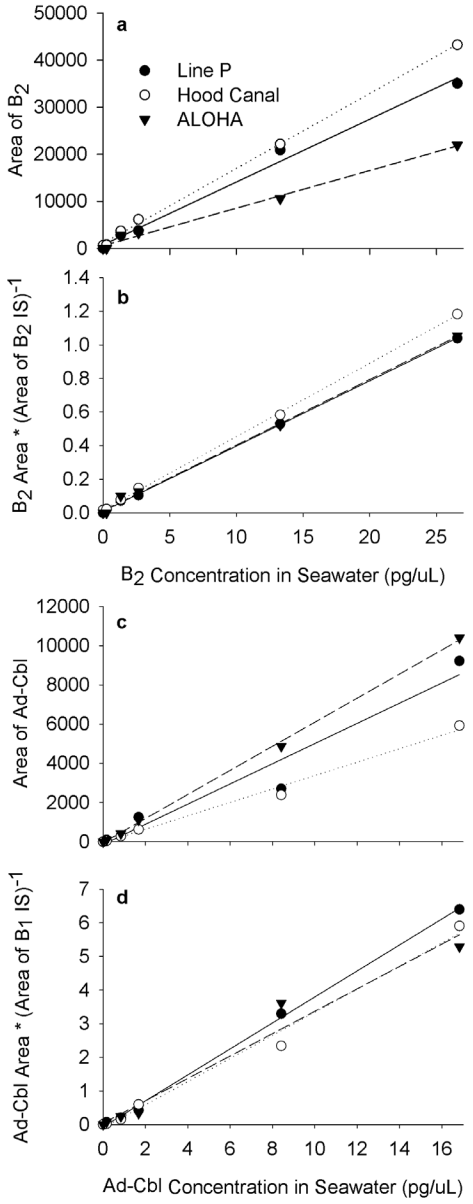

Figure 2.4: Example standard curves of B_2 (a, b) and Ado-Cbl (c, d) in three different matrices: Line P Station P4, closed circles with solid regression line; Hood Canal, open circles with dotted regression line; and ALOHA, closed triangles with dashed regression line. Plots of raw concentration vs peak areas, for B_2 (a) and Ado-Cbl (c) demonstrate the effect of matrix on instrument response. Plots of concentration vs peak area normalized to internal standard peak area for B_2 (b) and Ado-Cbl (d) demonstrate the uniformity of slopes obtained after normalization.

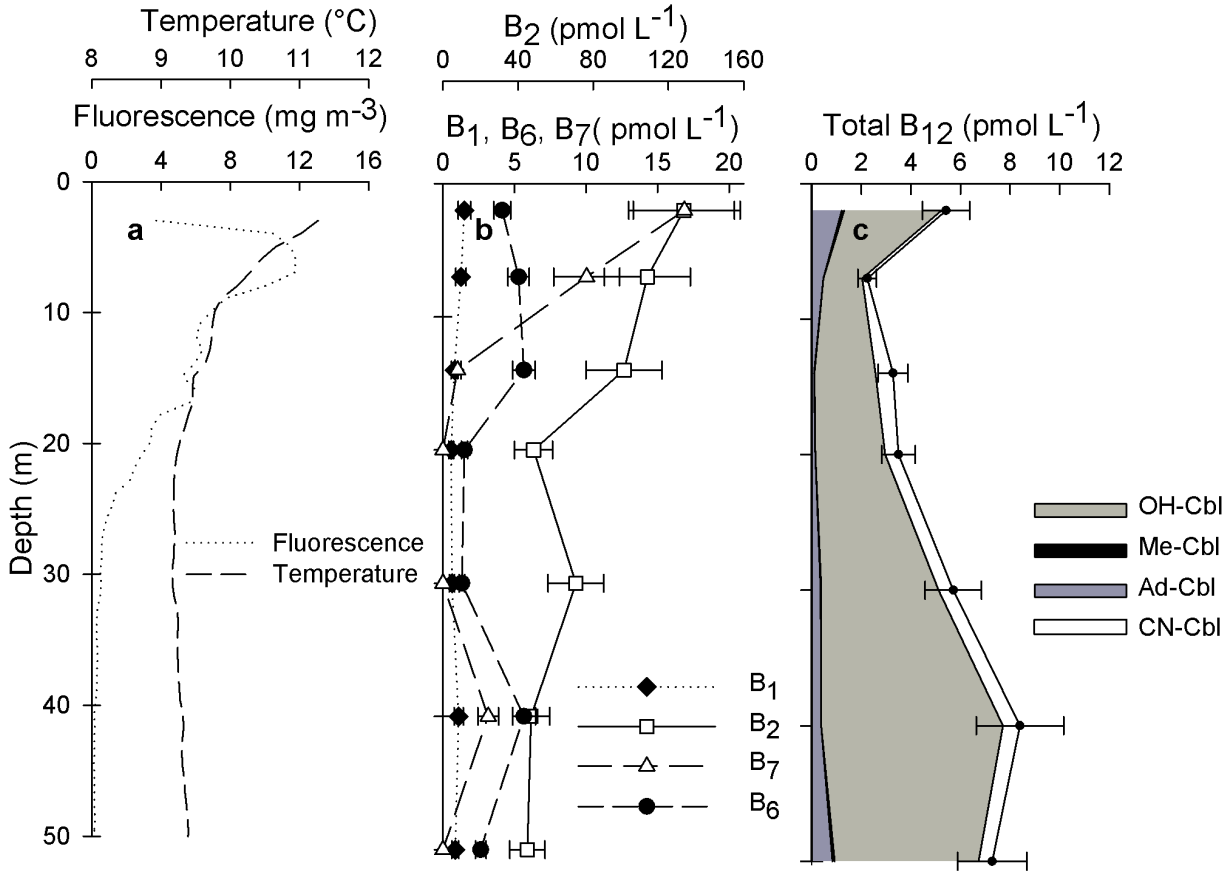

Figure 2.5: Depth profile of vitamin concentrations in Hood Canal. (a) Temperature and fluorescence; (b) B_1, B_2, B_6, B_7; (c) Total B_{12} with individual contributions (of CN-Cbl, OH-Cbl, Me-Cbl, and Ado-Cbl) to the pool shown by shaded regions. Data are the means of duplicate measurements for all vitamins. Error bars represent the standard deviation mainly associated with the preconcentration step, which introduces the largest source of error in the method.

2.8 Tables

Analyte	SRM	CE (V)	Cone volt. (V)	RT (min)
B_1	265.1 -> 122.1,*144.1*	12	6	0.63
B_1-IS	269.2 -> 122.1, *148.1*	12	2	0.63
B_6	170.1 -> 134.1,*79.6*	18	2	0.66
B_7	245.1 -> 97.1, *123.0*	28	4	2.14
B_7-IS	247.1 -> 99.0, *124.4*	26	2	2.14
B_2	377.2 -> 243.2, *172.1*	22	6	2.65
B_2-IS	383.2 -> 249.1, *175.1*	22	2	2.65
OH-Cbl	664.7 -> 147.1, *359.2*	34	42	3.1
CN-Cbl	678.6 -> 147.1, *359.2*	36	22	3.46
Ado-Cbl	790.7 -> 147.1, *665.6*	54	26	3.99
Me-Cbl	673.1 -> 147.1, *685.6*	38	46	4.52

Table 2.1: MS conditions and retention times (RT) for each analyte and internal standard (IS). Selective reaction monitoring (SRM) was used to identify and quantify each parent and daughter pair at the indicated collision energy (CE) and cone voltage. First listed daughter was used to quantify with second daughter (italicized) used to confirm identity

Analyte	% Recovery	LOD (pM)	LOQ (pM)	day %Precision	day %Precision	R^2 (w/IS)
B_1	55 ± 29	0.059	0.197	4.2	11.1	>0.99
B_6	101 ± 14	0.149	0.498	6	7.5	0.97*
B_7	67 ± 23	0.898	2.99	5.3	8.2	>0.99
B_2	93 ± 21	0.124	0.412	5	14	>0.99
OH-Cbl	88 ± 13	0.275	0.917	5.7	6.8	0.99
CN-Cbl	71 ± 4	0.014	0.046	6.9	9.1	0.99
Ado-Cbl	96 ± 13	0.138	0.459	5.9	15.5	>0.99
Me-Cbl	90 ± 24	0.024	0.08	2.3	36.9	0.97

Table 2.2: Results of recovery analysis (with standard deviation of triplicate analysis), limit of detection (LOD), limit of quantification (LOQ), Intra- and inter- day % precision, R^2 , and internal standards (IS) used. *For B_6, no internal standard was used.

Chapter 3

TWO DISTINCT POOLS OF B_{12} ANALOGS REVEAL COMMUNITY INTERDEPENDENCIES IN THE OCEAN

3.1 Abstract[1]

Organisms within all domains of life require the cofactor cobalamin (vitamin B_{12}), which is produced only by a subset of bacteria and archaea. Based on genomic analyses, cobalamin biosynthesis in marine systems has been inferred in three main groups: select heterotrophic Proteobacteria, chemoautotrophic Thaumarchaeota, and photoautotrophic *Cyanobacteria*. Culture work demonstrates that many *Cyanobacteria* do not synthesize cobalamin, but rather produce pseudocobalamin, challenging the connection between the occurrence of cobalamin biosynthesis genes and production of the compound in marine ecosystems. Here we show that cobalamin and pseudocobalamin coexist in the surface ocean, have distinct microbial sources, and support different enzymatic demands. Even in the presence of cobalamin, *Cyanobacteria* synthesize pseudocobalamin — likely reflecting their retention of an oxygen-independent pathway to produce pseudocobalamin, which is used in as a cofactor in their specialized methionine synthase (MetH). This contrasts a model diatom, *Thalassiosira pseudonana*, which transported pseudocobalamin into the cell but was unable to use pseudocobalamin in their homolog of MetH. Our genomic and culture analyses showed that marine Thaumarchaeota and select heterotrophic bacteria produce cobalamin. This indicates that cobalamin in the surface ocean is a result of *de novo* synthesis by heterotrophic bacteria or via modification of closely related compounds like cyanobacterially-produced pseudocobalamin. Deeper in the

[1]An edited version of this chapter was published as K. R. Heal, W. Qin, F. Ribalet, A. D. Bertagnolli, W. Coyote-Maestas, L. R. Hmelo, J. W. Moffett, A. H. Devol, E. V. Armbrust, D. A. Stahl, and A. E. Ingalls, "Two distinct pools of B_{12} analogs reveal community interdependencies in the ocean," *Proceedings of the National Academy of Sciences*, vol. 114, no. 2, pp.364–369, 2017.

water column, our study implicates Thaumarchaeota as the major producers of cobalamin based on genomic potential, cobalamin cell quotas, and abundance. Together, these findings establish the distinctive roles played by abundant prokaryotes in cobalamin-based microbial interdependencies that sustain community structure and function in the ocean.

3.2 Introduction

Cobalamin (vitamin B_{12}) is synthesized by a select subset of bacteria and archaea, yet organisms across all domains of life require it [17, 43, 15]. In the surface ocean, cobalamin auxotrophs (including most eukaryotic algae [15]) obtain the vitamin through direct interactions with cobalamin producers [15] or breakdown of cobalamin-containing cells [44, 45]. Interdependencies between marine cobalamin producers and consumers are critical in surface waters where primary productivity can be limited by the availability of cobalamin and the compound is short-lived [10, 46, 17]. The exchange of cobalamin in return for organic compounds is hypothesized to underpin mutualistic interactions between heterotrophic bacteria and autotrophic algae [15, 47, 10, 48]. The apparent pervasiveness of cobalamin biosynthesis genes in chemoautotrophic Thaumarchaeota and photoautotrophic *Cyanobacteria* genomes [49, 17, 50] raises the question of whether cobalamin production by these autotrophs may underlie additional, unexplored microbial interactions.

Cobalamin is a complex molecule with a central cobalt-containing corrin ring, an α ligand of 5,6-dimethylbenzimidizole (DMB), and a β ligand of either OH-, CN-, Me- or Ado- [51] (Fig. 3.1). Previous studies have shown that instead of producing cobalamin, *Cyanobacteria* produce pseudocobalamin [52, 53, 50], an analog of cobalamin in which adenine substitutes for DMB as the α ligand [51] (Fig. 3.1). Production of pseudocobalamin in a natural marine environment has not been shown, nor have reasons for the production of this compound in place of cobalamin been elucidated.

3.3 Results and Discussion

To explore the pervasiveness of cobalamin and pseudocobalamin supply and demand in marine systems, we determined the standing stocks of these compounds in microbial communities from surface waters across the North Pacific Ocean using liquid-chromatography mass spectrometry (LC-MS). Our LC-MS method [54] quantifies cobalamin and pseudocobalamin with different α ligands. We found that in the surface ocean, pseudocobalamin and cobalamin concentrations associated with organisms and detritus captured on a 0.2 μm filter (particulate fraction) were often of equal magnitude (Fig. 3.2b). Pseudocobalamin had peak concentrations within the euphotic zone at each station and was not detected at the bottom or below the euphotic zone. In contrast, cobalamin was measurable throughout the sampled waters and maintained similar or higher concentration from the lower euphotic zone to our deepest samples (Figs. 3.2a, 3.3).

The overlapping spatial distribution of cobalamin and pseudocobalamin suggests that these cofactors are produced in each other's presence, likely with different sources and sinks. To investigate correlations between *Cyanobacteria* and pseudocobalamin abundance, we compared observations of *Cyanobacteria* carbon inferred from flow cytometry with pseudocobalamin measurements taken at the same depth. In cases where we had both continuous measurements (by SeaFlow [55]) and discreet flow cytometry data, we used the discreet measurements, as collection of these samples was closely coupled in time and space to pseudocobalamin sampling ($n = 16$ for discreet, $n = 4$ for continuous). Pseudocobalamin concentrations are statistically correlated with carbon from *Synechococcus* and *Prochlorococcus* ($R^2 = 0.71$, $p < 0.001$), both in the surface and into the subsurface ocean (Fig. 3.2c), suggesting a primarily cyanobacterial source. No significant correlation existed between *Cyanobacteria* carbon and cobalamin concentrations (Fig. 3.4).

To identify the major producers of cobalamin and pseudocobalamin in the environment, we investigated representative marine isolates and then expanded our search into available genomes that encompass the phylogenetic diversity at our study site. As expected [17, 47],

two strains of marine Alphaproteobacteria with cobalamin synthesis genes (*Sulfitobacter* sp. SA11 and *Ruegeria pomeroyi* DSS-3) produced cobalamin, while the gammaproteobacterium *Vibrio fischerii* ES114 (which lacks cobalamin biosynthesis genes) did not (Table 3.1). Four pure strains of marine chemoautotrophic Thaumarchaeota (*Nitrosopumilus* spp. SCM1, HCE1, HCA1, and PS0) also produced cobalamin (Table 3.1), confirming earlier suggestions based on the presence of cobalamin biosynthesis genes in Thaumarchaeota genomes [49]. Like other *Cyanobacteria* [52, 53, 50], four axenic strains of marine *Cyanobacteria* (*Prochlorococcus* MED4 and MIT9313 and *Synechococcus* WH8102 and WH7803) produced pseudocobalamin (Table 3.1). In all of the cobalamin or pseudocobalamin producers, we detected compounds with β ligands Me-, Ado-, and OH-, but not CN- (Table 3.1).

The observed cell quotas of cobalamin or pseudocobalamin expressed on a per cellular carbon basis varied greatly among producers (Table 3.1). Laboratory cultures of Alphaproteobacteria and Procholorococcus strains had lower amounts of cobalamin or pseudocobalamin (less than 1,200 nmol cobalamin analog per mole carbon) than *Synechococcus* and Thaumarchaeota isolates (1,480-11,600 nmole cobalamin analog per mole carbon). Published values [56] for sea ice bacterial isolates estimated using a bioassay were highly variable (0.6-6,800 nmole cobalamin analog per mole carbon). In our environmental samples, we observed an average stoichiometry of 87 nmole pseudocobalamin per mole cyanobacterial carbon, lower than the cyanobacterial isolates (Fig. 3.2c). This suggests that the cellular stoichiometry of pseudocobalamin is variable and possibly influenced by environmental factors like nutrient availability and growth rate.

To expand the breadth of our survey beyond laboratory isolates, we inspected publically available whole genome sequences from bacteria and Thaumarchaeota for evidence of cobalamin biosynthesis. This analysis expands on previous work [57] while focusing on the phylogenetic groups present at our study site. We analyzed full genomes from the IMG database (https://img.jgi.doe.gov) from phylogenetic groups that encompassed $> 99.9\%$ of the Bacterial 16S rRNA gene sequences from our environmental samples to develop a systematic inference of cobalamin synthesis capacity (3,410 genomes). Alpha- and Gammapro-

teobacteria are hypothesized to be major marine cobalamin producers [49, 17] and 94% of the surveyed genomes from these groups that contain genes necessary for corrin biosynthesis (i.e., cbiA/cobB, cbiH/cobJ) [43] also have genes for DMB synthesis and activation (*bluB*, *cobT*) [58, 43, 59] (Fig. 3), consistent with the synthesis of cobalamin and the results from our representative cultures. All of the eight available high quality Thaumarchaeota genomes in the IMG database code for corrin and DMB biosynthesis genes (Fig. 3.5). Most of the lower quality, incomplete genomes available follow this same pattern (17/19, Dataset S1). No Thaumarchaeota genomes possess the *cobT* gene and thus must activate DMB through a different pathway. Of the 255 *Cyanobacteria*l whole genomes, 247 possessed genes for the synthesis of the corrin ring, but only one genome possessed an annotated *bluB* or *cobT* gene (Fig. 3.5) suggesting the vast majority of *Cyanobacteria* are unable to produce DMB, in agreement with a recent study that examined a subset of the available *Cyanobacteria* genomes [50].

All of the 49 *Prochlorococcus* genomes have genes for corrin synthesis without genes for DMB synthesis or activation and our analysis demonstrated that *Prochlorococcus* MED4, with its highly streamlined genome [60], has maintained these genes to synthesize pseudocobalamin. We propose this originates from an ancient specialization to the production and use of pseudocobalamin in lieu of cobalamin among *Cyanobacteria*. Biosynthesis of the corrin ring can occur via two separate pathways: an O_2-dependent or an O_2-independent pathway [57, 43]. DMB synthesis can also occur via two separate routes, the O_2-dependent bluB [58], or the O_2-independent and O_2-sensitive bza pathway [61]. Rhodobacters have the O_2-dependent corrin ring and DMB pathways [50], while Thaumarchaeota likely possess the O_2-independent pathway for the corrin ring and the O_2-dependent DMB pathway [49]. In some bacteria, pseudocobalamin can be produced if DMB synthesis is impaired; this is due to the natural presence of adenine in cells and enzyme substrate promiscuity that allows adenine to substitute for DMB in some organisms' CobT [62, 61, 63, 64, 65, 66]. *Cyanobacteria* use the O_2-independent pathway for corrin ring synthesis and neither pathway for DMB synthesis [50, 57] (Fig. 3.5). The use and production of cobalamin as a cofactor predates

oxygenic photosynthesis [67, 68]. Thus, possessing an O_2-independent mode for producing a cobalamin analog that is not impaired by O_2 may have been necessary for the success of oxygenic photosynthetic *Cyanobacteria* that were largely responsible for the rise of O_2 on earth and likely endured variable O_2 concentrations over time [69].

Cyanobacteria use pseudocobalamin as a cofactor in two enzymes: methionine synthase (MetH) and class II ribonucleotide reductase (NrdJ). The 3-D structure of MetH contains three β pleated sheets and two α helices that form a pocket for the DMB ligand of cobalamin in *E. coli* [70, 71]. Cyanobacterial MetH is predicted to form the same pocket [52]. However, conserved amino acids within this pocket in the cyanobacterial MetH differ from sequences of organisms known to use cobalamin (Figs. 3.6 and 3.7), suggesting a structure that preferentially binds pseudocobalamin in place of cobalamin as experimentally demonstrated in the cyanobacteria *Arthrospira* [52]. Many *Cyanobacteria* also code for O_2-independent *NrdJ* [72], which has limited sequence similarity to non-cyanobacterial *NrdJ* [73, 74]. Similar to pseudocobalamin biosynthesis in *Cyanobacteria*, NrdJ is both O_2-independent and O_2-insensitive and has been hypothesized as an important bridge between the pre-and post-oxygenated oceans [75]. The continued maintenance of both the biosynthetic pathway and pseudocobalamin-dependent enzymes throughout the diverse *Cyanobacteria* phylum (from *Arthrospira* and *Synechocystis* to the highly streamlined *Prochlorococcus*) suggests the production of pseudocobalamin is an ancient relic that persists in the oxic marine environment today.

For many eukaryotic algae, pseudocobalamin supports lower growth yields than cobalamin [76, 77, 50]. We examined the underlying cause of this reduced growth by supplying the model diatom *Thalassiosira pseudonana* (CCMP 1335) with pseudocobalamin and tracking it into the cells. Like others, we found that growth of *T. pseudonana* is limited at 1 pM cobalamin [46], and the addition of pseudocobalamin (at 200 pM) is unable to overcome this limitation [50, 77]. We observed that *T. pseudonana* takes up the inactive OH-pseudocobalamin from their growth media and converts it into the cofactor form Ado-pseudocobalamin, yet is unable to recover to cobalamin-replete growth rates (Figs. 3.8, 3.9, Table 3.2). The role of

Ado-cobalamin in diatoms is unclear, although *T. pseudonana* does code for Ado-cobalamin dependent methylmalonyl-CoA mutase (MCM) and actively transcribes a putative adenosyl-cobalamin transferase (which converts OH-cobalamin to Ado-cobalamin) under cobalamin limitation [46]. Previous studies suggest that when diatoms are starved for cobalamin, low Me-cobalamin (required for MetH activity) deprives cells of S-adenosylmethionine (SAM) [46, 46]. We found that when pseudocobalamin is supplied to cobalamin-limited *T. pseudonana*, they contain significantly less SAM per cell than cobalamin-replete conditions (Figs. 3.8 , 3.9, Table 3.2). The depressed levels of SAM and lack of detectable Me-pseudocobalamin within cells suggest that *T. pseudonana* are unable to efficiently use pseudocobalamin in their MetH. This same phenomenon has been demonstrated physiologically in marine cobalamin-dependent bacteria — cobalamin or methionine stimulate growth, but pseudocobalamin does not [78]. These combined findings demonstrate that marine *Cyanobacteria* achieve both independence and a competitive advantage by producing and requiring only pseudocobalamin; they meet their own need for their preferred cofactor while also not directly supplying cobalamin to other photoautotrophs.

Corrin synthesis genes in oligotrophic surface water metagenomes are dominated by pseudocobalamin-producing *Cyanobacteria* [49], suggesting that *de novo* production of true cobalamin in these regions is limited to a small subset of heterotrophic bacteria, including clades like Rhodobacterales that are commonly associated with eukaryotic algae [48]. Cobalamin auxotrophs may foster close physical or chemical relationships with these cobalamin producers, offsetting the expense of cobalamin biosynthesis in order to secure a consistent cobalamin source. Alternatively, organisms may employ techniques to use closely related compounds like pseudocobalamin. Many organisms without the biosynthetic capacity for *de novo* production of cobalamin have salvage or remodeling strategies for reactivating cobalamin analogs [79, 59, 50]. A recent study has shown that some eukaryotic algae have the genetic capacity to make cobalamin when supplied with pseudocobalamin and DMB (the α ligand of cobalamin), though this phenomenon may be limited by low DMB concentrations in natural seawater [50], and DMB biosynthesis seems to be limited to bacteria and archaea.

Of the 3,408 phylogenetically representative prokaryote genomes we surveyed, we found 73 genomes that did not have the biosynthetic pathway for the corrin moiety of cobalamin, but did have genetic capacity for cobalamin repair as well as synthesis and activation of DMB, which has no known role in cells beyond α ligand of cobalamin. These organisms include several heterotrophic bacteria known to occur in marine environments such as Methylophaga and Marinobacter (Dataset S3.1). Previous work has suggested that ligand-bound cobalt in the low latitudes may be cobalamin degradation or precursor compounds, which are present at concentrations much higher than has been observed for intact cobalamin [80, 17]. In these environments, organisms capable of salvaging and repairing cobalamin degradation products or precursors could be major suppliers of cobalamin to the environment and *de novo* biosynthesis may take a minor role. The prevalence of cobalamin salvage pathway transcripts in marine systems [10] suggests that synthesizing DMB or remodeling cobalamin is a common strategy that plays an important role connecting the production of pseudocobalamin to the growth of cobalamin auxotrophs in the low latitude oceans where *Cyanobacteria* are abundant.

We inspected genomes of marine prokaryotes to identity likely cobalamin producers in our study site. We combined bacterial 16S rRNA gene sequence-derived phylogenetic data with current knowledge of corrin biosynthesis genes (including our analysis) across different phylogenetic groups to infer the cobalamin biosynthetic capacity of organisms in the microbial communities at our study site as likely, unknown, or unlikely (Table 3.3). We quantified Thaumarchaeota by q-PCR [81, 82] and calculated their contribution to the microbial population by comparing this value to direct counts of prokaryotic cells determined for each sample. Our analysis at 5 targeted locations in the North Pacific suggested that Thaumarchaeota represent 30-80% of prokaryotes having likely or unknown genetic capacity to synthesize cobalamin in the lower euphotic zone and deeper (Fig. 3.10, Fig. 3.3). High cobalamin contents on a per carbon basis in cultured Thaumarchaeota implicate them as a major source of cobalamin in deeper waters (Table 3.1).

Like *Prochlorococcus*, marine Thaumarchaeota have maintained several genes committed

to the biosynthesis of cobalamin in their small genomes. Detection of MCM, NrdJ, BluB, and cobalamin biosynthesis proteins in the proteome of an oceanic thaumarchaeote with a highly streamlined genome, "Candidatus Nitrosopelagicus brevis", implies that cobalamin is actively produced and used in these microbes [83]. In Thaumarchaeota, the cobalamin-dependent MCM catalyzes a key step in their exceptionally energy-efficient pathway for carbon fixation [84, 85, 86]. The scarcity of cobalamin in the water column (often less than 1 pM [17]) and enzymatic demands like MCM may have necessitated that Thaumarchaeota retain the ability to synthesize cobalamin. Thaumarchaeota likely play a critical role in the microbial loop in the lower euphotic zone and deeper by providing an essential nutrient to cobalamin auxotrophs. In turn the auxotrophs provide substrates that promote Thaumarchaeota growth [87, 88] — most critically the ammonia required by the ammonia-oxidizing Thaumarchaeota [84].

While Thaumarchaeota and select heterotrophic bacteria synthesize cobalamin, undoubtedly benefiting from being their own source of their preferred cofactor, the production of dissimilar cobalamin analogs by marine *Cyanobacteria* is likely a result of their distinct ecological niches, enzymatic demands, and interactions with other cobalamin dependent organisms. Producers of cobalamin and related compounds thus play distinct roles in cobalamin-based microbial interdependencies that sustain primary productivity and shape marine community structure.

3.4 Materials and Methods

Environmental samples for cobalamin and pseudocobalamin, phytoplankton abundance, archaeal gene quantification, prokaryotic cell abundance, and DNA for 16S rRNA sequencing were collected in August 2013 along the historical transect Line P to ocean station PAPA, then following 150 W to the south into the North Pacific subtropical gyre sampling from the surface down to 300 m as shown in Figure 3.2. Culture and environmental samples (particulate material $< 0.2\mu$m) were analyzed for cobalamins, pseudocobalamins, and SAM using an organic solvent extraction [89] paired with LC-MS [54], both modified as described in

Supplemental Methods. Phytoplankton abundance and composition was taken continuously using SeaFlow [55], in addition to discreet samples taken at depth and analyzed by flow cytometry.

We investigated 11 axenic strains of marine prokaryotes for demonstrable evidence of cobalamin or pseudocobalamin production: four strains of *Nitrosopumilus* spp. (SCM1, HCA1, HCE1, and PS0), two strains of *Prochlorococcus* (MED4 and MIT9313), two strains of *Synechococcus* (WH7803 and WH8102), *Sulfitobacter* sp. SA11 [90], *Ruegeria pomeroyi* DSS-3, and *Vibrio fischerii* ES114. We used the model diatom *Thalassiosira pseudonana* to investigate the fate of pseudocobalamin in eukaryotic algae by culturing it under three conditions: cobalamin limited, cobalamin replete, and cobalamin limited with pseudocobalamin. To compare cobalamin- and pseudocobalamin-binding sites in MetH, we gathered MetH amino acid sequences from organisms known to use true cobalamin or pseudocobalamin as their cofactor. We then aligned the sequences and used a known crystal structure [91, 71] to visualize the binding pocket.

To investigate the distribution of cobalamin biosynthesis genes in marine prokaryotes, we used publically available full genomes from the IMG database that phylogenetically represent organisms at our study site and searched for cobalamin biosynthesis genes in these genomes (genomes and functions we used are listed in Dataset S3.1). We used quantitative polymerase chain reaction analysis (q-PCR) to quantify Thaumarchaeota and direct cell counts to quantify total prokaryotes in our environmental samples as previously described [82, 81, 92]. Environmental DNA for 16S rRNA sequencing was extracted, amplified, and sequenced as described in SI Methods. Operational taxonomic units were called using a 97% nucleotide identify threshold and taxonomic inference was based on the SILVA rRNA gene database. This yielded phylogenetic information we combined with the current knowledge of cobalamin-biosynthesis capacity (from literature and our whole genome analysis) to estimate the contribution of archaea to the prokaryotic community with the potential for cobalamin biosynthesis capacity at our sampling sites.

Further details on all aspects of the methods are given in SI Methods. Environmental

data presented and used in calculations are supplied in Dataset S3.2

3.5 Acknowledgements

We thank L. Truxal Carlson, R. Lionheart, A. Weid, R. Morales, and D. Rico for laboratory assistance; L. Truxal Carlson, D. French, R. Horak, A. Nelson, S. Amin, H. van Tol for sampling help in the field; the captain and crew of the R/V Kilo Moana for field sampling support; S. Chisholm, J. Becker, and S. Biller for supplying the axenic *Prochlorococcus* cells; B. Durham for supplying axenic *Synechococcus* cells. This work was supported by NSF award OCE1228770 to A.E.I. and D.A.S.; OCE1046017 to A.E.I., E.V.A., D.A.S., A.H.D., and J.W.M. This work was supported by a grant from the Simons Foundation (SCOPE Award ID 329108, A.E.I., E.V.A., S.C.). W.C-M. was supported by NSF REU award OCE-1046017AM004. K.R.H. was partially supported by an NSF Graduate Research Fellowship and a graduate student fellowship from the University of Washington Program on Climate Change.

3.6 Supplemental Methods

3.6.1 Cultures of representative cobalamin producers

We analyzed four axenic marine Thaumarchaeota strains in this study: *Nitrosopumilus* spp. SCM1, HCA1, HCE1, and PS0. All four strains were cultured in artificial medium as previously described [93, 94]. Triplicate growth experiments with these strains were carried out in dark without shaking at 25°C, except SCM1, which was grown at 30°C. Growth was monitored by NO_2^- production and microscopic cell counts as previously described [87]. Cultures of SCM1 (100 mL) were harvested at five different time points on 0.2 μm Nylon membrane filters (Millipore Co., MA, USA) to obtain samples representing actively growing cultures at different growth phases. Late exponential phase cells of HCE1, HCA1, and PS0 (100 mL) were harvested on 0.2 μm Durapore membrane filters (Millipore Co., MA, USA) with a vacuum filter system. To minimize photodegradation, cell harvesting was performed in the

dark. All samples were stored at -80°C until further analysis. For cell quotas reported we used an estimate of dry weight at 16-20 fg per cell [84], and an estimate of 48% carbon to dry weight to get cobalamin per unit carbon. For reported SCM1 quotas in Table 3.1, we used the exponentially growing cells; quotas at other growth points are given in Dataset S3.2.

We analyzed four representative axenic marine *Cyanobacteria*: *Prochlorococcus* (MED4 and MIT9313), and *Synechococcus* (WH7803 and WH8102). The *Prochlorococcus* strains were both cultured in natural seawater based Pro99 medium [95] prepared with 0.2 μm filtered, autoclaved seawater collected from Vineyard Sound, MA and supplemented with 10 mM sterile sodium bicarbonate. *Prochlorococcus* were grown in acid washed polycarbonate carboys under constant light conditions (40 μmol Q m^{-2} s^{-1} for MED4, 15 μmol Q m^{-2} s^{-1} for MIT9313) at 24 °C in the Chisholm lab (Massachusetts Institute of Technology). *Prochlorococcus* axenicity was verified by testing for contaminant growth in three purity broths (ProAC, ProMM, and MPTB) [96, 97, 98]. *Synechococcus* (WH7803 and WH8102) were cultured in artificial seawater based Pro99 medium [95] prepared with Turks Island Salt Solution and supplemented with 6 mM sterile sodium bicarbonate and 1 mM N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (TAPS) [99]. These strains were grown in combusted borosilicate glass tubes under 16:8 light:dark conditions (20 μmol Q m^{-2} s^{-1}) at 20 °C. We verified that *Synechococcus* were axenic by testing for contaminant growth in liquid MPTB and on 1/2YTSS plates as well as by SYBR Green staining and observation on Influx high-speed cell sorter flow cytometer (BD Biosciences, San Jose, CA). All *Cyanobacteria* cells were collected in mid to late exponential growth phase. For each culture, replicate cell biomass filters (5 for each *Prochlorococcus* strain, 3 for each *Synechococcus* strain) were collected by gently filtering 10 mL (for *Prochlorococcus*) or 24 mL (for *Synechococcus*) of culture onto a 0.2 μm nylon filter, and freezing at -80°C. Cell counts were performed by flow cytometry (for *Prochlorococcus*, GUAVA 8HT). For per carbon quotas for *Cyanobacteria*, we used a diameter range of 0.8-1.2 μm for *Prochlorococcus* MIT9313, 0.5-0.7 μm for MED4 [100] and diameters measured by flow cytometry for the *Synechococcus* with a conservative estimate of 325 fg C μm^{-3}[101].

Heterotrophic bacterial strains used in this study were *Sulfitobacter* strain SA11 [90], *Ruegeria pomeroyi* DSS-3, and *Vibrio fischerii*. All strains were cultured in an artificial seawater-tryptone media (30 g Instant Ocean salts, 5 g tryptone in 1 liter MQ water). For cobalamin and pseudocobalamin analysis, cultures were grown overnight at room temperature, with shaking at 250 rpm, using standard aseptic techniques. Overnight cultures were harvested by centrifugation (30 minutes at 4100 rcf, 4°C). Cell density was determined by enumerating plated culture forming units (CFUs). *R. pomeroyi* and *Sulfitobacter* SA11 were harvested in mid-late exponential growth phase; *V. fischerii* were harvested during stationary phase. For cell quotas, we used a previously measured 142 fg C per cell for *R. pomeroyi* [102]. We do not yet have a cellular carbon estimate for *Sulfitobacter* SA11, so we used a wide estimate of 32-292 fg C per cell previously measured from 13 species of marine Proteobacteria [102].

3.6.2 Diatom growth experiment

Cultures of *T. pseudonana* CCMP 1335 were maintained on previously described defined media [103] with f/2 concentrations of nutrients, with the exception of cobalamin, which was added to a final concentration of 1 pM OH-cobalamin (a concentration previously shown to limit the growth rate of *T. pseudonana* [46]). Although our diatom cultures are not axenic, *T. pseudonana* could not be maintained on cobalamin-free media (with or without pseudocobalamin), so any bacterial production of or conversion to cobalamin was not sufficient to disrupt our experiment. After observing the high Ado-cobalamin content of our experimental cells, we compared them to an axenic population (axenicity monitored by marine purity test broth [98]), and observed the same pattern in which Ado-cobalamin > OH-cobalamin > Me-cobalamin.

To assess the bioavailability of pseudocobalamin we used a single culture to inoculate a set of cultures with three different treatments: Control (1 pM OH-cobalamin, same as maintained conditions), +Cobalamin (200 pM OH-cobalamin), and +Pseudocobalamin (1 pM OH-cobalamin with 200 pM OH-pseudocobalamin). Triplicates of each treatment were

grown in 50 mL borosilicate culture tubes with 35 mL media at 20°C with a 12 hour light dark cycle. Cultures were monitored by fluorescence and harvested by gentle vacuum filtration onto 0.2 μm nylon filter at mid exponential phase. We took 1 mL samples (fixed with 1% formaldehyde) at four time points throughout the growth curve for cell counts and used a Beckman Coulter Z2 Particle Count and Size Analyzer (Beckman Coulter, Fullerton CA, USA) to measure diatom densities and diameters. To estimate carbon content of *T. pseudonana* in our experiment, we used the average observed diameter (5.36 μm) to calculate carbon using pg C = $0.109\text{V}^{0.991}$ [104, 56].

3.6.3 Genomic and phylogenetic analysis of cobalamin biosynthetic capacity

We analyzed full genomes from the IMG database (https://img.jgi.doe.gov) from phylogenetic groups that encompassed >99.9% of the Bacterial 16S reads from our environmental samples to develop a systematic inference of cobalamin synthesis capacity. We only included bacterial genomes of high quality (denoted finished or published) and removed all organisms noted as engineered. We included finished or published Thaumarchaeota genomes (including lower quality sequences), excluding Aigarchaeota *Caldiarchaeum* spp., whose phylogenetic placement is still unknown but likely outside of Thaumarchaeota [105]. This resulted in 3,410 full genomes, which we split into 6 phylogenetic groups: Flavobacteriia, Alphaproteobacteria, Gammaproteobacteria, Cyanobacteria, Thaumarchaeota, and a final group made up of 265 genomes from mixed Bacteria groups with poor phylogenetic coverage in the database such as Chlorobi and Deltaproteobacteria (see Dataset S3.1 for phylogenetic groups used, functional enzymes searched and lists of surveyed genomes in each subgroup). Next, we searched these genomes for annotations of functional enzymes related to cobalamin synthesis, and clustered them according to the annotated presence or absence of the genes of interest. Each phylogenetic group clustered easily into two groups: genomes with many genes related to the central corrin ring synthesis, and genomes with few genes related to the central corrin ring synthesis.

Beyond relying on annotations, we also searched for potential homologs of genes re-

lated to DMB synthesis and activation in Thaumarchaeota and Cyanobacteria genomes. No BLASTp [106, 107] hits below an E-value threshold of 10^{-5} with bit score greater than 50 could be identified as *cobT* in the 27 surveyed Thaumarchaeota genomes or 255 *Cyanobacteria*l genomes. Using the same thresholds, we were unable to identify any bluB homologs in any *Prochlorococcus* or *Synechococcus* genomes. The newly described anaerobic pathway for synthesizing DMB is not yet annotated so we searched *Prochlorococcus* and *Synechococcus* genomes for evidence of sequence similarity to *bzaC*, *bzaD*, and *bzaE* [61]. We were unable to find any BLASTp hits with an E-value threshold of 10^{-5} with bit score greater than 50 for *bzaC*. Although there were several BLASTp hits with a lower E-value for *bzaD* and *bzaE*, the hits never yielded an alignment that covered greater than 50% coverage of either protein with greater than 30% identical sequences. Given the low sequence similarity and the observation in the obligate anaerobe *Eubacterium limosum* bzaD and bzaE depend on bzaC to synthesize DMB [61], we concluded that this was not persuasive evidence for DMB synthesis capacity in *Prochlorococcus* or *Synechococcus*.

3.6.4 MetH alignment and comparison

We gathered MetH amino acid sequences from the *Cyanobacteria* genomes in our surveyed full genomes. We added in MetH sequences from organisms known to use true cobalamin as their cofactor: *E. coli*, *Pseudomonas*, *T. pseudonana*, and *Homo sapiens* [51, 76, 108]. Sequences were aligned using MAFFT v6.864 (http://www.genome.jp/tools/mafft/) [109] and trimmed to include only the cobalamin-binding motif [70, 71]. A selection of aligned sequences is shown in Figure 3.7. Using the cobalamin-binding motif of *Prochlorococcus* MIT9313, we used SWISS-MODEL [110] to visualize a rough proposed structure of MetH from *prochlorococcus* MIT9313 based on the same enzyme in *E. coli* [91] (37.6% identical in cobalamin binding domain, DOI: 10.2210/pdb3bul/pdb). For protein graphics, we used the UCSF Chimera Package [111] (Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, http://www.cgl.ucsf.edu/chimera) to edit the crystal structure to include pseudocobalamin in place of cobalamin. Using the built in

parameters on the UCSF Chimera Package, we elucidated possible hydrogen bonds between adenine and the protein as shown in Figure 3.6.

3.6.5 Field sampling

At sea, samples for particulate cobalamin and pseudocobalamin, archaeal gene quantification, prokaryotic cell abundance, and phytoplankton abundance were collected using Niskin bottles mounted onto a rosette system equipped with a conductivity-temperature-depth sensor package (CTD, Seabird SBE). Cobalamin and pseudocobalamin sampling was performed in a single cast at approximately 11 am local time at each location; fluorescence profiles are reported as uncorrected relative fluorescence units (RFU) for qualitative purposes only. Other reported measurements were taken no more than 7 hours earlier. Samples for DNA were collected using in-situ McLane large volume pumps collecting particles from the 0.2 – 1.6 μm size range by using a GF/A prefilter (Whatman) and a 0.2 μm Supor (Pall) filter.

3.6.6 Cobalamin, pseudocobalamin, and SAM extraction and quantification

At sea, 2 L samples for particulate analysis of cobalamins and SAM were filtered onto 0.2 μm Nylon membrane filters and frozen in combusted Al foil at -20°C (preliminary tests showed no difference between samples frozen at -20°C or -80°C). Culture samples were processed in the same way as environmental samples, but with smaller volumes as noted. For the archaea and bacteria cultures, we processed and analyzed blank media from all cultures according to previously described methods [54] and verified that there was no detectable cobalamin or pseudocobalamin in the starting media.

To extract metabolites from cellular matter in culture and environmental samples, we used an organic solvent extraction which has previously been used for marine environmental and culture samples [89, 112] paired with physical bead beating. We placed filters into bead beating tubes containing 100 and 400μm beads of equal volume, added 1 mL of cold (-20°C) acidic acetonitrile/methanol/water mixture (40:40:20 with 0.1% formic acid)[89]. Over the course of 20 minutes, we bead beat the samples for 40 seconds 3 times and kept

the samples in a -20°C freezer when possible. After centrifugation, the supernatant was removed and the filter was rinsed once with the 40:40:20 solvent and twice with methanol, centrifuging and combining the supernatants after each step. The supernatant and rinses were dried down under clean N_2 or under vacuum with minimal heat (less than 40°C). Due to the light-sensitive nature of these compounds, extractions were performed under dimmed lights and samples were protected from light exposure whenever possible. Without a consensus standard, it is difficult to ascertain if this is the most effective extraction, but our reproducibility suggests it is a consistent technique that yields comparable data between samples.

Samples were reconstituted with an internal standard mixture and quantified by standard addition curves using ultra-high pressure liquid chromatography coupled to a triple quadrupole mass spectrometer (Waters I-Class Aquity UPLC coupled to a Waters Xevo TQS) as previously described [54] but altered to include four additional transitions corresponding to the CN- OH- Me- and Ado-pseudocobalamin as well as SAM (Table 3.4). Internal standard concentrations were identical to previous work on the same LC-MS system [54]. Injection volumes were 40 μL for all environmental samples. Total cobalamin or pseudocobalmin values presented in all figures are the summed total of all detected α ligands for the corresponding cobalamin analog. Transitions, collision energies and cone voltages for the pseudocobalamins were optimized using a quantified pseudocobalamin stock from a recognized pseudocobalamin source, obtained and quantified as described in the following section. These quantified pseudocobalamin stocks were also used for the standard addition curves for quantification. We used a standard of SAM to optimize transitions, collision energies and cone voltages for this compound. See Figure 3.11 for example chromatograms.

Limits of detection in LC/MS analyses can vary among sample matrices (affecting background or analyte ion suppression), injection volume, chromatography quality, and instrument performance. Instead of using a pre-set concentration as our limit of detection, we assessed data quality on a batch-by-batch basis (i.e. environmental samples and culture samples were run at separate times and therefore treated as separate batches). Each peak

is subject to the following criteria during our quality control: peaks must be at the same retention time (± 0.2 min) as a standard, at least two daughters must be present, signal to noise ratio greater than five, and the integrated peak area at least five times greater than any peak found in the blank in the appropriate retention time window.

3.6.7 Preparation of pseudocobalamin stocks

Pseudocobalamin is not available for purchase in its most ecologically relevant forms (with β ligands of OH-, Me-, or Ado-). In order to quantify the compound, we concentrated and quantified pseudocobalamins from a known pseudocobalamin source: Spirulina powder (GNC SuperFoods Spirulina) [113]. First we extracted pseudocobalamin from approximately 6 grams of Spirulina powder using the extraction method described in this publication. The extract was dried and reconstituted in 20 mL Milli-Q water (>18.2 MΩ cm, MQ). A clean up step aimed at obtaining a pseudocobalamin-enriched fraction of the extract was performed by loading the aqueous mixture onto a conditioned C18 solid-phase extraction column (Waters, 35 mL capacity, 10 g resin). The column was rinsed with 40 mL MQ water, 20 mL 5% MeOH in MQ, and 20 mL 30% MeOH in MQ, before eluting the pseudocobalamin-containing fraction with 60 mL 50% MeOH in MQ.

Further purification of the pseudocobalamin-containing fraction was achieved by a two-step high-pressure liquid chromatography method with fraction collection (HPLC, Agilent 1100). For purification, we carried out 50 μl injections onto a C18 column (4.6x150 mm Agilent Eclipse). Initial conditions were 92% solvent A (water with 20 mM ammonium formate and 0.1% formic acid) and 8% solvent B (acetonitrile). The gradient ramped from 8% to 25% Solvent B over 25 minutes with a flow rate of 2 mL min^{-1}. Fractions were collected in 5 mL aliquots and each aliquot was analyzed for Me- and Ado-pseudocobalamin by UPLC MS/MS as described for sample quantification in this text (Waters Aquity UPLC coupled to Xevo TQS mass spectrometer). Pseudocobalamin-containing fractions were dried down and subjected to a second round of HPLC purification consisting of an isocratic HPLC run of 8% B and 13% B to isolate Ado- and Me-pseudocobalamin, respectively (modeled after

previous purifications [114]). Fractions containing each analyte were collected and pooled. At this point, our fractions are not pure (bulk UV-VIS does not show characteristic corrin ring peaks), but UPLC MS/MS analysis (the most sensitive analysis available) confirmed that fractions contained only the pseudocobalamin of interest and no detectable cobalamin forms, imperative to our diatom experiment. To make OH-pseudocobalamin, we exposed glass vials containing an aqueous solution of the Me-pseudocobalamin containing fraction to full spectrum light for 10 minutes [20]. Throughout the preparation of pseudocobalamin stocks, we took great care to protect the quantified stocks from light to avoid degradation. To make CN-pseudocobalamin, we exposed this fraction to 50:50 acetonitrile:water in full spectrum light for 10 minutes. This produced a mixture of both CN-pseudocobalamin and OH-pseudocobalamin.

3.6.8 Quantification of pseudocobalamin stocks

OH- and Me-pseudocobalamin were analyzed spectrophotometrically by HPLC- diode array detection (DAD) using 362 nm for OH- and 259 nm for Me-pseudocobalamin detection. To our knowledge, no UV-VIS data exist for Me- or OH- pseudocobalamin, so we used the observed extinction coefficients of Me- and OH- cobalamins (39,700 and 74,600 cm^{-1} mol^{-1} respectively) to quantify the corresponding pseudocobalamins as previously done for cobalamin analogs with CN β ligands [114], with chromatograms shown in Figure 3.12 . All data are therefore presented as cobalamin equivalents as noted in figure legends. Although cobalamin and pseudocobalamin likely have different extinction coefficients, this approach provides a first attempt to compare concentrations of these analogs. Furthermore, if UV-VIS based extinction coefficients become available in the future for Me- or OH-pseudocobalamin, a simple conversion at the wavelengths measured will provide true pseudocobalamin concentrations. Although we were able to confirm the structure and continue to monitor for Ado- and CN-pseudocobalamin by UPLC MS/MS, we were unable to purify enough Ado- or CN-pseudocobalamin to quantify by DAD due to the much lower concentration in Spirulina; we were only able to detect these forms in the *Prochlorococcus* and *Synechococcus* cultures

(where the peaks were 50-100 times smaller than the OH- or Me-pseudocobalamin peaks) and in the diatom experiment. The HPLC-DAD mobile phases and column used for quantification of pseudocobalamin standards mobile phases were the same as for purification but with a flow rate of 0.3 mL min^{-1} and a gradient from 9-32% B over 30 minutes.

3.6.9 Confirmation of pseudocobalamin structure

To confirm the presence of OH-, Me-, CN-, and Ado-pseudocobalamin, we obtained accurate mass data for our analytes using an LC-MS system consisting of a Waters Acquity UPLC coupled to a quadrupole time of flight mass spectrometer (Waters Xevo G2-S QTOF) in positive ion mode. We used observed accurate masses, fragmentation patterns, and retention times to confirm compound identities. Electrospray ionization source conditions for the QTOF method were the same as those used during sample quantification, as previously described [54]. We used MS/MS conditions and extracted the predicted accurate mass from associated spectra of each parent without collision energy (doubly charged; OH-, 659.7750; Me-, 667.7907; Ado-, 784.8220; CN-, 673.2805) and corresponding daughters (described in Table 3.4: adenine, 136.0623; adenine with sugar and phosphate, 348.0709; loss of β ligand, 660.2789 m/z). For each compound, we detected the doubly charged parent and each of its daughters at the same retention time with a 0.01 Dalton mass tolerance. Retention times matched those shown in Figure 3.11. As expected from previous work with CN-pseudocobalamin [61, 114], each of the pseudocobalamins eluted earlier than its corresponding cobalamin (same β ligand) in both the HPLC (purification and stock quantification) and UPLC (for stock confirmation and sample quantification, Figure 3.11, Figure 3.12) analyses.

3.6.10 Phytoplankton abundance and composition

While at sea, continuous measurements of picophytoplankton abundance and cell size were made using SeaFlow [55]. The instrument was equipped with a 457 nm 300 mW laser (Melles Griot). Forward light scatter (a proxy for cell size), orange and red fluorescence

were collected using a 457-50 bandpass filter, 572-27 bandpass filter and 692-40 band-pass filter, respectively. Seawater was prefiltered through a 100 μm stainless steel mesh (to eliminate large particles) prior to analysis. The flow rate of the water stream was set at 15 mL min^{-1} through a 200 μm nozzle for field and laboratory experiments; this corresponded to an analysis rate of 15 μL min^{-1} by the instrument [55]. A programmable syringe pump (Cavro XP3000, Hamilton Company) continuously injected fluorescent microspheres (1 μm, Polysciences) into the water stream as an internal standard. Data files were created every three minutes. Data were analyzed using the R package Popcycle version 0.2, which uses a SQLite relational database management system to retrieve flow cytometry data (https://github.com/uwescience/popcycle).

Discrete 2 mL samples for cytometry analysis were collected at different depths and fixed to a final concentration of 1% paraformaldehyde and 0.01% glutaraldehyde, frozen in liquid nitrogen, and analyzed onshore by an Influx high-speed cell sorter (BD Biosciences, San Jose, CA). A sequential bivariate manual gating scheme was used to cluster the different phytoplankton populations in both continuous and discrete samples. Phycoerythrin (PE)-positive cells (*Synechococcus*) were classified based on orange fluorescence. Light scattering and red fluorescence were used to distinguish *Prochlorococcus* and picoeukaryotes.

To estimate carbon content of environmental *Prochlorococcus* and *Synechococcus* populations, we first estimated *Prochlorococcus* and *Synechococcus* cell volumes using an empirical relationship between light scatter measured by the flow cytometer and cell volume measured by a Coulter Counter for different exponentially growing phytoplankton cultures of cell sizes with diameters ranging from 1 to 10 μm [115], and again used a conservative estimate of 325 fg C μm^{-3} [101]. The carbon per cell was multiplied by cell abundance to obtain the carbon content (μg L^{-1}) and averaged and binned into 0.5 by 0.5 degree bins.

3.6.11 q-PCR analysis of amoA

The archaeal ammonia monooxygenase subunit A (amoA) gene abundances were estimated by quantitative polymerase chain reaction analysis (q-PCR). Environmental samples (1 L) for

q-PCR were filtered on 0.2 μm Sterivex-GP filters (Millipore Co., Billerica, MA, USA) with a peristaltic pump, immediately frozen in liquid nitrogen and stored in -80°C freezer until analysis. DNA was extracted in the laboratory using the modified phenol-chloroform protocol and quantified using the ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) as described previously [82, 81, 116]. q-PCR targeting archaeal amoA gene copy number was performed using the Roche LightCycler FastStart DNA Master SYBR Green I kit and Roche LightCycler capillary system with the cycling conditions described previously [82, 81]. The primer sets used for targeting archaeal amoA genes were CrenAmoAQ-F and CrenAmoAModR [117]. The amplification efficiencies for q-PCR reactions were in the range of 96.6% to 97.4%.

3.6.12 Prokaryotic cell counts

Environmental water samples for total prokaryotic cell counts (50 mL) were fixed with 2% glutaraldehyde and filtered (1 ml) onto a 0.02 μm Anodisc 25 filter (Whatman, UK). Cells were stained with Moviol-SybrGreen mix and counted with an Olympus BHS/BHT system microscope on board [92]. At least 20 random fields of view with 10-100 SybrGreen stained cells per field were counted.

3.6.13 Environmental 16S sequencing and analysis

DNA extraction methods were adapted from the spin-column protocol for the Qiagen DNeasy Blood & Tissue Kit. A section from each 0.2 μm Supor filter (collected with large volume in situ McClane pumps) was placed in a falcon tube, submerged in 4 mL sucrose lysis buffer, and flash frozen in liquid nitrogen. After thawing, 200 μL of 1 mg/mL lysozyme was added, and the tube was incubated at 4°C for one hour, shaking every 10 minutes. Next, 930 μL 10% SDS, and 500 μL proteinase K were added, and the tube was placed in a shaking incubator at 55°C for two hours. After the incubation, 5 mL AL buffer and 5 mL 100% ethanol were added to the tube, and the solution was centrifuged through three Mini Spin Columns for one minute at 9000 rpm. Wash buffers AW1 and AW2 were then applied as described in the

Qiagen Blood & Tissue Kit Protocol. The DNA was eluted with 30 μL water, which was added to the spin column membrane for two minutes at room temperature, and centrifuged one minute at full speed into microfuge tubes. The elution step was repeated with another 30 μL water to yield 60 μL of eluent.

Genomic DNA was amplified for 16S rRNA genes with previously described primer sets [118], independently followed by sample specific barcoding using a microfluidic array (Fluidigm Access Array TM, Fluidigm INC Palo-Alto California). Barcoded samples were quantified using pico-green, diluted to equal-molar concentrations, pooled, and sequenced in paired-end mode on an Illumina MiSeq. Fastq files were trimmed (200 bp) and quality filtered in QIIME with default setting adjustments according to previous methods [119]. Quality sequences were then analyzed using the usearch software as suggested in 'UPARSE [120]. Briefly, operational taxonomic units (OTUs) were called using a 97% nucleotide identity threshold, and dereplicated sequences mapped to OTUs using "usearch_global" with the default settings.

Taxonomic inference was used based on comparison of OTU sequences to the SILVA rRNA gene database (version 119) using nucleotide basic-local alignment searches (i.e. BLASTN and BLAST 2.2.28+) [121]. Sample matrices were converted to biom format and taxonomic meta-data was incorporated from SILVA and summarized at varying resolutions using summarize_taxa.py in QIIME. Data from this study can be found in the European Nucleotide Archive under Accession Number PRJEB10943 (ERP012248) (Project Title 'Variable Influence of light on the activity of Thaumarchaea').

3.6.14 Estimation of Thaumarchaeota contribution to the microbial community with potential for cobalamin synthesis

We summarized the potential for cobalamin synthesis in important marine prokaryotes by using our full genome analysis (grouped taxonomically) and previous literature [49, 47, 122, 123, 17], results summarized in Table 3.3. Although inferences from phylogeny are most valuable when done using specific phylogenetic classification, many groups have few high

quality genomes available (i.e. the Actionobacter and Planctomycetes phyla). For groups with few genomes available, we used any genomes available at the wider classifications (Phylum and Class). When possible, we collected genomes at the Order level. Groups were generally considered unlikely, likely, or unknown to have cobalamin biosynthesis capacity if less than 10%, greater than 80%, or between 10-80% of the genomes had corrin biosynthetic capacity, respectively. Most of the genomes from the Pseudomonadales order with cobalamin biosynthetic capacity were from one species (Dataset S3.1), which is not marine. Of the marine genomes available (6 genomes), none have the biosynthetic capacity [17], thus we categorized them as unknown, rather than likely cobalamin producers. Similarly, no high quality genomes of the abundant SAR86 Clade are available, but other studies have shown that they are unlikely to have the genetic capacity for cobalamin biosynthesis [123], so we separated these organisms from the remaining Oceanospirillales.

We then applied the categories to our genomic 16S rRNA-derived taxonomy (after filtering out mitochondria and chloroplast reads). This step assumes that the categories of cobalamin biosynthesis capacity derived from high quality genomes in the IMG database reflect the microbial population present in our study site and yields the percentage of bacteria with the likely, unlikely, and unknown genomic potential for cobalamin biosynthesis. Using 16S rRNA-based abundances imperfectly describes bacterial community composition due to variable copy numbers of 16S among bacteria [124]. In the ocean, this can lead to an overestimation of taxa with high 16S rRNA copy numbers like Alteromonadales and a slight underestimation of taxa with low 16S rRNA copy numbers like Rhodobacterales [125].

To investigate the Thaumarchaeal contribution to cobalamin biosynthetic capacity, we first used the q-PCR amoA data to calculate an abundance of ammonia oxidizing archaea in the water column [82, 81], since there is one copy of amoA per cell in all characterized Thaumarchaeota [126]. We assumed that the abundance of ammonia oxidizing archaea was equal to the abundance of archaeal cells capable of cobalamin biosynthesis (marine group II euryarchaea are unlikely to have cobalamin biosynthesis capacity [127, 49] and are a small percentage of total prokaryotes [128].

We used the difference between the prokaryote cell counts and q-PCR amoA gene counts to estimate the bacterial abundance at each location, this may be a slight underestimation of bacterial contribution in the surface ocean as other archaeal groups are likely present. For the Cyanobacterial contribution, we multiplied the contribution of 16S reads attributed to *Cyanobacteria* (but not chloroplasts) by the bacterial cell abundance. In two instances, we had both 16S data and flow cytometry data (St. 4, 70m and St 2, 80m); Cyanobacterial contributions to the microbial population were low (0.5-3%) by direct observations and 16S data. For all other bacterial contributions, we took the percentage of bacteria with the different genomic potentials (likely, unlikely, unknown) for cobalamin biosynthesis and multiplied it by our cellular abundance. All data used in these calculations are reported in Dataset S3.2.

3.7 Figures

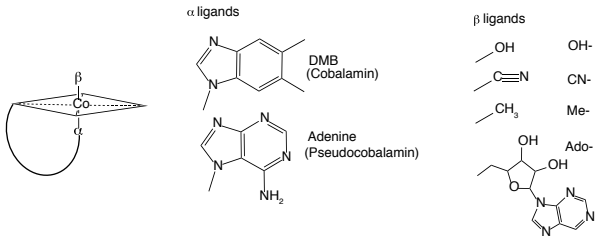

Figure 3.1: General form of cobalamin analogs. Schematic of conserved corrin ring with various upper (β) and lower (α) ligands. Structures of cobalamin analogs monitored in this study (8 total) are shown.

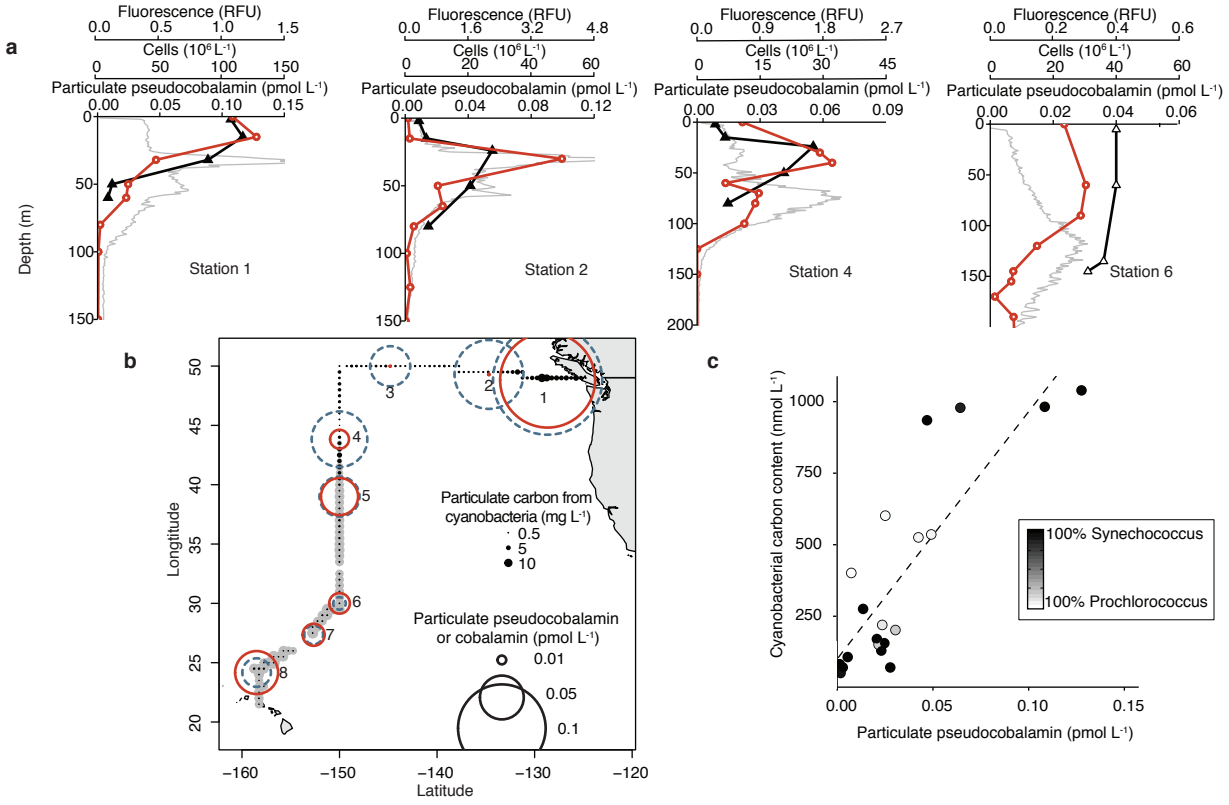

Figure 3.2: Pseudocobalmin and *Cyanobacteria* co-occurrence. (a) Depth profiles of particulate pseudocobalamin (orange circles), in situ chlorophyll a (grey) and cell abundance of *Synechococcus* (closed triangles) or *Prochlorococcus* (open triangles), whichever was the dominant *Cyanobacteria* at each station. b) Location of sampling with surface concentration of particulate cobalamin (dashed blue circles) and pseudocobalamin (orange circles). Closed circles are *Cyanobacteria*l (*Synechococcus*, black; *Prochlorococcus*, grey) carbon content calculated from cell abundance and size estimates (at 5 m via SeaFlow [55]). c) Correlation between environmental pseudocobalamin and calculated *Cyanobacteria* carbon content ($n = 20$, $R^2 = 0.71$, $p < 0.001$). Cobalamin and pseudocobalamin concentrations are summed values of the detected β ligands for these compounds (Ado-, Me-, and OH- for cobalamin; Me-, and OH- for pseudocobalamin), contributions of each beta ligand provided in Dataset S3.2. Pseudocobalamin concentrations are presented in cobalamin equivalents (see Supplemental Methods).

