UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Novel analytical approaches to investigate structure, source, and cycling of marine dissolved organic nitrogen

Permalink

https://escholarship.org/uc/item/66m7b5hb

Author

Ianiri, Hope L

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

**NOVEL ANALYTICAL APPROACHES TO INVESTIGATE
STRUCTURE, SOURCE, AND CYCLING OF MARINE
DISSOLVED ORGANIC NITROGEN**

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

DOCTOR OF PHILOSOPHY

in

OCEAN SCIENCES

by

Hope Ianiri

December 2021

The Dissertation of Hope Ianiri is approved:

_____

Professor Matthew D. McCarthy, Chair

_____

Associate Professor Pratigya Polissar

_____

Carolina Distinguished Professor Emeritus Ronald Benner

_____

Dr. Harris Mason

_____

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Novel analytical approaches to investigate structure, source, and cycling of
marine dissolved organic nitrogen

by

Hope Ianiri

Marine dissolved organic nitrogen (DON) represents the largest reactive nitro-
gen reservoir in the world and by far the largest pool of actively cycling reduced N
in the oceans. Yet, despite this huge pool of DON, primary production is nitrogen
limited in extensive oligotrophic regions of the surface ocean. The recalcitrant na-
ture of the vast majority of DON exerts a control on all marine food webs as well
as carbon sequestration by the ocean. However, despite significant research, the
mechanisms that lead to the accumulation of refractory DON (RDON) remain an
enigma.

Here, I address this knowledge gap by investigating the source and degradation
processes of a unique dissolved organic matter (DOM) sample set isolated from
the North Pacific and North Atlantic Subtropical Gyres. Using a new DOM isola-
tion scheme, I compare more traditionally studied younger, high molecular weight
(HMW) DOM with older, low molecular weight (LMW) solid phase extracted
(SPE) DOM. A novel aspect of this thesis is targeted investigation of LMW mate-
rial, which represents some of the oldest, and presumably most refractory, marine
DOM, but previously could not be isolated for direct analyses.

Chapters two and three utilize amino acids (AA) as powerful molecular level
proxies for proteinaceous marine DON source and cycling. In chapter two, I
coupled a new suite of D-AA bacterial tracers with AA-based proxies for bacte-
rial degradation and radiocarbon ages to investigate bacterial influence on DON

with age and ocean circulation. In chapter three, I made the first compound specific AA isotope measurements across the DON age/size spectrum to propose specific N sources as well as formation and degradation mechanisms of HMW and LMW SPE-DON. The results of these two chapters indicated that, contrary to our understanding of the wider DOC pool, production and degradation of proteinaceous HMW and LMW SPE-DON may be completely independent. While both appear dominated by bacterial molecules, I suggest HMW AA-containing molecules are surface produced and progressively degraded while LMW AA-containing molecules are produced relatively rapidly in the surface ocean and persist for millennia. These results are a substantial departure from current assumptions based on marine DOC and suggest a new interpretation for formation and cycling of refractory DON.

For my fourth chapter, I applied to DOM for the first time a novel solid-state nuclear magnetic resonance (NMR) spectroscopy technique, multiple cross polarization (multiCP) magic angle spinning (MAS). In contrast to traditional CP/MAS NMR, which was used for almost all past marine DOM NMR analyses, multiCP/MAS NMR is optimized to resolve non-protonated nuclei which likely represent most RDON and RDOC, such as heterocyclic nitrogen and carboxyl and aromatic carbon. These far more quantitative data of both HMW and LMW SPE-DON from the surface and deep ocean likely represent the most accurate picture of marine DOC and DON functionality to date. Based on these results, I suggest inherently stable molecules contribute to the refractory nature of both HMW and DOM. However, I also find that different groupings of biomolecules in DOC and DON are independently responsible for the recalcitrance in each pool.

This dissertation presents novel information regarding source, structure, and cycling of marine DON. I suggest specific production mechanisms for AA-containing

molecules in HMW and LMW DON and provide a new theory for bacterial control of RDON. By pairing these molecular level analyses with broader, advanced solid state NMR techniques, I additionally suggest new interpretations for the functional composition of RDOC and RDON. Overall, these new data imply a novel theory regarding production of the most refractory nitrogen containing molecules in the ocean and suggest a paradigm shift in our understanding of the marine DON pool.

To my family.

Acknowledgments

There are many people without whom this work would not have been possible. First, I am incredibly grateful to my advisor, Matthew McCarthy, for your constant support, advice, and feedback. Your enthusiasm for science is contagious, and I walk away from every meeting with you more excited about my research. I am also very grateful to my committee members, who have all been incredibly supportive and were instrumental in the completion of this dissertation. Pratigya, thank you for always being available to provide advice, whether it is science-based, career-related, or any other questions I may have. Ron, thank you for your invaluable insight since the start of my degree, which has helped me contextualize my work in terms of wider DOM research and has helped shape my dissertation. And Harris, thank you for spending so much time teaching me solid-state NMR, without which my fourth chapter would not have been possible, as well as your continued career advice and support.

Thank you to all past and present members of the McCarthy lab; I have learned so much from all of you and am very grateful for the thoughtful discussions, peer reviews, and practice talks. I am especially grateful to Stephanie Christensen, who has taught me so much about GC maintenance, GC-IRMS, and many other laboratory analyses. You were always available to answer any question I had, and I truly have no idea how our lab would function without you. Thank you also to Taylor Broek for guidance both in the lab and as I entered the complex field of dissolved organic matter research. Furthermore, I would like to thank all the members of the LLNL NMR facility, especially April Sawvel and Maxwell Marple, who continuously took time out of their busy schedules to help me with my NMR experiments.

Additionally, a huge thank you to my family and friends. I am incredibly for-

tunate to have such a strong and loving support system. I would especially like to thank my parents, who have provided endless love, support, advice, encouragement, and so much more. Without you this never would have been possible.

Chapter 1

Introduction

Dissolved organic matter

Marine dissolved organic matter (DOM) represents the largest pool of reduced carbon and nitrogen in the oceans. The subset of this pool which contains nitrogen, dissolved organic nitrogen (DON), is of particular importance to marine food webs and carbon sequestration, as a lack of usable nitrogen restricts primary production throughout much of the world's oceans. This suggests most marine DON is resistant to biological degradation, but despite its global importance, marine DON source, cycling processes, and chemical composition remain elusive.

Much of what is known about DON is inferred from an understanding of the wider studied dissolved organic carbon (DOC) pool. Dissolved organic carbon (DOC) is an incredibly complex mixture (Dittmar, 2015; Stubbins & Dittmar, 2014), and only a very small portion of the total DOC pool has been molecularly characterized. Instead, initial DOC observations relied on bulk concentration and isotopic measurements. DOC concentrations and Δ^{14}C values decrease exponentially with depth nearly everywhere in the ocean, with deep ocean DOC averaging 4,000 to 6,000 years old (Bauer et al., 1992; Williams & Druffel, 1987). Because of this, DOM is assumed to cycle as a "two-pool" model, with accumulation of semi-labile (SL), young DOM at the surface on top of consistent, background concentrations of old, refractory DOM throughout the water column. More recently, studies have shown that DOC Δ^{14}C age is directly correlated with molecular size; young, SL-DOM is mostly composed of high molecular weight (HMW) molecules while

old, RDOM is dominated by low molecular weight (LMW) molecules (Walker et al., 2011, 2014, 2016). Together, this work has been termed the size-age-reactivity continuum (reviewed by Benner and Amon, 2015; Walker et al., 2016).

Adding to the analytical challenge of studying DOM is the difficulty of isolating sufficient quantities for detailed analyses. For decades, ultrafiltration has routinely been used to isolate large volume samples of HMW, ultrafiltered (U)DOM (Benner et al., 1992), allowing for extensive study of the HMW DOM pool (Aluwihare et al., 2005; Benner et al., 1997; Kaiser & Benner, 2009; McCarthy et al., 1997, 1998, 2007). However, ultrafiltration only accounts for 20% to 40% of the total DOM pool, which is now known to be mostly SL, younger DOM. Additional methods for DOM isolation include solid phase extraction (SPE) with a range of organic sorbents (Dittmar et al., 2008) and reverse osmosis coupled with electrodialysis (RO/ED) (Koprivnjak et al., 2006, 2009), though neither method efficiently isolates solely LMW DOM. Thus, despite these advances, collecting enough LMW DOM for detailed analyses continues to be a limiting control on studies specific to RDOM.

Despite these challenges, many hypotheses have been put forward for the recalcitrant nature of DOM (Dittmar, 2015 and references therein). Recent studies increasingly focus on microbial alternation and resynthesis as main production mechanisms of RDOM. The microbial carbon pump (MCP) theory (Jiao et al., 2010, 2011) suggests that progressive reworking of DOM by microbes produces refractory DOC, "pumping" it into the deep ocean where it persists for thousands of years. In context of a size-age-reactivity continuum, this suggests bacteria are responsible for producing LMW RDOC from HMW, SL DOC.

The mystery of DON

Compared to DOC, one could argue DON has even greater uncertainties and

analytical challenges. Open ocean DON concentrations follow the same exponential trend with depth as DOC (Hansell & Carlson, 2001; Sipler & Bronk, 2015), and DON is also assumed to cycle according to a two-pool model. However, there is no known method to isolate only nitrogen containing molecules, meaning any detailed information regarding the DON pool must rely on nitrogen specific tools or molecular level tracers.

Some of the first studies specific to the DON pool relied on solid-state ^{15}N NMR. Initial ^{15}N NMR analyses found HMW DON to be composed almost entirely of amide N, and indicated that most subsurface HMW DON is proteinaceous material (Aluwihare et al., 2005; McCarthy et al., 1997). Surprisingly, despite the fact that amino acids (AA) are presumed to be labile biomolecules (Cowie & Hedges, 1994; Jørgensen et al., 2014), proteinaceous material has been shown to have radiocarbon ages older than that of ocean mixing (Loh et al., 2004). This present a puzzle: if most functionally characterized DON consists of supposedly labile biomolecules, why does the majority of DON appear to be refractory and persist in the ocean for millennia?

Novel isolation method and analyses of LMW SPE-DON

Up to now, detailed measurements specifically of RDON, which are necessary for understanding the persistence of most marine DON, have been limited by the lack of any known method to isolate only LMW material. Recently, however, a novel dual filtration method was introduced which couples ultrafiltration and solid phase extraction to individually isolate young, HMW DOM and old, LMW SPE-DOM (Broek et al., 2017). With the ability to collect significant amounts of LMW material, far more detailed analyses can be made on this size fraction which can then be interpreted in the context of the younger, semi-labile HMW material that has long been studied. Novel solid-state ^{15}N NMR of these samples have

shown LMW SPE-DON is completely different compositionally than previously believed and is almost entirely composed of complex heterocyclic nitrogen containing molecules (Broek et al., Submitted). It is suggested these molecules are breakdown products of autotrophic and heterotrophic bacterial molecules which are inherently stable and can persist for millennia. If true, these results have major implications for our understanding of DON cycling and suggest targeted investigation of heterocyclic nitrogen compounds may be the key to understanding RDON structure and cycling.

Dissertation overview

In this thesis, I apply novel analytical tools to HMW DON and LMW SPE-DON from the North Pacific and North Atlantic Subtropical Gyres. I pair molecular-level techniques with advanced bulk nitrogen structural analyses to directly investigate DON sources, degradation mechanisms, and detailed structure of HMW and LMW DON.

In chapter two, D and L-AA were measured as indicators of bacterial contribution to HMW and LMW SPE-DON. These bacterial proxies are interpreted in context of established AA proxies for bacterial resynthesis and degradation as well as radiocarbon age. This holistic approach yields novel information regarding bacterial source and cycling of these two pools and provides contextual information for the remainder of my thesis.

In chapter three, I applied compound specific isotope analysis of amino acids (CSI-AA) to evaluate nitrogen source and specific degradation mechanisms of HMW and LMW SPE-DON. This is the first application of CSI-AA to LMW DON, as well as to DON from the Atlantic Ocean. Using this technique, I suggest specific mechanisms of production and degradation for HMW and LMW AA-containing molecules.

4

In the fourth chapter of this dissertation, I applied cutting-edge solid-state ^{13}C and ^{15}N NMR techniques which overcome known quantitation issues with all past DOM NMR analyses. These results demonstrate that both HMW and LMW SPE-DON have more diverse structure than previously believed, with heterocyclic N contributing to both size fractions and a range of heterocyclic functionality dominating LMW SPE-DON. Based on these new analyses, I also find evidence that most aromatic DOC likely contains N. Overall, these results indicate that inherently stable DOC and DON structures contribute to the refractory nature of both DOM pools.

Collectively, by pairing specific, molecular-level analyses with broader, bulk nitrogen structural information, the results of this dissertation provide new information regarding the structure, formation, and cycling dynamics of marine DON. These new data represent a novel interpretation for the production of marine RDON and address important knowledge gaps regarding the chemical composition of both semi-labile and refractory DON.

Chapter 2

Distinct bacterial sources and cycling dynamics of HMW and LMW SPE-DON in the ocean

The contents of this chapter have been submitted to the journal *Marine Chemistry*.

2.1 Abstract

Amino acids (AA) represent the most abundant identifiable biomolecule class in marine dissolved organic nitrogen (DON) and provide powerful proxies for ON degradation state. D-AA are only produced in large quantities by bacteria, representing an ideal direct tracer for bacterially derived N. However, it remains unclear if D-AA accumulation and speciation in the ocean indicates that most DON arises from direct bacterial sources or from continual bacterial alteration of eukaryotic algal biosynthate, and which of these mechanisms predominantly controls the refractory DON (RDON) which accumulates in the deep sea. Here, we present the most extensive D-AA suite ever reported in younger, high molecular weight (HMW) DON, contrasted with older, low molecular weight (LMW) solid phase extracted (SPE) DON from the central Atlantic and Pacific Oceans. We evaluate D-AA in the context of multiple common AA-based proxies and bulk DOM radiocarbon (Δ^{14}C) data. Specifically, we assess if D-AA in HMW and LMW SPE-DON are most consistent with 1) preformed bacterial source signals, 2) progressive bacterial degradation/alteration of eukaryotic algal sources, or

3) gradual, continued resynthesis/addition of new bacterial biomolecules during ocean circulation. Our results suggest that AA-containing molecules in HMW and LMW SPE-DON are almost entirely distinct, with independent bacterial sources and degradation mechanisms. In HMW DON, all measured indices support a surface-produced, semi-labile component which is progressively altered with increasing radiocarbon age. In contrast, for LMW SPE-DON, AA-based proxies yielding conflicting results. Some proxies indicated LMW SPE-DON was less labile and more degraded than HMW DON, while others indicated less degradation and resynthesis to this size fraction, suggesting a disconnect in the mechanisms reflect by individual proxies. Limited change in AA composition or degradation state were observed in the subsurface ocean within either size fraction with increasing radiocarbon age, supporting the idea that HMW and LMW pools cycle independently. Together, our results suggest AA DON sources are almost entirely bacterial and are more diverse than previously believed, with much of the hydrolysable AA pool in the ocean not deriving directly from proteins as has been commonly assumed. Overall, these observations support the microbial nitrogen pump idea, with compositionally unique refractory components in both HMW and LMW material resisting degradation over millennial timescales.

2.2 Introduction

Marine dissolved organic matter (DOM) represents the largest pool of reduced carbon and nitrogen in the oceans, yet its long-term persistence remains an enigma. DOM is widely assumed to exist along a size-age-reactivity continuum, with high molecular weight (HMW) molecules representing young, semi-labile DOM while low molecular weight (LMW) molecules make up most of the old, refractory DOM (RDOM) (Walker et al., 2014, Walker et al., 2016, reviewed by Benner and Amon 2015). Increasing evidence suggests that bacteria are a key player in forming RDOM, which persists for thousands of years (Gruber et al., 2006; Ogawa et al., 2001; Yamashita & Tanoue, 2008). This concept, termed the microbial carbon pump (MCP, [Jiao et al., 2010, 2011]), has been studied at length (see reviewed literature in *The Microbial Carbon Pump in the Ocean*, edited by Jiao et al. 2011) and is supported by culturing studies (Gruber et al., 2006; Kawasaki & Benner, 2006; Lechtenfeld et al., 2015; Ogawa et al., 2001), in situ observations (Azam et al., 1983; Jørgensen et al., 2003; Nagata et al., 2003), and the incredible structural complexity of dissolved organic carbon (DOC) (Dittmar, 2015; Lechtenfeld et al., 2014). These observations, in context of a size-age-reactivity continuum, suggests that LMW RDOC is formed by progressive microbial degradation of semi-labile, HMW material.

However, work investigating bacterial production of dissolved organic nitrogen (DON), especially refractory DON (RDON), is far more limited. In large part this is due to the lack of methods to isolate the entire DON pool, such that molecular study of RDON formation and cycling requires N-specific molecular level tracers. Most previous DON structural work also focused only on HMW DON due to its ease of isolation. However, in contrast to the complex DOC pool, ^{15}N NMR studies indicated that most HMW DON is structurally simple,

appearing almost entirely in amide form (Aluwihare et al., 2005; McCarthy et al., 1997) and composed mostly of proteinaceous material at subsurface depths (Aluwihare et al., 2005). While this presents a paradox in terms of supposedly highly labile AA, these observations also underscore that AA-based proxies are the most powerful molecular-level biomarkers currently available to investigate the DON pool. While molecular-level analyses indicate AA are generally less than 10% of total DON (Benner, 2002; Kaiser & Benner, 2009), they still represent the largest identifiable biomolecules in DON. Thus, if bacteria form LMW RDON from semi-labile, HMW DON via a microbial nitrogen pump, analogous to the broader paradigm for the DOC pool (Yamaguchi & McCarthy, 2018), then a comparison of HMW versus LMW AA signatures should reflect progressive microbial processes.

Almost all protein AA (except for Glycine) are chiral, meaning they have an L and D-enantiomer that have the same molecular formulae but are non-superimposable mirror images of each other. While all living organisms produce L-AA, D-AA are produced in large quantities exclusively by prokaryotes (Radkov & Moe, 2014), meaning they are arguably the most direct bacterial biomarkers in marine DON (Broek et al., 2019; Kaiser & Benner, 2008; McCarthy et al., 1998). D-AA correlate with bacterial growth and degradation of organic matter (Kawasaki & Benner, 2006; Tremblay & Benner, 2009) and have been widely used to estimate contributions of bacterial OC and ON in sediments, particles and the DOM pool (Bourgoin & Tremblay, 2010; Kaiser & Benner, 2008; Tremblay & Benner, 2006). Selected D-AA are also highly abundant in marine DON, representing one of the central pieces of evidence for microbial origin (Broek et al., 2019; Kaiser & Benner, 2008; McCarthy et al., 1998).

However, despite this widespread use of D-AA as indicators of bacterial source and "degradation," the interpretation of D-AA patterns in terms of specific sources

and cycling of marine DON remain poorly understood. For example, marine DON D/L ratios do not increase with depth or age, as would be expected if continued bacterial alteration resulted in a higher relative abundance of D-AA (Broek et al., 2019; McCarthy et al., 1998; Pèrez et al., 2003). In addition, D-Alanine (Ala), by far the most abundant D-AA species in DON, often exhibits depth trends opposite from other D-AA (Broek et al., 2019; Kaiser & Benner, 2008) and has been suggested to be just as labile as its L-AA counterpart (Broek et al., 2019; Wang et al., 2020). These complex aspects suggest that a mechanistic understanding of bacterial processes responsible for marine RDON formation requires a clearer distinction between signatures of bacterial source versus poorly defined "degradation," which can encompass many processes such as bacterial biomass addition, alteration, or selective removal.

Evaluating multiple proxies for bacterial alteration and OM reactivity together represents one way to advance this mechanistic understanding. In addition to D-AA, multiple AA-based proxies have been developed to measure "degradation" based on observed differences in the AA composition of natural organic matter over a wide range of reactivity (Cowie & Hedges, 1994; Dauwe et al., 1999; Dauwe & Middelburg, 1998) and in incubation studies (Amon et al., 2001; Calleja et al., 2013; Davis et al., 2009). Individual AA proxies include nonprotein amino acids (NPAA) β-alanine (β-Ala) and γ-aminobutyric acid (γ-Aba), produced by the degradation of aspartic and glutamic acid, respectively (Cowie & Hedges, 1994; Davis et al., 2009; Lee & Cronin, 1982), and the abundance of glycine (Gly), from either preservation of diatom cell walls (Dauwe & Middelburg, 1998) or an increase in other Gly-rich bacterial molecules (Kaiser & Benner, 2009; Lehmann et al., 2020; Nguyen & Harvey, 1997). The degradation index (DI) is a widely applied index based on changes in mol% of multiple AA (Dauwe et al., 1999),

while the AA yield (%C-AA) is a broad indicator of general OM lability (Amon et al., 2001; Benner, 2002; Cowie & Hedges, 1994). Finally, the ΣV parameter is a relatively new, individual AA $\delta^{15}N$ -based proxy specific for bacterial resynthesis based on observations that heterotrophic bacterial resynthesis causes "scattering," or randomization of specific AA $\delta^{15}N$ values (McCarthy et al., 2007). Autotrophic biomass consistently has a ΣV value between 0 and 1, thus any ΣV values greater than this indicates bacterial processes causing isotopic fraction of AA regardless of compound type or OM matrix. Collectively, these AA-based proxies provide a range of potentially independent measures of bacterial source, reactivity, or degradation state of organic matter.

The overarching goal of this paper is to, for the first time, to take a holistic approach which combines the diverse information of these AA-proxies with a recently reported expanded suite of D-AA tracers (Broek et al., 2019). We hypothesize that together these data have the potential to unravel bacterial roles in DON preservation at a new mechanistic level. We report all these proxies for DON together and evaluate them in the context of both DOM molecular weight and $\Delta^{14}C$ from the surface to deep ocean in the central Atlantic and Pacific basins. This unique sample set allowed us to examine preformed bacterial source versus progressive bacterial degradation in the oldest vs. youngest components of the marine DON pool. We hypothesized that DON in the older waters in the deep Pacific would show evidence of more degradation at the molecular level compared to younger deep waters in the Atlantic, consistent with previous reports of lower concentrations of DOC and biochemicals in unfiltered deep water at HOT compared to BATS (Hansell & Carlson, 1998; Kaiser & Benner, 2009). Additionally, if it is assumed DON follows a size-age-reactivity continuum similar to DOC (Amon & Benner, 1994, 1996; Walker et al., 2014, 2016), we would expect DON molecular

composition to reflect progressive bacterial degradation of HMW DON to form more stable, refractory LMW DON.

2.3   Methods

2.3.1   DOM sample collection and molecular weight isolation

DOM samples were collected at in the North Atlantic Subtropical Gyre at BATS (31°40'N, 64°10'W) aboard the *R/V Atlantic Explorer* in August 2015 and May 2016 and in the North Pacific Subtropical Gyre at HOT Station ALOHA (22°45'N, 158°00'W) aboard the *R/V Kilo Moana* in August 2014 and May 2015 as described in Broek et al. (2017). Surface seawater samples were collected on each vessel via underway sampling systems at approximately 7.5 m on the *R/V Kilo Moana* and 2 m on the *R/V Atlantic Explorer*. Large volume subsurface samples (~1000 L to 4300 L) were collected via Niskin bottles at depths of 400 m, 850 m, and 2500 m. Full details of the sampling and sample isolation protocols are described in Broek et al. (2017). Briefly, all seawater was filtered through 53 µmol Nitex mesh and pumped through 0.2 µmol cartridge filters. Subsamples for total DOM were frozen in pre-combusted glass vials. HMW DOM was concentrated using large volume tangential-flow ultrafiltration (UF) using four spiral wound PES UF membranes with a molecular weight cut off of 2.5 kDa (GE Osmonics) and a concentration factor of approximately 1000. LMW SPE-DOM was collected via solid-phase extraction of the UF permeate using PPL sorbent (Agilent Bondesil PPL). After desalting via diafiltration (for HMW) and rinsing (for LMW), both fractions were lyophilized and stored as dry powder until analysis. An integrated subsample of the UF permeate was collected by sampling the permeate at constant time intervals throughout the ultrafiltration (Benner et al., 1997).

2.3.2 GC-MS and GC-IRMS sample preparation

To isolate AA for gas chromatograph mass spectrometry (GC-MS) and GC-isotope ratio mass spectrometry (IRMS) analysis, HMW and LMW SPE-DON samples were first hydrolyzed using liquid-state (6 N HCl) acid hydrolysis according to standard conditions (Calleja et al., 2013; Silfer et al., 1991) with protocols to minimize racemization blanks according to Kaiser and Benner (2005). Following the hydrolysis, a norleucine (Nle) internal standard was added to each sample, and hydrolysates were dried under N_2 gas at 60 °C. The dry samples were then redissolved in 0.1 N HCl, filtered with a 0.7 µm GFF filter, and purified using cation-exchange chromatography with Bio-Rad AG50W-X8 resin (200 - 400 mesh) and eluted with ammonium hydroxide according to Takano et al. (2010). Ammonium hydroxide was removed with N_2 gas, and samples were reprotonated with 0.2 N HCl at 110 °C for 5 minutes. Trifluoroacetyl isopropyl ester (TFAiP) derivatives were prepared according to Décima et al. (2017) and Silfer et al. (1991). Finally, AA were purified using liquid-liquid extraction after Ueda et al. (1989). Samples were stored at -20 °C until analysis, at which point they were dried under N_2 gas and dissolved in ethyl acetate for analysis.

2.3.3 GC-MS analysis and quantification of L- & D-AA

AA D/L ratios of HMW and LMW SPE-DON at BATS (2015, 2016) were measured using GC-MS analysis and structural identities verified based on MS fragmentation. An Agilent 7890A/5975B gas chromatograph mass spectrometer equipped with an Altech Chirasil-L-Val column (50m length, 0.25 mm diameter, 0.16 µm film thickness) was used to analyze D and L-AA in HMW UDON and LMW SPE-DON. 1 uL of sample was injected through a splitless inlet at 200 °C with helium gas carrier at 0.9 mL/min. AA were separated using a 3-ramp

temperature program: initial temperature 45 °C; ramp 1: 2 °C/min to 75 °C; ramp 2: 4 °C/min to 110 °C; ramp 3: 1 °C/min to 125, ramp 4: 4 °C/min to 200 °C, 2.5 min hold. Single ion monitoring was used to identify each AA using the following characteristic major ion fragments (m/z): L & D-Ala, 140; L & D-Valine (Val), 168; L-Threonine (Thr), 153; Glycine (Gly), 126; L-Isoleucine (Ile), L & D-Leucine (Leu), L-Nle, 182; L & D-Serine (Ser), 138; L-Proline (Pro), 166; L & D-asparagine + aspartic acid (Asx), 184; L & D- glutamine + glutamic acid (Glx), 180; L & D-Lysine (Lys) 180; L & D-Phenylalanine (Phe), 190; L & D-Tyrosine (Tyr), 203. For AA with the same characteristic fragment ion (Glx and Lys), identification was made based on retention time using external standards. Acid hydrolysis cleaves the terminal amine of glutamine and asparagine, converting them to glutamic acid and aspartic acid, meaning combined concentrations are measured and reported as "Glx" and "Asx." A response factor was calculated for each AA from a linear four-point calibration curve to determine AA concentration. Our large, concentrated samples of both HMW and LMW SPE-DON allowed for GC-MS analysis and mass-spectral confirmation of this expanded suite of D-AA.

2.3.4 GC-IRMS analysis of AA

δ^{15}N-AA measurements were made on a subset of the samples measured for AA D/L ratios, including HMW and LMW SPE-DON collected at HOT in 2015 and BATS in 2016. All isotopic analyses were completed at the University of California, Santa Cruz (UCSC) Stable Isotope Laboratory (SIL) according to established protocols of the McCarthy Lab (McCarthy et al., 2013; Yamaguchi & McCarthy, 2018). Following hydrolysis and column chromatography as detailed above, AA were further purified via HPLC and collected as separate fractions according to Broek et al. (2013) and Broek and McCarthy (2014). The purified

AA fractions were then recombined for further analysis. TFAiP derivatives were made, and samples were purified via liquid-liquid extraction as detailed above (Section 2.3.2). A Thermo Trace Ultra gas chromatograph coupled with a Finnigan MAT DeltaPlus XL IRMS at UCSC SIL was used for GC-IRMS analysis. AA were separated on a BPX-5 column (60 m x 0.32 mm, 1.0 µm film thickness) for δ^{15}N analysis. Samples were injected in triplicates. A total of twelve AA were measured, including Ala, Gly, Thr, Ser, Val, Leu, Ile, Pro, Asx, Glx, Phe, and Lys. The δ^{15}N of AA is reported relative to N$_2$ in air: δ^{15}N (‰) vs. air $=$ $[(^{15}N/^{14}N)_{sample}/(^{15}N/^{14}N)_{air}]$-1 x 1000.

2.3.5 HPLC sample preparation, analysis, and quantification of L- & D-AA

Concentrations of D- and L-enantiomers of AA in total DOM, UF permeate, HWM UDON, and LMW SPE-DON from HOT (2014, 2015) and BATS (2015, 2016) were determined using an Agilent 1260 ultrahigh performance liquid chromatography (UPLC) system equipped with a fluorescence detector (excitation: 330 nm; emission: 450nm) after Shen et al. (2017). HMW and LMW SPE-DON samples were hydrolyzed as detailed above using 6 N HCl at 110°C for 20 hours (Silfer et al., 1991). Total DOM and UF permeate samples were hydrolyzed using a vapor-phase technique with 6 N HCl at 150 °C for 32.5 min (Kaiser & Benner, 2005). Derivatization followed protocols of Kaiser and Benner (2005) using o-phthaldialdehyde and N-isobutyryl-L-cysteine (OPA/IBLC). Samples were then separated on a Poroshell 120 EC-C18 (4.6 x 100 mm, 2.7 µmol particles) column with a linear binary gradient starting with 100% potassium di-hydrogen phosphate (KH$_2$PO$_4$; 48 mmolŁ, pH $=$ 6.25) to 61% KH$_2$PO$_4$ and 39% methanol: acetonitrile (13:1, v/v) at 13.3 min, 46% KH$_2$PO$_4$ at 19.2 min, 40% KH$_2$PO$_4$ at

21.3 min, and 20% KH_2PO_4 at 22 min. Concentrations of L & D-Asx, L & D-Glx, L & D-Ser, L-His, Gly, L-Thr, β-Ala, L-Arg, L & D-Ala, γ-Aba, L-Tyr, L-Val, L-Phe, L-Ile, L-Leu, and L-Lys were determined from external standards of known concentrations. The limit of quantification is ~ 0.5 nmol/L.

2.3.6 Racemization Correction

Hydrolysis conditions for GC-MS, GC-IRMS, and HPLC analyses were identical to those reported in Kaiser and Benner (2005), and racemization was corrected using previously published values. For GC-MS and GC-IRMS analyses, racemization corrections for derivatization via TFAiP were determined independently for each amino acid by measuring the amount of D enantiomer produced in a Pierce H L Amino Acid standard, containing equivalent molar amounts of all AA investigated in this study (Thermo Scientific, 2.5 µmol/mL), derivatized in the same batch as samples.

Derivatization with OPA/IBLC (used for HPLC analyses) does not induce racemization. However, blank corrections were made for background enantiomeric values of the reagents. No racemization was observed during the column chromatography or base dry down (Broek et al., 2019). Thus, for all samples the total racemization blank is the sum of the hydrolysis blank and the derivatization blank. For HPLC analyses utilizing vapor phase hydrolysis, the average total racemization blank was 4.7% D (or equivalent to a D/L ratio of 0.05), while for HPLC analyses utilizing liquid-phase hydrolysis, the average total racemization blank was 1.9% (D/L ratio of 0.02). For GC-MS analyses, the average total racemization blank was 2.0% D (D/L ratio 0.02).

2.3.7 Data analysis

Analyses were completed in Microsoft Excel 365 and R (4.0.5) (R Core Team, 2021). Data were tested for normality using the Shapiro-Wilk test. Statistical differences between means were analyzed using the Welch's two-sample t-test for normally distributed data and the nonparametric Mann-Whitney U-test for non-normal data. To test for significant differences between the slope and y-intercept of linear regression lines, analysis of covariance (ANCOVA) was performed. A 95% confidence interval was used for all statistical tests. Principal component analysis was performed using AA-based proxies (D-AA yield, L-AA yield, %C-AA, Mol% Gly, %NPAA, and DI) and D/L ratios (D/L-Ala, D/L-Asx, D/L-Glx, D/L-Ser, D/L-Leu, D/L- Val, and D/L-Phe) of HMW and LMW SPE-DON. Variables that were not measured in all DOM samples (D/L-Tyr, D/L-Lys, and ΣV) were excluded from the analysis. All variables were scaled to unit variance prior to analysis.

Degradation index (DI) values were calculated according to Dauwe et al. (1999) using the formula DI $= \Sigma[(\text{AA}_i - \text{AVG AA}_i)/\text{SD AA}_i]$ x factor coefficient$_i$, where AVG AA$_i$ and SD AA$_i$ are the average and standard deviation of mol% for each AA of sample i. Factor coefficients from Kaiser and Benner (2009) specific for marine DOM were applied. The ΣV parameter was calculated according to McCarthy et al., 2007, using the equation $\Sigma\text{V} = (1/\text{n})^* \Sigma \text{Abs}_{(\chi i)}$, where χ_i is the offset in δ^{15}N of each trophic AA from the average δ^{15}N of all trophic AA. Total %D was calculated with and without D-Ala, as Total %D $= \Sigma[\text{D-AA}]$ nmol/L $/ \Sigma([\text{L-AA}] + [\text{D-AA}])$ nmol/L. Only AA measured in all samples (all L and D-AA except D-Tyr and D-Lys) were included in this calculation for direct comparison.

2.3.8   Terminology

"HMW DOM/N" and "LMW SPE-DOM/N" are used to refer to the individually isolated size fractions as described in these methods and in Broek et al. (2017). Thus, "HMW" refers to ultrafiltered DOM between 0.2µm and 2.5 kDa, which in this sample set had Δ^{14}C values ranging from -37.3‰ ± 3.8‰ (surface) to -365.7‰ ± 2.3‰ (2500 m) (240 to 3595 years) at HOT and -43.0‰ ± 3.2‰ (surface) to -304.2‰ ± 1.9‰ (2500m) (355 to 2915 years) at BATS (Broek et al., 2017, 2020). "LMW" refers to solid-phase extracted (PPL resin) DOM from the permeate of HMW DOM (smaller than 2.5 kDa), which had Δ^{14}C values ranging from -343.0‰ ± 2.3‰ (surface) to -577.6‰ ± 1.7‰ (2500 m) (3310 to 6860 years) at HOT and -316.1‰ ± 2.0‰ (surface) to -485.5 ± 1.9‰ (850m) (3050 to 5340 years) at BATS (Broek et al., 2017, 2020).

Most AA-based proxies utilized in this study have been used by past work to indicate generalized "degradation" or "degradation state." However, a goal of this paper is to try to understand the mechanistic underpinnings of these proxies. Due to the complex nature of bacterial degradation, we define some terms representing bacterial source and degradation mechanisms as used in the text below (Table S2.1): **"Bacterial source"** refers to prokaryotic biosynthate, as opposed to eukaryotic biosynthate. We use **"heterotrophic resynthesis"** to refer to any changes which occur within a bacterial cell resulting in new synthesis of bacterial biomolecules. Thus, heterotrophic resynthesis always results in a bacterial source, which could include intact cells, altered fragments of bacterial cells, or bacterial exudates. In contrast, we use **"degradation"** to refer to a range of microbial processes, including alteration of molecules, selective removal of labile biomolecules, or addition of new bacterial biomass. Notably, these degradation processes could be affecting eukaryotic biosynthate, meaning bacterial degradation does not nec-

essarily always result in a "bacterial source." For example, if bacterial degradation selectively removes labile biomolecules from eukaryotic biosynthate, the remaining material does not have any molecules synthesized by bacteria themselves, meaning there is no bacterial source.

2.4 Results

2.4.1 Total DOM [D-AA] and [L-AA]

Total hydrolysable L-AA and D-AA concentrations were greater at HOT than BATS in the surface, while at 2500 m concentrations were greater at BATS (Fig. 2.1, Table S2.2). At HOT, the maximal L-AA and D-AA concentrations were observed in the surface, averaging 209.7 $\pm$ 17.4 nM L-AA and 48.5 $\pm$ 2.3 nM D-AA between the two sampling years. L-AA and D-AA concentrations decreased substantially between the surface and 400 m, then continued to decrease slightly between 400 m and 2500 m. The lowest concentrations were observed at 2500 m, averaging 47.6 $\pm$ 6.8 nM L-AA and 8.9 $\pm$ 0.8 nM D-AA. Concentrations of both enantiomers were similar between sampling years.

At BATS, maximal D-AA concentrations were also observed at surface, with an average of 27.1 $\pm$ 3.3 nM between sampling years (Fig. 2.1, Table S2.2). Contrary to at HOT, there were notable differences in measured AA concentrations between different sampling years (Supplementary 2.7.1). The depth of maximal L-AA concentrations varied between sampling years; in 2016 (May) the highest L-AA concentrations were measured at the surface (138.5 nM) while in 2015 (August) the highest L-AA concentrations were measured at 400 m (166.8 nM). In 2015, minimum L-AA and D-AA concentrations were observed at 2500 m depth (53.9 nM and 13.2 nM, respectively), while in 2016, minimum L-AA and D-AA concentrations were observed at 850 m (66.4 nM and 15.3 nM, respectively).

2.4.2 Recovery of HMW and LMW SPE-DON fractions

Complete DOC and DON recovery data of our HMW and LMW SPE-DOM size fractions from total DOM were published previously (Broek, 2019; Broek et al., 2017, 2020). Briefly, the average total DOC recovery across all depths, stations, and sampling seasons was significantly greater in the LMW SPE-DOM fraction (26.7% ± 6.7%) than the HMW UDOM fraction (10.1% ± 3.2%) (Welch's two-sample t-test, p < 0.001) (Table 2.1). The average total DON recovery at all depths, stations, and sampling seasons was not significantly different in HMW UDOM verses LMW SPE-DOM, averaging 13.3% ± 5.4% (Welch's two-sample t-test, p = 0.138).

Hydrolysable AA recovery from total DOM across all depths, stations, and sampling seasons was significantly greater in the HMW UDOM size fraction than the LMW SPE-DOM size fraction (Welch's two-sample t-test, p < 0.001) (Table 2.1). There was no significant difference in AA recovery of HMW UDON at HOT versus BATS, with an average recovery of 25.3% ± 0.8% at both sites. In contrast, AA recovery of LMW SPE-DON was significantly greater at BATS (15.6% ± 3.4%) than HOT (9.9% ± 3.2%) (Welch's two-sample t-test, p = 0.0082). The combined AA recovery of HMW and LMW DOM was not significantly different between HOT and BATS, averaging 37.9 ± 9.2% over all depths and sampling seasons.

LMW SPE-DOM AA recovery was also calculated from the UF permeate subsample, representing total LMW DOM (measured in this study via HPLC). Notably, because a sample representing the entire permeate (thousands of liters of seawater) would have been impossible, our UF permeate samples represent an "integrated" permeate composite taken at regular time intervals during ultrafiltration (Section 2.3.1, Benner et al., 1997). However, the permeate gets more

concentrated throughout ultrafiltration, meaning oversampling at the beginning of ultrafiltration would result in a subsample less concentrated than the total, and oversampling at the end of ultrafiltration would result in a subsample more concentrated than the total. Thus, these permeate measurements likely represent only an approximate value of the actual permeate concentration. Still, we can use these values to estimate recovery of AA in the LMW SPE-DON size fraction, which was not statistically different at HOT and BATS, averaging $17.8\% \pm 5.34\%$ and $20.32\% \pm 5.79\%$, respectively.

2.4.3   L- & D-AA yield of HMW and LMW SPE-DON

By combing HPLC and GC-MS analyses, we measured a total of nine D-AA, thirteen L-AA, one achiral protein AA, and two NPAA. Because all samples investigated in this study (total DOM, HMW DOM, SPE-DON, and UF permeate) were analyzed via HPLC, for consistency, all analyses use concentration data as measured via HPLC when possible. The exceptions are the concentration of D-AA only measured by GC-MS (D-Val, D-Leu, D-Phe, D-Tyr, and D-Lys). Additionally, the D/L ratios of these D-AA only measured by GC-MS are reported as the ratio of D to L-AA concentration both measured by GC-MS. To compare this expanded suite of D and L-AA between ocean basins, we included previously published GC-MS data at HOT for D-AA which were not measured via HPLC (D-Leu, D-Val, and D-Phe, Broek et al., 2019). Overall, AA concentrations as measured by the two methods were similar (Supplementary 2.7.2, Fig. S2.1).

In the HMW size fraction, total D-AA yields (µmol/mgN) were significantly greater at HOT than BATS at all depths (Fig. 2.2, Table S2.3). HMW L-AA (µmol/mgN) yields were also greater at HOT than BATS, though these differences were not greater than analytical error at all depths (Fig. 2.2). In HMW DON, the

L-AA yield decreased from an average surface maximum of 6.1 ± 1.0 µmol/mgN at HOT and 5.6 ± 0.9 µmol/mgN at BATS to average deep (2500 m) values of 4.1 ± 0.4 µmol/mgN at HOT and 3.6 ± 0.5 µmol/mgN at BATS. The D-AA yield of HMW DON showed opposite behavior to the L-AA yield: D-AA yields increased from average surface values of 0.82 ± 0.01 µmol/mgN at HOT and 0.70 ± 0.08 µmol/mgN at BATS to average maximum values at 400 m of 0.90 ± 0.03 µmol/mgN at HOT and 0.71 ± 0.06 µmol/mgN at BATS. At both stations, minimum D-AA yields were observed at 2500 m depth, averaging 0.76 ± 0.02 µmol/mgN at HOT and 0.66 ± 0.06 µmol/mgN at BATS.

L-AA and D-AA yields were lower in LMW SPE-DON than HMW DON throughout the water column in both ocean basins (Fig. 2.2, Table S2.3). Unlike HMW DON, the L-AA and D-AA yields in LMW SPE-DON were similar within error between HOT and BATS at most depths. L-AA and D-AA yields were both greatest at the surface and decreased to the mesopelagic, then were relatively constant with depth. L-AA yields at the surface averaged 2.7 ± 0.2 µmol/mgN at HOT and 2.4 ± 0.2 µmol/mgN at BATS and at 2500m averaged 1.36 ± 0.06 µmol/mgN at HOT and 1.3 ± 0.1 µmol/mgN at BATS. LMW D-AA yields in the surface averaged 0.35 ± 0.03 µmol/mgN at HOT and 0.340 ± 0.003 µmol/mgN at BATS and at 2500m averaged 0.200 ± 0.004 µmol/mgN at HOT and 0.170 ± 0.001 µmol/mgN at BATS.

2.4.4 AA composition of HMW and LMW SPE-DON: D/L ratios and molar abundance

Few significant differences in average AA D/L ratios were observed between HOT and BATS within each size fraction (Fig. 2.3). In HMW DON, average D/L ratios of Leu and Phe were significantly greater at HOT compared to BATS

(Welch's two-sample t-test, p < 0.05). In LMW SPE-DON, no AA had significantly different D/L ratios between the two ocean basins. Similarly, depth profiles of D/L AA ratios were exceptionally consistent between BATS and HOT in both size fractions, with values similar within error for all AA at most depths (Fig. 2.4). In LMW SPE-DON, most AA had a maximum D/L ratio at 400 m, while depth trends in the HMW DON were more variable.

Within each ocean basin, average D/L ratios of most AA were significantly different between HMW and LMW SPE-DON (Fig. 2.3, Fig. 2.4). At HOT and BATS, D/L-Ala was uniquely and significantly greater in HMW DON compared to LMW SPE-DON (Mann Whitney U-test, p < 0.001), while D/L-Asx, D/L-Glx, D/L-Leu, D/L-Val, and D/L-Phe were significantly greater in LMW SPE-DON compared to HMW DON (Welch's two-sample t-test/Mann Whitney U-test, p < 0.001). At HOT, D/L-Ser was also significantly greater in LMW SPE-DON than HMW DON (Welch's two-sample t-test, p < 0.001). At BATS, D-Tyr and D-Lys were measured above blank values in almost all LMW SPE-DON samples; however, they were indistinguishable from blank values in multiple HMW samples (Fig. 2.4, Table S2.4). D-Tyr and D-Lys were not measured in previously published GC-MS analyses at HOT and cannot be observed by the HPLC method used here, thus are not reported in this ocean basin. Notably, while abiotic racemization of AA does occur over long time periods, these rates are too slow to account for the D/L ratios we observed (Bada, 1971). For example, Phe has the fast abiotic racemization rate of those published in Bada (1971), and one of the lowest D/L ratios in our data, yet in our oldest sample (Δ^{14}C = 6860 years), abiotic racemization would yield D/L ratio for Phe an order of magnitude less than what we measured (0.015 verses 0.15).

Within each size fraction, the average relative molar abundance of each AA

was also similar between HOT and BATS, with values for all AA at most depths within error at HOT and BATS (Fig. S2.2). However, relative molar abundance of individual AA was markedly different in HMW vs. LMW SPE-DON. At BATS, the average molar abundance of Ala and Ser was significantly higher in HMW DON than LMW SPE-DON (Welch's two-sample t-test, p <0.05), while Gly, Asx, Glx, Leu, Phe, and Tyr were higher in LMW SPE-DON than HMW DON (Welch's two-sample t-test, p < 0.05). At HOT, molar abundance of Ala, Ser, and Thr was significantly greater in HMW than LMW SPE-DON (Welch's two-sample t-test, p < 0.05), while Gly, Asx, Glx, Leu, Val, Phe, and Tyr were significantly greater in LMW SPE-DON than HMW DON (Welch's two-sample t-test/Mann Whitney U-test, p < 0.05).

D/L ratios and AA molar composition were also determined in the UF permeate, representing the total LMW DON pool. D/L ratios of the UF permeate were more variable with depth compared to D/L ratios of LMW SPE-DON or HMW DON, and the mean deviation of measurements made on two separate cruises were much larger (Fig. S2.2). At HOT, the depth averaged D/L ratio of Asx was lower in LMW SPE-DON compared to the UF permeate (Welch's two-sample t-test, p = 0.03). At BATS, depth averaged D/L ratios of Asx, Glu, and Ser were significantly lower in LMW SPE-DON compared to the UF permeate (Welch's two-sample t-test, p < 0.05). However, because of the low (natural abundance) concentrations of L and D-AA in the UF permeate, differences in D-AA concentrations equivalent to the limit of detection of this method would result in large differences in D/L ratios which would be considered statistically significant.

The overall AA composition of LMW SPE-DOM and the UF permeate was very similar (Fig. S2.3, Fig. S2.4). The difference in depth averaged molar abundance of individual AA ranged from 0.0% to 5.6%, with an average difference of $\pm$

1.4% across all samples at HOT and BATS. Significant differences in LMW SPE-DON versus permeate AA molar abundance were thus observed almost exclusively in AA with very low abundances. At HOT, the depth averaged molar abundance of Asx, Tyr, and Arg was significantly lower in LMW SPE-DOM, while Val, Phe, Lys, Ile, and Thr were significantly lower in the UF Permeate (Welch's two-sample t-test or Mann Whitney U-test, p < 0.05). At BATS, the depth averaged molar abundance of Tyr and Arg was significantly lower in LMW SPE-DOM while Leu, Lys, Ile, and Thr were significantly lower in the UF permeate (Welch's two-sample t-test or Mann Whitney U-test, p < 0.05). Still, while these differences are considered statistically significant within the analytical variation observed, we feel it is hard to put much weight on differences of $\sim$ 5% or less abundance. For example, concentration differences equivalent to the limit of detection would result in AA molar abundance differences between 0.65% to 1%, only slightly lower than the average difference of 1.4% we observed between LMW SPE-DON and the UF permeate.

2.4.5 Relative molar abundance vs. D/L ratio

The average relative molar abundance and average D/L ratio of each individual AA were significantly linearly correlated in HMW and LMW SPE-DON at HOT and BATS (r^2 = 0.99 – 0.88, p < 0.0001 – 0.0017) (Fig. 2.5). In both HMW and LMW SPE-DON, there was no statistically significant difference in slope (ANCOVA: HMW p = 0.39 LMW p = 0.71) or y-intercept (ANCOVA: HMW p = 0.36, LMW p = 0.45) of the linear regressions between HOT vs. BATS. The slope of the linear regression of all HMW DON samples was not statistically different than the slope of all LMW SPE-DON samples (ANCOVA, p = 0.15). However, the y-intercept of the HMW DON linear regression was significantly

25

lower than the y-intercept of the LMW SPE-DON linear regression (ANCOVA, p < 0.001).

2.4.6   Degradation state and reactivity of HMW and LMW SPE-DON

Within each ocean basin, most AA-based proxies were significantly different between HMW and LMW SPE-DON (Fig. 2.6, Table S2.3). However, individual parameters varied in their predictions for HMW vs. LMW degradation state. Due to the opposite trend of D-Ala compared to all other D-AA (Section 2.4.4), total %D was calculated without D-Ala, which we refer to as total %D_{NA}. Total %D_{NA} was greater in LMW SPE-DON than HMW DON at almost all depths in both ocean basins. The combined relative mol% of NPAA β-Ala and γ-Aba was also greater in LMW SPE-DON than HMW DON throughout the water column at HOT and BATS. The organic carbon normalized total hydrolysable AA yield (%C-AA) was significantly higher in HMW DON than LMW SPE-DON at all depths. Collectively, these proxies all indicated LMW SPE-DON is less labile and more degraded than HMW DON, resulting in a higher relative proportion of bacterially produced biomolecules.

In contrast, %D-Ala was greater in HMW than LMW SPE-DON all depths (Fig. 2.6), indicating a greater proportion of D-Ala containing bacterial molecules in this size fraction. Additionally, no significant differences in DI were observed between HMW and LMW SPE-DON in either ocean basin. Both size fractions had highest DI values (indicating least degradation) at the surface, though trends with depth were variable. ΣV values were significantly higher (indicating more resynthesis) in HMW DON than LMW SPE-DON throughout the water column at HOT and BATS. Together, these parameters indicated comparable or less bacterial resynthesis and degradation to LMW DOM.

Overall, within each size fraction, most AA-based proxies were very similar between ocean basins (Fig. 2.6). The only exceptions include %C-AA and the ΣV parameter in the HMW DON size fraction. %C-AA of HMW DON was significantly greater at HOT than BATS throughout the water column. HMW DON ΣV was greater (indicating more resynthesis) at HOT than BATS at all depths below the surface.

Principal component analysis of AA-based proxies and D/L-ratios was used to visualize compositional and degradation state differences between samples. The first principal component (PC) explained 73.9% of the variance and the second PC explained 13.2% of the variance. PC1 had the largest positive contributions from D/L-Val, Mol% Gly, %NPAA, D/L-Leu, D/L Asx and total $\%D_{NA}$ and largest negative contributions from %C-AA, L-AA yield, and D-AA yield (Fig. 2.7A). The contribution of each of these indices to PC1 was $\geq 8\%$, collectively accounting for 78% of the variance along PC1. PC2 had the largest positive contributions from D/L-Ala and D/L-Ser and largest negative contribution from DI. Together, these indices contributed 83% of the variance along PC2. HMW and LMW SPE-DON were clearly separated along PC1 (Fig. 2.7B). Surface and deep HMW DON were separated along PC2, while surface and deep LMW SPE-DON grouped together. There was no clear separation of HMW or LMW SPE-DON samples from HOT vs. BATS.

Degradation proxies which could be calculated from HPLC data alone (total $\%D_{NA}$ only including D-Asx, D-Glx, and D-Ser, %NPAA, Mol% Gly, %C-AA,%D-Ala, and DI) were also calculated for the UF permeate (Fig. S2.5). Overall, the magnitude and depth trends of most degradation proxies were similar between LMW SPE-DON and the UF permeate. When comparing depth averaged degradation proxies within each size fraction, at HOT, only %D-Ala was significantly

greater in the permeate compared to LMW SPE-DON (Mann Whitney U-test, p = 0.015). At BATS, DI was significantly lower in the permeate compared to LMW SPE-DON, while total $\%D_{NA}$ and %C-AA were significantly greater in the permeate (Welch's two-sample t-test, p < 0 .05).

2.4.7 Relationship between degradation indices, %D, and $\Delta^{14}C$

To investigate long-term changes in AA composition and reactivity to the DON pool with age, degradation indices were plotted against $\Delta^{14}C$ values of these same samples reported previously (Broek et al., 2017, 2020) (Fig. 2.8). In HMW DON, most measures of degradation were significantly correlated with $\Delta^{14}C$. Total $\%D_{NA}$, %NPAA, and mol% Gly were significantly negatively correlated with $\Delta^{14}C$, indicating an increase in these biomolecules with radiocarbon age (p < 0.001). In contrast, DI and%C-AA were significantly positively correlated with radiocarbon age, indicating an increase in degradation as recorded by DI and a decrease in total AA yield with radiocarbon age (p = 0.0028 & p = 0.0268, respectively). However, all these relationships all appear driven by the large offset between surface and subsurface ($\leq$ 400m) samples; if the surface samples are not included, only the relationship between $\Delta^{14}C$ and mol% glycine is still significant (Fig. S2.7).

In contrast, in the LMW SPE-DON size fraction, there are fewer significant correlations between degradation indices and $\Delta^{14}C$ (Fig. 2.8). $\Delta^{14}C$ was significantly positively correlated only with%C-AA and the D-AA yield (p = 0.0022 & p = 0.0144, respectively), indicating a decrease in D-AA and total AA yield with radiocarbon age. Mol% Gly was significantly negatively correlated with $\Delta^{14}C$, indicating an increase in Glycine with radiocarbon age (p = 0.0079). %NPAA was significantly positively correlated in only the deep ($\geq$400 m) samples (Fig. S2.3).

2.5 Discussion

2.5.1 Total hydrolysable amino acid concentrations at HOT and BATS

The L- and D-AA concentrations of total hydrolysable amino acids reported in this study (Fig. 2.1) are comparable to those reported previously at BATS and HOT (Kaiser & Benner, 2008). Earlier reports of L-AA concentrations at both BATS versus HOT are limited and suggest more variable concentrations (Lee & Bada, 1977; McCarthy et al., 1996). At the surface, the higher hydrolysable AA concentrations at HOT reported here and by Kaiser and Benner (2008) suggest either greater surface AA production or greater resistance of labile biochemicals compared to surface waters at BATS. While our data cannot distinguish exact mechanisms for these offsets, given that HOT is more oligotrophic than BATS, we suggest that more extensive microbial loop processes at this site may underlie surface accumulation of more refractory proteinaceous and AA-containing molecules. DOM in the subsurface at HOT is older than at BATS and may also be less degradable, which could lead to the observed greater concentrations of biochemicals at HOT. At 2500 m depth, AA concentrations are similar or greater at BATS than HOT (see Supplementary 2.7.1), a trend also documented by Kaiser and Benner (2008) and consistent with more extensive microbial removal of labile biomolecules in the older Pacific mesopelagic ocean waters.

2.5.2 D-AA and bacterial source of DON at HOT and BATS

We report here an expanded suite of D-AA in HMW and LMW SPE-DON. Only four D-AA (D-Ala, D-Asx, D-Glx, and D-Ser) have been commonly reported in almost all past marine DOM literature (Kaiser & Benner, 2008; McCarthy et al., 1998). Two additional D-AA, D-Leu and D-Val, have been reported in coastal and terrigenous DOM, however at near-blank levels (Bourgoin & Tremblay, 2010;

29

Hébert & Tremblay, 2017; Tremblay & Benner, 2006). Recently, D-Leu, D-Val, and D-Phe were confirmed for the first time in open ocean DOM based on large samples coupled with mass spectral verification (Broek et al., 2019). This study expands observations of these new D-AA to the Atlantic Ocean (BATS site) and further identifies two additional D-AA (D-Lys and D-Tyr), which to our knowledge have never been reported in any natural water. We show that D-AA within this greatly expanded suite of bacterial tracers fall into distinct groupings based on abundance and relative concentrations changes in HMW and LMW material.

All recently identified D-AA (D-Leu, D-Val, D-Phe, D-Tyr, D-Lys) are more abundant in LMW SPE-DON than HMW DON at both HOT and BATS. The significantly greater D/L ratios of D-Leu, D-Val and D-Phe in LMW SPE-DON compared to HMW DON in both ocean basin (Section 2.4.4, Fig. 2.4) supports previous work which hypothesized that these D-AA may be tracers specific for refractory bacterial AA containing-molecules (Broek et al., 2019). Similarly, PCA analysis indicated that of the seven most abundant D-AA measured in every sample (excluding D-Tyr and D-Lys), D/L-Leu and D/L-Val had the greatest positive contributions to PC1, associated with the LMW SPE-DON fraction (Fig. 2.7). Although D-Tyr and D-Lys could not be confirmed above blank values in all our HMW DOM samples, the D/L ratios of these two D-AA in LMW SPE-DON had very similar depth structure to the other newly identified D-AA (Fig. 2.4). While mass spectral data confirms authentic compounds, this observation provides confidence that these D/L ratios also follow similar oceanographic trends and suggests that D-Tyr and D-Lys may have similar sources as other newly identified D-AA. Past work has shown that heterotrophic bacteria can produce all the D-AA we report in this study, as well as others (Azúa et al., 2014; Cava et al., 2011; Lam et al., 2009; Zhang et al., 2016), suggesting bacteria are the most

likely source for all the D-AA we observed.

In contrast to all other D-AA, D-Ala was the only D-AA with a greater D/L ratio in HMW than in LMW SPE-DON, consistent with previous work (Broek et al., 2019; Kaiser & Benner, 2008) (Fig. 2.3, Fig. 2.4, Table S2.4). Similarly, our PCA indicates D/L-Ala is the only D/L ratio associated with HMW DON compared to LMW SPE-DON (Fig. 2.7). D-Ala is a key component of both autotrophic and heterotrophic bacterial peptidoglycan (Cava et al., 2011; Kaiser & Benner, 2008; Schleifer & Kandler, 1972), a HMW polymer which is hydrolyzed relatively rapidly in seawater (Jørgensen et al., 2003; Nagata et al., 2003). Our data is therefore consistent with past work suggesting that D-Ala could mostly trace bacterial peptidoglycan (Broek et al., 2019). However, we note that D-Ala is also found in other bacterial compounds (Kaiser & Benner, 2008), meaning other D-Ala containing molecules cannot be ruled out. Overall, because of the opposing trends observed for D-Ala vs. all other D-AA in our MW fractions, we suggest %D-Ala should be reported and considered separately in all DON data. In the text below, we therefore exclude D-Ala in discussion of the percentage of total D-AA (%D$_{NA}$).

While the AA composition of HMW DON and LMW SPE-DON were clearly distinct (Fig. 2.4, Fig. S2.4), the AA composition in both fractions was remarkably similar between BATS and HOT. The nearly identical depth profiles of D/L ratios and relative mol% contributions at the two stations within each size fraction (Fig. 2.4, Fig. S2.4) suggest that very similar microbial sources and removal processes define the composition of AA-containing material when comparing between similar depths throughout both ocean basins. This conclusion is also supported by the strong, apparently universal linear relationships between average D/L ratios and molar abundance in both HMW and LMW SPE-DON (r^2 0.88 to 0.99,

p < 1x10^{-4} to 0.0017) which are not statistically different between the two ocean basins (Fig. 2.5) (Broek et al., 2019). These relationships directly show that in both size fractions the most abundant chiral AA were also those with the highest bacterial contributions (as indicated by D/L ratio).

Together, these observations strongly suggest that bacterially produced AA dominate the entire dissolved AA pool throughout the ocean. It further suggests that at both BATS and HOT, similar bacterial processes have utilized and resynthesized most eukaryotic algal material in both HMW and LMW SPE-DON, while refractory D-AA containing molecules are selectively preserved. These processes result in incredibly similar D and L-AA composition when comparing between the two sites, indicating much of the AA-containing molecules in the subsurface of both size fractions may be refractory. Yet, at the same time, the composition of bacterial products appears to be clearly distinct between HMW and LMW SPE-DON pools. Overall, this suggests a difference between bacterial production of HMW versus LMW SPE-DON material, potentially indicating unique bacterial mechanisms produce long lived HMW and LMW SPE-DON.

2.5.3 Potential diversity of D-AA containing molecules in HMW vs. LMW SPE-DON

In contrast to the individual D/L-AA ratios, the total D and L-AA yields (μmol D or L-AA normalized to mg total organic nitrogen), provide information regarding the overall composition of DON (Table S2.1). Specifically, the L-AA yield is often used as an indicator of reactivity, with a higher L-AA yield indicating a greater proportion of the total ON is made of supposedly labile L-AA (Davis et al., 2009; Kaiser & Benner, 2009). In contrast, D-AA yields represent the relative proportion of total DON which is bacterially derived, with a higher D-AA yield

32

indicating greater bacterial contribution (Bourgoin & Tremblay, 2010; Kaiser & Benner, 2008; Tremblay & Benner, 2006, 2009) While the D-AA yield of individual AA in total DOM are often used to calculate the percent of bacterial DOC or DON, the lack of bacterial biomarker endmembers for HMW and LMW DON specifically mean this calculation is not appliable to our individually isolated size fractions (Supplementary 2.7.4). Still, a comparison of D and L-AA yield depth trends is informative about the reactivity of AA-containing molecules in both size fractions.

One major observation from the D-AA and L-AA yields is that D-AA present in HMW and LMW SPE-DON appear to not only have different compositions, but also very different reactivities in the different size fractions. An initial bacterial degradation study suggested that the four common D-AA (D-Ala, D-Glx, D-Ser, and D-Asx) all have similar reactivity to total bacterial DON (Kawasaki & Benner, 2006). However, bacterial incubations including some newly identified D-AA indicated greater bioavailability of "canonical" D-AA (D-Ala, D-Asx, D-Glx) vs. "non-canonical" D-AA (D-Leu, D-Met, D-Val) (Wang et al., 2020). Additionally, recent work investigating marine DOM D/L ratios with radiocarbon age also indicated varying reactivity of compounds containing D-Ala, D-Glx, D-Ser, and D-Asx vs. D-Leu, D-Val, and D-Phe (Broek et al., 2019). The difference in D-AA yield depth trends in HMW versus LMW SPE-DON reported here confirm and expand this idea (Fig. 2.2), implying a difference in the lability, and likely in molecular composition, of D-AA containing compounds in HMW versus LMW SPE-DON.

In HMW DON, the consistent or increasing D-AA yield in the upper ocean suggests D-AA containing compounds are more refractory than total HWW DON, while decreasing L-AA yields indicate L-AA containing compounds are preferen-

tially removed. This is further supported by the greater D-AA yield at HOT compared to BATS, which indicates greater persistence of D-AA containing compounds compared to total ON in the older waters at HOT. The increase in D/L ratios with depth of some AA in HMW DON (Ala, Asx, Ser, Val, Fig. 2.4) are also consistent with these observations, suggesting removal of L-AA compared to D-AA. These results are consistent with the general expectation that D-AA containing biomolecules are less labile than their L-AA counterparts.

In contrast to the HMW DON pool, the decrease in L and D-AA yields with depth in the LMW SPE-DON pool indicate both L and D-AA containing compounds are utilized preferentially compared to total LMW SPE-DON (Fig. 2.2). At the same time, the maximum D/L ratios observed at 400 m for nearly every LMW D-AA in both ocean basins (Fig. 2.4) imply a relative accumulation of LMW bacterially sourced molecules in the mesopelagic. Based on similar observations at HOT, Broek et al. (2019) suggested an input of fresh, heterotrophic bacterial material to the mesopelagic ocean. However, the decrease in D and L-AA yields we report here suggests net removal of both enantiomers throughout the mesopelagic, potentially indicating the D/L ratio maxima are due to enhanced removal of L-AA. Still, total AA yields can represent multiple addition and removal processes, meaning it is difficult to tell if addition of D-AA, enhanced removal of L-AA, or both result in the D/L ratio depth profiles observed here. Regardless, it is clear that unique processes are shaping the D-AA composition of LMW SPE-DON and HMW DON.

Because free AA are a very minor portion of total dissolved AA (Lee & Bada, 1977; McCarthy & Bronk, 2008), measurable changes in L-AA and D-AA yields reported here must be interpreted as either production or microbial degradation of L-AA and D-AA containing macromolecules. Thus, while AA are often interpreted

as indicating "proteinaceous" material, we hypothesize that the differences in relative reactivity of HMW and LMW SPE-DON D-AA we observe indicate different dominant D-AA containing compound classes in these two size fractions. For example, we would expect proteinaceous material (larger peptides and partially degraded protein fragments) and larger fragments of structural compounds (such as peptidoglycan), to be retained in our HMW size fraction (> 2.5 kDa). In contrast, compounds isolated in our LMW SPE-DON size fraction are more likely to either be less-labile, smaller, AA-containing natural products or else fragments of larger biomolecules (discussed further in Section 2.5.4). Together with the results above indicating differing reactivities to D-AA containing compounds in HMW DON and LMW SPE-DON, this suggests AA-containing compounds in HMW vs. LMW SPE-DON may be distinct, and many may in fact not be "proteinaceous." For example, lipopeptides, siderophores, pigments, and bacterial signaling molecules can all have both D and L-AA (Asano & Lübbehüsen, 2000; Cava et al., 2011; Kaiser & Benner, 2008; Radkov & Moe, 2014; Schleifer & Kandler, 1972). Further, the role of many D-AA produced by bacteria remains unknown, meaning there are also likely multiple additional D-AA containing molecule classes which remain to be characterized (Radkov & Moe, 2014).

2.5.4 Degradation signatures of HMW and LMW SPE-DON: Bacterial source or progressive degradation?

Progressive bacterial degradation of HMW DON

In the HMW DON pool, the very similar AA composition and predicted degradation state at HOT and BATS was inconsistent with expectations that the oldest deep waters at HOT would have the most "degraded" DON, as reflected in altered molecular composition (Fig. 2.6). Only the ΣV proxy demonstrates a significant

oceanographic offset, indicating more bacterial resynthesis in the deep Pacific than deep Atlantic (Fig. 2.6). Together with the similar AA composition between ocean basins discussed above (Section 4.2), most of these data suggest that essentially all bacterial processes occurred rapidly in the upper ocean, rather than continued bacterial alterations to HMW DON with ocean circulation. To further explore this idea, we investigated relationships between AA-based proxies and Δ^{14}C. If bacterial alteration of HMW DON with ocean circulation is progressive, as opposed to simply occurring in the biologically most active surface zone, then we would also expect degradation parameters to correlate with Δ^{14}C. We note that while total DOM Δ^{14}C likely does not exactly represent the Δ^{14}C of DON, Δ^{14}C of the proteinaceous fraction suggests it is at least a reasonable proxy (Loh et al., 2004).

The strong relationships observed between almost all AA-based proxies and Δ^{14}C for HMW DON indicate bacterial degradation is progressively changing the surface produced HMW DON with radiocarbon age (Fig. 2.8). Only the D-AA yield (in contrast to $\%D_{NA}$) is not significantly correlated with Δ^{14}C, which suggests changes to AA composition in this size fraction are due to net removal of L-AA rather than addition of D-AA. These changes in all bacterial and degradation proxies with Δ^{14}C imply that bacteria selectively remove labile L-AA from HMW DON, while changing the relative mol% contribution of AA (DI) and increasing the relative proportion of refractory D-AA, Gly, and the NPAA in HMW DON. We note that while a ΣV vs. Δ^{14}C correlation is missing from this analysis due to lack of sufficient data, based on the trends with depth in Fig. 2.6 we may expect a weak relationship.

Overall, these data suggests HMW DON is predominantly bacterially sourced, even in the surface ocean, consistent with past compound specific δ^{15}N -AA data

36

(Yamaguchi & McCarthy, 2018). However, AA composition and preservation appears to then be progressively altered via microbial degradation, consistent with expectations of a semi-labile HMW pool. Alternatively, protozoan grazing of heterotrophic bacteria is an additional, not mutually exclusive, explanation. Because most eukaryotes do not have enzymes required to digest D-AA (Asano & Lübbehüsen, 2000), protozoan heterotrophy of bacterial biomass might be expected to selectively remove labile L-AA containing molecules, leaving behind a HMW DON pool enriched by more refractory D-AA, Gly, and NPAA. Regardless, these data indicate that the main production mechanism of HMW DON is fresh bacterial production followed by progressive degradation and resynthesis, explaining together the changes we observe.