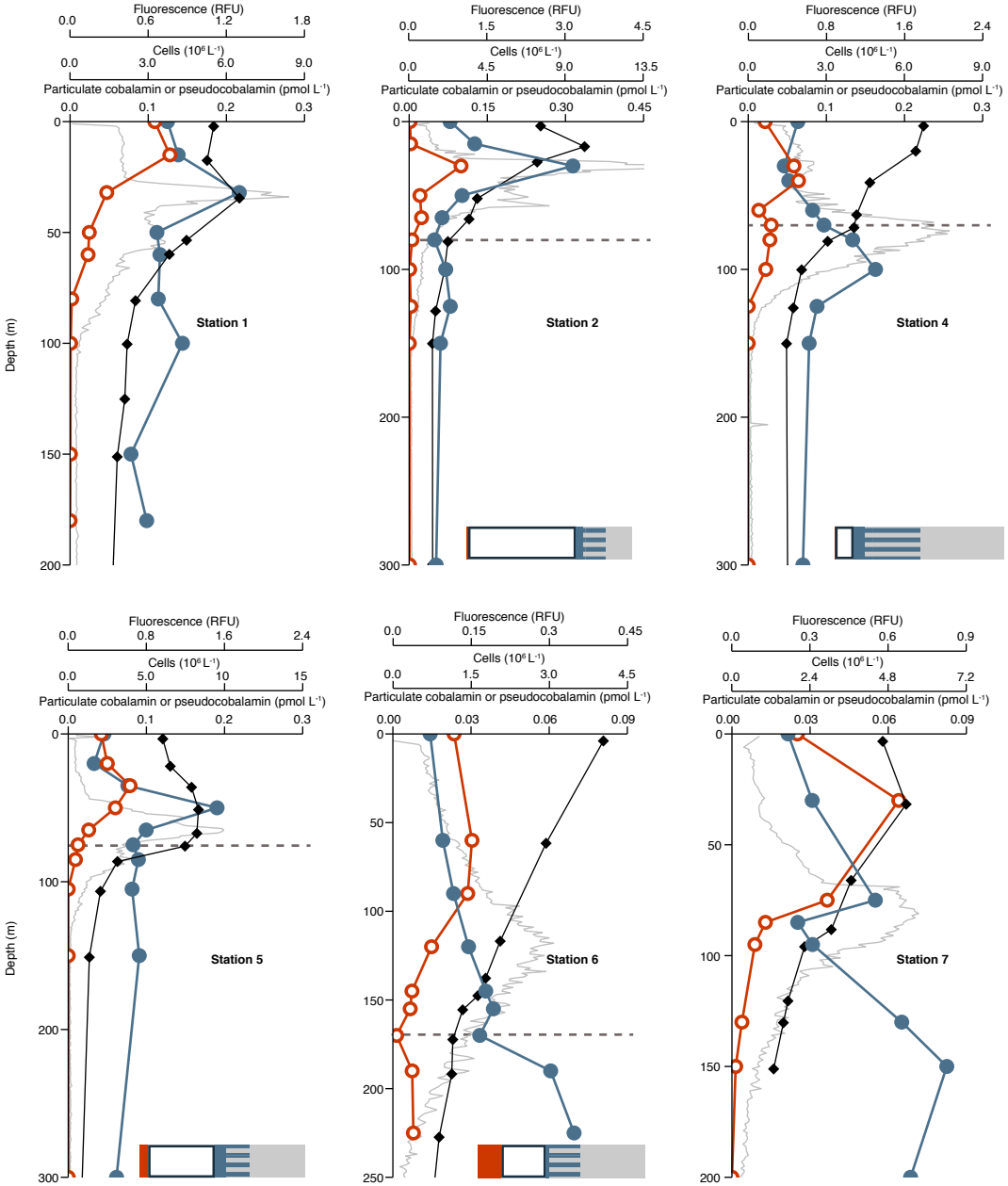

Figure 3.3: Caption on following page.

Figure 3.3. *Continued from previous page.* Pseudocobalamin and cobalamin producers at six additional stations. Cobalamin (closed blue circles), pseudocobalamin (orange open circles), prokaryote cell abundance (black diamonds), and in situ fluorescence (grey). Bar graphs with the percent of cells with the following predicted cobalamin strategies: unlikely producers (grey), bacteria with unknown cobalamin biosynthesis capacity (grey/blue striped), bacteria that likely possess the potential to produce cobalamin (blue), Thaumarchaeota (likely cobalamin producers, white), and *Cyanobacteria* (likely pseudocobalamin producers, orange) at the corresponding depths. Cobalamin and pseudocobalamin concentrations are summed values of the detected beta ligands as in Fig. 3.2. Pseudocobalamin concentrations are presented in cobalamin equivalents (see Supplemental Methods). Station numbers are as shown in Figure 3.2; dashed lines indicate the depth at which we analyzed prokaryotic diversity by 16S rRNA genes.

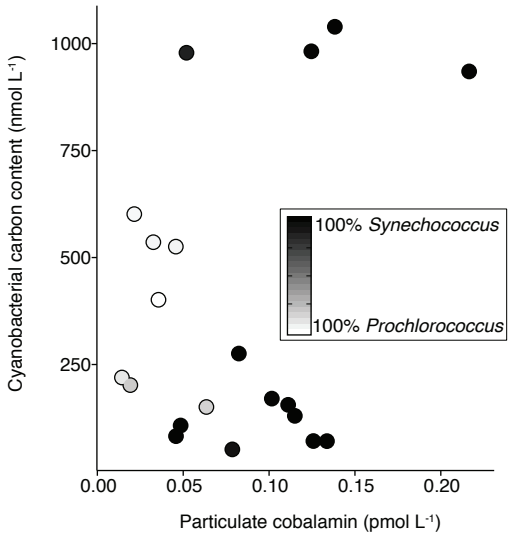

Figure 3.4: Environmental cobalamin and *Cyanobacteria* carbon. No significant correlation exists between calculated carbon from *Cyanobacteria* and cobalamin. Shade of circle is the source of the calculated *Cyanobacteria* carbon (from 100% *Synechococcus* to 100% *Prochlorococcus*). Cobalamin concentrations are summed values of the detected beta ligands for these compounds as in Figure 3.3.

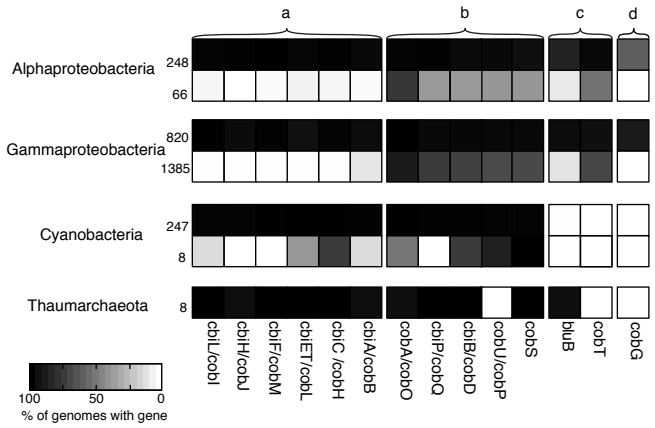

Figure 3.5: Presence or absence of annotated cobalamin biosynthesis genes in full-genome representatives. Four major groups of microbes contributing cobalamin biosynthesis genes in marine surface waters [49] are shown; each group is split into potential cobalamin producers (top row), and non-cobalamin producers (bottom row); with the number of genomes in each group. All of the high quality Thaumarchaeota genomes are potential cobalamin producers. The shade of each box indicates the percent of genomes with that gene. Genes are grouped as follows: a) corrin ring biosynthesis genes that are common to both O_2-dependent and O_2-independent pathways in bacteria and archaea, b) final synthesis and repair genes, c) genes for DMB synthesis and activation, d) O_2-dependent cobG in the O_2-dependent corrin ring biosynthesis pathway. In Archaea, cobY is used in place of cobU/cobP [129, 49]. For the list of genomes in each group, see Dataset S3.1, summarized in Table 3.1 .

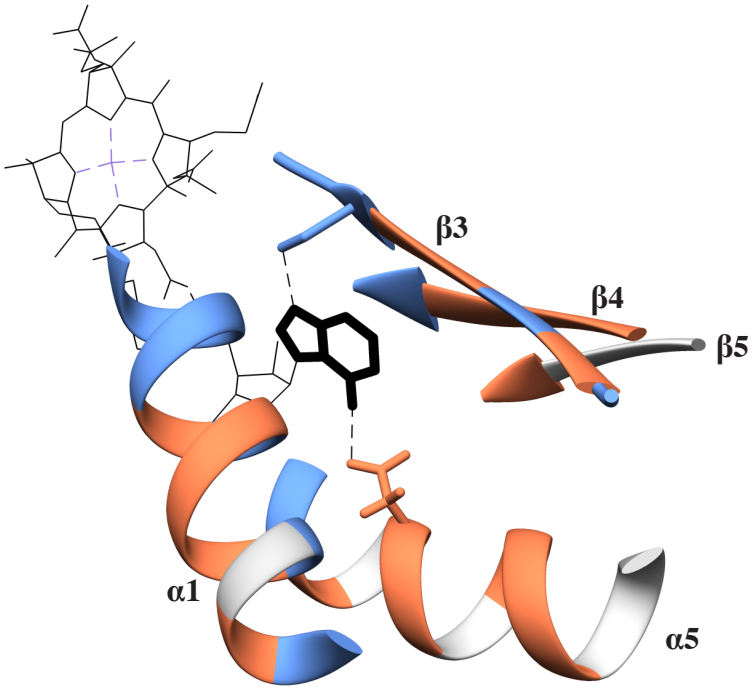

Figure 3.6: Structure shown is a representation of MetH in *Prochlorococcus* MIT9313 based on the crystal structure of the same enzyme in *E. coli* [91], with pseudocobalamin (in black, adenine ligand bolded, cobalt in violet) in place of cobalamin. Colors correspond to Figure 3.7. Dashed lines are possible hydrogen bonds between the adenine base and the enzyme; a well conserved serine residue previously suggested to hydrogen bond to DMB [70] (and likely adenine) on β3, and an aspartic acid residue found in all *Cyanobacteria* that may hydrogen bind to adenine on α5

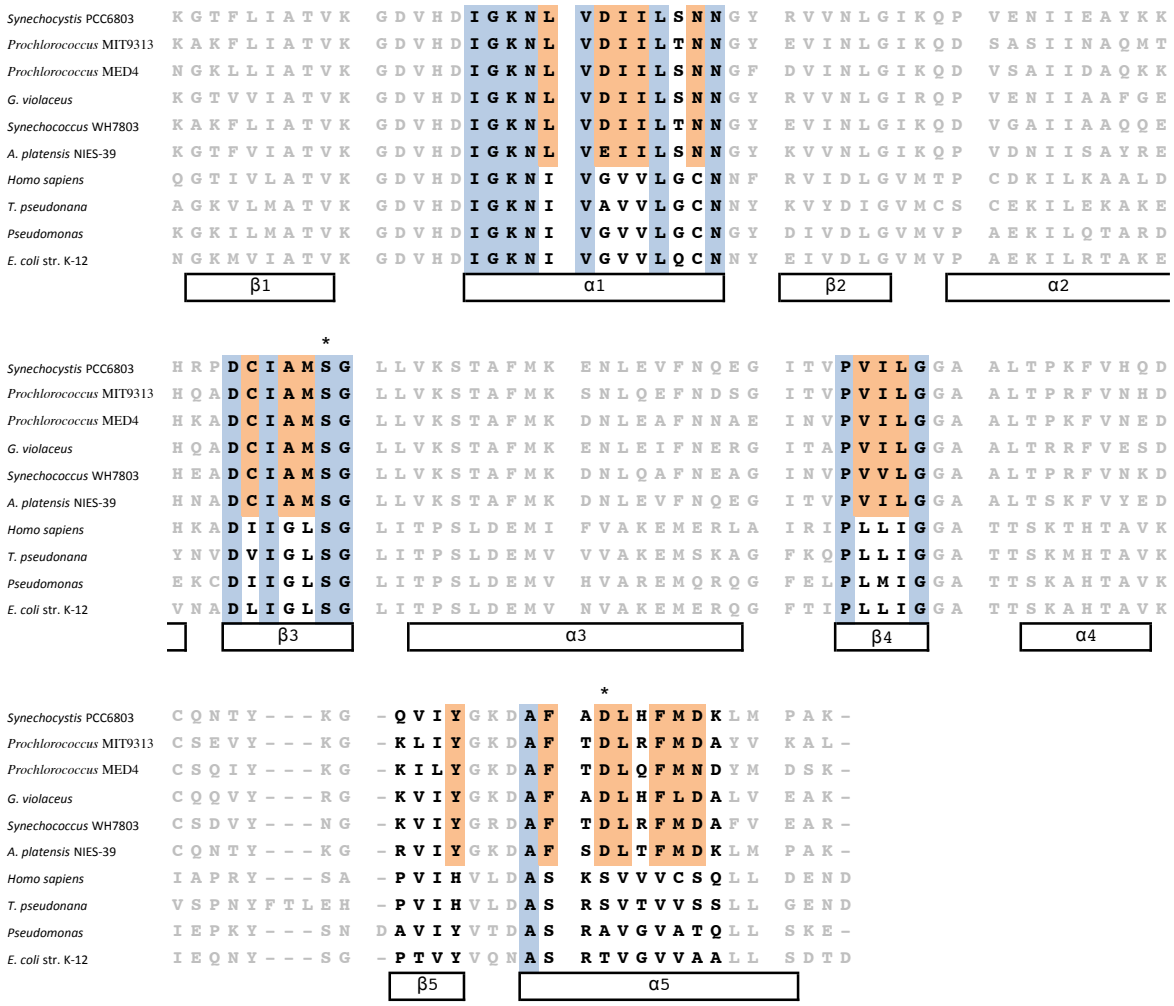

Figure 3.7: Alignment of cobalamin-binding motif in MetH from 6 *Cyanobacteria* (*Synechocystis* sp. PCC6803, *Prochlorococcus* spp. MIT9313 and MED4, *Gloeobacter violaceus* PCC 7421, *Synechococcus* WH7803, and *Arthrospira platensis* NIES-39) and 4 organisms known to use true cobalamin (*Homo sapiens*, *T. pseudonana*, *Pseudomonas*, and *E. coli*). MetH in *A. platensis* NIES-39 has been demonstrated to have a higher affinity to pseudocobalamin than cobalamin. Black amino acids are part of α-helixes or β-sheets in the pocket housing the base of cobalamin [70]; within that pocket, blue highlighted residues are conserved in both *Cyanobacteria* and the four cobalamin-users, and orange residues are conserved in >90% of surveyed *Cyanobacteria* genomes but are different than the cobalamin-using organisms. These colors are used in Figure 3.6 to better visualize the base-housing pocket. *amino acids that have potential to hydrogen-bond to pseudocobalamin's base, adenine.

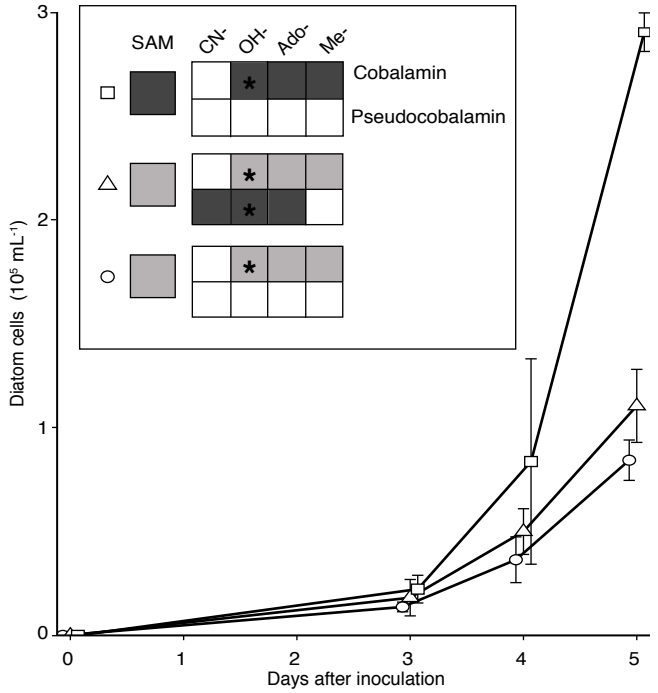

Figure 3.8: Growth of the diatom *T. pseudonana* with pseudocobalamin. (a) Growth of diatoms under Control (1 pM cobalamin, circle), +Pseudocobalamin (1 pM cobalamin with 200 pM pseudocobalamin, triangle), and +Cobalamin (200 pM cobalamin, square) treatments over time; error bars, s.d. from three replicates. Inset shows the final SAM, cobalamin, and pseudocobalamin cellular contents as compared to Control treatment on a per cell basis; not detected (white), similar range to control (light grey), or statistically higher than control (students unpaired t-test, $p < 0.05$, dark grey). Black stars indicate what forms of cobalamin or pseudocobalamin were added to each treatment. See Fig 3.9 for absolute values of cobalamins and SAM.

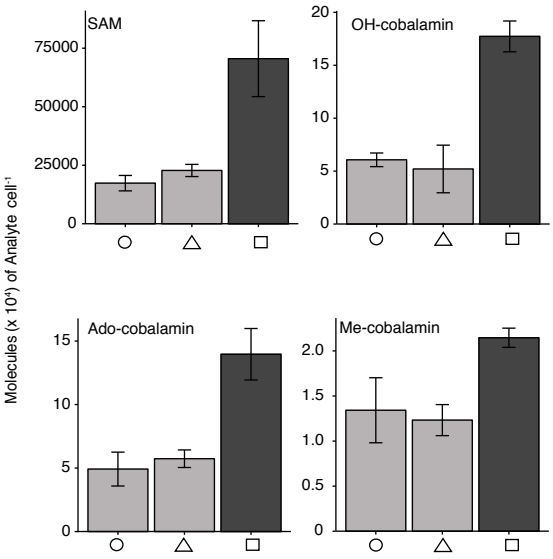

Figure 3.9: Cellular contents of cobalamin and SAM in diatom experiment. Molecules per cell of SAM and OH-, Ado- , and Me-cobalamin. Colors and symbols are as in Figure 3.8.

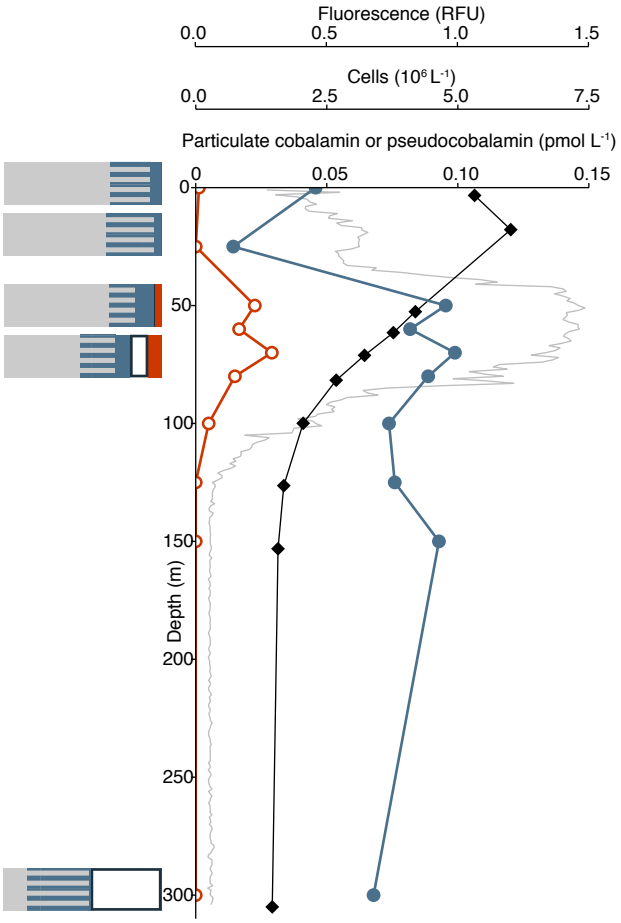

Figure 3.10: Pseudocobalamin and cobalamin producers at Station PAPA. Measurements of particulate cobalamin (closed blue circles), particulate pseudocobalamin (orange open circles), prokaryote cell abundance (black diamonds), and in situ fluorescence (grey). Bar graphs with the percent of cells with the following predicted cobalamin strategies: unlikely producers (grey), bacteria with unknown cobalamin biosynthesis capacity (grey/blue striped), bacteria that likely possess the potential to produce cobalamin (blue), Thaumarchaeota (likely to produce cobalamin, white), and *Cyanobacteria* (likely to produce pseudocobalamin, orange) at the corresponding depths. Cobalamin and pseudocobalamin concentrations are summed values of the detected beta ligands as in Figure 3.2. Pseudocobalamin concentrations are presented in cobalamin equivalents (see Supplemental Methods).

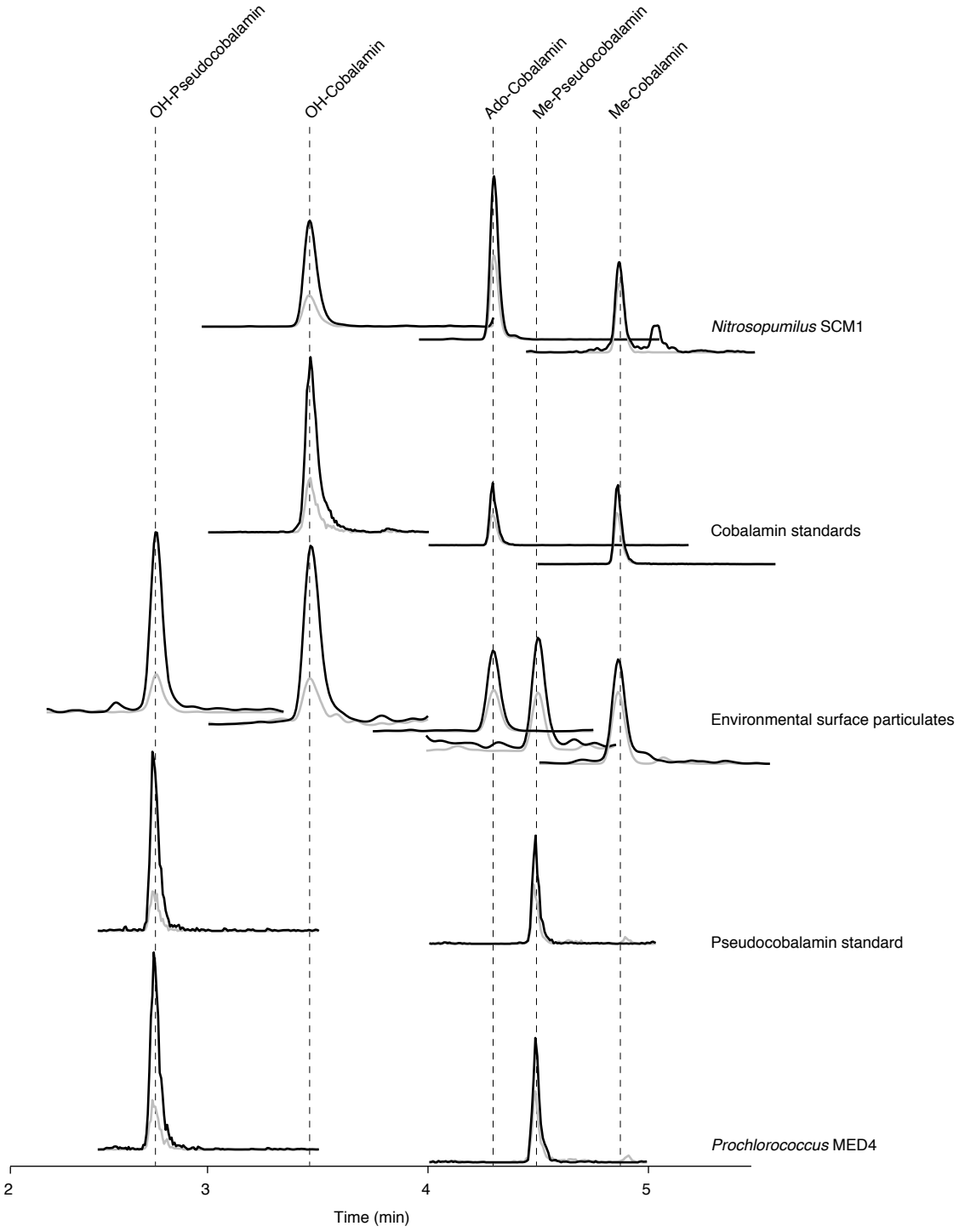

Figure 3.11: Caption on following page.

Figure 3.11. *Continued from previous page.* Example MS chromatograms. Two transitions from the 5 most common analytes (OH- and Me-pseudocobalamin and OH-, Me- and Ado-cobalamin) in standards, cultures (*Nitrosopumilus* SCM1 and *Prochlorococcus* MED4), and the environment (surface sample from Station 1). Both the primary (used for quantification, black) and secondary (used for identity confirmation, grey) fragment ions are shown with an arbitrary y-axis (ion intensity).

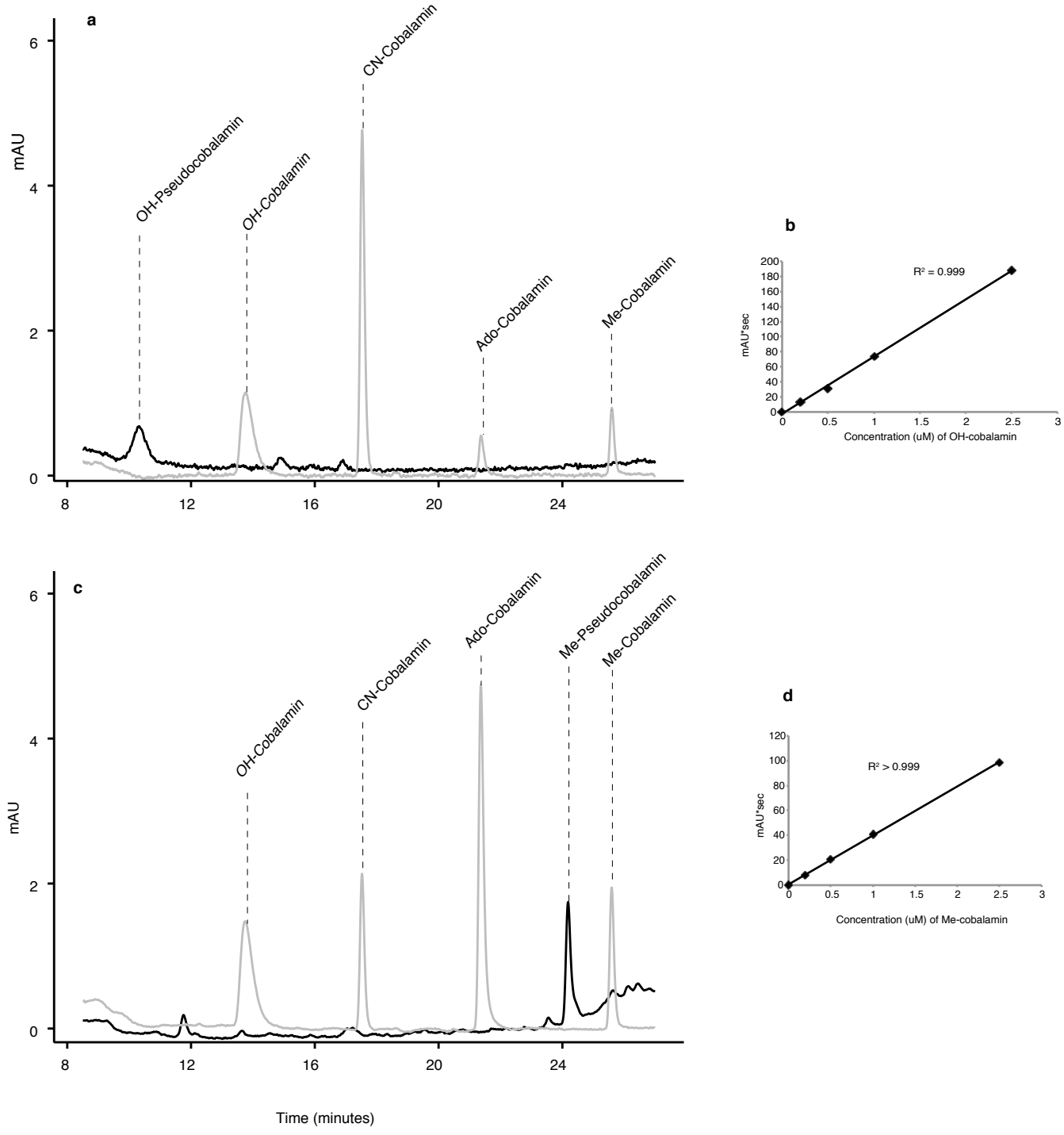

Figure 3.12: Chromatograms and standard curves from pseudocobalamin stock preparation and quantification. Chromatograms of 0.5 μM cobalamin standards (in grey) and pseudocobalamin stocks at (a) 362 nm or (c) 259 nm. Standard curves of OH-cobalamin (b, at 362 nm) and Me-cobalamin (d, at 259 nm) 0, 0.2, 0.5, 1, 2.5 μM.

3.8 Tables

			Molecules of cobalamin per cell			Molecules of pseudocobalamin per cell			nmole cobalamin analog per mole carbon
Phylum	Strain	*n*	OH-	Me-	Ado-	OH-	Me-	Ado-	
Proteobacteria	*Vibrio fischerii*	3	*nd*	*nd*	*nd*	*nd*	*nd*	*nd*	0
	Sulfitobacter sp. SA11	3	3 ± 0.1	4 ± 0.4	7 ± 3	*nd*	*nd*	*nd*	1-10
	R. Pomeryoi DSS-3	3	520 ± 290	120 ± 100	1200 ± 320	*nd*	*nd*	*nd*	240-260
Thaumarchaeota	*Nitrosopumilus sp.* HCE1	3	420 ± 19	52 ± 17	1600 ± 140	*nd*	*nd*	*nd*	4200-5300
	Nitrosopumilus sp. HCA1	3	1860 ± 14	366 ± 55	2252 ± 210	*nd*	*nd*	*nd*	9300-11600
	Nitrosopumilus sp. PS0	3	598 ± 65	139 ± 20	1548 ± 177	*nd*	*nd*	*nd*	4700-5900
	N. maritimus SCM1	3	670 ± 52	13 ± 4	680 ± 130	*nd*	*nd*	*nd*	2800-3500
Cyanobacteria	*Prochlorococcus* MED4	5	*nd*	*nd*	*nd*	240± 40	990 ± 90	trace	430-1190
	Prochlorococcus MIT9313	5	*nd*	*nd*	*nd*	360 ± 57	1600 ± 280	trace	130-450
	Synechococcus WH7803	3	*nd*	*nd*	*nd*	69 ± 13	16600 ± 2000	trace	2400-3000
	Synechococcus WH8102	3	*nd*	*nd*	*nd*	270 ± 43	9170 ± 1200	trace	1480-1870

Table 3.1: Summary of cultured archaea and bacteria in this study. Observed cellular quotas of major forms of cobalamin and pseudocobalamin with standard deviation with n replicates of each strain; nd = not detected, trace = detected a small peak but were unable to quantify (See Supplemental Methods).

Treatment	Molecules (x10^3) of cobalamin per cell				Molecules (x10^3) of pseudocobalamin per cell				nmole cobalamin per mole carbon	molecules (x10^8) SAM per cell
	OH-	Me-	Ado-	CN-	OH-	Me-	Ado-	CN-		
200 pM cobalamin	180 ± 14	21 ± 1.1	140 ± 20	*nd*	*nd*	*nd*	*nd*	*nd*	800 ± 85	7.1 ± 1.6
1 pM cobalamin	61 ± 6.4	13 ± 3.6	49 ± 13	*nd*	*nd*	*nd*	*nd*	*nd*	290 ± 74	1.7 ± .33
1 pM cobalamin + 200 pM pseudocobalamin	52 ± 22	12 ± 1.7	57 ± 6.9	*nd*	990 ± 531	*nd*	trace	trace	290 ± 56	2.2 ± .26

Table 3.2: Observed intracellular cobalamin and SAM contents of *T. pseudonana*. Three different conditions with standard deviation of 3 replicates of each strain; nd = not detected, trace = detected a peak but were unable to quantify (See Supplemental Methods).

Phylum	Class	Order	Clade	Genomes with corrin biosynthetic capacity		Applied Category
				This study	2014 Analysis‡	
Actinobacter				1/3		unknown
Bacteroidetes	Flavobacteriia			6/342	0/25	unlikely
Chlorobi				5/12		unknown
Chloroflexi				9/32		unknown
Cyanobacteria		Prochlorales		50/52	18/18	pseudocobalamin
Cyanobacteria		Chroococcales		100/104	20/21	pseudocobalamin
Cyanobacteria		All others		97/99		pseudocobalamin
Deferribacteres				4/6		unknown
Planctomycetes				5/28		unknown
Proteobacteria	Alphaproteobacteria	Rhodobacterales		162/200	43/50	likely
Proteobacteria	Alphaproteobacteria	Rhodospirillales		83/102		likely
Proteobacteria	Alphaproteobacteria	SAR11 Clade		0/12	0/11	unlikely
Proteobacteria	Betaproteobacteria			0/25		unknown
Proteobacteria	Deltaproteobacteria			106/159		unknown
Proteobacteria	Gammaproteobacteria	Altermonadales		12/226	3/49	unlikely
Proteobacteria	Gammaproteobacteria	Oceanospirillales	SAR86 Clade			unlikely§
Proteobacteria	Gammaproteobacteria	Oceanospirillales	All others	36/82	6/13	unknown
Proteobacteria	Gammaproteobacteria	Pseudomondales		749/790	0/6	unknown
Proteobacteria	Gammaproteobacteria	Salinisphaerales		1/2		unknown
Proteobacteria	Gammaproteobacteria	Thiotrichales		22/117	1/9	unknown
Proteobacteria	Gammaproteobacteria	Vibrionales		2/569	1/33	unlikely
Proteobacteria	Gammaproteobacteria	Xanthomonadales		0/419	0/1	unlikely
All others						unknown

Table 3.3: Cobalamin biosynthetic capacity deduced from taxonomy. Results are summarized from our full genome analysis (This study) and a previous study (2014 Analysis). Applied Category is the potential for cobalamin biosynthesis we used for each taxonomic group during our estimation of Thaumarchaeota contribution to biosynthesis of cobalamin. ‡ Sañudo-Wilhelmy *et. al.* (2014). § Dupont *et. al.* (2012).

Compound	Transitions monitored	CE (V)	CV (V)
OH-pseudocobalamin	659.8 -> **136.1**, 348.1	30	45
CN-pseudocobalamin	673.3 -> **136.1**, 348.1, 660.3	30	45
Me-pseudocobalamin	668.3 -> **136.1**, 348.1, 660.3	30	30
Ado-pseudocobalamin	785.3 -> **136.1**, 348.1, 660.3	30	50
S-adenosyl methionine	399.258 -> **136.1**, 250.2	26	2

Table 3.4: MS conditions used to monitor for pseudocobalamin and S-adenosylmethionine]. CE, collision energy in volts (V); CV, cone voltage in volts (V). For the pseudocobalamins, parent masses are all the $[M+2H]^{2+}$ except OH-pseudocobalamin which loses its OH- during ionization as shown previously [54]. 136 m/z, α ligand (adenine); 348 m/z α ligand with sugar and phosphate groups; 660 m/z doubly charged loss of the β ligand.

Chapter 4

METABOLIC CONSEQUENCES OF COBALAMIN SCARCITY IN DIATOMS AS REVEALED THROUGH METABOLOMICS

4.1 Abstract

Abstract[1]

Diatoms perform about 20% of photosynthesis on earth, play a crucial role in forming the base of the marine food web, and sequester anthropogenic carbon into the deep ocean through the biological pump. In some areas of the ocean, diatom growth is limited by the organic micronutrient cobalamin (vitamin B_{12}). Cobalamin affects phytoplankton growth rate and biomass yield, though the metabolic consequences of cobalamin limitation are not well understood. In a controlled laboratory setting, we grew two diatoms, *Thalassiosira pseudonana* and *Navicula pelliculosa*, under cobalamin limited and replete conditions to identify changes in metabolite pools under cobalamin limitation. Using targeted and untargeted metabolomics, we show that the cobalamin-dependent *T. pseudonana* experienced a metabolic cascade under cobalamin limitation that affects the central methionine cycle, the transsulfuration pathway, and the composition of its osmolyte pools. *N. pelliculosa*, a cobalamin-independent diatom, also shows evidence for changes in the methionine cycle when grown without cobalamin, though to a lesser extent than what we observed in *T. pseudonana*. Both diatoms showed a decrease in 5'-methylthioadenosine when limited for cobalamin, suggesting a disruption in the diatoms' polyamine biosynthesis. Furthermore, two acylcarnitines accumulated in *T. pseudonana* under cobalamin limitation, suggesting the use of an adenosylcobalamin-dependent enzyme, methylmalonyl CoA mutase, which is

[1]In preparation for submission as K. R. Heal, N. A. Kellogg, L. T. Carlson, R. M. Lionheart, and A. E. Ingalls, "Metabolic consequences of cobalamin scarcity in diatoms as revealed through metabolomics," *In prep.*

supported by observations of intracellular adenosylcobalamin in both diatoms. Overall, these changes in metabolite pools yield insight into the metabolic consequences of cobalamin limitation in diatoms and suggest that cobalamin starvation may have consequences for trophic interactions that are based on metabolite production by phytoplankton.

4.2 Introduction

Cobalamin (vitamin B_{12}) is an important micronutrient that can shape and control the primary productivity in marine ecosystems. Many species of algae require an exogenous source of the compound, which is produced through a biosynthetic pathway of over two dozen steps by select bacteria and archaea [78, 9, 43]. In the ocean, select heterotrophic bacteria and Thaumarachaeota are thought to be the main producers of cobalamin, while the majority of cyanobacteria produce pseudocobalamin, a closely related compound that is less bioavailable to eukaryotic phytoplankton [49, 130, 50]. Though few direct measurements of oceanic cobalamin (often < 1 pM) exist, perturbation experiments of natural populations suggest that the scarcity of cobalamin in the marine environment can control which phytoplankton species thrive [11, 10, 131, 54, 132].

In eukaryotic algae, the dependency on exogenous cobalamin, or *cobalamin auxotrophy*, arises from the compound's role as a cofactor in the methylcobalamin (Me-cobalamin)-dependent methionine synthase (MetH), a key enzyme in the synthesis and regeneration of methionine [133]. Some algae also have a cobalamin-independent isoform of the enzyme, MetE [134, 133]. MetE performs the same function as MetH, but much less efficiently [16, 135]. The current understanding of cobalamin-dependence in algae hinges on the presence of MetE, that is if algae have only MetH, they are dependent on exogenous cobalamin, but if algae have both MetE and MetH they can switch between enzymes and are cobalamin-independent [16, 133]. Both cobalamin-dependent and -independent eukaryotic algae can grow over a wide range of cobalamin concentrations [15, 77, 12, 13], and transcriptomic and proteomic techniques have been employed to elucidate the transcriptional and translational response that model diatoms have to cobalamin scarcity [134, 46, 16, 133]. These

analyses have revealed strategies diatoms employ to reduce cellular demand for cobalamin, increase cobalamin acquisition from the environment, and manage reduced functionality of cobalamin-dependent enzymes [134, 46, 16].

The important methyl donor, S-adenosyl methionine (SAM), is depleted in diatom cells grown under cobalamin limitation [130]; however, apart from SAM depeletion most of the hypothesized mechanisms from transcriptomic analyses have not yet been observed at the metabolite level. Specific hypotheses deduced from transcriptional changes include a decrease in the osmoregulators dimethylsulfoniopropionate (DMSP) and glycine betaine (GBT), an imbalance in the methionine cycle, and changes in polyamine biosynthesis, which all have yet to be validated. Beyond these hypotheses, all sequenced diatoms code for an adenosylcobalamin-dependent enzyme (methyl malonyl CoA mutase, MCM) [134, 133], but it is unclear if cobalamin availability affects diatoms' metabolism through decreased efficiency of this enzyme as well as the methionine synthase. Furthermore, it is not clear how algae cope with the accumulation of toxic compounds like S-adenosyl homocysteine (SAH) and homocysteine as seen in cobalamin-stressed algae [15]. While transcriptomic and proteomic analyses have provided valuable insight into cobalamin limitation in diatoms, there are likely additional key metabolic consequences that are not well-regulated or regulating post-translationally which are otherwise masked without direct metabolite measurements.

In this study, we used targeted and untargeted metabolomics to explore the metabolic consequences that cobalamin scarcity has on two diatoms: the centric, cobalamin-dependent *Thalassiosira pseudonana* and the pennate, cobalamin-independent *Navicula pelliculosa*. In our targeted approach, we obtain relative abundances (between samples) of known primary metabolites like amino acids, osmolytes, and methionine cycle intermediates [8]. In untargeted metabolomics, we hope a holistic understanding of both known and unknown metabolite pools and how they change under different conditions. Using a batch culture approach, we grew these two diatoms under low and replete cobalamin in both low and saturating light to parse out differences in metabolite pools that were due to growth rate changes. These experiments provide us the foundation we need to understand how important primary

producers experience and cope with cobalamin starvation in laboratory and natural settings.

4.3 Results

4.3.1 Growth rates

We grew both diatoms under both saturating and low light conditions with low or replete cobalamin (control). Cobalamin was provided as OH-cobalamin in all experiments. As expected, the effect of cobalamin scarcity on the growth rate was more drastic for cobalamin-dependent *T. pseudonana* than for cobalamin-independent *N. pelliculosa*. The growth rate of *T. pseudonana* was lowered by 38% and 48% when cobalamin-limited (1 pM initial cobalamin) compared to the cobalamin replete controls under saturating and low light, respectively ($p < 0.001$). *N. pelliculosa* experienced a more modest but statistically significant 20% and 14% decrease in growth rate when grown without cobalamin compared to cobalamin replete control under saturating and low light, respectively ($p < 0.001$, Table 4.1 and Figure 4.1). The low light conditions also affected *T. pseudonana* more drastically, with a 12% and 28% ($p = 0.01$) decrease under replete and low cobalamin, respectively. *N. pelliculosa* showed a 10% decrease ($p < 0.001$) in growth rate under replete cobalamin between light conditions, but no significant difference ($p = 0.75$) between the low and saturating light conditions for treatments without added cobalamin (Table 4.1 and Figure 4.1).

4.3.2 Targeted metabolomics

In our analyses, the concentration of a compound is directly related to the size of its peak, but comparisons cannot be made between compounds due to differences in ionization efficiency (i.e. a larger peak area for one compound does not necessarily mean that it is more abundant than another compound). Therefore, all quantitative information is relative between samples, not compounds.

We detected 75 metabolites in *T. pseudonana* and 55 metabolites in *N. pelliculosa* in our targeted approach (Datasets S4.1 and S4.1). To assess if our metabolomics approach

could discern different experimental conditions, we performed a two dimensional non-metric multidimensional scaling (NMDS) analysis. Using the relative concentrations that resulted from of our targeted analysis, an NMDS ordination resulted in a low stress value for the metabolomes of both organisms (Figure 4.2). For *T. pseudonana*, the non-metric approach was successfully able to discern low light from saturating light treatments as well as low cobalamin from replete cobalamin treatments (Table 4.3). Though the growth rate changed in *N. pelliculosa* under cobalamin starvation, the NMDS was not able to discern appreciable differences in the metabolite pools under cobalamin starvation in multivariate space (Table 4.3). Compounds that significantly contributed to the NMDS plot are reported in Table 4.2.

Of the targeted metabolites that we observed, 36% showed a significant difference in abundance between cobalamin treatments under both light regimes in *T. pseudonana*, while no compounds showed a significant difference in *N. pelliculosa* ($p < 0.05$, Table 4.4 and Figure 4.3).

4.3.3 Untargeted metabolomics

A targeted approach to metabolomics is inherently biased as the only data acquired are on a set of prescribed compounds chosen by the analyst. On the contrary, an untargeted approach is only biased towards compounds that are detectable in the analysis chosen. Similar to our targeted analyses, our untargeted analyses showed that many more mass features changed significantly under cobalamin limitation in *T. pseudonana* than in *N. pelliculosa* (Table 4.4, Figure 4.4). This less biased analysis corroborates our observation that the metabolome of *T. pseudonana* is more affected by cobalamin limitation than *N. pelliculosa*. Our multivariate approach yielded similar results to the targeted analyses (Figure 4.2, Table 4.3).

A major bottleneck in metabolomics research is the identification of detected peaks. Challenges remain in the field because there are outstanding gaps in metabolic pathways, existing databases have insufficient coverage in fragmentation spectra (MS^2), and absolute identification relies on authentic standards which can be prohibitively expensive or commercially unavailable [136]. We focussed our identification efforts on quality mass features

that showed a univariate response to cobalamin limitation in the hydrophilic fraction of our extraction, and we used the ranking system outlined in Sumner *et. al.* (2007). This ranking system matches to a compound's exact mass to charge (m/z), fragmentation, and retention time (and combinations thereof, when possible). As previously noted [137], these stringent guidelines yield a small subset of identifiable metabolites from an untargeted metabolomics analysis, but we can expect to see improvements in this yield in the future as metabolomics analyses become more common and the databases improve.

Using automated approaches, we were able to identify or putatively identify 1-2% of the quality mass features in *T. pseudonana* and *N. pelliculosa* (Datasets S4.3 and S4.4). With a more manual approach that we describe in our methods, we identified a large percentage of the mass features that showed a response to cobalamin limitation in the aqueous fraction of *T. pseudonana* (Table 4.6). Many of the mass features we identified matched those in our targeted analysis (i.e. glycine betaine, S-adenosylhomocysteine, and tryptophan), and the change in pool size of these compounds under cobalamin limitation matched in the signal from the targeted results. All the compounds explicitly discussed in this manuscript were putatively identified with m/z and MS^2 and validated with a newly purchased standard or matched to a standard in our existing targeted analysis [8].

4.3.4 Changes in specific metabolite pools

Cobalamins. We provided cobalamin in the hydroxocobalamin (OH-cobalamin) form only. We observed that under replete cobalamin, intracellular cobalamin concentrations (both OH-cobalamin and adenosylcobalamin (Ado-cobalamin)) were higher for *T. pseudonana* when provided with more cobalamin (Figure 4.3, Dataset S4.1). We also detected OH- and Ado-cobalamin in the replete cobalamin treatments of *N. pelliculosa* (Dataset S4.2).

Many of the compounds with significant differences between cobalamin treatments in *T. pseudonana* can be tied directly to changes in the methionine cycle, which has been previously hypothesized because cobalamin is used as the catalytic center in their central methionine synthase (MetH) enzyme [134, 46]. In the low cobalamin treatments, we see that

T. pseudonana experienced a major change in metabolite pools in the methionine cycle under cobalamin starvation (Figure 4.5). In cobalamin limited *T. pseudonana*, cystathionine was on average 42 times more abundant compared to cobalamin replete cells ($p = 0.02$, Figures 4.3, 4.6, and 4.7).

Methionine and SAM decreased while SAH increased in cobalamin-deprived *T. pseudonana* cells (Figures 4.3, 4.5, and 4.6). Methionine, SAM, and SAH showed the same general trends in *N. pelliculosa* under cobalamin treatment as *T. pseudonana*, though the differences in metabolite concentrations between cobalamin treatments were not statistically significant ($p = 0.055$, 0.26, and 0.26, for methionine, SAM, and SAH, respectively, Figures 4.3, 4.5, and 4.6). We did not detect cystathionine in *N. pelliculosa*.

Compatible solutes. Many compounds that have been identified as possible compatible solutes changed under cobalamin limitation in *T. pseudonana*. Dimethylsulfoniopropionate (DMSP) production appeared to be affected by cobalamin availability in *T. pseudonana*, as we observed on average 12 time less DMSP under cobalamin starvation ($p = 0.02$ under saturating light, $p = 0.04$ under low light, Figures 4.3 and 4.8). Glycine betaine (GBT) also decreased in cobalamin limited cells compared to cobalamin repelte controls (on average 11 times less GBT under cobalamin limitation, $p = 0.02$ under saturating light, $p = 0.007$ under low light, 4.3 and 4.9). Compounds related to GBT also decreased under cobalamin limitation (Figure 4.9), while compounds that are possible compatible solutes increased under cobalamin limitation in *T. pseudonana* (Figure 4.10). There were no obvious changes in the compatible solute pool of *N. pelliculosa*, though we did not detect GBT or DMSP in these diatoms.

5'-methylthioadenosine. In *T. pseudonana*, 5'-methylthioadenosine (MTA) was less abundant in cobalamin-limited cells (Figure 4.11, $p = 0.02$). We detected MTA in our untargeted analysis and confirmed its identity with commercial standards. Although not statistically significant, we observed a modest decrease in MTA in *N. pelliculosa* under cobalamin limitation (on average, about 60% as abundant under cobalamin limitation, Figures 4.6 and 4.11, $p = 0.15$).

Acylcarnitines. Our untargeted analysis revealed that two aceylcarninites, propionylcarnitine and butyrlcarnitine, accumulated in cells that were limited for cobalamin (Figure 4.12). In *T. pseudonana* this depletion was statistically significant for both molecules ($p < 0.05$). We observed the same general trend in *N. pelliculosa*, though the relationship was not statistically robust.

Although *N. pelliculosa* and *T. pseudonana* are roughly the same size, we saw that *N. pelliculosa* consistently contained less of the discussed molecules on a per cell basis (from 2 to 20 times less under replete cobalamin conditions). These are some of the first observations of metabolite pools in these organisms and our work demonstrates that intracellular concentrations of these primary metabolites are remarkably variable between species and growing conditions.

Several other compounds showed a univariate response to cobalamin limitation in *T. pseudonana* (Figures 4.3 and 4.4, Table 4.6). The concentrations of many amino acids were affected by cobalamin: glutamine and glutamic acid concentrations were roughly havled while tyrosine, and asparagine increased ($p < 0.05$ for all, Figure 4.3, Table 4.6, and Dataset S4.1). Tryptophan and its derivatives have been implicated in underpinning algal-bacterial interactions in the ocean [90], and our work suggests that cobalamin availability may have some influence on this chemical communication, as tryptophan increased under cobalamin limitation (Table 4.6). Interestingly, both purine nucleosides showed changes under cobalamin limitation, though in opposite directions (guanosine accumulated while adenosine decreased). Taken together, the metabolomic approach used in this study revealed far-reaching metabolic consequences of cobalamin limitation in these globally important organisms.

4.4 Discussion

4.4.1 Cobalamin in T. pseudonana and N. pelliculosa

Similar to previous work, we found that cobalamin limitation has a measurable effect on the growth rate of the cobalamin-independent diatom (in our case, *N. pelliculosa*), similar

to other cobalamin-independent diatoms [16] (Figure 4.1). Although *N. pelliculosa* does not have an available genome, in other diatoms, cobalamin independence arises from the presence of the *metE* gene in addition to the *metH* gene [133]. In the model pennate diatom *Phaeodactylum tricornutum*, the switch from using MetH to MetE resulted in a depression of growth rate similar to what we observe for *N. pelliculosa* [16]. Therefore, we hypothesize that *N. pelliculosa* encodes both *metE* and *metH*. Using MetE in place of MetH would relieve the organism of the cobalamin requirement when cobalamin is unavailable and explain the muted metabolomic response in the metabolite pools compared to cobalamin-dependent *T. pseudonana* as we observed in both multidimensional and univariate space (Figures 4.2, 4.3, and 4.4).

Intracellular cobalamins. Our observations of cobalamin variability in the two diatoms confirmed that *N. pelliculosa* transported cobalamin into cells when available and corroborated others' work demonstrating that the intracellular quota for both cobalamin-dependent and -independent diatoms is remarkably flexible, and not strictly stoichiometrically constrained [12]. The Ado-cobalamin form of cobalamin has been previously observed in *T. pseudonana* [130], but has not been well explained. Here we observed Ado-cobalamin in both strains of the diatoms that were supplied only OH-cobalamin, implying that the diatoms enzymatically added the adenosyl α ligand, and the genetic potential for this has been identified in all sequenced diatom genomes [133]. In addition to using cobalamin in the well studied MetH, this work suggests that the cobalamin is being used in a enzyme that uses Ado-cobalamin, the most likely candidate being MCM which has been identified in all sequenced diatom genomes [134, 133].

4.4.2 Disruption of the methionine cycle

The methionine cycle showed obvious changes under cobalamin limitation (Figures 4.3, 4.5, and 4.6). SAM, one of the compounds in the methionine cycle, is an enormously important molecule in many cellular processes where it acts as a methyl donor and radical source and likely only follows ATP in the variety of reactions in which it serves as a cofactor [138]. SAM

serves as a methylating agent for a wide variety of reactions in both primary and secondary metabolism, including DNA methylation (affecting gene expression) and osmoprotectant biosynthesis [139]. Once SAM has donated its methyl group during a methylation reaction, the resulting compound (SAH) must be re-methylated via homocysteine and then methionine in the central methionine cycle (Figure 4.6). SAM depletion has been hypothesized to be a major consequence of cobalamin starvation in diatoms by transcriptomic and proteomic inference [134, 46], and a recent study showed that this metabolite is less abundant in cobalamin-starved *T. pseudonana* cells [130]. We also found that SAM was generally less abundant under cobalamin starvation in *T. Pseudonana* (on average, 4 times less SAM under cobalamin limited conditions, Figures 4.3, 4.5, and 4.6), though the relationship was not statistically robust in a univariate comparison between treatments ($p = 0.1$) due to high variability. SAM also significantly contributed to the NMDS analysis of *T. pseudonana*'s metabolome ($p = 0.006$, Table 4.2), indicating that the compound plays a significant role in driving differences between the experimental treatments we tested.

We detected changes in the size of the pools of two other metabolites in the central methionine cycle. Methionine and SAH both showed significant differences between cobalamin treatments in *T. pseudonana* ($p = 0.04$, $p < 0.001$, respectively), where SAH was 10 times more abundant and methionine 10 times less abundant under cobalamin starvation (Figures 4.3, 4.5, and 4.6).

In *N. pelliculosa*, we observed a less intense disruption of the methionine cycle under cobalamin limitation (Figures 4.3, 4.5, and 4.6). These results suggest that MetE is not able to sustain the methionine cycle (and therefore growth rates) to the same degree as the cobalamin-dependent MetH in *N. pelliculosa*, as previously hypothesized for other cobalamin-independent diatoms [46, 16].

4.4.3 Changes in the transsulfuration pathway

Two intermediates in the methionine cycle, SAH and homocysteine, are potentially cytotoxic compounds due to their affinity for methyl groups, high reducing power, and structural

similarity to SAM or proteinous amino acids [140, 141, 142, 143], but no mechanisms have been proposed to cope with or prevent the build up of these metabolites during cobalamin starvation in diatoms. We were unable to detect homocysteine in any samples; this could be due to low concentration, metabolite instability, poor ionization efficiency, considerable matrix effect that confounds the instrumental signal, or some combination of these factors. However, using different methodologies, others have shown that homocysteine builds up under cobalamin starvation in a green algae [15], humans [144], and rats [145] indicating this is a common consequence of cobalamin limitation in cobalamin dependent organisms. In humans, a major consequence of cobalamin limitation is an increase in the traffic of metabolites through the transsulfuration pathway (the metabolic pathway that converts homocysteine to cysteine and vice versa) that results in an accumulation of cystathionine [146, 147](Figure 4.6).

Of the targeted metabolites we observed in this experiment, the sulfur-containing cystathionine was the most drastically affected by cobalamin availability, increasing greatly under cobalamin limitation in *T. Pseudonana*(Figures 4.3, 4.6, and 4.7). The transsulfuration pathway is poorly characterized in diatoms, and all four enzymes in the pathway may be encoded in *T. pseudonana*'s and other diatoms' genomes (Table 4.5). Regardless of the particular enzymes involved, our cystathionine data suggest that either more cystathionine is being produced or less cystathionine is being converted to cysteine or homocysteine, demonstrating a change in the transsulfuration pathway as a result of cobalamin limitation, which may be a strategy to prevent or cope with the accumulation of homocysteine and SAH.

In other eukaryotes, the activity of at least one of the enzymes in the transsulfuration pathway is regulated allosterically. Specifically, the production of cystathionine from homocysteine via cystathionine-β-synthase (CBS, labeled 1 on Figure 4.6) depends on the concentration of small metabolites in the methionine cycle [19, 148]. The regulation of the other enzymes in the transsulfuration pathway is less understood, but if they are also regulated allosterically, it may explain why previous transcriptomic and proteomic studies have not uncovered a change in the transsulfuration pathway in cobalamin-limited diatoms [46]. Though

we do not fully understand the metabolic pathways related to cystathionine biosynthesis in diatoms, detecting this compound in *T. pseudonana* demonstrates the power of measuring metabolites in concert with transcriptomic and proteomic studies when investigating the physiological responses and adaptations of these important organisms.

We did not detect any cystathionine in any of the *N. pelliculosa* cultures, though this could be an issue of detection limit rather than lack of the compound in the organism. Nevertheless, unlike the disruption of the methionine cycle, the build of cystathionine does not appear to be a widespread response to cobalamin limitation in diatoms and may be limited to those organisms that experience cobalamin limitation more drastically due to auxotrophy.

4.4.4 Depletion of major osmolyte pools

Two of *T. pseudonana*'s abundant pools of osmolytes, dimethylsulfoniopropionate (DMSP) and glycine betaine (GBT), are greatly reduced under cobalamin starvation (Figures 4.3, 4.8, and 4.9). Osmolytes are large pools of polar compounds that algae use to balance osmotic pressure due to their saline environment; in *T. pseudonana*, DMSP and GBT are two of the three largest peaks we observed in the targeted analyses, indicating they are very abundant molecules and supporting their role as osmolytes. This observation is consistent with other published metabolomes of *T. pseudonana* [149]. Though the biosynthetic pathway for DMSP is not fully elucidated in diatoms, it has been shown that DMSP requires methylation from SAM and derives directly from methionine (Figure 4.8). Thus, DMSP production essentially removes methionine from the methionine cycle [150]. Our work suggests that cobalamin availability directly affects DMSP production in *T. pseudonana*, as this compound dramatically decreases under cobalamin limitation 4.3 and 4.8). Like previous work, we see a change in intracellular DMSP under the different light regimes, where cultures grown under saturating light had more DMSP when compared to low light cultures, supporting the osmolyte's dual role as an antioxidant [151]. Due to our experimental design, we cannot rule out that the change we observed in DMSP between cobalamin treatments is

due to changes in growth rate as the pattern followed growth rate changes. DMSP is the precursor for climatically important gas dimethylsulfide [152], and our results suggest that in order to constrain DMSP production in algae we must understand how these organisms experience cobalamin limitation in the environment.

We also observed a dramatic decrease of GBT under cobalamin limitation (Figures 4.3 and 4.9), supporting former conjectures that GBT production in diatoms depends on the presence of copious amounts of SAM [134], which is true for other organisms [153]. In higher plants, GBT is synthesized from choline, which receives three methyl groups from SAM [154, 155]. Choline is also less abundant in *T. pseudonana* under cobalamin limitation, though not to such a dramatic degree as GBT (on average two times less abundant, $p = 0.049$ under saturating light, $p = 0.041$ under low light) (Figure 4.9, Supplemental Table 2). We detected another choline derivative, glycerophosphocholine, in our untargeted analysis; this compound also decreased under cobalamin limitation (Figure 4.9, Table 4.6). In our targeted analysis, we do not see an obvious replacement for GBT and DMSP in the osmolyte pools of *T. pseudonana* under cobalamin limitation, but our untargeted analysis revealed that two proposed osmolytes, homarine and proline betaine, increased under cobalamin limitation (Figure 4.10).

Both DMSP and GBT have been proposed as important energy and carbon sources for heterotrophic bacteria living near algae in the surface ocean due to the high concentrations of these molecules inside algae [156, 157], and the compound can change the metabolism of marine heterotrophs [137]. Cobalamin availability clearly influences which osmolyte pools diatoms produce and therefore can provide to a wider community. In this way, cobalamin availability may change which members of the heterotrophic community can thrive, even if those organisms are not directly reliant on cobalamin.

We did not detect DMSP nor GBT in the metabolome of *N. pelliculosa*, showing that unlike *T. pseudonana* and many other eukaryotic phytoplankton, *N. pelliculosa* does not appear to accumulate DMSP or GBT as major osmolytes. We did observe choline, which under low light conditions showed a slight decrease under cobalamin limitation, though the

relationship was not statistically significant (about two times less abundant, $p = 0.5$, Figures 4.3).

The diversity of osmolytes used by organisms is an outstanding area of research in evolutionary and molecular biology, and this work sheds new insight regarding the role of cobalamin availability on the osmolyte suite used by algae. The apparent dependencies of GBT and DMSP on SAM and methionine would put extra strain on the methionine cycle (and therefore demand more cobalamin) compared to using SAM-independent osmolytes like small carbohydrates or amino acids like proline. The pervasive use of GBT and DMSP in diatoms despite this SAM dependency suggests that these compounds act as superior chemical osmolytes or have functions beyond osmoprotection like antioxidant capacity or predatory defense as has been proposed for DMSP [158]. Another possibility is that there has been little evolutionary pressure for cobalamin-dependent organisms like *T. pseudonana* to use SAM-independent osmolytes because there has been a consistent source of cobalamin to these organisms; this is supported by the likely loss of the *metE* gene and subsequent cobalamin dependence [133]. It is tempting to speculate that the lack of DMSP and GBT in *N. pelliculosa* arose from the organism's lack of a consistent source of cobalamin over evolutionary timescales, but this pattern does not hold true across all cobalamin-independent diatoms. For example, the cobalamin-independent *P. tricornutum* produces both GBT and DSMP as major osmolytes [159], despite being able to survive without cobalamin and maintaining the *metE* gene [16]. Nevertheless, we find it very possible that cobalamin availability has played a role in shaping the osmolyte suites used by algae over evolutionary timescales.

4.4.5 Depletion of MTA

In *T. Pseudonana*, MTA was about half as abundant in cobalamin-limited cells (Figure 4.11). This compound is a sulfur-containing nucleoside that is produced from SAM (via decarboxylated-SAM) during polyamine biosynthesis (Figure 4.6). The available diatom genomes have strong matches to characterized enzymes in the polyamine biosynthesis steps that results in the production of MTA (Table 4.5). Polyamines are likely essential for growth

of all organisms and are the building blocks of long chain polyamines (LCPAs) that diatoms use to build their silica frustules [160]. Due to the high demand that polyamine biosynthesis has for SAM, others have hypothesized that polyamine biosynthesis could be disrupted by cobalamin limitation [46], and our observed decrease in MTA corroborate this hypothesis. Our extraction process does not extract the polyamines within the frustule structure, so we cannot measure these molecules using this metabolomics approach. In other organisms, MTA is recycled back to methionine via the methionine salvage pathway [161] (Figure 4.6). The methionine salvage pathway is not well characterized in diatoms, though *T. pseudonana* and other diatoms likely possess some, if not all, of the pathway (Table 4.5). Thus, it is also possible that the observed change of MTA is a result of an increased efficiency in the methionine salvage pathway. The modest decrease of MTA in *N. pelliculosa* under cobalamin limitation suggests that this phenomena is not restricted to cobalamin-dependent diatoms.