Finally, one caveat to the observation of progressive bacterial degradation is that if we consider only data for samples between 400 m and 2500 m, most relationships between degradation proxies and Δ^{14}C in HMW DON were no longer significant (Fig. S2.6). Additionally, PCA analysis of D/L ratios and degradation proxies indicate surface HMW DON has a distinct composition from all subsurface ($\geq$ 400 m) samples (Fig. 2.7). Considering the expected relative lability of proteinaceous HMW DON, typically considered the semi-labile component of marine DON (Amon et al., 2001; Amon & Benner, 1996), the lack of changes to amino acid composition with further age beyond the mesopelagic (representing a radiocarbon age difference >2000 between mesopelagic and deep waters at HOT) might be seen as surprising. Instead, it appears that by the time advected HMW DON reaches the deep ocean, AA-containing molecules remaining are relatively stable and undergo little measurable compositional change with time. Only the ΣV offset in deep waters between the ocean basins would seem in contrast to this possibility, though this offset is only based on a few data points. Still, this could

potentially suggest that if continuing slow bacterial remineralization is the main mechanism for this older HMW DON, CSI-AA patterns alone continue to reflect this process in very old DOM.

Overall, for HMW DON, our results are generally consistent with a "two-pool" model; reactive semi-labile HMW DON appears restricted to the upper water column, while an additional "background" HMW DON pool appears to be relatively refractory and unchanging. This is consistent with radiocarbon measurements of "protein-like" HMW DON, estimating this material to be 3,000-4,000 years old (Loh et al., 2004). Additionally, a similar concept was proposed based on solid-state NMR results, which suggested HMW DON is composed of two chemically distinct pools with varying reactivities (Aluwihare et al., 2005). If almost all HMW DON is in fact proteinaceous material (Aluwihare et al., 2005; McCarthy et al., 1997), these results would further imply bacterial alteration of proteinaceous material produces stable compounds which can persist for thousands of years. Indeed, the observation that most AA-based degradation proxies behave in similar, expected ways in the HMW DON fraction may itself be further evidence for a predominantly proteinaceous N pool in HMW material.

Direct bacterial source of LMW SPE-DON

The discrepancy between degradation state of LMW SPE-DON as predicted by different degradation parameters suggests a disconnect in how individual parameters reflect "degradation" in this size fraction (Fig. 2.6). As noted above (Section 2.4.6), most parameters linked to D or L-AA yields indicated increased degradation or bacterial influence in the LMW SPE fraction compared to the HMW fraction at all depths (%D_{NA}, %NPAA, mol% Gly, and %C-AA), corresponding with expectations based on older average radiocarbon ages for this size

fraction (Section 2.3.8). However, the observations that the δ^{15}N-based ΣV parameter indicated less bacterial resynthesis in LMW SPE-DON at all depths (with values close to autotrophic algae), along with substantially lower %D-Ala and similar degradation (DI) to HMW DON, were unexpected (Fig. 2.6). We suggest a few possible interpretations consistent with the data reported here, which are not mutually exclusive.

One is the potentially different composition of LMW AA-containing molecules hypothesized above (Section 2.5.3). If substantial LMW nitrogenous material is in fact bacterial, but not proteinaceous, then the "baseline" levels of many degradation parameters might be different in comparison with proteinaceous N which dominates HMW material. This possibility is supported by a lack of significant relationships between %D-Ala and total %D_{NA} versus DI in the LMW SPE-DON size fraction, which suggests DI does not record bacterially mediated changes in this size fraction (Supplementary 2.7.2, Fig. S2.7). Similarly, if LMW AA containing molecules have a distinct nitrogenous composition from those in HMW, then it is at least plausible ΣV may not record the same changes in LMW SPE-DON. Finally, unique biomolecular sources of D-AA containing compounds in LMW SPE-DON would also be consistent with previous work reporting D-AA content in the LMW DON pool (Broek et al., 2019; Kaiser & Benner, 2008).

A second possibility is that patterns of D-AA, NPAA, and most AA yields in LMW SPE-DON reflect well preserved bacterial source signatures, irrespective of composition, which are created in the surface ocean and then undergo little subsequent bacterial alteration on millennial timescales. Production of refractory, LMW DOM can occur via direct bacterial release during growth or viral lysis (Gruber et al., 2006; Ogawa et al., 2001), consistent with LMW D-AA representing a "fresh" bacterial source signal. In this scenario, bacterial production would

produce LMW molecules with inherently lower L-AA yields and greater contribution of bacterially derived molecules (D-AA and NPAA) compared to HMW molecules. This interpretation could explain the low ΣV values, which are close to ranges for autotrophic algae (McCarthy et al., 2007), but also to de-novo AA synthesis in heterotrophic bacteria (Yamaguchi et al., 2017). The limited relationships between $\Delta^{14}C$ values and bacterial proxies in LMW SPE-DON (Fig. 2.8) further support this idea, and could imply that at least D-AA, DI and NPAA are a predominantly a surface bacterial source signature. Indeed, the significant decrease in D-AA yield with $\Delta^{14}C$ (Fig. 2.8f) is the opposite of what would be expected if relative D-AA contribution in LMW SPE-DON were increasing with age. Similarly, PCA analysis shows no separation of LMW SPE-DON samples by depth or ocean basin, indicating all LMW SPE-DON samples investigated here have a similar composition and degradation state (Fig. 2.7). Together, these data support that idea that D-AA in LMW SPE-DON could be predominantly an indicator of refractory bacterial surface sources, with no real impact from further progressive heterotrophic degradation or resynthesis.

We note, however, that the depth and sampling limitations inherent in these data set does not tightly constrain the timescale of surface-produced bacterial sources. Even the youngest HMW material isolated from the surface has average $\Delta^{14}C$ ages of decades to hundreds of years, while the LMW material is far older, by mass balance representing a mixture of newly produced LMW SPE-DON and refractory background material. Our data set therefore suggests that in contrast to HMW DON, no major compositional changes occur in LMW SPE-DON over timescales of ocean mixing, however any alteration of LMW AA containing molecules on decadal scales would appear "preformed" in our sample set. Previous work has demonstrated that bacteria can produce complex, RDOM on a

timescale of days to weeks, the majority of which is uncharacterizable (Gruber et al., 2006; Ogawa et al., 2001). If production and subsequent degradation of bacterial biomolecules in the ocean were similar, such changes could never be observed over the long timescales our sample comparisons here can address.

Taken as a whole, the data presented here indicates LMW AA-containing molecules, at least those isolated via SPE, may be distinct from those in HMW DON, representing compound classes other than proteinaceous material. This could, at least in part, explain the discrepancies we report between different AA-based "degradation" proxies in LMW SPE-DON. However, an additional explanation is if LMW AA-containing molecules are relatively rapidly produced in the surface (compared to timescales of ocean mixing), such that bacterial biomolecules represent a source signature in LMW SPE-DON. If true, this would suggest a departure from the current assumption that progressive bacterial degradation produces refractory, LMW DON molecules, and instead suggests LMW AA-containing molecules represent a surface, prokaryotic source completely independent of HMW DON production and degradation. We suggest future work on LMW DON should test this hypothesis. Specifically, compound specific isotope analysis may have potential to tease apart the mechanisms responsible for bacterial changes to the HMW and LMW DON pools (McCarthy et al., 2007; Yamaguchi & McCarthy, 2018).

Finally, while these data are specific to the LMW SPE-DON isolated in this study, it is unclear if these conclusions apply to the total LMW DON pool. The overall similarity of the AA composition (Fig. S2.4) and degradation state (Fig. S2.5) of the UF permeate and LMW SPE-DON suggests our LMW SPE-DON samples are fairly representative of AA in bulk LMW DON. However, notable differences in D/L ratios of some AA in LMW SPE-DON versus the UF permeate

may indicate distinctions in AA composition between the two (Fig. S2.3). Still, as noted in Section 2.4.4, differences in AA concentration at the limit of the detection for the HPLC method used here would result in significant differences in D/L ratios and AA molar abundance for natural abundance samples (UF permeate and total DOM). This highlights a benefit of the large-scale isolation method applied here, allowing analysis of much more concentrated samples. Additionally, due to reasons discussed in Sections 2.4.2 and 2.4.5, it is possible the concentrations of the UF permeate we collected could differ somewhat compared to the total permeate from ultrafiltration. Nevertheless, while we cannot confidently say whether these results apply to the total LMW DON pool, the data presented here suggests it is plausible. Regardless, while our HMW and LMW SPE-DON samples only represent a subset of the total DON pool, they still make up the greatest proportion of total DON and total dissolved AA every directly characterized to date.

2.5.5   Two independent pools of DON

Taken together, a comparison of D/L ratios, relative AA mol% data, and degradation state proxies for HMW and LMW SPE-DON shows clear compositional differences between these two size fractions (Fig. 2.3, Fig. 2.6, Fig. S2.4). PCA analyses further indicated that HMW and LMW SPE-DON are compositionally distinct (Fig. 2.7). Still, at the same time, within each size fraction, AA composition was remarkably similar between the Atlantic and Pacific Oceans (Fig. 2.3, Fig. 2.6, Fig. S2.4). These results appear to contradict expectations based on a common size-age-reactivity continuum theory which has been widely applied to the DOC pool (Amon & Benner, 1994, 1996; Walker et al., 2014, 2016). Specifically, if microbial activity were progressively degrading HMW DON to LMW DON, we would expect to see signatures of continual bacterial degradation link-

ing both pools. While our LMW SPE-DON only represents a subset of the total DON pool, comparisons of the AA composition and degradation of total LMW DON (via the UF permeate) and LMW SPE-DON do not indicate drastic differences. Thus, while molecular size is an underlying operational parameter in these data set, rather than something we explicitly set out to test, we expected to see an approximate continuum of degradation and D-AA composition with radiocarbon ages between the HMW and LMW SPE-DON pools. Instead, despite the large overlap in age of our isolated size fractions (Section 2.3.8), all parameters and D-AA composition were distinct between the two size fractions at every depth throughout the water column in both ocean basins. These differences suggest AA-containing molecules in HMW and LMW SPE-DON cycle independently.

While these results are contradictory to expectations based on size-age-reactivity data of DOC, they are consistent with studies specific to the DON pool. At HOT, recent work reported individual AA D/L ratios were not correlated throughout the HMW and LMW SPE-DON pools with radiocarbon age, suggesting a disconnect between D-AA composition of HMW and LMW SPE-DON (Broek et al., 2019). Similarly, Knapp et al. (2012) also suggested that nitrogen isotope signature of HMW and LMW SPE-DON cannot be explained by formation of LMW SPE-DON via degradation of HMW DON. Instead, they suggested direct formation of LMW SPE-DON from the degradation of PON. This would be consistent with our hypothesis that most LMW AA containing molecules may be directly released from bacteria and undergo little further heterotrophic bacterial resynthesis, in direct contrast to HMW DON. If true, together this implies that completely independent mechanisms exist for the formation of HMW and LMW bacterial DON, which are largely decoupled from the dominant formation processes hypothesized for DOC.

2.6 Conclusions

We report here an expanded suite of D-AA measured directly in both HMW and LMW SPE-DON at BATS. These data are interpreted in the context of existing data at HOT and paired with DOM radiocarbon ages and multiple bacterial AA-based proxies. Our new data confirms recent observations that three novel D-AA are ubiquitous in marine DON (D-Leu, D-Val, and D-Phe). Additionally, we report two additional D-AA never previously reported in any natural water (D-Tyr and D-Lys). D/L ratios of all five D-AA are significantly greater in the older, LMW SPE-DON size fraction, suggesting they may represent new tracers for bacterially produced refractory DON. The clearly distinct D-AA compositions within HMW and LMW SPE-DON pools at all ocean depths suggests AA containing molecules in DON may be more diverse than previously believed, with distinct sources and reactivities which are characterized at least in part by molecular weight. We hypothesize these differences are indicative of unique D-AA containing nitrogenous molecules in the marine DON pool, with AA-containing molecules in HMW DON that are dominated by proteinaceous material while AA-containing molecules in LMW SPE-DON represent compounds other than peptides.

By then coupling D-AA composition with multiple measures of bacterial source, degradation and radiocarbon age, we evaluated how different measures of DON reactivity and degradation change over time in the HMW and LMW SPE-DON pools. All AA-based proxies indicated heterotrophic bacterial alteration of HMW DON throughout the mesopelagic, supporting the paradigm of a surface produced, semi-labile HMW DON component which is then progressively altered by bacteria. In contrast, in the LMW SPE-DON pool, AA-based proxies yielded conflicting results regarding degradation state, with some indicating LMW SPE-DON is more refractory and degraded than HMW DON, while others indicated LM SPE-DON

is less degraded. Additionally, few relationships were observed between Δ^{14}C and degradation proxies. We hypothesize that together these data indicate that bacterially derived LMW AA-containing molecules are intrinsically more refractory, and therefore less susceptible to transformation or degradation than HMW bacterial AA-containing molecules. While these results are most directly appliable to the LMW SPE-DON isolated in this study, together our HMW and LMW SPE-DON size fractions represent the greatest proportion of the total DON pool every directly characterized. Additionally, measurements of total LMW DON (the UF permeate) indicate similar AA composition and degradation state of LMW SPE-DON and total LMW DON, suggesting that these conclusions may apply to the total LMW DON pool as well.

Taken together, our D-AA, Δ^{14}C, degradation proxies, and AA molar abundance data suggest that the DON cycling may be fundamentally different than that of most DOC. Specifically, the distinct signatures of HMW and LMW SPE-DON, coupled with consistent AA composition and degradation state parameters in the deep, older background pool of both size fraction suggests completely separate pools of material. This interpretation runs counter to expectations that the AA composition of HMW DON would progressively shift towards that of LMW DON, as might be expected in a size-age-reactivity continuum model. Instead, our results imply independent composition, sources, and cycling of AA-containing molecules in HMW and LMW SPE-DON. These results suggest a potential paradigm shift for DON cycling, suggesting the most refractory, AA-containing compounds in LMW DON may not result from progressive bacterial breakdown of HMW DON into LMW components. Our data also supports a growing body of work suggesting that amide functionality in marine DON may not be limited to proteinaceous material, but likely encompass a diverse range of

other N-containing molecules. Future work identifying the composition of non-proteinaceous nitrogenous material will aid in determining potential sources and degradation processes of refractory organic nitrogen.

Main text figures and tables
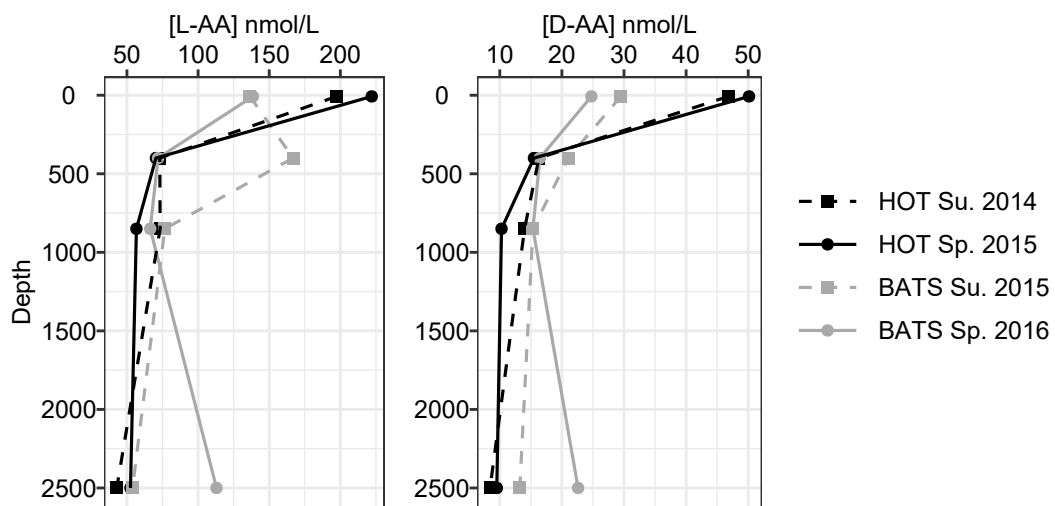
Figure 2.1: Total dissolved hydrolysable D-AA and L-AA concentration (nmol/L) at HOT (black) and BATS (grey) measured in summer ("Su.," dashed squares) and spring ("Sp.," solid circles) cruises.
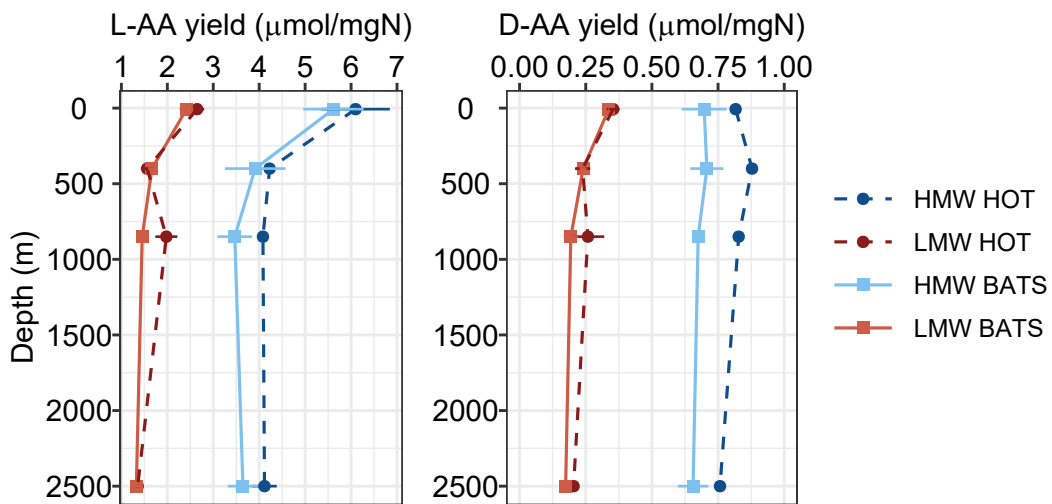
Figure 2.2: D-AA and L-AA yields normalized to mg total organic nitrogen (µmol/mgN) in HMW (blue) and LMW (red) DON at HOT (circles, dashed line) and BATS (squares, solid line). Error bars represent the mean deviation of spring and summer cruise data and are smaller than symbol where not visible.
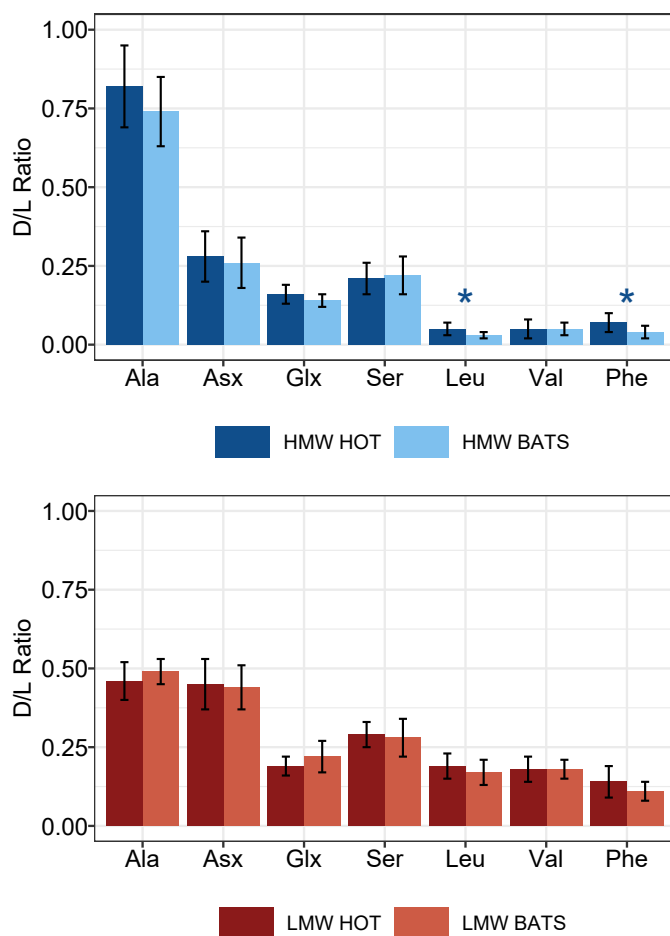
Figure 2.3: Average D/L ratios of seven D-AA measured at HOT and BATS. In the HMW DON pool, D/L ratios of D-Leu and D-Phe are significantly greater at HOT than BATS (denoted by an asterisk). In the LMW DON pool, there are no statistically significant differences in individual AA D/L between HOT and BATS. Error bars represent the standard deviation of all depths and spring and summer cruise data ($n = 8$).
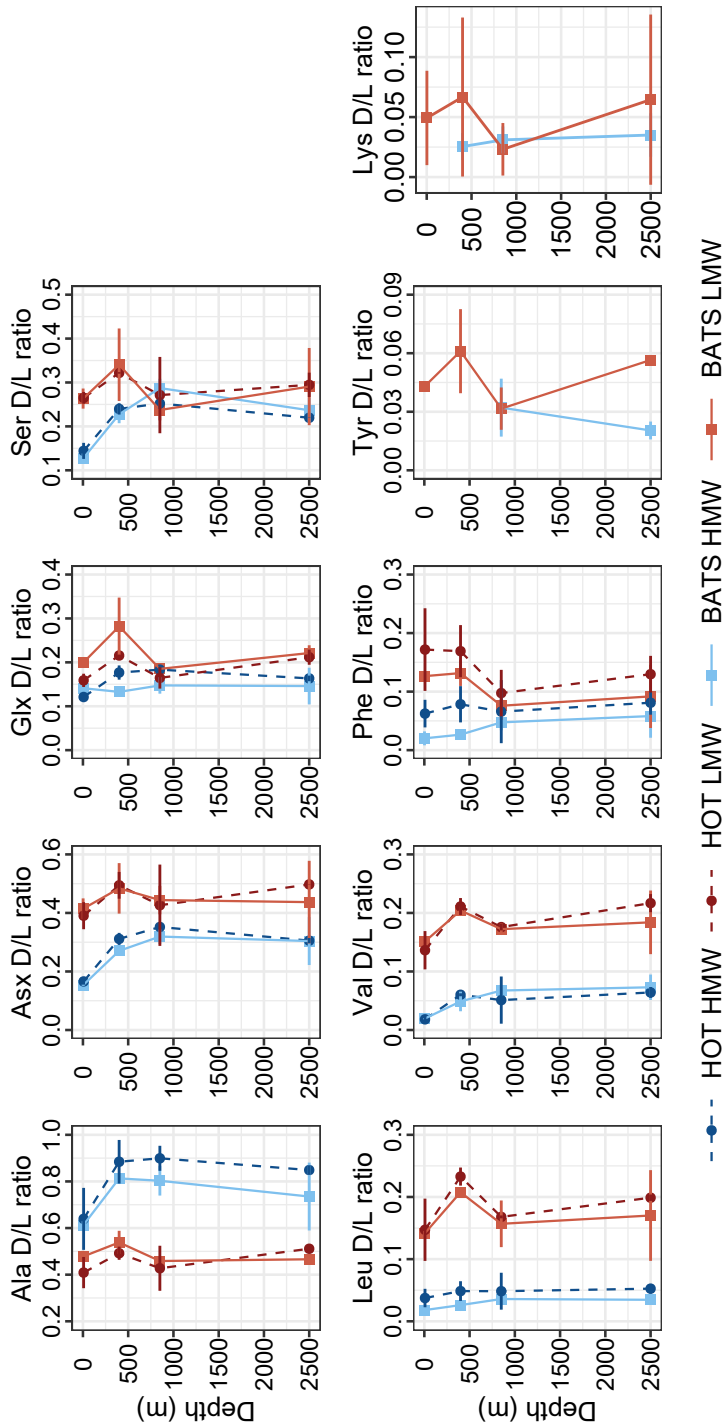
Figure 2.4: D/L ratios of four commonly reported (top row) and five newly confirmed (bottom row) D-AA. Depth profiles of all D/L ratios were similar between HOT (circles, dashed line) and BATS (squares, solid line) for both HMW (blue) and LMW DON (red). In the LMW DON size fraction, the maximum D/L ratio is almost always observed at 400m depth. Error bars represent the mean deviation of summer and spring cruise data and are smaller than symbol where not visible. Tyr and Lys were not measured at HOT (see Section 3.4). Missing values indicate where D/L ratio was indistinguishable from blank values.
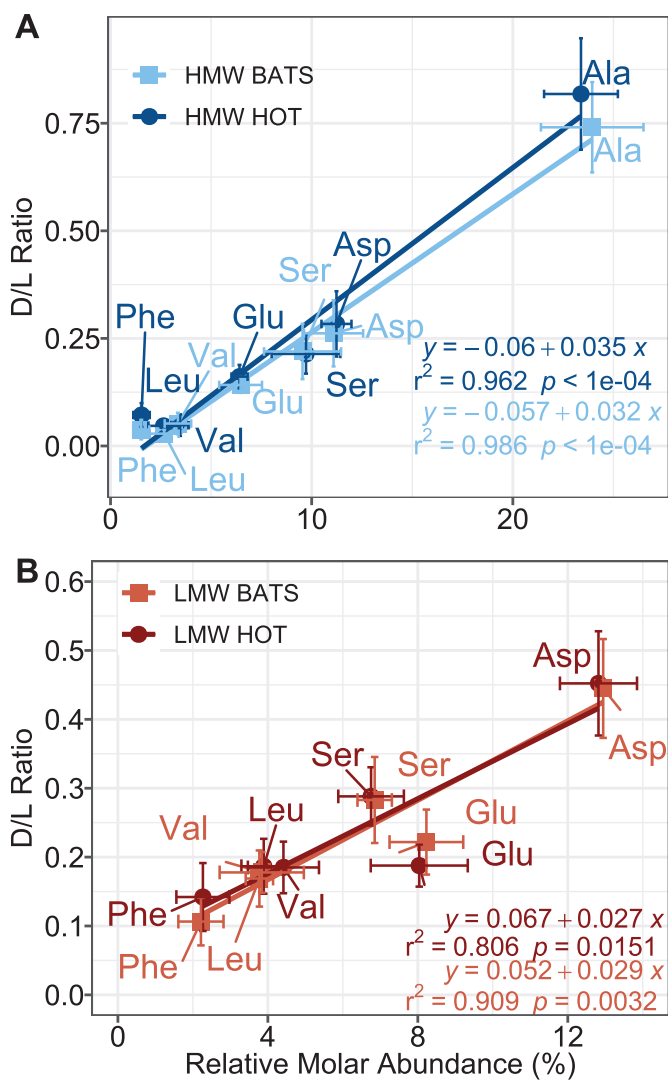
Figure 2.5: Average D/L ratio and average percent relative molar abundance of individual AA was significantly linearly correlated in HMW (A) and LMW (B) DON. For both size fractions, slope and y-intercept of linear regressions are not statistically different at HOT vs. BATS. Error bars represent the standard deviation of all HMW or LMW samples measured in that ocean basin (n = 8).
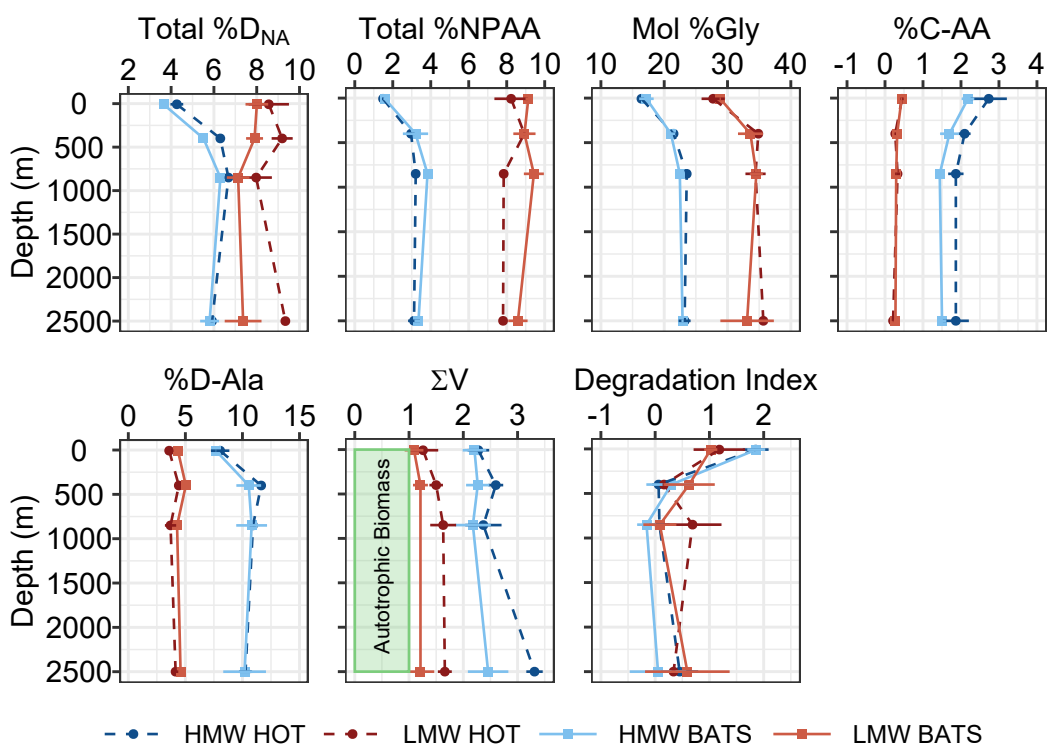
Figure 2.6: Depth profiles of common AA-based proxies in HMW (blue) and LMW (red) DON at HOT (circles, dashed line) and BATS (squares, solid line). Total %D without D-Ala (%D$_{NA}$),total % non-protein AA (NPAA), mol% Gly, and %C-AA (top row) all indicate lower reactivity and greater bacterial source contribution in LMW DON compared to HMW DON. In contrast, %D-Ala, ΣV, and DI index (bottom row) indicate similar or greater bacterial source, resynthesis, and degradation in HMW DON than LMW DON. Error bars represent the mean deviation of summer and spring cruise data for all proxies except ΣV. ΣV was only measured on spring cruise data at depths 400 m to 2500 m, and error bars represent the propagated analytic error associated with triplicate isotopic measurements of each AA. Error bars smaller than symbol were not visible.
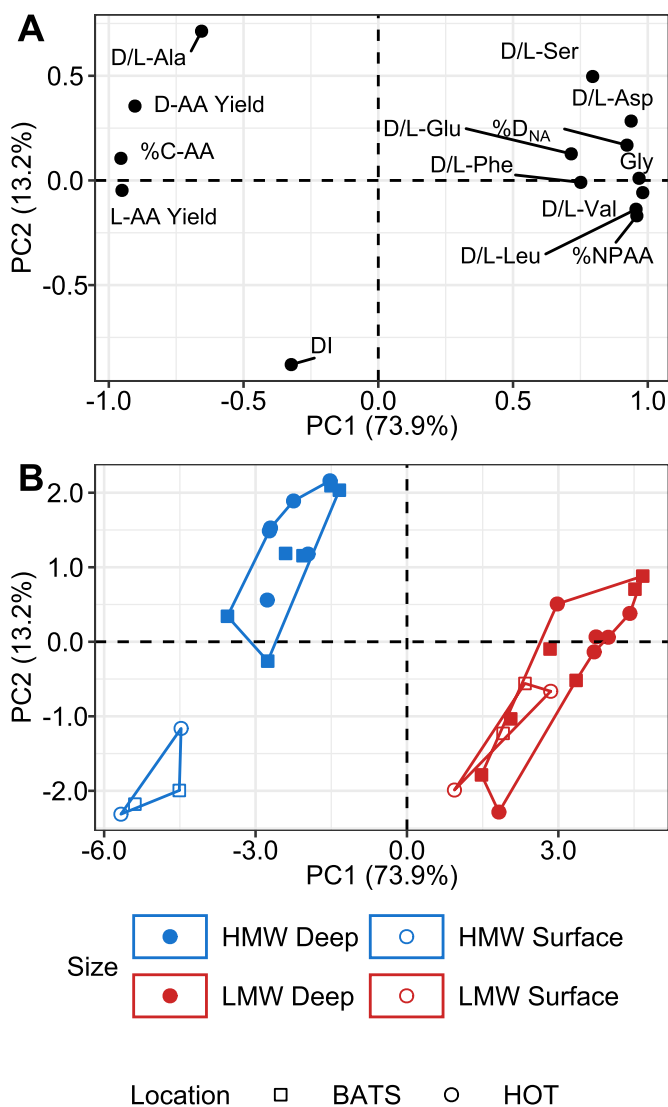
Figure 2.7: Principal component analysis (PCA) of AA-based proxies and D/L ratios of HMW (blue) and LMW (red) DON from BATS (squares) and HOT (circles). A) Loadings of D/L ratios and degradation indices. Only variables that were measured in all samples are included in the PCA. B) PCA scores show clear separation of HMW and LMW DON. Surface (open symbols) HMW DON is clearly separated from deep (filled symbols) HMW DON while surface LMW DON groups with deep LMW DOM.
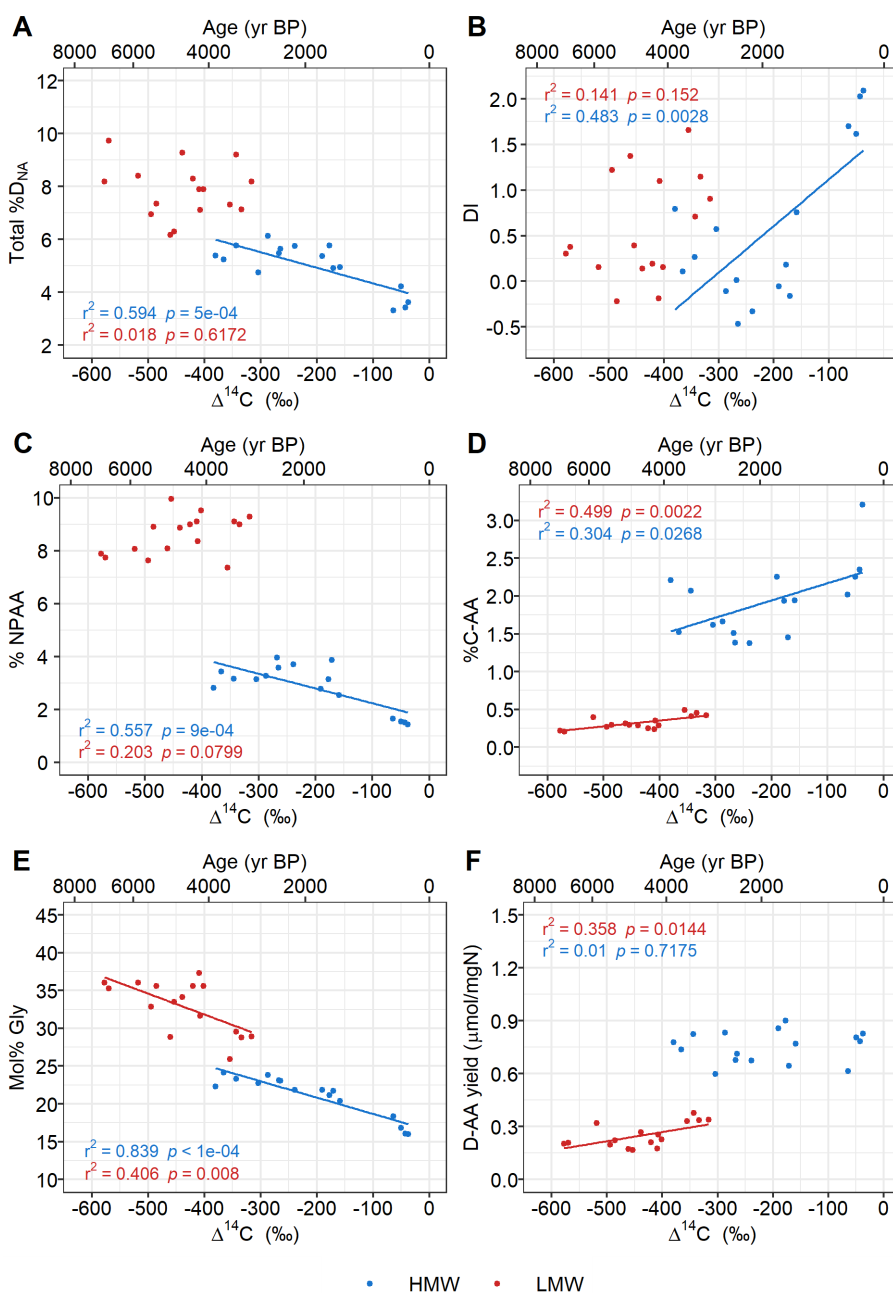
Figure 2.8: Linear regressions of Δ^{14}C vs. common AA-based proxies. (A)Total %D$_{NA}$ (excluding D-Ala), (B) Degradation Index (DI), (C)% non-protein amino acids (NPAA), (D) %C-AA (E), mol% Gly, and (F) total D-AA yield (µmol/mgN). Upper x-axis represents approximate age in years calculated from radiocarbon values according to: Age = 8033^ln(1 + Δ^{14}C/1000). HMW (red) and LMW (blue) include data from both HOT and BATS, with significant linear regressions indicated by solid regression lines. Almost all indices have significant regressions with Δ^{14}C in HMW DON, while few are significant in LMW DON.