4.4.6 Accumulation of acylcarnitines

In *T. pseudonana*, propionylcarnitine showed a similar pattern as SAH, with the cobalamin-limited cells having nearly seven times more propionylcarnitine than our cobalamin-replete cultures (Figure 4.12). Accumulation of propionylcarnitine has been observed in cobalamin-limited mammals [162, 163], and is used as a diagnostic for cobalamin deficiency in newborns [146, 163]. We also observed an increase in butyrlcarnitine under cobalamin limitation (Figure 4.12), though to a lesser degree. An increase in butyrlcarnitine has also been observed in cobalamin-deficient animals, to a lesser extent than a concurrent increase in propionylcarnitine [164], similar to what we observed. In mammals, the production of acylcarnitines (like propionylcarnitine and butyrlcarnitine) is a method for liberating Coenzyme A (CoA) from acyl-CoAs when the use of an acyl-CoA is impaired. Previous work has hypothesized that increased propionylcarnitine observed under cobalamin limitation is a result of the inability to process propionyl- or methylmalonyl-CoA through the cobalamin-dependent enzyme methylmalonyl-CoA mutase (MCM) [146] (Figure 4.13), and there is strong genomic evidence for this pathway in *T. pseudonana* and other diatoms (Table 4.5). Some eukaryotic phy-

toplankton encode MCM in their genomes, including *T. pseudonana* [134, 133], though the presence of MCM in diatom genomes does not result in an absolute cobalamin requirement — for instance, the model diatom *P. tricornutum* codes for and transcribes the enzyme, but can thrive without cobalamin [16, 165]. Overall, the role and importance of MCM in phytoplankton is unclear, but our data suggest that *T. pseudonana* use MCM and experienced a decreased efficiency of the pathway under cobalamin limitation.

In contrast to *T. pseudonana*, we do not see a statistically significant increase in propionylcarnitine or butrylcarnitine in *N. pelliculosa* under cobalamin limitation, though the general trends are the same (Figures 4.12 and 4.13). This could be a result of a lack of MCM used by *N. pelliculosa*, or an alternate pathway that is able to perform the same function as MCM in the absence of cobalamin (analogous to MetE's replacement of MetH). All of the sequenced diatoms code for MCM [46, 133], but with such a small number of genomes available, it is still possible that *N. pelliculosa* is an exception and does not code for this enzyme. Nevertheless, *N. pelliculosa* does not appear to experience the same disruption of MCM under cobalamin limitation as *T. pseudonana*.

4.4.7 Conclusions

Our work demonstrated that the pools of many metabolites in diatoms are affected by the availability of cobalamin. Many of the changes we observed could be directly linked to the role that cobalamin plays as a coenzyme in MetH, including an imbalanced methionine cycle, altered transsulfuration activity, and rearrangement of osmolyte pools. Our untargeted approach that two revealed acylcarnitines increased under cobalamin limitation, which we propose to be directly linked to the Ado-cobalamin dependent MCM. Overall, the affected metabolite pools were more dramatically changed in the cobalamin-dependent *T. pseudonana* than in the cobalamin-independent *N. pelliculosa*, supporting overall growth rate changes. In some areas of the natural environment, diatoms likely experience ephemeral or sustained cobalamin limitation and this work revealed many metabolite pools that may be affected by the availability of this organic micronutrient.

4.5 Methods

4.5.1 Culture maintenance

Two diatoms were chosen for laboratory studies, *Thalassiosira pseudonana* CCMP1335 and *Navicula pelliculosa* CCMP 543, both from National Center for Marine Algae and Microbiota (NCMA, `ncma.bigelow.org`). Both cultures were maintained at 13°C with a 12 hr light:dark cycle (light levels reported in Table 4.1)in 50 mL combusted borosilicate tubes with 35 mL filter sterilized artificial seawater media (41 g instant ocean salts in 1 L MilliQ water with f/2 nutrients added, except cobalamin). *T. pseudonana* requires cobalamin for growth, while *N. pelliculosa* does not require cobalamin [15, 77]; therefore, *T. pseudonana* was maintained in 1 pM OH-cobalamin, while *N. pelliculosa* was maintained on cobalamin-free media. Both species were maintained for at least 4 transfers on maintenance media before starting the experiment and shown to be axenic via marine purity test broth [98].

4.5.2 Experimental conditions

Using a single inoculum, we inoculated nine 35-mL cultures under each of the following conditions: replete cobalamin, saturating light; replete cobalamin, low light; low cobalamin, saturating light; low cobalamin, low light. For all media, we also prepared a media blank (uninoculated media treated the same as the samples). Light was manipulated using a neutral density photographic lighting film (Lee Filters, Burbank, CA, USA), and photosynthetically available radiation (PAR) was measured with a LI-250A Light Meter (LI-COR Biosciences, Lincoln, NE, USA). Light levels and intial cobalamin concentrations for each treatment and organisms are given in Table 4.1.

We monitored growth of these four conditions by relative fluorescence over the course of 12 days, taking chlorophyll *a* fluorescence measurements at approximately 11 am each day with a Turner Designs model 10-AU fluorometer (*in vivo* chlorophyll optical kit). To calculate growth rate, we used days 3–9 for *N. pelliculosa*, days 1–5 for high light treatments of *T. pseudonana*, and days 3–7 of low light treatments of *T. pseudonana* using *in vivo*

fluorescence ($n = 9$ for each treatment). Using a combusted glass apparatus and gentle vacuum filtration, we harvested two 35 mL cultures onto one 47 mm 0.2 μm PTFE filters when the cultures were in exponential phase (on days noted on Figure 4.1). All cultures were harvested at approximately 11 AM. Throughout exponential growth, we took 1 mL samples (fixed with 1% formaldehyde) for cell counts. We used a Beckman Coulter Z2 Particle Count and Size Analyzer (Beckman Coulter) to measure *T. pseudonana* densities and diameters. *N. pelliculosa* is a colonial species of diatom, so we determined cellular density with the inverted microscope technique as previously described [166]. For each species at each light level, we determined the *in vivo* fluorescence vs cell concentration ($R^2 > 0.9$ for all) and converted our fluorescence at harvest to cell density on filters. Between treatments, we did not observe any significant differences in the size of *T. pseudonana* or *N. pelliculosa*, so we used cell counts as a proxy for biomass for each species for normalization (see data processing section).

4.5.3 Metabolite extraction

Filters of samples and media blanks were extracted as described in detail in Boysen *et. al.* (2018), which resulted in a more non-polar organic fraction and a highly polar aqueous fraction. To aid in normalization, we added a cocktail of internal standards (listed in Table 4.7) before and after extraction.

4.5.4 Data acquisition

To obtain metabolomic data, we separated compounds using reverse phase (RP, for the organic and aqueous fractions) and hydrophilic interaction liquid chromatography (HILIC, for aqueous fraction only), using the exact specifications as previously described [8]. After extraction, samples were run within 24 hours for HILIC and within 96 hours for RP; between extraction and analysis, samples were stored at -80°C. We obtained targeted and untargeted metabolomics data on triple quadrupole (TQS) and Q-Exactive (QE) mass spectrometers (MSs), respectively. Full scan (non fragmented mass to charge (m/z)) data acquisition from the QE and all data acquisition from the TQS are described fully in Boysen *et. al.* (2018).

Throughout the run, we ran a pooled sample several times in order to monitor signal stability and train normalization.

For data collected for the untargeted analysis, we collected fragmentation spectra (MS^2) on our pooled samples. For samples separated via HILIC, separate injections were analyzed in positive and negative ion modes analyzed by data dependent acquisition (DDA). MS^2 spectra were collected from *m/z* 50–500 at resolution of 30,000 at *m/z* 200. For each MS1 scan, DDA was set to perform on the top five most abundant ions with dynamic exclusion of 20 seconds using a normalized collision induced dissociation at 35 V. For RP, samples were only run in positive mode; DDA was performed in the same manner but with a dynamic exclusion time of 10 seconds due to narrower peaks than in HILIC.

4.5.5 Data processing

Targeted data

Targeted data (from the TQS) were subject to an in-house quality control, blank subtracted, and normalized via best-matched internal standard (B-MIS) normalization [8], using a 20% improvement to the relative standard deviation (RSD) of each mass feature in a pooled sample as a criteria to apply normalization, that is, $(RSD_{final} - RSD_{raw}) / RSD_{raw}) < 0.20$. We arrived at the 20% improvement criteria using tools in the B-MIS normalization package [8]. Like our previous work [8], we did not normalize any mass features that had a RSD_{raw} of $< 10\%$ in the raw pooled area, instead defaulting to the raw area. This normalization process resulted in adjusted peak areas that minimize obscuring variation—the non-biological variation inherent to LC-MS analyses introduced during analysis. We then normalized the adjusted peak areas to cell counts of each biological replicate to account for different amount of cellular material on each filter. For subsequent analyses, the normalized adjusted peak areas were used. For univariate and multivariate statistical analyses (see sections below), we only included compounds that were detected in at least two treatments. For our targeted analysis, when targeted compounds were detected in only a subset of the treatments or

replicates (or small peaks were removed during the quality control step), we assigned a value for the remaining treatments or replicates that represents an upper estimate of how large a peak could be present and still remain below our detection limit ($3\times$ *peak area in blank*$+100$) . This value underwent the same normalization (B-MIS and to cell counts).

Untargeted data

Data collected for MS and MS^2 from the QE were first converted to .mzxml using MSConvert [167] with positive and negative scans processed separately. For MS data, each fraction (organic on RP, aqueous on RP, aqueous on HILIC (positive scans), and aqueous on HILIC (negative scans)) were processed separately through XMCS [168, 169, 170] with parameters optimized via Isotopologue Parameter Optimization [171] (Table 4.8). For each sample, this resulted in a list of peak areas of mass features (peaks with unique retention time (RT) and m/z). We filtered out any mass features with an average peak area less than three times larger than the methodological blank or with an average peak size less than 100, based on visual examination. Next we normalized for obscuring variation using B-MIS normalization [8], using the same criteria as in our targeted analyses. We calculated the coefficient of variation (CV) of each mass feature in injections of the pooled sample and removed peaks that did not demonstrate acceptable replicability (CV $> 30\%$).

Next, we identified mass features that were likely ^{13}C, ^{15}N, or ^{34}S isotopologues of other mass features because we did not want to give these compounds extra weight in our statistical analyses. To do this, we searched for expected differences in m/z and intensity between mass features for each of the aforementioned isotopes within each three second (for RP) or six second (for HILIC) corrected retention time window. We excluded these ^{13}C, ^{15}N, or ^{34}S mass features from downstream analyses.

Like in the targeted analysis, we normalized the adjusted peak areas of quality peaks to cell counts of each replicate to account for different amount of cellular material on each filter, and for subsequent analyses, the normalized adjusted peak areas of the filtered mass features were used.

MS^2 data processing

For quality mass features, we searched the .mzXML files collected for DDA analysis for MS^2 scans that match the parent *m/z* (at 0.5 Da tolerance) and retention time (at 10 and 20 second tolerance for RP and HILIC chromatography, respectively). The isolation chamber of the QE is low resolution, so it is possible that a parent scan will match at 0.5 Da but the fragmentation spectra present is from a different parent *m/z* with the same nominal mass—this is especially problematic for low abundance peaks. Therefore, for each matched MS^2 scan (at 0.5 Da), we filtered out any scans where the parent *m/z* was not present in the high-resolution fragmentation spectra at 0.02 Da tolerance. If multiple scans were found, we summed the scans' intensities for each fragment across the scans as fragments' intensities can vary between scans. We filtered out peaks that contributed less than 0.5% of the intensity of the most intense fragment.

Compound identification

Using the ranking system outlined in Sumner *et. al.* (2007), we attempted to identify the quality mass features present in our sample sets. First we removed any mass features that were plausible contaminants by searching for common contaminants [172]. For the remaining quality mass features, we searched an internal database of standards run in the exact manner on our instruments for matches to exact *m/z* and retention time, yielding an *unequivocal identification* (confidence level 1); for current list of the standards run in this exact manner see `https://github.com/kheal/Example_Untargeted_Metabolomics_Workflow/blob/master/Ingalls_Lab_Standards.csv`).

For mass features without a standard match but with MS^2, we searched against the publicly available LC/MS MS^2 spectral databases, including MassBank [173], Global Natural Products Social Molecular Networking [174], MetiTree [175], RIKEN tandem mass spectral database [176], and the Human Metabolome Database [177] (all downloaded from `http://mona.fiehnlab.ucdavis.edu/downloads` on November 17, 2017). We searched for spectra

with matching parent m/z and a cosine similarity > 0.8 from an ESI spectrum, using the same algorithms as MassBank [173]. When a compound in the MassBank database matched m/z and MS^2, we assigned a *putative identification* (confidence level 2). For some of these compounds, we later obtained standards and upgraded their identification to confidence level 1. For mass features without confidence levels 1 or 2 identification, we searched for possible matches to compound in the KEGG database based only on m/z, these identifications are considered *possible identification* (confidence level 3). Our extraction analysis resulted in both a highly polar aqueous and a more non-polar organic fraction, which were analyzed separately. Although there are many more mass features in the organic fraction (Table 4.8), most of the primary metabolites are extracted into the aqueous fraction [8], and we focused our untargeted analysis on this fraction to understand the primary metabolic response to cobalamin limitation. For many mass features, the aforementioned identification attempts were not fruitful, so we attempted to identify quality mass features that showed a univariate response to cobalamin limitation from our aqueous extraction, using a manual approach with the larger, but not publically available, Metlin database [178]. We searched mass features by m/z (assuming M+H, M+NH_4, or M+Na for positive ionization, M-H, M+Cl for negative ionization, at a 5 ppm tolerance) and compared MS^2 when obtained by DDA and available in the database. We assigned putative identifications using the same confidence levels as discussed above.

4.5.6 Data analysis

Multivariate statistics

For all multivariate statistics, we used data that were standardized to z-scores for each compound or mass feature, where $z = (X - \mu)/\sigma$, where z is the z-score, X is the adjusted normalized peak area in each replicate, μ is the mean peak area, and σ is the standard deviation of the peak areas across each sample set (separated by organism). We employed two multivariate analyses to analyze changes in the metabolomes of *T. pseudonana* and *N.*

pelliculosa. We used a non-metric dimensional scaling (NMDS) approach [179] based on a euclidean distance to analyze differences in the treatments. We chose a non-metric approach due to our low sample numbers, high variable numbers (at least 58 compounds), and the non-normal nature of our dataset to avoid overfitting common in metric approaches of ordination in metabolomics [180]. We assessed dimensionality of the NMDS by examining a scree plot and calculated the probability with a monte carlo permutation. We paired this NMDS with an analysis of similarities (ANOSIM) to determine whether our metabolite data could parse out differences between light and cobalamin-stressed diatoms, using 999 permutations. All multivariate statistics were performed in R using the vegan package (V2.4-2).

Univariate statistics

We used an unpaired t-test to compare growth rates between experimental treatments. For targeted and untargeted analyses, compounds (from our targeted analysis) and mass features (from our untargeted analysis) were investigated for significant differences between low and replete cobalamin treatments and low and high light treatments in each organism. For all univariate statistics on metabolomes, we corrected p values for false discovery rate [181] and report those corrected values throughout the text. We calculated fold changes and p-values (via an unpaired t-test) between low and replete cobalamin treatments (regardless of light status) and low and high light treatments (regardless of cobalamin status, $n = 6$). Finally, we calculated fold changes and p values between cobalamin treatments within each light status and between light treatments for each cobalamin status ($n = 3$ for each treatment). For mass features in the untargeted data, we visually examined peaks that had resulted in significant p values between cobalamin treatments and removed any mass features that did not have quality peak shapes or integrations from downstream analyses.

4.5.7 Gene searches

To corroborate our metabolomics analysis we performed a blastp search on the *T. pseudonana* genome [182] for genes related to transsulfuration, polyamine biosynthesis, methionine

salvage, and propionylcarnitine metabolism, using e-value threshold 10^{-80}. To gain an understanding if these pathways were well distributed among diatoms, we also searched publicly available genomes from *Phaeodactylum tricornutum* [183], *Fragilariopsis cylindrus*, *Thalassiosira oceanica* [184], and *Fistulifera solaris* [185].

4.6 Acknowledgements

The authors would like to acknowledge B. Durham, J. Young, R. Lundeen for input on the manuscript; E. Armbrust and G. Rocap for input on experimental design. This work was supported by grants from the Simons Foundation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I.), NSF OCE-1228770 and OCE-1205232 to A.E.I., NSF GRFP to K.R.H.

4.7 Figures

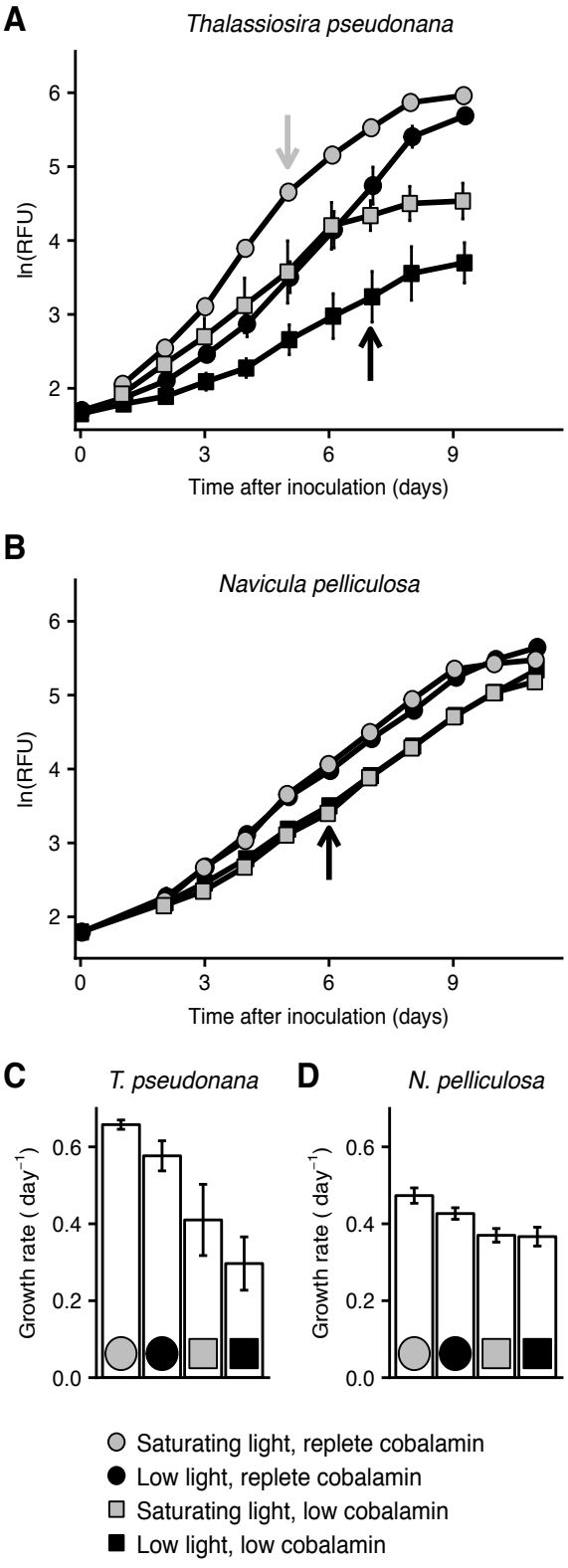

Figure 4.1: Caption on following page.

Figure 4.1. *Continued from previous page.* Growth curves of *T. pseudonana* (A) and *N. pelliculosa* (B) under the four experimental conditions. *T. pseudonana* grown under saturating light were harvested at grey arrow. *T. pseudonana* grown under low light and all *N. pelliculosa* cultures were harvested at black arrows. RFU = relative fluorescence units, error bars are standard deviation and are often smaller than the markers ($n = 9$ for all pre-harvest time points, $n = 3$ for post-harvest time points). Average growth rate with standard deviation for *T. pseudonana* (C) and *N. pelliculosa* (D), $n = 9$. For each species, growth rates were significnatly different between all treatments (t-test, $p < 0.05$), except *N. pelliculosa* low light and saturating light at low cobalamin, which were not statistically different.

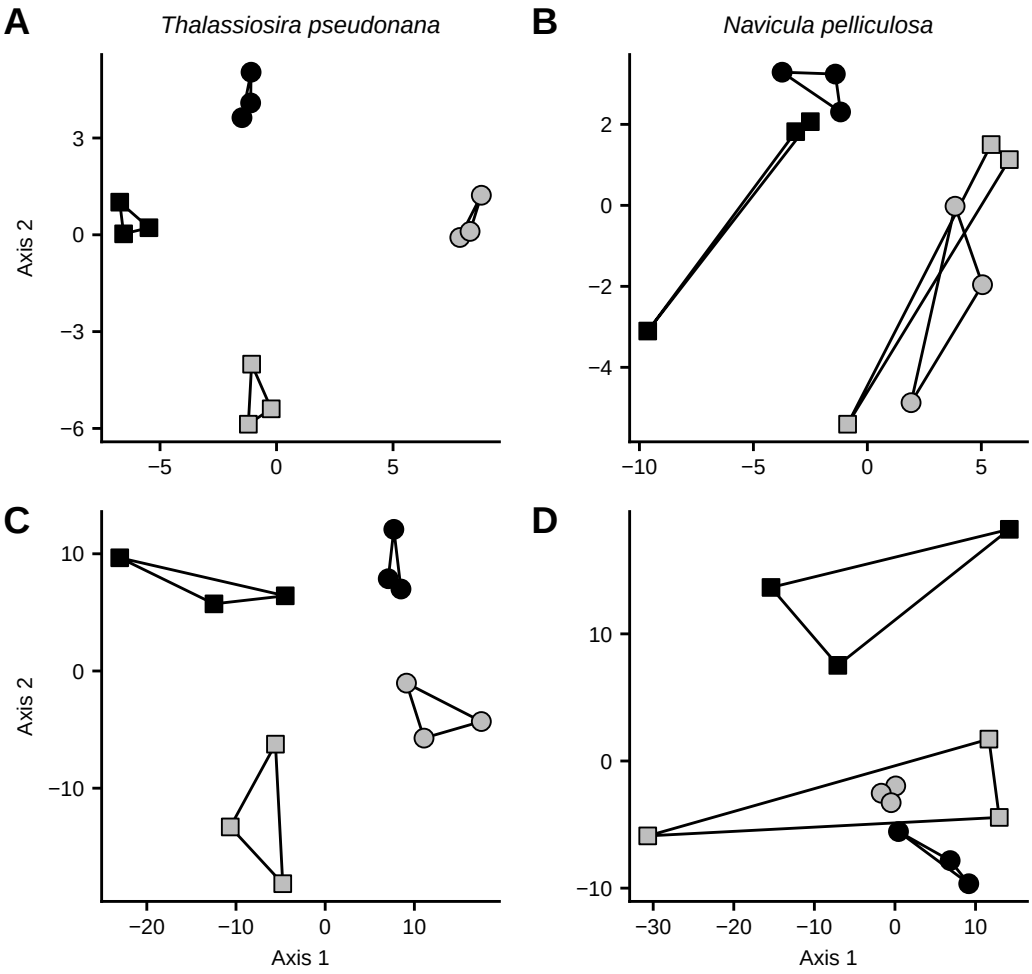

Figure 4.2: Results from NMDS analysis in targeted and untargeted analyses show clear differentiation between treatments in *T. pseudonana*, less so for *N. pelliculosa*. Panels A and B show the locations of the samples in NMDS space for targeted analyses; panels C and D are the same analysis on the 500 largest mass features from the untargeted analyses. Treatment markers are as in Figure 4.1. For panels A, B, C, D, stresses were 0.026, 0.085, 0.083, 0.053, respectively; and p values < 0.05 for all.

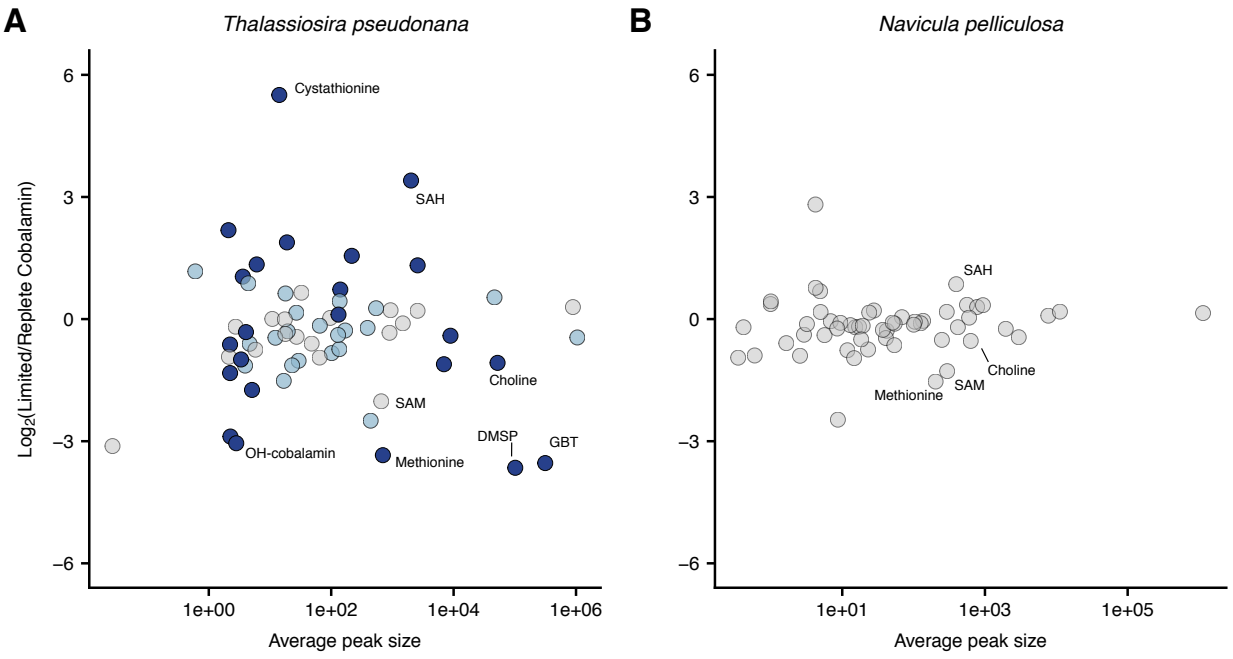

Figure 4.3: Results from the targeted analyses for *T. pseudonana* and *N. pelliculosa*. In each panel, every compound detected with the targeted analysis is depicted as a dot; light blue compounds are significantly different between cobalamin treatments in either saturating or low light conditions, blue compounds are significantly different under both light conditions ($p < 0.05$). Y-axis is $\log_2$(average peak size of replete cobalamin treatments / average peak size of replete cobalamin treatments). X-axis is average peak area of compound after adjustment and normalization; note that x-axis is log scaled. Specific compounds of discussion are labeled.

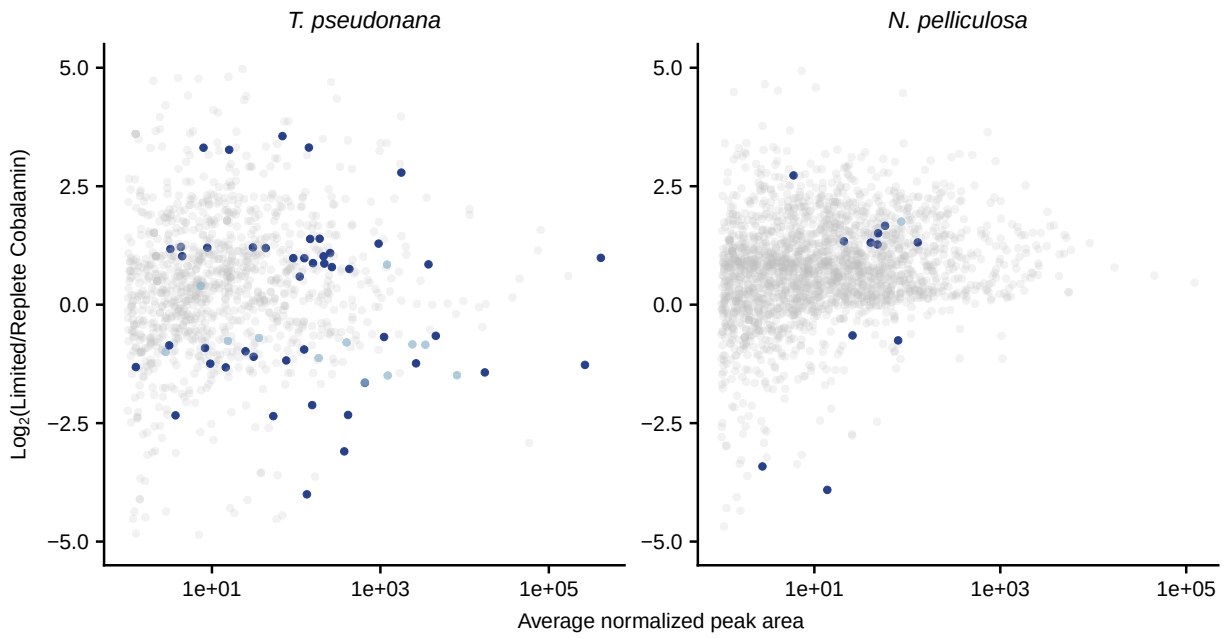

Figure 4.4: Overall results from the untargeted analysis from the aqueous extraction, highlighting a cobalamin limitation signal in individual mass features. Each mass feature is depicted as a dot; light blue features are significantly different between cobalamin treatments in either saturating or low light conditions, dark blue mass features are significantly different under both light conditions ($p < 0.05$). Y-axis is $\log_2$(average peak size of replete cobalamin treatments / average peak size of replete cobalamin treatments), regardless of light status. X-axis is average peak area of mass feature, note that x-axis is log scaled.

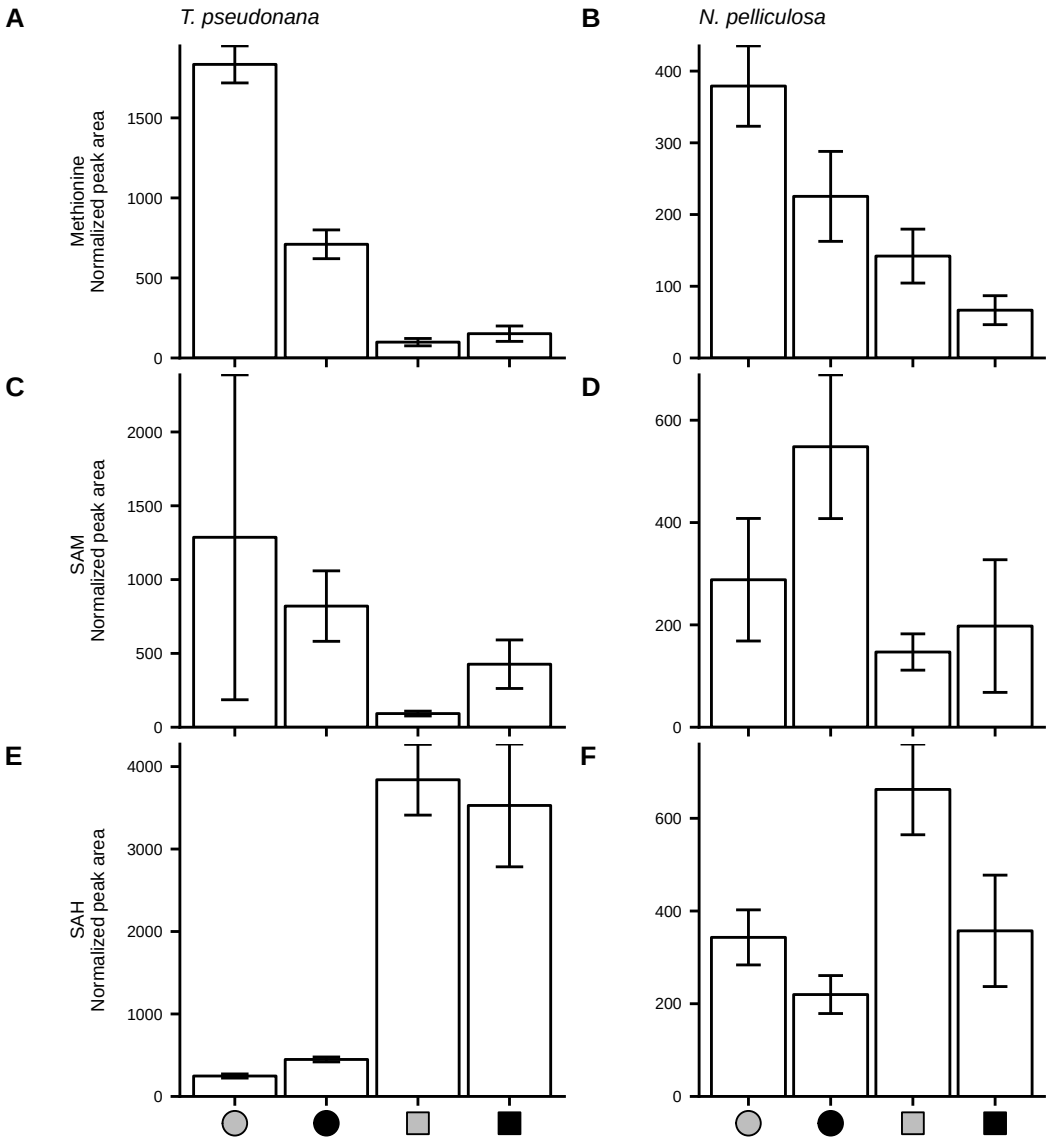

Figure 4.5: Normalized peak areas of methionine (top row), SAM (middle row), and SAH (bottom row) in *T. pseudonana* (A, C, and E) and *N. pelliculosa* (B, D, and F), summarized by each of the growing conditions (labeled the same as Figure 4.1). Error bars are standard deviation with $n = 3$. All these data were obtained in our targeted analysis.

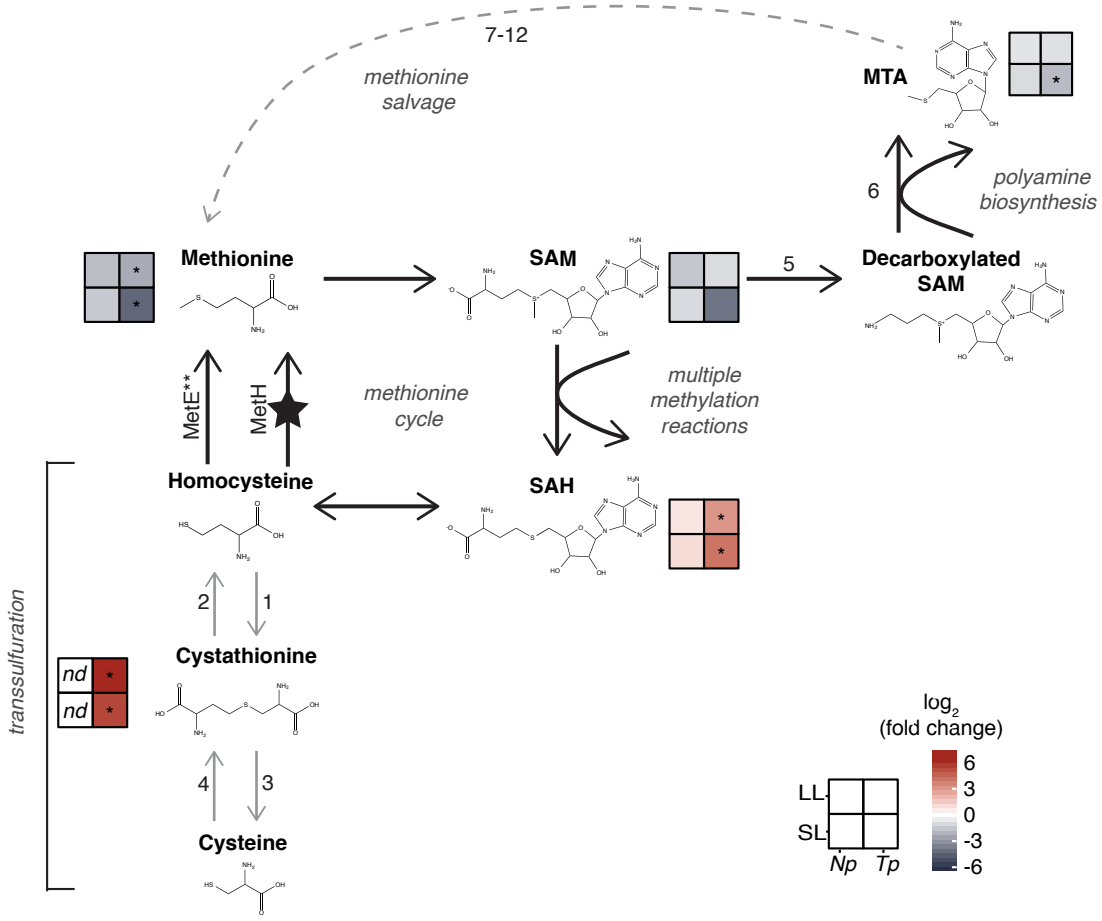

Figure 4.6: Changes in the methionine cycle and related pathways as observed in metabolite concentrations. Log_2(fold change) in metabolite concentrations are shown between cobalamin treatments in high light (HL) and saturating light (SL) conditions and *T. pseudonana* (*Tp*) and *N. pelliculosa* (*Np*) as noted in the key. Significant differences in concentrations between cobalamin treatments in individual light treatments are designated with an asterisk ($n = 3$, $p < 0.05$). Reactions with corresponding enzymes that are well described in diatoms are shown in black, and reactions with known enzymes but without strong genomic evidence in diatoms, numbers correspond to enzymes reported in Table 4.5. Compounds that were not detected are noted with *nd*. Note that the methionine salvage pathway consists of multiple enzymes (dashed arrow). **MetE is only present in some diatoms; it is likely in *N. Pelliculosa* (see Discussion section), but certainly is not encoded in *T. pseudonana*.

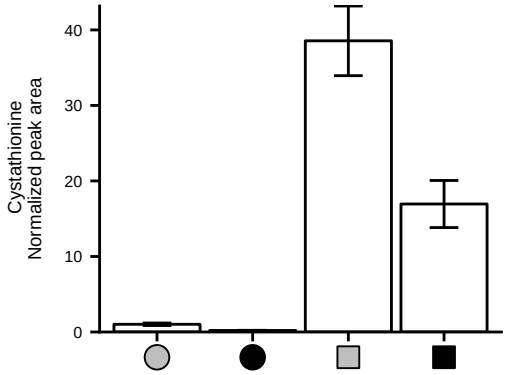

Figure 4.7: Normalized peak areas of cystathionine in *T. pseudonana* summarized by each of the growing conditions (labeled the same as Figure 4.1). Error bars are standard deviation with $n = 3$. We detected cystathionine in our targeted analysis.

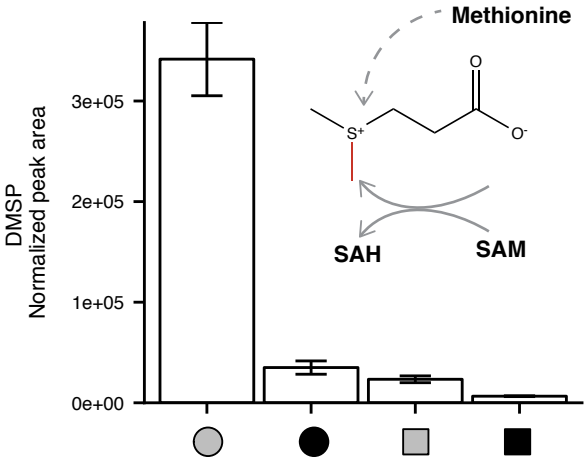

Figure 4.8: Normalized peak areas of DMSP in *T. pseudonana* summarized by each of the growing conditions (labeled the same as Figure 4.1). Inset shows the structure of DMSP and its connection to the methionine cycle, where it is directly derived from methionine and is methylated via SAM (red methyl groups). Error bars are standard deviation with $n = 3$.

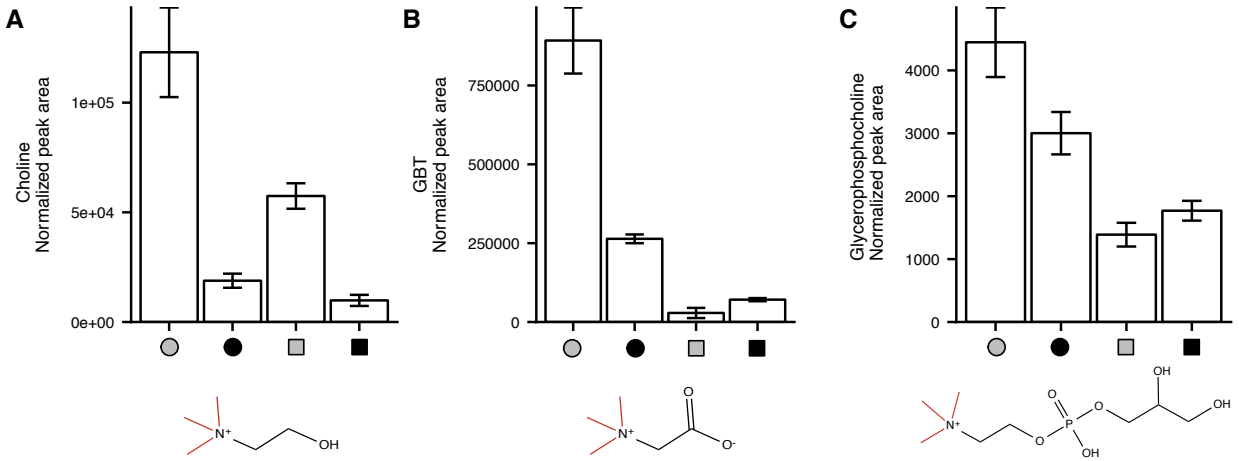

Figure 4.9: Normalized peak areas of choline (A), GBT (B), and glycerophosphocholine (C) *T. pseudonana* summarized by each of the growing conditions (labeled the same as Figure 4.1), with structures. In each structure, the red methyl groups are likely directly from SAM. Error bars are standard deviation with $n = 3$. Data shown for GBT and DMSP are from our targeted analysis; glycerophosphocholine was detected in our untargeted analysis.

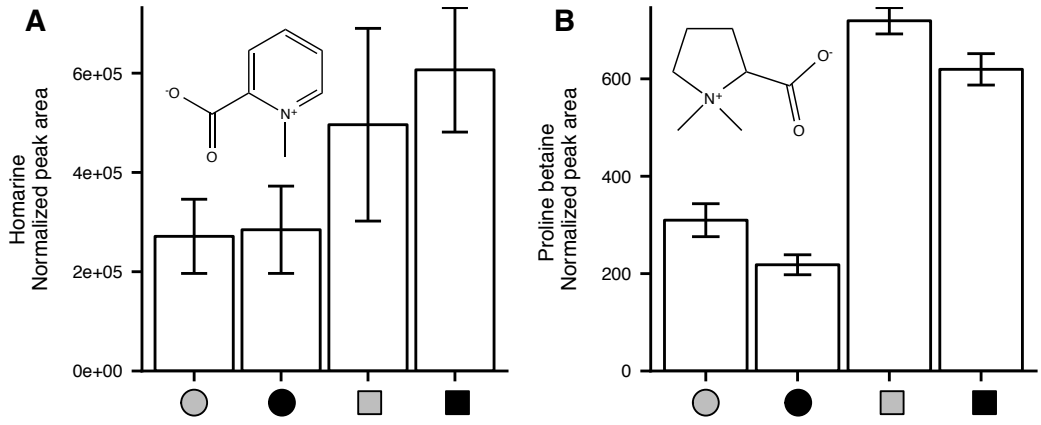

Figure 4.10: Normalized peak areas of homarine (A) and proline betaine (B) in *T. pseudonana* summarized by each of the growing conditions (labeled the same as Figure 4.1), with structures in insets. Error bars are standard deviation with $n = 3$. Theses compounds were detected in our untargeted analysis.

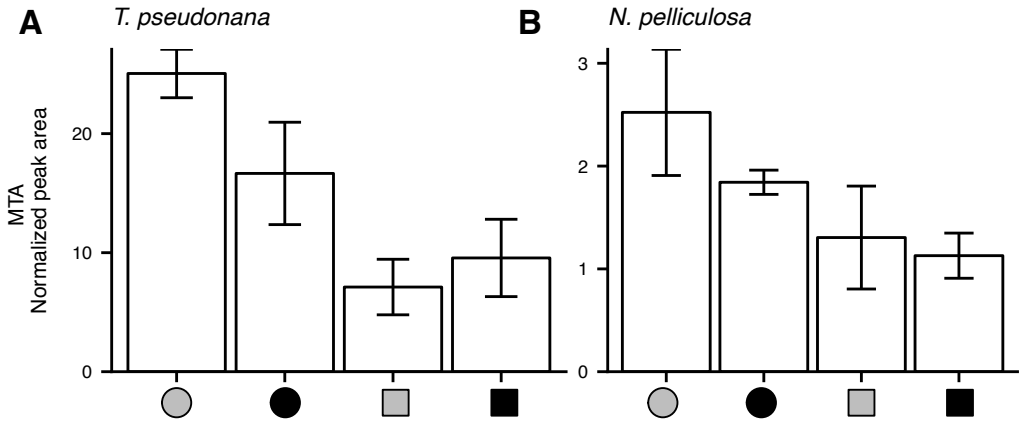

Figure 4.11: Normalized peak areas of MTA in *T. pseudonana* (A) and *N. pelliculosa* (B), summarized by each of the growing conditions (labeled the same as Figure 4.1). Error bars are standard deviation with $n = 3$. MTA was detected in our untargeted analysis.

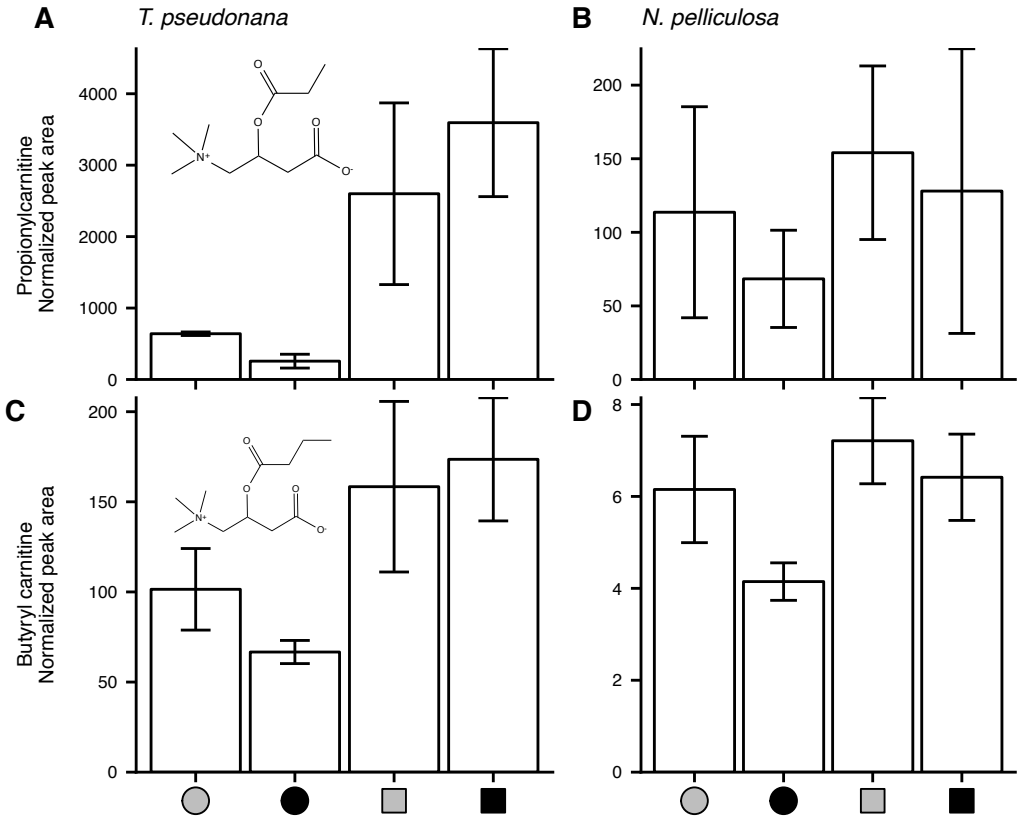

Figure 4.12: Normalized peak areas of propionylcarnitine (A, B) and butyrylcarnitine (C, D) in *T. pseudonana* and *N. pelliculosa*, summarized by each of the growing conditions (labeled the same as Figure 4.1, with structures). Error bars are standard deviation with $n = 3$. We detected these compounds in our untargeted analysis and confirmed the structures with authentic standards.

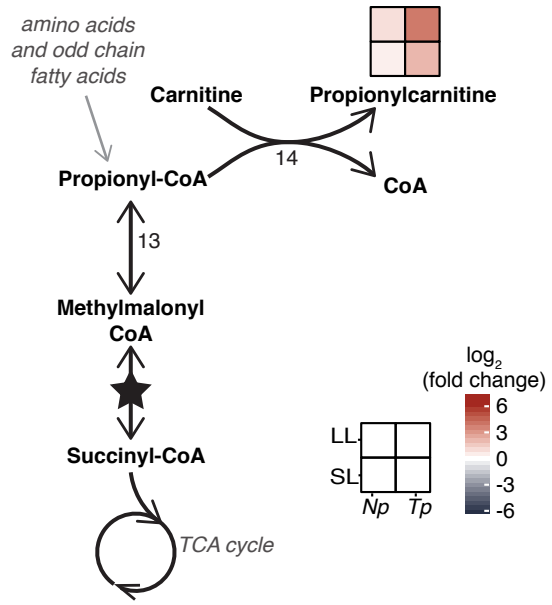

Figure 4.13: Accumulation of propionylcarnitine and our proposed mechanism in relation to MCM decreased efficiency based on Hannibal *et. al.* (2016). Log_2(fold change) in metabolite concentrations are shown between cobalamin treatments in saturating light (SL) and low light (LL) conditions and *T. pseudonana* (*Tp*) and *N. pelliculosa* (*Np*) as noted in the key. The enzymatic conversion that is cobalamin-dependent (MCM) is starred. Numbers next to conversions correspond to genes reported in Table 4.5.

4.8 Tables

Organism	Treatment	Light level (µmol photons m^{-2} sec^{-1})	Cobalamin (pM)	µ ± sd (day^{-1})	*n*
T. pseudonana	Saturating light, replete cobalamin	120	200	0.66 ± 0.01	9
	Saturating light, low cobalamin	120	1	0.41 ± 0.09	9
	Low light, replete cobalamin	50	200	0.58 ± 0.04	9
	Low light, low cobalamin	50	1	0.30 ± 0.07	9
N. pelliculosa	Saturating light, replete cobalamin	120	200	0.47 ± 0.02	9
	Saturating light, low cobalamin	120	0	0.37 ± 0.02	9
	Low light, replete cobalamin	50	200	0.43 ± 0.01	9
	Low light, low cobalamin	50	0	0.36 ± 0.03	9

Table 4.1: Experimental conditions and growth rates (μ) for *T. pseudonana* and *N. pelliculosa* under the four experimental conditions.

Compound	*p*	Organism
Adenosine	0.001	*T. pseudonana*
Adenosyl Homocysteine	0.0015	*T. pseudonana*
Adenosyl Methionine	0.0059	*T. pseudonana*
Alanine	2.00E-04	*T. pseudonana*
AMP	0.0074	*T. pseudonana*
Arachidonic Acid	0.0094	*T. pseudonana*
Asparagine	1.00E-04	*T. pseudonana*
Aspartic acid	1.00E-04	*T. pseudonana*
Betaine	5.00E-04	*T. pseudonana*
cGMP	6.00E-04	*T. pseudonana*
Chitobiose	0.0055	*T. pseudonana*
Choline	1.00E-04	*T. pseudonana*
Citric Acid	0.042	*T. pseudonana*
Coenzyme B12	1.00E-04	*T. pseudonana*
Cystathionine	2.00E-04	*T. pseudonana*
Cytosine	8.00E-04	*T. pseudonana*
DHA	0.0112	*T. pseudonana*
DHAP	0.0231	*T. pseudonana*
DHPS	0.002	*T. pseudonana*
Dimethyl Glycine	8.00E-04	*T. pseudonana*
DMSP	1.00E-04	*T. pseudonana*
EPA	0.0011	*T. pseudonana*
Ergosterol	5.00E-04	*T. pseudonana*
FAD	0.0034	*T. pseudonana*
Fructose 6 phosphate	1.00E-04	*T. pseudonana*
Gluconic Acid	0.0192	*T. pseudonana*
Glucose 6 phosphate	2.00E-04	*T. pseudonana*
Glutamic acid	1.00E-04	*T. pseudonana*
glycerol 3 phosphate	3.00E-04	*T. pseudonana*
GMP	5.00E-04	*T. pseudonana*
Guanine	1.00E-04	*T. pseudonana*
Guanosine	4.00E-04	*T. pseudonana*
Heavy Threonine	0.0027	*T. pseudonana*
Homarine	1.00E-04	*T. pseudonana*
Hydroxo B12	2.00E-04	*T. pseudonana*
Kynurenine	2.00E-04	*T. pseudonana*
Lysine	0.0017	*T. pseudonana*
Malic Acid	1.00E-04	*T. pseudonana*
Methionine	1.00E-04	*T. pseudonana*
NAD	3.00E-04	*T. pseudonana*
NADP	1.00E-04	*T. pseudonana*
Orotic Acid	0.0376	*T. pseudonana*
Picolinic Acid	0.0299	*T. pseudonana*
Proline	1.00E-04	*T. pseudonana*
Pyridoxal Phosphate	0.0402	*T. pseudonana*
Serine	7.00E-04	*T. pseudonana*
Thiamine pyrophosphate	4.00E-04	*T. pseudonana*
Threonine	3.00E-04	*T. pseudonana*
trans Hydroxyl proline	6.00E-04	*T. pseudonana*
Tryptophan	2.00E-04	*T. pseudonana*
Tyrosine	3.00E-04	*T. pseudonana*
UDP-glucose	0.0424	*T. pseudonana*
Uridine	0.0053	*T. pseudonana*
Vitamin B2	3.00E-04	*T. pseudonana*

Compound	*p*	Organism
Adenine	0.0087	*N. pelliculosa*
Alanine	3.00E-04	*N. pelliculosa*
AMP	1.00E-04	*N. pelliculosa*
Asparagine	0.0055	*N. pelliculosa*
Aspartic acid	0.0017	*N. pelliculosa*
Choline	0.0019	*N. pelliculosa*
Citric Acid	0.036	*N. pelliculosa*
EPA	0.0079	*N. pelliculosa*
Ergosterol	0.0107	*N. pelliculosa*
Glucosylglycerol	0.0019	*N. pelliculosa*
Glutamic acid	5.00E-04	*N. pelliculosa*
Glutamine	0.0019	*N. pelliculosa*
GMP	0.0302	*N. pelliculosa*
Guanosine	0.0425	*N. pelliculosa*
Hydroxo B12	0.0145	*N. pelliculosa*
Isoleucine	0.034	*N. pelliculosa*
Lysine	0.0015	*N. pelliculosa*
Orotic Acid	0.0422	*N. pelliculosa*
PEP	0.0062	*N. pelliculosa*
Phenylalanine	7.00E-04	*N. pelliculosa*
Proline	0.0032	*N. pelliculosa*
Pyridoxal Phosphate	0.005	*N. pelliculosa*
Serine	0.0047	*N. pelliculosa*
Thiamine monophosphate	0.0026	*N. pelliculosa*
thiamine pyrophosphate	0.0166	*N. pelliculosa*
Threonine	1.00E-04	*N. pelliculosa*
trans Hydroxyl proline	0.0032	*N. pelliculosa*
Tryptophan	5.00E-04	*N. pelliculosa*
Tyrosine	0.0032	*N. pelliculosa*
UDP-glucosamine	0.0473	*N. pelliculosa*
UDP-glucose	1.00E-04	*N. pelliculosa*
Uridine	0.0024	*N. pelliculosa*
Valine	0.0022	*N. pelliculosa*
Vitamin B2	1.00E-04	*N. pelliculosa*
Xanthine	1.00E-04	*N. pelliculosa*

Table 4.2: Compounds that significantly contributed ($p < 0.05$) to the NMDS analysis of *T. pseudonana* and *N. pelliculosa* targeted metabolomes.

	Variable	ANOSIM Statistic	*p*	Analysis
T. pseudonana	Light	0.602	0.003	Targeted
	Cobalamin	0.507	0.006	Targeted
	Light	0.380	0.008	Untargeted
	Cobalamin	0.685	0.004	Untargeted
N. pelliculosa	Light	0.657	0.003	Targeted
	Cobalamin	0.187	0.064	Targeted
	Light	-0.078	0.876	Unargeted
	Cobalamin	0.324	0.005	Unargeted

Table 4.3: Results from ANOSIM analysis in *T. pseudonana* and *N. pelliculosa*. ANOSIM statistic and p values are given for differentiating metabolomes depending on light or cobalamin in the targeted and untargeted analyses.

Organism	Analysis		Quality Mass Features	% of Mass Features with Cobalamin Signal
T. pseudonana	Untargeted	Aqueous	3348	2.81
		Organic	9596	4.23
	Targeted		63	36.51
N. pelliculosa	Untargeted	Aqueous	3107	0.45
		Organic	7915	0.11
	Targeted		55	0

Table 4.4: Summarized results of targeted and untargeted metabolomics analyses. Percent of the mass features or compounds that showed a difference ($p < 0.05$ between the cobalamin treatments).

Pathway	Gene	T. pseudonana	P. tricornutum	F. cylindrus	T. oceanica	F. solaris
Transsulfuration Pathway	cystathionine beta-synthase (1)	?	+	?	+	?
	cystathionine gamma-synthase (2)	?	+	?	?	?
	cystathionine gamma-lyase (3)	+	+	nh	+	nh
	cystathionine beta-lyase (4)	?	+	?	?	?
Polyamine synthesis	S-adenosylmethionine decarboxylase (5)	+	+	+	+	+
	spermidine synthase (6)	+	+	+	+	+
Methionine salvage	MTA/SAH nucleosidase (7)	+	+	+	+	+
	S-methylthioribose kinase (8)	?	+	?	?	?
	methylthioribose-1-phosphate isomerase (9)	+	+	+	+	+
	methylthioribulose-1-phosphate dehydrase (10)	?	?	?	?	?
	enolase-phosphatase E1 (11)	?	?	?	nh	?
	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (12)	?	?	?	nh	?
Propionylcarnitine metabolism	propionyl-CoA carboxylase (7)	+	+	+	+	+
	carnitine acyltransferase (7)	+	+	+	+	+

Table 4.5: Summarized results from gene searches in *T. pseudonana* and other publicly available diatom genomes. We search genes related to transsulfuration, polyamine synthesis, methionine salvage, and propionylcarnitine synthesis as discussed in the text. Code for presence of genes in genomes: +, very likely present with E value $< 10^{-80}$; ?, gene possibly present, but bad E value or coverage; nh, no reasonable hit. Full results of gene searches including queries, E values, and accession numbers are found in Dataset S4.5. Numbers in parentheses correspond to Figures 4.6 and 4.13.

m/z	RT (min)	Column, polarity	Log_2 (fold change) SL	LL	Best identification [ion]	Confidence	Evidence
385.1288	10.55	HILIC, positive	3.364	3.248	S-adenosyl homocysteine [M+H]	1	*m/z*, RT, standard
218.1386	6.97	HILIC, positive	2.021	3.808	Propionyl-L-carnitine [M+H]	1	*m/z*, MS^2, RT, standard
175.1189	18.42	HILIC, positive	1.796	0.779	Arginine [M+H]	1	*m/z*, RT, standard
205.0973	1.95	RP, positive	1.377	1.411	Tryptophan [M+H]	1	*m/z*, RT, standard
584.3941	6.23	HILIC, positive	1.318	0.671		4	
144.102	7.3	HILIC, positive	1.215	1.506	Proline betaine [M+H]	1	*m/z*, RT, standard
147.0996	2.38	RP, positive	1.104	1.320		4	
149.1093	2.37	RP, positive	1.081	1.317		4	
166.1387	2.37	RP, positive	1.078	1.280		4	
166.1359	2.37	RP, positive	1.075	1.334		4	
149.1121	2.37	RP, positive	1.030	1.416		4	
182.0813	0.99	RP, positive	0.990	1.228	Tyrosine [M+H]	1	*m/z*, RT, standard
133.0608	11.37	HILIC, positive	0.904	0.695	Asparagine [M+H]	1	*m/z*, RT, standard
278.1611	6.24	HILIC, positive	0.897	0.786		4	
562.4104	9.46	HILIC, positive	0.885	0.849		4	
138.0548	6.22	HILIC, positive	0.871	1.092	Homarine [M+H]	1	*m/z*, RT, standard
188.0706	8.28	HILIC, positive	0.819	0.972	Tryptophan [2M+H]	1	*m/z*, RT, standard
370.1628	1.07	RP, positive	0.811	1.283		4	
366.0511	9.39	HILIC, positive	0.809	1.225		4	
116.3942	9.46	HILIC, positive	0.742	0.419		4	
232.1544	1.36	RP, positive	0.643	1.381	(iso)Butyryl carnitine [M+H]**	1	*m/z*, MS^2, RT, standard
247.0577	10.24	HILIC, positive	0.455	1.189		4	
148.0603	11.94	HILIC, positive	-0.897	-0.367	Glutamic acid [M+H]	1	*m/z,* RT, standard
490.7438	2.26	RP, positive	-1.005	-0.743		4	
339.2279	0.92	RP, positive	-1.102	-0.705		4	
613.1596	0.91	RP, positive	-1.188	-1.331	Glutathione disulfide [M+H]	1	*m/z*, MS^2, RT, standard
364.0651	13.83	HILIC, positive	-1.436	-0.680	GMP [M+H]	1	*m/z*, RT, standard
120.0657	11.3	HILIC, positive	-1.452	-0.851	Homoserine [M+H]	1	*m/z*, RT, standard
187.1229	6.26	HILIC, positive	-1.469	-2.231	Calligonine [M+H]	3	*m/z*
155.0012	8.14	HILIC, negative	-1.601	-1.251		4	
258.11	11.47	HILIC, positive	-1.679	-0.763	Glycerophosphocholine [M+H]	1	*m/z*, MS^2, RT, standard
298.0969	2.13	RP, positive	-1.817	-0.802	MTA [M+H]	1	*m/z*, MS^2, RT, standard
309.0438	9.85	HILIC, negative	-2.073	-2.178		4	
275.1026	0.84	RP, positive	-2.470	-2.183		4	
178.0715	7.91	HILIC, negative	-2.623	-1.922		4	
278.5725	0.9	RP, positive	-2.697	-1.884		4	
116.0707	7.9	HILIC, negative	-4.543	-2.004	Glycine betaine [M-H]	1	*m/z*, RT, standard

Table 4.6: Mass features with significant differences between cobalamin treatments in *T. pseudonana* that were detected in our untargeted analysis of the aqueous fraction ($p < 0.05$). Identification annotations with compound names and ions (in brackets)observed. Identification confidence based on the rating from Sumner *et. al.* (2007), where confidence level 1 is *unequivocal* and verified with standards, level 2 is *putative* and supported by a match based on MS^2 and *m/z*, level 3 should be considered a *possible* match based on *m/z* in the KEGG database, and level 4 yielded no helpful matches. Log_2(fold change) between cobalamin treatments under saturating light (SL), or low light (LL). Many adducts of mass features in this list also showed significant differences between treatments which are not repeated here, see Dataset S4.3 Table 4 for full results including those adducts with annotations. RT = retention time. **isobutrylcarnitine and butyrlcarnitine are indistinguishable in our analysis, so this mass feature may be either or a combined signal from both.