Table 2.1: Total hydrolysable amino acids (AA), organic carbon (OC), and organic nitrogen (ON) recoveries of HMW UDOM and LMW SPE-DOM (PPL extract of ultrafiltration permeate). Recovery values are relative to hydrolysable AA, OC, and ON concentrations in 0.2 μm filtered seawater.

Location	Type	%AA	±	%OC	±	%ON	±	n
HOT	HMW UDOM	27.42	8.6	10.58	4.2	13.61	4.1	8
BATS	HMW UDOM	22.9	7.1	9.74	1.9	11.08	3.7	7
HOT	LMW SPEDOM	9.94	3.2	23.38	6.1	13.12	4.7	8
BATS	LMW SPEDOM	15.63	3.7	31.16	5.4	16.29	7.5	7
HOT	Total DOM	37.36	9.2	33.96	7.3	26.73	6.3	16
BATS	Total DOM	38.53	8	40.9	5.7	27.37	8.3	14

2.7 Supplementary Materials

2.7.1 Seasonal variation in BATS AA concentrations

At BATS, total DOM L-AA concentrations at 400 m were $\sim$ 2x greater in Summer 2015 than Spring 2016 (Fig. 2.1). Additionally, at 2500 m, both L and D-AA concentrations were $\sim$ 2x greater in Spring 2016 than Summer 2015 (Fig. 2.1). These seasonal differences are somewhat surprising considering total DOC radiocarbon ages for the spring 400 m and 2500 m samples were 3105 years and 3930 years, respectively (Broek et al., 2020). Still, the greater values we report here are consistent with most recent previous work. L-AA concentrations at BATS measured in June 2001 were even higher at 400 m than we report in Summer 2015, and similarly increase between the surface and 400 m (Kaiser & Benner, 2008). Deep water concentrations measured in the same study were also similar to the L and D-AA concentrations we report at 2500m in Spring 2016. One possibility for the differences we observed is that the AA concentration of recently down welled deep water at BATS varies seasonally depending on surface production. This could explain why deep-water AA concentrations reported in previous work vary by about 70nM (70-140 nM) (Kaiser & Benner, 2008; Lee & Bada, 1977; McCarthy et al., 1996). Alternatively, varying oceanographic conditions prior to each cruise could influence AA concentrations at 400 m.

2.7.2 GC-MS vs. HPLC analysis of L-AA and D-AA

As noted in the main text, AA concentrations of total DOM, the UF permeate, HMW DON, and LMW SPE-DON were all measured via HPLC. In addition to these analyses, AA concentrations were measured at BATS via GC-MS, and for comparison we also present previously published AA concentration data of the samples we collected at HOT measured via GC-MS (Broek et al., 2019). Most

AA reported here are measured by both methods (L & D-Ala, L & D-Asx, L & D-Glx, L & D-Ser, Gly, L-Thr, L-Tyr, L-Val, L-Phe, L-Ile, L-Lys), though L-His, L-Arg, β-Ala, and γ-Aba were only measured by HPLC and D-Val, D-Phe, D-Leu, D-Tyr, and D-Lys were only measured by GC-MS (D-Tyr and D-Lys were only measured in this study and not in previous work (Broek et al., 2019)). Considering non-protein AA β-Ala and γ-Aba and additional D-AA D-Val, D-Phe, D-Leu, D-Tyr, and D-Lys are all key compounds for understanding bacterial contribution to marine DON, the combination of the two methods allows for the most comprehensive analysis of bacterial biomarkers in marine DON.

Overall, the magnitude and depth trends of L-AA and D-AA concentrations as measured by the two methods were similar. Depth averaged L-AA and D-AA concentrations measured via HPLC versus GC-MS (only including AA which are measured by both methods) for HMW and LMW SPE-DON in each respective ocean basin across both seasons (n = 8) were not significantly different for HMW DON at HOT (Welch's two-sample t-test, L-AA: p = 0.31, D-AA: p = 0.29), HMW DON at BATS (Welch's two-sample t-test, L-AA: p = 0.28, D-AA: p = 0.28), or LMW SPE-DON at HOT (Welch's two-sample t-test, L-AA: p = 0.62, D-AA: p = 0.29). However, depth averaged L-AA and D-AA concentrations were significantly less as measured by GC-MS compared to HPLC for LMW SPE-DON at BATS (Welch's two-sample t-test, L-AA: $p < 0.001$, D-AA: $p < 0.001$). It is unclear if the differences observed in LMW SPE-DON at BATS are due to low recovery via GC-MS or higher concentrations as calculated by HPLC. Based on TDAA concentrations of total DOM and HMW DON, we may expect surface L and D-AA concentrations at BATS to be lower than those measured at HOT, while deep concentrations may be similar. This would suggest L and D-AA concentrations in LMW SPE-DON at BATS are in between the measured concentrations of the

two methods.

2.7.3   AA-based proxies and mechanisms of bacterial alteration

A closer consideration of different degradation and bacterial source proxies may help resolve the apparent discrepancy between predicted degradation state of HMW DON and LMW SPE-DON by different proxies. As noted in the in main text, while many of the applied AA-based proxies are interpreted as indicators of bacterial source, overall ON reactivity, or generalized bacterial degradation (Table S2.1), specific mechanistic explanations for most are lacking. DI in particular is used as a quantitative measurement of overall bacterial "degradation", but it is also a fully empirical proxy based on observed AA mol% changes in a continuum of detrital OM (Dauwe et al., 1999). The relationship between DI versustotal %D can inform if these proxies truly reflect relative accumulation of bacterial biomass in HMW and LMW SPE-DON (Fig. S2.7).

In the HMW DON size fraction, both %D-Ala and total $\%D_{NA}$ were significantly negatively correlated with DI ($p < 0.0001$) (Fig. 2.8), indicating a greater proportion of all D-AA are correlated with more degradation as measured by DI (more degradation = more negative DI). These strong relationships are the first direct evidence that a greater relative proportion of bacterial DON correlates with more extensive molar AA distribution shifts (more negative DI) (Fig. 2.8). Additionally, our PCA analysis demonstrates that DI contributes most to PC2, along which surface and deep HMW DON are clearly separated, with more positive DI (less degradation) associated with surface HMW DON and more negative DI in deep HMW DON (Fig. 2.7). Together, these observations provide compelling support for the assumption that DI records bacterially mediated transformations in HWM DON.

In contrast, for LMW SPE-DON, no significant relationships between %D-Ala or total $\%D_{NA}$ and DI were observed (Fig. S2.7), indicating that the overall AA molar changes recorded by DI are not linked to accumulation of bacterial D-AA containing molecules in LMW SPE-DON. We also note that the DOM-specific DI calculation used here was developed based on observed differences in AA molar abundance of planktonic DOM and total DOM from the North Pacific (Kaiser & Benner, 2009), which may not be representative of the LMW SPE-DON size fraction. This is consistent with our hypothesis that D-AA may trace fundamentally different nitrogenous classes in the different MW fractions (Section 2.5.3) and suggests the interpretation of DI in the LMW SPE-DON samples collected here is uncertain (Supplementary 2.7.2).

Like DI, ΣV was developed based on observed changes to proteinaceous dominated ON with bacterial resynthesis (McCarthy et al., 2007). Thus, while ΣV values might be able to directly address this idea, unfortunately our ΣV data is too sparse for meaningful correlations. However, the different trends in DI versus ΣV with depth in both the HMW and LMW SPE-DON pools (Fig. 2.6) does support the idea that these proxies trace different bacterial mechanisms. While it is possible the above observations are a result of selectivity of the PPL resin used to collect our LMW SPE-DON samples, measurements of the UF permeate indicate this is likely not the case. Specifically, the similar DI values observed in LMW SPE-DON and the UF permeate at most depths, as well as the inconsistent DI depth trends (Fig. S2.5) and the lack of any correlation between DI and $\%D_{NA}$ in the UF permeate (data not shown) instead suggest that changes recorded by DI may also not be linked to accumulation of bacterial D-AA molecules in the bulk LMW SPE-DON pool.

2.7.4 Bacterial contribution to Total, HMW, and LMW SPE-DON

Percent bacterial OC and ON of total, HMW, and LMW DOM were calculated according to Tremblay and Benner (2006) using the equation: Bacterial C or N (%)= (Biomarker$_{DOM}$/Biomarker$_{Bacterial}$)*100, where Biomarker$_{DOM}$ is the C or N normalized yield of D-Ala, D-Asx, or D-Glx in DOM samples and Biomarker$_{Bacterial}$ is the C or N normalized yield of that D-AA in freshly produced bacterial DOM. Biomarker$_{Bacterial}$ values for bacterial DOM represent yields of each respective D-AA in total bacterial DOM from Kaiser and Benner (2008).

The % bacterial OC and ON for HMW and LMW SPE-DON varied greatly depending on which D-AA was used for the calculation, and often yielded values significantly greater than 100% for HMW DON (Fig. S2.4). Bourgoin and Tremblay (2010) similarly found bacterial contributions greater than 100% for HMW UDON when using bacterial endmembers calculated from total bacterial DOM, as well as variable results with each biomarker. These results indicate bacterial biomarker endmembers calculated for the total DOM pool likely do not reflect the yields of these biomarkers specifically in the HMW and LMW SPE-DON size fractions, consistent with the clearly different D-AA contributions. Instead, in order to apply this calculation to isolated DOM fractions, bacterial DOM endmembers specific to those size fractions must also be determined.

The percent of bacterial DOC and DON in total DOM also varied greatly with each D-AA biomarker, ranging from an average of 30% to 64% of total DOC and 32% to 100% of total DON. The large range in values we observe depending on chosen biomarker and with depth is inconsistent with the only previous report of the percent of bacterial derived DOC and DON at BATS and HOT, which found 21% to 29% of DOC and 45% to 54% of DON to be derived from bacteria regardless of ocean basin, depth, or which D-AA was used (Kaiser and Benner

2008). These results are consistent, however, with a recent report of bacterial contributions to sedimentary OM calculated from Mur and D-Glx, which found highly variable and unreasonable ($> 800\%$) values depending on chosen biomarker (Lehmann et al., 2020).

The use of D-AA as quantitative biomarkers relies on a few assumptions, which, if not met, could explain the large range in values we observe here. First, the bacterial DOM endmember must be representative of the system investigated. The endmember values applied here represent a mix of freshly produced bacterial DOM from the surface and deep Pacific Ocean and coastal Atlantic Ocean, which we would expect to be reasonable endmembers for total DOM from HOT and BATS. Second, the biomarkers used (D-Asx, D-Glx, D-Ala) must have a similar reactivity to total bacterial carbon or nitrogen, meaning the ratio of biomarker to total bacterial biomass remains constant during degradation. As discussed in the main text (Section 2.5.3), while some incubation experiments support this theory (Kawasaki & Benner, 2006), we hypothesize this may not be the case for all D-AA in the more complex open ocean. A similar suggestion was made by Lehman et al., 2020, who hypothesized selective preservation of peptidoglycan or other bacterial biomolecules would lead to non-quantitative and variable overestimations of bacterial contribution. The large range in values we observed depending on which biomarker was applied is consistent with this hypothesis and provides additional support for a range of reactivity for dissolved D-AA and bulk bacterial OC/ON. We suggest that assumptions behind this approach should be thoroughly tested before D-AA yields can be used as quantitative biomarkers.

Supplementary Figures and Tables
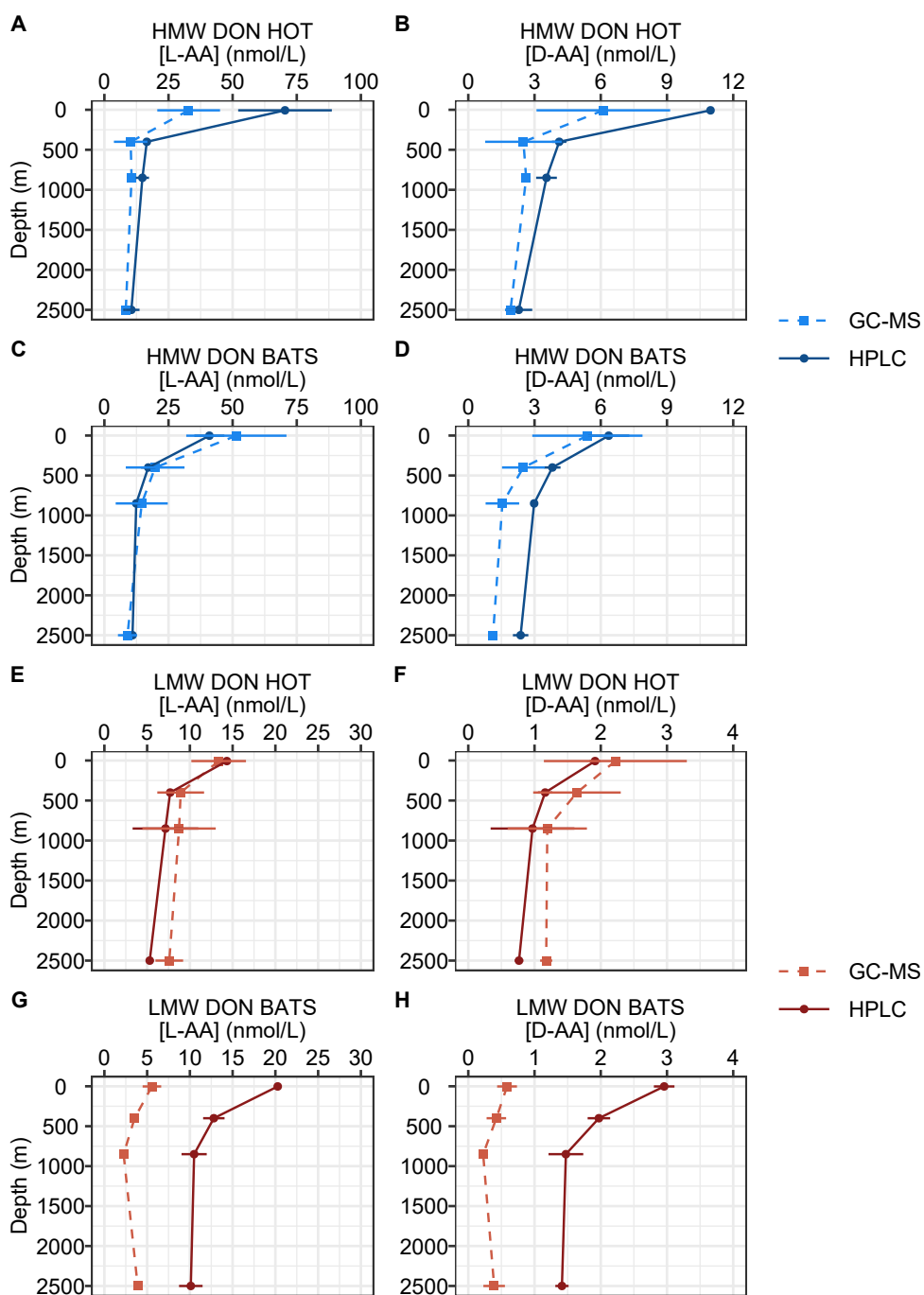
Figure S2.1: Depth profiles of total L-AA and D-AA concentration measured by HPLC (darker shade, solid) and GC-MS (lighter shade, dashed) in HMW DON (blue), LMW SPE-DON (red). Only L and D-AA measured by both methods are included. L-AA concentration also includes the achiral AA Glycine. Error bars represent mean deviation of summer and spring cruise data and are smaller than symbol where not visible.
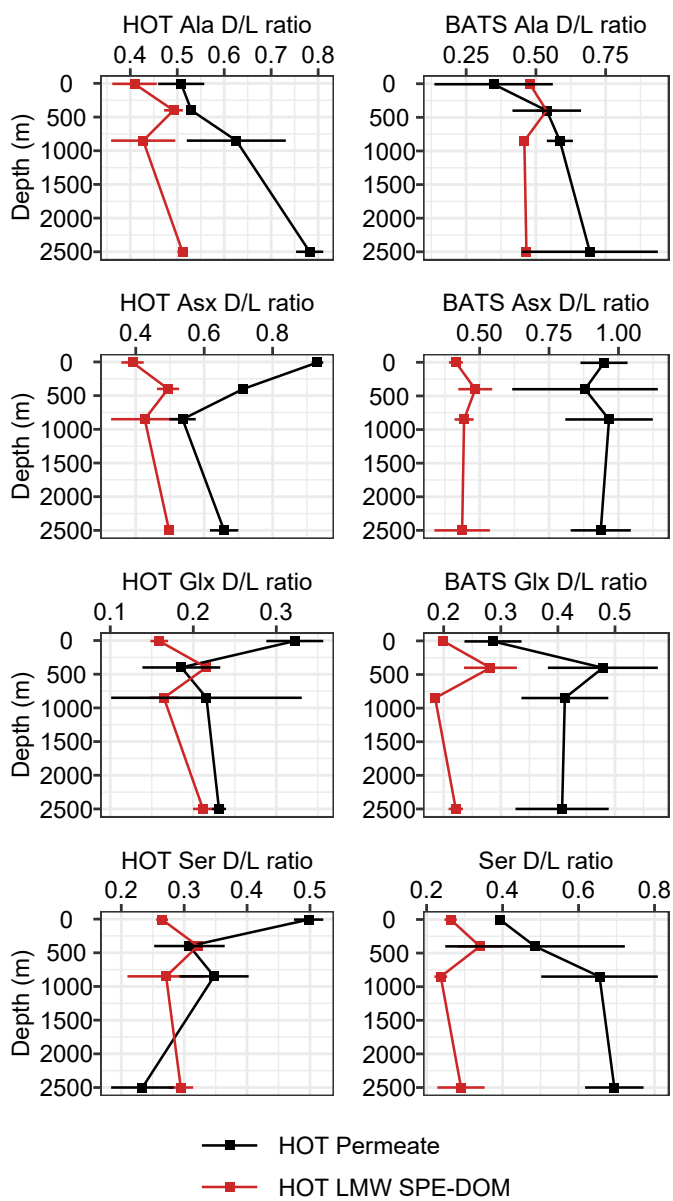
Figure S2.2: Depth profiles of D/L ratios measured via HPLC in LMW SPE-DON (red) and UF permeate (black). Error bars represent mean deviation of summer and spring cruise data and are smaller than symbol where not visible.

64
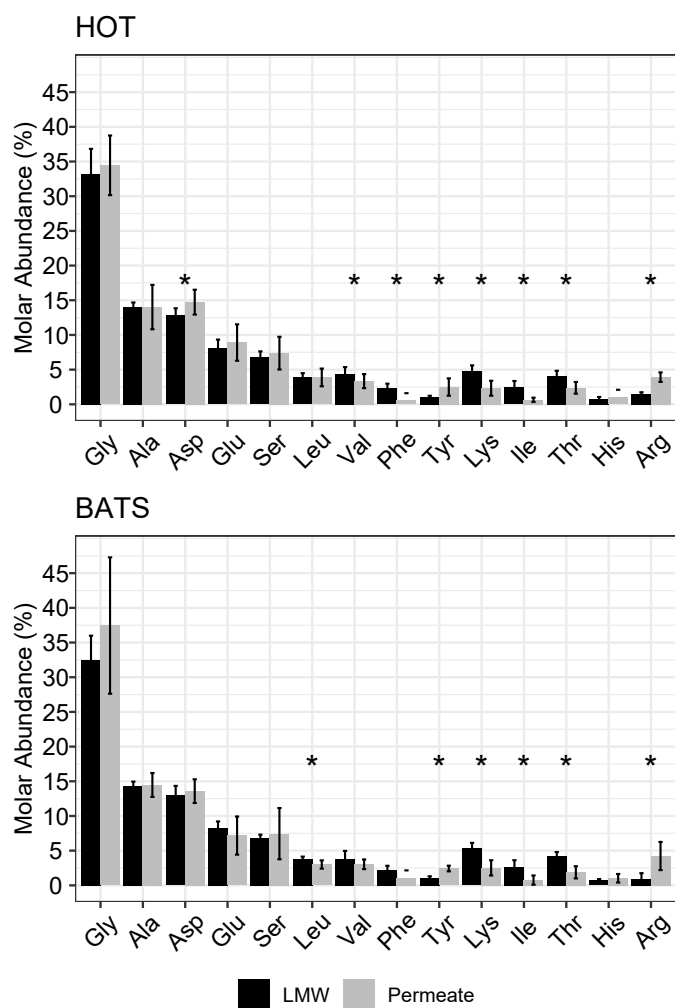
Figure S2.3: Depth averaged percent relative molar abundance of individual AA measured by HPLC in the ultrafiltration permeate (grey) and LMW DON collected via solid phase extraction of the ultrafiltration permeate (black) at HOT (left) and BATS (right). Error bars represent the standard deviation of all depths and spring and summer cruise samples (n = 8).
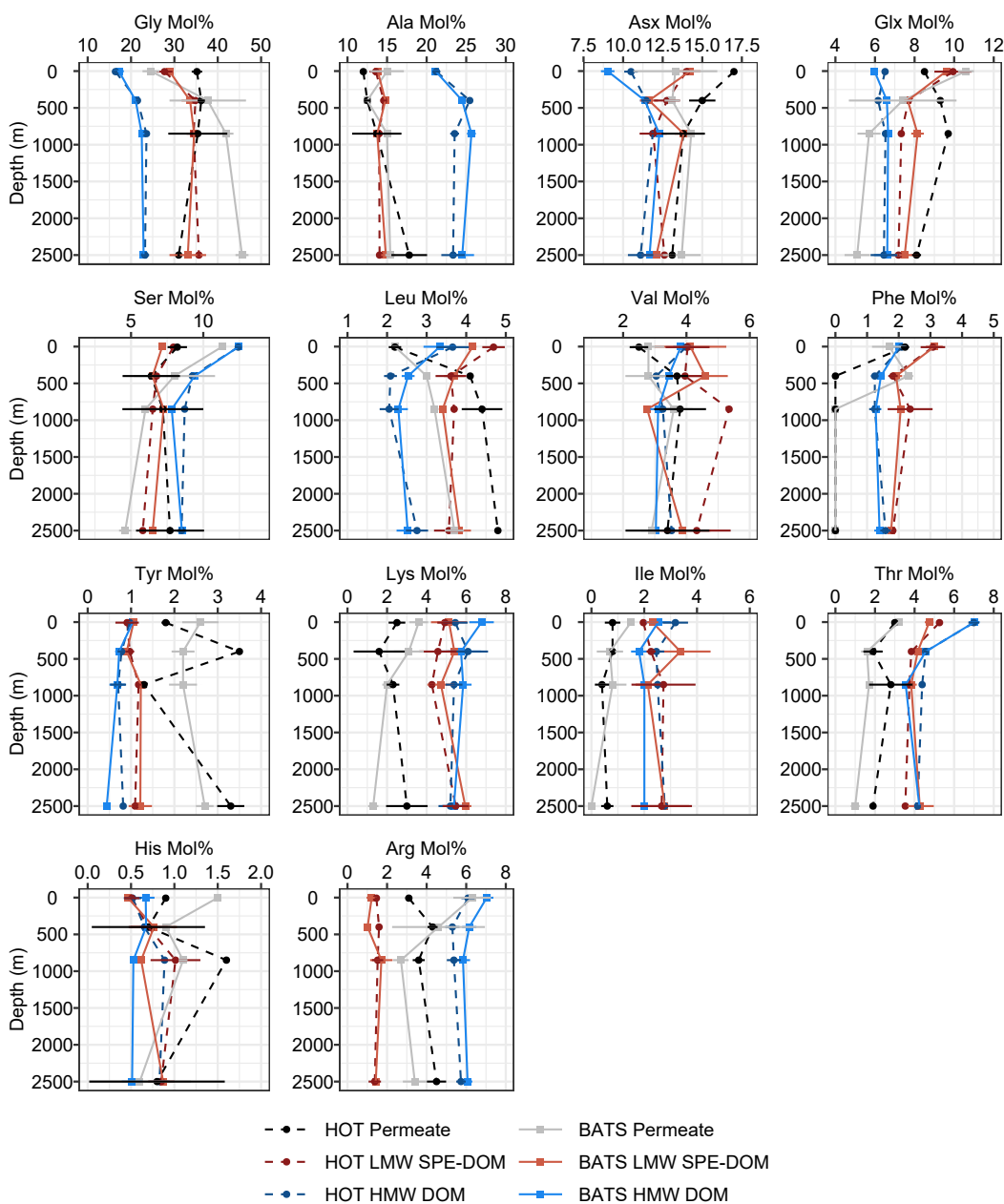
Figure S2.4: Depth profiles of percent relative molar abundance of individual AA measured by HPLC in HMW DON (blue), LMW SPE-DON (red) and UF permeate (grey/black). Error bars represent mean deviation of summer and spring cruise data and are smaller than symbol where not visible.
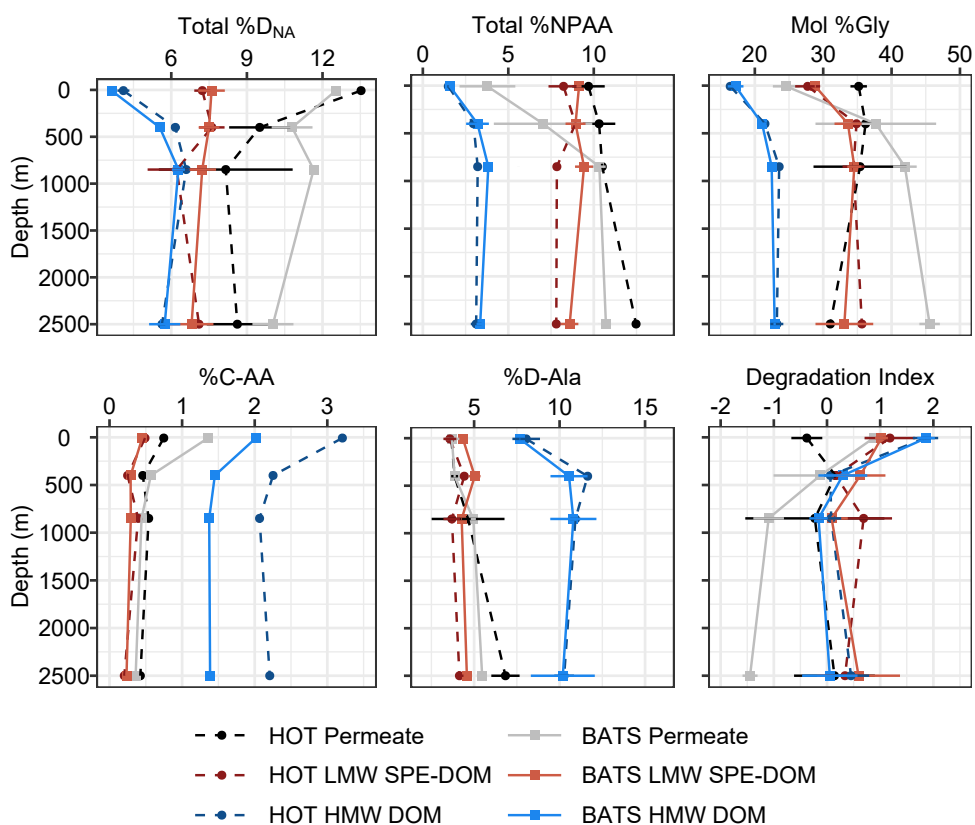
Figure S2.5: Depth profiles of common AA-based proxies which can be measured via HPLC in HMW DON (blue) and LMW SPE-DON (red) and UF permeate (grey/back) at HOT (circles, dashed line) and BATS (squares, solid line). Total %D without D-Ala (%D_{NA}) only includes D-Asx, D-Glx, and D-Ser. Error bars represent mean deviation of summer and spring cruise data and are smaller than symbol where not visible.
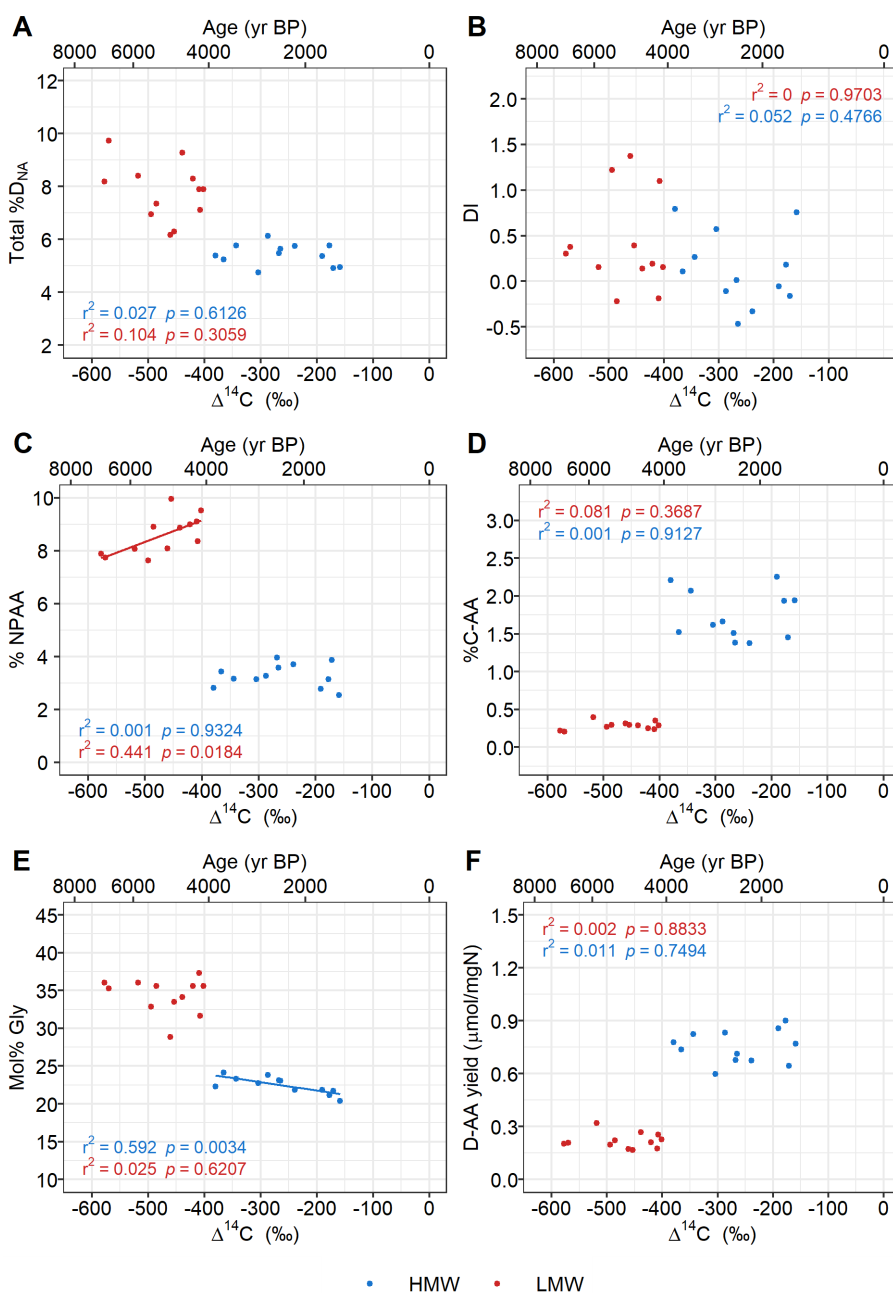
Figure S2.6: Linear regressions of $\Delta^{14}C$ vs. common AA-based proxies in HMW DON (blue) and LMW SPE-DON (red) collected from 400 m to 2500 m deep. (A) Total $\%D_{NA}$ (excluding D-Ala), (B) Degradation Index (DI), (C)% non-protein amino acids (NPAA), (D) %C-AA (E), mol% Gly, and (F) total D-AA yield (μmol/mgN). Upper x-axis represents approximate age in years calculated from radiocarbon values according to: Age = $8033\hat{\ }\ln(1 + \Delta^{14}C/1000)$. HMW and LMW DON include data from both HOT and BATS, with significant linear regressions indicated by solid regression lines. Without surface samples, very few significant relationships are observed.
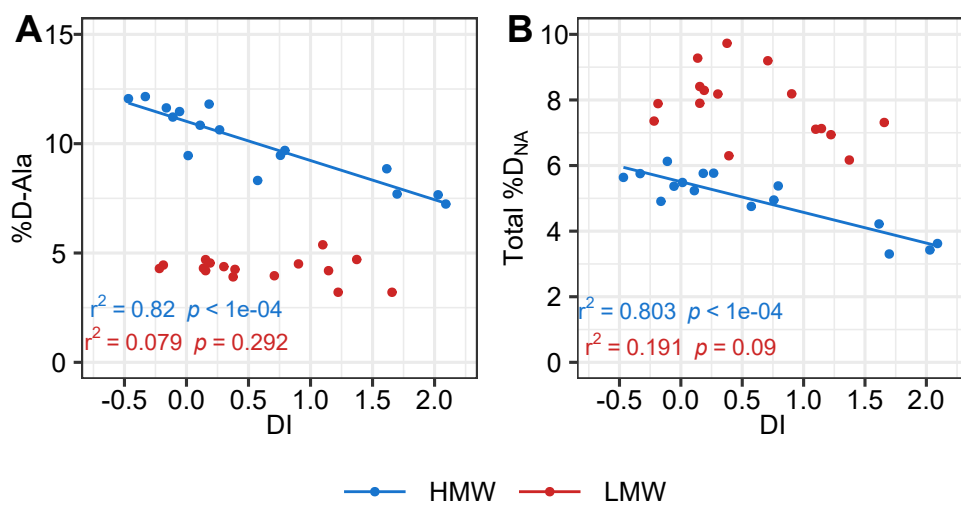
Figure S2.7: (A)%D-Ala and (B) total %D$_{NA}$ (excluding D-Ala) are significantly correlated with DI in HMW DON (blue) but not in LMW SPE-DON (red). Each size fraction includes data from HOT and BATS. Significant linear regressions indicated by solid regression lines.
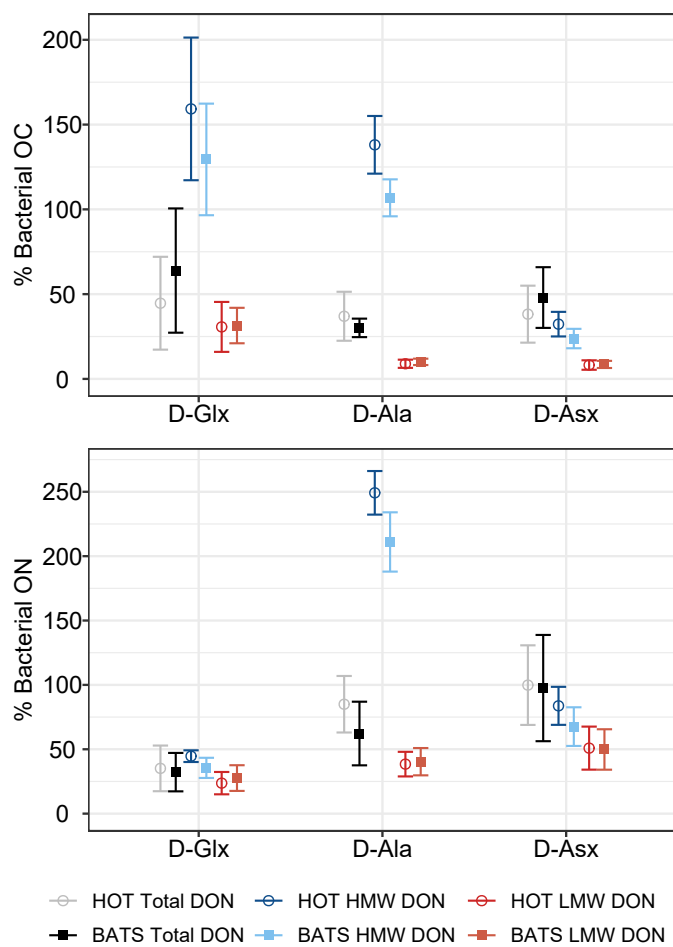
Figure S2.8: Calculated total percent of organic carbon (left) and organic nitrogen (right) of bacterial origin in total (black), HMW (blue), and LMW (red) DON at BATS (filled squares) and HOT (open circles). Bacterial yields vary with different biomarkers used, and often indicate unreasonable results. Error bars represent the standard deviation of all depths and spring and summer cruise samples (n = 8).