	Standard	Fraction	Column	z	Inj Conc (nM)	Extracted *m/z*
Extraction Standards	Indole 3 Acetic Acid, D_5	Aqueous	RP	+1	500	181.1021
	Arachidonic Acid, D_8	Organic	RP	+1	2000	313.2977
	Tryptamine, D_4	Aqueous	RP	+1	500	164.1246
	L-Cysteic Acid, D_3	Aqueous	HILIC	-1	1000	171.0161
	Sulfolactic Acid, $^{13}C_3$	Aqueous	HILIC	-1	500	171.9913
	Sulfoacetic Acid, $^{13}C_2$	Aqueous	HILIC	-1	500	140.9774
	DL-Histidine, ^{15}N	Aqueous	HILIC	+1	500	157.0738
	DL-Alanine, D_3	Aqueous	HILIC	+1	5000	93.0738
	L-Tryptophan, D_3	Aqueous	RP	+1	100	208.1161
	L-Isoleucine, ^{15}N	Aqueous	HILIC	+1	500	133.0989
	Isethionic Acid, $^{13}C_2$	Aqueous	HILIC	-1	500	126.9981
	Taurine, D_4	Aqueous	HILIC	-1	500	128.0325
Injection Standards	L-Phenylalanine, D_8	Aqueous	RP	+1	100	174.1365
	Riboflavin-dioxopyrimidine (Vitamin B_2), $^{13}C_4$, $^{15}N_2$	Aqueous	RP	+1	100	383.1531
	L-Methionine, D_3	Aqueous	HILIC	+1	500	153.0722
	Acetyl CoA, $^{13}C_2$	Aqueous	HILIC	-1	500	810.1252

Table 4.7: Isotope-labeled standards used for extraction standards and injection standards. The fraction, column, and ion mode (z) used for analysis for each standard are displayed as well as the injection concentration and extracted *m/z*. Extraction standards were spiked during extraction, injection standards were spiked just before analysis by LC-MS.

			HILIC (negative)	HILIC (positive)	RP (aqueous)	RP (organic)
LC-MS		RT range (sec)	30-1494	30-1494	50-600	300-1070
		Resolution	60,000	60,000	120,000	120,000
XCMS parameters		peakwidth (sec)	10-150	10-150	4-30	4-30
		S:N threshold	10	10	10	10
		PPM	5	5	3	3
		mzdiff	0.002	0.002	0.002	0.002
		mzwid	0.004	0.004	0.004	0.004
		bw	10	2	1	1
Results	*T. pseudonana*	PPS	150	502	1568	54053
		initial MFs	1372	3782	2680	11361
		Quality MFs	297	523	348	3128
	N. pelliculosa	PPS	140	810	4563	58748
		initial MFs	629	4574	2345	12452
		Quality MFs	192	1457	389	3936

Table 4.8: Experimental parameters and basic results of untargeted analysis. Retention time (RT) range and resolution for the different chromatography and ion mode configurations. Parameters used for XCMS. Results include the peak picking score (PPS) from IPO, initial number of mass features (MFs), number of MFs after filtering out peaks with CV $> 30\%$ after B-MIS normalization and likely isotopologues as described in text (Quality MFs).

Chapter 5

COMMUNITY METABOLOMICS ACROSS A NATURAL GRADIENT IN THE NORTH PACIFIC OCEAN

5.1 Abstract

In the surface ocean, carbon is fixed by phytoplankton and respired by the entire marine community at an astonishingly high rate. At any point in time, the difference between these two processes yields a carbon pool in surface particles that is a combination of both freshly fixed and partially degraded material. On a molecular level, we have a limited knowledge of the small molecules, or metabolites, within this pool. Specific metabolites have been shown to be responsible for fueling respiration, maintaining organismal interactions, and transferring energy throughout the microbial community. Metabolomics, or the direct observation and quantification of the small molecules that are the result of cellular activity, provides an important lens through which we can begin to assess the standing stocks of small compounds that likely fuel a great deal of heterotrophic activity in the surface ocean. Here I describe community metabolomes of particulate material in the North Pacific Ocean. Using both targeted and untargeted metabolomics, I identify metabolites in the particulate carbon pool and explore their latitudinal distributions. This analysis reveals several compounds that have not been previously recognized as abundant components of the marine organic carbon pool. On the basis of hydrographic and nutrient characteristics, the samples were from three distinct oceanographic regimes along a transect in the North Pacific: the southern subtropical domain, the subtropical frontal zone, and the transition zone. I found that the community metabolome showed nuanced differences between the regimes. The community metabolome in surface waters of the subtropical domain was remarkably consistent even when sampled weeks apart, while the northern regions showed a patichier and less reproducible community

metabolome. Some individual compounds showed distinct patterns between oceanographic regimes, including homarine, an abundant molecule that can contribute up to 0.8% of the total particulate carbon pool in marine surface waters. Glutamic acid and glutamine showed opposite patterns in the oceanographic regimes, suggesting differences in community-level nitrogen assimilation in these different regimes. Overall, this study offers a new perspective into particulate carbon composition in oceanographic research, reveals important carbon pools that may fuel the microbial loop, and suggests an altered community-level nitrogen assimilation capacity over the North Pacific transition zone.

5.2 Introduction

An astounding 100 million tons of carbon are estimated to be fixed by phytoplankton on a daily basis [186], roughly equivalent to half the total biomass of humans on earth [187]. Some of this carbon is respired by phytoplankton themselves, and about half of it is respired by marine bacteria and the associated heterotrophic community on a daily basis [1]. The delicate balance between carbon fixation and respiration in the surface ocean is responsible for fueling marine food webs and sequestering carbon into the deep ocean through the biological pump.

The resulting organic carbon in the surface ocean is a chemically diverse mixture, from small biomolecules like simple sugars to large polymers like structural proteins. At the molecular level, surprisingly little is known about the composition of organic matter in seawater. Previous work has shown that different chemical pools of carbon are known to be more resistant to degradation than others, where freshly fixed biomolecules like amino acids are thought to be some of the most bioavailable for degradation [188]. Others have demonstrated that small molecules, or metabolites, can be carbon, nutrient, or energy sources for heterotrophs [47, 137], maintain phytoplankton-bacterial interactions [90, 47], or be important micronutrients for members of the microbial community [130, 10]. Observing metabolites offers a unique perspective into the cycling of marine organic carbon and can be harnessed to answer long standing questions regarding the chemical makeup of the carbon pool that is cycled quickly among microbes in the surface ocean.

Recently, the observable pools of marine organic carbon has extended to these small biomolecules through analytical advancements in liquid chromatography-mass spectrometry (LC-MS) and *metabolomics* in general. Our lab has made advances in our ability to detect and measure relative concentrations of metabolites in the challenging marine environment [8]. We measure both a targeted suite of metabolites (about 200 analytes) that includes compounds like amino acids, osmolytes, and primary metabolites. Beyond the targeted metabolites, we also obtain quantitative and qualitative data through an untargeted approach, observing a huge array of metabolites that ionize well on our analytical platform. Through both approaches, we use statistical analyses to explore relationships between compounds and between metabolites and environmental parameters. Overall, this provides an important link between taxonomy and phenotype.

Metabolomics has been used to investigate physiological changes in marine organisms under laboratory conditions [189, 137, 193, 149], though it has not been widely applied to whole communities. One study investigated how metabolite pools change between long term observational sites in the English Channel [194] using an untargeted approach, and a few studies have explored relationships between targeted metabolites in sinking particles and across oceanographic regimes in the western Atlantic [195]. This work revealed compounds that have not been previously recognized in marine particulate matter and reported distributions of these compounds over depth and geographical space. In this chapter, I explore how surface metabolite pools change across a major oceanographic gradient in temperature, salinity, nutrients, and community composition: the *transition zone* of the North Pacific Ocean.

This *transition zone* extends from Japan to North America and arises from large-scale ocean circulation. Instead of equal mixing, however, the northern and southern edges of the transition zone present more like fronts with rapid changes in thermohaline structure and biological species composition [196, 197]. These areas have long been of interest to commercial fisheries as important fishing grounds [198, 199].

Nutrients change dramatically across these oceanographic regimes, where nitrogen and

phosphorus limit primary production in the south subtropical waters and iron is limiting in the northern subarctic gyres [198]. The production of cobalamin (vitamin B_{12}) changes, where in the subtropical south *Prochlorococcus* produce a closely related, but less bioavailable form of the micronutrient (pseudocobalamin), and in the north the pool is mostly true cobalamin [130]. The different nutrient availability coincides with a gradient in primary producers across the transition zone, where in the south the cyanobacteria *Prochlorococcus* dominate, while *Synechococcus* and picoeukaryotes dominate to the north [200].

5.2.1 Goals of this study

The goal of this study is to investigate the ramifications of these different oceanographic regimes on the metabolite pools in surface organic particulate matter. I explore major pools of metabolites in each regime, examine if these pools of metabolites change within the transition zone, and discuss the consequences of particulate matter composition on the microbial community. I also analyze which metabolites show the greatest differences between oceanographic regimes and discuss which metabolic processes may be affecting or may be affected by the observed differences. Finally, I offer a new set of observations of organic matter that yield a wealth of hypotheses to fuel future research.

5.3 Methods

5.3.1 Sample Collection

Samples were collected for metabolomics of particulate material at locations shown in Figure 5.1 while aboard the Ka'imikai-O-Kanoloa research vessel from April 20 to May 2, 2016. At each sampling location, either duplicate or triplicate filters were collected from 15 m water depth as previously described for environmental metabolomics [8] using either niskins or the underway intake. Table 5.1 has complete descriptions of the samples collected for metabolomics. In short, samples (10–15 L each) were collected into polycarbonate carboys, filtered onto a 147 mm 0.2 μm PTFE filters using a peristaltic pump. Filters were flash

frozen in liquid N_2 and stored at -80°C until extraction.

5.3.2 Metabolite extraction

Metabolites were extracted as previously described [8]. Briefly, filters were bead beaten in 1:1:2 methanol:water:dichloromethane and separated into two fractions: a polar aqueous extract (from the methanol and water fraction), and less polar organic extract (from the dichloromethane fraction). I used the same internal standard suite as in Boysen *et. al.* (2018) to train normalization and monitor instrument stability.

5.3.3 Liquid chromatography

Samples were run on two chromatographic separations using the same solvents, columns, and gradients as previously reported [8]. In brief, the organic fractions of the samples were run in reversed phase (RP), while the aqueous fractions were run on both RP and using hydrophilic interaction chromatography (HILIC). This sample set was particularly prone to obscuring variability, so I diluted the aqueous samples 1 part sample, 2 parts water and the organic fraction of the samples with 1 part sample, 3 parts injection solvent (1:1 water:acetonitrile). I injected 2 μL for HILIC analysis, 5 μL for RP analysis on the aqueous fraction, 5 μL for RP on organic fraction for targeted analysis, and 10 μL for RP on organic fraction for untargeted analysis, based on a dilution series monitoring the internal standard response.

5.3.4 Mass spectrometry

Each LC-sample configuration (RP-organic, RP-aqueous, HILIC-aqueous) was run on two mass spectrometers. I used a waters Xevo triple quadrupole (TQS) run in selected reaction monitoring, SRM, for targeted analysis. For the untargeted analysis, I used a Thermo Q-Exactive (QE) in full scan for quantitative data, or data dependent acquisition (DDA), for fragmentation. The mass spectrometers were run as in Boysen *et. al.* (2018). For untargeted analysis, pooled samples were run in DDA mode for MS^2 fragmentation as described in Heal

et. al. (in prep).

5.3.5 Data processing

Targeted data

As in Heal *et. al.* (in prep), I subjected the targeted data to an in-house quality control, and normalized via best-matched internal standard (B-MIS) normalization [8]. As in Boysen *et. al.* (2018), I used a 40% improvement to the relative standard deviation (RSD) of each mass feature in a pooled sample as a criteria to apply normalization, and I did not normalize any mass features that had a RSD of $< 10\%$ in the raw pooled area. I then normalized for different amount of volume filtered to get a peak area per volume filtered (WV normalized). Finally, I normalized to the concentration of particulate carbon in each sample (PC). For all statistical analyses (see sections below), I only included compounds that were detected in at least half of the samples.

Untargeted data

Untargeted metabolomics data were converted with MSConvert [167] and processed through XMCS [168, 169, 170] as described in Heal *et. al.* (in prep), using the same parameters for XCMS and methodological blank filtering as in that work. I removed peaks where the average peak area across samples was less than 1000, based on visual examination. I normalized for obscuring variation using B-MIS normalization [8], using the same criteria as in the targeted analyses. Like in Heal *et. al.* (in prep), I removed peaks that did not demonstrate acceptable replicability in the pooled samples (coefficient of variance (CV) $> 30\%$); I also removed peaks that showed greater average variability between biological replicates than over the whole sample set as in previous work [137]. I performed three types of subsequent normalization, to WV, PC, and to the MS total usable signal (MSTUS, the sum of all quality peaks after isotope deconvolution [201], see below). For subsequent analyses, the normalized adjusted peak areas of the filtered (quality) mass features were used.

MS^2 data processing. For quality mass features, I searched for corresponding MS^2 scans in the data dependent acquisition (DDA) files and applied a filter to remove low abundance fragments in the exact manner as reported in Heal *et. al.* (in prep)

Isotope and adduct deconvolution. In untargeted metabolomics, multiple mass features can correspond to one metabolite due to natural abundance isotopes, adducts, or multiply charged ions. Like in Heal *et. al.* (in prep), to avoid putting extra statistical weight on these isotopes and adducts, I identified mass features that were likely ^{13}C, ^{15}N, or ^{34}S isotopologues of other mass features. I extended this search to include adducts of Na^+, NH_4^+, K^+, (for positive ionization) and Cl^- (for negative ionization), as well as for doubly charged ions of mass features whose M+H ion was present. I performed these searches within each three second (for RP) or six second (for HILIC) corrected retention time window, and discarded these mass features from downstream statistical analyses. These mass features are reported in Dataset S5.1.

Manual peak quality verification. Despite best efforts, the peak picking algorithms often result in many false positives or poorly integrated peaks [202, 203]. Therefore, I visually inspected the largest 200 peaks in each fraction and discarded features that did not show quality peak shapes or integrations from downstream analyses. I also visually inspected mass features that resulted in significant p values during the ANOVA and t-tests between oceanographic regimes, again discarding any poor quality peaks (see *Statistics* section). Any compounds that are explicitly discussed in the text or reported in tables have been manually verified as well-integrated peaks.

Mass feature identification. As in Heal *et. al.* (in prep), I used the ranking system outlined in Sumner *et. al.* (2007), to attempt to identify the quality mass features present in my sample sets in an automated fashion. In short, I removed any mass features that were possible contaminants [172], searched an internal database of compounds with known exact exact m/z and retention time on the LC-MS configurations, searched against the publicly available LC-MS MS^2 spectral databases [173, 174, 175, 176, 177], and searched for possible matches to compounds in the KEGG database [204, 205] (based only on m/z).

Manual integrations. XCMS peak picking algorithms assume a normal gaussian shape for peaks [168, 169, 170], which often results in poor integrations for compounds that do not achieve this shape during chromatographic separation; these peaks are then removed during the CV filter or manual peak quality verification. Therefore, for a few compounds that I know are present at high abundances in marine particles and did not show quality integrations [195, 47], I manually extracted ion chromatograms from the QE associated with those compounds using Skyline [206]. These compounds underwent the same normalization as the other untargeted data and were added to the dataset after confirming they were not detected during the automated peak picking. Compounds that were manually extracted from the untargeted data include dimethylsulfoniopropionate (DMSP) and 2,3-dihydroxypropane1-sulfonate (DHSP).

Absolute concentrations. For compounds for which our lab has own standards, I calculated absolute concentrations. To do this, I used the following formula.

$$Concentration = \frac{Area}{RF} \times \frac{Vol_{\text{reconst}}}{Vol_{\text{filtered}}} \times \frac{1}{(RF_{\text{ratio}})}$$

Where RF is the response factor ($\frac{Area}{concentration}$) of each compound at known concentration in water. Standards are run before and after each run on each instrument, therefore, an RF for each compound is obtained within each batch. Vol_{reconst} is the volume that the samples were reconstituted into (in this case, 380 μL); Vol_{filtered} is the volume filtered in the field (reported in Table 5.1, generally between 10 and 15 L); RF_{ratio} is the $\frac{RF_{\text{matrix}}}{RF_{\text{water}}}$ these compounds in a matrix of marine particulates (as reported in Boysen *et. al.* (2018)). A few compounds do not have reported RF_{ratio}, if these compounds were detected on HILIC, which experiences less ion suppression [8], I assumed $RF_{\text{ratio}} = 1$. For compounds without known RF_{ratio} detected on RP, I do not attempt to quantify as the RF_{ratio} on RP is highly variable [8]. Errors associated with RF and ion suppression were propagated.

5.3.6 Statistics

Multivariate statistics

For all multivariate statistics, data were standardized to z-scores. All multivariate statistics were performed in R using the vegan package (V2.4-2).

Oceanographic provinces. To define which oceanographic provinces the samples belonged to, I used the hydrographic and nutrient data (salinity, temperature, NO_3^-, PO_4^{2-}, and SiO_4^{2-}) at each of the sampling locations and performed a average-linkage hierarchical clustering analysis based on a Euclidean distance matrix. I tested the significance of resulting clusters using a multiscale bootstrap resampling with 500 bootstraps.

Metabolomics. I used two multivariate analyses to investigate changes in the community metabolome across the sample set. First I used a non-metric dimensional scaling [179] based on a euclidean distance. I assessed dimensionality of the NMDS by examining a scree plot and calculated the probability with a monte carlo permutation. I then paired this NMDS with an analysis of similarities (ANOSIM, [207]) to determine whether the oceanographic provinces could be parsed out in the metabolite data, using 999 permutations. I also performed ANOSIM analysis on other variables that the samples covered including time of collection, northbound or southbound collection, day of extraction, and method of collection (underway intake or niskin).

Univariate statistics

I performed an analysis of variance (ANOVA) test on each compound or mass feature to test for differences between oceanographic provinces. I used an unpaired t-test to test between each of the oceanographic provinces (subtropical domain, subtropical frontal zone, and transition zone). I corrected p values for false discovery rate [181]; only the corrected values are reported throughout the text.

5.3.7 Supporting biological, physical, and chemical measurements

In addition to the metabolomic data presented in this chapter I used data generously collected, processed, and provided by other groups. Hydrographic data (temperature, salinity) were collected via an underway CTD and flow cytometry data were collected via SeaFlow [55]. These data were processed in the Armbrust lab by François Ribalet at the University of Washington. Inorganic nutrients (NO_3^-, PO_4^{2-}, and SiO_4^{2-}) were collected and processed by the Karl lab at the University of Hawai'i Māona. Particulate carbon, nitrogen, and phosphorus were collected and processed by the White lab at Oregon State University. Satellite-derived sea surface temperature and chlorophyll were obtained from MODIS-Aqua Level-2 Ocean Color data, accessed January 30, 2018.

5.4 Results

5.4.1 Defining oceanographic provinces

Hydrographic data split our samples into three major provinces using a average-linkage hierarchical clustering based on a Euclidean distance matrix derived from temperature, salinity, and inorganic nutrients (Figure 5.2). These zones correspond well with previously documented provinces in the area (see [197, 208]). Therefore, for the remainder of this chapter, and in accordance with previous literature [197], I will use the terminology subtropical domain (SD), subtropical frontal zone (SFZ), and transition zone (TZ) to describe the three zones, from south to north (Figures 5.2 and 5.3).

As previously observed [200, 196, 130], these zones show differences in the primary producers present. The SD is dominated by *Prochlorococcus*, while the SFZ and TZ regions have significant *Synechococcus* and picoeukaryotes, respectively (Figure 5.3). Inorganic nutrients are also significantly different between zones, where SD have very low macronutrients, which increase northbound, with a substantial increase between the SFZ and TZ. The physical, chemical, and biological attributes in each zone are summarized in Table 5.2.

5.4.2 Targeted metabolites

Of roughly 200 compounds monitored, 79 compounds were detected in the targeted analysis, which are reported as relative abundances between samples in Dataset S5.2 and summarized by station or zone in Figure 5.4. Among the detected compounds are amino acids, compatible solutes, and primary metabolites. My data generally agree with the few previous publications regarding significant components of community metabolomes of marine particulates [195]. A few compounds that the method usually targets were consistently overloaded in the targeted approach (glycine betaine (GBT), homarine, and dimethylsulfoniopropionate (DMSP)), so quantitative comparisons cannot be made between samples using the targeted data, though it can be done with the untargeted data (see *Untargeted metabolomics* section below).

Multivariate analyses

I used an NMDS analysis to visualize metabolic differences between the zones (Figure 5.5). In multivariate space, the relative metabolite concentrations of the targeted metabolites that were detected in every sample (63 metabolites) were able to discern between zones when normalized to either WV or PC (ANOSIMS $p < 0.001$ for both normalizations), though there was substantial overlap between the northern two regimes. Pairwise ANOSIMS tests between zones showed that this method was able to discern samples between each zone with both normalization methods (SD vs SFZ, SD vs TZ, and SFZ vs TZ, Table 5.3).

Additional ANOSIM analyses revealed that other variables covered by the sample set also attributed to variability of the community metabolome in multivariate space (Table 5.4). Specifically, whether samples were collected via underway intake or niskin may some effect though it does not hold between normalization techniques ($p = 0.004$ for WV normalization, $p = 0.2$ for PC normalization). Similar results were found for whether samples were collected during the day or night ($p = 0.043$ for WV normalization, $p = 0.088$ for PC normalization). Whether the samples were collected on the northbound or southbound portion of the transect significantly attributed to their multivariate location in both normalization techniques ($p <$

0.001 for WV normalization, $p = 0.017$ for PC normalization).

Univariate analyses

Of the 79 metabolites detected in the samples using the targeted analysis, 35 showed significant differences in a 3-way ANOVA between the oceanographic provinces when using WV normalized data (Figure 5.4). Generally speaking, many compounds were more abundant in the SFZ and TZ than in the SD (Figure 5.4 and 5.6), though much of this signal may be due to differences in biomass between the two regimes as the pattern rarely holds when normalized to PC (Figure 5.7). Many more compounds were significantly different between SD and the more northern zones than between the northern zones (SFZ and TZ), using either WV or PC normalization (Figures 5.6 and 5.7, Table 5.6). This corroborates the NMDS analysis, which shows much more overlap between SFZ and TZ than SD and the other two zones. When normalized to WV, compounds with significant differences between oceanographic regimes are depicted with an asterisk in Figure 5.4; a selection of these are more closely inspected in the *Discussion* section.

Quantification of targeted metabolite pools

Through the targeted analysis, I was able to absolutely quantify 58 compounds. These compounds were quantified from the TQS data, with the exception of GBT and homarine which were quantified from the QE data. The quantified values for the 60 compounds are listed in Table 5.5. Overall, I was able to identify 0.55–1.3% of the particulate carbon pool (Figure 5.8) as belonging to of these 60 metabolites.

5.4.3 Untargeted metabolomics

An untargeted metabolomics approach offers a more holistic approach to investigating the small molecules in particulate organic matter [209]. Overall, I detected 16,822 quality mass features in the untargeted metabolomics analysis of the aqueous extractable metabolite pool.

This does not include the 5,804 mass features that were identified as adducts or isotopes of other mass features (reported in Dataset S5.1). Using automated approaches, I was able to putatively identify (confidence level 1 or 2) 72 mass features, and offer possible matches (confidence level 3) for 578 mass features. The remaining mass features did not match to any compounds in the publically available MS^2 databases or in the KEGG pathways under the search parameters. Full results are provided in Dataset S5.3.

Identification of largest peaks

One of my main motivations in this chapter is to identify compounds that are present in high concentrations in particulate organic matter but underappreciated in microbial ecology. Of the largest 100 peaks in the aqueous fraction of the sample set, I was able to identify approximately 37% using combined automated and manual techniques (Table 5.7). In all samples, the largest peak detected was GBT. GBT was followed by DMSP, homarine, and trimethylammonium propionate (TMAP) as the largest four peaks. Together, these mass features comprise nearly half of the MS total usable signal of the aqueous extractable metabolites pool (Figure 5.9).

Multivariate analyses

As for the targeted analysis, I performed a NMDS analysis on the largest 100 mass features in the aqueous fraction detected in the untargeted approach (Figure 5.10). With this less-biased approach, I obtained similar results using a pairwise ANOSIMS tests between zones — all zones were significantly different in multivariate space regardless of normalization technique ($p < 0.05$, Table 5.3).

Supporting the targeted analysis, the untargeted data analysis revealed that whether the samples were collected on the northbound or southbound portion of the transect significantly attributed to their multivariate location in both normalization techniques ($p = 0.02$ for WV normalization, $p < 0.001$ for PC normalization, and $p = 0.005$ for MSTUS normalization). Other variables tested (whether samples were collected via underway intake

or niskin, whether samples were collected during the day or night, or which day the sample was extracted on) did not have a consistent multivariate signal, which also corroborates the targeted analysis (Table 5.4).

Univariate analyses

I performed *t*-tests to investigate differences between provinces on each individual mass feature. Of over 16,000 mass features detected in the aqueous fractions of the samples using the targeted analysis, 7.5% showed significant differences in a 3-way ANOVA between the oceanographic provinces when using WV normalized data. Similar to the targeted analysis, more compounds were significantly different between SD and the northern zones than between the northern zones (Table 5.6, Figures 5.11, 5.12), using either WV, PC, or MSTUS normalization. Tables 5.8, 5.9, and 5.10 report mass features that were significantly different between oceanographic regimes in either PC or WV normalized data from the aqueous extract. A selection of these mass features are more closely inspected in the *Discussion* section.

5.5 Discussion

5.5.1 Metabolite profile differences between provinces

Regardless of normalization technique or statistical perspective (univariate or multivariate), I found broad patterns in metabolite profiles between the three oceanographic regimes of interest. The NMDS analyses revealed that the SD samples were more similar to each other (less in-group spread) than the other two stations (Figures 5.5 and 5.10), showing that the metabolite pools in the SD were remarkably reproducible, even when collected weeks apart. Station-to-station differences were high in the SFZ and TZ in the largest detectable peaks (Figure 5.9) and in multivariate space (Figure 5.5, 5.10). The results suggest a biologically patchier environment in the northern portion of the sampling area, with heterogeneity in community composition, community activity, or a combination of both. This

finding is supported by less replicable measurements within the regeime on other parameters like particulate carbon, net community production, and cell density seen in previous studies [200].

The metabolite profiles of the SFZ and TZ showed significant overlap in multivariate space (Figures 5.5 and 5.10), with few mass features showing a distinct differences in abundance between the regimes in a univariate comparison (Figures 5.6, 5.7, 5.11, and 5.12). I found this similarity (and the overall similarity of the community metabolomes across the samples) surprising as the nutrient availability between the two areas is different (Table 5.2), and the community composition is distinct between the two regimes (Figure 5.3, Lim *et. al.* (in prep)). Previous work of pure cultures of marine organisms has shown that metabolite pools change under different nutrient regimes [189, 137, 149, 151], and recent work from our lab demonstrates that closely related organisms like two centric diatoms diatoms can produce different metabolite pools (unpublished data). The community metabolome data suggest that functional diversity is more consistent between the two northern sampling regimes compared to the large differences in phylogenetic diversity [208]. It is important to note, however, the analysis only yields a snapshot in time — it is possible that the turnover rates of different metabolites are different between regimes but the resulting standing stock is similar. Nevertheless, these are some of the first data to explore metabolite pools over the transition zone in the North Pacific, and the findings suggest that over this oceanographic transect, the metabolite pools in the surface ocean are remarkably consistent.

5.5.2 Largest peaks

One of my main research objectives was to identify compounds present in high concentrations in particulate organic matter. Figure 5.9 shows the relative contribution of the largest 20 peaks in the untargeted analysis from the aqueous fraction and Table 5.7 reports the top 100 peaks with the best identification. The size of the peak in a sample is related to both the concentration of the compound and the relative ease in which the compound ionizes. Therefore, we should interpret these "large" peaks as compounds that are both abundant

and efficient ionizers. There are likely more compounds that are in high abundance that do not ionize as easily and therefore do not present as large peaks. I discuss some of these exceptions in the following paragraphs. Nevertheless, the compounds presented in both Figure 5.9 and Table 5.7 are abundant in marine particulate matter.

Many of the largest peaks in the untargeted analysis are zwitterionic compounds known to act as compatible solutes in microorganisms (i.e. GBT, DMSP, homarine). Some have been recognized as contributing to the metabolite pools in marine particulate matter previously [195, 47, 210], but to my knowledge, many have not been previously measured in marine particles. For instance, homarine is recognized as a compatible solute in marine invertebrates [211, 212], and has been observed more recently in microalgae [189, 213], but has never been reported in the open ocean. Other probable compatible solutes like proline betaine, trigonelline, and dimethylsulfonioacetate (DMS-Ac) have received less attention in the marine environment but were abundant in the samples. Although most of the largest peaks did not show a univariate difference between oceanographic regimes (Figures 5.11, 5.12), nuanced differences of metabolite pools in multivariate space were apparent in the NMDS analysis of the largest 100 peaks (Figure 5.10, Table 5.3).

The quantitative analysis yielded absolute concentrations on 58 compounds that were detected in all samples. Though I was able to putatively identify more compounds with m/z and MS^2 matches, without authentic standards I cannot quantify these compounds. Homarine and GBT were the most abundant of these quantifiable compounds (Table 5.5), corroborating the untargeted analysis which showed these as some of the largest peaks (Figure 5.9). The targeted analysis also showed that the simple sugar sucrose and the nucleic acid base guanine were abundant, with concentrations ranging from 10 to 800 pM, despite their much smaller peak areas compared to the zwitterions GBT or trigonelline (Table 5.5, Dataset S5.2), due to the fact that neither sucrose nor guanine ionize as efficiently. Other compounds were also found to have particulate abundances in the tens of pM. These cover many of the quantifiable compounds in Figure 5.9, but also include glutathione disulfide, guanosine, citrulline, and homoserine that were not among the 100 largest peaks detected in

the untargeted analysis (Table 5.5, Table 5.7). These are the first measurements of most of these compounds in the Pacific Ocean.

Remaining unknown compounds present in high abundance

Many of the mass features with the largest peak areas in the samples are without identifications(Table 5.7, Figure 5.9). For example, the mass feature "I155.0815R8.72" has a *m/z* of 155.08146 and major peak in its MS^2 fragmentation spectra of *m/z* 109.07644 (Figure 5.9, Table 5.7). In addition to being a large peak, its distribution is oceanographically interesting; it is more abundant in the TZ than in the SD (Table 5.9). At the time of writing this work, there are no hits in the METLIN or MassBank databases that obviously match to a compound with that parent mass that also has a fragment of *m/z* 109.07644. There are several matches to the parent mass in both databases, and the isotope distribution of this peak paired with the low mass error hint that the empirical formula for this mass feature could be $C_7H_{10}N_2O_2$. However, there are several different compounds with that same parent mass nor can I rule out that it is a water or ammonium adduct. Therefore, without more diagnostic information (a matched spectrum or retention time), I can only provide a possible molecular formula. Without an identification, I cannot offer much in the way of ecological or biochemical interpretation. Nevertheless, I hope it encourages the field of metabolomics to continue their efforts on compound identification.

5.5.3 Glycine betaine and related compounds

In the untargeted analysis, the largest peak detected in all of the samples was identified as glycine betaine (GBT, Figure 5.13). GBT has been recognized as the dominant compatible solute in many organisms, from bacteria to higher plants [216, 217]. Particulate GBT is present at 0.7–3 nM in the samples, representing 0.1–0.4% of the PC pool (Table 5.5, Figures 5.8, 5.14, and 5.15). GBT has been previously quantified in marine particulates in coastal N-rich waters at concentrations from $<$ 2nM to 15 nM [210], as well as in the open ocean of the Atlantic at lower concentrations than these findings (3.8–580 pM, [195]). GBT has a

relatively low C:N (5:1), which is lower than classic Redfieldian particulate matter (6.6:1), and GBT represents 0.1–0.5% of the PN pool (Table 5.5). GBT production has been demonstrated in many diverse organisms, but the accumulation to high intracellular concentrations is not universal among marine taxa. For instance, recent work showed that not all diatoms produce GBT [189] and many cyanobacteria do not appear to have the genetic capacity for GBT synthesis [218]. In a recent survey of over 20 marine photoautotrophs, our lab found that GBT is not detectable in several cultures of *Prochlorococcus* and some species of marine diatoms (unpublished data).

Three compounds closely related to GBT were also detected in the sample sets: dimethylglycine (DMG), choline, and trimethylammonium propionate (TMPA, putatively identified as shown in Figure 5.16, structures shown in Figure 5.13). Choline is a precursor to GBT, while DMG is produced from GBT (Figure 5.13). Virtually nothing is known about the biochemistry of TMAP; its structural similarity to GBT suggests it has as similar function in cells (Figure 5.13), and it has been detected in some marine microalgae before [213]. Beyond this, no work has investigated the role of TMAP in microbes, natural environments, or cellular biochemistry.

GBT is more abundant in the SD than in the TZ, while choline, DMG, and TMPA do not show any latitudinal differences in WV normalized abundances (Figure 5.14). None of these compounds have significantly different contribution to the particulate carbon pools between oceanographic zones (Figure 5.15). Despite their close relations (either biochemically or structurally), these compounds do not share patterns latitudinally. This suggests that the sources and sinks for these compounds that may not be consistent in the different oceanographic regimes. The enzymology of GBT production and its conversion to DMG is relatively well characterized — in future work, I can harness transcriptomic data to investigate differences in these chemical conversions in the different regimes in future work.

I found the abundance and prevalence of GBT across the sampling scheme quite surprising. Most studies of GBT suggest that it accumulates in cells as it is used as a compatible solute. Many organisms are able to sustain a pool of different types of compatible solutes,

and the diversity of compatible solutes used by organisms is an outstanding area of research in evolutionary and molecular biology. To us, it seems counterintuitive that GBT, a nitrogen-rich compatible solute, would be a widely used in the N-limited SD, especially since there are abundant marine organisms that are able to live without it [218]. Furthermore, recent work has shown that in some organisms, GBT production depends on availability of cobalamin [189], which is low in the oligotrophic surface ocean (often < 0.5 pM, unpublished data). The prevalence of GBT in the N-limited SD suggests that GBT may have secondary functions beyond osmoprotection that has placed pressure on the maintenance of GBT accumulation in marine microbes. I offer two hypothetical roles that GBT could fulfill that may explain its high intracellular concentrations observed in this study and in others [210, 195].

One possible function of GBT is as an alternative methyl donor in cells when methionine stores are low. Some organisms are able to regenerate methionine from homocysteine by demethylating GBT via betaine-homocysteine methyltransferase (BHMT, Figure 5.13, [219]). Methionine scarcity can occur in cobalamin auxotrophs when cobalamin supply is low because cobalamin is a cofactor in their methionine synthase (MetH) [189, 134]. By this logic, if an organism with BHMT maintains a cellular pool of GBT, the cell may be able to replenish methionine stores without the aid of the cofactor cobalamin; this has been shown in higher organisms [220]. BHMT has primarily been investigated in mammals, though the gene is not exclusive to eukaryotes [220, 221]. Further work can be done to test the hypothesis of GBT as an alternate methyl donor in marine microbes by searching genomes and transcriptomes for co-occurance of *bhmt* and *metH* in the same organism. Intriguingly, across dozens of Red Sea metagenomes, bacterial *bhmt* has been shown to correlate with surface particles [222], and the ubiquious and genomically streamlined "*Candidatus* Pelagibacter ubique" has maintained this gene [221]. In fact, *P. ubique* uses BHMT exclusively for methionine generation from homocysteine (unlike most bacteria that use MetH or MetE [223]), indicating that GBT is available to these ubiquitous organisms. The high concentrations of GBT I observed corroborates this work.

Many compatible solutes can stabilize proteins, and this is thought to be a major role

of the compounds in cells (beyond maintaining osmotic pressure in a saline environment) [224]. However, there have been several laboratory experiments that demonstrate that not all compatible solutes perform equally in regards to protein stabilization [225]. GBT, in particular, has been highlighted as an unusually effective protein stabilizer for bacterial and eukaryotic Photosystem II complex [226]. Papageorgiou *et. al.* (1995) suggest this may be because of its small size, zwitterionic state at physiologic pH, and its lack of a lone electron pair that could interact with metal centers of proteins. Compared to alternative compatible solutes with a lower C:N, perhaps GBT is better able to stabilize important proteins like those involved in photosynthesis. If this is the case, its prevalence in the sample set could be explained by the need to prevent protein turnover in the harsh, high light, environment of the surface ocean.

5.5.4 DMSP and DMS-Ac

DMSP was the second largest peak detected in most of the samples in the untargeted analysis (Table 5.7 and Figure 5.9). I detected particulate DMSP at concentrations in the 100s of pM to a few nM across the sample set, which is considerably less than previous particulate DMSP measurements in the open ocean [227]. DMSP is readily volatilized during sample processing and I suspect that the sample preparation used in this study may result in some loss of the compound; therefore I do not consider DMSP to be a quantifiable compound with the methodology presented here. DMSP is sulfur-containing molecule that is used as an osmoprotectant and antioxidant in phytoplankton, where it can accumulate to 100s of mM within the cytosol [228]; its structure is shown in Figure 5.17. Like GBT, DMSP has been shown to be affected by cobalamin availability in cultures cobalamin-dependent phytoplankton [189]. It is also the precursor to the climatically-active gas dimethylsulfide (DMS, Figure 5.17) and thus received much attention in the oceanographic community [152]. In some culture studies, DMSP is shown to replace GBT under nitrogen stress [229, 230]. On a community level, however Keller *et. al.* (2004) did not observe this replacement in coastal waters. Although my DMSP measurements are rightfully suspect, I do not observe

any obvious trends in either DMSP or GBT from the N-limited SD into the TZ.

I putatively identified a compound of similar structure, dimethylsulfionoacetate (DMS-Ac), as a prominent peak in the samples (Figures 5.17, 5.16, and 5.9, Table 5.7). Without a commercially available standard, I was unable to obtain absolute concentrations for the compound. This compound has been observed in pure cultures of eukaryotic phytoplankton [213] and corals in the Heron Island reef flat [231], but it has received little attention in biochemical or ecological studies. If microbes degrade DMS-Ac in the same way as DMSP, DMS-Ac could also be a significant precursor to DMS (Figure 5.17).

5.5.5 Other betaines

Several other betaines were among the largest peaks, including proline betaine, betonicine (putatively identified), and arsenobetaine (putatively identified) (Table 5.7 and Figures 5.9 and 5.18). Both proline betaine and betonicine have been identified as compatible solutes in bacteria and plants, though the production and catabolism of these compounds is not well studied in any organisms [232]. The biosynthesis of proline betaine has been mostly elucidated in higher plants, where it originates from ornithine and proline, though the final methylation step has not yet been characterized in any organism [233, 234]. I found that betonicine was both more abundant and contributed significantly more to the PC pool in the SD than in the TZ; proline betaine followed a similar pattern (Figures 5.19 and 5.20).

Interestingly, betonicine has been shown to be produced by the red alga *Ahnfeltiopsis flabelliformis* and possessed a stimulatory effect on bacterial quorum sensing, though a much higher concentration was needed to obtain a similar result to that from the well studied and structurally unrelated molecule noctanoyl-dl-homoserine lactone [235]. Without a commercially available standard, I am unable to quantify this compound, but if the compound is abundant in marine systems as the analysis suggests, it could prove an important mediator of bacterial quorum sensing in marine systems. Like many of these compounds, very little is known about the biochemistry or enzymology of betonicine — there are no reports of microbes producing betonicine, though macroalage [235] and corals [231] have been shown to

accumulate the compound at levels that would imply its role as a compatible solute. Since the enzymology of betonicine and proline betaine production are poorly understood, exploring organismal contributions to these pools is not possible at the moment using genomic-based tools. However, I can mine existing metabolomics datasets produced in the lab (both from Heal *et. al.* (in prep) and other unpublished work) as well as data contributed to data repositories (i.e. the MetaboLights [236]) to search organisms' metabolomes for evidence of production of these interesting compounds.

I putatively identified an arsenic-based betaine in the samples from the untargeted analysis (Table 5.7, Figure 5.19). Arsenobetaine was first detected in marine animals over thirty years ago [237], and it is the most abundant form of arsenic in marine animal tissues [238]. Scientific debates remain on whether the compound is produced *in situ* in animals, in the algae marine animals consume (and thus bio accumulated), or by the bacterial community that is associated with algal or animal microbiomes. There are multiple proposed biochemical pathways of production of arsenobetaine and little consensus in the field [238, 239]. Bacterial degradation of arsenobetaine has been observed in pure cultures — this work has revealed the biochemical pathway but the enzymes involved are not yet characterized [240]. To my knowledge, arsenobetaine has not been investigated in microalgae or been observed in the open ocean. The structural similarity of arsenobetaine to the ubiquitous GBT (Figures 5.18 and 5.13), begs the question of interchangeability of arsenobetaine with GBT in terms of compatible solute properties under nitrogen limitation. This nascent hypothesis is supported by the opposite distributions that GBT and arsenobetaine have — arsenobetaine is more abundant in the N-limited SD whereas GBT is more abundant in the northern areas (Figures 5.19 and 5.14).

Two aromatic betaine compounds, homarine and trigonelle, were found to be among the largest peaks detected in the samples (Table 5.7 and Figure 5.9). These aromatic compounds are poorer ionizers than GBT, DMSP, or proline betaine, and I was surprised to find that homarine was present at 0.2 to 5 nM in marine particles, occupying up to 0.8% of the total PC pool (Table 5.5, Figures 5.8 and 5.21). Homarine has been detected in marine particles

from surface coastal waters at similar concentrations ($\sim$ 1 nM) [210]. I also detected the isomer of homarine, trigonelline, at lower concentrations (2 to 200 pM, Table 5.5, Figure 5.22); to my knowledge, trigonelline has not been detected in any environmental sample. Though the biosynthetic pathways of these compounds have not been fully characterized, some early work from the 1970's suggests that in animals, homarine is decarboxylated from quinolinic acid, which is produced from tryptophan [241]; the exact enzyme that performs the decarboxylation has not yet been characterized. To my knowledge, no published work has attempted to identify biosynthetic pathway or enzymes involved in homarine or trigonelline biosynthesis in marine microbes, so it is not surprising that these molecules have not been highlighted through genomic or transcriptomic techniques in marine systems.

In the SD, homarine is much less abundant than in the SFZ or TZ ($p < 0.05$, Figures 5.21 and 5.11, and Tables 5.8 and 5.9). Unlike GBT that appears to follow biomass, homarine also contributes less to the PC pool in the TZ and SFZ than in the SD ($p < 0.05$, Figures 5.21 and 5.12, Tables 5.8 and 5.9). Trigonelline was less abundant in the SD than the TZ, but did not contribute to more of the PC pool, showing a distinct pattern from homarine (Figures 5.21 and 5.22). These patterns suggest distinct sources or sinks for these structurally similar compounds.

In the few studies that have investigated homarine in marine particulates or marine microbial isolates, homarine is much less abundant than GBT [210, 229, 230, 213], which does not hold in these samples. Both of these compounds are zwitterionic nitrogenous betaines that likely serve (at least in part) as compatible solutes, though homarine has a considerably higher C:N (7:1) than GBT (5:1). In a survey of over 40 marine organisms, I saw that high homarine accumulation in cells was limited to select eukaryotic algae (unpublished data). I hypothesize that the increased homarine in the SFZ and TZ is a result of the changing phytoplankton community [208] and changing contribution to the total PC pool by eukaryotic phytoplankton (i.e. increasing picoeukaryotes, Figure 5.3). On the other hand, from culture work I know that homarine concentrations within cells can be affected by nutrient availability [189], so the quantitative differences at the community level are likely

due to both a change in the community composition present and the way the community is experiencing environmental stressors. It is also possible that the turnover times of GBT and homarine are very different, if particulate homarine has a longer residence time than particulate GBT (perhaps due to a slower secretion rate), the standing stock could be of similar magnitude with equal sources.

5.5.6 Glutamic acid and glutamine

Among other amino acids, I detected glutamic acid and glutamine at high abundances in all of the samples (glutamic acid 7–50 pM, glutamine 2–8 pM, Table 5.5). Glutamine was more abundant in the SD than in the SFZ or TZ with either normalization techniques (Figures 5.4, 5.6, 5.7, and 5.23). Glutamic acid was less abundant in the SD than in the SFZ or TZ with WV normalization (Figure 5.4, Tables 5.8 and 5.9), but it did not contribute differentially to the PC pool in the different regimes (Tables 5.8 and 5.9). During ammonia assimilation (either directly or after nitrate or nitrite reduction), glutamic acid is produced from alpha ketoglutarate via glutamate synthetase. Afterwards, glutamate can assimilate another molecule of ammonia to produce glutamine via glutamine synthetase. In pure cultures of phytoplankton, the glutamine:glutamic acid ratio has been interpreted as a biomarker for nitrogen stress, where a lower ratio was observed when the cells were stressed for nitrogen [242, 243]. Flynn *et. al.* (1989) attribute this change in glutamine:glutamic acid due to the role these molecules play in inorganic nitrogen assimilation. Those authors proposed a low glutamine:glutamic acid ratio may indicate that intracellular N is insufficient for protein synthesis. However, the same authors note in a mixed community that the glutamine:glutamic acid is a function of the nutrient status of both predator and prey [242].

If I interpret the glutamine:glutamic acid in the environment as a community signal for N-limitation, it would indicate that the community SD is more N-limited than in the SFZ or in the TZ, which is counter-intuitive to the concentrations of nitrate over the transect (Figure 5.25). There is much literature that supports that primary production in the SD is N-limited; but it is not clear if the SFZ or TZ are also N-limited, though preliminary incuba-

tion studies from this same field campaign suggest they are. Regardless, I hypothesize that the changed glutamine:glutamic acid ratio is a reflection of the community nitrogen assimilatory potential, where the southern community is dominated by oligotrophic organisms that have the ability to access and thrive on very low concentrations of ambient nitrogen. In the north, I hypothesize that the community is also nitrogen limited (despite the higher concentrations of NO_3^-), but have the biochemical capacity to assimilate more nitrogen, if it were available. Follow up studies including manipulative incubations could confirm or disprove this hypothesis. Alternatively, the changes in glutamine and glutamic acid concentrations could be a result of community-level processes that these compounds are involved in beyond NH_4^+ assimilation, like building proteins or glutathione synthesis. Alternate N sources (i.e. direct assimilation of organic nitrogen or urea) could also alter the glutamic acid:glutamine in cells, which would suggest that there are changing nitrogen sources (from inorganic to organic) along the transect. Finally, it is well known that different microbes have different free amino acid contents and that even within one type of organisms during different times of their cell cycle the amino acid content can change dramatically [244]. Therefore, free amino acid concentrations may change with community composition or activity (independent of nitrogen assimilation potential).

5.5.7 Metabolites as fuel for the microbial loop

Although these measurements are of the metabolites associated with the particulate pool of organic matter, all the components of this pool contribute to the dissolved organic matter pool after cell lysis or sloppy feeding. Once in the dissolved organic carbon pool, heterotrophic bacteria may be able to use these compounds as their source of carbon, nutrients, and energy [45]. Some of the compounds I have discussed here been implicated as important currency between marine microbes as fuel for the microbial loop by direct measurements and physiological experiments. For instance, DMSP is well studied with regards to microbial respiration and degradation [137]. The sulfonate (R)-2,3-dihydroxypropane-1-sulfonate (DHPS, which I detected at 0.1–6 pM) has been shown to underpin a co-culture of a marine

bacterium and diatom and be present in concentrations in marine waters [47].

Other compounds I detected have been hypothesized as important substrates for bacterial respiration through genome-based techniques. For example, GBT transporters are among the most dominant proteins expressed by *P. ubique* in the Sargasso Sea [245], the Red Sea [222], and throughout the water column in the Atlantic [246], indicating this abundant molecule is a carbon and energy source to the microbial community. The bacterial degradation pathway for proline betaine was recently elucidated in three alphaproteobacteria [247]. In this work, Kumar *et. al.* (2014) searched all sequenced bacterial genomes and found that of those with the genomic capacity for proline betaine degradation, 20% were marine bacteria, including the genomically streamlined open ocean bacterium *Pelagibacter bermudensis*. Both GBT and proline betaine are tractable elements of the microbial loop — through existing techniques, I can mine transcriptomic datasets to understand the sources and sinks of these molecules over biogeographical regimes. By combining these transcriptomic and metabolomics techniques, I can identify key organisms and compounds that govern microbial respiration in the surface ocean.

Many of the molecules detected and discussed in this chapter have no recognized roles in microbial ecology or microbial food webs. As a community, we know virtually nothing about the genetic machinery needed to biosynthesize or catabolize molecules like homarine, trigonelline, TMAP, DMS-Ac, and betonicine. I hope this work encourages a newfound dedication to understanding the biochemical pathways of catabolism for these compounds and exploring how these abundant molecules contribute to fueling the microbial loop in the surface ocean.

5.6 Conclusions

The work in this chapter aimed to present a new set of observations to the marine organic carbon pool. I investigated the metabolite pools of marine particulates over a transect into the transition zone of the North Pacific. This analysis showed that there are significant differences in the metabolite pools present in the different oceanographic regimes sampled

and I explore a few of these differences. Generally, the SD was consistently more stable pool of metabolites than the TZ and SFZ. The community metabolomes in the TZ and SFZ were more similar to each other and showed a patchier distribution. I calculated absolute concentrations on sixty metabolites in marine particulate material, providing some of the first data on many of the compounds in this oceanic basin. I was particularly interested in attempting to uncover molecules that are highly abundant and may be fueling microbial respiration. I offer several zwitterionic betaines as possible underappreciated molecules in microbial food webs based on their high abundances. This work opens doors for future research to understand which compounds mediate carbon and energy cycling in the surface ocean through the microbial loop by directly observing the molecules that accumulate in marine microbes.

5.7 Conclusions

The work in this chapter aimed to present a new set of observations to the marine organic carbon pool. I investigated the metabolite pools of marine particulates over a transect into the transition zone of the North Pacific. This analysis showed that there are significant differences in the metabolite pools present in the different oceanographic regimes sampled and I explore a few of these differences. Generally, the SD was consistently more stable pool of metabolites than the TZ and SFZ. The community metabolomes in the TZ and SFZ were more similar to each other and showed a patchier distribution. I calculated absolute concentrations on sixty metabolites in marine particulate material, providing some of the first data on many of the compounds in this oceanic basin. I was particularly interested in attempting to uncover molecules that are highly abundant and may be fueling microbial respiration. I offer several zwitterionic betaines as possible underappreciated molecules in microbial food webs based on their high abundances. This work opens doors for future research to understand which compounds mediate carbon and energy cycling in the surface ocean through the microbial loop by directly observing the molecules that accumulate in marine microbes.

5.8 Acknowledgements

I would to thank R. Lim, R. Grossman, B. Durham, and the rest of the science and crew on KOK1606 for collecting the samples presented in this chapter. N. Kellog, R. Lionheart, A. Weid, and A. Boysen helped in the lab for sample extraction; L. Truxal Carlson led data acquistion. SeaFlow data were provided by F. Ribalet from the Armburst group, particulate carbon data were generously provided by A. White, and inorganic nutrients were shared from the Karl lab. The entire Ingalls lab provided helpful insight into data analysis and interpretation. This work was supported by grants from the Simons Foundation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I.), NSF OCE-1228770 and OCE-1205232 to A.E.I., NSF GRFP to K.R.H.

5.9 Figures

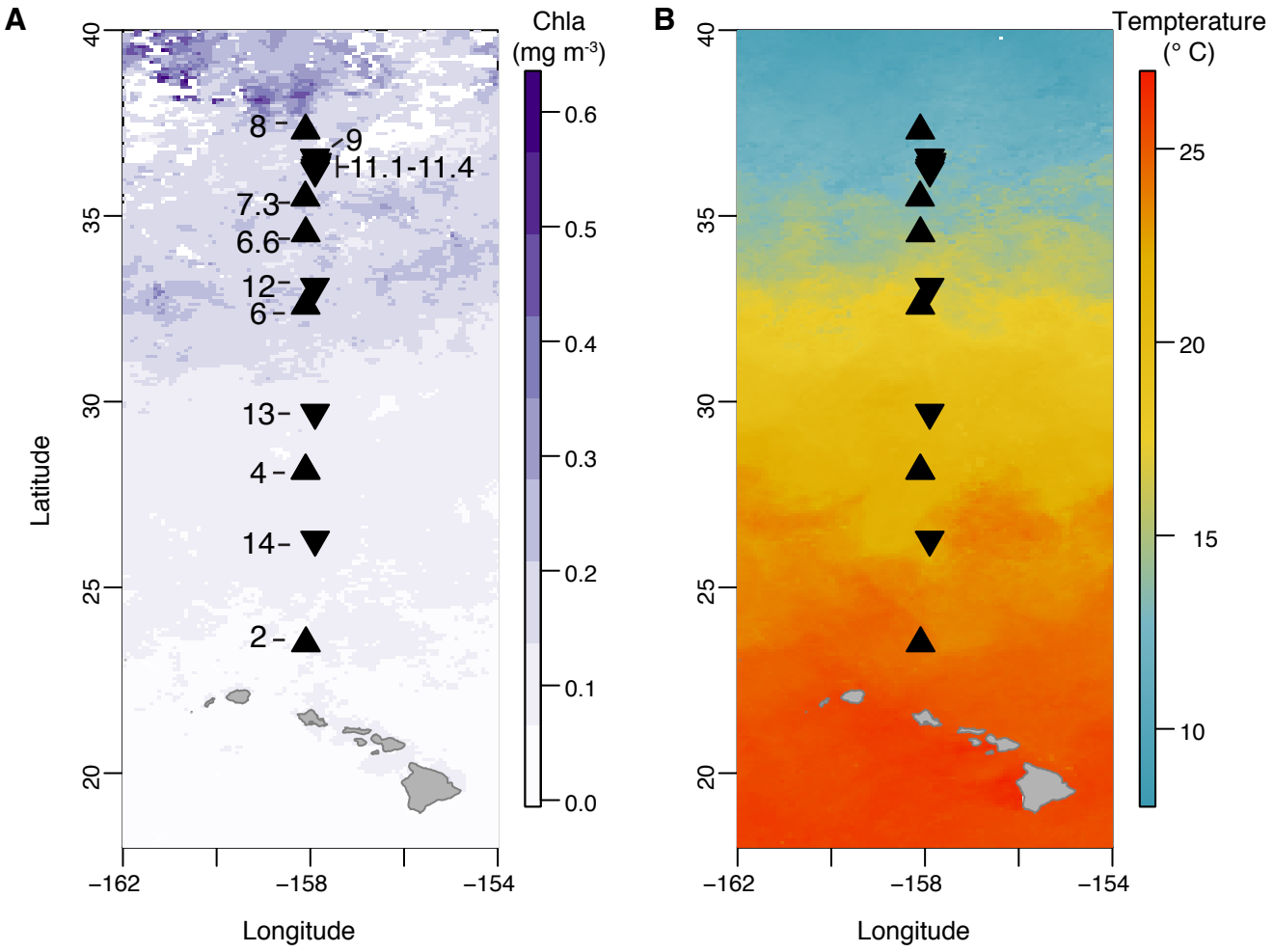

Figure 5.1: Locations of samples collected for metabolomic analysis. Samples were collected northbound (triangles pointing up), and southbound (triangles pointing down), separated for visualization — all samples were collected at approximately -158 W. Colors behind sampling locations correspond to chlorophyll a (A) or sea surface temperature (B) from MODIS satellite 9 km, 8-day composites over the sampling period.

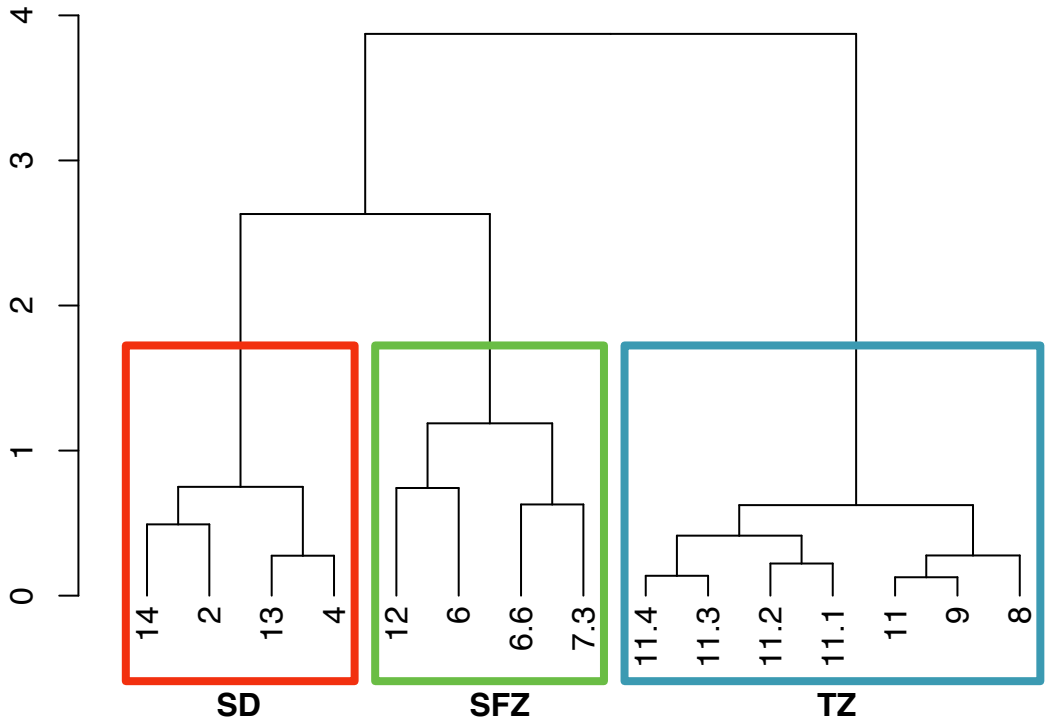

Figure 5.2: Average-linkage dendrogram of the sampling locations based on a Euclidean distance matrix on environmental parameters (salinity, temperature, NO_3^-, PO_4^{2-}, and SiO_4^{2-}). Using a multiscale bootstrap resampling with 500 bootstraps, the three boxed clusters are statistically different from one another ($p < 0.05$). These correspond to three ocean regimes discussed in the text, the subtropical domain (SD, red), subtropical frontal zone (SFZ, green), and transition zone (TZ, blue). Locations of station numbers are shown in Figure 5.1.

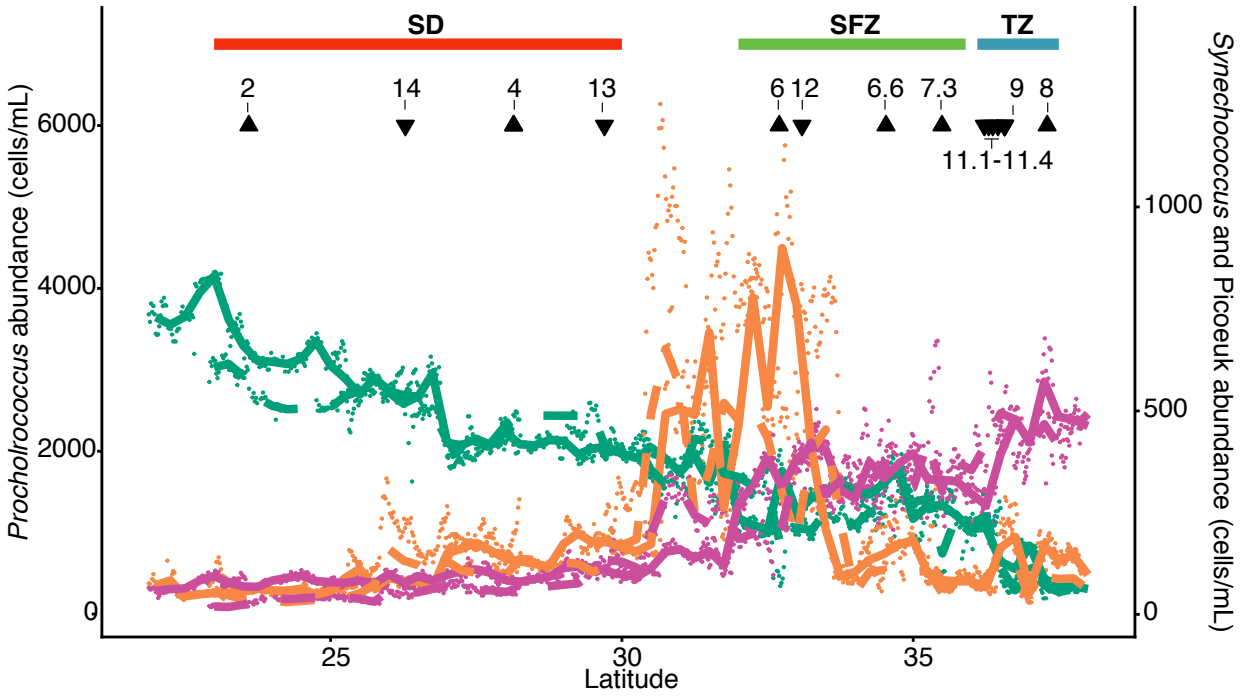

Figure 5.3: Major photoautotrophic abundances over the transect. *Prochlorococcus* (green), *Synechococcus* (orange), and picoeukaryotes (magenta) as monitored by SeaFlow [55]. Dots are 0.1 degree binned averages, thick lines are 1 degree binned averages (solid line is northbound, dashed is southbound). Locations of the samples with station numbers shown with triangles (northbound collected, upward facing triangles; southbound collected, downward facing triangles).

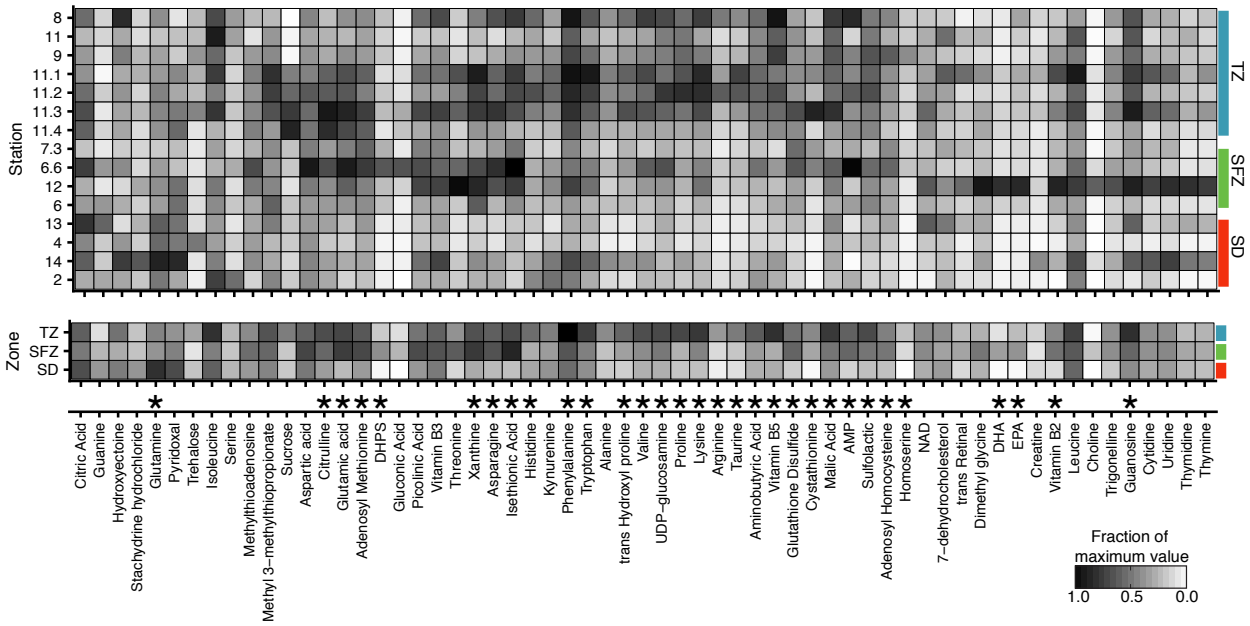

Figure 5.4: Summarized results from the targeted metabolite analysis when normalized to WV. Metabolites that were detected in all samples are shown summarized by Station (top tiles), or Zone (bottom tiles), with means shown. The shade of each tile corresponds to the fraction to the maximum value for that metabolite across all stations. Metabolites that had significantly different abundances between zones (ANOVA test with $p < 0.05$) are noted with an asterisk.

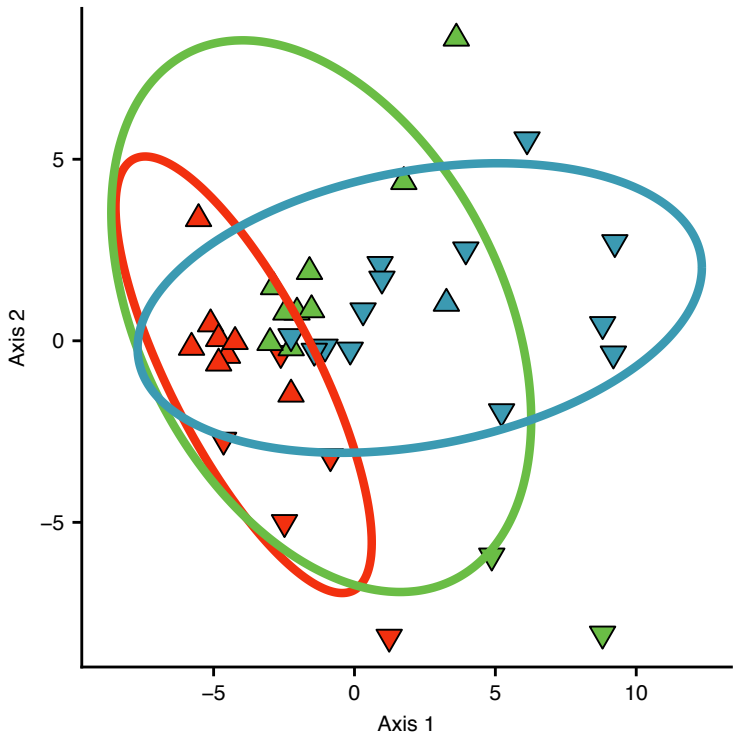

Figure 5.5: Unconstrained ordination NMDS analysis performed on Euclidean distance matrix of targeted analytes. Samples are colored by zone: SD, red; SFZ, green; and TZ, blue. Samples collected on the northbound transect, upward facing triangles; southbound transect, downward facing triangles. Ellipses represent 95% confidence intervals for each group, stress = 0.136, $p < 0.01$.

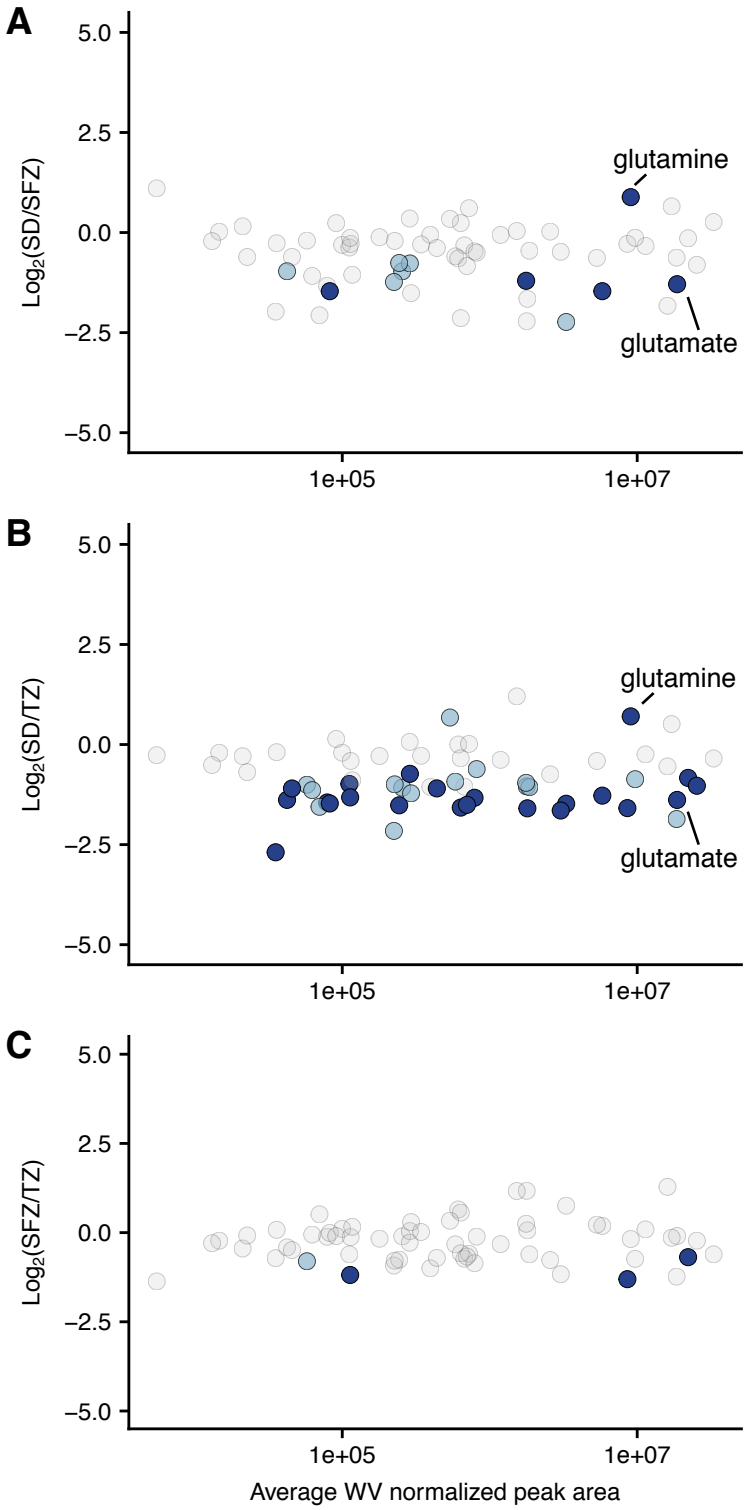

Figure 5.6: Caption next page.

Figure 5.6. *Continued from previous page.* Overall results from the targeted analysis, highlighting mass features that show significant differences between oceanographic regimes. These data have been normalized to WV. Each mass feature is depicted as a dot; blue features are significantly different between oceanographic regimes (by unpaired t-test) mass features with p between 0.01 and 0.05 are shown in light blue, and mass features with $p < 0.01$ are shown in dark blue. All other mass features are show in grey. Y-axis is $\log_2$(fold change) of average peak size between the different oceanographic regimes, either SD and SFZ (A), SD and TZ (B), or SFZ and TZ (C). X-axis is average peak area of mass feature; note that x-axis is log scaled.

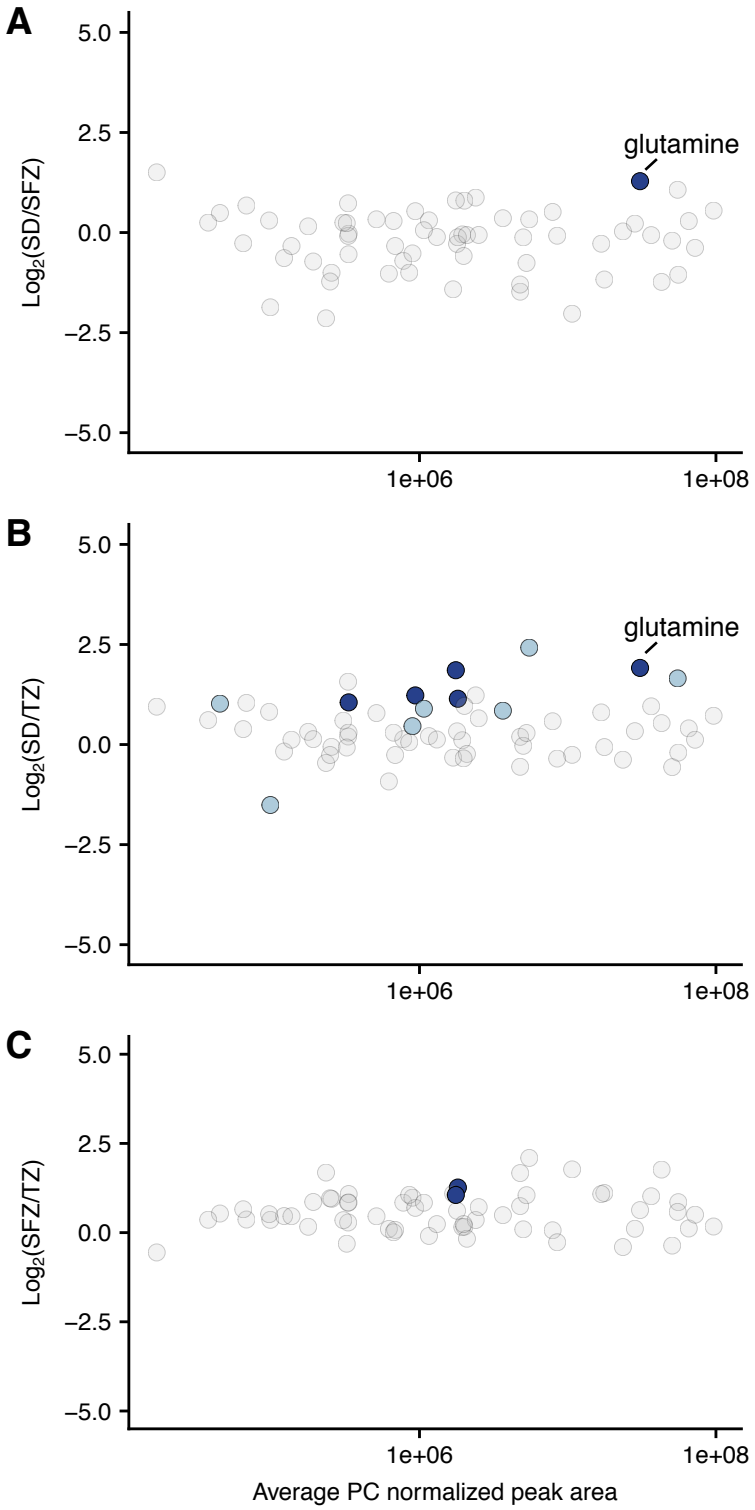

Figure 5.7: Overall results from the targeted analysis, highlighting compounds that show significant differences between oceanographic regimes when normalized to PC. Colors and axes are as in Figure 5.6.

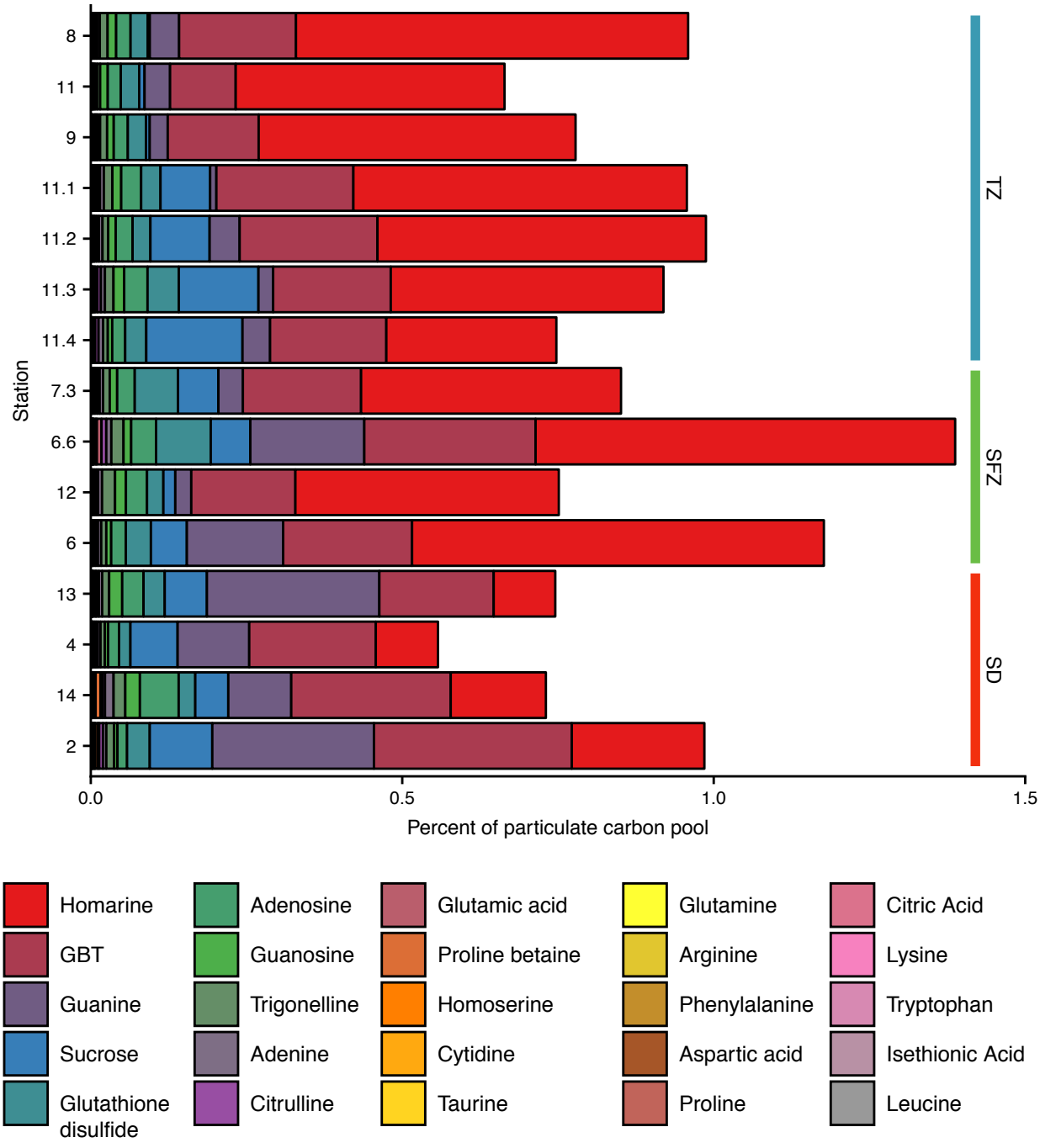

Figure 5.8: Most abundant quantifiable compounds as a percent of the particulate carbon pool at each station. Stations are ordered by latitude (northernmost at top). These data are from the TQS, except homarine and GBT, which were overloaded on the TQS. Displayed is the mean contribution of each compound to the PC pool of each station. These metabolites are all unequivocally identified.

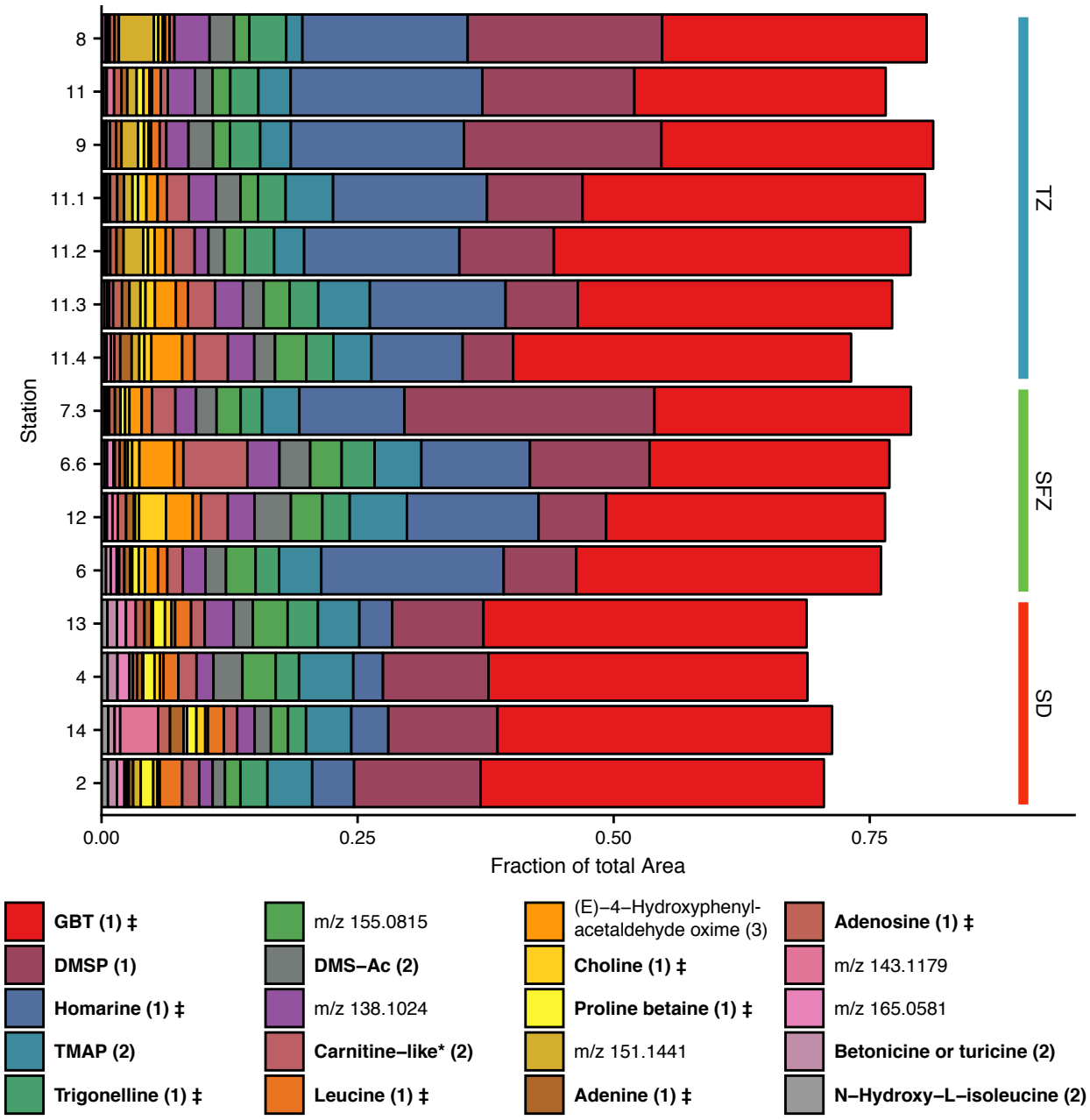

Figure 5.9: Caption next page.

Continued from previous page. Largest 20 peaks in the untargeted analysis of quality peaks detected in the aqueous fraction in the untargeted analysis. Stations are ordered by latitude (northernmost at top). The mean area of each mass feature in each station as a fraction of the mean MS total usable signal of each station is displayed. The best identifications are listed with confidence in parentheses, using previously published criteria [136], bolded identifications are unequivocal (level 1) or have both m/z and MS^2 matches in previous work (level 2). Non-bolded are a match on m/z only to a compound in the KEGG database (level 3) or unidentified (The m/z of this mass feature is displayed (level 4)). Despite efforts to deconvolute isotopes and adducts prior to data interpretation, some of the mass features with largest areas were duplicates of other mass features and are not reported here or in Table 5.7, but are retained in Dataset S5.3. *This mass feature has the same m/z and MS^2 to carnitine but not the same retention time. †I can quantify these mass features.

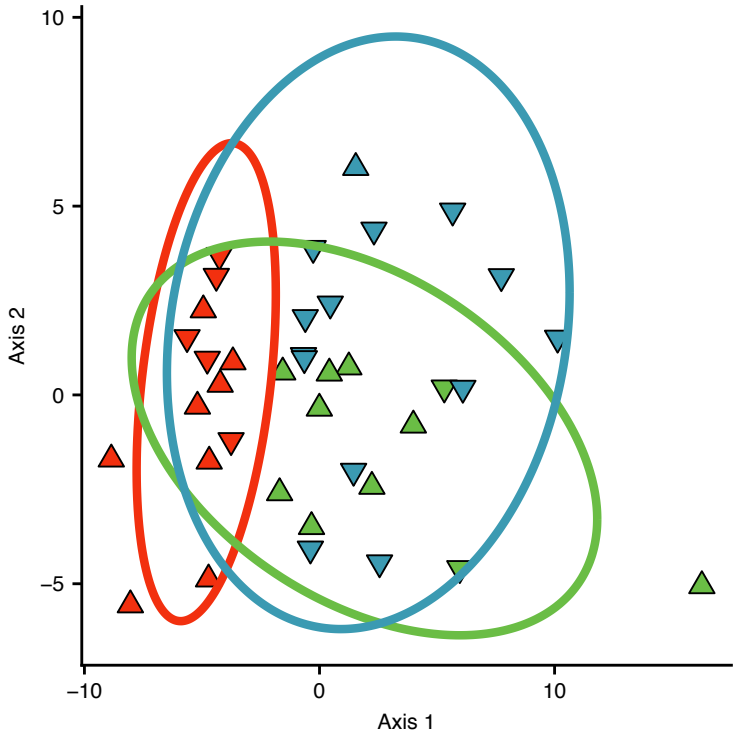

Figure 5.10: Unconstrained ordination NMDS analysis performed on Euclidean distance matrix of largest 100 peaks detected in the aqueous fraction in the untargeted analysis. Markers and ellipses are as in Figure 5.5. Stress = 0.150, $p < 0.01$.

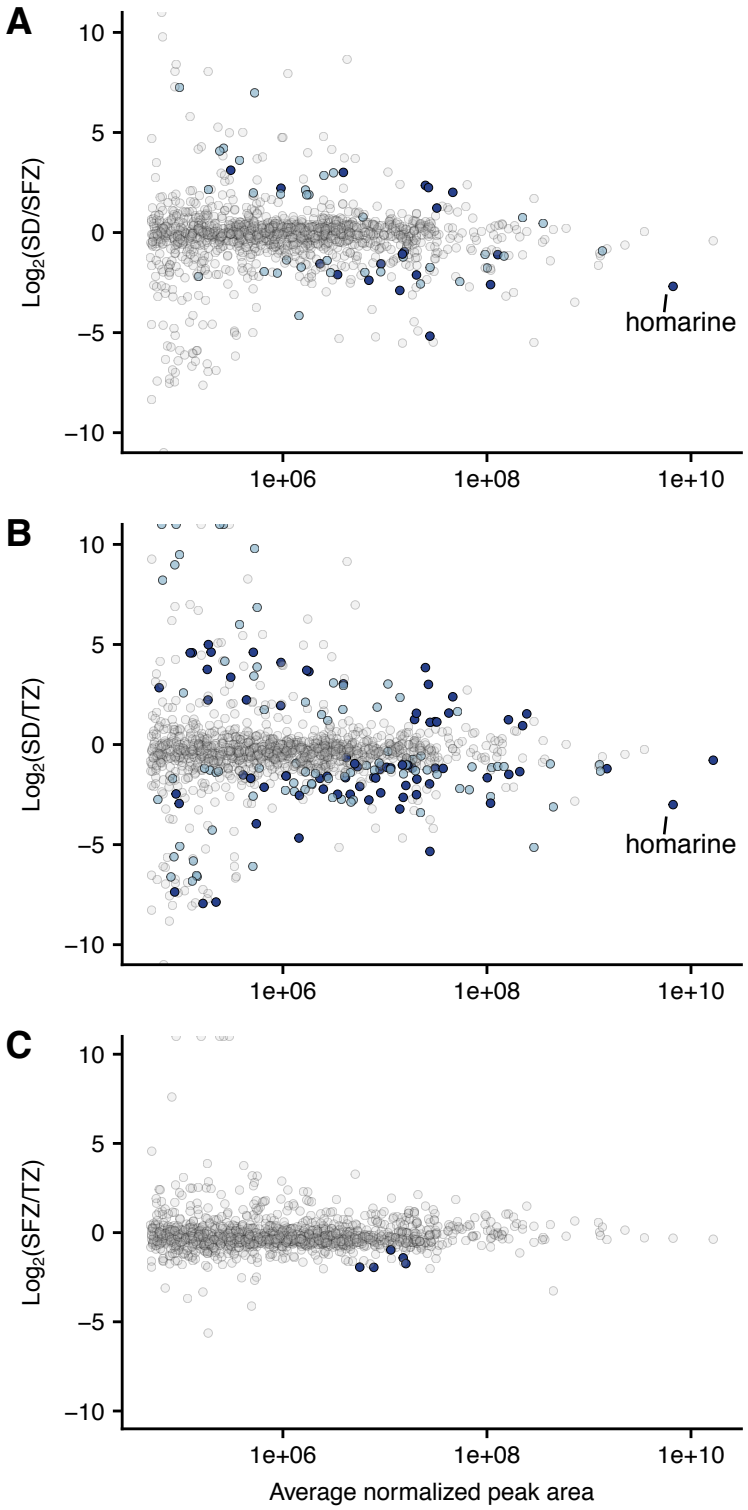

Figure 5.11: Overall results from the untargeted analysis from the aqueous extraction, highlighting mass features that show significant differences between oceanographic regimes when normalized to WV. Colors and axes are as in Figure 5.6.

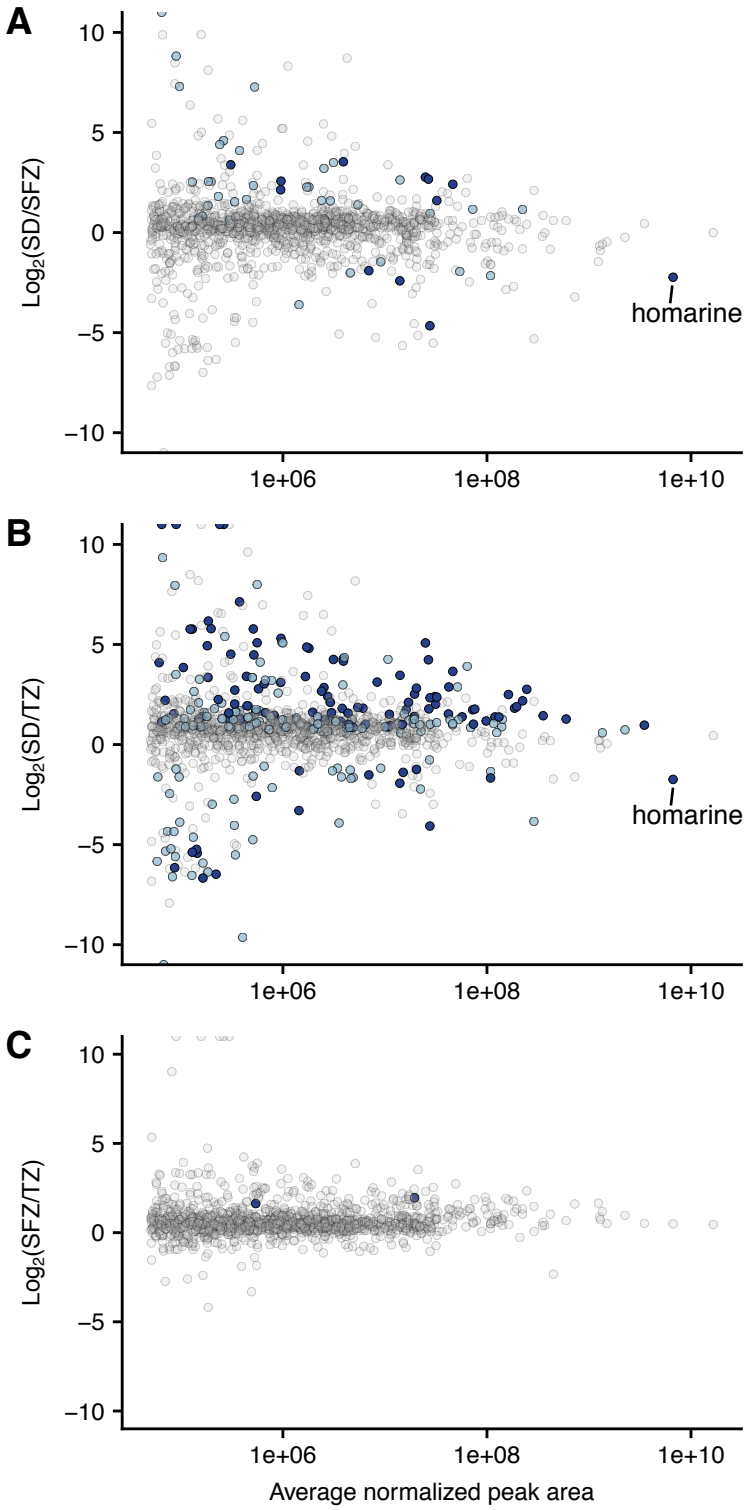

Figure 5.12: Overall results from the untargeted analysis from the aqueous extraction, highlighting mass features that show significant differences between oceanographic regimes when normalized to PC. Colors and axes are as in Figure 5.6.

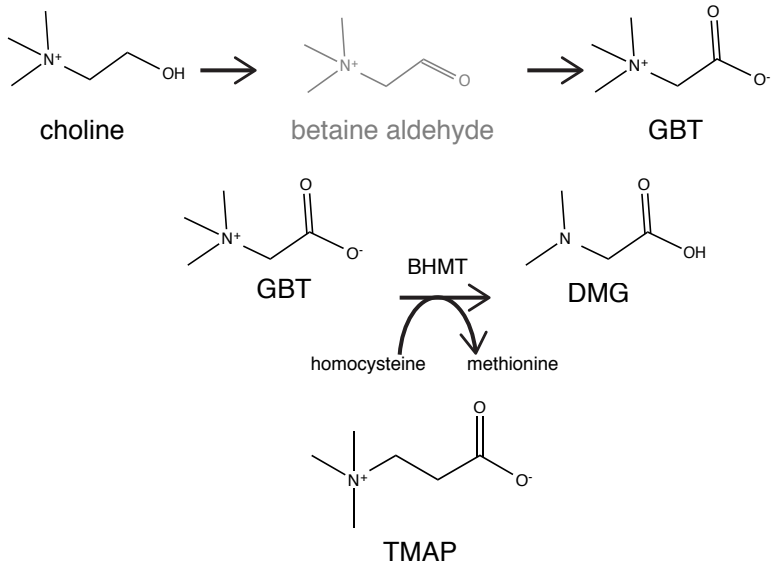

Figure 5.13: Structures of GBT and related compounds discussed in the text. Greyed out compounds were not detected in the analysis.

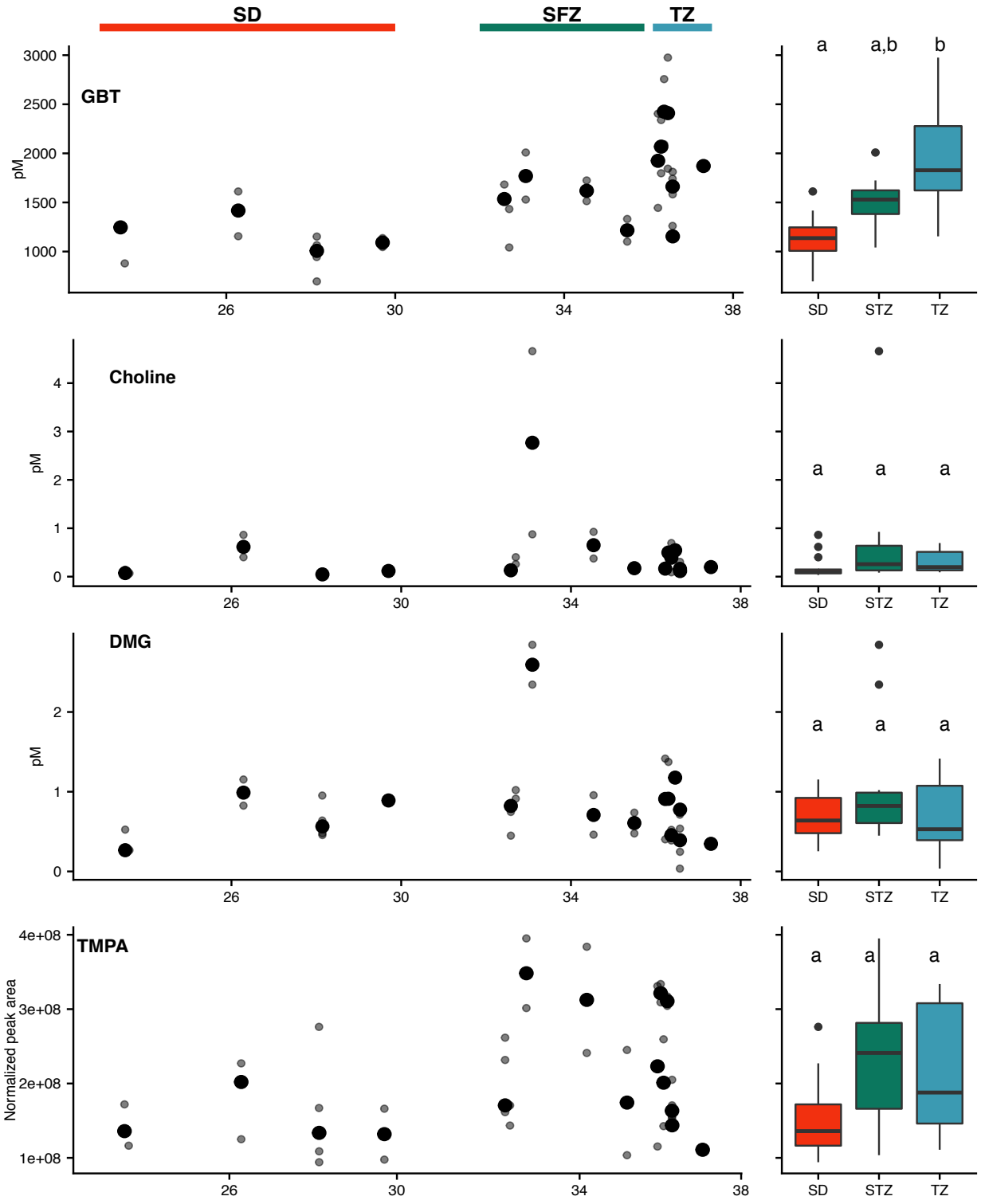

Figure 5.14: Caption next page.

Figure 5.14. *Continued from previous page.* Latitudinal trends of GBT and related compounds. Structures of these compounds are shown in Figure 5.13. Each sample is a grey dot, median of each station shown as a larger black dot. These samples have been normalized to WV. Box and whisker plots on the right show inter-quartile range (IQR) with median plotted. Whiskers are 1.5 time the IQR, with all outliers shown. Letters above box and whisker plots show statistically significant ($p < 0.05$) differences in pairwise comparisons. Oceanographic regions shown on top as different colors.

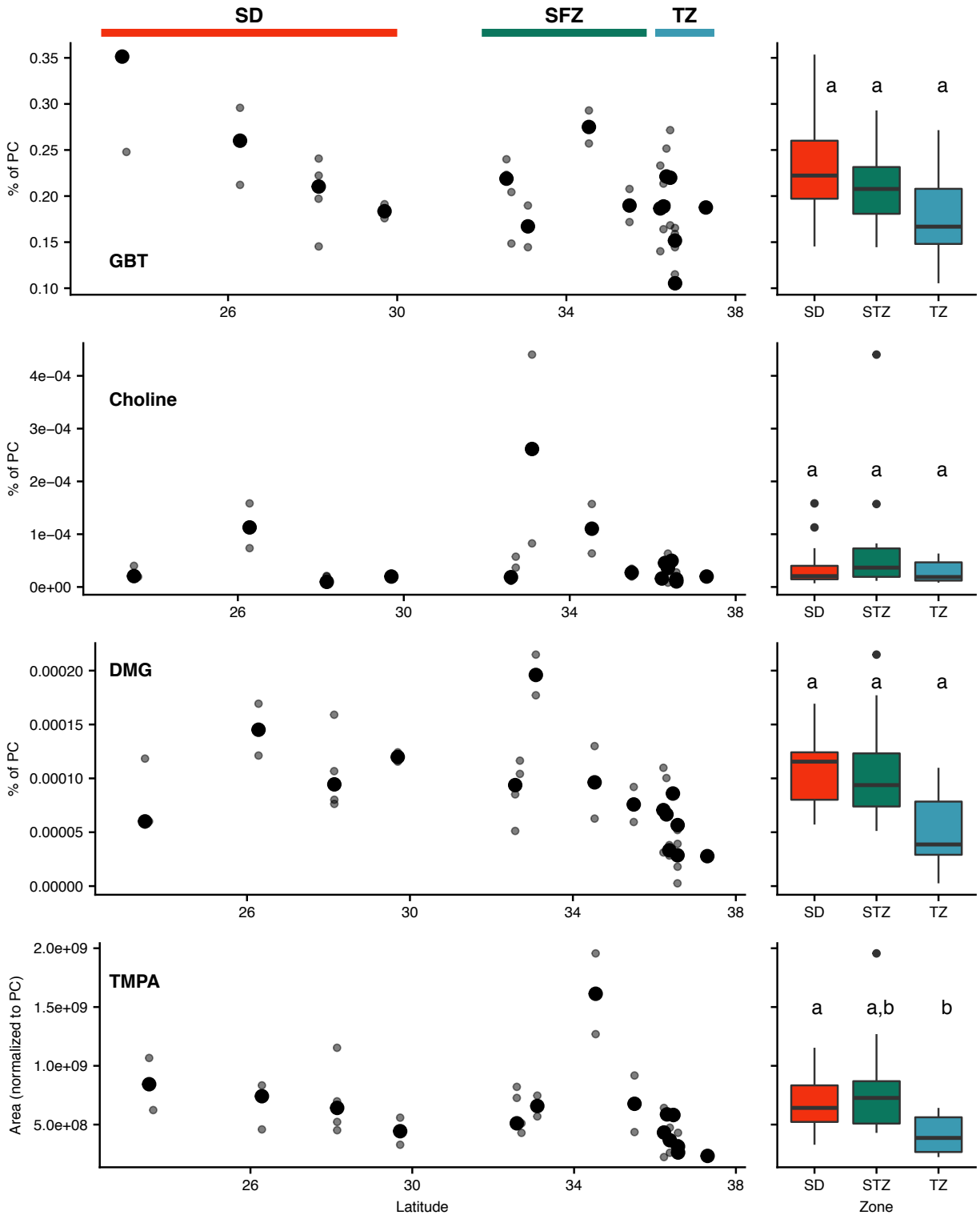

Figure 5.15: Latitudinal trends of GBT and related compounds when normalized to PC. Dots and colors are as in Figure 5.14.

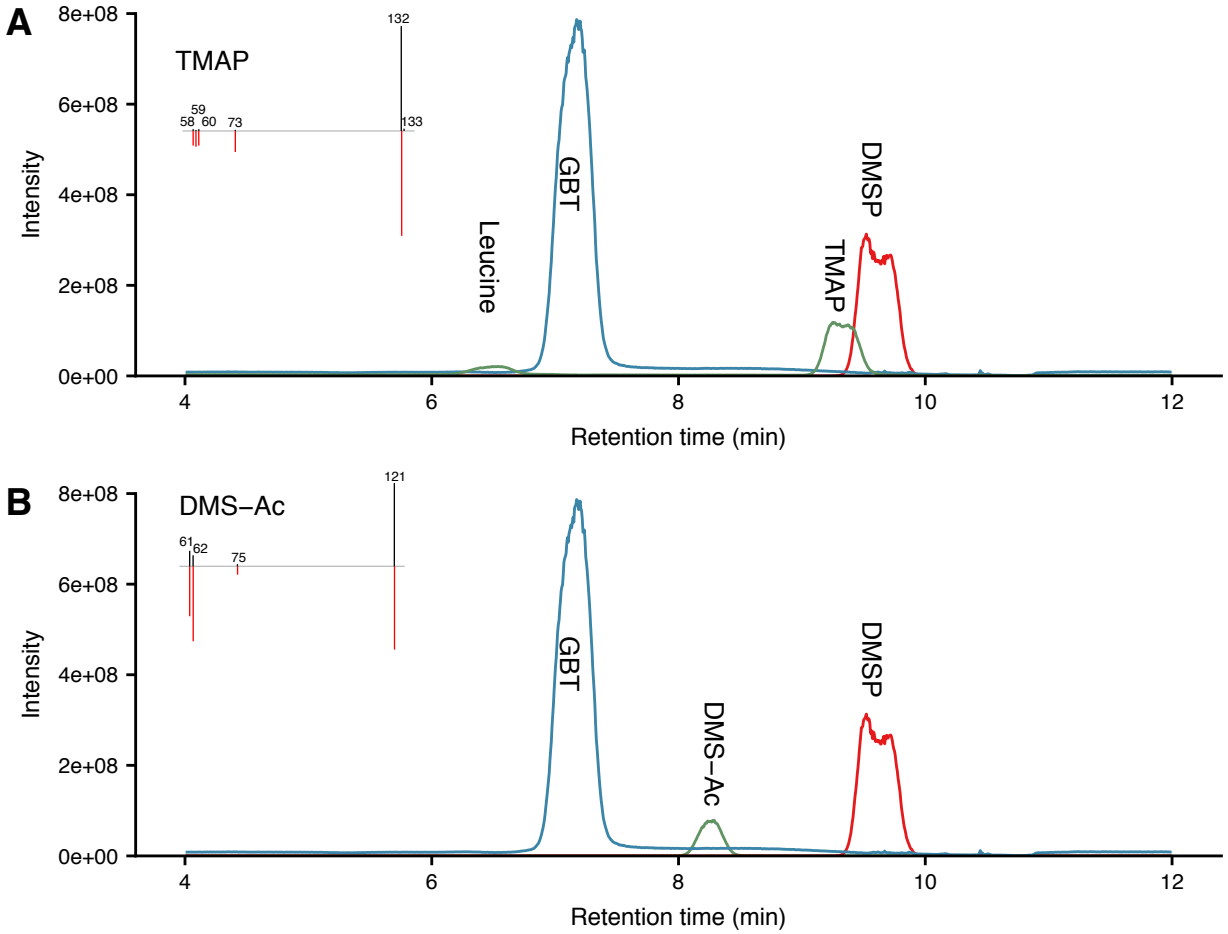

Figure 5.16: Extracted ion chromatograms of mass features putatively identified as TMAP (A, green) and DMS-Ac (B, green) match previously published chromatography in relation to GBT (blue) and DMSP (red) [213]. Spectra (insets) from the samples (black) also match well to previously published spectra (red) for these compounds [213].

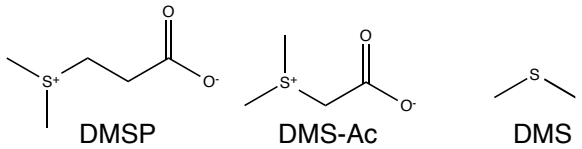

Figure 5.17: Structures of DMSP, DMS-Ac, and DMS.

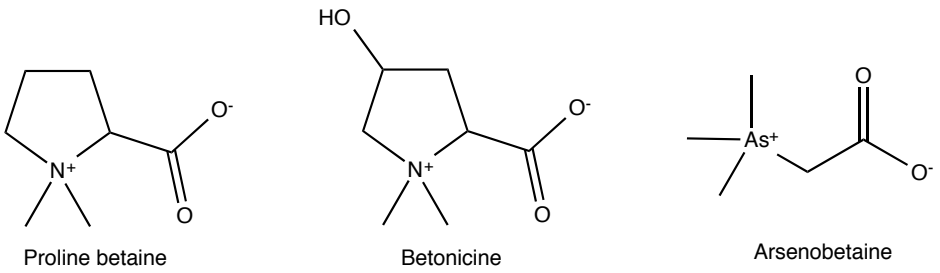

Figure 5.18: Structures of additional betaines discussed in text.

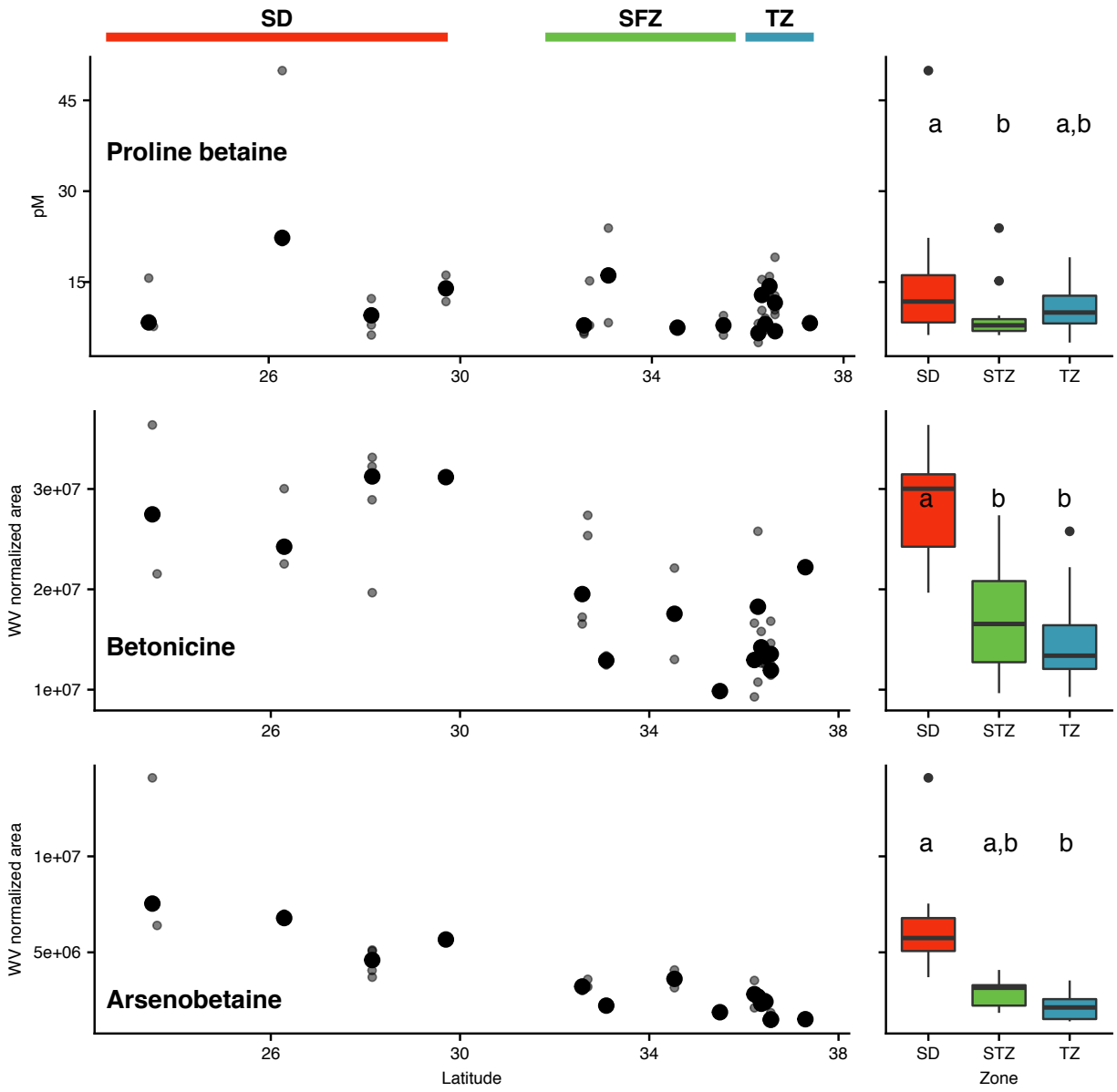

Figure 5.19: Latitudinal trends of additional betaines normalized to WV. Dots, colors, and annotated box plots are as in Figure 5.14.

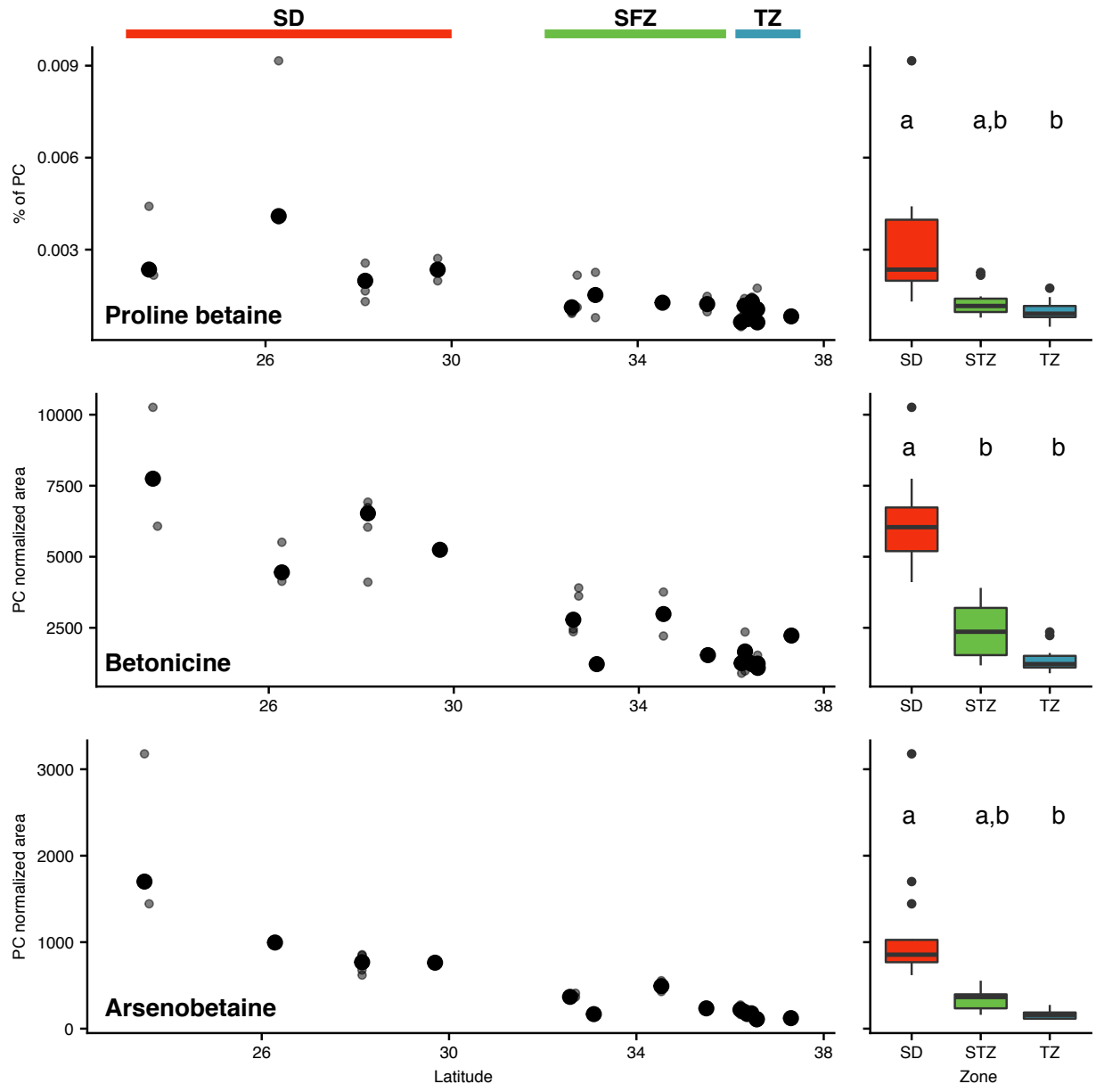

Figure 5.20: Latitudinal trends of additional betaines normalized by PC. Dots, colors, and annotated box plots are as in Figure 5.14.

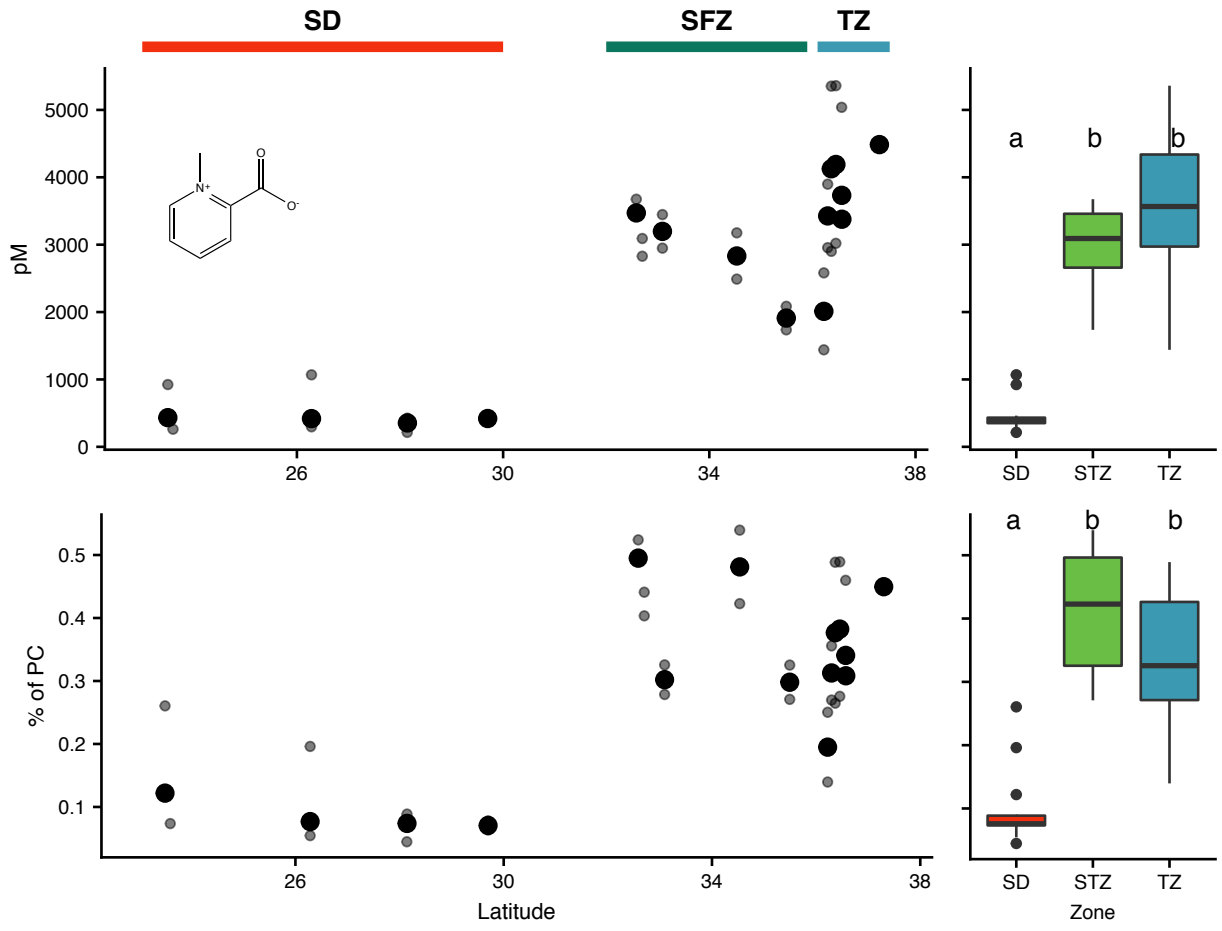

Figure 5.21: Homarine concentrations across the transect. Dots and colors are as in Figure 5.14. Samples were either normalized to WV (A), or to PC (B). The structure of homarine is also shown.

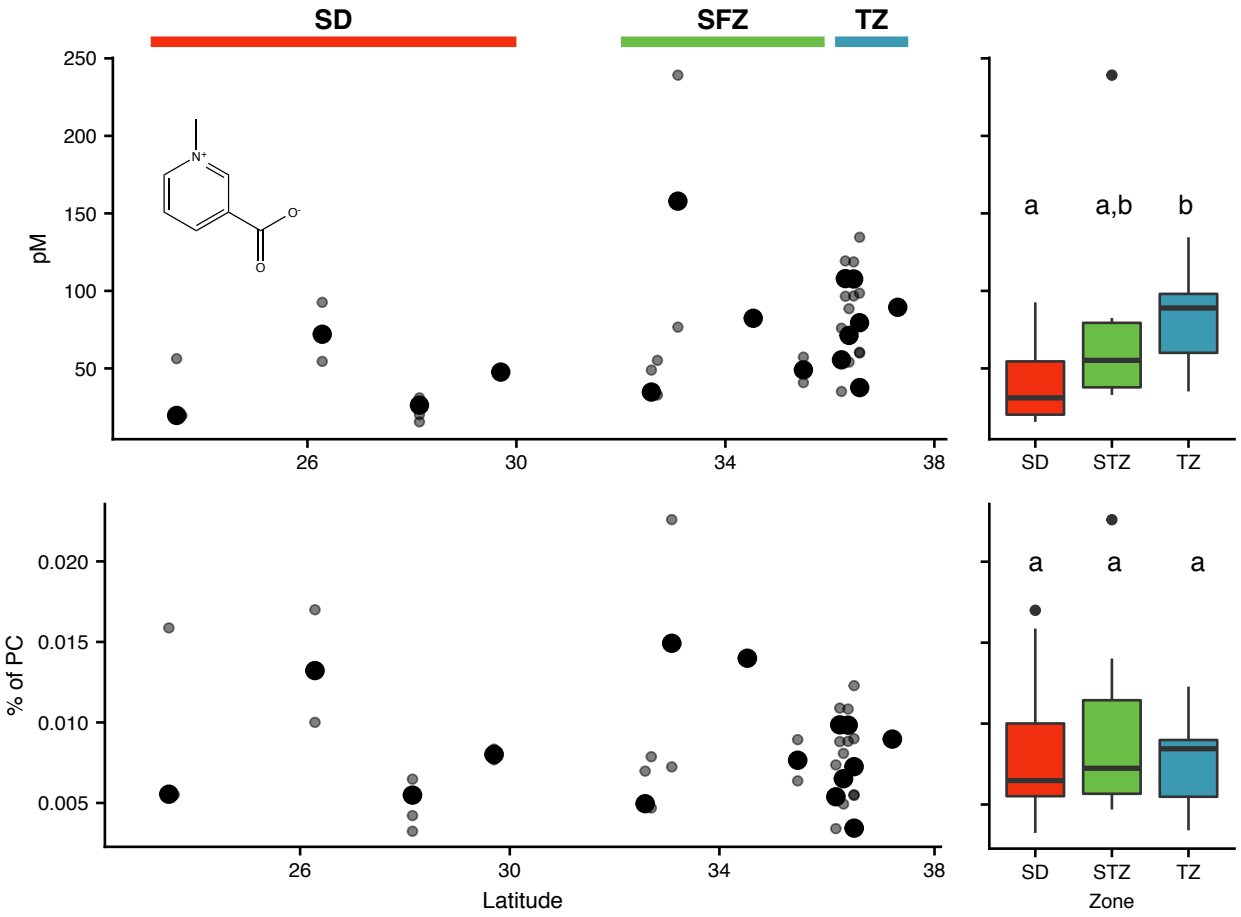

Figure 5.22: Trigonelline concentrations in all the samples. Dots and colors are as in Figure 5.14. Samples were either normalized to WV (A), or to PC (B). The structure of trigonelline is also shown.

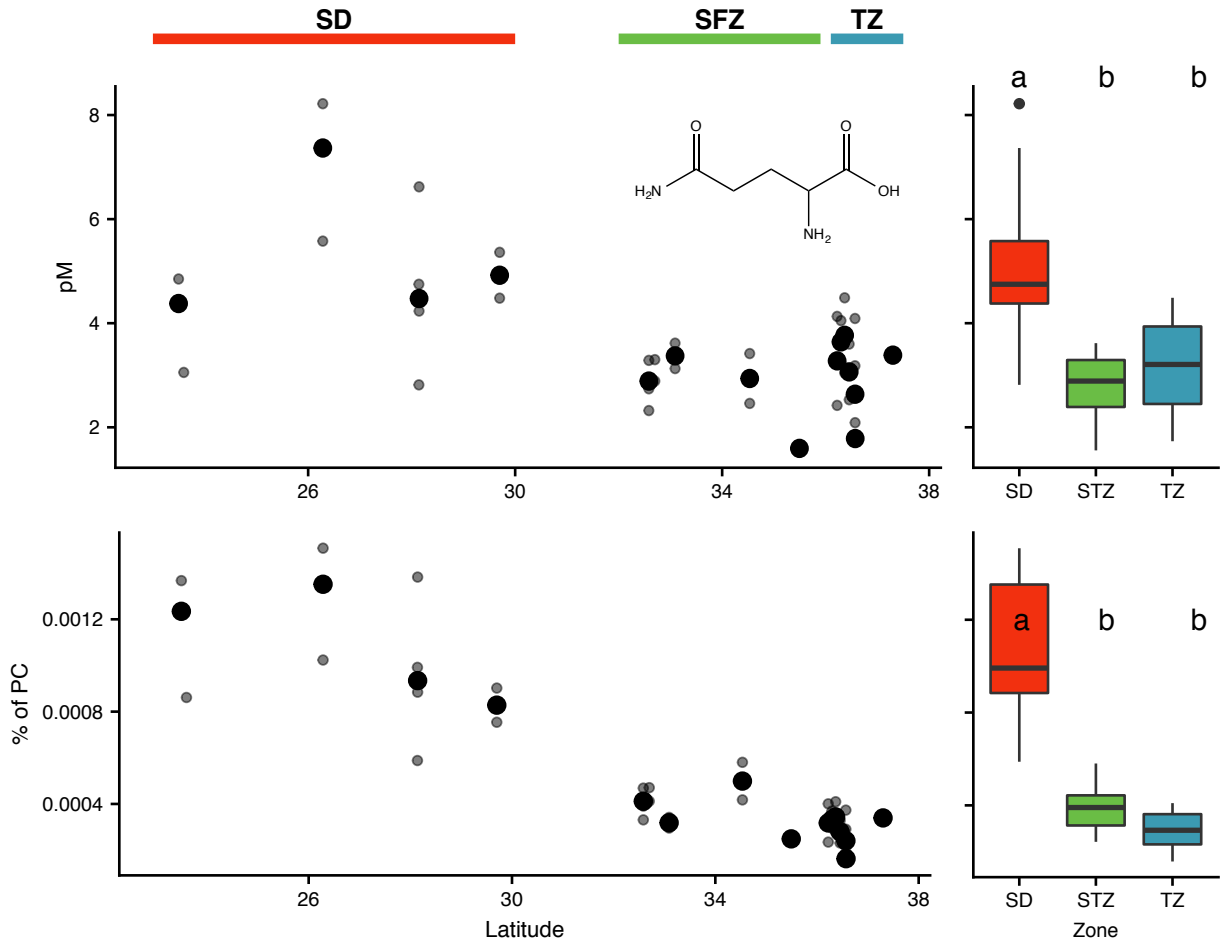

Figure 5.23: Glutamine concentrations in all the samples. Dots and colors are as in Figure 5.14. Samples were either normalized to WV (A), or to PC (B). The structure of glutamine is also shown.