Table S2.1: Summary of degradation and bacterial source proxies discussed in the text. Amino acid proxies (AA proxy) as we have used them in the paper are listed then with the basic assumed underlying process or property they trace and notes on specific mechanistic interpretation. The latter clarifies the mechanism (or mechanisms, where no single mechanism is known or widely accepted) underlying changes in each.

AA Proxy	Process or Property	Interpretation
%D (without Ala)	Bacterial source or degradation	Addition or selective preservation of bacterially derived molecules, typically more refractory (may or may not be proteinaceous)
%D-Ala	Bacterial source or degradation	Addition or selective preservation of bacterially derived molecules, typically more labile (hypothesized main peptidoglycan component)
%NPAA	Bacterial degradation	Product of the bacterial hydrolysis of amino acids
DI	Bacterial degradation	Changes in mol% of AA away from distribution typical in primary production. Caused by bacterial activity (could be either selective preservation or bacterial addition)
ΣV	Bacterial resynthesis	Metabolic resynthesis of primary production AA by heterotrophic bacterial activity; addition of new bacterial-derived proteinaceous material
%C-AA	Reactivity/lability or bacterial degradation	Relative proportion of OC which is AAs - indicates relative OM "freshness" or bacterial degradation
D-AA Yield	Bacterial Source	Relative bacterial contribution to ON
L-AA Yield	Reactivity	Proportion of organic N which is labile AA
Mol% Gly	Bacterial source or degradation	Increase of mol% Gly by either selective removal or addition of new bacterial biomolecules

Table S2.2: Hydrolysable L-AA and D-AA concentrations in total DON from HOT and BATS.

Location	Year	Depth (m)	[L-AA] (nM)	[D-AA] (nM)
HOT	2014	7.5	197.44	46.81
HOT	2014	400	73.08	16.3
HOT	2014	850	73.32	13.99
HOT	2014	2500	42.85	8.43
HOT	2015	7.5	222.05	50.13
HOT	2015	400	70.36	15.5
HOT	2015	850	56.66	10.28
HOT	2015	2500	52.48	9.52
BATS	2015	7.5	136.44	29.49
BATS	2015	400	166.82	21.03
BATS	2015	850	76.59	15.34
BATS	2015	2500	53.94	13.22
BATS	2016	7.5	138.55	24.75
BATS	2016	400	71.88	16.45
BATS	2016	850	66.35	15.34
BATS	2016	2500	112.95	22.6

Table S2.3: Calculated degradation indices and AA yields normalized to total organic N in HMW and LMW SPE-DON at BATS and HOT. D and L-AA yields are in µmol/mgON.

Site	Year	Size Fraction	Depth (m)	DI	Δ^{14}C	%D	%D$_{NA}$	D-AA Yield	L-AA Yield	ΣV	%N-AA	%C-AA	Mol% Gly	β-Ala%	γ-Aba%
HOT	2014	HMW	7.5	2.09	-37.32	11.15	4.15	0.828	6.615	2.26	12.22	3.21	15.98	1	0.43
HOT	2014	HMW	400	-0.1	-190.4	17.42	6.67	0.856	4.178	ND	8.18	2.25	21.83	2.38	0.39
HOT	2014	HMW	850	0.27	-343.7	16.7	6.83	0.824	4.2	ND	8.32	2.07	23.33	2.64	0.52
HOT	2014	HMW	2500	0.79	-379.7	15.51	6.23	0.778	4.269	ND	8.39	2.21	22.28	2.42	0.38
HOT	2015	HMW	7.5	1.61	-49.96	13.6	4.9	0.805	5.152	2.37	9.94	2.25	16.83	1.15	0.4
HOT	2015	HMW	400	0.18	-177.3	18.17	6.82	0.9	4.065	2.60	8.3	1.94	21.12	2.65	0.5
HOT	2015	HMW	850	-0.1	-286.9	18.03	7.28	0.832	3.803	3.04	7.6	1.66	23.79	2.77	0.49
HOT	2015	HMW	2500	0.11	-365.7	16.8	6.05	0.737	3.614	3.31	7.11	1.52	24.17	2.89	0.54
BATS	2015	HMW	7.5	2.03	-43.02	11.61	3.87	0.785	5.933	ND	11.35	2.35	16.04	1.13	0.38
BATS	2015	HMW	400	0.76	-158.5	15.27	5.77	0.774	4.253	ND	8.33	1.94	20.4	2.08	0.46
BATS	2015	HMW	850	0.01	-267.5	16.19	6.38	0.683	3.42	ND	6.59	1.51	23.09	3.42	0.55
BATS	2015	HMW	2500	0.57	-304.3	13.46	5.38	0.601	3.857	ND	7.18	1.62	22.72	2.67	0.47
BATS	2016	HMW	7.5	1.7	-64.12	11.45	3.81	0.615	4.777	2.21	9.3	2.02	18.34	1.23	0.42
BATS	2016	HMW	400	-0.2	-171	16.99	5.84	0.647	3.182	2.27	6.42	1.45	21.73	2.74	1.13
BATS	2016	HMW	850	-0.3	-239.2	18.37	6.91	0.676	3.034	2.18	6.29	1.37	21.84	3.2	0.49
BATS	2016	HMW	2500	-0.5	-265.1	18.15	6.75	0.715	3.254	2.46	6.75	1.39	23.07	2.87	0.7
HOT	2014	LMW	7.5	1.66	-354.8	11.65	8.59	0.329	2.653	ND	4.51	0.49	25.9	5.65	1.7
HOT	2014	LMW	400	0.19	-420.4	13.58	9.32	0.21	1.363	ND	2.33	0.25	35.58	7.37	1.63
HOT	2014	LMW	850	0.15	-518	13.53	9.34	0.32	1.991	ND	3.45	0.39	36.02	6.26	1.8
HOT	2014	LMW	2500	0.38	-569.7	13.53	9.33	0.207	1.106	ND	1.92	0.2	35.29	6.31	1.42
HOT	2015	LMW	7.5	0.71	-343	13.98	10.04	0.378	2.233	1.71	3.84	0.41	29.56	7.07	2.04

Continued on next page

73

Location	Year	Size Fraction	Depth (m)	DI	$\Delta^{14}C$	%D	$\%D_{NA}$	D-AA Yield	L-AA Yield	ΣV	%N-AA	%C-AA	Mol% Gly	β-Ala%	γ-Aba%
HOT	2015	LMW	400	0.14	-438.8	14.09	9.84	0.267	1.543	1.63	2.73	0.29	34.14	6.76	2.11
HOT	2015	LMW	850	1.22	-494.3	10.52	7.22	0.196	1.56	1.63	2.61	0.27	32.83	4.92	2.7
HOT	2015	LMW	2500	0.3	-577.6	13.32	9.08	0.201	1.303	1.66	2.27	0.22	36.03	6.22	1.66
BATS	2015	LMW	7.5	0.9	-316.1	13.87	9.8	0.356	2.23	ND	3.85	0.42	28.94	6.76	2.52
BATS	2015	LMW	400	1.1	-407.2	13.44	8.51	0.251	1.722	ND	1.21	0.35	31.66	6.34	2.01
BATS	2015	LMW	850	0.39	-453.6	10.7	6.68	0.156	1.304	ND	2.27	0.3	33.52	5.23	4.73
BATS	2015	LMW	2500	1.37	-460.7	11.67	7.27	0.175	1.382	ND	1.1	0.32	28.87	4.62	3.46
BATS	2016	LMW	7.5	1.15	-333.6	12.65	8.76	0.34	2.388	1.09	4.09	0.45	28.76	6.93	2.06
BATS	2016	LMW	400	0.15	-401.5	13.91	9.59	0.229	1.537	1.21	2.67	0.29	35.58	6.83	2.69
BATS	2016	LMW	850	-0.2	-485.5	11.78	7.78	0.208	1.56	ND	2.66	0.29	35.58	6.22	2.68
BATS	2016	LMW	2500	-0.2	-408.8	13.48	9.37	0.177	1.207	1.21	2.11	0.23	37.32	6.91	2.19

Table S2.4: D/L ratios of all AA with D enantiomers in HMW UDON and LMW SPE-DON at BATS and HOT. D/L-Leu, D/L-Val, and D/L-Phe at HOT were previously published in Broek et al., 2019.

Location	Year	Size Fraction	Depth	D/L-Ala	D/L-Asp	D/L-Glu	D/L-Ser	D/L-Leu	D/L-Val	D/L-Phe	D/L-Tyr	D/L-Lys
HOT	2014	HMW	2	0.55	0.16	0.11	0.13	0.03	0.01	0.05	ND	ND
HOT	2014	HMW	400	0.82	0.3	0.17	0.23	0.04	0.06	0.06	ND	ND
HOT	2014	HMW	850	0.86	0.34	0.18	0.24	0.03	0.02	0.03	ND	ND
HOT	2014	HMW	2500	0.83	0.3	0.16	0.22	0.05	0.07	0.07	ND	ND
HOT	2015	HMW	2	0.73	0.17	0.13	0.16	0.05	0.02	0.08	ND	ND
HOT	2015	HMW	400	0.95	0.33	0.19	0.25	0.06	0.06	0.1	ND	ND
HOT	2015	HMW	850	0.94	0.36	0.19	0.26	0.07	0.08	0.1	ND	ND
HOT	2015	HMW	2500	0.86	0.31	0.17	0.22	0.06	0.06	0.1	ND	ND
BATS	2015	HMW	2	0.62	0.15	0.14	0.12	0.02	0.02	0.01	ND	ND
BATS	2015	HMW	400	0.81	0.26	0.13	0.21	0.02	0.04	0.03	ND	0.02
BATS	2015	HMW	850	0.76	0.3	0.13	0.27	0.04	0.06	0.04	0.04	0.02
BATS	2015	HMW	2500	0.63	0.25	0.12	0.22	0.03	0.06	0.03	0.01	0.02
BATS	2016	HMW	2	0.6	0.15	0.14	0.14	0.02	0.02	0.03	0	0.01
BATS	2016	HMW	400	0.81	0.28	0.13	0.24	0.03	0.06	0.03	0.01	0.03
BATS	2016	HMW	850	0.85	0.34	0.16	0.3	0.03	0.07	0.06	0.02	0.04
BATS	2016	HMW	2500	0.84	0.36	0.18	0.26	0.04	0.09	0.08	0.02	0.05
HOT	2014	LMW	2	0.36	0.36	0.15	0.26	0.11	0.11	0.12	ND	ND
HOT	2014	LMW	400	0.47	0.46	0.22	0.32	0.22	0.2	0.14	ND	ND
HOT	2014	LMW	850	0.5	0.52	0.18	0.33	0.17	0.18	0.07	ND	ND
HOT	2014	LMW	2500	0.5	0.5	0.22	0.31	0.2	0.21	0.11	ND	ND
HOT	2015	LMW	2	0.46	0.42	0.17	0.27	0.18	0.16	0.22	ND	ND

Continued on next page

75

Location	Year	Size Fraction	Depth	D/L-Ala	D/L-Asp	D/L-Glu	D/L-Ser	D/L-Leu	D/L-Val	D/L-Phe	D/L-Tyr	D/L-Lys
HOT	2015	LMW	400	0.51	0.53	0.21	0.32	0.24	0.22	0.2	ND	ND
HOT	2015	LMW	850	0.36	0.33	0.15	0.21	0.17	0.17	0.13	ND	ND
HOT	2015	LMW	2500	0.52	0.5	0.2	0.27	0.2	0.23	0.15	ND	ND
BATS	2015	LMW	2	0.49	0.44	0.2	0.28	0.14	0.15	0.12	0.04	0.02
BATS	2015	LMW	400	0.57	0.42	0.33	0.28	0.2	0.2	0.12	0.05	0.02
BATS	2015	LMW	850	0.46	0.41	0.19	0.22	0.13	0.17	0.06	0.02	0.01
BATS	2015	LMW	2500	0.45	0.34	0.23	0.23	0.12	0.15	0.05	0.02	0.01
BATS	2016	LMW	2	0.46	0.39	0.2	0.25	0.14	0.15	0.13	0.04	0.08
BATS	2016	LMW	400	0.5	0.54	0.24	0.4	0.21	0.21	0.15	0.08	0.11
BATS	2016	LMW	850	0.46	0.48	0.18	0.25	0.18	0.18	0.09	0.04	0.04
BATS	2016	LMW	2500	0.48	0.54	0.21	0.35	0.22	0.22	0.13	0.1	0.11

Chapter 3

Compound-specific isotope analysis of amino acids: novel insight to production and degradation mechanisms of HMW and LMW SPE-DON

3.1 Abstract

Stable nitrogen isotopes (δ^{15}N) are a potentially powerful tool to understand marine dissolved organic nitrogen (DON) cycling. However, δ^{15}N measurements of total DON are restricted to the surface ocean where inorganic N concentrations are negligible. In contrast, more detailed δ^{15}N measurements can be made on isolated DON size fractions from throughout the entire water column. Compound-specific isotope analysis of amino acids (CSI-AA) has significantly greater information potential than bulk δ^{15}N measurements and may represent a transformational new tool to understand proteinaceous marine DON source and degradation mechanisms. In this study, we compare δ^{15}N amino acid (δ^{15}N-AA) patterns in selectively isolated high molecular weight (HMW) and for the first time low molecular weight (LMW) DON from the surface to 2500 m deep in the Atlantic and Pacific Subtropical Gyres. Our overarching goal was to use the new information potential from CSI-AA to investigate sources and transformation processes of LMW refractory DON, which makes up the vast majority of the marine DON reservoir.

We observed distinct δ^{15}N-AA patterns of HMW and LMW SPE-DON, indicating unique formation and bacterial alteration mechanisms for AA in each size fraction. Consistent with previous work, these data suggest that proteinaceous HMW DON in the surface ocean derives mainly from heterotrophic bacterial sources utilizing subsurface nitrate. However, new δ^{15}N-AA data also indicates a previously overlooked protozoan trophic step, consistent with microbial loop processes of HMW DON. Additionally, we find support for a previous hypothesis that surface and mesopelagic HMW DON production are decoupled, with proteinaceous mesopelagic HMW DON primarily derived from heterotrophic resynthesis of suspended PON.

In contrast to the HMW DON pool, these first LMW SPE-DON δ^{15}N-AA patterns suggest completely independent cycling processes, and likely molecular composition, of this AA in this size fraction. δ^{15}N-AA patterns of LMW SPE-DON at all depths and both ocean basins appear surprisingly like those of autotrophs, with signs of limited bacterial resynthesis or microbial loop trophic transfer. Together with previous work indicating a substantial contribution of heterotrophic bacterial biomolecules in these same LMW SPE-DON samples, this suggests there is a direct heterotrophic bacterial source to proteinaceous LMW SPE-DON with limited isotopic fractionation via degradation or resynthesis. We hypothesize that *de novo* synthesis of LMW SPE-DON followed by direct exudation or viral lysis of bacterial LMW molecules are most consistent with these isotopic signatures.

Within each size fraction, δ^{15}N-AA patterns were incredibly similar between HOT and BATS, suggesting these findings can be extended to marine DON at least throughout subtropical gyres, if not globally. If true, our hypotheses indicate a significant shift in our understanding of the production and degradation mechanisms of marine DON. We suggest production of LMW, AA-containing molecules

is not linked to production and degradation of HMW DON, and instead bacteria rapidly produce refractory, LMW AA-containing molecules in the surface ocean. Considering LMW SPE-DON makes up most of the marine DON reservoir, these new data may have wide-reaching implications for the production of refractory DON. Finally, these results demonstrate the impressive potential of CSI-AA as a new tool to study sources and transformation of the DON pool.

3.2 Introduction

Throughout extensive oligotrophic regions of the surface ocean, a lack of usable nitrogen restricts primary production. However, dissolved organic nitrogen (DON) persists in measurable quantities throughout the ocean at all depths, representing the world's largest reactive nitrogen reservoir. The recalcitrant nature of most marine DON, especially in the upper ocean, thus exerts a vital control on marine food webs and carbon sequestration by the ocean. Still, despite its global importance, a detailed understanding of marine DON source, cycling processes, and chemical composition remain elusive. This is largely because studies of the DON pool require nitrogen-specific, molecular-level tracers. Stable nitrogen isotopes (δ^{15}N) represent one such nitrogen-specific tool that is potentially invaluable to probing DON sources and cycling. However, there is no known method to isolate DON from dissolved inorganic nitrogen (DIN), meaning measurements of bulk DON δ^{15}N are restricted to the surface ocean where DIN concentrations are generally negligible. Because of this, to date, δ^{15}N applications to marine DON remain extremely limited and are constrained to a small portion of the total ocean.

Isolating fractions of DOM from seawater can circumvent some of these issues, allowing for both nitrogen isotopic and detailed molecular study. However, there is no known isolation method to recover the entire DOM pool. For decades, most DON studies focused on the high molecular weight (HMW) size fraction that can be isolated in large concentrations from seawater via ultrafiltration (Benner et al., 1992). ^{15}N NMR studies of HMW DON indicate the functional composition is almost entirely amide, with proteinaceous material dominating HMW DON at all subsurface depths (Aluwihare et al., 2005; McCarthy et al., 1997). Global offsets in δ^{15}N values of total and HMW DON vary with the δ^{15}N and concentration

of NO_3^-, supporting recent biological production of a subset of the DON pool (Broek et al., Submitted; Knapp et al., 2011). Additionally, $\delta^{15}N$ values of HMW DON are generally greater than PON from the same region, potentially suggesting degradation via processes with substantial isotopic fractionation (Knapp et al., 2012; Yamaguchi & McCarthy, 2018).

However, it is now generally understood that dissolved HMW molecules tend to be younger, more labile compounds, while low molecular weight (LMW) DOM is more representative of most DOM. An abundance of studies supports the production of older, refractory, LMW SPE-DOC from younger, semi-labile HMW DOC, collectively termed the "size-age-reactivity continuum" (Amon and Benner, 1996; Benner and Amon, 2015; Walker et al., 2016; Walker et al., 2016). Combined with evidence for microbial production of refractory DOC molecules (Jiao et al., 2011, Jiao et al., 2010) this suggests bacteria progressively form refractory DOC molecules. However, it is unclear if a similar mechanism exists for marine DON. Bulk $\delta^{15}N$ measurements of LMW SPE-DON indicate uniformly lower $\delta^{15}N$ values than HMW DON, suggesting the production of LMW SPE-DON from HMW DON is unlikely (Broek et al., Submitted; Knapp et al., 2012). Instead, the authors suggested the two pools either have different nitrogen sources or cycle and degrade independently. These theories are inconsistent with expectations based on a "size-age-reactivity" continuum and would have significant implications for our interpretation of DON source and cycling. However, the specific formation and degradation mechanisms of marine DON, especially the quantitatively more important LMW SPE-DON pool, remain to be confidently understood.

Compound-specific isotope analysis of amino acids (CSI-AA) recently emerged as a promising new tool that can resolve independent process that are confounded with bulk $\delta^{15}N$ measurements. Amino acids (AA) are synthesized via unique, indi-

vidual pathways, meaning δ^{15}N CSI-AA patterns can distinguish between different mechanisms of organic matter (OM) alteration (Ohkouchi et al., 2017). Considering most identifiable nitrogen in HMW DON contains amide functional groups (Aluwihare et al., 2005; McCarthy et al., 1997), as does a significant portion of recoverable LMW SPE-DON (Broek et al., Submitted), CSI-AA has promise for investigating formation and degradation processes of marine DON. Additionally, dissolved proteinaceous material has radiocarbon ages older than that of ocean mixing, implying these molecules are long-lived in the ocean (Loh et al., 2004).

δ^{15}N-AA fractionation provides specific information regarding OM transformation processes and "baseline" inorganic δ^{15}N source. For example, metazoan trophic position (TP) can be calculated based on a predicable relationship between AA which fractionate with trophic transfer (trophic AA, Tr-AA) and those which do not (source AA, Src-AA) (Chikaraishi et al., 2009). Additionally, δ^{15}N-AA data provide multiple ways of quantifying heterotrophic bacterial resynthesis mechanisms (Ohkouchi et al., 2017; Yamaguchi et al., 2017). Some examples include a TP equation specific for heterotrophic protozoa (Décima et al., 2017), the predictable fractionation of all AA which occurs during extracellular protein hydrolysis (Hannides et al., 2013; Yamaguchi & McCarthy, 2018), and the ΣV parameter, which reflects the scattering of trophic δ^{15}N-AA values during heterotrophic bacterial resynthesis (McCarthy et al., 2007). Finally, the δ^{15}N value of phenylalanine (δ^{15}N$_{\text{Phe}}$) does not undergo isotopic fractionation during metazoan or protozoan trophic transfer (Chikaraishi et al., 2009; Décima et al., 2017) or microbial degradation (Fogel & Tuross, 1999; Yamaguchi & McCarthy, 2018), and as such provides an excellent proxy for the δ^{15}N value of inorganic N source (Batista et al., 2014; Sherwood et al., 2014; Vokhshoori & McCarthy, 2014).

Despite the wide information potential for DON source and transformation

mechanisms, previous CSI-AA measurements of marine DON are incredibly limited and are restricted to HMW DON. The pioneering δ^{15}N-AA measurements of HMW DON demonstrated promise for this technique but suffered from poor chromatography and unresolved AA due to technology at the time (McCarthy et al., 2007). More recently, δ^{15}N-AA analysis of three HMW DON samples from the North Pacific Subtropical Gyre indicated novel interpretations for HMW DON source and cycling (Yamaguchi & McCarthy, 2018). Based on δ^{15}N-AA patterns, the authors suggested that HMW DON in the surface ocean is derived from direct heterotrophic bacterial sources while mesopelagic HMW DON derived from heterotrophic resynthesis of mesopelagic PON. While novel results, this study was restricted to only three HMW DON samples ranging surface to 915 m in one ocean basin. Additionally, and perhaps most importantly, this study was limited to HMW DON, representing the more rapidly cycling, semi-labile pool of marine DON. To understand the refractory nature of most marine DON, the older, more representative LMW DON pool must be directly investigated.

In this study, we utilize CSI-AA as a novel, nitrogen-specific tool to make detailed δ^{15}N measurements of HMW and LMW SPE-DON from the surface to 2500m deep at both the Hawaii Ocean Time Series (HOT) and Bermuda Atlantic Time Series (BATS). By coupling measurements of both HMW and LMW SPE-DON for the first time, we aim to directly assess size-age-reactivity relationships within the DON pool. These first CSI-AA results of LMW DON indicate LMW proteinaceous material appears surprisingly autotrophic and may be less "degraded" than HMW DON from the same location. We hypothesize novel production and transformation mechanisms for LMW SPE-DON, which, compared with our findings for HMW DON, suggest proteinaceous HMW and LMW SPE-DON cycle completely independently.

3.3 Methods

3.3.1 DOM sample collection and molecular weight isolation

HMW and LMW SPE-DON samples were collected in the North Pacific Subtropical Gyre at HOT Station ALOHA (22°45'N, 158°00'W) aboard the *R/V Kilo Moana* in August 2014 and May 2015, and in the North Atlantic Subtropical Gyre at BATS station (31°40'N, 64°10'W) aboard the *R/V Atlantic Explorer* in May 2016 (Broek et al. 2017). Surface samples were collected via underway sampling systems at 7.5 m water depth on the *R/V Kilo Moana*, and at 2 m water depth aboard the *R/V Atlantic Explorer*. Large volume subsurface samples (~ 1000 L to 4000 L) were collected via CTD casts at 400 m, 850 m, and 2500 m. All seawater was prefiltered through 53 µm Nitex mesh and then pumped through 0.2 µm cartridge filters. Subsamples were then collected for TOC and TON analyses in pre-combusted glass vials.

HMW UDOM and LMW SPE-DOM fractions were separately isolated as described previously (Broek et al., 2017). Briefly, HMW DOM was concentrated using large volume tangential-flow ultrafiltration using four spiral wound PES UF membranes with a MW cut off of 2.5 k Da (GE Osmonics) and a concentration factor of ~ 1000. LMW SPE-DOM was subsequently isolated via solid phase extraction (SPE) of the ultrafiltration permeate (Agilent Bondesil PPL). After desalting via diafiltration (for HMW) and rinsing (for LMW), both fractions were lyophilized and stored as dry powder until analysis. In addition to the above cruise dates, surface subsamples for total DON were also collected at HOT Station ALOHA in February 2014 and at BATS in August 2015. Full details of the sampling procedure, DOM isolation protocols, and mass balance and isotopic examinations of LMW SPE-DOM versus total LMW SPE-DOM are described in Broek et al. (2017).

3.3.2   Bulk isotopic analyses

δ^{15}N values of nitrate and TDON were measured in the Knapp lab at Florida State University via persulfate oxidation and the denitrifier method (Knapp et al., 2005; Sigman et al., 2001). $\delta^{15}NO_3^-$ was measured only for samples with $[NO_3^-]$ > 1.0 µM. TDON δ^{15}N were measured only in the upper 250 m where $[NO_3^-] < 3$ µM. Total dissolved nitrogen (TDN) was oxided to NO_3^- via persulfate oxidation, acidified, and the δ^{15}N value was calculated via mass balanced by subtracting the product of the δ^{15}N and $[NO_3^-]$ from the δ^{15}N and [TDN] and correcting for the isotope effect of the persulfate oxidizing reagent blank (Knapp et al., 2005). Duplicate measurements were made of each sample. The average standard deviation of DON δ^{15}N values, which represents the propagated error of DON and NO_3^- δ^{15}N measurements, was $\pm$ 0.2‰. δ^{15}N is reported relative to N_2 in air: δ^{15}N (‰) vs. air $= [(^{15}N/^{14}N)_{sample}/ \ (^{15}N/^{14}N)_{air}]$-1 x 1000.

Bulk δ^{15}N measurements of HMW and LMW SPE-DOM samples were measured at the University of California, Santa Cruz Stable Isotope Laboratory (UCSC-SIL) via elemental analyzer isotope ratio mass spectrometry (EA-IRMS), using a Carlo Erba CHNS-O EA1108-elemetnal analyzer interfaced to a ConFlo III device and ThermoFinnigan Delta Plus XP isotope ratio mass spectrometer (Thermo Fisher Scientific). All bulk isotope samples were analyzed in triplicate and values corrected according to UCSC-SIL standard protocols. Analytical error associated with this measurement was $\pm$ 0.2‰ for HMW and LMW SPE-DOM.

3.3.3   Sample preparation and HPLC purification of HMW and LMW SPE-DON for CSI-AA

DON samples were subjected to intensive upstream purification prior to δ^{15}N-AA measurement, following protocols first described in Yamaguchi & McCarthy

(2018). Briefly, HMW and LMW SPE-DOM subsamples were hydrolyzed in liquid phase (6 N HCl, 20 hrs) according to standard procedures (Calleja et al., 2013; Silfer et al., 1991). A norleucine internal standard was added to each sample to monitor procedural losses and isotopic values. Hydrolysates were dried under N_2 gas at 60° C, redissolved in 0.1 M HCl and filtered with a 0.2 µm GFF filter. AA were then first purified using cation-exchange chromatography with Bio-Rad AG50W-X8 resin (200-400 mesh) according to Takano et al. (2010). Samples were reprotonated with 0.2 N HCl by heating at 110° C for 5 minutes. To further isolate target AA for precise δ^{15}N-AA measurement within the highly complex DOM hydrolysate matrix, individual AA were then separated via high pressure liquid chromatography (HPLC; Shimadzu Scientific Instruments, Inc.) coupled with an evaporative light scattering detector (ELSD-LT II, Sedex 85LT; SEDERE). AA were collected from multiple separate chromatography regions based on HPLC retention times for AA standards to exclude major impurities according to Yamaguchi and McCarthy 2018. The separate AA groupings were then combined, and the solvent removed using a Jouan centrifugal evaporator at 60° C.

3.3.4 Compound specific isotope analysis of amino acids

δ^{15}N-AA measurements were made via GC-IRMS, using standard protocols in the McCarthy Lab (McCarthy et al., 2013). Briefly, Trifluoroacetyl isopropyl ester (TFAiP) derivatives were prepared after Silfer et al. (1991), and AA derivatives were purified using liquid-liquid extraction following Ueda et al. (1989). Samples were then stored at -20° C in TFA/ethyl acetate until GC-IRMS analysis. Compound specific isotopic analyses were made at UCSC-SIL using Thermo Trace Ultra gas chromatograph coupled with a Finnigan MAT DeltaPlus XL IRMS. Samples were dried under N_2 gas and redissolved in ethyl acetate. AA deriva-

tives were separated on a BPX-5 column (60m x 0.32 mm, 1.0 μm film thickness). Samples were injected in duplicates or triplicates unless noted in the text. Instrument performance was monitored using both external and internal AA standards, and directly measured δ^{15}N-AA values were corrected based on bracketing external standards (McCarthy et al., 2007). Based on this protocol a total of nine AA could be reproducibly measured in both HMW and LMW SPE-DON hydrolysates above a 100 mV threshold (see Fig. S3.2 for chromatograms): Alanine (Ala), Serine (Ser), Glycine (Gly), Theronine (Thr), Valine (Val), Isoleucine (Ile), Leucine, (Leu), Aspartic Acid + Asparagine (Asx), Glutamic Acid + Glutamine (Glx), with these latter combined peaks for Asx and Glx resulting from cleavage of terminal amines in glutamine (Gln) and asparagine (Asn) during hydrolysis. In addition, Phenylalanine (Phe), Tyrosine (Tyr) and Lysine (Lys) were present in most samples but δ^{15}N values could not reproducibly be measured in all samples due to low concentration (Table S3.2). The standard deviation of triplicate or duplicate measures for HMW and LMW SPE-DON averaged $\pm$ 0.65‰ for δ^{15}N-AA measurements; analytical variation for each individual AA is given in Table S3.2.