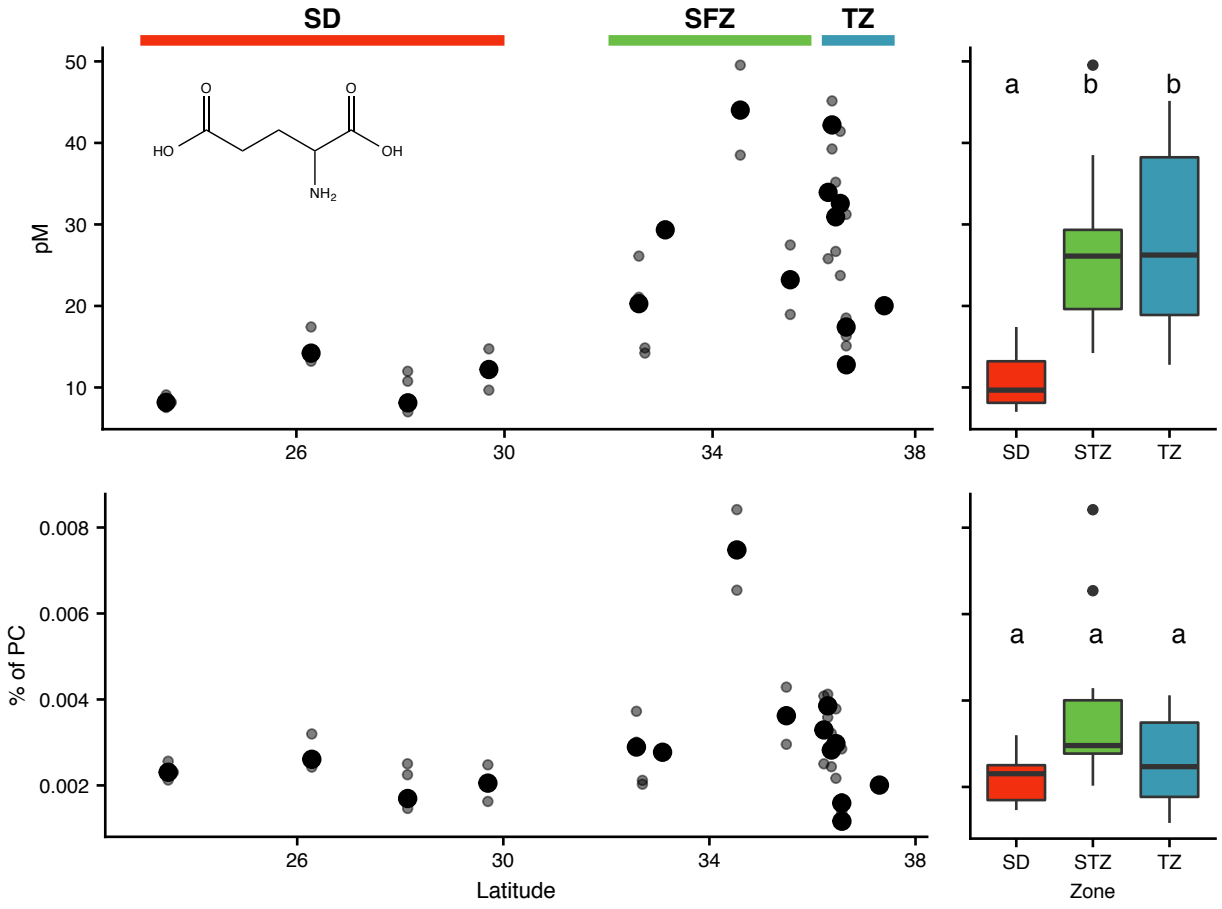

Figure 5.24: Glutamic acid concentrations in all the samples. Dots and colors are as in Figure 5.14. Samples were either normalized to WV (A) or to PC (B).The structure of glutamic acid is also shown.

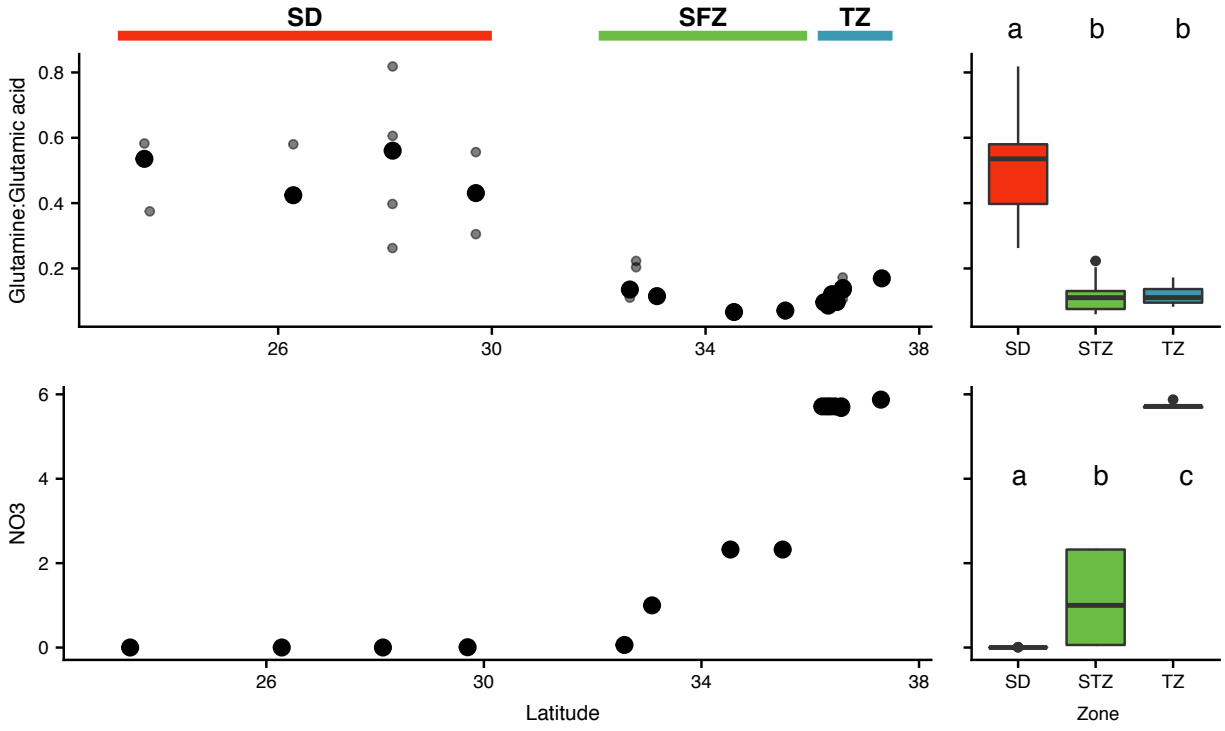

Figure 5.25: Glutamine:glutamic acid ratio (top row) NO_3^- concentrations in all the samples. Dots and colors are as in Figure 5.21.

5.10 Tables

Sample ID	Zone	Collection method	Station	NBorSB	Collection time (UTC)	Volume (L)	Latitude	Longitude	Temperature (C)	Salinity	PO_4^{2-} (μM)	NO_3^- (μM)	SiO_4^{2-} (μM)
S2C1_B	SD	Niskin	2	NB	4/20/16 15:13	11	23.50	-157.99	24.1	35.098	0.054	0.001	1.243
S2C1_C	SD	Niskin	2	NB	4/20/16 15:13	11	23.50	-157.99	24.1	35.098	0.054	0.001	1.243
S2C3_A	SD	Niskin	2	NB	4/21/16 1:22	9.5	23.60	-157.96	23.4	35.238	0.054	0.001	1.243
S4C1_A	SD	Niskin	4	NB	4/22/16 15:08	11.5	28.14	-158.00	20.1	35.171	0.045	0.001	1.300
S4C1_B	SD	Niskin	4	NB	4/22/16 15:08	11.5	28.14	-158.00	20.1	35.171	0.045	0.001	1.300
S4C1_C	SD	Niskin	4	NB	4/22/16 15:08	11.5	28.14	-158.00	20.1	35.171	0.045	0.001	1.300
S4C2_A	SD	Niskin	4	NB	4/23/16 1:57	10	28.14	-158.00	20.8	35.190	0.045	0.001	1.300
S4C2_B	SD	Niskin	4	NB	4/23/16 1:57	10	28.14	-158.00	20.8	35.190	0.045	0.001	1.300
S13C1_A	SD	Niskin	13	SB	5/1/16 15:02	10	29.70	-158.00	21.0	35.184	0.024	0.008	2.007
S13C1_B	SD	Niskin	13	SB	5/1/16 15:02	10	29.70	-158.00	21.0	35.184	0.024	0.008	2.007
S14C1_A	SD	Niskin	14	SB	5/2/16 15:05	10	26.28	-158.00	23.1	35.343	0.025	0.002	1.636
S14C1_B	SD	Niskin	14	SB	5/2/16 15:05	10	26.28	-158.00	23.1	35.343	0.025	0.002	1.636
S14C1_C	SD	Niskin	14	SB	5/2/16 15:05	10	26.28	-158.00	23.1	35.343	0.025	0.002	1.636
S6C1_A	SFZ	Niskin	6	NB	4/24/16 15:09	11	32.58	-158.00	16.6	34.625	0.070	0.061	3.438
S6C1_B	SFZ	Niskin	6	NB	4/24/16 15:09	11	32.58	-158.00	16.6	34.625	0.070	0.061	3.438
S6C1_C	SFZ	Niskin	6	NB	4/24/16 15:09	11	32.58	-158.00	16.6	34.625	0.070	0.061	3.438
S6C2_A	SFZ	Niskin	6	NB	4/25/16 1:33	10.5	32.70	-157.99	17.3	34.789	0.070	0.061	3.438
S6C2_B	SFZ	Niskin	6	NB	4/25/16 1:33	10.5	32.70	-157.99	17.3	34.789	0.070	0.061	3.438
U3_A	SFZ	Underway	6.6	NB	4/25/16 19:45	15.5	34.53	-158.00	15.3	34.508	0.219	2.323	5.418
U3_B	SFZ	Underway	6.6	NB	4/25/16 19:45	15	34.53	-158.00	15.3	34.508	0.219	2.323	5.418
U4_A	SFZ	Underway	7.3	NB	4/26/16 2:33	13.5	35.49	-158.01	14.0	34.271	0.219	2.323	5.418
U4_B	SFZ	Underway	7.3	NB	4/26/16 2:33	12	35.49	-158.01	14.0	34.271	0.219	2.323	5.418
S12C1_A	SFZ	Niskin	12	SB	4/30/16 15:05	10	33.09	-158.00	16.2	34.506	0.131	1.000	4.744
S12C1_B	SFZ	Niskin	12	SB	4/30/16 15:05	10	33.09	-158.00	16.2	34.506	0.131	1.000	4.744
S8C1_A	TZ	Niskin	8	NB	4/26/16 15:10	10.5	37.30	-158.00	11.5	34.165	0.485	5.875	9.157
S9C1_A	TZ	Niskin	9	SB	4/27/16 15:12	11.5	36.57	-158.00	12.1	34.154	0.507	5.677	9.371
S9C1_B	TZ	Niskin	9	SB	4/27/16 15:12	11	36.57	-158.00	12.1	34.154	0.507	5.677	9.371
S9C2_A	TZ	Niskin	9	SB	4/28/16 3:43	10	36.57	-158.00	12.1	34.063	0.507	5.677	9.371
S9C2_B	TZ	Niskin	9	SB	4/28/16 3:43	10	36.57	-158.00	12.1	34.063	0.507	5.677	9.371
S11C1_A	TZ	Niskin	11	SB	4/29/16 6:40	10	36.57	-158.00	12.3	34.067	0.513	5.715	9.153
U6_A	TZ	Underway	11.1	SB	4/29/16 10:00	10	36.46	-157.96	13.5	34.264	0.513	5.715	9.153
U6_B	TZ	Underway	11.1	SB	4/29/16 10:00	10.5	36.46	-157.96	13.5	34.264	0.513	5.715	9.153
U7_A	TZ	Underway	11.2	SB	4/29/16 11:14	10	36.37	-157.97	13.0	34.181	0.513	5.715	9.153
U7_B	TZ	Underway	11.2	SB	4/29/16 11:14	10	36.37	-157.97	13.0	34.181	0.513	5.715	9.153
U8_A	TZ	Underway	11.3	SB	4/29/16 12:15	10	36.30	-157.99	13.9	34.358	0.513	5.715	9.153
U8_B	TZ	Underway	11.3	SB	4/29/16 12:15	10	36.30	-157.99	13.9	34.358	0.513	5.715	9.153
U9_A	TZ	Underway	11.4	SB	4/29/16 13:19	10	36.22	-158.00	14.3	34.401	0.513	5.715	9.153
U9_B	TZ	Underway	11.4	SB	4/29/16 13:19	10	36.22	-158.00	14.3	34.401	0.513	5.715	9.153

Table 5.1: Caption on following page.

Table 5.1. *Continued from previous page.* Descriptions of individual samples. Information includes Sample ID (as used in full results); Zone (which oceanographic province, as determined in Figure 5.2; Collection method (niskin or underway); Station; NBorSB (northbound or southbound collection); Collection time (in coordinated universal time); Volume filtered (L); Latitude; and Longitude. Also included are the physical and chemical parameters used to determine oceanographic provinces, including salinity, temperature, NO_3^-, PO_4^{2-}, and SiO_4^{2-}.

		SD	SFZ	TZ
Nutrients	PO_4^{2-} (μM)	< 0.06	0.07 - 0.22	> 0.48
	NO_3^- (μM)	< 0.01	0.06 - 2.4	> 5.5
	SiO_4^{2-} (μM)	< 2.1	3.4 - 5.5	> 9
	Temperature (ºC)	> 20.5	16.5 - 13.9	< 14.3
	Salinity	> 35.15	34.2 - 34.7	< 34.4
Community	*Prochlorococcus*	high	mid	low
	Synechococcus	low	high	low
	Picoeukaryotes	low	low	high

Table 5.2: Descriptions of the oceanographic provinces used in this study. SD = subtropical domain, SFZ = subtropical frontal zone, TZ = transition zone, as discussed in the text.

Data type	Normalization	Test Zones	*R* stat	*p*
Targeted	WV	All three zones	0.23	0.001
Targeted	PC	All three zones	0.22	0.001
Untargeted	WV	All three zones	0.44	0.001
Untargeted	MSTUS	All three zones	0.58	0.001
Untargeted	PC	All three zones	0.46	0.001
Targeted	WV	SD vs SFZ	0.17	0.003
Targeted	PC	SD vs SFZ	0.1	0.035
Untargeted	WV	SD vs SFZ	0.5	0.001
Untargeted	MSTUS	SD vs SFZ	0.66	0.001
Untargeted	PC	SD vs SFZ	0.36	0.001
Targeted	WV	SD vs TZ	0.37	0.001
Targeted	PC	SD vs TZ	0.32	0.001
Untargeted	WV	SD vs TZ	0.62	0.001
Untargeted	MSTUS	SD vs TZ	0.78	0.001
Untargeted	PC	SD vs TZ	0.67	0.001
Targeted	WV	SFZ vs TZ	0.16	0.009
Targeted	PC	SFZ vs TZ	0.24	0.001
Untargeted	WV	SFZ vs TZ	0.16	0.018
Untargeted	MSTUS	SFZ vs TZ	0.31	0.002
Untargeted	PC	SFZ vs TZ	0.39	0.001

Table 5.3: ANOSIM results exploring multivariate differences between oceanographic regimes. Analyzed data are either targeted (all compounds detected in all samples through the targeted analysis) or untargeted (the largest 100 mass features from the aqueous fraction of the untargeted analysis); Normalization is technique for normalization, WV = water volume, PC = particulate carbon, MSTUS = MS total usable signal; R statistic and p values from ANOSIM analysis on distance matrices as described in Figures 5.5 and 5.10.

Data type	Normalization	Test variables	*R* stat	*p*
Targeted	WV	AMorPM	0.02	0.141
Targeted	WV	Day_Extract	-0.01	0.533
Targeted	WV	SampType	0.28	0.004
Targeted	WV	NBorSB	0.16	0.001
Targeted	PC	AMorPM	0.05	0.043
Targeted	PC	Day_Extract	0	0.479
Targeted	PC	SampType	0.07	0.199
Targeted	PC	NBorSB	0.07	0.017
Untargeted	WV	AMorPM	0.05	0.088
Untargeted	WV	Day_Extract	0.11	0.018
Untargeted	WV	SampType	0.31	0.003
Untargeted	WV	NBorSB	0.09	0.02
Untargeted	MSTUS	AMorPM	0.04	0.156
Untargeted	MSTUS	Day_Extract	0.07	0.077
Untargeted	MSTUS	SampType	0.11	0.092
Untargeted	MSTUS	NBorSB	0.13	0.005
Untargeted	PC	AMorPM	0.06	0.049
Untargeted	PC	Day_Extract	0.05	0.142
Untargeted	PC	SampType	0.08	0.194
Untargeted	PC	NBorSB	0.24	0.001

Table 5.4: ANOSIM results exploring multivariate differences between variables other than oceanographic regimes. Data type; Normalization; R statistic; and p values are as in Table 5.3. Test variables: AMorPM = whether samples were taken during the day or night; Day_Extract = which day (of a total of 4) the samples were extracted; NBorSB = whether the samples were collected on the northbound or southbound leg of the journey; SampType = whether samples were collected via niskin or underway intake.

Compound	pM	% of PC	% of PN
Homarine	200 - 5000	0.06 - 0.8	0.07 - 0.6
GBT	700 - 3000	0.1 - 0.4	0.1 - 0.5
Guanine	70 - 3000	0.006 - 0.5	0.03 - 3
Sucrose	10 - 800	0.003 - 0.2	
Adenosine	10 - 300	0.006 - 0.07	0.02 - 0.3
Glutathione Disulfide	10 - 200	0.01 - 0.1	0.02 - 0.2
Trigonelline	20 - 200	0.005 - 0.03	0.004 - 0.03
Guanosine	7 - 100	0.003 - 0.03	0.01 - 0.1
Adenine	10 - 100	0.002 - 0.03	0.01 - 0.2
Citrulline	8 - 60	0.001 - 0.009	0.003 - 0.03
Glutamic acid	7 - 50	0.001 - 0.008	0.001 - 0.01
Homoserine	3 - 100	5e-04 - 0.007	7e-04 - 0.01
Taurine	4 - 60	2e-04 - 0.002	7e-04 - 0.007
Proline betaine	5 - 50	7e-04 - 0.01	6e-04 - 0.01
Glutamine	2 - 8	2e-04 - 0.002	3e-04 - 0.004
Cytidine	0.06 - 10	2e-05 - 0.004	5e-05 - 0.009
Aspartic acid	1 - 4	9e-05 - 6e-04	1e-04 - 9e-04
Isethionic Acid	0.6 - 4	4e-05 - 3e-04	
Arginine	0.5 - 10	1e-04 - 0.001	4e-04 - 0.004
Proline	0.3 - 5	6e-05 - 4e-04	8e-05 - 5e-04
Phenylalanine	0.5 - 2	2e-04 - 6e-04	1e-04 - 5e-04
Citric Acid	0.4 - 2	6e-05 - 5e-04	
DHPS	0.1 - 6	2e-05 - 6e-04	
Lysine	0.3 - 2	6e-05 - 3e-04	1e-04 - 5e-04
DMG	0.04 - 3	3e-06 - 2e-04	4e-06 - 3e-04
Serine	0.2 - 4	2e-05 - 6e-04	4e-05 - 0.001
Valine	0.2 - 2	4e-05 - 3e-04	5e-05 - 3e-04
Leucine	0.2 - 1	4e-05 - 3e-04	5e-05 - 3e-04
Threonine	0.09 - 2	1e-05 - 2e-04	2e-05 - 3e-04
Malic Acid	0.2 - 1	3e-05 - 1e-04	
Xanthine	0.1 - 1	1e-05 - 2e-04	6e-05 - 8e-04
Vitamin B3	0.09 - 1	2e-05 - 2e-04	2e-05 - 2e-04
Asparagine	0.02 - 1	3e-06 - 2e-04	1e-05 - 4e-04
Alanine	0.05 - 2	7e-06 - 2e-04	2e-05 - 4e-04
Choline	0.03 - 5	7e-06 - 4e-04	9e-06 - 5e-04
Tryptophan	0.06 - 0.9	3e-05 - 3e-04	4e-05 - 4e-04
Uridine	0.06 - 1	2e-05 - 4e-04	3e-05 - 6e-04
AMP	0.007 - 0.9	3e-06 - 3e-04	9e-06 - 8e-04
Histidine	0.04 - 0.9	1e-05 - 2e-04	4e-05 - 8e-04
Tyrosine	0.01 - 0.8	5e-06 - 1e-04	4e-06 - 1e-04
Thymine	0.03 - 1	9e-06 - 1e-04	2e-05 - 3e-04
Aminobutyric Acid	0.04 - 0.6	8e-06 - 1e-04	1e-05 - 2e-04
Thymidine	0.02 - 1	7e-06 - 2e-04	1e-05 - 3e-04
Gluconic Acid	0.004 - 2	8e-07 - 3e-04	
Isoleucine	0.05 - 0.3	9e-06 - 1e-04	9e-06 - 1e-04
Methionine	0.007 - 0.4	1e-06 - 5e-05	1e-06 - 7e-05
Adenosyl Methionine	0.04 - 0.3	2e-05 - 1e-04	4e-05 - 3e-04
Pyridoxal	0.02 - 0.3	5e-06 - 8e-05	3e-06 - 7e-05
UDP-glucosamine	0.04 - 0.3	3e-05 - 1e-04	3e-05 - 1e-04
trans Hydroxyl proline	0.004 - 0.3	9e-07 - 5e-05	1e-06 - 7e-05
Cystathionine	0.003 - 0.5	9e-07 - 7e-05	2e-06 - 1e-04
Adenosyl Homocysteine	0.009 - 0.3	5e-06 - 8e-05	2e-05 - 2e-04
Vitamin B5	0.03 - 0.2	8e-06 - 5e-05	5e-06 - 4e-05
Trehalose	2e-04 - 0.3	5e-08 - 2e-04	
NAD	0.001 - 0.2	2e-06 - 1e-04	4e-06 - 3e-04
Vitamin B2	0.002 - 0.2	2e-06 - 8e-05	3e-06 - 1e-04
7-dehydrocholesterol	0.01 - 0.2	6e-06 - 1e-04	
Kynurenine	4e-04 - 0.009	1e-07 - 5e-06	2e-07 - 7e-06

Table 5.5: Continued on following page.

Table 5.5. *Continued from previous page.* Summary of absolute abundances of quantifiable compounds. Reported is the observed range of concentrations for each compound in the samples, as well as the range of contribution to the particulate organic carbon and nitrogen pools (as percent), ordered by mean concentration. Full results for each sample are given in Dataset S5.4.

Normalization	Data Aquisition	ANOVA	SD vs STF	SD vs TZ	STF vs TZ
WV	Untargeted	7.53% (1266/16822)	0.68% (114/16822)	3.87% (651/16822)	0.03% (5/16822)
PC	Untargeted	14.99% (2521/16822)	0.45% (76/16822)	15.74% (2648/16822)	0.02% (4/16822)
MSTUS	Untargeted	13.54% (2277/16822)	0.9% (152/16822)	7.48% (1259/16822)	0.03% (5/16822)
WV	Targeted	44.3% (35/79)	15.19% (12/79)	50.63% (40/79)	5.06% (4/79)
PC	Targeted	30.38% (24/79)	1.27% (1/79)	20.25% (16/79)	2.53% (2/79)

Table 5.6: Summary of univariate statistics in targeted and untargeted analysis under different normalization schemes. Percent of mass features (for untargeted analysis) or compounds (for targeted analysis) that showed a significant differences across the oceanographic regimes (ANOVA), of between individual oceanographic regimes (SD vs STF, SD vs TZ, and STF vs TZ), as described in Figure 5.11.

MF_Frac	Best Identification	*m/z*	rt	MS^2
I118.0864R7.19_HILICPos	GBT (1)	118.0863989	431	
DMSP_HILICPos	DMSP (1)	135.0473	570	
I138.0549R5.35_HILICPos	Homarine (1)	138.0548868	321	
I132.102R9.3_HILICPos	Trimethylammonium propionate (2)	132.101973	558	132.10195, 100; 133.10526, 3.2; 58.06587, 1.2; 60.08152, 1; 73.02906, 0.7; 59.0737, 0.6
I138.055R7.72_HILICPos	Trigonelline (1)	138.054973	463	
I155.0815R8.72_HILICPos		155.0814682	523	155.08147, 100; 156.08267, 4.2; 109.07644, 2; 110.08218, 1.4; 111.08986, 0.7
I121.032R8.27_HILICPos	DMS-Ac (2)	121.0319755	496	121.032, 100; 61.0114, 18.6; 62.01923, 13.2; 75.02695, 2.6; 122.03822, 2.3; 63.0271, 2.1; 103.02171, 0.9
I138.1024R7.58_HILICPos		138.1023915	455	138.05241, 100; 137.04588, 17.8; 139.123, 7.2; 139.05821, 3.5; 138.10263, 2.1; 139.14278, 1.4; 93.13782, 1.4; 139.08672, 0.9; 94.06559, 0.9; 92.13151, 0.9; 139.09798, 0.7; 138.13649, 0.7; 97.07645, 0.6
I162.1125R8.11_HILICPos	Carnitine-like (2)	162.1124785	486	162.11246, 100; 58.0659, 4.3; 163.11581, 3.7; 59.07372, 2.3
I132.102R6.53_HILICPos	Leucine (1)	132.102017	392	
I152.0707R9.08_HILICPos	(E)-4-Hydroxyphenylacetaldehyde oxime (3)	152.0706519	545	152.0706, 100; 153.07112, 3.9; 107.07332, 1.3; 152.05659, 0.8; 134.0601, 0.8; 152.04733, 0.6
I104.1074R10.68_HILICPos	Choline (1)	104.1073724	641	
I144.1019R6.44_HILICPos	Proline betaine (1)	144.1019499	386	
I151.1441R6.61_HILICPos		151.1441453	397	76.07634, 100; 152.07069, 4.5; 58.06591, 3.1; 59.07373, 3; 151.14436, 2.7; 150.0775, 2.6; 77.07966, 1.3; 152.04729, 0.9; 151.08678, 0.9; 151.08095, 0.6
I136.0618R4.6_HILICPos	Adenine (1)	136.0618268	276	
I268.104R4.35_HILICPos	Adenosine (1)	268.1040173	261	
I143.1179R9.29_HILICPos		143.1179003	557	143.08156, 100; 144.10201, 27.6; 143.1543, 18.4; 100.11252, 16.5; 97.07655, 12.3; 101.07149, 9.8; 72.08144, 8.1; 143.1179, 6.7; 119.988, 5.6; 58.06586, 4.8; 144.08504, 2.7; 144.03257, 2.5; 114.11544, 1.9; 102.06656, 1.8; 125.07105, 1.7; 144.0806, 1.2; 142.12279, 1.2; 156.66719, 0.8; 122.82355, 0.7; 137.9987, 0.7; 105.49532, 0.7; 142.78804, 0.7; 90.48379, 0.7; 101.9774, 0.7; 100.00436, 0.6; 163.30692, 0.6; 102.05906, 0.6; 56.28045, 0.6; 144.21475, 0.6; 147.0152, 0.6; 69.38133, 0.6; 55.12312, 0.6; 88.78605, 0.6; 56.40947, 0.6
I165.0581R8.33_HILICPos		165.0581105	500	165.05791, 100; 100.0248, 47.5; 61.02916, 42.9; 166.08633, 40.5; 82.01439, 29.6; 123.04052, 18.7; 101.02541, 13.1; 71.01348, 11.6; 63.02659, 11.5; 83.0165, 9.7; 103.03937, 9.7; 118.03518, 9.2; 57.03428, 9; 75.04473, 7.9; 89.04151, 7.1; 102.02234, 6.4; 124.03861, 5.2; 141.05092, 4.3; 84.01191, 3.7; 164.11777, 3.2; 166.06155, 2.6; 142.04901, 2.6; 125.03929, 2.5; 119.03594, 2.4; 165.10237, 2.3; 133.03189, 2.3; 85.02901, 1.6; 124.08722, 1.6; 165.11328, 1.5; 59.04988, 1.5; 120.03269, 1.2; 56.05026, 1.1; 148.08703, 1; 164.06677, 1; 77.04258, 1; 123.95489, 0.9; 119.03059, 0.6
I160.0969R8.37_HILICPos	Betonicine or turicine (2)	160.096881	502	160.09673, 100; 161.10011, 3.7; 88.07622, 1.1; 101.06006, 0.7
I148.0968R8.21_HILICPos	N-Hydroxy-L-isoleucine (2)	148.0968289	492	148.09677, 100; 147.04749, 5.3; 58.06589, 4.8; 104.07097, 4.3; 149.10021, 3; 102.09179, 2.5; 85.02897, 2.4; 123.96453, 1.8; 148.05115, 1; 61.01136, 0.8; 147.12098, 0.6; 146.98028, 0.6
I176.1282R7.85_HILICPos		176.1281837	471	176.12811, 100; 117.0549, 19.5; 60.08153, 19.3; 85.029, 12.8; 158.08123, 3.7; 177.13155, 3.4; 176.09189, 1.3; 71.04982, 1; 75.04476, 0.9; 159.08451, 0.7
I126.904R4.8_HILICNeg		126.9039935	288	126.90401, 100; 85.03849, 1.4; 127.89758, 0.9; 58.02827, 0.7; 125.89881, 0.6
I116.0709R8.8_HILICPos	Proline (1)	116.0709257	528	

Table 5.7: Continued on following page.

Table 5.7 continued.

MF_Frac	Best Identification	*m/z*	rt	MS^2
I225.9825R8.48_HILICPos		225.9824926	509	161.93994, 100; 179.9505, 85.8; 184.95595, 39.7; 202.96656, 38.8; 143.92939, 23.5; 96.08128, 12.8; 225.98219, 10.2; 208.18072, 10.1; 226.11858, 6.8; 180.17487, 5.6; 102.90322, 4.4; 226.10882, 4; 123.09187, 4; 225.12289, 3.7; 225.20714, 3.6; 226.17989, 3.6; 197.09218, 3; 218.45924, 2.9; 139.12289, 2.7; 163.14798, 2.6; 162.94077, 2.5; 180.95152, 2.4; 113.10753, 2; 184.09454, 1.7; 185.95741, 1.7; 208.17056, 1.6; 223.14052, 1.6; 182.12891, 1.5; 203.96838, 1.3; 185.07716, 1.3; 70.06581, 1.2; 227.13895, 1.1; 225.19571, 1.1; 83.08611, 1; 81.07046, 1; 107.08603, 0.9; 166.08463, 0.8; 137.0267, 0.7; 95.086, 0.7; 209.1842, 0.7; 226.07379, 0.7; 138.92418, 0.6; 218.47255, 0.6; 165.13873, 0.6; 97.10156, 0.6; 121.10145, 0.6; 184.10811, 0.6
I236.1492R9.91_HILICPos		236.1492202	594	236.14916, 100; 100.1125, 4.1; 144.10192, 3.7; 237.1526, 2.5; 60.08096, 1.6; 87.0446, 1.2; 58.06591, 1; 75.04472, 1; 103.03936, 0.6
I141.0659R7.2_HILICPos	Methylimidazoleacetic acid (3)	141.0658822	432	141.06696, 100; 140.06827, 9.8; 74.09713, 3.3; 142.06915, 3.2; 97.07656, 2.6
I161.0631R9.25_HILICPos		161.0631374	555	161.06315, 100; 99.04454, 49.8; 63.02707, 47; 57.03429, 31.6; 160.09687, 29.4; 162.11242, 11.2; 61.01225, 7.1; 71.04984, 5.7; 161.10009, 4.7; 81.03413, 4.5; 162.06655, 2.6; 100.04789, 1.4; 161.03802, 1.4; 103.03958, 1.2; 160.13326, 1.1; 162.10311, 1.1; 162.04302, 1; 138.01675, 0.9; 144.06566, 0.8; 60.08161, 0.7
I148.0605R12.12_HILICPos	Glutamic acid (1)	148.0604571	727	
I180.1019R8.38_HILICPos	Phenacetin (2)	180.1019072	503	180.10187, 100; 56.96556, 4.5; 181.10541, 3.5; 97.9911, 2.7; 121.08884, 1.8; 181.16999, 1.1
I148.0912R2.33_CyanoAq		148.0911669	140	148.10725, 100; 147.09932, 5; 149.10911, 4.7; 123.96465, 3.6; 149.02352, 3.6; 124.01518, 1; 121.09535, 0.9; 120.08901, 0.9
I262.7388R8.46_HILICNeg		262.7388196	508	78.9177, 100; 262.73862, 95; 80.91567, 61.9; 221.11324, 6.7; 106.03985, 5.4; 263.73972, 5.2; 82.03983, 4.4; 195.06726, 2.5; 65.01329, 2.5; 208.06229, 2.4; 80.91299, 2; 66.00852, 1.9; 181.05157, 1.6; 196.06206, 1.6; 236.09398, 1.4; 193.05147, 1.4; 130.04019, 1.3; 209.08287, 1.3; 221.08253, 1.2; 123.39746, 1.2; 93.88864, 1.2; 270.42453, 1.2; 92.02436, 1.1; 96.66238, 1.1; 87.81324, 1.1; 135.94133, 1.1; 56.96875, 1; 281.11609, 1; 73.09868, 1; 89.01331, 1; 172.09268, 1; 73.2831, 1; 60.62236, 1
I707.9976R14.44_HILICPos		707.9976088	867	
I161.0921R9.24_HILICPos	N(gamma)-Acetyldiaminobutyrate (3)	161.0921209	554	161.06315, 100; 99.04454, 49.8; 63.02707, 47; 57.03429, 31.6; 160.09687, 29.4; 162.11242, 11.2; 61.01225, 7.1; 71.04984, 5.7; 161.10009, 4.7; 81.03413, 4.5; 162.06655, 2.6; 100.04789, 1.4; 161.03802, 1.4; 103.03958, 1.2; 160.13326, 1.1; 162.10311, 1.1; 162.04302, 1; 138.01675, 0.9; 144.06566, 0.8; 60.08161, 0.7
I236.1493R8.23_HILICPos		236.1492911	494	236.14915, 100; 212.89696, 7.5; 189.88111, 6.9; 237.1525, 2.2; 58.06591, 2; 60.08154, 1.9; 235.91299, 1.6; 87.04463, 1.3; 131.07044, 1.2; 75.0447, 1; 57.03428, 0.7; 59.07374, 0.6; 171.87037, 0.6
I252.1092R3.46_HILICPos	Deoxyadenosine (3)	252.1091636	208	136.0618, 100; 117.05491, 7.4; 99.0445, 2.7; 241.98735, 2; 137.0652, 1.5; 252.20668, 1.5; 62.98244, 1; 212.03279, 1; 217.78203, 0.9; 235.1805, 0.9; 252.10921, 0.9; 73.02917, 0.9; 223.97739, 0.8; 251.1544, 0.8; 265.00394, 0.6
I139.0868R7.52_HILICPos		139.0867526	451	138.05502, 100; 139.05834, 45.7; 138.05274, 25.7; 139.123, 11.2; 140.06821, 2.5; 94.06562, 2.3; 93.13782, 1.7; 139.1428, 1.1; 140.05944, 1.1; 110.05944, 1.1; 139.08659, 1; 139.09799, 1; 92.05003, 0.7; 74.09706, 0.7; 56.96555, 0.6

MF_Frac	Best Identification	*m/z*	rt	MS^2
l252.1441R10.53_HILICPos		252.1441108	632	104.10743, 100; 252.14412, 78.9; 132.10202, 19.9; 251.15427, 7.9; 252.20738, 6.7; 252.1571, 4.4; 223.81908, 4.2; 253.14809, 2.6; 105.11068, 2.5; 242.99089, 2.4; 153.73599, 1.9; 50.66918, 1.7; 128.48116, 1.7; 234.25095, 1.5; 235.17966, 1.5; 149.64406, 1.5; 111.10906, 1.5; 115.67265, 1.4; 82.54028, 1.4; 53.38991, 1.4; 106.48422, 1.4; 56.23027, 1.4; 237.37634, 1.4; 81.20952, 1.4; 121.77345, 1.3; 84.99016, 1.3; 62.80082, 1.3; 55.54034, 1.2
l156.1132R8.68_HILICPos		156.1131656	521	155.08148, 100; 156.08175, 23.6; 109.07642, 3.8; 110.08195, 3.1; 111.08985, 1.3; 156.10057, 1.1; 137.07112, 0.7; 157.08665, 0.6
l143.0816R7.97_HILICPos	Ectoine (2)	143.0815933	478	143.08154, 100; 97.07651, 26.6; 144.10197, 5.9; 143.15434, 5.5; 100.11253, 4.4; 100.03974, 3.7; 101.07139, 2.9; 143.1179, 2.6; 102.05541, 2.5; 144.08539, 2.3; 72.08145, 1.8; 99.02437, 1.2; 125.07114, 0.9; 144.08151, 0.7; 98.07987, 0.6
l128.0343R7.03_HILICNeg	L-Pyroglutamic acid (1)	128.0342966	422	
l235.1653R7.18_HILICPos		235.1652918	431	118.08653, 100; 212.89704, 24.6; 189.88106, 22.7; 144.99303, 9.7; 235.91298, 6; 234.07686, 3.1; 128.95368, 2.7; 112.07361, 2; 187.88291, 1.8; 171.87046, 1.6; 210.89899, 1.6; 199.87671, 1.3; 235.08043, 1.2; 145.99637, 1.1; 211.90228, 1.1; 235.10737, 1; 153.85997, 0.9; 194.88691, 0.9; 222.89253, 0.6; 132.10199, 0.6
l333.1293R11.28_HILICPos		333.1293282	677	333.12909, 100; 284.0497, 14.7; 230.12606, 11.5; 274.11601, 10.8; 266.03928, 10.4; 186.09975, 7.5; 310.06569, 7; 303.11838, 5.5; 334.13226, 4.9; 255.09733, 4.7; 300.09494, 4.5; 300.095, 4.1; 136.0623, 3.6; 219.45999, 3.1; 265.08163, 3; 244.13296, 2.9; 318.10545, 2.6; 307.06567, 2.6; 289.05515, 2.4; 287.12396, 1.6; 219.44679, 1.5; 223.79799, 1.4; 230.08957, 1.3; 289.05481, 1.3; 256.10574, 1.2; 122.19897, 1; 128.95351, 1; 292.89105, 1; 59.63515, 0.8; 223.80867, 0.7; 213.0956, 0.7; 212.11546, 0.7; 283.97977, 0.6; 137.06511, 0.6; 186.09154, 0.6; 187.19127, 0.6; 74.41245, 0.6; 53.36844, 0.6; 56.43122, 0.6; 80.40791, 0.6; 271.13068, 0.6; 333.91791, 0.6; 195.91821, 0.6; 118.11964, 0.6; 59.76548, 0.6
l127.0729R6.63_HILICPos	Melamine (2)	127.072859	398	127.07283, 100; 127.08672, 24.6; 126.07774, 7.4; 85.05147, 6.8; 127.06188, 5.7; 109.07639, 5.6; 128.07915, 4.6; 126.10277, 4; 127.0505, 4; 83.06098, 3.7; 111.05566, 3.7; 126.06642, 2.7; 128.07021, 2.2; 126.08862, 1.8; 110.07166, 1.7; 85.07659, 1.2; 126.11074, 1.2; 127.09811, 1; 59.06111, 0.8; 127.12323, 0.7; 58.06585, 0.6; 126.98246, 0.6; 69.0341, 0.6; 86.06051, 0.6
l162.1125R6.83_HILICPos	Carnitine-like (2)	162.1125331	410	162.11247, 100; 118.08652, 42; 163.11579, 3.6; 100.07725, 3.2; 121.95819, 2.9; 60.08146, 1.5; 163.09802, 1.5; 162.04282, 1.1; 58.0659, 1; 119.08984, 0.9; 161.10017, 0.9; 138.99285, 0.9; 139.02684, 0.9; 121.01636, 0.7; 162.10341, 0.6
DHPS_HILICNeg	DHPS	155.002	588	

Table 5.7 continued.

MF_Frac	Best Identification	*m/z*	rt	MS^2
I157.9075R9.18_HILICNeg		157.9074618	551	130.0401, 100; 157.05121, 92.8; 131.03945, 63.6; 65.01333, 47.9; 115.02912, 45.7; 158.05057, 37.9; 106.03999, 32.1; 78.91774, 29.2; 61.98717, 28.4; 91.02905, 24.1; 159.06929, 23.7; 89.0134, 21.1; 132.05575, 19.3; 157.908, 19; 158.03525, 16.2; 116.02436, 15.1; 158.00749, 13.1; 158.07146, 11.3; 157.86208, 11.1; 157.01534, 10; 66.00953, 8.9; 118.04004, 8.7; 92.04839, 8.3; 92.02428, 7.3; 90.00864, 7.3; 93.0528, 5.8; 159.0305, 5.5; 119.04789, 5.1; 116.03198, 4.4; 157.1227, 4.2; 158.95537, 3.9; 131.06053, 3.9; 105.0447, 3.8; 117.04486, 3.6; 65.99737, 3.5; 113.0597, 3.1; 131.02425, 2.5; 132.0194, 2.1; 67.02897, 2.1; 144.04311, 2; 77.01334, 1.7; 156.86501, 1.6; 143.04797, 1.6; 115.05025, 1.5; 117.01972, 1.5; 81.04464, 1.5; 102.00873, 1.5; 103.02911, 1.4; 88.01815, 1.4; 66.01673, 1.4; 64.0181, 1.3; 132.03886, 1.2; 158.89239, 1.2; 158.90653, 1.1; 158.01857, 1.1; 92.03237, 1.1; 90.01672, 1; 79.02896, 0.9; 118.05257, 0.9; 159.05571, 0.8; 156.91078, 0.8; 117.02774, 0.8; 90.03366, 0.7; 50.00249, 0.7; 74.99273, 0.6; 156.8942, 0.6; 64.00549, 0.6
I193.0894R10.12_HILICPos		193.0893909	607	131.07036, 100; 194.11751, 66.7; 135.04405, 22.4; 57.03424, 16.4; 103.03934, 14.7; 71.04983, 14.2; 58.0659, 9.2; 193.13382, 8; 115.96445, 7.8; 113.0601, 7.4; 194.08461, 5.6; 99.04459, 5.4; 61.02917, 4.8; 72.08151, 4.1; 165.13885, 3.8; 192.17484, 3.7; 192.07715, 3.6; 192.12274, 3.5; 192.13847, 3.4; 128.01953, 1.9; 193.0697, 1.9; 193.16972, 1.8; 117.05454, 1.3; 71.01353, 1.3; 195.9892, 1.2; 151.03529, 1.1; 123.01141, 1.1; 132.07381, 1; 59.04987, 1; 156.99062, 1; 193.14276, 0.9; 103.80431, 0.9; 141.74794, 0.9; 99.48976, 0.9; 184.00227, 0.9; 157.52861, 0.8; 178.25116, 0.8; 52.52118, 0.8; 87.38175, 0.8; 82.3164, 0.8; 180.9104, 0.8; 65.83685, 0.8; 63.69588, 0.7; 73.72759, 0.7; 51.03787, 0.7; 67.11774, 0.7
I197.9728R10.62_HILICNeg		197.9727682	637	162.83834, 100; 197.80754, 40.9; 160.84132, 38.8; 72.94194, 20; 171.06695, 8.9; 70.94448, 5.5; 89.96853, 5.5; 65.01331, 5.4; 74.93823, 5.2; 157.05122, 4.9; 156.05591, 4.4; 198.07781, 3.2; 106.03993, 2.9; 130.04009, 2.5; 105.04466, 2.5; 132.05576, 2; 144.0558, 1.9; 87.97144, 1.6;
I121.0399R3.54_HILICPos		121.0398909	212	121.03986, 100; 93.04524, 18; 122.07141, 3.9; 122.0413, 3.4; 122.09657, 2.3; 120.08106, 1.3; 120.09087, 1.1; 66.03454, 1; 121.07638, 0.8; 81.04531, 0.6; 122.06029, 0.6
I143.0816R8.4_HILICPos	Ectoine (2)	143.0815862	504	143.08153, 100; 102.0554, 9.1; 97.0765, 8.7; 144.10194, 7.1; 143.15432, 6.3; 100.11253, 5.7; 101.07145, 3.1; 143.11795, 2.6; 72.08144, 2; 144.08537, 1.8; 99.02437, 1.5; 144.0825, 1.3; 85.02904, 1.3; 125.07113, 0.9; 58.06586, 0.6; 100.03974, 0.6; 144.03233, 0.6

Table 5.7 continued.

MF_Frac	Best Identification	*m/z*	rt	MS^2
I267.0726R11.17_HILICNeg	Inosine (3)	267.0725633	670	87.00763, 100; 101.02334, 66.9; 267.07239, 46.9; 59.01266, 45.3; 71.01268, 38.7; 106.03993, 36.2; 65.01336, 24.8; 89.02328, 23.6; 237.06181, 22.8; 166.05331, 20.8; 75.00769, 19.7; 105.01821, 16.7; 85.02824, 16.7; 66.00861, 16.2; 132.05573, 15.8; 82.03983, 15.2; 130.04016, 14.9; 157.05136, 14; 171.0668, 13.9; 113.0233, 13.8; 198.07771, 12.8; 73.02834, 11.6; 119.0341, 11.5; 156.056, 11.3; 120.0556, 6; 131.03499, 5.9; 61.98724, 5.9; 129.01842, 5.6; 105.04465, 5.6; 223.89104, 5.4; 107.03526, 5.3; 213.07832, 5.2; 144.05537, 5.1; 81.04465, 5; 240.08826, 5; 92.02432, 4.6; 121.05091, 4.6; 145.05051, 4.6; 83.02383, 4.6; 133.05109, 4.5; 142.04041, 4.5; 87.00475, 4.4; 169.05116, 4.4; 83.01272, 4.4; 219.0509, 4.3; 197.08218, 4.3; 148.05081, 4.2; 121.03965, 4.2; 159.06712, 4.2; 65.99729, 4.1; 223.87837, 4; 181.05141, 4; 225.07748, 3.9; 118.04009, 3.8; 224.09419, 3.8; 147.0553, 3.8; 183.06709, 3.8; 91.02905, 3.7; 196.06245, 3.5; 158.04716, 3.3; 134.03479, 3.2; 119.06056, 3; 72.40837, 3; 242.88736, 2.9; 101.73653, 2.9; 139.20265, 2.9; 92.92683, 2.9; 61.76694, 2.8; 96.97018, 2.8; 219.55209, 2.7; 224.73079, 2.7; 268.71353, 2.7; 199.0728, 2.7; 72.62506, 2.7; 130.50117, 2.6; 60.53605, 2.6; 73.73257, 2.6; 98.91158, 2.6; 184.33243, 2.5; 123.42855, 2.5; 60.37306, 2.5; 66.97381, 2.5; 78.34675, 2.5; 209.02884, 2.4; 88.15565, 2.4; 220.96109, 2.4; 53.91296, 2.4; 71.03653, 2.2
I204.1232R7.22_HILICPos	Acetyl-L-carnitine (1)	204.1231804	433	
I176.1281R10.33_HILICPos		176.1281461	620	176.1281, 100; 117.05488, 10.9; 60.08152, 9.9; 177.13156, 3.1; 116.10727, 2.7; 55.05502, 1.5; 116.07079, 1.3; 99.04455, 1.2; 57.03426, 0.7
I160.0969R7.56_HILICPos	Betonicine or turicine (2)	160.0969025	454	160.09681, 100; 116.04734, 11.6; 160.03892, 3; 160.13305, 2.9; 160.10809, 2.8; 161.10014, 2.8; 118.0865, 2; 159.09179, 1.4; 72.08142, 1.1; 114.09158, 1; 142.08618, 0.7
I193.0531R10.87_HILICPos	Monodechloroaminopyrrolnitrin (3)	193.0530528	652	131.03404, 100; 113.0237, 56.5; 63.02704, 42.7; 194.11765, 26.2; 85.02898, 22.9; 193.05305, 22.5; 128.0195, 13.6; 135.04428, 7.5; 75.02695, 7.5; 130.05002, 7.3; 151.03531, 6.2; 103.02172, 5.8; 73.02911, 5.8; 110.00912, 4.2; 193.13374, 4.1; 129.0197, 4.1; 115.96432, 3.2; 58.06593, 2.9; 102.05529, 2.5; 165.13879, 2.5; 172.06039, 2.2; 132.03738, 2.1; 146.02988, 2.1; 194.08443, 1.9; 100.0248, 1.9; 130.01697, 1.8; 152.03667, 1.6; 61.01142, 1.5; 192.12321, 1.2; 192.1746, 1.1; 129.01488, 0.9; 72.08141, 0.8; 194.12791, 0.8; 111.00975, 0.7; 121.03186, 0.7; 123.40791, 0.7; 192.13838, 0.7; 101.02379, 0.6; 147.03093, 0.6; 123.01137, 0.6; 149.04453, 0.6; 153.03296, 0.6; 193.10825, 0.6
I317.0553R12.1_HILICNeg	Prekinamycin (3)	317.0552535	726	317.0549, 100; 116.92745, 66.9; 80.96398, 57.3; 225.00706, 51.9; 164.98569, 24; 221.04355, 17.7; 194.90558, 14.9; 94.97974, 13.7; 106.03983, 12.4; 152.98558, 11.6; 221.03024, 5.7; 148.99037, 5; 125.02339, 4.9; 61.98717, 4.8; 318.05862, 4.1; 296.6268, 3.7; 135.14255, 3.7; 230.50868, 3.5; 82.03979, 3.3; 140.51195, 3.3; 107.85749, 3.3; 97.02851, 3.2; 299.63806, 3.2; 134.97481, 3.1; 164.51566, 3; 95.44386, 2.9; 105.91322, 2.9; 60.44644, 2.7; 124.33031, 2.7; 113.77394, 2.7
I179.0048R7.02_HILICPos	Arsenobetaine (2)	179.0048127	421	179.00476, 100; 156.00804, 17.9; 178.12267, 11.5; 179.02379, 9; 119.99046, 7.8; 56.96555, 5.2; 119.04934, 4.1; 179.0636, 3.7; 97.99183, 3.5; 58.06587, 2.6; 180.00832, 2.6; 179.11783, 2.5; 179.12662, 2.2; 179.1541, 1.8; 160.99415, 1.7; 178.13323, 1.5; 178.15889, 1.2; 179.09252, 1.1; 136.99427, 1.1; 137.08212, 1; 140.06854, 0.8; 178.08636, 0.6; 180.06059, 0.6
I160.037R5.34_HILICPos		160.0369543	320	116.0474, 100; 160.03707, 14.5; 159.12104, 6.5; 160.13336, 3.8; 117.05075, 2.5; 160.09669, 2.1; 160.08723, 1.8; 72.0815, 1.2; 157.07372, 0.9; 120.0811, 0.6; 159.09198, 0.6; 159.11385, 0.6

Table 5.7 continued.

Table 5.7 continued.

MF_Frac	Best Identification	*m/z*	rt	MS^2
I137.0459R6.2_HILICPos	Hypoxanthine (2)	137.0458775	372	137.04588, 100; 138.05502, 19.8; 138.05203, 11.7; 136.06192, 3.4; 137.10742, 3; 95.06089, 2.9; 138.1026, 2; 136.087, 1.8; 137.08237, 1.7; 137.07117, 1.3; 136.11226, 0.9; 96.04488, 0.7; 138.04272, 0.7; 110.07154, 0.7; 119.03568, 0.6; 136.07587, 0.6; 94.06567, 0.6
I283.9066R8.4_HILICPos		283.9066272	504	NA
I139.0503R3.54_HILICPos	Urocanate (3)	139.0502618	213	121.03987, 100; 139.123, 76.6; 139.05025, 11; 139.08671, 7.9; 139.09794, 6; 140.06824, 5.1; 138.10261, 4.5; 140.08173, 4.1; 97.07649, 4.1; 122.04237, 4; 98.09685, 3.5; 111.09209, 3.3; 140.12638, 2.8; 122.07154, 2.7; 98.07172, 1.6; 140.107, 1.5; 56.96556, 1.4; 74.09708, 1.3; 138.06625, 0.9; 140.11851, 0.8; 96.08135, 0.7; 138.09018, 0.6
I174.022R5.45_HILICPos	Sulfanilic acid (2)	174.0220161	327	174.02198, 100; 128.12993, 11.1; 174.11226, 4.3; 127.12478, 2.6; 175.0258, 1.7; 174.10351, 1.2; 174.13643, 1.1; 114.0916, 0.9; 110.06043, 0.7; 93.05791, 0.7; 174.1284, 0.6
I149.0631R10.15_HILICPos		149.0631499	609	87.0446, 100; 148.10596, 85.2; 149.10841, 19.8; 149.08211, 11.1; 63.02703, 10.7; 148.09698, 7.8; 149.07092, 6.5; 150.09151, 4.7; 150.10278, 3.9; 148.08698, 3.7; 148.12276, 3.2; 88.04803, 2.8; 150.12794, 1.9; 149.06361, 1.8; 100.07604, 1.1; 89.04893, 0.7; 120.04459, 0.7; 148.13358, 0.7; 93.27065, 0.7; 116.09648, 0.6; 66.21967, 0.6; 149.6387, 0.6; 169.29317, 0.6; 107.11597, 0.6; 93.38021, 0.6; 149.02353, 0.6; 152.21892, 0.6; 68.3215, 0.6; 60.08178, 0.6; 50.91764, 0.6
I244.0828R9.94_HILICNeg		244.0828365	596	150.05518, 100; 167.05811, 70.2; 82.03997, 20.9; 110.06013, 19.3; 65.01331, 14.9; 220.20808, 14.7; 185.0688, 12.8; 142.08621, 12.5; 106.03988, 11.5; 81.04465, 10.3; 132.05551, 9.6; 182.08144, 9.6; 244.08275, 9.5; 66.00853, 9.3; 175.06148, 8.9; 121.05099, 8.9; 244.87312, 8.5; 136.05057, 8.4; 83.02374, 8.4; 120.05578, 7.8; 134.03502, 5.1; 133.03961, 5.1; 133.05084, 4.9; 109.0397, 4.8; 200.09242, 4.7; 160.05074, 4.7; 99.00759, 4.6; 148.05078, 4.3; 159.06656, 4; 157.0508, 4; 122.03528, 3.7; 186.88232, 3.6; 119.0239, 3.5; 175.09846, 3.5; 126.05511, 3.4; 226.07146, 3.3; 161.08246, 3.3; 71.01262, 3.2; 128.89101, 3.2; 147.06627, 3.1; 65.99729, 3.1; 135.0666, 3.1; 121.0284, 2.9; 201.07751, 2.9; 135.0558, 2.8; 120.01904, 2.7; 107.03548, 2.5; 154.05022, 2.5; 82.0285, 2.5; 161.04561, 2.5; 59.0127, 2.5; 150.06584, 2.5; 121.03946, 2.5; 174.06657, 2.4; 73.0282, 2.3; 61.47541, 2.3; 123.044, 2.3; 122.05999, 2.2; 202.07281, 2.2; 82.40207, 2.2; 205.29005, 2.2; 88.80005, 2.1; 61.99092, 2.1; 91.02917, 2.1; 82.81013, 2.1; 158.07213, 2; 55.17418, 2
I139.0616R3.64_HILICPos		139.0615822	219	121.0406, 100; 139.12301, 71.8; 139.05026, 11.1; 139.08671, 7.4; 139.09792, 6; 140.0683, 4.9; 138.1026, 4.3; 97.07652, 4; 122.04247, 4; 140.08179, 3.6; 98.09688, 3.2; 111.09207, 3.1; 140.1264, 2.8; 122.07155, 2.6; 140.10697, 1.6; 98.07175, 1.6; 74.09707, 1.4; 56.96557, 1; 138.06625, 0.8; 96.08129, 0.8; 138.0903, 0.7; 116.97213, 0.7; 140.11836, 0.7
I196.083R5.03_HILICPos		196.0830337	302	196.08297, 100; 181.05948, 60.4; 138.05502, 35.3; 195.18558, 6.2; 165.06481, 4.8; 195.12142, 4; 154.05331, 3.4; 59.06118, 3.1; 139.0585, 2.7; 195.08123, 2; 197.08716, 1.8; 197.12839, 1.6; 179.11783, 1.5; 195.11247, 1.5; 196.14449, 1.4; 196.1893, 1.2; 173.99789, 1.2; 89.06023, 1.1; 117.95991, 0.8; 195.14912, 0.8; 136.04726, 0.7; 109.07644, 0.6; 156.04922, 0.6
I170.1176R8.07_HILICPos		170.1175901	484	171.11278, 100; 170.11756, 46.9; 125.1075, 41.3; 153.07377, 2.4; 128.95358, 1.7; 169.09709, 1.6; 171.14905, 1.3; 100.11253, 1.2; 126.09147, 1.2; 154.0863, 1; 170.06458, 0.9; 170.10051, 0.9; 153.10224, 0.9; 142.08627, 0.7; 152.03757, 0.7; 170.12865, 0.6

MF_Frac	Best Identification	*m/z*	rt	MS^2
I209.0844R8.07_HILICPos		209.0843609	484	69.03418, 100; 145.00941, 16; 129.05472, 10.1; 117.95982, 8.9; 209.08327, 8.2; 63.027, 7.6; 209.16156, 7.6; 209.09317, 7.2; 168.02554, 7.2; 163.02, 7; 75.04472, 5.5; 97.02886, 5.5; 209.2014, 4.3; 147.0655, 4; 209.12843, 3.6; 208.09866, 3.2; 149.02682, 2.9; 209.01509, 2.8; 115.03919, 2.6; 219.89507, 2.5; 186.03619, 2.3; 208.16991, 1.8; 149.05972, 1.8; 191.08958, 1.8; 208.13329, 1.8; 73.02905, 1.5; 158.98618, 1.4; 70.03751, 1.4; 135.97018, 1.3; 208.95677, 1.2; 128.9538, 1.2; 60.08153, 1.1; 221.23969, 1.1; 168.07645, 1.1; 209.10364, 1; 96.08123, 0.9; 139.1229, 0.9; 192.1496, 0.8; 221.25227, 0.8; 210.11237, 0.6; 85.029, 0.6; 209.17502, 0.6; 101.06005, 0.6
I130.05R11.1_HILICPos	4-Oxoproline (3)	130.050048	666	130.04871, 100; 84.04593, 76; 112.03673, 20; 114.09164, 18.8; 130.08636, 15; 129.06603, 13.2; 130.1593, 10.1; 131.11806, 9.5; 129.1024, 7.4; 85.02906, 6.7; 131.03738, 5.1; 83.06093, 3.6; 102.05549, 2.7; 113.02361, 2.7; 130.06519, 2.2; 130.09745, 1.9; 111.05544, 1.3; 59.06113, 1; 131.08141, 0.9; 148.67453, 0.9; 113.07124, 0.9; 65.85174, 0.9; 123.34173, 0.8; 71.37144, 0.8; 133.68968, 0.8; 57.25389, 0.8; 90.64259, 0.8; 134.1745, 0.8; 125.9045, 0.8; 129.07736, 0.8; 91.15017, 0.7; 69.54247, 0.7; 93.25442, 0.7; 86.29333, 0.7; 85.93565, 0.7; 68.38651, 0.7; 107.66518, 0.7; 61.00084, 0.7; 76.37401, 0.7; 63.58412, 0.7; 71.51575, 0.7; 68.71929, 0.7; 51.34872, 0.7; 100.24065, 0.7; 65.43417, 0.7; 104.36656, 0.7; 106.03456, 0.7; 79.99648, 0.7; 71.54296, 0.6; 94.40057, 0.6; 112.3596, 0.6; 64.71566, 0.6; 79.14597, 0.6; 95.16517, 0.6; 55.98895, 0.6; 89.41602, 0.6; 63.96688, 0.6; 54.94443, 0.6; 104.64256, 0.6; 60.28411, 0.6; 71.06061, 0.6; 62.43143, 0.6; 71.15626, 0.6
I158.1176R5.48_HILICPos		158.117615	329	158.1176, 100; 157.17003, 6.5; 158.09705, 5.5; 158.15408, 4.1; 159.12089, 3.1; 72.08149, 2.1; 98.09679, 1.7; 95.04962, 1.6; 116.08204, 1.4; 141.00045, 1.3; 99.05571, 1.2; 158.17072, 1; 116.04729, 1; 158.04786, 0.9; 141.13855, 0.8; 98.07167, 0.6; 58.0659, 0.6
I150.0413R8.14_HILICNeg	Guanine (1)	150.0412713	488	
I206.1389R6.76_HILICPos	Pantothenol (3)	206.1388583	406	206.13869, 100; 60.08099, 13.7; 132.10201, 10.8; 147.0652, 8.6; 115.96436, 8.3; 76.07633, 7.2; 119.07056, 6.6; 87.04462, 5.5; 115.03927, 3.6; 89.0603, 3.1; 206.19044, 2.3; 123.95492, 2.1; 156.99069, 2; 217.00867, 2; 206.04573, 1.8; 207.14194, 1.7; 216.99493, 1.4; 59.04989, 1.4; 206.07584, 0.9; 83.0133, 0.9; 206.16577, 0.9; 223.78368, 0.9; 133.97475, 0.8; 59.07369, 0.8; 156.00798, 0.7; 88.05238, 0.6; 85.02898, 0.6
I177.0435R9.8_HILICPos		177.0435426	588	176.12814, 100; 178.12263, 84.6; 117.05489, 62.2; 119.0494, 27.2; 177.13136, 15.5; 60.08151, 13; 58.06592, 10.2; 135.04395, 6.7; 161.04596, 6.3; 178.0722, 6.2; 85.02885, 6; 177.11324, 5.4; 79.02186, 4.9; 177.05817, 2.9; 176.07343, 2.4; 135.04854, 2; 177.07721, 2; 118.05798, 1.6; 84.862, 1.5; 99.04434, 1.5; 177.10262, 1.4; 117.00082, 1.4; 189.32565, 1.3; 143.57462, 1.3; 178.15959, 1.2; 60.24946, 1.2; 63.55116, 1.2; 70.38406, 1.1; 60.45156, 1.1
I157.0973R7.33_HILICPos		157.0972657	440	157.0972, 100; 157.17, 17.5; 113.10772, 6.8; 115.96437, 6.1; 156.04213, 3.6; 158.09941, 3; 158.15392, 2.9; 58.06587, 2.6; 157.13371, 2.3; 111.09206, 2.3; 95.06079, 1.7; 72.08144, 1.7; 116.08203, 1.1; 157.04552, 1; 158.1176, 0.9; 156.99055, 0.8; 156.13853, 0.8; 98.07172, 0.7; 59.06113, 0.7; 99.05576, 0.7; 70.06576, 0.7; 95.04966, 0.6; 112.07608, 0.6; 74.93818, 0.6
I124.0064R10.88_HILICNeg	Taurine (1)	124.006394	653	

Table 5.7 continued.