3.3.5 Equations, parameters, and data analysis

$\delta^{15}\textbf{N}_{\textbf{Bulk}}$ refers to the δ^{15}N value of all ON in total, HMW, or LMW SPE-DON. $\delta^{15}\textbf{N}_{\textbf{THAA}}$ (THAA = total hydrolysable AA) represents the average δ^{15}N of total proteinaceous material and is calculated as the mole percent weighted sum of δ^{15}N-AA values: $\delta^{15}N_{THAA} = \Sigma\ (\delta^{15}N_{AA} * \text{mol\%}_{AA})$. The average propagated error of $\delta^{15}N_{THAA}$ for our entire sample set is $\pm$ 0.2‰. $\delta^{15}\textbf{N}_{\textbf{Protein - Bulk}}$ reflects the offset in δ^{15}N values of total proteinaceous material ($\delta^{15}N_{THAA}$) and total ON and is calculated as $\delta^{15}N_{THAA} - \delta^{15}N_{Bulk}$. The δ^{15}N value of all ON besides AA ($\delta^{15}\textbf{N}_{\textbf{Other}}$) was calculated according to Yamaguchi and McCarthy (2018)

by isotope mass balance: $\delta^{15}N_{Other} = [(\delta^{15}N_{Bulk} - \delta^{15}N_{THAA}) * (\%ON_{AA}/100)]$ / $(1 - \%ON_{AA} / 100)$, where $\delta^{15}N_{Bulk}$ is the $\delta^{15}N$ value of total dissolved organic nitrogen for that size fraction and $\%ON_{AA}$ is the nitrogen normalized yield of measured AA. The average propagated error of $\delta^{15}N_{Other}$ is $\pm 1.0‰$. The "classic" trophic position **($TP_{Classic}$)** is calculated according to Chikaraishi et al. (2009): $TP_{Classic} = (\delta^{15}N_{Glx} - \delta^{15}N_{Phe} - 3.4) / (7.16 + 1)$. The protozoan trophic position **($TP_{Protist}$)** is calculated according to Décima et al. (2017): $TP_{Protist} = (\delta^{15}N_{Ala} - \delta^{15}N_{Phe} - 3) / (5.6 + 1)$. **The ΣV parameter**, quantifying the "scattering" of tropic $\delta^{15}N$ values with heterotrophic resynthesis, was calculated according to McCarthy et al. (2007) as follows: $\Sigma V = (1/n)* \Sigma Abs_{(\chi i)}$, where χ_i is the offset in $\delta^{15}N$ of each trophic AA from the average $\delta^{15}N$ of all trophic AA.

All calculations were done in Microsoft Excel or R Studio version 4.0.3. Data was tested for normality using a Shapiro Wilk Test and visually using QQ plots, and parametric statistical tests with a 95% confidence interval were used unless otherwise noted. Due to the small sample size, for most statistical tests samples from both ocean basins were grouped by size fraction for better statistical power.

3.3.6 Terminology and definitions

DOM size fraction classification: We primarily refer to the individually isolated size fractions by MW and use the terms **"HMW DOM/N"** and **"LMW SPE-DOM/N"** to refer to the selectively isolated size fractions described above. However, they also represent different age/reactivity classes. Thus, **"HMW"** here refers to ultrafiltered DOM 0.2 µm to 2.5 kDa (Broek et al., 2017) with radiocarbon ages of surface HMW DOM were 530 years at BATS and averaged 295 years at HOT (spring and summer samples). At 2500m, HMW DOM was 2475 years at BATS and 3595 years at HOT (Broek et al., 2017, Broek et al., 2020).

"LMW" refers to solid phase (SPE) extracted DOM from the permeate of HMW DOM (smaller than 2.5 kDa) with radiocarbon ages at the surface of 3260 years at BATS and 3310 years at HOT, and at 2500m of 4420 years at BATS and 6860 years at HOT (Broek et al., 2017, Broek et al., 2020).

Degradation and mechanisms: Regarding bacterial consumption and alteration of organic nitrogen, we use **"heterotrophic resynthesis"** to refer to AA synthesized within a bacterial cell during heterotrophy, presumably from external AA, as tracked by the ΣV parameter (McCarthy et al., 2007). ***de novo* synthesis** refers to AA synthesized from autotrophy using mineral N sources. **External enzyme hydrolysis** of AA refers to hydrolysis and selective removal of AA occurring entirely outside a bacterial cell, typically accompanied by isotopic fractionation (Hannides et al., 2013; Yamaguchi & McCarthy, 2018). Finally, in discussing common AA degradation parameters (e.g., DI index) we use **"degradation"** in the mechanistically non-specific way it is commonly used in the literature, to refer to the totality of processes potentially involved in bacterial heterotrophy. This includes structural alteration of molecules, selective or non-selective removal of biomolecules outside a bacterial cell, as well as respiration within a bacterial cell.

CSI-AA groupings: We group and discuss individual AA according to common δ^{15}N classifications in literature, which are based on δ^{15}N relative fractionation with trophic transfer (recently reviewed by McMahon & McCarthy (2016)). Measured trophic AA (Tr-AA) included Glx, Asx, Ala, Ile, Leu, Val, and Pro. Measured source AA (Src-AA) include Phe, Tyr and Lys (Chikaraishi et al., 2007; McClelland & Montoya, 2002). Gly and Ser are no longer considered to be Src-AA and are now best classified as "intermediate" AA (Int-AA). Thr is an AA with unique δ^{15}N systematics and does not have a universal classification, however, is labeled here as "metabolic" (Met-AA) following Germain et al. (2013)

89

and McCarthy et al. (2013).

3.4 Results

3.4.1 DON and AA recovery in MW size fractions

Extensive recovery data for the HMW and LMW SPE-DON size fractions is given in Broek et al. (2017), Broek (2019) and Ianiri et al. (Submitted). The relevant data is summarized here. For the samples investigated in this study, the percent of total OC recovered in our HMW size fraction was 11.1% ± 3.7%, while the percent of total OC recovered in the LMW SPE-DOM size fraction was 26.8% ± 8.3%. The percent of total ON recovered in HMW DON was 12.5% ± 4.8% and in LMW SPE-DON was 13.9% ± 7.2%. The percent of total AA recovered in HMW DON was 25.6% ± 8.8% and in LMW SPE-DON was 11.7% ± 4.9%.

3.4.2 δ^{15}N of total DON and NO_3^-

We report $\delta^{15}N_{\text{Bulk}}$ values of total DON from the surface ocean at HOT and between surface to 250 m at BATS (Fig. 3.1, Table S3.1). At HOT, $\delta^{15}N_{\text{Bulk}}$ values of surface, total DON averaged 3.7‰ ± 0.2‰ across two cruises. At BATS, $\delta^{15}N_{\text{Bulk}}$ values of surface, total DON were lower than at HOT for the same depth, averaging 3.0‰ ± 0.1‰ across two cruises. Below the surface, total DON $\delta^{15}N_{\text{Bulk}}$ values at BATS represent measurements from only one cruise. Between 25 m and 250 m at BATS, total DON $\delta^{15}N_{\text{Bulk}}$ ranged from 2.3 ± 0.2‰ to 3.2 ± 0.2‰, though there was no clear depth trend. $\delta^{15}N$ of NO_3^- at BATS was consistently 2.3‰ from 150 m to 250 m (the depth range for which NO_3^- could be measured in this study) (Fig. 3.1).

3.4.3 δ^{15}N of ON size fractions

δ^{15}N$_{\text{Bulk}}$ values of HMW and LMW SPE-DON are reported elsewhere (Broek et al., 2017, Submitted), however, are briefly summarized here to provide context for our compound specific δ^{15}N measurements (Fig. 3.1, Table S3.1). At HOT, δ^{15}N$_{\text{Bulk}}$ values of HMW DON measured over two cruises averaged 6.7 $\pm$ 0.1‰ while LMW SPE-DON δ^{15}N$_{\text{Bulk}}$ values were lower and averaged 3.5 $\pm$ 0.1‰. Similarly, at BATS, δ^{15}N$_{\text{Bulk}}$ of HMW DON measured over two cruises averaged 5.4 $\pm$ 0.2‰ while LMW SPE-DON δ^{15}N$_{\text{Bulk}}$ values were lower and averaged 3.0 $\pm$ 0.2‰. When averaged over both ocean basins and all depths, δ^{15}N$_{\text{Bulk}}$ of LMW SPE-DON was significantly lower than δ^{15}N$_{\text{Bulk}}$ of HMW DON by an average of 2.8‰ (Welch's two-sample t-test, p $<$ 0.001).

δ^{15}N$_{\text{THAA}}$ is a proxy for the average δ^{15}N value of total proteinaceous material (Section 3.3.6). At HOT, δ^{15}N$_{\text{THAA}}$ of HMW DON increased from average surface values of 6.1 $\pm$ 0.07‰ to average subsurface (400 m – 2500 m) values of 10.3 $\pm$ 0.1‰ (Fig. 3.2A, Table S3.2). δ^{15}N$_{\text{THAA}}$ of LMW SPE-DON at HOT did not change significantly with depth and averaged 5.8‰ $\pm$ 0.1‰ throughout the water column. At BATS, δ^{15}N$_{\text{THAA}}$ of HMW DON also increased from 4.7‰ $\pm$ 0.2‰ in the surface and to an average subsurface (400 m – 2500 m) value of 8.8‰ $\pm$ 0.1‰. δ^{15}N$_{\text{THAA}}$ of LMW SPE-DON at BATS averaged 4.6‰ $\pm$ 0.2‰ throughout the water column. Similar to δ^{15}N$_{\text{Bulk}}$ values, δ^{15}N$_{\text{THAA}}$ of HMW DON was significantly greater than LMW SPE-DON by 2.9‰ when averaged across both ocean basins and throughout the water column (Welch's two-sample t-test, p $=$ 0.0018).

On average, δ^{15}N$_{\text{THAA}}$ of HMW and LMW SPE-DON is greater than δ^{15}N$_{\text{Bulk}}$, resulting in mostly positive δ^{15}N$_{\text{Protein - Bulk}}$ values (Fig. 3.2B). However, the magnitude of the offset varies with depth, size fraction, and ocean basin. In the

surface, $\delta^{15}N_{THAA}$ and $\delta^{15}N_{Bulk}$ of HMW DON are very similar at both HOT and BATS, with $\delta^{15}N_{Protein\,-\,Bulk}$ values averaging -0.3‰ $\pm$ 0.2‰ at HOT and to 0.6‰ $\pm$ 0.2‰ at BATS. In both ocean basins, $\delta^{15}N_{Protein\,-\,Bulk}$ of HMW DON increases dramatically between the surface and 400 m, with average subsurface $\delta^{15}N_{Protein\,-\,Bulk}$ values of 3.4‰ $\pm$ 0.2‰ at HOT and 2.9‰ $\pm$ 0..25‰ at BATS. In the LMW SPE-DON pool, $\delta^{15}N_{Protein\,-\,Bulk}$ at both HOT and BATS was also lowest in the surface than at subsurface depths, though the increase with depth was less than observed in the HMW DON pool. Across both ocean basins, surface $\delta^{15}N_{Protein\,-\,Bulk}$ values averaging 1.8‰ $\pm$ 0.2‰ at HOT and 1.3‰ $\pm$ 0.1‰ at BATS while subsurface $\delta^{15}N_{Protein\,-\,Bulk}$ values averaged 2.5‰ $\pm$ 0.2‰ at HOT and 1.8‰ $\pm$ 0.2‰ at BATS. For both size fractions, subsurface $\delta^{15}N_{Protein\,-\,Bulk}$ values were greater at HOT than BATS.

$\delta^{15}N_{Other}$, or the $\delta^{15}N$ value of all N besides THAA (calculated via mass balance, Section 3.3.6), was within error of $\delta^{15}N_{Bulk}$ values at all depths and in both ocean basins for both HMW and LMW SPE-DON (Fig. S3.3, Table S3.2). This reflects the low percentage of total ON which is recoverable as THAA ($\leq$ 10%).

3.4.4 δ^{15}N CSI-AA patterns

δ^{15}N-AA patterns of HMW DON in both basins are similar to those previously reported in the Pacific Ocean for comparable depths (Fig. 3.3, Table S3.3) (McCarthy et al., 2007; Yamaguchi & McCarthy, 2018). δ^{15}N values of most AA in HMW DON were greater at HOT than BATS when comparing between the same depths, consistent with the significantly greater $\delta^{15}N_{THAA}$ values at HOT compared to BATS. HMW DON δ^{15}N-AA values of almost all AA were greater in the subsurface ($\geq$ 400 m) compared to the surface at both HOT and BATS, though the magnitude of the deep – surface offset varied for individual AA (Fig.

92

3.4). The average δ^{15}N-AA deep – surface offset of all AA across all depths at HOT was 2.7‰ ± 2.0‰ while at BATS the average offset was 2.6‰ ± 1.4‰. Ala was notable in that it consistently had the greatest δ^{15}N-AA value for all HMW DON samples, in contrast to typical primary production patterns, with maximum values at 2500m of 17.4‰ ± 0.5‰ at BATS and 18.5‰ ± 0.2‰ at HOT.

In contrast, the first LMW SPE-DON δ^{15}N-AA data were completely different from those of HMW DON in both ocean basins (Fig. 3.3, Table S3.3). δ^{15}N-AA values were generally greater at HOT than BATS for comparable depths, again consistent with trends in δ^{15}N$_{THAA}$. However, individual AA deep – surface offsets were smaller in LMW SPE-DON compared to those observed for HMW DON (Fig. 3.4). At BATS, most LMW SPE-DON AA had greater δ^{15}N-AA values at depth ($\geq$ 400 m) compared to the surface, with an average deep – surface across all depths of 1.1‰ ± 0.8‰. At HOT, there was more variation in the magnitude of individual offsets, with an average offset of -0.5‰ ± 0.9‰ for all depths.

To directly compare δ^{15}N-AA patterns between ocean basins and size fractions, δ^{15}N-AA values were also normalized to THAA (δ^{15}N-AA – δ^{15}N$_{THAA}$) (Fig. S3.4). Normalization to THAA does not alter the δ^{15}N-AA pattern but shifts the measured δ^{15}N-AA values to the same scale, allowing a direct comparison regardless of the δ^{15}N value of the inorganic N source. Normalized δ^{15}N-AA patterns of HMW DON showed some variation with depth, while LMW SPE-DON δ^{15}N-AA normalized were more similar in both ocean basins and at all depths.

3.4.5 δ^{15}N-AA parameters for N source and resynthesis

The ΣV parameter for heterotrophic bacterial resynthesis had an average value for HMW DON from all depths in both ocean basins of 2.4‰ ± 0.1‰, significantly greater than the average ΣV of LMW SPE-DON, 1.3‰ ± 0.1‰ (Mann-Whitney

U-Test, p < 0.0001) (Fig. 3.5A, Table S3.2). Within each size fraction, ΣV was greater at HOT compared to BATS at most depths. $TP_{Classic}$, a proxy for metazoan trophic position, was between the expected values for primary producers (1.0) and primary consumers (2.0) for both HMW and LMW SPE-DON at all depths and in both ocean basins (Fig. 3.5B, Table S3.2). The average $TP_{Classic}$ was not significantly different between the two size fractions (HMW DON: 1.3 $\pm$ 0.1, LMW SPE-DON: 1.3 $\pm$ 0.1, Welch's two-sample *t*-test, p = 0.91). $TP_{Protist}$, a proxy for protozoan trophic position, was more than one full trophic position higher than $TP_{Classic}$ for HMW DON, averaging 2.6 $\pm$ 0.1 at all depths and in both ocean basins (Fig. 3.5C, Table S3.2). The average $TP_{Protist}$ of LMW SPE-DON was similar to $TP_{Classic}$ of LMW SPE-DON, averaging 1.5 $\pm$ 0.1. $TP_{Protist}$ of LMW SPE-DON was significantly less then HMW DON by 1.1 (Welch's two-sample *t*-test, p <0.0001).

$\delta^{15}N_{Phe}$, a proxy for $\delta^{15}N$ of the baseline inorganic N source, followed approximately similar patterns at HOT and BATS (Fig. 3.6, Table S3.3). At HOT, $\delta^{15}N_{Phe}$ of surface, HMW DON was not significantly different in the spring and summer, measuring 1.5‰ $\pm$ 0.9‰ and 1.3 $\pm$ 0.8‰, respectively (Fig. 3.6A, Fig. S3.5). $\delta^{15}N_{Phe}$ of HMW DON increased to a maximum value of 7.0‰ (single injection) at 850m. $\delta^{15}N_{Phe}$ of HMW DON at 2500 m could not be resolved due to low concentrations. In the LMW SPE-DON pool, $\delta^{15}N_{Phe}$ of surface DON at HOT was greater in the spring than the summer, measuring 3.3‰ $\pm$ 0.5‰ and 1.9‰ $\pm$ 0.7‰, respectively. Subsurface $\delta^{15}N_{Phe}$ values of LMW SPE-DON at were significantly lower than subsurface $\delta^{15}N_{Phe}$ values of HMW DON. A minimum $\delta^{15}N_{Phe}$ value of LMW SPE-DON at HOT was measured at 400 m (0.9‰ $\pm$ 1.4‰), while at both 850 m and 2500 m $\delta^{15}N_{Phe}$ was 2.4‰.

At BATS, $\delta^{15}N_{Phe}$ of HMW DON in the surface was 1.8‰ $\pm$ 0.4‰, within

error of surface HMW DON $\delta^{15}N_{Phe}$ values at HOT (Fig. 3.6B). $\delta^{15}N_{Phe}$ of HMW DON increased to a maximum of 4.6‰ ± 0.8‰ at 400 m, then decreased again with depth. $\delta^{15}N_{Phe}$ of LMW SPE-DON in the surface was 1.4‰ ± 0.6‰, within error to surface HMW DON $\delta^{15}N_{Phe}$ measured in the summer. Similar to HOT, $\delta^{15}N_{Phe}$ of LMW SPE-DON at 400 m was less than $\delta^{15}N_{Phe}$ of HMW DON at the same depth, measuring 2.4‰ ± 1.7‰. LMW SPE-DON $\delta^{15}N_{Phe}$ values could not be resolved at deeper depths at BATS due to low concentrations. Including data from both ocean basins, no significant relationship in HMW or LMW SPE-DON is observed between $\delta^{15}N_{Phe}$ and ΣV (HMW DON: $r^2 = 0.23$, p = 0.23, LMW SPE-DON: $r^2 = 0.23$, p = 0.28) or $\delta^{15}N_{Phe}$ and $TP_{Protist}$ (HMW DON: $r^2 = 0.001$, p = 0.95, LMW SPE-DON: $r^2 = 0.23$, p = 0.27) (Fig. S3.6).

3.5 Discussion

3.5.1 New implications for novel N compound classes in HMW and LMW SPE-DON

Much of our understanding regarding DON composition is derived from early ^{15}N NMR studies, which indicated HMW DON is almost entirely composed of amide functional groups (Aluwihare et al., 2005; McCarthy et al., 1997). Based on these results, it has generally been assumed that non-recoverable AA makes up most of DON, with amino sugar (AS) contributions in the surface (Aluwihare et al., 2005). However, this has presented a long-standing conundrum in DON research: if DON is supposedly composed of labile biomolecules, why is a substantial portion unavailable to biological utilization? The new isotope data presented here for $\delta^{15}N_{Bulk}$, $\delta^{15}N_{THAA}$, and $\delta^{15}N_{Other}$ (Section 3.3.6) strongly contradict past assumptions of an entirely AA and AS composition, and instead indicate that additional unknown N compound classes may make up an important fraction of

both HMW and LMW SPE-DON. In the discussion sections below, by comparing the δ^{15}N values of these different N compound classes, we hypothesize alternative sources for non-AA N in these two size fractions.

"Other" HMW DON: novel amide molecules or heterocyclic N?

If almost all HMW amide was composed of proteinaceous material, as has long been assumed (at least in the subsurface), then we would expect the δ^{15}N$_{\text{THAA}}$ parameter, a proxy for the δ^{15}N value of proteinaceous ON (section 2.5) to directly reflect the δ^{15}N$_{\text{Bulk}}$ value of HMW material. However, the positive δ^{15}N$_{\text{Protein − Bulk}}$ values we report at all subsurface depths (Fig. 3.2B) implies an additional, non-AA pool of N with a unique δ^{15}N signature and cycling processes.

These subsurface δ^{15}N$_{\text{Protein − Bulk}}$ values we observed (2.6‰ to 3.7‰) are consistent with previous CSI-AA studies spanning a range of matrices, including autotrophic and heterotrophic cells (δ^{15}N$_{\text{Protein − Bulk}}$ ∼ 2.3‰ to 3.5‰: Macko et al., 1987, McCarthy et al., 2013, Pan et al., 2008, Batista et al., 2014), sedimentary ON (δ^{15}N$_{\text{Protein − Bulk}}$ ∼ 2.9‰ to 4.6‰: Batista et al., 2014), HMW DON (δ^{15}N$_{\text{Protein − Bulk}}$ ∼ 2.3 ± 1‰: Yamaguchi et al., 2017), and suspended PON (δ^{15}N$_{\text{Protein − Bulk}}$ ∼ 3.6 ± 1.2‰: Yamaguchi et al., 2017). Previous work suggested this constant offset can be explained by proteinaceous material being generally enriched compared to other nitrogenous biomolecules (Macko et al., 1987, Batista et al., 2014). Based on ^{15}N NMR analyses indicating most DON (Aluwihare et al., 2005; McCarthy et al., 1997), sedimentary ON (Knicker & Hatcher, 1997), and PON (Knicker, 2000) is amide, it was generally assumed the "other nitrogenous biomolecules" in all the above matrices were AS. A substantial AS contribution to suspended PON was also hypothesized by Yamaguchi et al. (2018) based on δ^{15}N values of different N compound classes, which found δ^{15}N$_{\text{THAA}}$ of suspended PON

increased significantly between surface and subsurface depths while $\delta^{15}N_{\text{Other}}$ was relatively constant. The authors hypothesized this was because external enzyme hydrolysis increased $\delta^{15}N_{\text{THAA}}$ values but not the $\delta^{15}N$ of AS because there is no breakage of a C-N bond.

For the HMW DON pool specifically, while hydrolysable AS yields are generally $\leq 10\%$ of total ON (Benner and Kaiser 2003, Kaiser and Benner 2009), experiments measuring acetic acid produced via mild acid hydrolysis of HMW DON found a significant portion of AS may be resistant to acid hydrolysis, leading to revised estimates that up to half of surface HMW DON could be AS (Aluwihare et al., 2005). However, the $\delta^{15}N$ compound class data we report here for HMW DON do not support the hypothesis that significant AS contributions cause the $\delta^{15}N_{\text{Protein - Bulk}}$ offsets we observed. Instead, $\delta^{15}N_{\text{Protein - Bulk}}$ is near zero at the surface in both ocean basins, where AS contribution to HMW DON is expected to be greatest (Aluwihare et al., 2005; Kaiser & Benner, 2009). These very similar $\delta^{15}N_{\text{THAA}}$ and $\delta^{15}N_{\text{Bulk}}$ values we observe in the surface ocean suggest that either the $\delta^{15}N$ value of AS is comparable to $\delta^{15}N_{\text{THAA}}$, or that abundances of AS are lower than hypothesized based on mild acid hydrolysis experiments. Lower abundances of AS is also consistent with measured AS recoveries via acid hydrolysis, which typically range from $\sim 60\%$ to 80% (Benner & Kaiser, 2003), suggesting AS concentrations are likely close to those recovered at the molecular level. Thus, while AS may contribute to the "other" HMW ON, in contrast to previous hypotheses for sediments and PON, it appears they cannot account for the $\delta^{15}N_{\text{Protein - Bulk}}$ values we observe at all subsurface depths.

An additional possibility consistent with both N isotopic and NMR data is a significant subsurface contribution of unknown amide-containing structures in the HMW DON pool. Recently, advanced ^{15}N NMR spectroscopy of Synechococcus-

97

derived HMW DON revealed a range of previously unidentified amide-N-containing molecules, including N-methyl amides, primary amides, and novel AS structures (Cao et al., 2017). The presence of these compounds in cyanobacterial-derived HMW DON suggests a range of "other-amide" molecules that could also contribute marine HMW DON. If these cyanobacterially-derived compounds, or their degradation products, have lighter δ^{15}N values than δ^{15}N$_{THAA}$ and persist longer than other N-containing biomolecules, they could be responsible for the δ^{15}N$_{Protein - Bulk}$ values we observed at depth. While we acknowledge the importance of such compounds is speculative, at the same time, the identification of such compounds in culture, coupled with the importance of cyanobacterial production in all open ocean regions, makes this a clearly testable hypothesis for future research.

Finally, we note that ^{15}N CP/MAS NMR analyses of Pacific Ocean indicate HMW DON may have a small contribution of heterocyclic N (Aluwihare et al., 2005; Broek et al., 2020; McCarthy et al., 1997). While past work indicated these structures only account for a minor portion of total HMW DON, CP/MAS NMR is known to significantly underestimate unsaturated N moieties (Smernik & Baldock, 2005). Further, heterocyclic N structures derived from marine primary producers are estimated to have a lower δ^{15}N value than proteinaceous material (Higgins et al., 2010), meaning a greater contribution of these compounds in the deep ocean could contribute to the lower δ^{15}N$_{Bulk}$ vs. δ^{15}N$_{THAA}$ values. If traditional CP/MAS NMR has substantially underestimated unsaturated nitrogen contribution to HMW DON, then these isotopic observations suggest this could account for at least some of δ^{15}N$_{Protein - Bulk}$ values we observed at subsurface depths.

Overall, the identity of this "other" N fractions is likely the key to understanding the sources and dynamics of most HMW DON In the ocean. While past work

suggested this material was mostly AS in the surface and proteinaceous material at subsurface depths, these new isotopic data now call this into question. The positive $\delta^{15}N_{\text{Protein - Bulk}}$ values at all subsurface depths strongly indicate non-THAA N is not proteinaceous. Additionally, the near zero $\delta^{15}N_{\text{Protein - Bulk}}$ in the surface indicate AS cannot account for the isotopic differences between $\delta^{15}N_{\text{THAA}}$ and $\delta^{15}N_{\text{Other}}$. Instead, it we suggest it more likely that either unknown amide compound classes or underestimated heterocyclic N could be major unrecognized components of HMW DON through the water column. Considering $\delta^{15}N_{\text{Protein - Bulk}}$ values are greatest in the deep ocean and older waters at HOT, this "other" N may represent some of the most persistent HMW DON and warrants further future study.

"Other" LMW SPE-DON: δ^{15}N of proteinaceous vs. heterocyclic LMW SPE-DON

In the LMW SPE-DON pool, the significant offsets we observed between $\delta^{15}N_{\text{THAA}}$ and $\delta^{15}N_{\text{Bulk}}$ (Fig. 3.2B) support recent new data suggesting a complete reevaluation of marine subsurface DON composition. The first ^{15}N CP/MAS NMR experiments of LMW SPE-DON indicated amide compounds represent $\sim$ 25% of total N, with novel heterocyclic N compounds representing the remaining $\sim$ 75% (Broek et al., Submitted). It is therefore likely that the offset in $\delta^{15}N_{\text{THAA}}$ and $\delta^{15}N_{\text{Bulk}}$ reflects the δ^{15}N of LMW SPE-DON compounds with heterocyclic N functionality. While the exact identity of this newly identified heterocyclic fraction has yet to be determined, heterocyclic ring structures are generally intrinsically stable, potentially leading to recalcitrance (Higgins et al., 2010, 2011; Knicker, 2004). Considering refractory molecules are also expected to accumulate over ocean circulation, the greater $\delta^{15}N_{\text{Protein - Bulk}}$ we observed at

HOT compared to BATS could be consistent with heterocyclic N compounds accumulating in the LMW SPE-DON pool. Compound specific δ^{15}N measurements targetting heterocyclic N molecules in marine DON will help unravel potential sources and formation mechanisms of these compounds which appear to dominate LMW RDON.

3.5.2 Baseline N sources to HMW and LMW SPE-DON

A major application of CSI-AA is the ability to determine the δ^{15}N value of the inorganic N source to proteinaceous ON, central to understanding marine DON production mechanisms and its role in marine nitrogen cycling. δ^{15}N$_{\text{Phe}}$ is commonly used as a proxy for baseline inorganic N source of proteinaceous material because it fractionates only minimally with trophic transfer or microbial resynthesis (Chikaraishi et al., 2009; Décima et al., 2017; McCarthy et al., 2007) and serves as a proxy for the δ^{15}N of inorganic N even in aged or degraded substrates (Batista et al., 2014; Sherwood et al., 2014; Yamaguchi & McCarthy, 2018). The lack of any significant relationships between δ^{15}N$_{\text{Phe}}$ and microbial biomarkers such as ΣV or TP$_{\text{Protist}}$ provides additional support that δ^{15}N$_{\text{Phe}}$ represents the δ^{15}N-AA of inorganic N source to proteinaceous HMW and LMW SPE-DON (Fig. S3.6).

HMW DON N source

Recently, a new interpration for surface production of HMW DON was suggested from CSI-AA data (Yamaguchi & McCarthy, 2018). The authors measured a single δ^{15}N$_{\text{Phe}}$ value for surface HMW DON collected in winter from the central Pacific Ocean of 5.2‰ ± 1.7‰. Based on this data, Yamaguchi & McCarthy (2018) suggested surface HMW DON is derived from production in the

deep chlorophyll maximum using relatively ^{15}N enriched nitrate, a new paradigm for HMW DON origin. This hypothesis also served to explain past elevated δ^{15}N values of HMW DOM across the Pacific Basin (Benner et al., 1997) and the lack of correlation between N$_2$ fixation and DOM δ^{15}N values (Knapp et al., 2011, 2012; Meador et al., 2007).

The surface δ^{15}N$_{\text{Phe}}$ values we observed for HMW DON in spring at BATS (1.8‰ ± 0.4‰) and spring and summer at HOT (1.5‰ ± 0.9‰ and 1.3‰ ± 0.8‰, respectively) (Fig. 3.6) are significantly lower than the previous δ^{15}N$_{\text{Phe}}$ measurement of surface HMW DON in winter (Fig. S3.5). It is possible the lower δ^{15}N$_{\text{Phe}}$ values we report here compared to previous work imply greater reliance on surface production fueled by N$_2$ fixation or recycled N during our sampling months. However, both fixed N$_2$ and surface PON have average source δ^{15}N values of -2‰ to 0‰ (Carpenter et al., 1997; Hannides et al., 2013, 2020; Yamaguchi & McCarthy, 2018), significantly lower than the surface δ^{15}N$_{\text{Phe}}$ values we observed. Additionally, previous work has found no relationship between N$_2$ fixation rates and the δ^{15}N value of total, HMW, or LMW SPE-DON, suggesting N$_2$ fixation is either not a major direct source of marine DON, or DON produced via N$_2$ fixation is rapidly utilized (Knapp et al., 2005, 2011, 2012; Meador et al., 2007).

Instead, we therefore hypothesize the difference in our δ^{15}N$_{\text{Phe}}$ values of HMW DON compared to those measured by Yamaguchi & McCarthy (2018) are due to seasonal differences in the δ^{15}N of NO$_3^-$. Seasonal increases in nitrogen fixation in subtropical gyres lead to lower δ^{15}N values of subsurface NO$_3^-$ (Casciotti et al., 2008). Between 150 – 250 m, summertime δ^{15}N of NO$_3^-$ ranges from 2‰ to 2.5‰ at BATS (Knapp et al., 2005) (Table S3.1, Fig. 3.6B) and 1.5‰ to 4‰ at HOT (Casciotti et al., 2008; Knapp et al., 2011). These values are within error of the

$\delta^{15}N_{Phe}$ values we report for surface HMW and LMW SPE-DON collected during the spring and summer at HOT and BATS (Fig. 3.6, Fig. S3.5). Overall, these new observations thus strongly support the hypothesis of Yamaguchi & McCarthy (2018), suggesting nitrate sourced from the deep chlorophyll maximum is main inorganic N source for proteinaceous HMW DON.

LMW SPE-DON N source

These first $\delta^{15}N_{Phe}$ measurements of LMW SPE-DON are within error of those reported for HMW DON (Fig. 3.6), suggesting the same paradigm outlined above for the HMW DON pool also applies to LMW proteinaceous material, despite evidence for different composition (Ianiri et al., Submitted). There are no previous CSI-AA measurements of LMW SPE-DON to compare with, so we cannot evaluate if there is a seasonal offset in $\delta^{15}N_{Phe}$ values of LMW SPE-DON. However, seasonal variation in the LMW SPE-DON size fraction would be more surprising than for HMW DON considering the old radiocarbon ages of all LMW SPE-DOM samples (~ 3500 years in the surface of both ocean basins). At the same time, elevated concentrations of Phe in surface LMW SPE-DON compared to those at depth (Ianiri et al., Submitted) suggests a substantial new input of Phe in the surface ocean which could potentially record a seasonal signature. Additionally, the higher $\delta^{15}N_{Phe}$ values we observe at HOT in the spring vs. summer ($\sim 1.4‰$ difference) provide some support for a seasonal signal in inorganic N source to this size fraction. Additional CSI-AA measurements of LMW SPE-DON in different seasons could help constrain these two possibilities. Moreover, radiocarbon measurements of N-containing compounds, such as AA, would be helpful in determining cycling rates of LMW proteinaceous material and if a seasonal signal is reasonable.