MF_Frac	Best Identification	*m/z*	rt	MS^2
I154.0532R5.23_HILICPos		154.0532475	314	154.05328, 100; 153.06996, 8.7; 153.13859, 6.6; 154.09974, 4.1; 154.04457, 2.7; 153.10241, 2.4; 58.06589, 2.3; 155.05666, 1.6; 59.0737, 1.5; 95.06085, 1.3; 73.08925, 1.2; 154.07872, 1.2; 155.08169, 1.1; 112.08733, 1.1; 154.08659, 0.8; 154.13416, 0.8; 154.12268, 0.8; 153.11345, 0.8; 154.14151, 0.7; 155.11818, 0.7; 125.07911, 0.7; 137.07109, 0.6; 97.08437, 0.6; 136.08672, 0.6
I175.119R17.83_HILICPos	Arginine (1)	175.1190291	1070	
I170.0684R16.91_HILICPos		170.0683824	1014	
I252.083R12.11_HILICPos		252.0830154	727	211.05626, 100; 206.05098, 83; 165.05096, 79.6; 188.04031, 64.3; 142.0349, 61.9; 183.06152, 55.3; 241.98708, 42.4; 229.06705, 38.5; 160.04568, 29.4; 251.1537, 28.2; 104.10741, 25.7; 223.97691, 23; 252.20717, 21.7; 141.04701, 21.7; 252.1441, 19.1; 170.02975, 18.7; 265.00345, 18.5; 235.18065, 17.7; 252.93094, 15.3; 134.03, 9.5; 242.99065, 9.2; 234.07262, 8.8; 206.07753, 8.1; 223.79031, 7.4; 124.02462, 7.3; 139.07172, 6.7; 224.98085, 6.5; 252.08237, 6.4; 250.98756, 6.3; 212.05843, 5.8; 251.1402, 5.5; 223.80229, 5.5; 251.24809, 5.1; 264.85352, 4.9; 136.06197, 4.7; 252.15735, 4.5; 216.50221, 4.4; 97.59184, 4.1; 225.95148, 4.1; 75.37668, 4.1; 56.35294, 4; 74.23139, 4; 115.17371, 3.9; 89.9743, 3.9; 177.17516, 3.8; 81.8637, 3.8; 81.62167, 3.8; 105.17507, 3.5
I271.1288R9.7_HILICPos	Bacilysin (3)	271.1288395	582	271.12869, 100; 225.12325, 22.7; 238.09469, 9.1; 256.10535, 8.5; 225.087, 6.7; 272.13243, 4.6; 212.09483, 3.6; 235.10765, 3.2; 253.11859, 3; 219.65848, 2.4; 170.08118, 2.3; 210.09988, 2; 226.05959, 1.9; 177.10226, 1.7; 211.1076, 1.6; 231.06459, 1.6; 203.0814, 1.6; 124.07586, 1.2; 221.17937, 1.1; 217.0968, 1; 270.27869, 0.9; 208.04826, 0.9; 183.07625, 0.6; 71.08623, 0.6; 152.07019, 0.6; 223.10741, 0.6
I294.975R14.44_HILICPos		294.9750456	866	NA
I136.0508R7.09_HILICPos		136.0507939	425	95.0609, 100; 136.0605, 90.2; 136.08704, 55.8; 136.05071, 37.3; 90.0457, 28.1; 137.04578, 26.5; 136.11207, 21; 136.07575, 18.2; 64.03026, 16.7; 135.09673, 15.4; 137.10742, 13; 109.07638, 7.8; 137.08246, 6; 137.07103, 5.6; 85.02903, 4; 135.09155, 3.3; 136.02154, 2.7; 135.06663, 2.4; 136.10303, 2.4; 110.07183, 2.3; 134.03624, 1.3; 95.55695, 1.3; 78.20219, 1.2; 120.09462, 1.2; 108.05609, 1.1; 69.17332, 1.1; 98.77731, 1.1; 57.96737, 1.1; 65.48399, 1.1; 115.39027, 1.1; 104.68243, 1.1; 96.13373, 1.1; 64.26604, 1.1; 117.35629, 1.1; 98.36456, 1.1; 103.62711, 1.1; 92.29008, 1; 76.42995, 1; 55.78654, 1; 108.55973, 1; 101.46804, 1; 50.62292, 1; 65.11148, 1; 134.95015, 1; 79.94245, 1; 98.20818, 1; 85.62806, 1; 53.17469, 1; 143.88516, 1; 60.33573, 1; 58.50202, 1; 133.40071, 1; 96.13895, 1; 97.76218, 1; 58.89027, 1; 67.6197, 0.9; 60.14315, 0.9; 65.31797, 0.9; 78.61413, 0.9
I152.9831R10.73_HILICPos		152.9831085	644	NA
I322.8476R5.22_HILICPos		322.8475757	313	276.81552, 100; 299.83124, 41.2; 322.84738, 6.1; 322.19299, 3.6; 187.88322, 3.2; 322.28619, 3.2; 223.78508, 3.1; 212.0956, 3.1; 294.19666, 2.4; 258.80527, 2.3; 202.76772, 2.3; 348.30933, 2.2; 281.8205, 2; 286.81158, 2; 223.77122, 2; 256.35901, 2; 226.43054, 1.7; 280.05182, 1.7; 258.16913, 1.7; 212.53238, 1.7; 122.32446, 1.7; 142.10965, 1.6; 101.33846, 1.6; 199.3942, 1.6; 65.93016, 1.6; 304.13373, 1.6; 263.18723, 1.6; 171.53821, 1.6; 309.25775, 1.6; 112.65562, 1.5; 116.96413, 1.5; 73.90466, 1.5
I137.0459R7.49_HILICPos	Hypoxanthine (3)	137.0458872	449	138.05308, 100; 137.04588, 5.7; 94.06564, 1.2; 110.06046, 0.8

Table 5.7 continued.

Table 5.7. *Continued from previous page.* Mass features with the largest peak areas in the aqueous fraction, listed in order of peak size. MF_Frac = mass feature and fraction (RP or HILIC), mz = m/z, rt = retention time (in seconds). The best identification match and confidence value [136] for that match are given. For compounds with non putative identifications (all confidence levels > 1, MS^2s are given if acquired.

MF_Frac	*m/z*	rt (sec)	log2(SD/SFZ) WV	log2(SD/SFZ) PC	*p* value WV	*p* value PC	Best Identification
I550.0958R2.2_HILICPos	550.0958	132	8.4063	8.8171	0.0519	0.0466	
I299.1678R4.24_CyanoAq	**299.1678**	**254**	**7.2518**	**7.2973**	**0.0413**	**0.0439**	
I533.0521R2.2_HILICNeg	**533.0521**	**132**	**6.9807**	**7.2718**	**0.0435**	**0.0327**	
I493.0815R2.22_HILICNeg	**493.0815**	**133**	**4.2120**	**4.5987**	**0.0455**	**0.0425**	
I456.275R7.08_CyanoAq	**456.2750**	**425**	**4.1317**	**4.2559**	**0.0228**	**0.0124**	
I495.0783R2.22_HILICNeg	**495.0783**	**133**	**4.0737**	**4.3966**	**0.0443**	**0.0439**	
I253.9168R1.77_HILICNeg	**253.9168**	**106**	**3.6096**	**4.0959**	**0.0459**	**0.0425**	
I188.0378R3.96_HILICPos	**188.0378**	**237**	**3.1136**	**3.3861**	**0.0005**	**0.0000**	
I225.0791R7.77_HILICPos	**225.0791**	**466**	**3.0047**	**3.5367**	**0.0000**	**0.0018**	
I191.0737R8.52_HILICPos	**191.0737**	**511**	**2.9832**	**3.4961**	**0.0459**	**0.0387**	**2-Oxo-7-methylthioheptanoic acid; (3)**
I192.9131R3.72_HILICNeg	**192.9131**	**223**	**2.8523**	**3.2088**	**0.0266**	**0.0330**	
I209.0843R7.18_HILICPos	209.0843	431	2.4106	2.6267	0.1073	0.0489	
I206.1389R6.76_HILICPos	**206.1389**	**406**	**2.3565**	**2.7670**	**0.0001**	**0.0009**	**Pantothenol (3)**
I209.0844R8.07_HILICPos	**209.0844**	**484**	**2.2453**	**2.6680**	**0.0003**	**0.0018**	
I188.0377R4.48_HILICPos	**188.0377**	**269**	**2.2166**	**2.5675**	**0.0005**	**0.0015**	
I640.0344R2.18_HILICPos	640.0344	131	2.1571	2.5399	0.0519	0.0385	
I642.0313R2.17_HILICPos	**642.0313**	**130**	**2.1453**	**2.5487**	**0.0459**	**0.0368**	
I225.0791R8.9_HILICPos	225.0791	534	2.1263	2.5092	0.0443	0.0743	
I623.0078R6.46_CyanoAq	623.0078	388	2.0969	2.5281	0.0579	0.0419	
I160.0969R7.56_HILICPos	**160.0969**	**454**	**2.0100**	**2.4104**	**0.0002**	**0.0098**	**Betonicine (2)**
I624.9882R2.17_HILICNeg	**624.9882**	**130**	**1.9813**	**2.3490**	**0.0497**	**0.0332**	
I188.0377R7.78_HILICPos	**188.0377**	**467**	**1.9217**	**2.1348**	**0.0124**	**0.0062**	
I620.9939R2.17_HILICNeg	**620.9939**	**130**	**1.8892**	**2.2760**	**0.0459**	**0.0262**	
I622.9911R2.17_HILICNeg	**622.9911**	**130**	**1.8841**	**2.2723**	**0.0491**	**0.0289**	
I506.3174R1.94_HILICPos	506.3174	116	1.3594	1.8059	0.1351	0.0382	
I263.1062R11.5_HILICPos	263.1062	690	1.2613	1.6551	0.0518	0.0210	
I201.1234R10.69_HILICPos	201.1234	642	1.2604	1.6042	0.0776	0.0220	
I170.1176R8.07_HILICPos	**170.1176**	**484**	**1.2241**	**1.6009**	**0.0003**	**0.0098**	
I251.099R11.1_HILICPos	251.0990	666	1.1949	1.5863	0.0564	0.0212	
I462.291R1.97_HILICPos	462.2910	118	1.1227	1.5374	0.1293	0.0490	
I210.0724R11.1_HILICPos	210.0724	666	0.9659	1.3797	0.0684	0.0210	
I285.6464R1_CyanoAq	285.6464	60	0.8430	1.3577	0.4021	0.0425	
I174.1126R6.89_HILICPos	174.1126	413	0.7701	1.1714	0.0435	0.1078	Swainsonine (3)
I160.0969R8.37_HILICPos	**160.0969**	**502**	**0.7449**	**1.1509**	**0.0210**	**0.0209**	**Betonicine (2)**
I157.9075R9.18_HILICNeg	157.9075	551	0.7228	1.1624	0.2199	0.0221	
I157.0973R7.33_HILICPos	157.0973	440	0.5294	0.9495	0.2728	0.0393	
I144.1019R6.44_HILICPos	144.1019	386	0.4594	0.8579	0.0210	0.0500	Proline betaine (1)
I364.8782R0.89_CyanoAq	364.8782	54	0.3269	0.8173	0.3912	0.0261	
I350.9153R0.89_CyanoAq	350.9153	54	0.2397	0.7006	0.4623	0.0332	
I189.005R0.89_CyanoAq	189.0050	54	0.2334	0.6862	0.2345	0.0115	
I254.0426R0.89_CyanoAq	254.0426	54	0.1711	0.6178	0.4608	0.0463	
I155.0815R8.72_HILICPos	155.0815	523	-0.9164	-0.5852	0.0401	0.4071	
I102.0549R12.14_HILICNeg	102.0549	729	-0.9548	-0.6634	0.0026	0.3541	
I84.045R12.12_HILICPos	84.0450	727	-1.0763	-0.7994	0.0078	0.3084	
I494.0888R0.93_CyanoAq	494.0888	56	-1.0902	-0.5674	0.0191	0.2216	
I156.1132R8.68_HILICPos	156.1132	521	-1.0920	-0.7244	0.0241	0.3758	
I148.0605R12.12_HILICPos	148.0605	727	-1.1766	-0.9088	0.0241	0.3158	Glutamic acid (1)
I155.0815R6.1_HILICPos	155.0815	366	-1.2238	-0.8730	0.0413	0.3391	
I162.0776R1.13_CyanoAq	162.0776	68	-1.3690	-1.0578	0.0385	0.1868	
I189.1598R15.41_HILICPos	189.1598	924	-1.3930	-0.9267	0.0401	0.0887	N6,N6,N6-Trimethyl-L-lysine (3)
I162.0763R12.28_HILICPos	162.0763	737	-1.5574	-1.2684	0.0028	0.1020	O-Acetyl-L-homoserine(3)
I189.1235R6.93_HILICPos	189.1235	416	-1.5664	-1.2890	0.0080	0.2191	N6-Acetyl-L-lysine (3)
I214.0388R5.25_HILICNeg	214.0388	315	-1.7263	-1.3314	0.0413	0.1836	
I195.0765R10.08_HILICPos	195.0765	605	-1.7363	-1.3799	0.0282	0.2789	
I143.0816R7.97_HILICPos	143.0816	478	-1.7696	-1.4276	0.0124	0.2430	Ectoine (2)
I316.0457R12.61_HILICNeg	316.0457	757	-1.9530	-1.7691	0.0337	0.1066	

Table 5.8: Continued on following page.

MF_Frac	*m/z*	rt (sec)	log2(SD/STZ) WV	log2(SD/STZ) PC	*p* value WV	*p* value PC	Best Identification
I249.1347R5.35_HILICPos	**249.1347**	**321**	**-1.9617**	**-1.4651**	**0.0210**	**0.0163**	**8-Cyclopentyl-1,3-dimethylxanthine (2)**
I416.1778R9.11_HILICPos	416.1778	546	-1.9881	-1.6823	0.0142	0.1285	
I414.1638R9.08_HILICNeg	414.1638	545	-2.0059	-1.6795	0.0282	0.1714	
I416.178R0.93_CyanoAq	416.1780	56	-2.0265	-1.7014	0.0413	0.1720	
I241.0034R5.34_HILICPos	241.0034	321	-2.1018	-1.6659	0.0004	0.0701	
I96.0449R5.35_HILICPos	96.0449	321	-2.1192	-1.7206	0.0016	0.0547	
I240.1092R0.94_CyanoAq	240.1092	56	-2.1922	-1.8257	0.0485	0.0978	Dihydrobiopterin (3)
I93.0578R5.36_HILICPos	**93.0578**	**321**	**-2.3851**	**-1.9036**	**0.0005**	**0.0063**	
I160.037R5.34_HILICPos	**160.0370**	**320**	**-2.4548**	**-1.9482**	**0.0419**	**0.0180**	
I116.0474R5.33_HILICPos	116.0474	320	-2.5227	-2.0178	0.0522	0.0276	
I145.0608R10.18_HILICPos	145.0608	611	-2.5693	-2.3407	0.0380	0.2055	
I138.0549R5.35_HILICPos	**138.0549**	**321**	**-2.6880**	**-2.2341**	**0.0001**	**0.0058**	**Homarine (1)**

Table 5.8. *Continued from previous page.* Mass features with significant differences between SD and STF (t-test, $p < 0.05$). MF_Frac = mass feature and fraction (RP or HILIC, aqueous or organic), mz = m/z, rt = retention time (in seconds), $\log_2$(fold change) when normalized to WV and PC, as well as p values when normalized to WV and PC. The best identification match and confidence value [136] for that match are given. If a compound is bolded, $p < 0.05$ with both normalizations, otherwise $p < 0.05$ under either PC or WV normalization.

MF_Frac	*m/z*	rt (sec)	log2(SD/TZ) WV	log2(SD/TZ) PC	*p* value WV	*p* value PC	Best Identification
I297.1885R5.9_CyanoAq	297.1885	354	10.3327	11.3514	0.0604	0.0196	
I533.0521R2.2_HILICNeg	**533.0521**	**132**	**9.7926**	**10.8985**	**0.0151**	**0.0043**	
I299.1678R4.24_CyanoAq	**299.1678**	**254**	**9.4872**	**10.6161**	**0.0138**	**0.0069**	
I456.275R7.08_CyanoAq	**456.2750**	**425**	**9.1358**	**10.2480**	**0.0054**	**0.0009**	
I533.0693R5.95_CyanoAq	**533.0693**	**357**	**8.9787**	**10.0913**	**0.0434**	**0.0173**	
I531.0723R5.95_CyanoAq	**531.0723**	**357**	**8.2134**	**9.3407**	**0.0461**	**0.0188**	
I467.2832R6.23_CyanoAq	467.2832	374	8.0055	9.2918	0.0824	0.0418	
I425.1404R2.22_HILICNeg	425.1404	133	6.8937	7.9530	0.0910	0.0435	
I275.9414R1.7_HILICNeg	**275.9414**	**102**	**6.8517**	**7.9950**	**0.0327**	**0.0142**	
I253.9168R1.77_HILICNeg	**253.9168**	**106**	**5.9966**	**7.1310**	**0.0132**	**0.0052**	
I642.0313R2.17_HILICPos	**642.0313**	**130**	**4.9935**	**6.1694**	**0.0050**	**0.0023**	
I640.0344R2.18_HILICPos	**640.0344**	**131**	**4.6182**	**5.7897**	**0.0064**	**0.0024**	
I624.9882R2.17_HILICNeg	**624.9882**	**130**	**4.6088**	**5.7759**	**0.0047**	**0.0017**	
I623.0078R6.46_CyanoAq	**623.0078**	**388**	**4.5833**	**5.7701**	**0.0070**	**0.0028**	
I625.0048R6.46_CyanoAq	**625.0048**	**388**	**4.5813**	**5.7668**	**0.0086**	**0.0036**	
I448.1663R12.31_HILICPos	**448.1663**	**739**	**4.1629**	**5.4008**	**0.0434**	**0.0146**	
I188.0377R4.48_HILICPos	**188.0377**	**269**	**4.0999**	**5.3054**	**0.0002**	**0.0001**	
I276.9109R1.79_HILICNeg	276.9109	107	3.9834	5.1109	0.0630	0.0111	
I277.9097R1.79_HILICNeg	277.9097	107	3.9468	5.0771	0.0727	0.0146	
I429.1257R12.32_HILICNeg	429.1257	739	3.8741	5.0854	0.0132	0.0049	
I206.1389R6.76_HILICPos	**206.1389**	**406**	**3.8439**	**5.0778**	**0.0001**	**0.0001**	**Pantothenol (3)**
I325.1761R2.97_CyanoAq	**325.1761**	**178**	**3.7540**	**4.9365**	**0.0028**	**0.0010**	
I622.9911R2.17_HILICNeg	**622.9911**	**130**	**3.7094**	**4.8687**	**0.0052**	**0.0014**	
I620.9939R2.17_HILICNeg	**620.9939**	**130**	**3.6471**	**4.8061**	**0.0049**	**0.0013**	
I479.7914R1.79_HILICNeg	**479.7914**	**108**	**3.4228**	**4.4779**	**0.0235**	**0.0034**	
I188.0378R3.96_HILICPos	**188.0378**	**237**	**3.3616**	**4.5134**	**0.0004**	**0.0000**	
I297.1885R2.25_HILICPos	297.1885	135	3.2968	4.3492	0.0708	0.0189	
I191.0737R8.52_HILICPos	**191.0737**	**511**	**3.0788**	**4.2507**	**0.0170**	**0.0050**	**2-Oxo-7-methylthioheptanoic acid (3)**
I225.0791R7.77_HILICPos	**225.0791**	**466**	**3.0335**	**4.2834**	**0.0000**	**0.0002**	
I207.0686R8.88_HILICPos	**207.0686**	**533**	**3.0252**	**4.2579**	**0.0485**	**0.0211**	
I209.0844R8.07_HILICPos	**209.0844**	**484**	**3.0047**	**4.2364**	**0.0001**	**0.0002**	
I168.0517R1.05_CyanoAq	168.0517	63	2.9503	4.1157	0.0874	0.0180	
I259.1291R9.17_HILICPos	**259.1291**	**550**	**2.9489**	**4.1697**	**0.0200**	**0.0058**	
I235.1001R1.21_CyanoAq	**235.1001**	**73**	**2.8405**	**4.0952**	**0.0031**	**0.0016**	
I162.1125R6.83_HILICPos	162.1125	410	2.7232	3.9090	0.1466	0.0477	L-carnitine (2)
I280.1579R6.68_CyanoAq	**280.1579**	**401**	**2.5794**	**3.8501**	**0.0132**	**0.0042**	
I160.0969R7.56_HILICPos	**160.0969**	**454**	**2.3869**	**3.6553**	**0.0001**	**0.0005**	**Betonicine (2)**
I209.0843R7.18_HILICPos	**209.0843**	**431**	**2.3516**	**3.4589**	**0.0473**	**0.0054**	
I263.1062R11.5_HILICPos	**263.1062**	**690**	**2.2314**	**3.4067**	**0.0026**	**0.0004**	**1,6-Dimethoxypyrene (3)**
I285.6464R1_CyanoAq	**285.6464**	**60**	**2.2271**	**3.3515**	**0.0060**	**0.0003**	
I594.37R1.9_HILICPos	594.3700	114	2.2005	3.5020	0.1095	0.0430	
I184.1698R6.47_CyanoAq	184.1698	388	2.1458	3.3365	0.1187	0.0339	
I225.0791R8.9_HILICPos	**225.0791**	**534**	**2.1146**	**3.3985**	**0.0167**	**0.0082**	
I491.224R11.34_HILICPos	491.2240	680	2.0908	3.3495	0.0644	0.0209	
I196.1698R7.18_CyanoAq	196.1698	431	2.0824	3.2635	0.1246	0.0344	
I196.1698R7.49_CyanoAq	196.1698	450	2.0151	3.2014	0.1343	0.0377	
I478.0533R11.15_HILICPos	478.0533	669	1.9747	3.2123	0.0668	0.0261	
I188.0377R7.78_HILICPos	**188.0377**	**467**	**1.9457**	**3.1031**	**0.0041**	**0.0003**	
I303.1551R9.17_HILICPos	**303.1551**	**550**	**1.8680**	**3.1202**	**0.0215**	**0.0061**	
I332.0829R11.15_HILICPos	**332.0829**	**669**	**1.7531**	**2.9806**	**0.0397**	**0.0142**	
I363.1444R8.12_HILICPos	**363.1444**	**487**	**1.7450**	**3.0377**	**0.0146**	**0.0072**	**Gibberellin A8-catabolite (3)**
I285.6462R12.04_HILICPos	285.6462	723	1.6802	2.7840	0.0518	0.0039	
I192.9131R3.72_HILICNeg	192.9131	223	1.6668	2.8475	0.0970	0.0053	
I267.0726R11.17_HILICNeg	**267.0726**	**670**	**1.6473**	**2.8798**	**0.0489**	**0.0108**	**Inosine (3)**
I179.0048R7.02_HILICPos	**179.0048**	**421**	**1.5738**	**2.8778**	**0.0066**	**0.0080**	**Arsenobetaine (2)**
I185.0922R8.02_HILICPos	**185.0922**	**481**	**1.5635**	**2.8173**	**0.0066**	**0.0010**	
I165.0581R8.33_HILICPos	**165.0581**	**500**	**1.5336**	**2.7564**	**0.0016**	**0.0002**	
I175.0789R1_CyanoAq	175.0789	60	1.5320	2.7211	0.0612	0.0073	
I201.1234R10.69_HILICPos	**201.1234**	**642**	**1.4974**	**2.6661**	**0.0198**	**0.0007**	
I325.0188R11.15_HILICNeg	325.0188	669	1.3516	2.5699	0.1398	0.0358	

Table 5.9: Continued on following pages.

			log2(SD/TZ)		*p* value		
MF_Frac	*m/z*	rt (sec)	WV	PC	WV	PC	Best Identification
I371.2655R2.22_HILICPos	371.2655	133	1.2660	2.6498	0.1091	0.0328	
I196.004R5.46_HILICPos	**196.0040**	**327**	**1.2019**	**2.4032**	**0.0171**	**0.0005**	
I150.0413R8.14_HILICNeg	**150.0413**	**488**	**1.1439**	**2.3969**	**0.0001**	**0.0002**	**Guanine (1)**
I170.1176R8.07_HILICPos	**170.1176**	**484**	**1.1275**	**2.3707**	**0.0003**	**0.0004**	
I157.0973R7.33_HILICPos	**157.0973**	**440**	**1.1084**	**2.3377**	**0.0004**	**0.0002**	
I359.1815R1.4_CyanoAq	359.1815	84	1.0687	2.2099	0.0930	0.0007	
I506.3174R1.94_HILICPos	506.3174	116	0.9994	2.2488	0.1227	0.0030	
I160.0969R8.37_HILICPos	**160.0969**	**502**	**0.9439**	**2.1847**	**0.0001**	**0.0002**	**Betonicine (2)**
I251.099R11.1_HILICPos	251.0990	666	0.9076	2.1095	0.0587	0.0010	
I209.0478R11.07_HILICPos	209.0478	664	0.8937	2.1067	0.1223	0.0069	2-(2'-Methylthio)ethylmalic acid (3)
I448.2754R2.02_HILICPos	448.2754	121	0.8306	2.0879	0.1934	0.0318	Sisomicin (3)
I130.05R11.1_HILICPos	130.0500	666	0.7746	2.0154	0.0928	0.0056	4-Oxoproline (3)
I462.291R1.97_HILICPos	462.2910	118	0.7562	2.0235	0.1527	0.0034	
I126.904R4.8_HILICNeg	126.9040	288	0.6885	1.8159	0.1274	0.0011	
I414.1624R0.94_CyanoAq	414.1624	56	0.6786	2.0707	0.2587	0.0419	
I321.148R1.96_HILICPos	321.1480	118	0.6656	1.7916	0.3110	0.0362	
I148.0968R8.21_HILICPos	148.0968	492	0.6627	1.8953	0.0626	0.0010	N-Hydroxy-L-isoleucine (3)
I418.2647R2.02_HILICPos	418.2647	121	0.6587	1.9397	0.1661	0.0054	
I210.0724R11.1_HILICPos	210.0724	666	0.6422	1.8551	0.0975	0.0008	
I374.2385R2.07_HILICPos	374.2385	124	0.6217	1.9289	0.2100	0.0153	
I193.0894R10.12_HILICPos	193.0894	607	0.6048	1.7715	0.1806	0.0006	
I170.5703R16.89_HILICPos	170.5703	1013	0.6038	1.8074	0.1174	0.0003	
I131.0704R10.12_HILICPos	131.0704	607	0.5819	1.7469	0.1849	0.0005	Pantolactone (2)
I170.0684R16.91_HILICPos	170.0684	1014	0.5788	1.7816	0.1301	0.0004	
I157.9075R9.18_HILICNeg	157.9075	551	0.5439	1.7378	0.1359	0.0005	
I174.1126R6.89_HILICPos	174.1126	413	0.5368	1.8148	0.0651	0.0056	Swainsonine (3)
I203.0077R11.11_HILICNeg	203.0077	666	0.5183	1.6791	0.3107	0.0308	
I300.8705R4.81_HILICNeg	300.8705	288	0.4690	1.5909	0.3016	0.0079	
I228.0979R6.34_HILICPos	228.0979	380	0.4322	1.5267	0.3815	0.0186	Deoxycytidine (3)
I204.1596R0.89_CyanoAq	204.1596	53	0.4249	1.6464	0.2717	0.0033	
I459.3522R5.96_CyanoAq	459.3522	358	0.4208	1.7383	0.3455	0.0233	
I453.344R5.96_CyanoAq	453.3440	358	0.4191	1.7619	0.2764	0.0174	
I241.1216R5.2_HILICPos	241.1216	312	0.4143	1.5912	0.5162	0.0039	
I360.2019R2.15_HILICPos	360.2019	129	0.4019	1.6554	0.2819	0.0159	
I385.1097R5.45_HILICPos	385.1097	327	0.3976	1.6110	0.1847	0.0007	
I269.1595R2.2_HILICPos	269.1595	132	0.3863	1.7000	0.4079	0.0356	
I360.223R2.14_HILICPos	360.2230	129	0.3751	1.6275	0.3532	0.0292	
I622.3646R1.94_HILICPos	622.3646	116	0.3481	1.5779	0.2557	0.0017	
I204.1595R8.76_HILICPos	204.1595	525	0.3458	1.5674	0.2937	0.0008	
I198.0623R0.96_CyanoAq	198.0623	57	0.3316	1.6122	0.5233	0.0176	
I174.022R5.45_HILICPos	174.0220	327	0.3216	1.5289	0.1922	0.0001	Sulfanilic acid (2)
I309.1367R1.04_CyanoAq	309.1367	63	0.3158	1.4890	0.3459	0.0052	
I233.0788R1.05_CyanoAq	233.0788	63	0.3098	1.5370	0.3497	0.0079	
I242.1135R4.91_HILICPos	242.1135	295	0.2939	1.3885	0.5661	0.0216	
I217.1048R1.05_CyanoAq	217.1048	63	0.2863	1.5140	0.3923	0.0099	
I190.1439R0.83_CyanoAq	190.1439	50	0.2683	1.4614	0.4848	0.0024	
I439.1861R6.35_HILICPos	439.1861	381	0.2664	1.3835	0.5171	0.0108	
I241.1437R1.85_HILICPos	241.1437	111	0.2266	1.3616	0.6141	0.0259	
I173.0921R8.33_HILICPos	173.0921	500	0.1986	1.3976	0.5406	0.0079	
I144.1019R6.44_HILICPos	144.1019	386	0.1925	1.4399	0.1555	0.0007	Proline betaine (1)
I262.7388R8.46_HILICNeg	262.7388	508	0.1895	1.3799	0.5263	0.0062	
I275.1103R1.32_CyanoAq	275.1103	79	0.1851	1.4007	0.5941	0.0155	
I364.8782R0.89_CyanoAq	364.8782	54	0.1755	1.3983	0.2457	0.0003	
I264.7368R8.47_HILICNeg	264.7368	508	0.1677	1.3754	0.5584	0.0018	
I176.1281R10.33_HILICPos	176.1281	620	0.1626	1.3840	0.4622	0.0023	
I608.3859R5_CyanoAq	608.3859	300	0.1614	1.4010	0.6438	0.0215	
I652.4119R1.77_HILICPos	652.4119	106	0.1077	1.3616	0.8008	0.0437	
I159.0765R10.12_HILICPos	159.0765	607	0.1075	1.2982	0.6539	0.0013	5-Hydroxyectoine (3)
I241.1183R1.89_HILICPos	241.1183	113	0.0996	1.2644	0.8373	0.0483	
I242.1501R0.88_CyanoAq	242.1501	53	0.0987	1.2875	0.6483	0.0006	

Table 5.9 continued.

MF_Frac	*m/z*	rt (sec)	log2(SD/TZ) WV	log2(SD/TZ) PC	*p* value WV	*p* value PC	Best Identification
I270.1548R3.45_HILICPos	270.1548	207	0.0964	1.3296	0.7941	0.0167	
I268.0858R5.15_CyanoAq	268.0858	309	0.0891	1.3211	0.7537	0.0137	S-Ribosyl-L-homocysteine (3)
I289.1393R10.28_HILICPos	289.1393	617	0.0834	1.2674	0.8786	0.0393	
I193.0531R10.87_HILICPos	193.0531	652	0.0579	1.2437	0.8857	0.0103	Monodechloroaminopyrrolnitrin (3)
I221.0599R5.15_CyanoAq	221.0599	309	0.0536	1.2919	0.8554	0.0144	
I609.3888R1.79_HILICPos	609.3888	107	0.0494	1.2889	0.9143	0.0461	
I350.9153R0.89_CyanoAq	350.9153	54	0.0474	1.2641	0.7567	0.0002	
I189.005R0.89_CyanoAq	189.0050	54	0.0442	1.2729	0.4830	0.0001	
I225.9825R8.48_HILICPos	225.9825	509	0.0428	1.2578	0.9147	0.0304	
I229.9781R8.46_HILICPos	229.9781	508	0.0366	1.2496	0.9279	0.0264	
I244.0828R9.94_HILICNeg	244.0828	596	0.0329	1.2326	0.9367	0.0468	
I132.102R6.53_HILICPos	132.1020	392	0.0109	1.2746	0.9630	0.0047	Isoleucine (2)
I261.1311R1.29_CyanoAq	261.1311	77	0.0076	1.2360	0.9824	0.0209	
I564.3597R4.78_CyanoAq	564.3597	287	0.0066	1.2407	0.9868	0.0284	
I254.0426R0.89_CyanoAq	254.0426	54	0.0065	1.2365	0.9514	0.0002	
I253.1279R3.45_HILICPos	253.1279	207	-0.0019	1.2470	0.9961	0.0230	
I120.1022R4.22_HILICPos	120.1022	253	-0.0069	1.1320	0.9889	0.0369	
I608.3858R1.79_HILICPos	608.3858	107	-0.0110	1.2361	0.9769	0.0291	
I227.1754R2.2_HILICPos	227.1754	132	-0.0125	1.2376	0.9748	0.0334	1,8-Diazacyclotetradecane-2,9-dione (3)
I565.363R4.78_CyanoAq	565.3630	287	-0.0262	1.2279	0.9507	0.0475	
I348.7227R5.36_CyanoAq	348.7227	322	-0.0458	1.2139	0.8953	0.0326	
I252.1441R10.53_HILICPos	252.1441	632	-0.0582	1.1771	0.8489	0.0050	
I146.0601R2.63_HILICPos	146.0601	158	-0.0700	1.1902	0.8200	0.0269	
I244.0929R7.89_HILICPos	244.0929	473	-0.0861	1.0578	0.8571	0.0476	Cytidine gamma-Glutamyl-beta-cyanoalanine (3)
I238.0674R12.49_HILICPos	238.0674	749	-0.0991	1.1494	0.6656	0.0036	
I577.973R14.46_HILICPos	577.9730	868	-0.1032	1.1479	0.7301	0.0374	
I340.9469R0.89_CyanoAq	340.9469	53	-0.1247	1.1186	0.3921	0.0010	
I330.191R2.19_HILICPos	330.1910	131	-0.1434	1.1593	0.6309	0.0322	
I264.9572R3.23_HILICPos	264.9572	194	-0.1567	1.0354	0.6353	0.0356	
I565.3628R1.8_HILICPos	565.3628	108	-0.1647	1.0763	0.6109	0.0284	
I709.9941R14.45_HILICPos	709.9941	867	-0.1660	1.0927	0.5437	0.0271	
I285.9045R8.4_HILICPos	285.9045	504	-0.1767	1.0149	0.5876	0.0183	
I281.9087R8.41_HILICPos	281.9087	505	-0.1770	1.0131	0.5876	0.0190	
I130.1592R4.02_CyanoAq	130.1592	241	-0.1796	0.9302	0.6572	0.0308	
I494.0888R0.93_CyanoAq	494.0888	56	-0.1813	1.0777	0.7413	0.0489	
I163.5793R4.94_CyanoAq	163.5793	297	-0.1893	1.0076	0.4532	0.0008	
I436.7858R8.42_HILICPos	436.7858	505	-0.1912	0.9904	0.5534	0.0019	
I227.1755R2.17_CyanoAq	227.1755	130	-0.1965	1.0114	0.5925	0.0157	1,8-Diazacyclotetradecane-2,9-dione (3)
I707.9976R14.44_HILICPos	707.9976	867	-0.1990	1.0485	0.5117	0.0348	
I283.9066R8.4_HILICPos	283.9066	504	-0.2007	0.9882	0.5754	0.0370	
I564.3595R1.8_HILICPos	564.3595	108	-0.2055	1.0285	0.4919	0.0259	
I822.9914R14.45_HILICPos	822.9914	867	-0.2055	1.0493	0.5389	0.0495	
I197.9728R10.62_HILICNeg	197.9728	637	-0.2169	1.0164	0.1815	0.0032	
I403.1349R13.93_HILICPos	403.1349	836	-0.2177	0.9283	0.5370	0.0205	
I575.9771R14.45_HILICPos	575.9771	867	-0.2207	1.0028	0.2925	0.0064	
I215.7812R8.86_HILICNeg	215.7812	531	-0.2300	1.0061	0.0582	0.0003	
I139.0157R7.24_HILICNeg	139.0157	435	-0.2320	1.0355	0.6039	0.0430	(S)-2-Hydroxypropylphosphonate (3)
I211.786R8.85_HILICNeg	211.7860	531	-0.2377	0.9985	0.0581	0.0003	
I425.0251R8.37_HILICPos	425.0251	502	-0.2401	0.9814	0.4633	0.0256	
I80.9157R8.39_HILICNeg	80.9157	504	-0.2472	0.9682	0.1192	0.0001	
I322.1877R1.19_CyanoAq	322.1877	71	-0.2698	1.0115	0.3600	0.0477	
I305.1573R1.6_CyanoAq	305.1573	96	-0.2732	0.9489	0.3791	0.0375	
I257.7338R8.79_HILICNeg	257.7338	527	-0.2766	0.9427	0.0507	0.0002	
I152.0568R1.08_CyanoAq	152.0568	65	-0.2844	0.8756	0.3485	0.0046	
I310.14R1.31_CyanoAq	310.1400	78	-0.2903	0.9804	0.4709	0.0257	
I504.3384R4.14_CyanoAq	504.3384	248	-0.3013	0.9159	0.3551	0.0497	
I341.8308R8.4_HILICPos	341.8308	504	-0.3017	0.8850	0.3624	0.0138	
I294.145R1.19_CyanoAq	294.1450	72	-0.3045	0.9466	0.5311	0.0355	
I461.3155R3.62_CyanoAq	461.3155	217	-0.3228	0.8869	0.3095	0.0461	
I417.2891R2.77_CyanoAq	417.2891	166	-0.3335	0.8876	0.2885	0.0437	

Table 5.9 continued.

			log2(SD/TZ)		*p* value		
MF_Frac	*m/z*	rt (sec)	WV	PC	WV	PC	Best Identification
I541.3954R7.98_CyanoAq	541.3954	479	-0.3369	0.9004	0.1806	0.0196	
I428.2283R2.22_HILICPos	428.2283	133	-0.3372	0.9358	0.3032	0.0462	
I314.1518R5.6_CyanoAq	314.1518	336	-0.3437	0.8662	0.2289	0.0186	
I259.7317R8.79_HILICNeg	**259.7317**	**527**	**-0.3449**	**0.8748**	**0.0105**	**0.0002**	
I399.7549R8.4_HILICPos	399.7549	504	-0.3485	0.8365	0.3507	0.0229	
I180.1019R8.38_HILICPos	180.1019	503	-0.3514	0.8897	0.1937	0.0107	Phenacetin (2)
I440.7816R8.4_HILICPos	440.7816	504	-0.3566	0.8502	0.2921	0.0099	
I521.3368R1.81_HILICPos	521.3368	109	-0.3601	0.8586	0.2502	0.0330	
I520.3334R1.81_HILICPos	520.3334	109	-0.3628	0.8571	0.2453	0.0278	
I296.1413R5.6_CyanoAq	296.1413	336	-0.3701	0.8437	0.1488	0.0108	GYKI (3)
I371.2277R1.89_HILICPos	371.2277	113	-0.3715	0.8296	0.2055	0.0236	
I216.9718R10.92_HILICPos	216.9718	655	-0.3836	0.8189	0.2331	0.0118	
I397.757R8.4_HILICPos	397.7570	504	-0.3857	0.8094	0.3120	0.0304	
I143.0816R8.4_HILICPos	143.0816	504	-0.4029	0.8430	0.3444	0.0425	Ectoine(2)
I321.1314R1.98_HILICPos	321.1314	119	-0.4094	0.7748	0.1944	0.0327	
I477.3104R1.83_HILICPos	477.3104	110	-0.4321	0.7735	0.1957	0.0373	
I390.26R1.89_HILICPos	390.2600	113	-0.4331	0.7610	0.1431	0.0183	
I415.2541R1.86_HILICPos	415.2541	111	-0.4601	0.7378	0.1685	0.0310	TMAP (2)
I132.102R9.3_HILICPos	132.1020	558	-0.4872	0.7360	0.1534	0.0230	
I351.1435R7.18_HILICPos	351.1435	431	-0.5343	0.6620	0.1534	0.0199	
I160.0969R6.82_HILICPos	**160.0969**	**409**	**-0.5729**	**0.6615**	**0.0214**	**0.0302**	**5-Acetamidopentanoate Calystegin A3 (3)**
I116.9849R8.39_HILICPos	116.9849	503	-0.5902	0.6338	0.1196	0.0474	
I155.0815R8.72_HILICPos	155.0815	523	-0.6060	0.5877	0.0568	0.0243	
I236.1493R8.23_HILICPos	236.1493	494	-0.6253	0.5954	0.1453	0.0338	
I196.097R8.81_HILICPos	**196.0970**	**529**	**-0.6551**	**0.5741**	**0.0012**	**0.0153**	**2-Phenyl-1,3-propanediol monocarbamate (3)**
I174.1126R8.06_HILICPos	174.1126	483	-0.7401	0.5367	0.0321	0.2148	Swainsonine Swainsonine Swainsonine (3)
I118.0864R7.19_HILICPos	118.0864	431	-0.7832	0.4463	0.0060	0.0533	GBT (1)
I211.0565R12.12_HILICPos	211.0565	727	-0.9240	0.2558	0.0492	0.3163	
I102.0549R12.14_HILICNeg	102.0549	729	-0.9298	0.2589	0.0045	0.1677	
I229.0671R12.12_HILICPos	229.0671	727	-0.9352	0.2315	0.0353	0.3345	
I564.426R7.17_HILICPos	564.4260	430	-0.9574	0.2927	0.0006	0.2136	
I104.1074R10.68_HILICPos	104.1074	641	-0.9660	0.1956	0.0423	0.4288	
I105.0906R10.66_HILICPos	105.0906	640	-0.9669	0.1953	0.0400	0.4213	
I121.032R8.27_HILICPos	121.0320	496	-1.0072	0.1993	0.0245	0.4257	DMS-Ac (3)
I479.1923R4.35_HILICPos	479.1923	261	-1.0144	0.0972	0.0484	0.7503	
I84.045R12.12_HILICPos	84.0450	727	-1.0221	0.1650	0.0055	0.3618	
I329.1742R7.17_HILICPos	329.1742	430	-1.0269	0.2244	0.0067	0.4313	
I188.0707R7.54_HILICPos	188.0707	452	-1.0441	0.1510	0.0344	0.6280	Deethylatrazine (3)
I189.1235R6.93_HILICPos	189.1235	416	-1.0645	0.1169	0.0050	0.5323	N6-Acetyl-L-lysine N6-Acetyl-L-lysine (3)
I148.0605R12.12_HILICPos	148.0605	727	-1.1014	0.0903	0.0196	0.6982	Glutamic acid (1)
I282.1184R11.26_HILICPos	282.1184	676	-1.1055	0.0577	0.0037	0.7945	
I204.1232R7.22_HILICPos	204.1232	433	-1.1158	0.1200	0.0155	0.6302	Acetyl-L-carnitine (1)
I288.1918R7.18_HILICPos	288.1918	431	-1.1378	0.1025	0.0007	0.6666	
I139.0868R7.52_HILICPos	139.0868	451	-1.1550	0.1091	0.0143	0.7244	
I326.0563R6.58_HILICPos	326.0563	395	-1.1733	0.1187	0.0024	0.7400	
I177.033R4.41_HILICPos	177.0330	265	-1.1777	0.0127	0.0125	0.9695	
I345.0462R0.88_CyanoAq	345.0462	53	-1.1824	0.0321	0.0234	0.9152	
I158.1176R5.48_HILICPos	158.1176	329	-1.1936	0.0485	0.0060	0.8391	
I138.055R7.72_HILICPos	138.0550	463	-1.2055	0.0331	0.0008	0.8913	Trigonelline (1)
I376.9346R7.43_HILICNeg	376.9346	446	-1.2426	-0.0086	0.0231	0.9779	
I136.0395R7.72_HILICNeg	136.0395	463	-1.2443	-0.0046	0.0012	0.9866	Anthranilate (3)
I130.0864R6.66_HILICNeg	130.0864	400	-1.2650	-0.0268	0.0125	0.9436	L-Leucine (3)
I287.2049R2.22_HILICPos	287.2049	133	-1.2769	-0.0744	0.0196	0.8398	
I284.0991R8.89_HILICPos	284.0991	534	-1.2817	-0.2197	0.0435	0.5709	Guanosine (3)
I298.957R7.41_HILICNeg	298.9570	445	-1.3066	-0.1042	0.0456	0.7063	
I138.1024R7.58_HILICPos	138.1024	455	-1.3292	-0.1679	0.0337	0.6112	
I293.2113R9.66_CyanoAq	293.2113	579	-1.3417	-0.1160	0.0400	0.6790	Etherolenic acid (3)
I116.0709R8.8_HILICPos	116.0709	528	-1.3541	-0.1318	0.0053	0.6888	Proline (1)
I431.1003R4.6_HILICPos	431.1003	276	-1.3656	-0.2216	0.0233	0.5042	
I293.2112R9.77_CyanoAq	293.2112	586	-1.4058	-0.1872	0.0355	0.5052	Etherolenic acid (3)

Table 5.9 continued.

			log2(SD/TZ)		*p* value		
MF_Frac	*m/z*	rt (sec)	WV	PC	WV	PC	Best Identification
I245.0434R10.56_HILICNeg	245.0434	634	-1.4074	-0.2196	0.0147	0.4879	
I519.9359R6.55_HILICPos	519.9359	393	-1.4440	-0.1269	0.0167	0.7671	
I207.1704R6.56_HILICPos	207.1704	394	-1.4649	-0.0538	0.0461	0.9475	
I175.119R17.83_HILICPos	175.1190	1070	-1.4884	-0.1777	0.0142	0.6808	Arginine (1)
I236.1492R9.91_HILICPos	236.1492	594	-1.4913	-0.2710	0.0012	0.2353	
I229.1549R0.89_CyanoAq	229.1549	53	-1.5166	-0.3414	0.0086	0.2487	
I162.0776R1.13_CyanoAq	162.0776	68	-1.5614	-0.4298	0.0353	0.2760	
I606.0756R12.98_HILICNeg	606.0756	779	-1.5736	-0.3438	0.0004	0.1318	UDP-glucosamine (1)
I189.1598R15.41_HILICPos	189.1598	924	-1.5850	-0.3433	0.0004	0.1712	N6,N6,N6-Trimethyl-L-lysine (3)
I247.0578R10.61_HILICPos	247.0578	637	-1.5944	-0.4063	0.0143	0.2424	
I219.1703R6.51_HILICPos	219.1703	390	-1.6390	-0.2889	0.0086	0.6363	
I193.1549R6.6_HILICPos	193.1549	396	-1.6555	-0.3887	0.0049	0.3259	
I143.0816R7.97_HILICPos	143.0816	478	-1.6573	-0.4252	0.0028	0.2568	Ectoine ID:KO002834 (2)
I366.0514R8.81_HILICPos	366.0514	528	-1.6725	-0.4129	0.0016	0.3247	
I608.0889R12.98_HILICPos	608.0889	779	-1.6844	-0.4386	0.0010	0.1668	UDP-N-acetyl-alpha-D-glucosamine (3)
I162.0763R12.28_HILICPos	162.0763	737	-1.6897	-0.5384	0.0022	0.1120	O-Acetyl-L-homoserine (3)
I238.1149R17.83_HILICPos	238.1149	1070	-1.6912	-0.4069	0.0338	0.3640	
I343.7137R1.49_CyanoAq	343.7137	90	-1.7002	-0.5363	0.0423	0.3293	
I134.0813R10.18_HILICPos	134.0813	611	-1.7284	-0.4700	0.0059	0.1508	N-Hydroxy-L-valine (3)
I327.1585R8.82_HILICPos	327.1585	529	-1.9181	-0.5988	0.0245	0.2800	
I377.1457R4.03_HILICPos	377.1457	242	-1.9577	-0.9708	0.0473	0.1270	Riboflavin (3)
I158.0813R7.81_HILICPos	**158.0813**	**468**	**-1.9662**	**-0.7667**	**0.0004**	**0.0160**	**2-Hydroxyethylclavam (3)**
I447.2461R4.66_CyanoAq	447.2461	280	-1.9868	-0.8963	0.0448	0.1591	
I311.2218R1.7_HILICPos	311.2218	102	-2.0022	-0.7684	0.0164	0.0927	13(S)-HPOT; 9(S)-HPOT; 2(R)-HPOT (3)
I179.024R6.59_HILICPos	179.0240	396	-2.0446	-0.7153	0.0005	0.1387	
I206.0351R6.59_HILICPos	206.0351	395	-2.0932	-0.7567	0.0017	0.1591	
I316.0457R12.61_HILICNeg	**316.0457**	**757**	**-2.1317**	**-1.0811**	**0.0030**	**0.0491**	
I160.037R5.34_HILICPos	160.0370	320	-2.2008	-0.9868	0.0100	0.0618	
I208.0333R6.59_HILICPos	208.0333	396	-2.2252	-0.8875	0.0040	0.1339	
I382.0653R12.08_HILICPos	382.0653	725	-2.2442	-1.1554	0.0400	0.0691	
I317.0553R12.1_HILICNeg	317.0553	726	-2.2588	-1.1060	0.0324	0.0560	Prekinamycin (3)
I219.9271R10.87_HILICNeg	219.9271	652	-2.2865	-1.0451	0.0233	0.0791	
I217.93R10.87_HILICNeg	217.9300	652	-2.3433	-1.1348	0.0225	0.0656	
I249.1347R5.35_HILICPos	**249.1347**	**321**	**-2.4129**	**-1.1832**	**0.0034**	**0.0170**	**8-Cyclopentyl-1,3-dimethylxanthine (2)**
I358.2087R3.38_CyanoAq	**358.2087**	**203**	**-2.4677**	**-1.2250**	**0.0001**	**0.0147**	
I241.0034R5.34_HILICPos	**241.0034**	**321**	**-2.4829**	**-1.1784**	**0.0019**	**0.0120**	
I116.0474R5.33_HILICPos	**116.0474**	**320**	**-2.4850**	**-1.2705**	**0.0018**	**0.0151**	
I96.0449R5.35_HILICPos	**96.0449**	**321**	**-2.5064**	**-1.2435**	**0.0001**	**0.0011**	
I241.0119R13.69_HILICNeg	**241.0119**	**821**	**-2.5408**	**-1.3121**	**0.0043**	**0.0088**	**6-Deoxy-5-ketofructose 1-phosphate (3)**
I423.0919R12.08_HILICPos	**423.0919**	**725**	**-2.5775**	**-1.5751**	**0.0417**	**0.0496**	**Lactose 6'-phosphate alpha (3)**
I235.1653R7.18_HILICPos	**235.1653**	**431**	**-2.5986**	**-1.3457**	**0.0220**	**0.0375**	
I155.0815R6.1_HILICPos	**155.0815**	**366**	**-2.6405**	**-1.3909**	**0.0001**	**0.0005**	
I292.0727R10.9_HILICNeg	**292.0727**	**654**	**-2.6542**	**-1.3153**	**0.0123**	**0.0260**	
I216.0633R11.68_HILICPos	**216.0633**	**701**	**-2.7362**	**-1.6195**	**0.0325**	**0.0332**	**sn-Glycero-3-phosphoethanolamine (3)**
I289.18R9.85_CyanoAq	**289.1800**	**591**	**-2.7488**	**-1.6158**	**0.0211**	**0.0316**	**6beta-Hydroxy-17beta-estradiol (3)**
I214.0484R11.68_HILICNeg	**214.0484**	**701**	**-2.7692**	**-1.6496**	**0.0308**	**0.0310**	**sn-Glycero-3-phosphoethanolamine (3)**
I93.0578R5.36_HILICPos	**93.0578**	**321**	**-2.7702**	**-1.5178**	**0.0001**	**0.0003**	
I219.0616R5.35_HILICPos	**219.0616**	**321**	**-2.8754**	**-1.7007**	**0.0109**	**0.0145**	
I280.674R0.89_CyanoAq	**280.6740**	**54**	**-2.9464**	**-1.6191**	**0.0090**	**0.0291**	
I138.0549R5.35_HILICPos	**138.0549**	**321**	**-3.0014**	**-1.7435**	**0.0001**	**0.0002**	**Homarine (1)**
I151.1441R6.61_HILICPos	151.1441	397	-3.1145	-1.7209	0.0308	0.0627	
I279.0592R11.7_HILICPos	279.0592	702	-3.1770	-2.1543	0.0603	0.0419	
I308.1605R5.35_HILICPos	**308.1605**	**321**	**-3.2208**	**-1.9323**	**0.0002**	**0.0002**	
I145.0608R10.18_HILICPos	**145.0608**	**611**	**-3.4044**	**-2.2250**	**0.0328**	**0.0227**	
I338.1392R5.52_CyanoAq	338.1392	331	-3.7066	-2.4539	0.0694	0.0473	
I316.0563R10.89_HILICPos	**316.0563**	**653**	**-3.9565**	**-2.5881**	**0.0060**	**0.0039**	
I488.2414R1.26_CyanoAq	488.2414	76	-4.1435	-3.0476	0.0652	0.0377	
I419.6799R6.47_CyanoAq	**419.6799**	**388**	**-4.2729**	**-2.9834**	**0.0223**	**0.0124**	
I375.2136R4.1_CyanoAq	375.2136	246	-4.3372	-2.7321	0.0669	0.0459	
I350.1462R5.34_HILICPos	**350.1462**	**321**	**-4.6710**	**-3.2974**	**0.0006**	**0.0004**	

Table 5.9 continued.

MF_Frac	*m/z*	rt (sec)	log2(SD/TZ) WV	log2(SD/TZ) PC	*p* value WV	*p* value PC	Best Identification
I394.6729R6.02_CyanoAq	394.6729	361	-4.9482	-3.6186	0.0795	0.0403	
I380.1632R6.07_CyanoAq	**380.1632**	**364**	**-5.0832**	**-3.8812**	**0.0379**	**0.0176**	
I176.1282R7.85_HILICPos	**176.1282**	**471**	**-5.1361**	**-3.8407**	**0.0485**	**0.0259**	
I329.207R1.71_HILICPos	329.2070	103	-5.1390	-3.9206	0.0896	0.0481	
I466.2302R2.16_CyanoAq	**466.2302**	**130**	**-5.3124**	**-3.8505**	**0.0126**	**0.0073**	
I297.0846R5.35_HILICPos	**297.0846**	**321**	**-5.3408**	**-4.0729**	**0.0047**	**0.0025**	
I377.6816R3.98_CyanoAq	**377.6816**	**239**	**-5.6079**	**-4.3485**	**0.0459**	**0.0281**	
I511.2407R3.6_CyanoAq	511.2407	216	-5.6134	-4.0342	0.0762	0.0368	
I331.2054R4.21_CyanoAq	331.2054	253	-5.6372	-4.3413	0.0928	0.0449	
I490.2589R5.48_CyanoAq	**490.2589**	**329**	**-5.7592**	**-4.3307**	**0.0400**	**0.0178**	
I424.2081R2.64_CyanoAq	**424.2081**	**159**	**-5.8159**	**-4.6321**	**0.0500**	**0.0213**	
I262.627R1.3_CyanoAq	**262.6270**	**78**	**-5.9627**	**-4.6066**	**0.0100**	**0.0053**	
I370.6952R2.68_CyanoAq	**370.6952**	**161**	**-6.0864**	**-4.7563**	**0.0376**	**0.0162**	
I558.7576R5.14_CyanoAq	558.7576	309	-6.0934	-5.1037	0.0983	0.0466	
I296.1837R1.4_CyanoAq	296.1837	84	-6.2525	-5.2188	0.0758	0.0339	
I523.2406R4.99_CyanoAq	**523.2406**	**299**	**-6.2860**	**-5.0249**	**0.0188**	**0.0068**	
I630.2563R6.94_CyanoAq	630.2563	417	-6.3229	-4.7333	0.0527	0.0303	
I577.1657R5.3_CyanoAq	**577.1657**	**318**	**-6.5425**	**-5.2392**	**0.0258**	**0.0098**	
I575.1679R5.3_CyanoAq	**575.1679**	**318**	**-6.6003**	**-5.4303**	**0.0242**	**0.0090**	
I589.782R5.43_CyanoAq	**589.7820**	**326**	**-6.6120**	**-5.2083**	**0.0311**	**0.0126**	
I306.4913R3.77_CyanoAq	306.4913	226	-6.6690	-5.5117	0.0935	0.0431	
I402.1687R5.48_CyanoAq	402.1687	329	-6.6747	-5.4710	0.0669	0.0295	
I380.4976R6.06_CyanoAq	380.4976	364	-6.7473	-5.8355	0.0539	0.0224	
I589.2805R5.44_CyanoAq	**589.2805**	**326**	**-6.8219**	**-5.3774**	**0.0146**	**0.0058**	
I502.7509R4.52_CyanoAq	502.7509	271	-6.9295	-5.5979	0.0527	0.0221	
I512.2441R3.6_CyanoAq	512.2441	216	-6.9707	-5.3273	0.0916	0.0425	
I575.2718R6.02_CyanoAq	**575.2718**	**361**	**-7.0987**	**-6.2029**	**0.0461**	**0.0181**	
I470.2256R5.55_CyanoAq	470.2256	333	-7.2470	-5.9195	0.0885	0.0405	
I651.7973R6.19_CyanoAq	**651.7973**	**372**	**-7.3658**	**-6.1554**	**0.0083**	**0.0037**	**Triiodothyronine (3)**
I459.2336R3.77_CyanoAq	459.2336	226	-7.4339	-6.3703	0.0744	0.0322	
I358.719R1.96_CyanoAq	**358.7190**	**118**	**-7.5086**	**-6.5464**	**0.0472**	**0.0208**	
I510.2201R1.96_CyanoAq	**510.2201**	**118**	**-7.7865**	**-6.9487**	**0.0338**	**0.0128**	
I578.1689R5.3_CyanoAq	**578.1689**	**318**	**-7.8361**	**-6.6772**	**0.0448**	**0.0175**	
I320.1713R2.37_CyanoAq	**320.1713**	**142**	**-7.8704**	**-6.4800**	**0.0071**	**0.0029**	
I548.2723R4.91_CyanoAq	**548.2723**	**294**	**-7.9415**	**-6.6714**	**0.0064**	**0.0029**	
I605.819R6.37_CyanoAq	605.8190	382	-8.0367	-6.6061	0.0673	0.0303	
I605.3175R6.37_CyanoAq	605.3175	382	-8.0643	-6.5365	0.0641	0.0288	
I549.2757R4.9_CyanoAq	**549.2757**	**294**	**-8.1446**	**-7.0810**	**0.0119**	**0.0050**	
I359.1525R4.27_CyanoAq	359.1525	256	-8.3361	-7.3904	0.0954	0.0427	
I587.2892R1.5_CyanoAq	**587.2892**	**90**	**-9.1064**	**-8.1579**	**0.0168**	**0.0066**	**Phycocyanobilin (3)**
I576.1711R5.3_CyanoAq	**576.1711**	**318**	**-9.4291**	**-8.0732**	**0.0413**	**0.0153**	
I293.613R4.91_CyanoAq	**293.6130**	**294**	**-9.8197**	**-8.9994**	**0.0109**	**0.0045**	
I329.2073R2.82_CyanoAq	329.2073	169	-11.0991	-9.6306	0.0626	0.0280	
I299.1967R2.47_CyanoAq	299.1967	148	-11.1528	-10.1830	0.0742	0.0349	

Table 5.9. *Continued from previous pages.* Mass features with significant differences between SD and TZ (t-test $p < 0.05$). Abbreviations are as in Table 5.8.

			log2(STF/TZ)		*p* value	
MF_Frac	*m/z*	rt (sec)	WV	PC	WV	PC
I494.0888R0.93_CyanoAq	494.0888	56	0.9089	1.6451	0.4642	0.0105
I204.1596R0.89_CyanoAq	204.1596	53	0.8151	1.6286	0.3803	0.0146
I326.0563R6.58_HILICPos	326.0563	395	-0.9674	-0.1224	0.0487	0.6523
I155.0815R6.1_HILICPos	155.0815	366	-1.4167	-0.5178	0.0303	0.3063
I179.024R6.59_HILICPos	179.0240	396	-1.7371	-0.8361	0.0316	0.2758
I206.0351R6.59_HILICPos	206.0351	395	-1.9346	-1.0592	0.0487	0.2755
I219.1703R6.51_HILICPos	219.1703	390	-1.9493	-1.0064	0.0487	0.2755

Table 5.10: Mass features with significant differences between STZ and TZ (t-test $p < 0.05$). Abbreviations are as in Table 5.8.

BIBLIOGRAPHY

[1] F. Azam, T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyer-Reil, and F. Thingstad, "The ecological role of water-column microbes in the sea," *Marine Ecology*, vol. 10, no. 3, pp. 257–263, 1983.

[2] M. J. Kainz, M. T. Arts, and A. Mazumder, "Essential fatty acids in the planktonic food web and their ecological role for higher trophic levels," *Limnology and Oceanography*, vol. 49, no. 5, pp. 1784–1793, 2004.

[3] S. Wright, S. Jeffrey, and R. Mantoura, *Phytoplankton pigments in oceanography: guidelines to modern methods*. Unesco Pub., 2005.

[4] A. E. Ingalls, S. R. Shah, R. L. Hansman, L. I. Aluwihare, G. M. Santos, E. R. M. Druffel, and A. Pearson, "Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon," *Proceedings of the National Academy of Sciences*, vol. 103, no. 17, pp. 6442–6447, 2006.

[5] K. Kaiser and R. Benner, "Biochemical composition and size distribution of organic matter at the Pacific and Atlantic time-series stations," *Marine Chemistry*, vol. 113, no. 1-2, pp. 63–77, 2009.

[6] S. Peulvé, J. W. De Leeuw, M. A. Sicre, M. Baas, and A. Saliot, "Characterization of macromolecular organic matter in sediment traps from the northwestern Mediterranean Sea," *Geochimica et Cosmochimica Acta*, vol. 60, no. 7, pp. 1239–1259, 1996.

[7] E. C. Minor, S. G. Wakeham, and C. Lee, "Changes in the molecular-level characteristics of sinking marine particles with water column depth," *Geochimica et Cosmochimica Acta*, vol. 67, no. 22, pp. 4277–4288, 2003.

[8] A. Boysen, K. Heal, L. Carlson, and A. Ingalls, "Best-matched internal standard normalization in liquid chromatography-mass spectrometry metabolomics applied to environmental samples.," *Anal. Chem.*, 2018.

[9] M. T. Croft, M. J. Warren, and A. G. Smith, "Algae need their vitamins," *Eukaryot. Cell*, vol. 5, no. 8, pp. 1175–1183, 2006.

[10] E. M. Bertrand, J. P. McCrow, A. Moustafa, H. Zheng, J. B. McQuaid, T. O. Delmont, A. F. Post, R. E. Sipler, J. L. Spackeen, and K. Xu, "Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge," *Proc.Nat. Acad. Sci.*, vol. 112, pp. 9938–9943, 2015.

[11] E. M. Bertrand, M. A. Saito, J. M. Rose, C. R. Riesselman, M. C. Lohan, A. E. Noble, P. A. Lee, and G. R. Di Tullio, "Vitamin B_{12} and iron colimitation of phytoplankton growth in the Ross Sea," *Limnol. Oceanogr.*, vol. 52, p. 1079, 2007.

[12] Y. Z. Tang, F. Koch, and C. J. Gobler, "Most harmful algal bloom species are vitamin B_1 and B_{12} auxotrophs," *Proc. Nat. Acad. Sci.*, vol. 107, pp. 20756–20761, 2010.

[13] M. R. Droop, "Vitamin B_{12} and marine ecology. IV. The kinetics of uptake, growth and inhibition in *Monochrysis lutheri*," *Journal of The Marine Biological Association of the United Kingdom*, vol. 48, pp. 689–733, 1968.

[14] C. Panzeca, A. Tovar-Sanchez, S. S. Agustí, I. Reche, C. M. Duarte, G. T. Taylor, and S. A. Sañudo-Wilhelmy, "B vitamins as regulators of phytoplankton dynamics," *EOS, Transactions American Geophysical Union*, vol. 87, no. 52, pp. 593–596, 2006.

[15] M. T. Croft, A. D. Lawrence, E. Raux-Deery, M. J. Warren, A. G. Smith, and A. D. L. Martin T. Croft Evelyn Raux-Deery, Martin J. Warren, Alison G. Smith, "Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria," *Nature*, vol. 438, pp. 90–93, 2005.

[16] E. M. Bertrand, D. M. Moran, M. R. McIlvin, J. M. Hoffman, A. E. Allen, and M. A. Saito, "Methionine synthase interreplacement in diatom cultures and communities: Implications for the persistence of B_{12} use by eukaryotic phytoplankton," *Limnol. Oceanogr.*, vol. 58, pp. 1431–1450, 2013.

[17] S. A. Sañudo-Wilhelmy, L. Gómez-Consarnau, C. Suffridge, and E. A. Webb, "The role of B vitamins in marine biogeochemistry," *Annual Review of Marine Science*, vol. 6, no. 1, pp. 339–367, 2014.

[18] G. Cooley, B. Ellis, V. Petrow, G. H. Beaven, E. R. Holiday, and E. A. Johnson, "The chemistry of anti-pernicious anaemia factors," *Journal of Pharmacy and Pharmacology*, vol. 3, pp. 271–285, 1951.

[19] R. Banerjee, R. Evande, Ö. Kabil, S. Ojha, and S. Taoka, "Reaction mechanism and regulation of cystathionine β-synthase," *Biochim. Biophys. Acta - Proteins Proteom.*, vol. 1647, no. 1-2, pp. 30–35, 2003.

[20] A. Juzeniene and Z. Nizauskaite, "Photodegradation of cobalamins in aqueous solutions and in human blood," *Journal of Photochemistry and Photobiology B: Biology*, vol. 122, pp. 7–14, 2013.

[21] A. F. Carlucci and S. B. Silbernagel, "Bioassay of seawater: 1. A ^{14}C uptake method for the determination of concentrations of vitamin B_{12} in seawater," *Canadian Journal of Microbiology*, vol. 12, pp. 175–183, 1966.

[22] K. C. Haines and R. R. L. Guillard, "Growth of Vitamin B12-Requiring marine diatoms in mixed laboratory culutres with vitamin B12-producing marine bacteria," *Journal of Phycology*, vol. 10, pp. 245–252, 1974.

[23] M. Okbamichael and S. A. Sañudo-Wilhelmy, "A new method for the determination of vitamin B_{12} in seawater," *Analytica Chimica Acta*, vol. 517, pp. 33–38, 2004.

[24] S. A. Sañudo-Wilhelmy, L. S. Cutter, R. Durazo, E. A. Smail, L. Gómez-Consarnau, E. A. Webb, M. G. Prokopenko, W. M. Berelson, and D. M. Karl, "Multiple B-vitamin depletion in large areas of the coastal ocean," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 14041–14045, 2012.

[25] A. Suárez-Suárez, A. Tova-Sanchez, R. R. Rosselló-Mora, A. Tovar-Sánchez, and R. R. Rosselló-Mora, "Determination of cobalamins (hydroxo-, cyano-, adenosyl- and methylcobalamins) in seawater using reversed-phase liquid chromatography with diode-array detection," *Analytica Chimica Acta*, vol. 701, no. 1, pp. 81–85, 2011.

[26] H. A. Schwertner, S. Valtier, and V. S. Bebarta, "Liquid chromatographic mass spectrometric (LC/MS/MS) determination of plasma hydroxocobalamin and cyanocobalamin concentrations after hydroxocobalamin antidote treatment for cyanide poisoning," *Journal of Chromatography B*, 2012.

[27] H. Chassaigne and R. Łobiński, "Direct species-selective determination of cobalamins by ionspray mass spectrometry and ionspray tandem mass spectrometry," *Analyst*, vol. 123, pp. 131–137, 1998.

[28] A. Szterk, M. Roszko, K. Małek, M. Czerwonka, and B. Waszkiewicz-Robak, "Application of the SPE reversed phase HPLC/MS technique to determine vitamin B_{12} bio-active forms in beef," *Meat Science*, vol. 91, pp. 408–413, 2012.

[29] P. Viñas, I. López-García, M. B. Bravo, and M. Hernández-Córdoba, "Multi-walled carbon nanotubes as solid-phase extraction adsorbents for the speciation of cobalamins in seafoods by liquid chromatography," *Analytical and Bioanalytical Chemistry*, vol. 401, pp. 1393–1399, 2011.

[30] K. W. Phinney, C. A. Rimmer, J. B. Thomas, L. C. Sander, K. E. Sharpless, and S. A. Wise, "Isotope dilution liquid chromatography-mass spectrometry methods for fat- and water-soluble vitamins in nutritional formulations," *Analytical Chemistry*, vol. 83, pp. 92–98, 2010.

[31] S. S. Kumar, R. S. Chouhan, and M. S. Thakur, "Trends in analysis of vitamin B_{12}," *Analytical Biochemistry*, vol. 398, pp. 139–149, 2010.

[32] Z. Chen, B. Chen, and S. Yao, "High-performance liquid chromatography/electrospray ionization-mass spectrometry for simultaneous determination of taurine and 10 water-soluble vitamins in multivitamin tablets," *Analytica Chimica Acta*, vol. 569, pp. 169–175, 2006.

[33] A. Leporati, D. Catellani, M. Suman, R. Andreoli, P. Manini, and W. Niessen, "Application of a liquid chromatography tandem mass spectrometry method to the analysis of water-soluble vitamins in Italian pasta," *Analytica Chimica Acta*, vol. 531, pp. 87–95, 2005.

[34] A. Gentili, F. Caretti, G. D'Ascenzo, S. Marchese, D. Perret, D. Di Corcia, and L. M. Rocca, "Simultaneous determination of water-soluble vitamins in selected food matrices by liquid chromatography/electrospray ionization tandem mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 22, pp. 2029–2043, 2008.

[35] D. Guggisberg, M. C. Risse, and R. Hadorn, "Determination of vitamin B_{12} in meat products by RP-HPLC after enrichment and purification on an immunoaffinity column," *Meat Science*, vol. 90, pp. 279–283, 2012.

[36] M. Okbamichael and S. A. Sañudo-Wilhelmy, "Direct determination of vitamin B_1 in seawater by solid-phase extraction and high-performance liquid chromatography quantification," *Limnololy and Oceanography: Methods*, vol. 3, pp. 241–246, 2005.

[37] B. W. Beadle, D. A. Greenwood, and H. R. Kraybill, "Stability of thiamine to heat I. Effect of pH and buffer salts in aqueous solutions," *Journal of Biological Chemistry*, vol. 149, pp. 339–347, 1943.

[38] S. Mohsin, M. Zumwalt, and I. Singh, "Quantitative analysis of water-soluble B-vitamins in cereal using rapid resolution LC/MS/MS application," *Food Analysis. Agilent Technologies Inc., USA*, pp. 1–12, 2008.

[39] M. Csuros, *Environmental Sampling and Analysis for Technicians*. Boca Raton: CRC press, 1994.

[40] G. F. M. Ball, *Vitamins in Foods: Analysis, Bioavailability, and Stability*. CRC Press, 2005.

[41] L. E. H. Gerards and S. Balt, "Acid induced decomposition of coenzyme B_{12} and variants," *Recueil des Travaux Chimiques des Pays-Bas*, vol. 111, pp. 411–412, 1992.

[42] P. Carini, E. O. Campbell, J. Morré, S. A. Sañudo-Wilhelmy, J. C. Thrash, S. E. Bennett, B. Temperton, T. Begley, and S. J. Giovannoni, "Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea," *The ISME Journal*, vol. 8, no. 8, pp. 1727–1738, 2014.

[43] M. J. Warren, E. Raux, H. L. Schubert, and J. C. Escalante-Semerena, "The biosynthesis of adenosylcobalamin (vitamin B_{12})," *Nat. Product Rep.*, vol. 19, pp. 390–412, 2002.

[44] M. R. Droop, "Vitamins, phytoplankton and bacteria: Symbiosis or scavenging?," *Journal of Plankton Research*, vol. 29, pp. 107–113, 2007.

[45] F. Azam, "Microbial control of oceanic carbon flux: The plot thickens," *Science*, vol. 280, pp. 694–696, 1998.

[46] E. M. Bertrand, A. E. Allen, C. L. Dupont, T. M. Norden-Krichmar, J. Bai, R. E. Valas, and M. A. Saito, "Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein," *Proc. Nat. Acad. Sci.*, vol. 109, no. 26, pp. E1762–E1771, 2012.

[47] B. P. Durham, S. Sharma, H. Luo, C. B. Smith, S. A. Amin, S. J. Bender, S. P. Dearth, B. A. S. Van Mooy, S. R. Campagna, E. B. Kujawinski, E. V. Armbrust, and M. A. Moran, "Cryptic carbon and sulfur cycling between surface ocean plankton," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 2, pp. 453–457, 2015.

[48] M. B. Cooper and A. G. Smith, "Exploring mutualistic interactions between microalgae and bacteria in the omics age," *Current Opinion in Plant Biology*, vol. 26, no. Figure 1, pp. 147–153, 2015.

[49] A. C. Doxey, D. A. Kurtz, M. D. J. Lynch, L. A. Sauder, and J. D. Neufeld, "Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production," *ISME J.*, vol. 9, pp. 461–471, 2014.

[50] K. E. Helliwell, A. D. Lawrence, A. Holzer, U. J. Kudahl, S. Sasso, B. Kräutler, D. J. Scanlan, M. J. Warren, and A. G. Smith, "Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B_{12}," *Curr. Biol.*, vol. 26, no. 8, pp. 999–1008, 2016.

[51] P. Renz, "Biosynthesis of the 5, 6-dimethylbenzimidazole moiety of cobalamin and of the other bases found in natural corrinoids," in *Chemistry and Biochemistry of B_{12}* (R. Banerjee, ed.), pp. 557–575, New York: John Wiley and Sons, 1999.

[52] Y. Tanioka, E. Miyamoto, Y. Yabuta, K. Ohnishi, T. Fujita, R. Yamaji, H. Misono, S. Shigeoka, Y. Nakano, and H. Inui, "Methyladeninylcobamide functions as the cofactor of methionine synthase in a Cyanobacterium, *Spirulina platensis* NIES-39," *FEBS Letters*, vol. 584, pp. 3223–3226, 2010.

[53] Y. Tanioka, Y. Yabuta, R. Yamaji, S. Shigeoka, Y. Nakano, F. Watanabe, and H. Inui, "Occurrence of pseudovitamin B_{12} and its possible function as the cofactor of cobalamin-dependent methionine synthase in a Cyanobacterium *Synechocystis sp.* PCC6803," *Journal of Nutritional Science and Vitaminology*, vol. 55, pp. 518–521, 2009.

[54] K. R. Heal, L. T. Carlson, A. H. Devol, E. Armbrust, J. W. Moffett, D. A. Stahl, and A. E. Ingalls, "Determination of four forms of vitamin B_{12} and other B vitamins in seawater by liquid chromatography/tandem mass spectrometry," *Rapid Commun. Mass Spectrom.*, vol. 28, no. 22, pp. 2398–2404, 2014.

[55] J. E. Swalwell, F. Ribalet, and E. Armbrust, "SeaFlow: A novel underway flow cytometer for continuous observations of phytoplankton in the ocean," *Limnology and Oceanography: Methods*, vol. 9, pp. 466–477, 2011.

[56] G. T. Taylor and C. W. Sullivan, "Vitamin B_{12} and cobalt cycling among diatoms and bacteria in Antarctic sea ice microbial communities," *Limnology and Oceanography*, vol. 53, p. 1862, 2008.

[57] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand, "Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes," *Journal of Biological Chemistry*, vol. 278, no. 42, pp. 41148–41159, 2003.

[58] M. E. Taga, N. A. Larsen, A. R. Howard-Jones, C. T. Walsh, and G. C. Walker, "BluB cannibalizes flavin to form the lower ligand of vitamin B_{12}," *Nature*, vol. 446, pp. 449–453, 2007.

[59] J. C. Escalante-Semerena, "Conversion of cobinamide into adenosylcobamide in bacteria and archaea," *Journal of bacteriology*, vol. 189, no. 13, pp. 4555–4560, 2007.

[60] F. Partensky and L. Garczarek, "Prochlorococcus: advantages and limits of minimalism," *Annual Review of Marine Science*, vol. 2, pp. 305–331, 2010.

[61] A. B. Hazra, A. W. Han, A. P. Mehta, K. C. Mok, V. Osadchiy, T. P. Begley, and M. E. Taga, "Anaerobic biosynthesis of the lower ligand of vitamin B_{12}," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 34, pp. 10792–10797, 2015.

[62] T. S. Crofts, E. C. Seth, A. B. Hazra, and M. E. Taga, "Cobamide structure depends on both lower ligand availability and CobT substrate specificity," *Chemistry & Biology*, vol. 20, no. 10, pp. 1265–1274, 2013.

[63] C.-G. G. Cheong, J. C. Escalante-Semerena, and I. Rayment, "Structural investigation of the biosynthesis of alternative lower ligands for cobamides by nicotinate mononucleotide: 5, 6-Dimethylbenzimidazole phosphoribosyltransferase from *Salmonella enterica*," *Journal of Biological Chemistry*, vol. 276, no. 40, pp. 37612–37620, 2001.

[64] C.-G. Cheong, J. C. Escalante-Semerena, and I. Rayment, "The three-dimensional structures of nicotinate mononucleotide: 5, 6-dimethylbenzimidazole phosphoribosyltransferase (CobT) from *Salmonella typhimurium* complexed with 5, 6-dimethybenzimidazole and its reaction products determined to 1.9 Å resolution," *Biochemistry*, vol. 38, pp. 16125–16135, 1999.

[65] P. J. Anderson, J. Lango, C. Carkeet, A. Britten, B. Kräutler, B. D. Hammock, and J. R. Roth, "One pathway can incorporate either adenine or dimethylbenzimidazole as an α-axial ligand of B_{12} cofactors in *Salmonella enterica*," *Journal of Bacteriology*, vol. 190, pp. 1160–1171, 2008.

[66] B. Keck and P. Renz, "*Salmonella typhimurium* forms adenylcobamide and 2-methyladenylcobamide, but no detectable cobalamin during strictly anaerobic growth," *Archives of Microbiology*, vol. 173, pp. 76–77, 2000.

[67] S. A. Benner, A. D. Ellington, and A. Tauer, "Modern metabolism as a palimpsest of the RNA world," *Proceedings of the National Academy of Sciences*, vol. 86, no. 18, pp. 7054–7058, 1989.

[68] A. Lazcano, "Planetary change and biochemical adaptation: Molecular evolution of corrinoid and heme biosyntheses," *Hematology*, vol. 17, pp. s7–s10, 2012.

[69] H. D. Holland, "The oxygenation of the atmosphere and oceans," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, pp. 903–915, 2006.

[70] M. L. Ludwig and P. R. Evans, "X-ray crystallography of B_{12} enzymes: Methylmalonyl-CoA mutase and methionine synthase," in *Chemistry and Biochemistry of B_{12}* (R. Banerjee, ed.), pp. 595–632, New York, NY: John Wiley and Sons, 1999.

[71] C. L. Drennan, R. G. Matthews, and M. L. Ludwig, "Cobalamin-dependent methionine synthase: The structure of a methylcobalamin-binding fragment and implications for other B_{12}-dependent enzymes," *Current Opinion in Structural Biology*, vol. 4, pp. 919–929, 1994.

[72] M. D. Sintchak, G. Arjara, B. A. Kellogg, J. Stubbe, and C. L. Drennan, "The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer," *Nature Structural Biology*, vol. 9, pp. 293–300, 2002.

[73] D. Lundin, S. Gribaldo, E. Torrents, B.-M. Sjöberg, and A. M. Poole, "Ribonucleotide reduction-horizontal transfer of a required function spans all three domains," *BMC Evolutionary Biology*, vol. 10, 2010.

[74] F. K. Gleason and N. E. Olszewski, "Isolation of the gene for the B_{12}-dependent ribonucleotide reductase from *Anabaena sp.* strain PCC 7120 and expression in *Escherichia coli*," *Journal of Bacteriology*, vol. 184, pp. 6544–6550, 2002.

[75] A. M. Poole, D. T. Logan, and B.-M. Sjöberg, "The evolution of the ribonucleotide reductases: Much ado about oxygen," *Journal of Molecular Evolution*, vol. 55, pp. 180–196, 2002.

[76] R. R. L. Guillard, "B_{12} specificity of marine centric diatoms," *Journal of Phycology*, vol. 4, pp. 59–64, 1968.

[77] L. Provasoli and A. F. Carlucci, "Vitamins and growth regulators," in *Algal physiology and biochemistry* (W. D. P. Stewart, H. G. Baker, H. Beevers, and F. R. Whatey, eds.), Botanical Monographs, pp. 741–787, Berkeley and Los Angeles: University of California Press, 1974.

[78] W. A. Ayers, "Specificity of the vitamin B_{12} requirement in certain marine bacteria," *J. Bacteriol.*, vol. 80, p. 744, 1960.

[79] E. C. Seth and M. E. Taga, "Nutrient cross-feeding in the microbial world," *Frontiers in Microbiology*, vol. 5, no. JULY, pp. 1–6, 2014.

[80] M. A. Saito, G. Rocap, and J. W. Moffett, "Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome," *Limnol. Oceanogr*, vol. 50, pp. 279–290, 2005.

[81] R. E. A. Horak, W. Qin, A. J. Schauer, E. V. Armbrust, A. E. Ingalls, J. W. Moffett, D. A. Stahl, and A. H. Devol, "Ammonia oxidation kinetics and temperature sensitivity of a natural marine community dominated by Archaea," *The ISME Journal*, vol. 7, pp. 2023–2033, 2013.

[82] H. Urakawa, W. Martens-Habbena, and D. A. Stahl, "High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method," *Applied and Environmental Microbiology*, vol. 76, pp. 2129–2135, 2010.

[83] A. E. Santoro, C. L. Dupont, R. A. Richter, M. T. Craig, P. Carini, M. R. McIlvin, Y. Yang, W. D. Orsi, D. M. Moran, and M. A. Saito, "Genomic and proteomic characterization of Candidatus Nitrosopelagicus brevis: An ammonia-oxidizing archaeon from the open ocean," *Proceedings of the National Academy of Sciences*, p. 201416223, 2015.

[84] H. Urakawa, W. Martens-Habbena, and D. A. Stahl, "Physiology and genomics of ammonia-oxidizing archaea," in *Nitrification* (B. B. Ward, M. G. Klotz, and D. J. Arp, eds.), pp. 117–156, Washington, DC: ASM Press, 2011.

[85] C. B. Walker, J. R. de la Torre, M. G. Klotz, H. Urakawa, N. Pinel, D. J. Arp, C. Brochier-Armanet, P. S. G. Chain, P. P. Chan, A. Gollabgir, J. Hemp, M. Hugler, E. A. Karr, M. Konneke, M. Shin, T. J. Lawton, T. Lowe, W. Martens-Habbena, L. A. Sayavedra-Soto, S. M. Sievert, A. C. Rosenzweig, G. Manning, D. A. Stahl, M. Hügler, E. A. Karr, M. Könneke, M. Shin, T. J. Lawton, T. Lowe, W. Martens-Habbena, L. A. Sayavedra-Soto, D. Lang, S. M. Sievert, A. C. Rosenzweig, G. Manning, and D. A. Stahl, "*Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 19, no. 19, p. 107, 2010.

[86] M. Könneke, D. M. Schubert, P. C. Brown, M. Hügler, S. Standfest, T. Schwander, L. S. von Borzyskowski, T. J. Erb, D. A. Stahl, and I. A. Berg, "Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO_2 fixation," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 8239–8244, 2014.

[87] W. Qin, S. A. Amin, W. Martens-Habbena, C. B. Walker, H. Urakawa, A. H. Devol, A. E. Ingalls, J. W. Moffett, E. V. Armbrust, and D. A. Stahl, "Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 12504–12509, 2014.

[88] J.-G. Kim, S.-J. Park, J. S. Sinninghe Damsté, S. Schouten, W. I. C. Rijpstra, M.-Y. Jung, S.-J. Kim, J.-H. Gwak, H. Hong, O.-J. Si, S. Lee, E. L. Madsen, and S.-K. Rhee, "Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 7888–7893, 2016.

[89] J. D. Rabinowitz and E. Kimball, "Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*," *Analytical Chemistry*, vol. 79, no. 16, pp. 6167–6173, 2007.

[90] S. A. Amin, L. R. Hmelo, H. M. van Tol, B. P. Durham, L. T. Carlson, K. R. Heal, R. L. Morales, C. T. Berthiaume, M. S. Parker, and B. Djunaedi, "Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria," *Nature*, vol. 522, pp. 98–101, 2015.

[91] S. Datta, M. Koutmos, K. A. Pattridge, M. L. Ludwig, and R. G. Matthews, "A disulfide-stabilized conformer of methionine synthase reveals an unexpected role for the histidine ligand of the cobalamin cofactor," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 4115–4120, 2008.

[92] M. Lunau, A. Lemke, K. Walther, W. Martens-Habbena, and M. Simon, "An improved method for counting bacteria from sediments and turbid environments by epifluorescence microscopy," *Environmental Microbiology*, vol. 7, pp. 961–968, 2005.

[93] M. Könneke, A. E. Bernhard, R. José, C. B. Walker, J. B. Waterbury, D. A. Stahl, J. R. de la Torre, C. B. Walker, J. B. Waterbury, and D. A. Stahl, "Isolation of an autotrophic ammonia-oxidizing marine archaeon," *Nature*, vol. 437, no. 7058, pp. 543–546, 2005.

[94] W. Qin, L. T. Carlson, E. V. Armbrust, A. H. Devol, J. W. Moffett, D. A. Stahl, and A. E. Ingalls, "Confounding effects of oxygen and temperature on the TEX86 signature of marine Thaumarchaeota," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, pp. 10979–10984, 2015.

[95] L. R. Moore, A. Coe, E. R. Zinser, M. a. Saito, M. B. Sullivan, D. Lindell, K. Frois-Moniz, J. Waterbury, S. W. Chisholm, K. FroisMoniz, J. Waterbury, and S. W. Chisholm, "Culturing the marine cyanobacterium *Prochlorococcus*," *Limnology and Oceanography: Methods*, vol. 5, no. 10, pp. 353–362, 2007.

[96] J. J. Morris, R. Kirkegaard, M. J. Szul, Z. I. Johnson, and E. R. Zinser, "Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by 'helper' heterotrophic bacteria," *Applied and Environmental Microbiology*, vol. 74, pp. 4530–4534, 2008.

[97] P. M. Berube, S. J. Biller, A. G. Kent, J. W. Berta-Thompson, S. E. Roggensack, K. H. Roache-Johnson, M. Ackerman, L. R. Moore, J. D. Meisel, and D. Sher, "Physiology and evolution of nitrate acquisition in *Prochlorococcus*," *The ISME Journal*, vol. 9, pp. 1195–1207, 2015.

[98] M. A. Saito, J. W. Moffett, S. W. Chisholm, and J. B. Waterbury, "Cobalt limitation and uptake in *Prochlorococcus*," *Limnol. Oceanogr.*, vol. 47, no. 6, pp. 1629–1636, 2002.

[99] J. Jeffrey Morris and E. R. Zinser, "Continuous hydrogen peroxide production by organic buffers in phytoplankton culture media," *Journal of Phycology*, vol. 49, pp. 1223–1228, 2013.

[100] A. W. Thompson, K. Huang, M. A. Saito, and S. W. Chisholm, "Transcriptome response of high-and low-light-adapted *Prochlorococcus* strains to changing iron availability," *The ISME Journal*, vol. 5, pp. 1580–1594, 2011.

[101] M. D. DuRand, R. J. Olson, and S. W. Chisholm, "Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 48, pp. 1983–2003, 2001.

[102] A. E. Zimmerman, S. D. Allison, and A. C. Martiny, "Phylogenetic constraints on elemental stoichiometry and resource allocation in heterotrophic marine bacteria," *Environmental microbiology*, vol. 16, pp. 1398–1410, 2014.

[103] J. P. Sachs and O. E. Kawka, "The influence of growth rate on $^{2}H/^{1}H$ fractionation in continuous cultures of the coccolithophorid *Emiliania huxleyi* and the diatom *Thalassiosira pseudonana*," *PLOS One*, vol. 10, 2015.

[104] D. J. S. Montagnes, J. A. Berges, P. J. Harrison, and F. Taylor, "Estimating carbon, nitrogen, protein, and chlorophyll a from volume in marine phytoplankton," *Limnology and Oceanography*, vol. 39, pp. 1044–1060, 1994.

[105] J. P. Beam, Z. J. Jay, M. C. Schmid, D. B. Rusch, M. F. Romine, R. de M Jennings, M. A. Kozubal, S. G. Tringe, M. Wagner, and W. P. Inskeep, "Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous streamer'community," *The ISME Fournal*, vol. 10, pp. 210–224, 2016.

[106] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, pp. 3389–3402, 1997.

[107] S. F. Altschul, J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schäffer, and Y. Yu, "Protein database searches using compositionally adjusted substitution matrices," *FEBS Journal*, vol. 272, pp. 5101–5109, 2005.

[108] P. Lengyel, R. Mazumder, and S. Ochoa, "Mammalian methylmalonyl isomerase and vitamin B_{12} coenzymes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 46, pp. 1312–1318, 1960.

[109] K. Katoh, K.-i. Kuma, H. Toh, and T. Miyata, "MAFFT version 5: Improvement in accuracy of multiple sequence alignment," *Nucleic Acids Research*, vol. 33, pp. 511–518, 2005.

[110] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: A web-based environment for protein structure homology modeling," *Bioinformatics*, vol. 22, pp. 195–201, 2006.

[111] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimeraa visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, pp. 1605–1612, 2004.

[112] M. C. Kido Soule, K. Longnecker, W. M. Johnson, and E. B. Kujawinski, "Environmental metabolomics: Analytical strategies," *Marine Chemistry*, vol. 177, pp. 374–387, 2015.

[113] F. Watanabe, H. Katsura, S. Takenaka, T. Fujita, K. Abe, Y. Tamura, T. Nakatsuka, and Y. Nakano, "Pseudovitamin B_{12} is the predominant cobamide of an algal health food, spirulina tablets," *Journal of Agricultural and Food Chemistry*, vol. 47, pp. 4736–4741, 1999.

[114] R. H. Allen and S. P. Stabler, "Identification and quantitation of cobalamin and cobalamin analogues in human feces," *The American Journal of Clinical Nutrition*, vol. 87, pp. 1324–1335, 2008.

[115] F. Ribalet, J. Swalwell, S. Clayton, V. Jiménez, S. Sudek, Y. Lin, Z. I. Johnson, A. Z. Worden, and E. V. Armbrust, "Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre," *Proceedings of the National Academy of Sciences*, vol. 112, pp. 8008–8012, 2015.

[116] W. Martens-Habbena, W. Qin, R. E. A. Horak, H. Urakawa, A. J. Schauer, J. W. Moffett, E. Armbrust, A. E. Ingalls, A. H. Devol, and D. A. Stahl, "The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger," *Environmental Microbiology*, vol. 17, pp. 2261–2274, 2015.

[117] T. J. Mincer, M. J. Church, L. T. Taylor, C. Preston, D. M. Karl, and E. F. DeLong, "Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre," *Environmental Microbiology*, vol. 9, pp. 1162–1175, 2007.

[118] J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, and J. I. Gordon, "QIIME allows analysis of high-throughput community sequencing data," *Nature Methods*, vol. 7, pp. 335–336, 2010.

[119] N. A. Bokulich, S. Subramanian, J. J. Faith, D. Gevers, J. I. Gordon, R. Knight, D. A. Mills, and J. G. Caporaso, "Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing," *Nature Methods*, vol. 10, pp. 57–59, 2013.

[120] R. C. Edgar, "Search and clustering orders of magnitude faster than BLAST," *Bioinformatics*, vol. 26, pp. 2460–2461, 2010.

[121] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, pp. 403–410, 1990.

[122] H. J. Tripp, J. B. Kitner, M. S. Schwalbach, J. W. H. Dacey, L. J. Wilhelm, and S. J. Giovannoni, "SAR11 marine bacteria require exogenous reduced sulphur for growth," *Nature*, vol. 452, pp. 741–744, 2008.

[123] C. L. Dupont, D. B. Rusch, S. Yooseph, M.-J. Lombardo, R. A. Richter, R. Valas, M. Novotny, J. Yee-Greenbaum, J. D. Selengut, D. H. Haft, A. L. Halpern, R. S. Lasken, K. Nealson, R. Friedman, and J. C. Venter, "Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage," *The ISME Journal*, vol. 6, no. 6, pp. 1186–1199, 2012.

[124] L. D. Crosby and C. S. Criddle, "Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity," *Biotechniques*, vol. 34, pp. 790–803, 2003.

[125] T. Větrovský and P. Baldrian, "The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses," *PloS one*, vol. 8, p. e57923, 2013.

[126] M. Stieglmeier, R. J. E. Alves, and C. Schleper, "The phylum Thaumarchaeota," in *The Prokaryotes*, pp. 347–362, Springer, 2014.

[127] V. Iverson, R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales, and E. V. Armbrust, "Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota," *Science*, vol. 335, pp. 587–590, 2012.

[128] M. B. Karner, E. F. DeLong, and D. M. Karl, "Archaeal dominance in the mesopelagic zone of the Pacific Ocean," *Nature*, vol. 409, pp. 507–510, 2001.

[129] J. D. Woodson, R. F. Peck, M. P. Krebs, and J. C. Escalante-Semerena, "The cobY gene of the archaeon *Halobacterium sp.* strain NRC-1 is required for de novo cobamide synthesis," *Journal of Bacteriology*, vol. 185, pp. 311–316, 2003.

[130] K. R. Heal, W. Qin, F. Ribalet, A. D. Bertagnolli, W. Coyote-Maestas, L. R. Hmelo, J. W. Moffett, A. H. Devol, E. V. Armbrust, D. A. Stahl, and A. E. Ingalls, "Two distinct pools of B_{12} analogs reveal community interdependencies in the ocean," *Proc. Nat. Acad. Sci.*, vol. 114, no. 2, pp. 364–369, 2017.

[131] N. R. Cohen, K. A. Ellis, W. G. Burns, R. H. Lampe, N. Schuback, Z. Johnson, S. Sañudo-Wilhelmy, and A. Marchetti, "Iron and vitamin interactions in marine diatom isolates and natural assemblages of the Northeast Pacific Ocean," *Limnol. Oceanogr.*, vol. 62, pp. 2076–2096, 2017.

[132] S. A. Sañudo-Wilhelmy, C. J. Gobler, M. Okbamichael, and G. T. Taylor, "Regulation of phytoplankton dynamics by vitamin B_{12}," *Geophys. Res. Lett.*, vol. 33, no. 4, pp. 10–13, 2006.

[133] K. E. Helliwell, G. L. Wheeler, K. C. Leptos, R. E. Goldstein, and A. G. Smith, "Insights into the evolution of vitamin B_{12} auxotrophy from sequenced algal genomes," *Mol. Biol. Evol.*, vol. 28, pp. 2921–2933, 2011.

[134] E. M. Bertrand and A. E. Allen, "Influence of vitamin B auxotrophy on nitrogen metabolism in eukaryotic phytoplankton," *Frontiers Microbiol.*, vol. 3, pp. 1–16, 2012.

[135] J. C. Gonzalez, R. V. Banerjee, S. Huang, J. S. Sumner, and R. G. Matthews, "Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from *Escherichia coli*: Two solutions to the same chemical problem," *Biochem.*, vol. 31, no. 26, pp. 6045–6056, 1992.

[136] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, and J. Harnly, "Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Inititative (MSI)," *Metabolomics*, vol. 3, no. 3, pp. 211–221, 2007.

[137] W. M. Johnson, M. C. Kido Soule, and E. B. Kujawinski, "Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP," *ISME J.*, vol. 10, pp. 2304–2316, 2016.

[138] S. C. Lu, "S-Adenosylmethionine," *Int. J. Biochem. Cell Biol.*, vol. 32, no. 4, pp. 391–395, 2000.

[139] S. Roje, "S-Adenosyl-l-methionine: Beyond the universal methyl group donor," *Phytochemistry*, vol. 67, no. 15, pp. 1686–1698, 2006.

[140] D. R. Hoffman, D. W. Marion, W. E. Cornatzer, and J. A. Duerre, "S-Adenosylmenthionine and S-Adenosylhomocysteine metabolism in isolated rate liver," *J. Biol. Chem.*, vol. 255, no. 22, pp. 10822–10827, 1980.

[141] H. Jakubowski, "Molecular basis of homocysteine toxicity in humans," *Cellular and Molecular Life Sciences CMLS*, vol. 61, pp. 470–487, 2004.

[142] A. J. Roe, C. O'Byrne, D. McLaggan, and I. R. Booth, "Inhibition of *Escherichia coli* growth by acetic acid: A problem with methionine biosynthesis and homocysteine toxicity," *Microbiol.*, vol. 148, pp. 2215–2222, 2002.

[143] J. Selhub, "Homocysteine metabolism.," *Ann. Rev. Nut.*, vol. 19, pp. 217–246, 1999.

[144] R. H. Allen, P. Stablzr, G. Savage, and J. Lindenbaum, "Metabolic abnormalities in cobalamin (vitamin B_{12}) and folate deficiency," *FASEB J.*, vol. 7, no. 14, pp. 1344–1353, 1993.

[145] S. P. Stabler, D. A. Sampson, L. P. Wang, and R. H. Allen, "Elevations of serum cystathionine and total homocysteine in pyridoxine-, folate-, and cobalamin-deficient rats," *J. Nutr. Biochem.*, vol. 8, no. 5, pp. 279–289, 1997.

[146] L. Hannibal, V. Lysne, A.-L. Bjørke-Monsen, S. Behringer, S. C. Grünert, U. Spiekerkoetter, D. W. Jacobsen, and H. J. Blom, "Biomarkers and algorithms for the diagnosis of vitamin B_{12} deficiency," *Frontiers Mol. Biosci.*, vol. 3, no. June, 2016.

[147] S. P. Stabler, J. Lindenbaum, D. G. Savage, and R. H. Allen, "Elevation of serum cystathionine levels in patients with cobalamin and folate deficiency," *Blood*, vol. 81, no. 12, pp. 3404–13, 1993.

[148] J. D. Finkelstein, W. E. Kyle, J. J. Martin, and A.-M. Pick, "Activation of cystathionine β-synthase by adenosylmethionine and adenosylethionine," *Biochem. Biophys. Res. Commun.*, vol. 66, no. 1, pp. 81–81, 1975.

[149] E. B. Kujawinski, K. Longnecker, H. Alexander, S. T. Dyhrman, C. L. Fiore, S. T. Haley, and W. M. Johnson, "Phosphorus availability regulates intracellular nucleotides in marine eukaryotic phytoplankton," *Limnol. Oceanogr. Lett.*, vol. 2, no. 4, pp. 119–129, 2017.

[150] D. A. Gage, D. Rhodes, K. D. Nolte, W. A. Hicks, T. Leustek, A. J. L. Cooper, and A. D. Hanson, "A new route for synthesis of dimethylsulphoniopropionate in marine algae," *Nature*, vol. 387, no. 6636, pp. 891–894, 1997.

[151] N. L. Kettles, S. Kopriva, and G. Malin, "Insights into the regulation of DMSP synthesis in the diatom *Thalassiosira pseudonana* through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen," *PLoS ONE*, vol. 9, no. 4, 2014.

[152] M. O. Andreae and H. Raemdonck, "Dimethyl sulfide in the surface ocean and the marine atmosphere: A global view," *Science*, vol. 221, no. 4612, pp. 744–747, 1983.

[153] L. V. Kurepin, A. G. Ivanov, M. Zaman, R. P. Pharis, S. I. Allakhverdiev, V. Hurry, and N. P. Hüner, "Stress-related hormones and glycinebetaine interplay in protection of photosynthesis under abiotic stress conditions," *Photosyn. Res.*, vol. 126, no. 2-3, pp. 221–235, 2015.

[154] P. S. Summers and E. A. Weretilnyk, "Choline synthesis in spinach in relation to salt stress," *Plant Physiol.*, vol. 103, no. 4, pp. 1269–1276, 1993.

[155] J. A. Stekol, E. I. Anderson, and S. Weiss, "S-Adenosyl-L-Methionine in the synthesis of choline, creatine, an cysteine *in vivo* and *in vitro*," *J. Biol. Chem.*, vol. 233, no. 2, pp. 425–29, 1958.

[156] R. P. Kiene, L. J. Linn, and J. A. Bruton, "New and important roles for DMSP in marine microbial communities," *J. Sea Res.*, vol. 43, no. 3-4, pp. 209–224, 2000.

[157] J. Sun, L. Steindler, J. C. Thrash, K. H. Halsey, D. P. Smith, A. E. Carter, Z. C. Landry, and S. J. Giovannoni, "One carbon metabolism in SAR11 pelagic marine bacteria," *PLoS ONE*, vol. 6, no. 8, 2011.

[158] P. H. Yancey, "Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses," *J. Exp. Biol.*, vol. 208, no. 15, pp. 2819–2830, 2005.

[159] A. Spielmeyer and G. Pohnert, "Influence of temperature and elevated carbon dioxide on the production of dimethylsulfoniopropionate and glycine betaine by marine phytoplankton," *Mar. Environ. Res.*, vol. 73, pp. 62–69, 2012.

[160] A. J. Michael, "Molecular machines encoded by bacterially-derived multi-domain gene fusions that potentially synthesize, N-methylate and transfer long chain polyamines in diatoms," *FEBS Lett.*, vol. 585, no. 17, pp. 2627–2634, 2011.

[161] E. Albers, "Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5'-methylthioadenosine," *IUBMB Life*, vol. 61, no. 12, pp. 1132–1142, 2009.

[162] E. P. Brass and S. P. Stabler, "Carnitine metabolism in the vitamin B_{12} deficient rat," *Biochem. J.*, vol. 255, no. 1, pp. 153–9, 1988.

[163] K. Sarafoglou, J. Rodgers, A. Hietala, D. Matern, and K. Bentler, "Expanded newborn screening for detection of vitamin B_{12} deficiency," *JAMA*, vol. 305, no. 12, pp. 1198–1200, 2011.

[164] E. Kelmer, G. D. Shelton, D. A. Williams, C. G. Ruaux, M. E. Kerl, and D. P. O'Brien, "Organic acidemia in a young cat associated with cobalamin deficiency," *J. Vet. Emerg. Crit. Care*, vol. 17, no. 3, pp. 299–304, 2007.

[165] U. Maheswari, K. Jabbari, J.-L. Petit, B. M. Porcel, A. E. Allen, J.-P. Cadoret, A. De Martino, M. Heijde, R. Kaas, J. La Roche, P. J. Lopez, V. Martin-Jézéquel, A. Meichenin, T. Mock, M. Schnitzler Parker, A. Vardi, E. V. Armbrust, J. Weissenbach, M. Katinka, and C. Bowler, "Digital expression profiling of novel diatom transcripts provides insight into their biological functions," *Genome Biol.*, vol. 11, no. 8, p. R85, 2010.

[166] J. W. G. Lund, C. Kipling, and E. D. Le Cren, "The inverted microscope method of estimating algal numbers and the statisical basis of estimations by counting," *Hydrobiol.*, vol. 11, no. 2, pp. 143–170, 1958.

[167] M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb, and P. Mallick, "A cross-platform toolkit for mass spectrometry and proteomics," *Nat. Biotechnol.*, vol. 30, no. 10, pp. 918–920, 2012.

[168] H. P. Benton, E. J. Want, and T. M. D. Ebbels, "Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data," *Bioinform.*, vol. 26, no. 19, pp. 2488–2489, 2010.

[169] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, and G. Siuzdak, "XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification," *Anal. Chem.*, vol. 78, no. 3, pp. 779–787, 2006.

[170] R. Tautenhahn, C. Böttcher, and S. Neumann, "Highly sensitive feature detection for high resolution LC/MS," *BMC Bioinform.*, vol. 9, no. 1, p. 504, 2008.

[171] G. Libiseller, M. Dvorzak, U. Kleb, E. Gander, T. Eisenberg, F. Madeo, S. Neumann, G. Trausinger, F. Sinner, T. Pieber, and C. Magnes, "IPO: A tool for automated optimization of XCMS parameters," *BMC Bioinf.*, vol. 16, no. 1, p. 118, 2015.

[172] B. O. Keller, J. Sui, A. B. Young, and R. M. Whittal, "Interferences and contaminants encountered in modern mass spectrometry," *Anal. Chim. Acta*, vol. 627, no. 1, pp. 71–81, 2008.

[173] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka, "MassBank: A public repository for sharing mass spectral data for life sciences," *J. Mass Spectrom.*, vol. 45, no. 7, pp. 703–714, 2010.

[174] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrov, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, and N. Bandeira, "Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking," *Nat. Biotech.*, vol. 34, no. 8, pp. 828–837, 2016.

[175] A. Vaniya and O. Fiehn, "Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics," *Trends in Anal. Chem.*, vol. 69, pp. 52–61, 2015.

[176] Y. Sawada, R. Nakabayashi, Y. Yamada, M. Suzuki, M. Sato, A. Sakata, K. Akiyama, T. Sakurai, F. Matsuda, T. Aoki, M. Y. Hirai, and K. Saito, "RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database," *Phytochem.*, vol. 82, pp. 38–45, 2012.

[177] D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M. A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. MacInnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel, and L. Querengesser, "HMDB: The human metabolome database," *Nucleic Acids Res.*, vol. 35, no. SUPPL. 1, pp. 521–526, 2007.

[178] C. A. Smith, G. O'Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan, and G. Siuzdak, "Metlin A Metabolite Mass Spectral Database," *Ther. Drug Monit.*, vol. 27, no. 6, pp. 747–751, 2005.

[179] J. B. Kruskal and M. Wish, "Multidimensional scaling," in *Sage University Paper series on Quantitative Applications in the Social Sciences* (E. M. Uslaner, ed.), vol. 07-011, pp. 7–28, Beverly Hills and London: Sage Publications, 1978.

[180] E. Saccenti, H. C. Hoefsloot, A. K. Smilde, J. A. Westerhuis, and M. M. Hendriks, "Reflections on univariate and multivariate analysis of metabolomics data," *Metabolomics*, vol. 10, no. 3, pp. 361–374, 2013.

[181] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Royal Stat. Soc. Ser. B (Methodolog.)*, vol. 57, pp. 289–300, 1995.

[182] E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, and M. Bechner, "The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism," *Science*, vol. 306, pp. 79–86, 2004.

[183] C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J. P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De

Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M. P. O. L. Secq, C. Napoli, M. Obornik, M. S. Parker, J. L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. Von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van De Peer, and I. V. Grigoriev, "The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes," *Nature*, vol. 456, no. 7219, pp. 239–244, 2008.

[184] M. Lommer, M. Specht, A. S. Roy, L. Kraemer, R. Andreson, M. A. Gutowska, J. Wolf, S. V. Bergner, M. B. Schilhabel, U. C. Klostermeier, R. G. Beiko, P. Rosenstiel, M. Hippler, and J. LaRoche, "Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation," *Genome Biol.*, vol. 13, no. 7, 2012.

[185] T. Tanaka, Y. Maeda, A. Veluchamy, M. Tanaka, H. Abida, E. Maréchal, C. Bowler, M. Muto, Y. Sunaga, M. Tanaka, T. Yoshino, T. Taniguchi, Y. Fukuda, M. Nemoto, M. Matsumoto, P. S. Wong, S. Aburatani, and W. Fujibuchi, "Oil Accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome," *Plant Cell Online*, vol. 27, no. 1, pp. 162–176, 2015.

[186] P. G. Falkowski and C. Wilson, "Phytoplankton productivity in the North Pacific ocean since 1900 and implications for absorption of anthropogenic CO_2," *Nature*, vol. 355, pp. 741–743, 1992.

[187] S. C. Walpole, D. Prieto-Merino, P. Edwards, J. Cleland, G. Stevens, and I. Roberts, "The weight of nations: An estimation of adult human biomass," *BMC Public Health*, vol. 12, no. 1, p. 1, 2012.

[188] T. I. Eglinton and D. J. Repeta, "Organic matter in the contemporary," in *The oceans and marine geochemistry* (H. Elderfield, ed.), vol. 6, ch. 6.06, p. 145, Elsevier, 2006.

[189] K. R. Heal, N. A. Kellogg, L. T. Carlson, R. M. Lionheart, and A. E. Ingalls, "Metabolic consequences of cobalamin scarcity in diatoms as revealed through metabolomics," *In prep for Protist*, 2018.

[190] C. H. Johnson, J. Ivanisevic, H. P. Benton, and G. Siuzdak, "Bioinformatics: The next frontier of metabolomics," *Analytical Chemistry*, vol. 87, no. 1, pp. 147–156, 2015.

[191] S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea,

C. Cruaud, F. D'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, and P. Bork, "Ocean plankton. Structure and function of the global ocean microbiome.," *Science (New York, N.Y.)*, vol. 348, no. 6237, p. 1261359, 2015.

[192] S. Engelen, P. Hingamp, M. Sieracki, C. Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Bescot, I. Probert, M. Carmichael, J. Poulain, and S. Romac, "Eukaryotic plankton diversity in the sunlit ocean," *Science*, vol. 348, no. 6237, pp. 1261605–1/11, 2015.

[193] B. P. Durham, S. P. Dearth, S. Sharma, S. A. Amin, C. B. Smith, S. R. Campagna, E. V. Armbrust, and M. A. Moran, "Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system," *Environmental Microbiology*, vol. 19, no. 9, pp. 3500–3513, 2017.

[194] C. A. Llewellyn, U. Sommer, C. L. Dupont, A. E. Allen, and M. R. Viant, "Using community metabolomics as a new approach to discriminate marine microbial particulate organic matter in the western English Channel," *Progress in Oceanography*, vol. 137, pp. 421–433, 2015.

[195] W. M. Johnson, *Linking microbial metabolism and organic matter cycling through metabolite distributions in the ocean.* PhD thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2017.

[196] J. J. Polovina, E. A. Howell, D. R. Kobayashi, and M. P. Seki, "The transition zone chlorophyll front updated: Advances from a decade of research," *Progress in Oceanography*, vol. 150, pp. 79–85, 2017.

[197] G. I. Roden, "Subarctic-subtropical transition zone of the North Pacific: Large-scale aspects and mesoscale structure," *NOAA (National Oceanic and Atmospheric Administration) Technical Report NMFS (National Marine Fisheries Service)*, pp. 1–38, 1991.

[198] W. G. Pearcy, "Biology of the transition region," *NOAA (National Oceanic and Atmospheric Administration) Technical Report NMFS (National Marine Fisheries Service)*, pp. 39–55, 1991.

[199] R. M. Laurs and R. J. Lynn, "North Pacific Albacore Ecology and Oceanography," *NOAA Technical Report NMFS-105*, pp. 69–87, 1991.

[200] L. W. Juranek, P. D. Quay, R. A. Feely, D. Lockwood, D. M. Karl, and M. J. Church, "Biological production in the NE Pacific and its influence on air-sea CO_2 flux: Evidence from dissolved oxygen isotopes and O_2/Ar," *Journal of Geophysical Research: Oceans*, vol. 117, no. 5, pp. 1–23, 2012.

[201] B. M. Warrack, S. Hnatyshyn, K.-H. Ott, M. D. Reily, M. Sanders, H. Zhang, and D. M. Drexler, "Normalization strategies for metabonomic analysis of urine samples," *J. Chromatogr. B*, vol. 877, no. 5, pp. 547–552, 2009.

[202] O. D. Myers, S. J. Sumner, S. Li, S. Barnes, and X. Du, "Detailed investigation and comparison of the XCMS and MZmine2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data," *Analytical Chemistry*, vol. 89, no. 17, pp. 8689–8695, 2017.

[203] J. B. Coble and C. G. Fraga, "Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery," *Journal of Chromatography A*, vol. 1358, pp. 155–164, 2014.

[204] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Research*, vol. 44, no. D1, pp. D457–D462, 2016.

[205] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: New perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–D361, 2017.

[206] B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, and M. J. MacCoss, "Skyline: An open source document editor for creating and analyzing targeted proteomics experiments," *Bioinformatics*, vol. 26, no. 7, pp. 966–968, 2010.

[207] K. R. Clarke, "Non-parametric multivariate analyses of changes in community structure," *Australian Journal of Ecology*, vol. 18, pp. 117–143, 1993.

[208] R. Lim, "Investigating microbial community composition and diversity across the North Pacific," *University of Washington, School of Oceanography Master's Thesis*, In preparation for submission.

[209] Gary J. Patti, O. Yanes, and G. Siuzdak, "Metabolomics: The apogee of the omics trilogy," *Nature Reviews Molecular Cell Biology*, vol. 13, pp. 263–269, 2012.

[210] M. D. Keller, P. A. Matrai, R. P. Kiene, and W. K. Bellows, "Responses of coastal phytoplankton populations to nitrogen additions: dynamics of cell-associated dimethylsulfoniopropionate (DMSP), glycine betaine (GBT), and homarine," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 61, no. 5, pp. 685–699, 2004.

[211] P. Polychronopoulos, P. Magiatis, A. L. Skaltsounis, F. Tillequin, E. Vardala-Theodorou, and A. Tsarbopoulos, "Homarine, a common metabolite in edible mediterranean molluscs: Occurrence, spectral data and revision of a related structure," *Natural Product Letters*, vol. 15, no. 6, pp. 411–418, 2001.

[212] I. K. Suwetja, K. Hori, K. Miyazawa, and K. Ito, "Changes in content of ATP-related compounds, trigonelline in marine invertebrates during ice storage," *Nippon Susian Gakkaish*, vol. 55, no. 3, pp. 559–566, 1989.

[213] B. Gebser and G. Pohnert, "Synchronized regulation of different zwitterionic metabolites in the osmoadaption of phytoplankton," *Marine Drugs*, vol. 11, no. 6, pp. 2168–2182, 2013.

[214] D. Petras, I. Koester, R. Da Silva, B. M. Stephens, A. F. Haas, C. E. Nelson, L. W. Kelly, L. I. Aluwihare, and P. C. Dorrestein, "High-resolution liquid chromatography tandem mass spectrometry enables large scale molecular characterization of dissolved organic matter," *Frontiers in Marine Science*, vol. 4, no. December, 2017.

[215] K. Longnecker, J. Futrelle, E. Coburn, M. C. K. Soule, and E. B. Kujawinski, "Environmental metabolomics: Databases and tools for data analysis," *Marine Chemistry*, vol. 177, pp. 366–373, 2015.

[216] M. Ashraf and M. R. Foolad, "Roles of glycine betaine and proline in improving plant abiotic stress resistance," *Environmental and Experimental Botany*, vol. 59, no. 2, pp. 206–216, 2007.

[217] B. Landfald and A. R. Strom, "Choline-glycine betaine pathway confers a high level of osmolyte tolerance in *Escherichia coli*," *Journal of Bacteriology*, vol. 165, no. 3, pp. 849–855, 1986.

[218] D. J. Scanlan, M. Ostrowski, S. Mazard, A. Dufresne, L. Garczarek, W. R. Hess, A. F. Post, M. Hagemann, I. Paulsen, and F. Partensky, "Ecological genomics of marine picocyanobacteria," *Microbiology and Molecular Biology Reviews*, vol. 73, no. 2, pp. 249–299, 2009.

[219] W. A. Klee, H. H. Richards, and G. L. Cantoni, "The synthesis of methionine by enzymatic transmethylation," *Biochimica et Biophysica Acta*, vol. 54, pp. 157–164, 1961.

[220] R. S. Ganu, Y. Ishida, M. Koutmos, S. O. Kolokotronis, A. L. Roca, T. A. Garrow, and L. B. Schook, "Evolutionary analyses and natural selection of betaine-homocysteine S-methyltransferase (BHMT) and BHMT2 genes," *PLoS ONE*, vol. 10, no. 7, pp. 1–19, 2015.

[221] E. Poiata, M. M. Meyer, T. D. Ames, and R. R. Breaker, "A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria," *RNA*, vol. 15, no. 11, pp. 2046–2056, 2009.

[222] L. R. Thompson, G. J. Williams, M. F. Haroon, A. Shibl, P. Larsen, J. Shorenstein, R. Knight, and U. Stingl, "Metagenomic covariation along densely sampled environmental gradients in the Red Sea," *ISME Journal*, vol. 11, no. 1, pp. 138–151, 2017.

[223] W. M. Patrick, "Bacterial methionine biosynthesis," *Microbiology*, vol. 160, no. 2014, pp. 1571–1584, 2014.

[224] A. D. Brown and J. Simpson, "Water relations of sugar-tolerant yeasts: The role of intracellular polyols.," *Journal of General Microbiology*, vol. 72, no. 3, pp. 589–591, 1972.

[225] P. Mohanty, H. Hayashi, G. C. Papageorgiou, and N. Murata, "Stabilization of the Mn-cluster of the oxygen-evolving complex by glycinebetaine," *Biochimica et Biophysica Acta*, vol. 1144, no. 1, pp. 92–96, 1993.

[226] G. C. Papageorgiou and N. Murata, "The unusually strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving Photosystem II complex," *Photosynthesis Research*, vol. 44, no. 3, pp. 243–252, 1995.

[227] S. Belviso, P. Buat-Menard, J. P. Putaud, B. C. Nguyen, H. Claustre, and J. Neveux, "Size distribution of dimethylsulfoniopropionate (DMSP) in areas of the subtropical Atalntic Ocean and the Mediterranean Sea," *Marine Chemistry*, vol. 44, pp. 55–71, 1993.

[228] D. C. Yoch, "Dimethylsulfoniopropionate: Its sources, role in the marine food web, and biological degradation to dimethylsulfide," *Society*, vol. 68, no. 12, pp. 5804–5815, 2002.

[229] M. D. Keller, R. P. Kiene, P. A. Matrai, and W. K. Bellows, "Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures," *Marine Biology*, vol. 135, no. 2, pp. 237–248, 1999.

[230] M. D. Keller, R. P. Kiene, P. A. Matrai, and W. K. Bellows, "Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. II. N-limited chemostat cultures," *Marine Biology*, vol. 135, no. 2, pp. 249–257, 1999.

[231] H. B. Swan, E. S. Deschaseaux, G. B. Jones, and B. D. Eyre, "The relative abundance of dimethylsulfoniopropionate (DMSP) among other zwitterions in branching coral at Heron Island, southern Great Barrier Reef," *Analytical and Bioanalytical Chemistry*, vol. 409, no. 18, pp. 4409–4423, 2017.

[232] A. Bashir, T. Hoffmann, B. Kempf, X. Xie, S. H. J. Smits, and E. Bremer, "Plant-derived compatible solutes proline betaine and betonicine confer enhanced osmotic and temperature stress tolerance to *Bacillus subtilis*," *Microbiology*, vol. 160, no. 2014, pp. 2283–2294, 2014.

[233] L. Marion, "The Pyrrolidine Alkaloids," in *The alkaloids* (R. H. F. Manske, ed.), vol. 6, Chemist, pp. 31–34, 1960.

[234] K. Kishor, P. B, S. Sangam, R. N. Amrutha, P. S. Laxmi, K. R. Naidu, K. R. S. S. Rao, S. Rao, K. J. Reddy, P. Theriappan, and N. Sreenivasulu, "Regulation of proline biosynthesis, degradation,uptake and transport in higher plants: Its implications in plant growth and abiotic stress tolerance," *Current Science*, vol. 88, no. December 2015, pp. 424–438, 2005.

[235] H. B. Liu, K. P. Koh, J. S. Kim, Y. Seo, and S. Park, "The effects of betonicine, floridoside, and isethionic acid from the red alga *Ahnfeltiopsis flabelliformis* on quorum-sensing activity," *Biotechnology and Bioprocess Engineering*, vol. 13, no. 4, pp. 458–463, 2008.

[236] K. Haug, R. M. Salek, P. Conesa, J. Hastings, P. De Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. González-Beltrán, S. A. Sansone, J. L. Griffin, and C. Steinbeck, "MetaboLights - An open-access general-purpose repository for metabolomics studies and associated meta-data," *Nucleic Acids Research*, vol. 41, pp. D781–D786, 2013.

[237] K. Hanaoka, T. Fujita, M. Matrsuura, S. Tagawa, and T. Kaise, "Identification of arsenobetaine as a major arsenic compound in muscle of two demersal sharks, shortnose dogfish," *Comp. Biochem. Physiol.*, vol. 86B, no. 4, pp. 681–682, 1987.

[238] J. M. Neff, "Ecotoxicology of arsenic in the marine environment," *Environmental Toxicology and Chemistry*, vol. 16, no. 5, pp. 917–927, 1997.

[239] S. Foster and W. Maher, "Arsenobetaine and thio-arsenic species in marine macroalgae and herbivorous animals: Accumulated through trophic transfer or produced *in situ*?," *Journal of Environmental Sciences (China)*, vol. 49, pp. 131–139, 2016.

[240] R. O. Jenkins, A. W. Ritchie, J. S. Edmonds, W. Goessler, N. Molenat, D. Kuehnelt, C. F. Harrington, and P. G. Sutton, "Bacterial degradation of arsenobetaine via dimethylarsinoylacetate," *Archives of Microbiology*, vol. 180, no. 2, pp. 142–150, 2003.

[241] E. R. Hall and S. Gurin, "Experiments in marine biochemistry: Homarine metabolism in *Penaeu Durarum*," *The Journal of Biological Chemistry*, vol. 250, no. 17, pp. 6943–6946, 1975.

[242] K. J. Flynn, D. M. J. Dickson, and O. A. Al-Almoudi, "The ratio of glutamine:glutamate in mircroalgae: A biomarker for N-status suitable for use at natural cell densities," *Journal of Plankton Research*, vol. 11, no. 1, pp. 165–170, 1989.

[243] K. J. Flynn and J. Fielder, "Changes in intracellular and extracellular amino acids during the predation of the chlorophyte *Dunaliella primolecta* by the heterotrophic dinoflagellate *Oxyrrhis marina* and the use of the glutamine/glutamate ratio as an indicator of nutrient status in mix," *Marine Ecology Progress Series*, vol. 53, no. 2, pp. 117–127, 1989.

[244] G. yuan Chen, H. wei Liao, Y. J. Tseng, I. lin Tsai, and C. hua Kuo, "A matrix-induced ion suppression method to normalize concentration in urinary metabolomics studies using flow injection analysis electrospray ionization mass spectrometry," *Analytica Chimica Acta*, vol. 864, no. 33, pp. 21–29, 2014.

[245] S. M. Sowell, L. J. Wilhelm, A. D. Norbeck, M. S. Lipton, C. D. Nicora, D. F. Barofsky, C. A. Carlson, R. D. Smith, and S. J. Giovanonni, "Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea," *ISME Journal*, vol. 3, no. 1, pp. 93–105, 2009.

[246] K. Bergauer, A. Fernandez-Guerra, J. A. L. Garcia, R. R. Sprenger, R. Stepanauskas, M. G. Pachiadaki, O. N. Jensen, and G. J. Herndl, "Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics," *Proceedings of the National Academy of Sciences*, vol. 115, no. 3, pp. E400–E408, 2018.

[247] R. Kumar, S. Zhao, M. W. Vetting, B. M. K. Wood, A. Sakai, K. Cho, J. Solbiati, S. C. Almo, J. V. Sweedler, M. P. Jacobson, J. A. Gerlt, and J. E. Cronan, "Prediction and biochemical demonstration of a catabolic pathway for the osmoprotectant proline betaine," *mBio*, vol. 5, no. 1, pp. 1–13, 2014.

Appendix A

SUPPLEMENTAL DATASETS AND TABLES

All of the listed datasets and tables are available as downloadable excel files.

A.1 Supplemental Datasets for Chapter 3

A.1.1 Dataset S3.1

Genomes and gene search material. Shown are full genomes and functions searched in the IMG database (`https://img.jgi.doe.gov`) to make Figure 3.5

A.1.2 Dataset S3.2

Raw data files of environmental and culture data used to make Figures 3.2, 3.3, 3.4, and 3.8.

A.2 Supplemental Datasets for Chapter 4

A.2.1 Dataset S4.1

Full results from targeted metabolomics analysis of *T. pseudonana*. Treatments are noted as RCobSL (replete cobalamin, saturating light); RCobLL (replete cobalamin, low light); LCobSL (low cobalamin, saturating light); and LCobLL (low cobalamin, low light), each has three replicates (AB, CD, and EF). Reported peak areas are adjusted (via B-MIS normalization [8]) and then normalized to biomass.

A.2.2 Dataset S4.2

Full results from targeted metabolomics analysis of *N. pelliculosa*. Treatments and values are the same as in Dataset S4.1.

A.2.3 Dataset S4.3

Full results from untargeted metabolomics analysis of *T. pseudonana*. Treatments and values are the same as in Dataset S4.1. MF_Frac = mass feature and fraction (RP or HILIC, aqueous or organic), mz = *m/z*, rt = retention time (in seconds), fraction = aqueous or organic.

A.2.4 Dataset S4.4

Full results from untargeted metabolomics analysis of *N. pelliculosa*. Treatments and values are the same as in Dataset S4.1. Abbreviations as in Dataset S4.3.

A.3 Supplemental Datasets for Chapter 5

A.3.1 Dataset S5.1

Mass features from untargeted analysis that correspond to isotopes and adducts of other mass features. MF_Frac is the ID for mass feature; *m/z*; and retention time (RT, in seconds); IsotopeMatch is adduct or isotope and MF_Frac of associated matching MF_Frac. Samples are as labeled in Table 5.1. Reported peak areas are adjusted (via B-MIS normalization [8]) and then normalized to WV.

A.3.2 Dataset S5.2

Full results from our targeted metabolomics analysis. Samples are as labeled in Table 5.1. Reported peak areas are adjusted (via B-MIS normalization [8]) and then normalized to WV.

A.3.3 Dataset S5.3

Full results from untargeted metabolomics analyses. Samples are as in Table 5.1. MF_Frac = mass feature and fraction (RP or HILIC, aqueous or organic), mz = *m/z*, rt = retention time (in seconds), fraction = aqueous or organic.

A.3.4 Dataset S5.4

Full results (each sample) of compounds which were detected in our sample set that we were able to offer absolute concentrations. Samples are as in Table 5.1. All concentrations are in pM.