Overall, these first $\delta^{15}N_{Phe}$ measurements of LMW SPE-DON indicate a similar nitrogen source to proteinaceous HMW DON, implying production of proteinaceous LMW SPE-DON also occurs deeper in the water column primarily using nitrate. These findings have important implications for understanding and modeling open ocean nitrogen cycling, suggesting even in oligotrophic regions of the ocean nitrate supports production of semi-labile and persistent marine DON. Additionally, a similar N source to both size fractions implies the difference in $\delta^{15}N_{THAA}$ values must be due to different cycling processes rather than production from inorganic N with different $\delta^{15}N$ values. Still, as noted above, because $\delta^{15}N_{Phe}$ is a direct proxy for proteinaceous N source, it is unclear if these results can be extrapolated to other LMW N compound classes. $\delta^{15}N$ measurements of other N compound classes could aid in determining if there are similar N sources to the entire LMW SPE-DON pool.

3.5.3   Autotrophic vs. heterotrophic sources to HMW DON

HMW DON has long been observed to have substantially elevated $\delta^{15}N$ values compared to both local production and LMW SPE-DON material (Benner et al., 1997; Broek et al., Submitted; Knapp et al., 2012). However, to what degree this offset is connected to specific sources or fractionation with degradation remains unclear. One suggested possibility is that HMW DON is predominantly algal biosynthate but has undergone partial degradation with strong fractionation (Knapp et al., 2011). This theory for elevated $\delta^{15}N$ values of HMW DON is analogous to mechanisms suggested for the universal increase in upper ocean suspended PON $\delta^{15}N$ values between the surface and subsurface depths (Altabet, 1988, 1989). Subsequent CSI-AA analyses of suspended PON found that the increase in $\delta^{15}N$ values of PON is consistent with a microbial external enzyme hy-

drolysis mechanism (Hannides et al., 2013; Yamaguchi & McCarthy, 2018). The signature of this mechanism is nearly equivalent increases in all δ^{15}N-AA values (except for Lysine, which also has non-amide bond N), therefore preserving the algal-like δ^{15}N-AA pattern while also increasing δ^{15}N values. Thus, if a similar mechanism is responsible for the elevated δ^{15}N values of HMW DON compared to PON, a comparison of HMW DON and PON δ^{15}N-AA signatures should reflect this signature of external enzyme hydrolysis.

Contrary to this hypothesis, the first work directly comparing the CSI-AA pattern of a single surface HMW DON sample with those of suspended PON did not observe evidence of external enzyme hydrolysis, and instead reported CSI-AA patterns consistent with a direct bacterial origin (Yamaguchi & McCarthy, 2018). Our new data from both ocean basins strongly supports, and also expands, this interpretation. While the δ^{15}N-AA values of surface HMW DON are greater than surface suspended PON from the same region (Yamaguchi & McCarthy, 2018) by an average of 3.4‰, there is significant variation in individual δ^{15}N-AA offsets (data not shown), indicating external degradation of PON is not the primary mechanism of HMW DON formation in the surface ocean.

Instead, our CSI-AA measurements provide additional evidence for a direct bacterial source for HMW DON in the surface ocean. Specifically, the low $TP_{Classic}$ of HMW DON (~ 1.5) suggests classic metazoan trophic transfer is only a minor resynthesis mechanism of the HMW DON pool (Fig. 3.5B, Table S3.2). In contrast, the high ΣV (~ 2.3) indicates production via microbial resynthesis, supporting a direct microbial source (Fig. 3.5A, Table S3.2). A direct heterotrophic bacterial source is also consistent with previous work measuring D-AA content to HMW DON. HMW DON had significant contributions of D-AA, many of which are not observed in PON, indicating heterotrophic bacterial origins (Broek et

al., 2019; Ianiri et al., Submitted; Kaiser & Benner, 2008; McCarthy, 1998). Additionally, similar δ^{15}N values of D and L-Ala in HMW DON suggest a primarily bacterial source for both D and L-AA (Yamaguchi & McCarthy, 2018). Finally, resynthesis by heterotrophic bacteria increases δ^{15}N$_{\mathrm{THAA}}$ by $\sim 3‰ - 6‰$ (Calleja et al., 2013), meaning microbial production could also cause the elevated δ^{15}N values of HMW DON. Overall, these results strongly support the previous hypothesis that heterotrophic bacterial production appears to be the dominant production mechanism of surface, proteinaceous HMW DON. A surface bacterial source, rather than degradation of autotrophic-sourced particles, has important implications for understanding and modeling DON production and degradation in the upper ocean.

Linking protozoan heterotrophy and HMW DON production

In addition to a microbial source to HMW DON, we also find new evidence for a far more important role of protozoan feeding than has been previously recognized. The TP$_{\mathrm{Protist}}$ of HMW DON, based on δ^{15}N fractionation of Ala (Décima et al., 2017; Gutiérrez-Rodríguez et al., 2014), is one full trophic position greater than TP$_{\mathrm{Classic}}$ (Fig. 3.5), suggesting protozoan grazing is a previously overlooked microbial loop step for the HMW DON pool. However, δ^{15}N fractionation of Ala also occurs with heterotrophic bacterial resynthesis (Calleja et al., 2013) and results in an increase in ΣV values (Fig. S6), meaning it can be difficult to distinguish between these different microbial loop processes.

Our preferred hypothesis for how heterotrophic protozoa and bacteria could be interacting to produce HMW DON with the observed δ^{15}N-AA signals is that protozoan grazing on heterotrophic bacterial biomass produces HMW DON. Considering only a few eukaryotic organisms express enzymes to digest D-AA (Asano

& Lübbehüsen, 2000), protozoan grazing of heterotrophic bacteria would likely preferentially utilize L-AA. The HMW DON pool left behind would thus have increased D/L ratios and ΣV, consistent with the δ^{15}N-AA data we observed. It is estimated that protozoan grazing is the most significant contributor to marine DOM production, particularly in oligotrophic gyres (Nagata, 2000), and that protozoa consume up to 75% of heterotrophic bacteria biomass daily (Caron et al., 1991), both consistent with this theory. However, in this scenario HMW DON is not ingested and resynthesized by protozoa, but instead is the remainder from what protozoa do not consume. This means we would not expect to see evidence of an additional protozoan trophic position. If this is the case, the $TP_{Protist}$ of $\sim 2 - 2.5$ is actually due to heterotrophic bacterial resynthesis increasing the δ^{15}N value of Ala, rather than protozoan heterotrophy. Alternatively, it is possible HMW DON could be produced via sloppy feeding of mesozooplankton on protozoa, in which case the increased $TP_{Protist}$ values would be due to protozoan heterotrophy (Supplementary 3.7.1).

Considering the complex nature of heterotrophic bacterial metabolism, additional CSI-AA culture and field studies of microbial loop processes would help constrain δ^{15}N-AA patterns associated with heterotrophic metabolism and evaluate our hypothesis. Previous δ^{15}N-AA studies evaluating heterotrophic bacterial resynthesis (Calleja et al., 2013) and protozoan heterotrophy (Gutierrez-Rodriguez et al, 2014, Decima et al., 2017) are limited, and to our knowledge there is no previous work evaluating multiple microbial loop steps. Additionally, targeted analyses of δ^{15}N fractionation of D and L-Ala could help identify specific metabolic processes forming semi-labile HMW DON, as it appears to be a key diagnostic indicator of HMW DON resynthesis.

3.5.4   Unique source to mesopelagic HMW DON

Compared to surface HMW DON, subsurface (≥ 400 m) HMW DON in both ocean basins has a unique δ^{15}N-AA signature and greater δ^{15}N values of all AA (Fig. 3.4), resulting in a significant increase in δ^{15}N$_{\text{THAA}}$ between the surface and mesopelagic (Fig. 3.2A). As noted above, past work attributed the increase in suspended PON δ^{15}N-AA values between the surface and mesopelagic to external heterotrophic bacterial degradation after production in the surface ocean (Hannides et al., 2013; Yamaguchi & McCarthy, 2018). However, unlike PON, there is no known direct connection for surface and mesopelagic DON. Instead, our surface, mesopelagic and deep samples each represent unique water masses between which dissolved material is not expected to mix. Furthermore, while δ^{15}N-AA values of all AA are greater at subsurface depths than in the surface, the magnitude of this increase is variable for individual AA (Fig. 3.4). This variability is inconsistent with the uniform enrichment of all AA expected from external hydrolysis (Section 3.5.3). Finally, similar to all other AA, δ^{15}N$_{\text{Phe}}$ is greater at all subsurface depths compared to the surface (average increase of 5.1‰ at HOT and 2.3‰ at BATS). This difference, in absence of evidence for the external enzyme hydrolysis mechanisms, suggests independent sources to surface and deep HMW DON.

A similar increase in δ^{15}N-AA values of HMW DON with depth was also observed previously for two samples in the central Pacific (Yamaguchi & McCarthy, 2018). They suggested three hypotheses consistent with their data: 1) the δ^{15}N-AA signature of mesopelagic HMW DON could represent a globally averaged, background pool of HMW DON, 2) δ^{15}N-AA signatures of mesopelagic HMW DON could be "pre-formed," reflecting formation in the surface source waters of different source water masses, or 3) decoupling of surface and mesopelagic

semi-labile DON, with mesopelagic DON produced via heterotrophic resynthesis of more labile suspended PON. With our expanded dataset spanning multiple depths across two ocean basins, we can now assess these hypotheses.

Regarding the first hypothesis, if δ^{15}N-AA signatures of HMW DON represent a well-mixed, global average, then we would expect all samples from below the surface in both ocean basins to have the same δ^{15}N-AA values. While the ^{15}N-AA patterns of all our HMW DON samples are quite similar at all subsurface depths in both basins, individual δ^{15}N-AA values vary between ocean basins and with depth (Fig. 3.3, Fig. 3.4). This indicates HMW DON δ^{15}N-AA values do not represent well mixed global averages.

Alternatively, if δ^{15}N-AA patterns of HMW DON were "pre-formed" source water signatures, we would expect δ^{15}N-AA patterns at each depth to reflect those of HMW DON from the surface waters where each water mass was formed. For example, at both HOT and BATS, waters at 400 m originated in high-latitude regions of each respective ocean basin, waters at 850 m at HOT originated in the Antarctic, and deep waters (2500 m) at BATS are down welled from high latitudes while deep waters at HOT are transported via thermohaline circulation. Unfortunately, because there is no CSI-AA data of HMW DON from high-latitude regions, a direct comparison of HMW DON δ^{15}N-AA signatures with those of source waters is not possible. However, we can compare δ^{15}N$_{\text{Phe}}$ values of HMW DON from each depth with δ^{15}N values of nitrate and/or primary production from the respective surface source waters, assuming δ^{15}N$_{\text{Phe}}$ should be equivalent to δ^{15}N of baseline N. Large discrepancies are observed between most subsurface δ^{15}N$_{\text{Phe}}$ values and calculated δ^{15}N values of nitrate or plankton at high latitudes (McMahon et al., 2013, Rafter et al., 2019), suggesting preformed δ^{15}N-AA signatures from these source waters is also unlikely (See Supplementary 3.7.2).

Instead, our new data is most consistent with mesopelagic production of HMW DON from microbial resynthesis of PON. ΣV values of HMW DON at 400 m are greater than those of suspended PON at the same depths (Yamaguchi & McCarthy, 2018), supporting microbial production of mesopelagic HMW DON from PON. Additionally, at HOT, $\delta^{15}N_{\mathrm{Phe}}$ of mesopelagic (400 m and 850 m) HMW DON is within error of $\delta^{15}N_{\mathrm{Phe}}$ values of suspended PON collected at similar depths (Hannides et al, 2013, Yamaguchi et al., 2017) strongly supporting suspended PON as a potential source. Similarly, although there are no CSI-AA measurements of suspended PON in the Atlantic, $\delta^{15}N_{\mathrm{Phe}}$ of mesopelagic HMW DON is within error of the $\delta^{15}N$ of suspended PON in the mesopelagic Sargasso Sea (Altabet et al., 1988). While we know of no CSI-AA measurements of sinking PON from the central Pacific or Atlantic Gyres, bulk $\delta^{15}N$ values of sinking PON tend to be lower than that of suspended PON at subsurface depths (Altabet, 1988; Dore et al., 2002). Thus, considering the similarity of our measured $\delta^{15}N_{\mathrm{Phe}}$ values of HMW DON and literature $\delta^{15}N_{\mathrm{Phe}}$ and $\delta^{15}N_{\mathrm{Bulk}}$ values of suspended PON from the same depths, we suggest suspended PON is a more likely source than sinking PON.

Formation of subsurface HMW DON via microbial production from suspended PON would also be consistent with previous work in oligotrophic gyres. Elevated D/L ratios of mesopelagic HMW DON indicate a heterotrophic bacterial source to this material (Broek et al., 2019; Ianiri et al., Submitted), consistent with the increase in ΣV values we observed. Additionally, HMW DOM AA concentrations were observed to increase between 200 m and 300 m in the North Pacific Subtropical Gyre, suggesting an input of HMW AA at this depth (Kaiser & Benner, 2012). At the same time, high%D-Asx values of both HMW DON and suspended PON in the subsurface signify a substantial heterotrophic bacterial contribution

in the upper mesopelagic, indicating bacterial metabolism is causing the increase in AA concentrations (Kaiser & Benner, 2012). Similarly, measurements of bacterial abundance and activity also indicate mesopelagic microbial metabolism is enhanced due to POM, producing subsurface semi-labile DOM (Hansman et al., 2009; Nagata, 2000). Heterotrophic protists are also abundant in the mesopelagic (Pernice et al., 2015) and were estimated to consume $\sim 70\%$ of mesopelagic prokaryotic biomass at these depths (Fukuda et al., 2007), consistent with elevated $TP_{Protist}$ values. Finally, radiocarbon data is also consistent with these observations. Walker et al., (2016) suggested suspended POM as a major source of mesopelagic HMW DOM based on similar subsurface (615m) radiocarbon ages in the Central Pacific Ocean (suspended POM 915 yr BP to 3921.3 yr BP, HMW DOM 2934.3 yr BP). Similarly, Repeta & Aluwihare (2006) found neutral sugars in mesopelagic HMW DOM had younger $\Delta^{14}C$ ages than DIC at the same depths, suggesting they were sourced from POM. Lastly, ocean basin offsets in $\Delta^{14}C$ age of HMW DOM are less than expected due to aging with ocean circulation, indicating an addition of recently produced HMW DON below the surface (Broek et al., 2020).

Collectively, these results are consistent with the hypothesis that microbial resynthesis and protozoan grazing of mesopelagic suspended PON are important production mechanisms of semi-labile DON in the oceans twilight zone. These data support a surface to mesopelagic coupling in the marine nitrogen cycle, indicating some PON transported to the subsurface ocean in oligotrophic gyres enters the longer-lived DON pool. However, while it may have been predicted that the larger flux of more labile, sinking particles into the mesopelagic would have dominated this coupling, instead, the very similar $\delta^{15}N_{Phe}$ values of mesopelagic HMW DON and suspended PON strongly support suspended PON as the most

likely source. While it is unclear why older, more degraded suspended PON would represent the major source to mesopelagic DON production, it is possible the connection between particulate and dissolved ON operates along a size continuum, similar to that suggested for the greater DOC pool (Benner & Amon, 2015). Additional CSI-AA experiments of sinking PON could help constrain the dynamics between these pools of N.

3.5.5   LMW SPE-DON: Seemingly autotrophic with limited reworking

Perhaps the most novel aspect of this study is the new δ^{15}N-AA patterns of LMW SPE-DON, representing the first CSI-AA data of this size fraction (Fig. 3.3). Based on the Δ^{14}C ages of our LMW SPE-DOM samples (-333‰ to -577‰, or 3260 to 6860 years, (Broek et al., 2017, Submitted)), and in context of a size-age-reactivity framework, we anticipated LMW SPE-DON to represent the most microbially altered proteinaceous DON. Additionally, previous work demonstrated LMW SPE-DON has even more heterotrophic bacterially derived D-AA and non-protein AA than HMW DON (Kaiser and Benner et al., 2008, Broek et al. 2019, Ianiri et al., Submitted), suggesting heterotrophic bacteria are the dominant source to LMW SPE-DON. Thus, we hypothesized LMW SPE-DON would have δ^{15}N-AA patterns clearly deviated from the well-known patterns of autotrophs and high ΣV values.

Contrary to these expectations, the fact that CSI-AA patterns of LMW SPE-DON are very similar to autotrophic patterns suggests proteinaceous LMW SPE-DON is both surface produced and better preserved than HMW DON by most δ^{15}N-AA proxies (Fig. 3.3, Fig. 3.5). The limited range of trophic δ^{15}N-AA values and the low $\text{TP}_{\text{Classic}}$, $\text{TP}_{\text{Protist}}$, and ΣV values are instead all more consistent with δ^{15}N-AA signatures of autotrophs than of bacterially degraded material (Décima

et al., 2017; McCarthy et al., 2007, 2013; McClelland & Montoya, 2002). Indeed, scatterplots of $TP_{Protist}$ and $TP_{Classic}$ vs. ΣV all show LMW SPE-DON plotting with ultrafiltered PON, supporting this interpretation (Fig. S6). Additionally, both absolute and normalized $\delta^{15}N$-AA patterns and values of LMW SPE-DON are similar regardless of location or depth, with depth offsets in $\delta^{15}N$-AA values much smaller than those observed in HMW DON (Fig. 3.4, Fig. S3.4). Together, these results suggest that most proteinaceous LMW SPE-DON has remarkably similar sources and formation processes, with production in the surface ocean followed by limited reworking with age and ocean circulation.

The limited variation in most $\delta^{15}N_{Phe}$ values of subsurface LMW SPE-DON (Fig. 3.6) also supports a similar source to LMW SPE-DON throughout the water column in both ocean basins. While there may be some variability in $\delta^{15}N_{Phe}$ values in the upper water column at HOT, we are hesitant to put too much emphasis on one point (400 m) which has relatively large error bars. Instead, the similarity of all other $\delta^{15}N_{Phe}$ values measured at HOT, especially those at 850 m and 2500 m, may suggest that $\delta^{15}N_{Phe}$ represents an average, background value of the inorganic N source to proteinaceous LMW SPE-DON. While unfortunately we only have $\delta^{15}N_{Phe}$ values for the upper water column (surface and 400 m) at BATS, the two datapoints we do have are also within error of each other, and to most $\delta^{15}N_{Phe}$ values at HOT, supporting this possibility. If true, this could provide an informative constraint on where in the world's oceans refractory DON molecules are produced. While the global average $\delta^{15}N$ of NO_3^- is ~ 5‰, the average $\delta^{15}N_{Phe}$ of LMW SPE-DON throughout oligotrophic gyres is ~ 2‰. The offset in these two values could suggest that persistent proteinaceous LMW SPE-DON is preferentially produced from an N source with $\delta^{15}N$ values slightly lower than mean ocean nitrate, but higher than N_2 fixation or recycled N. Annually

112

averaged NO_3^- from oligotrophic gyres has $\delta^{15}N$ values been $2 - 3‰$, within this range (Knapp et al., 2005). Additional $\delta^{15}N$-AA measurements of LMW SPE-DON from other oceanic environments, such as coastal margins and high latitude regions, as well as deep waters in the Atlantic Ocean, could help determine if $\delta^{15}N_{Phe}$ values are long term global averages or represent a local signal.

Taken together, the lack of isotopic evidence for major bacterial degradation or resynthesis in the LMW SPE-DON pool is contradictory to our expectations that LMW SPE-DON would represent the more microbially altered DON size fraction. These results are also unexpected in context of previously published D-AA signatures of LMW SPE-DON, which suggested a dominant heterotrophic bacterial source to LMW, proteinaceous material (Broek et al., 2019; Ianiri et al., Submitted; Kaiser & Benner, 2008). However, there are two potential production mechanisms which could be consistent with the isotopic and bacterial biomarker observations.

The first is *de novo* synthesis by heterotrophic bacteria. *De novo* synthesis should have CSI-AA patterns similar to those of autotrophic production (Yamaguchi et al., 2017), but could still produce the observed D-AA in LMW SPE-DON (Broek et al., 2019; Ianiri et al., Submitted; Kaiser & Benner, 2008) (Fig. 3.8). LMW SPE-DON molecules could then be released from heterotrophic bacteria via processes with limited fractionation such as direct exudation or viral lysis of heterotrophic bacterial cells. Incubation experiments have shown *de novo* synthesis by heterotrophic bacteria can rapidly (< 48 hours) produce persistent, refractory DOM compounds, supporting this mechanism (Gruber et al., 2006; Ogawa et al., 2001). Additionally, we note that in one experiment, refractory, LMW compounds produced by bacteria did not derive from the breakdown of HMW compounds, but rather were directly released from heterotrophic bacteria (Gruber et al., 2006),

consistent with the apparent separation of MW pools we have observed across our data set. Furthermore, while most AA-containing compounds released via viral lysis are rapidly recycled, a portion may enter the refractory DON pool (up to 17%, Middelboe & Jørgensen, 2006). If heterotrophic bacteria were utilizing subsurface NO_3^- as their main inorganic N source, and directly released LMW proteinaceous molecules were resistant to further degradation, the above mechanisms would be consistent with all the isotopic data we observed.

Alternatively, it is possible cyanobacterial sources may be more important than previously recognized. Previous work has shown Synechococcus directly releases RDOM which is resistant to heterotrophic degradation (Zhang et al., 2021; Zhao et al., 2017), supporting cyanobacterial production of refractory molecules. While early δ^{13}C CSI-AA work suggested cyanobacteria as a possible direct source to at least HMW DON (McCarthy et al., 2004), subsequent measurements of D-AA in cyanobacterial cultures indicated cyanobacteria do not produce the full range of D-AA observed in LMW SPE-DON (Kaiser & Benner, 2008). However, to our knowledge, measurements of D-AA have only been made in one natural assemblage of cyanobacteria (Trichodesmium sp., Kaiser & Benner, 2008). Further, many of the less common D-AA observed in LMW SPE-DON have so far only been observed in heterotrophic bacterial cultures under specific environmental stressors (Cava et al., 2011), meaning if cyanobacteria also produce these D-AA, they may not be observed in standard lab cultures. Additionally, measurement of many of the novel D-AA unique to DON (as opposed to PON) have required identification via MS mass fragmentation patterns (Broek et al., 2019; Ianiri et al., Submitted; Lam et al., 2009), suggesting previous work measuring D-AA in cyanobacteria via HPLC may not have detected minor D-AA. We suggest that future experiments using high resolution mass spectral identification of all D-AA

in culture, or isolation and measurements of natural cyanobacterial assemblages will be necessary to determine if cyanobacteria could be a potential source of D-AA in LMW SPE-DON.

Regardless of the exact mechanism, overall, our data indicates production of proteinaceous LMW SPE-DON occurs mainly in the surface ocean with limited reworking during ocean circulation. Direct production of LMW SPE-DON in the surface ocean is a major departure from the current understanding of LMW SPE-DOC production, which suggests continuous bacterial reworking of semi-labile DOC produces refractory molecules (Amon & Benner, 1996; Benner & Amon, 2015; Jiao et al., 2011). However, our interpretation is supported by previous work targeting the LMW SPE-DON pool. For example, recent work investigating the relationship between D-AA, AA-based degradation indices, and radiocarbon similarly suggests AA-containing molecules in LMW SPE-DON do not undergo progressive microbial alteration with age (Broek et al., 2019; Ianiri et al., Submitted). Additionally, based on low $\delta^{15}N_{Bulk}$ values of LMW SPE-DON compared to HMW DON, Knapp et al. (2012) suggested LMW SPE-DON production and degradation is independent of HMW DON cycling. Instead, they hypothesized that LMW SPE-DON was produced via mechanisms with limited isotopic fractionation, consistent with our hypothesized model based on CSI-AA patterns. Finally, a direct bacterial source of relatively refractory DON could explain past observations of preferential degradation of DOC compared to DON in surface oligotrophic gyres (Abell et al., 2000, 2005; Emerson & Hayward, 1995). Together, this past work combined with our CSI-AA data suggests production of refractory, LMW SPE-DON may be largely decoupled from production and cycling dynamics of RDOC.

Notably, these results are specific to AA in the LMW SPE-DON size fraction

isolated and measured in this study, and it is unclear if these results can be extrapolated to the entire LMW SPE-DON pool. However, comparisons of AA composition and degradation state in LMW SPE-DON and the ultrafiltration permeate (representing total LMW SPE-DON) were very similar (Ianiri et al., Submitted), suggesting it is likely these same mechanisms may be applicable to all LMW AA.

3.5.6 New hypotheses for DON source and cycling

Taken together, our CSI-AA data of HMW and LMW SPE-DON spanning two oligotrophic gyres indicate distinct production and degradation mechanisms of these two size fractions. Based on these data, we suggest separate hypothesized conceptual models for production of proteinaceous HMW DON and LMW SPE-DON (Fig. 3.7, Fig. 3.8, respectively).

Our model for HMW DON (Fig. 3.7) is a modified version of a previous hypothesis suggested by (Yamaguchi & McCarthy, 2018), and assumes that $\delta^{15}N_{Phe}$ values represent the $\delta^{15}N$ value of baseline NO_3^-. In this model, eukaryotic algae at the base of the euphotic zone utilize subsurface NO_3^-. Heterotrophic bacteria then resynthesize labile ON produced via this primary production. While previous work suggested heterotrophic bacteria could also be acquiring NO_3^- directly (Yamaguchi & McCarthy, 2018), we suggest this is less likely, as heterotrophic biomass produced via *de novo* synthesis from inorganic N would have $\delta^{15}N$-AA patterns like autotrophs rather than the microbial $\delta^{15}N$-AA patterns we observed for HMW DON (Yamaguchi et al., 2017). Heterotrophic resynthesis has been shown to increase $\delta^{15}N_{THAA}$ of HMW DON by $\sim 3‰$ to 6‰ (Calleja et al., 2013), which could explain the higher $\delta^{15}N_{THAA}$ and $\delta^{15}N_{Bulk}$ values of HMW DON compared to $\delta^{15}N_{Phe}$. Finally, we suggest an additional heterotrophic protozoan

116

trophic step. While there are multiple ways heterotrophic bacteria and protozoa could produce the CSI-AA signals we observe (Section 3.5.3, Supplementary 3.7.1), our preferred hypothesis is that HMW DON is left behind from protozoan grazing on heterotrophic bacterial biomass. Together, these mechanisms would result in a HMW DON pool with increased ΣV and $TP_{Protist}$ values, low $TP_{Classic}$, and elevated D/L ratios.

In contrast, $\delta^{15}N$-AA isotopic signatures indicate production of proteinaceous LMW SPE-DON is completely independent from the HMW DON pool. In our model for the production of proteinaceous LMW SPE-DON (Fig. 3.8), heterotrophic bacteria utilize subsurface nitrate directly for *de novo* synthesis. LMW SPE-DON is then produced via processes with limited fractionation such as direct exudation or viral lysis of heterotrophic bacterial cells. This would result in lower $\delta^{15}N_{THAA}$ values than HMW DON and $\delta^{15}N$-AA patterns similar to those of autotrophs, while still consistent with the elevated D/L ratios indicating heterotrophic production. This mechanism would also be consistent with the offset in $\delta^{15}N_{Bulk}$ values observed between HMW and LMW SPE-DON (Broek et al., Submitted; Knapp et al., 2012) and the similar $\delta^{15}N_{Bulk}$ values of LMW SPE-DON to the $\delta^{15}N$ value of NO_3^- in oligotrophic gyres.

Notably, it is unclear if these same cycling processes apply to the entire marine DON pool or only to AA-containing molecules. The differences in $\delta^{15}N_{THAA}$ and $\delta^{15}N_{Other}$ we observed for both HMW and LMW SPE-DON could indicate differences in production or degradation mechanisms of proteinaceous DON versus other dissolved N compound classes (Section 3.5.1). For the LMW SPE-DON pool specifically, given that heterocyclic nitrogen has recently been identified as a new and potentially major component of this size fraction, understanding sources and cycling to this material, and to what degree it mirrors that of THAA, will be

critical.

3.6   Conclusions

In this study, we demonstrate that δ^{15}N-AA measurements provide unprecedented detail regarding proteinaceous HMW and LMW SPE-DON source and cycling mechanisms. Differences in δ^{15}N$_{\text{Bulk}}$, δ^{15}N$_{\text{THAA}}$, and δ^{15}N$_{\text{Other}}$ indicate multiple, independently cycling pools of HMW and LMW SPE-DON. However, in contrast to past work, we suggest AS cannot account for the differences in δ^{15}N$_{\text{THAA}}$ and δ^{15}N$_{\text{Bulk}}$ and must either have similar δ^{15}N values to proteinaceous material or do not accumulate appreciably in surface HMW DON. Instead, we suggest unidentified amide molecules and possibly some heterocyclic N make up the "other" HMW DON which accumulates in the subsurface and has a lower δ^{15}N than proteinaceous material. In contrast, most "other" LMW SPE-DON likely consists of heterocyclic N molecules which also have a lower δ^{15}N value compared to LMW proteinaceous material. We suggest these "other" ON pools may represent some of the most refractory marine DON and warrant further study.

The CSI-AA results presented here, including two DON size fractions spanning four depths in both the Pacific and Atlantic gyres, represents by far the most expansive CSI-AA study of marine DON, as well as the first CSI-AA data of LMW SPE-DON or DON from outside the Pacific Ocean. For the HMW DON pool, our results both confirm and expand recent CSI-AA results on a smaller dataset. Consistent with these previous results, we find evidence that microbial metabolism utilizing subsurface nitrate is the main production mechanism of proteinaceous HMW DON in the surface ocean. Additionally, we find evidence for microbial production of mesopelagic HMW DON from suspended PON, supporting a particle source to HMW DON in the mesopelagic and deep ocean. Collectively, while

these results provide evidence for a coupling between surface and mesopelagic N cycling, this connection is not via the marine DON pool or resynthesis of sinking PON, as may be expected. Instead, these results suggest production of surface and mesopelagic HMW DON must be considered independently, with different sources and production mechanisms. Still, microbial loop processes appear to be the primary source to proteinaceous HMW DON at all depths, suggesting microbial resynthesis and protozoan grazing are key to the persistence of HMW DON throughout oligotrophic gyres. Overall, these conclusions have significant implications for our understanding of the marine N cycle and production of semi-labile ON.

Finally, these first CSI-AA measurements of LMW SPE-DON provide detailed information regarding source and cycling processes of this size fraction which were previously obscured by bulk measurements. Contrary to our expectations that proteinaceous LMW SPE-DON would have the most "degraded" δ^{15}N-AA patterns, LMW SPE-DON had δ^{15}N-AA patterns similar to those of autotrophic organisms and appeared better preserved than HMW DON as measured by δ^{15}N-AA parameters. Based on these data, we suggest proteinaceous LMW SPE-DON is directly produced in the surface ocean and subject to limited further microbial resynthesis. Taken as a whole, our data thus implies production and degradation mechanisms of HMW and LMW SPE-DON are completely independent and likely indicate AA-containing molecules in the two size fractions have very different chemical compositions. These conclusions are inconsistent with the "size-age-reactivity" framework which is widely applicable to the DOC pool, and suggest RDOC and RDON cycling may be relatively decoupled. Overall, these conclusions have significant implications for our understanding of marine DON production and recycling, suggesting the refractory nature of most DON may be due to direct

bacterial production of degradation-resistant molecules rather than progressive microbial degradation.

Main text figures
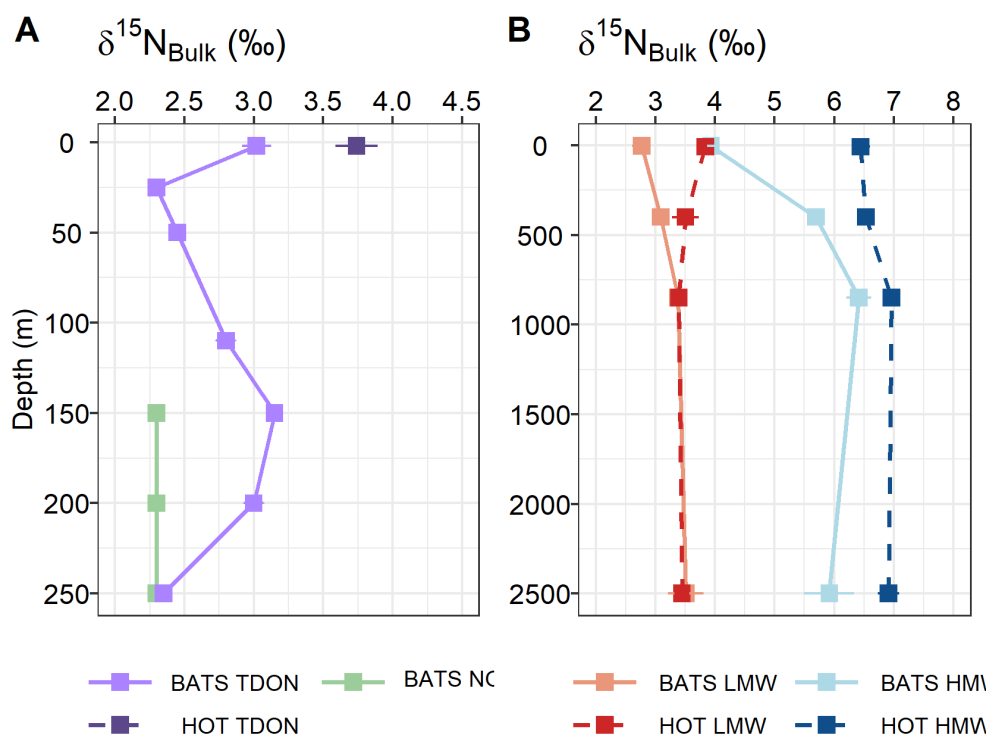
Figure 3.1: Depth profiles of $\delta^{15}N_{Bulk}$ for A) total DON (purple) and NO_3^- (green) and B) HMW DOM (blue) and LMW SPE-DOM (red) from HOT (dashed lines, darker shade) and BATS (solid lines, lighter shade). Surface total DON (TDON) and all HMW and LMW SPE-DON $\delta^{15}N$ values represent averages of spring and summer cruises and error bars represent the propagated analytical error of triplicate measurements. Below the surface, TDON at BATS was only measured during the summer cruise. For these points, error bars represent the analytical error of triplicate measurements. Error bars are smaller than symbol where not visible.
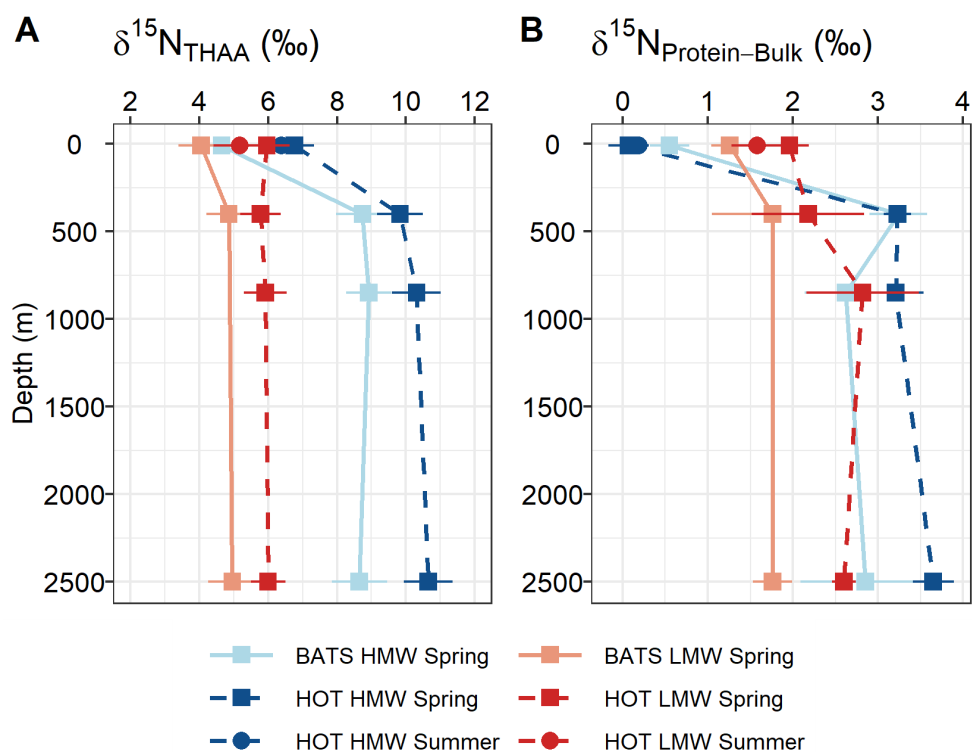
Figure 3.2: Depth profiles of A) $\delta^{15}N_{THAA}$ and B) the offset of $\delta^{15}N_{Protein-Bulk}$ of HMW DON (blue) and LMW SPE-DON (red) at HOT (dashed lines, darker shade) and BATS (solid lines, lighter shade). HMW and LMW SPE-DON were measured throughout the water column in spring (squares) and at the surface in the summer at HOT (circles). Error bars for $\delta^{15}N_{THAA}$ represent the propagated error of the standard deviation of duplicate or triplicate $\delta^{15}N$-AA and mol% measurements. $\delta^{15}N_{Protein-Bulk}$ error bars represent the propagated error of the two measurements. Error bars are smaller than symbol where not visible.
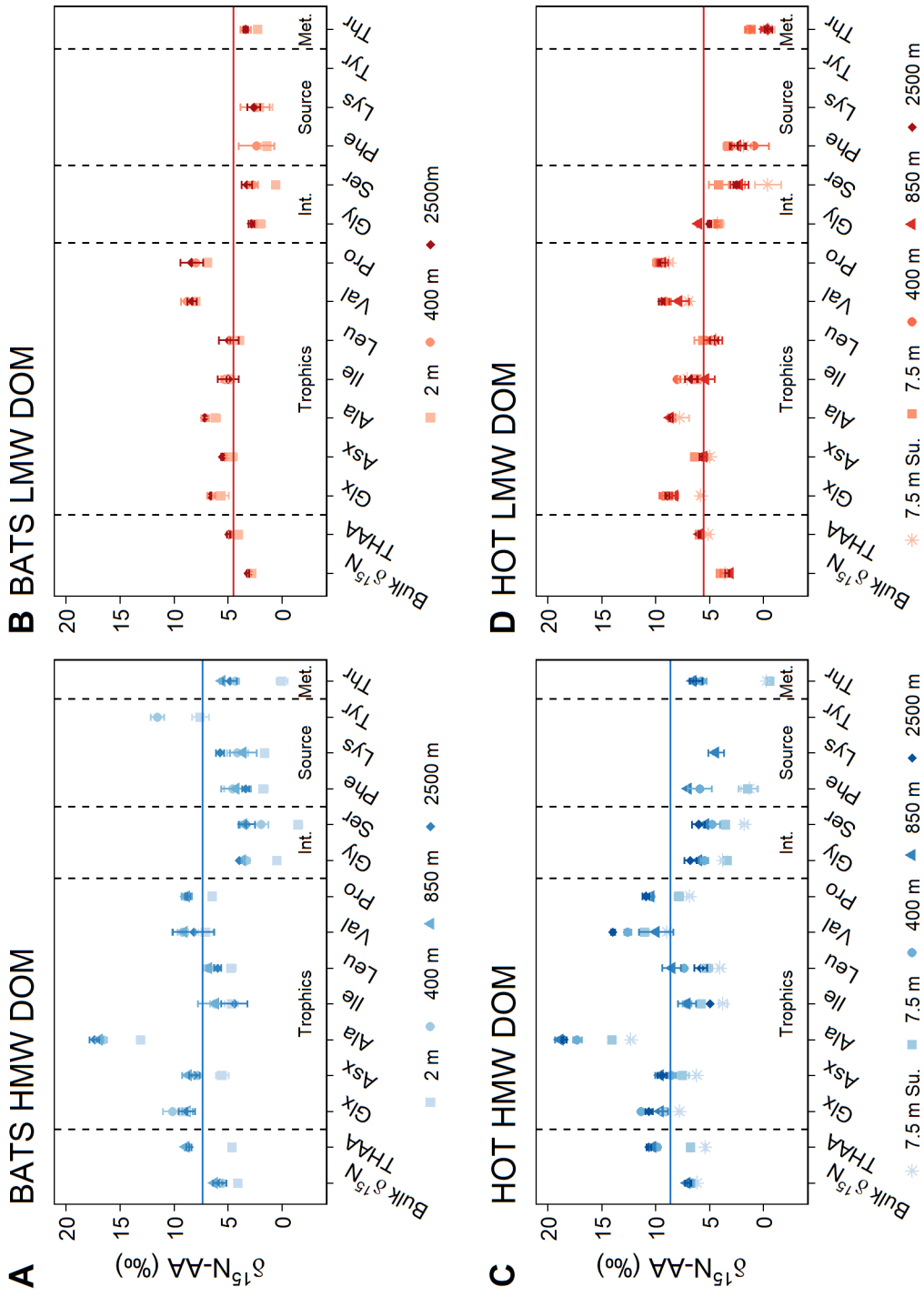
Figure 3.3: δ^{15}N-AA patterns of A) HMW DON at BATS, B) LMW SPE-DON at BATS, C) HMW DON at HOT, D) LMW SPE-DON at HOT. HMW DON δ^{15}N-AA values generally increase from surface (lightest shade) to depth (darkest shade) while LMW SPE-DON δ^{15}N-AA values are similar at all depths. Solid horizontal line indicates average δ^{15}N$_{THAA}$ for all samples in that size fraction/ocean basin. Error bars represent the standard deviation of triplicate measurements unless otherwise noted in the text. AA are grouped as described in text (Section 3.3.6). "Int." stands for Intermediate AA, while "Met." stands for Metabolic. Samples labeled "Su." were collected during summer cruises, while all others were collected on spring cruises.
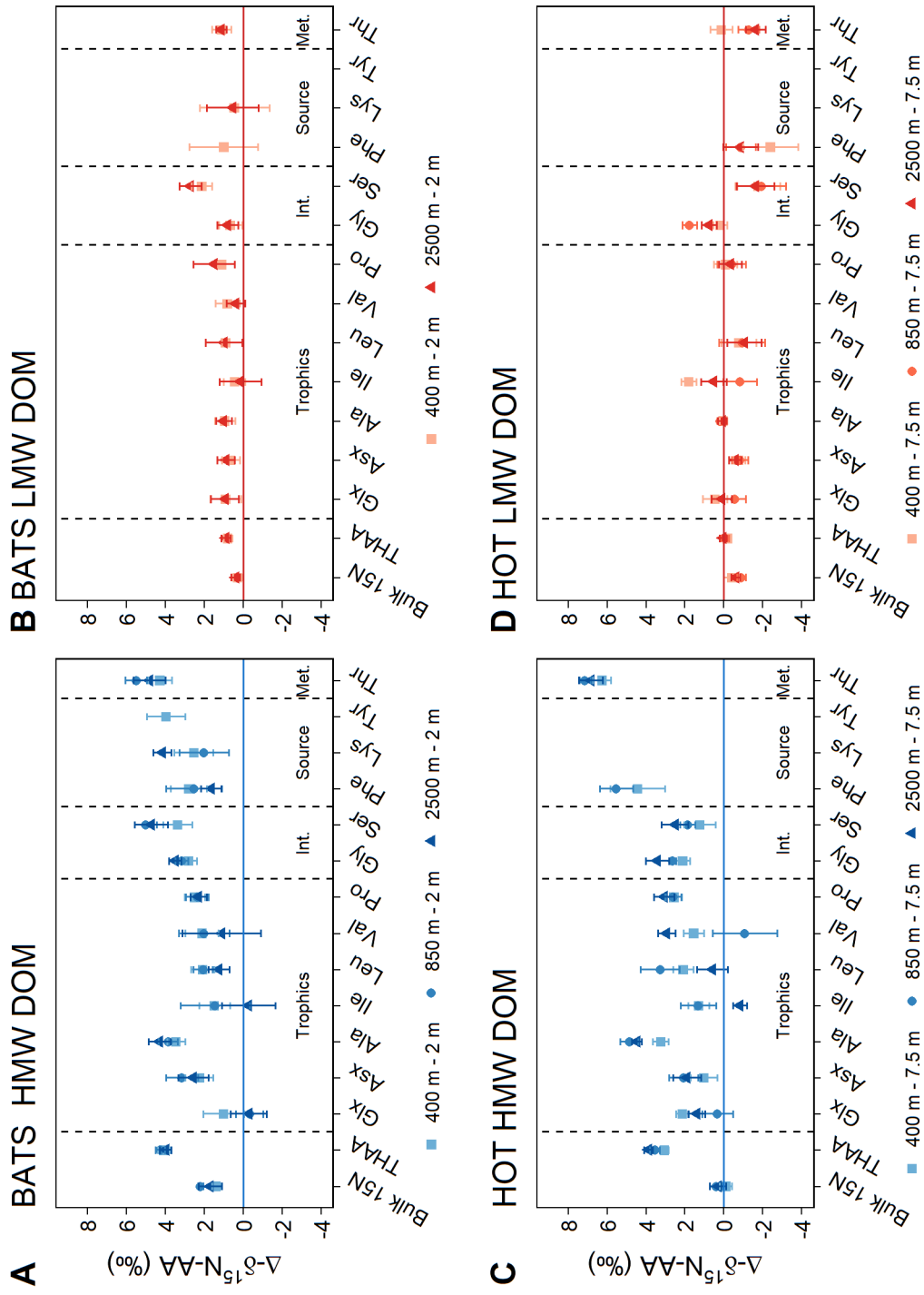
Figure 3.4: Difference in deep − surface δ^{15}N-AA values of A) HMW DON at BATS, B) LMW SPE-DON at BATS, C) HMW DON at HOT, and D) LMW SPE-DON at HOT. Error bars represent the propagated error of the standard deviation of triplicate measurements unless otherwise noted in the text. AA are grouped as described in the text (Section 3.3.6). "Int." stands for Intermediate AA, while "Met." stands for Metabolic. All data is from spring cruises.
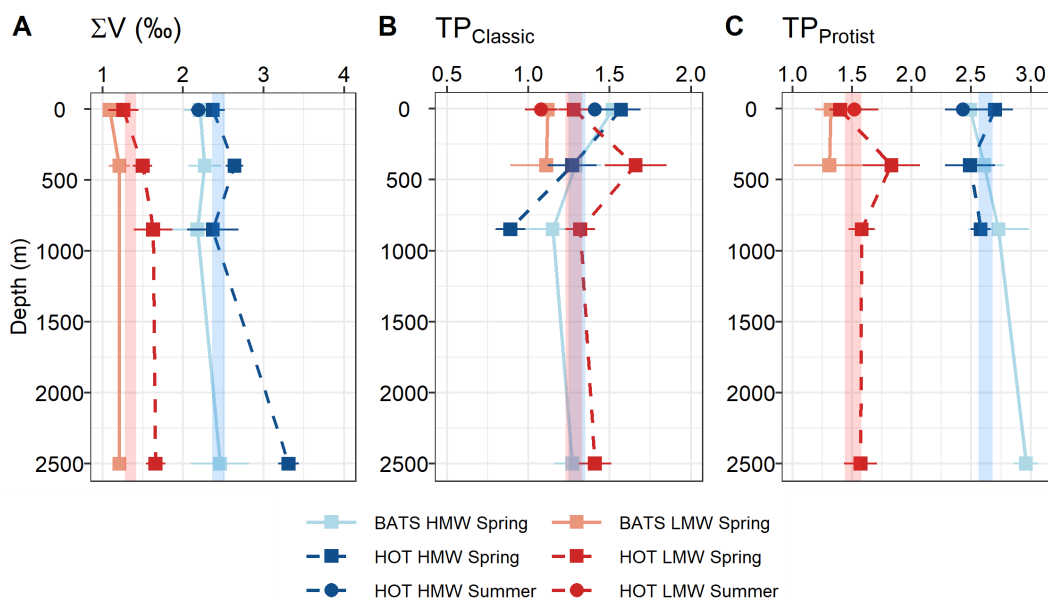
Figure 3.5: δ^{15}N-AA parameters for resynthesis and degradation of HMW DON (blue) and LMW SPE-DON (red) at HOT (dashed lines, darker shade) and BATS (solid lines, lighter shade) measured in the summer (circles) and spring (squares). Shaded regions indicate average of all HMW or LMW SPE-DON samples from both ocean basins $\pm$ one standard deviation. A) ΣV, a proxy for heterotrophic bacterial resynthesis, suggests more resynthesis to HMW DON than LMW SPE-DON throughout the water column. B) $TP_{Classic}$, or metazoan trophic position, of both HMW and LMW SPE-DON indicates mostly autotrophic sources, while C) $TP_{Protist}$, or protozoan trophic position, indicates an additional protozoan trophic step only to HMW DON. Error bars represent the propagated error of the standard deviation of triplicate measurements unless otherwise noted in the text.
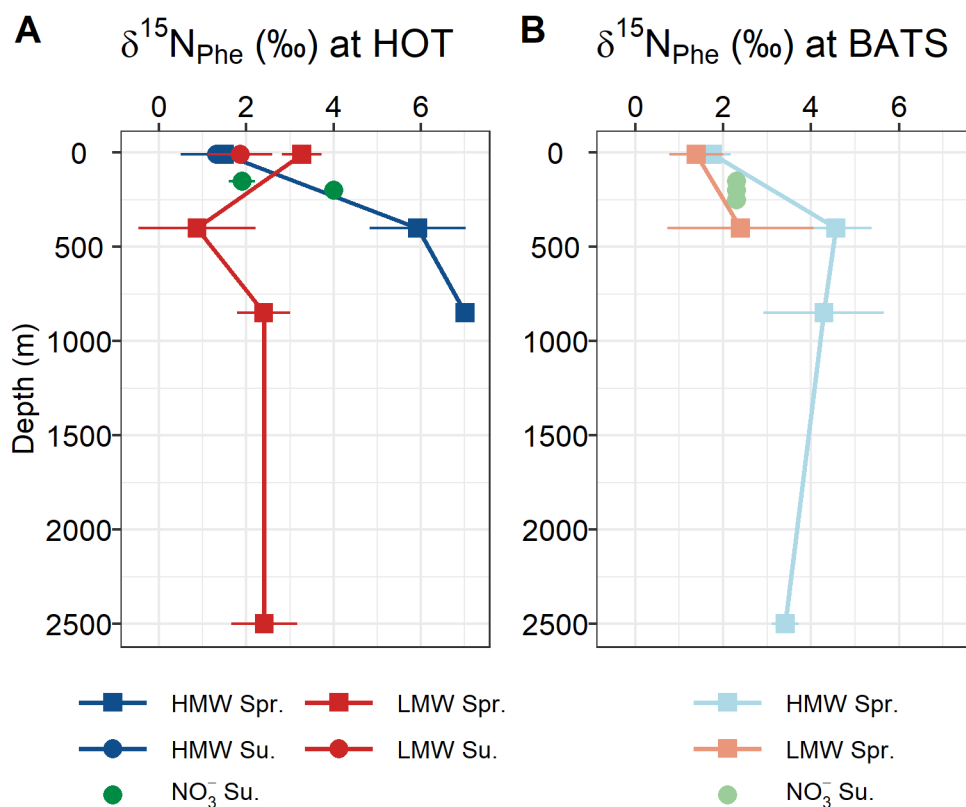
Figure 3.6: $\delta^{15}N_{Phe}$, a proxy for the $\delta^{15}N$ value of baseline N, A) at HOT and B) at BATS. Surface $\delta^{15}N_{Phe}$ of HMW (blue) and LMW SPE-DON (red) measured in spring (squares, "Spr.") and summer (circles, "Su.") are within error of summer measurements of $\delta^{15}N$ of NO_3^- made in this study at BATS (light green) and literature values from HOT (dark green, Casciotti et al., 2008 and Knapp et al., 2011). HMW DON $\delta^{15}N_{Phe}$ values in the subsurface are significantly greater than surface values, while LMW SPE-DON $\delta^{15}N_{Phe}$ is relatively consistent throughout the water column.

126
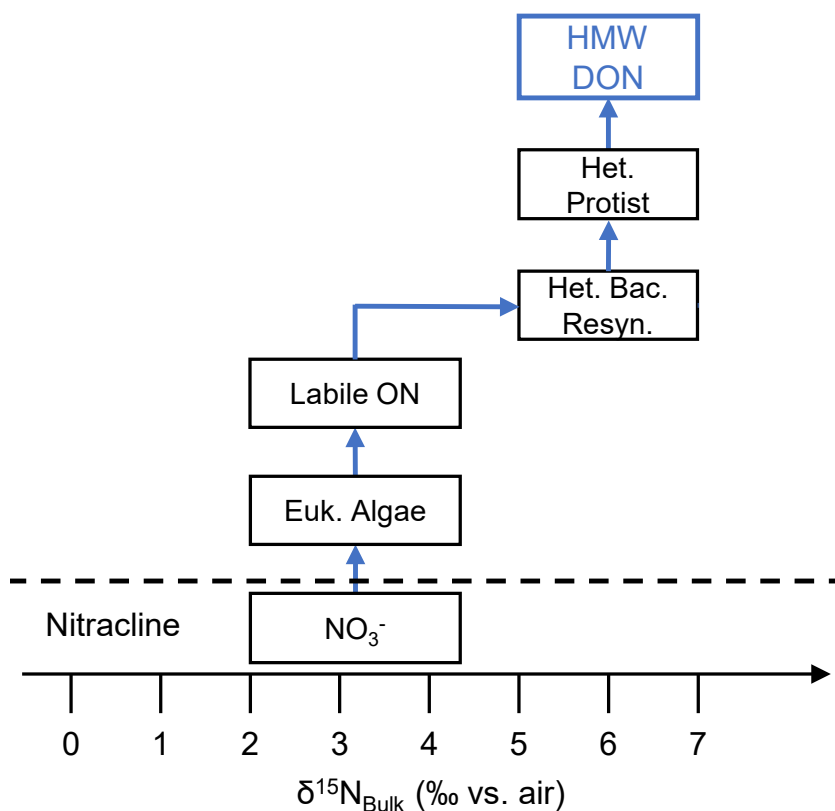
Figure 3.7: Conceptual diagram of hypothesized production and cycling mechanisms of proteinaceous HMW DON in the surface ocean (Section 3.5.6). Schematic is a modified version of model originally proposed by Yamaguchi & McCarthy (2018), and assumes $\delta^{15}N_{Phe}$ tracks baseline NO_3^- values. Autotrophic organisms utilize subsurface NO_3^-, producing labile ON. Heterotrophic bacterial consume and resynthesize this ON, isotopically fractionating AA. Epsilon for fractionation via heterotrophic bacteiral resynthesis was estimated from Calleja et al. (2013). Finally, protozoan grazing of heterotrophic bacteria produces semi-labile HMW DON. Y-axis value of NO_3^- is an approximate estimate for the North Pacific Subtropical Gyre (NPSG) during summer (Casciotti et al., 2008, Knapp et al., 2011). $\delta^{15}N$ values of eukaryotic algae, labile ON, and heterotrophic bacteria assume complete utilization of NO_3^- in oligotrophic regions. While these values are specific to the NPSG, we suggest mechanisms of production are similar throughout oligotrophic gyres. Abbrevations include: Het. Protist (heterotrophic protists), Het. Bac. Resyn. (heterotrophic bacterial resynthesis), Labile ON (labile organic nitrogen), Euk. Algae (eukaryotic algae).
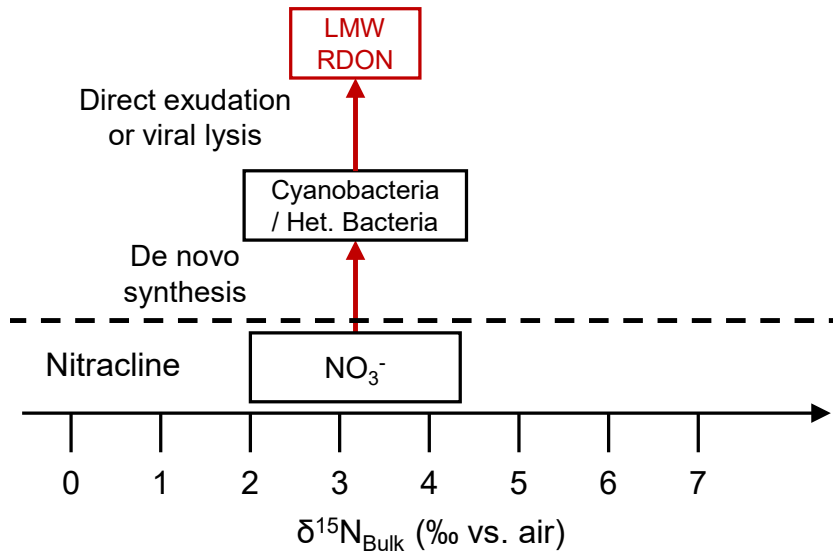
Figure 3.8: Conceptual diagram of hypothesized production and cycling mechanisms of proteinaceous LMW SPE-DON in the surface ocean. Schematic assumes $\delta^{15}N_{Phe}$ tracks baseline NO_3^- values. In contrast to HMW DON, we suggest LMW SPE-DON is sourced from heterotrophic bacteria via processes with limited isotopic fractionation. Heterotrophic bacteria utilize subsurface nitrate directly for *de novo* synthesis, producing $\delta^{15}N$-AA patterns similar to autotrophic organisms. They then release LMW SPE-DON via processes with limited fractionation, such as viral lysis or direct exudation. This results in lower $\delta^{15}N_{Bulk}$ and $\delta^{15}N_{THAA}$ values than HMW DON. Y-axis value of NO_3^- is an approximate estimate for the North Pacific Subtropical Gyre (NPSG) during summer (Casciotti et al., 2008, Knapp et al., 2011). $\delta^{15}N$ values of heterotrophic bacteria assume complete utilization of NO_3^- in oligotrophic regions. While these values are specific to the NPSG, we suggest mechanisms of production are similar throughout oligotrophic gyres. Abbreviations include Het. Bacteria (heterotrophic bacteria).

3.7 Supplementary

3.7.1 $\delta^{15}N_{Ala}$, ΣV, and $TP_{Protist}$

Ala is generally the most abundant amino acid in HMW DON, has the highest D/L ratio in HMW DON (Broek et al., 2019, Ianiri et al., Submitted), and by far the greatest $\delta^{15}N$ value for all HMW DON samples. Ala's significance to HMW DON and apparent connection to microbial metabolism poses the question of how influential Ala is to ΣV. Indeed, a strong correlation between ΣV and $TP_{Protist}$ is observed in HMW DON, as well as throughout the entire ON pool (UPON, HMW, and LMW SPE-DON) (Fig. S6). However, this relationship is no longer significant if ΣV is calculated without Ala (Fig. S6). The lower ΣV values of HMW DON without Ala suggests Ala is fractionated more than any other AA by microbial loop processes in HMW DON. This provides strong support for Ala as a key indicator of microbial alteration of HMW DON.

As noted in the main text (Section 3.5.1), the strong relationship between $\delta^{15}N_{Ala}$ and ΣV raises the question if Ala fractionation is due to microbial resynthesis, protozoan heterotrophy, or both. Our preferred hypothesis is that protozoan heterotrophy of microbial biomass produces HMW DON (Fig. 3.7). However, an alternative mechanism is if protozoan heterotrophy, not microbial heterotrophy, consumes eukaryotic algae utilizing NO_3^- from the base of the food web. In this case, protozoan metabolism would fractionate $\delta^{15}N$ of Ala, leading to the increased $TP_{Protist}$ and ΣV of HMW DON. However, to date, there is no evidence that protozoa produce or utilize D-AAs, meaning this process could not result in the D-AA content in HMW DON. To produce the D-AA signal of HMW DON, we suggest heterotrophic bacteria would then resynthesize leftovers from the grazing of mesozooplankton on protozoa. HMW DON would then be produced from heterotrophic bacterial biomass by processes such as viral lysis.

3.7.2 $\delta^{15}N_{Phe}$ compared to high latitude primary production

To determine if subsurface HMW DON δ^{15}N-AA patterns are a preformed source signature, we compare $\delta^{15}N_{Phe}$ values with δ^{15}N values of NO_3^- or marine primary productivity where each water mass originated. This comparison assumes that the $\delta^{15}N_{Phe}$ value represents the δ^{15}N value of baseline N at each source region and is persevered with further ocean circulation. At HOT and BATS, the 400 m sample represents North Pacific Intermediate water and Subtropical Mode Water, respectively (Talley, 1993; You, 2003), each of which originate in high latitude waters of the respective ocean basin. At BATS, North Atlantic Deep Water (2500 m) (Talley, 1996), also originates at high latitudes in the Atlantic Ocean. In contrast, waters at 850 m at HOT represent Antarctic Intermediate Water (Santoso & England, 2004), the 2500 m sample represent older, North Pacific Deep Water (Talley, 2013), both of which most recently upwelled in the Antarctic Ocean.

In the Atlantic Ocean, it is estimated high latitude δ^{15}N values of plankton range from $\sim 6‰$ to 8‰ (McMahon et al., 2013) and δ^{15}N of NO_3^- ranges from $\sim 6‰$ to 10‰ between the surface and 200 m (Rafter et al., 2019), substantially greater than $\delta^{15}N_{Phe}$ values we observed throughout the water column at BATS (Fig. 3.6). At HOT, δ^{15}N of NO_3^- between the surface to 200 m is estimated to range from 7‰ to 10‰ in high latitude regions, again higher than $\delta^{15}N_{Phe}$ at 400 m. In contrast, estimated δ^{15}N NO_3^- values in the Southern Ocean ($\sim 6‰$ in the surface, Rafter et al (2019)) are closer to the $\delta^{15}N_{Phe}$ values we observe at intermediate waters at HOT. Still, the discrepancy between $\delta^{15}N_{Phe}$ and most estimated δ^{15}N NO_3^- values, as well as the similar δ^{15}N-AA patterns of HMW DON at all depths and both ocean basins, suggests unique, preformed δ^{15}N-AA signatures from source waters is unlikely.

Supplementary Figures and Tables
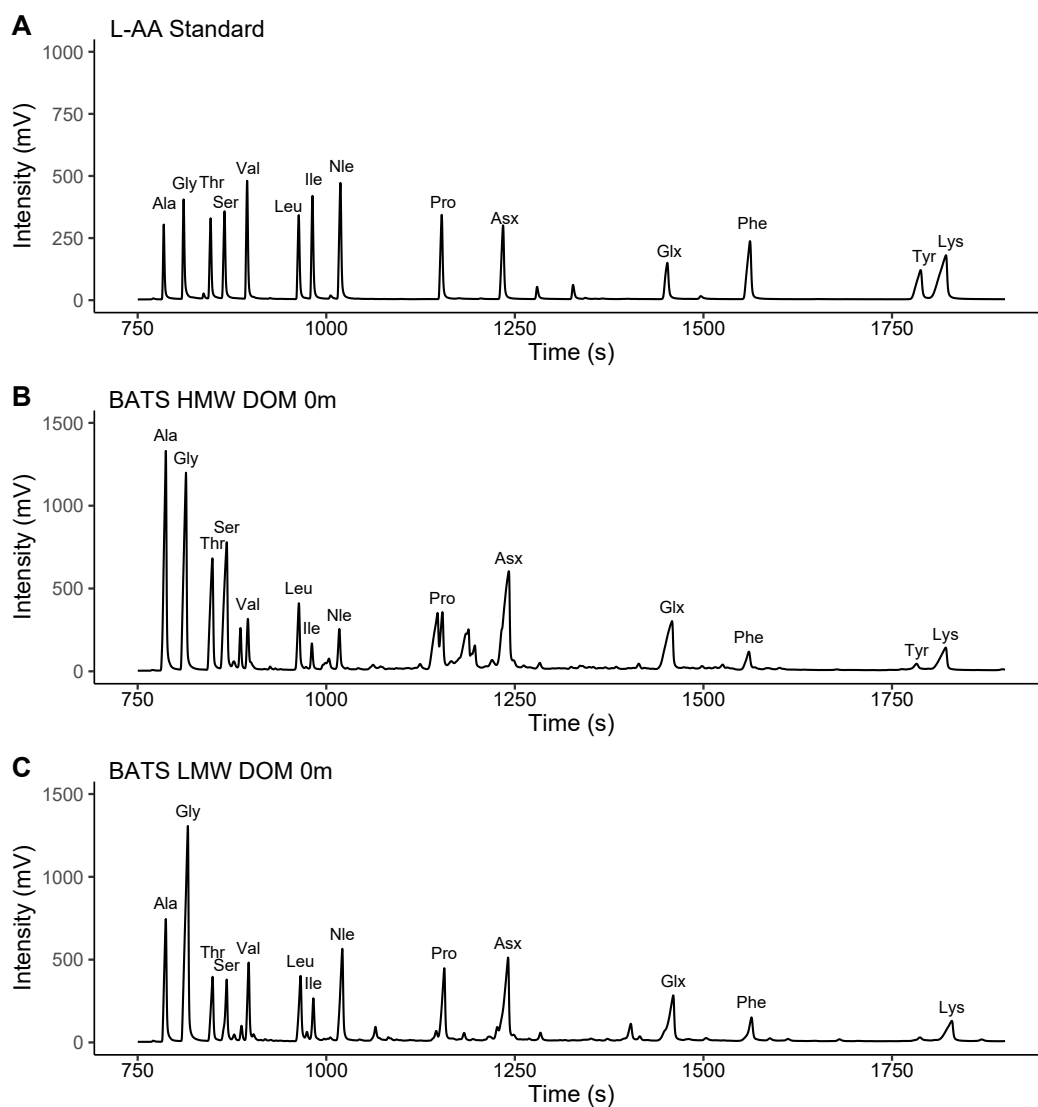
Figure S3.1: δ^{15}N GC-IRMS chromatograms of A) an L-AA standard, B) HMW DON from the surface ocean at BATS, and C) LMW SPE-DON from the surface ocean at BATS. Substantial upstream purification of DOM samples (Section 3.3.3) resulted in few N-containing molecules besides target AA in either size fraction.
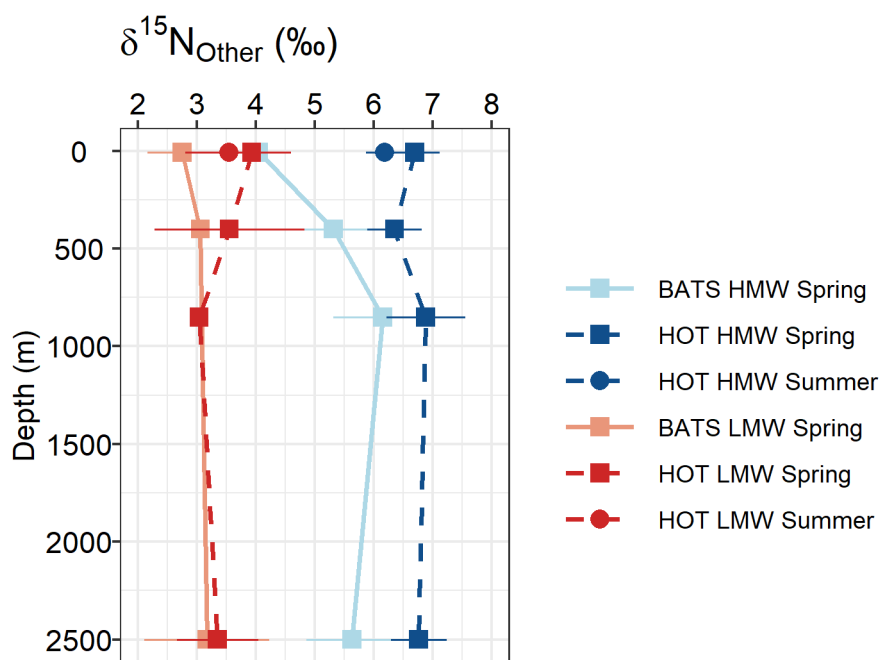
Figure S3.2: $\delta^{15}N_{Other}$ ($\delta^{15}N$ of ON besides hydrolysable amino acids) in HMW DOM (blue) and LMW SPE-DOM (red) from HOT (dashed lines, darker shade) and BATS (solid lines, lighter shade) measured in spring (squares) and summer (circles). $\delta^{15}N_{Other}$ is very similar to $\delta^{15}N_{Bulk}$. Error bars represent propagated error of triplicate measurements.
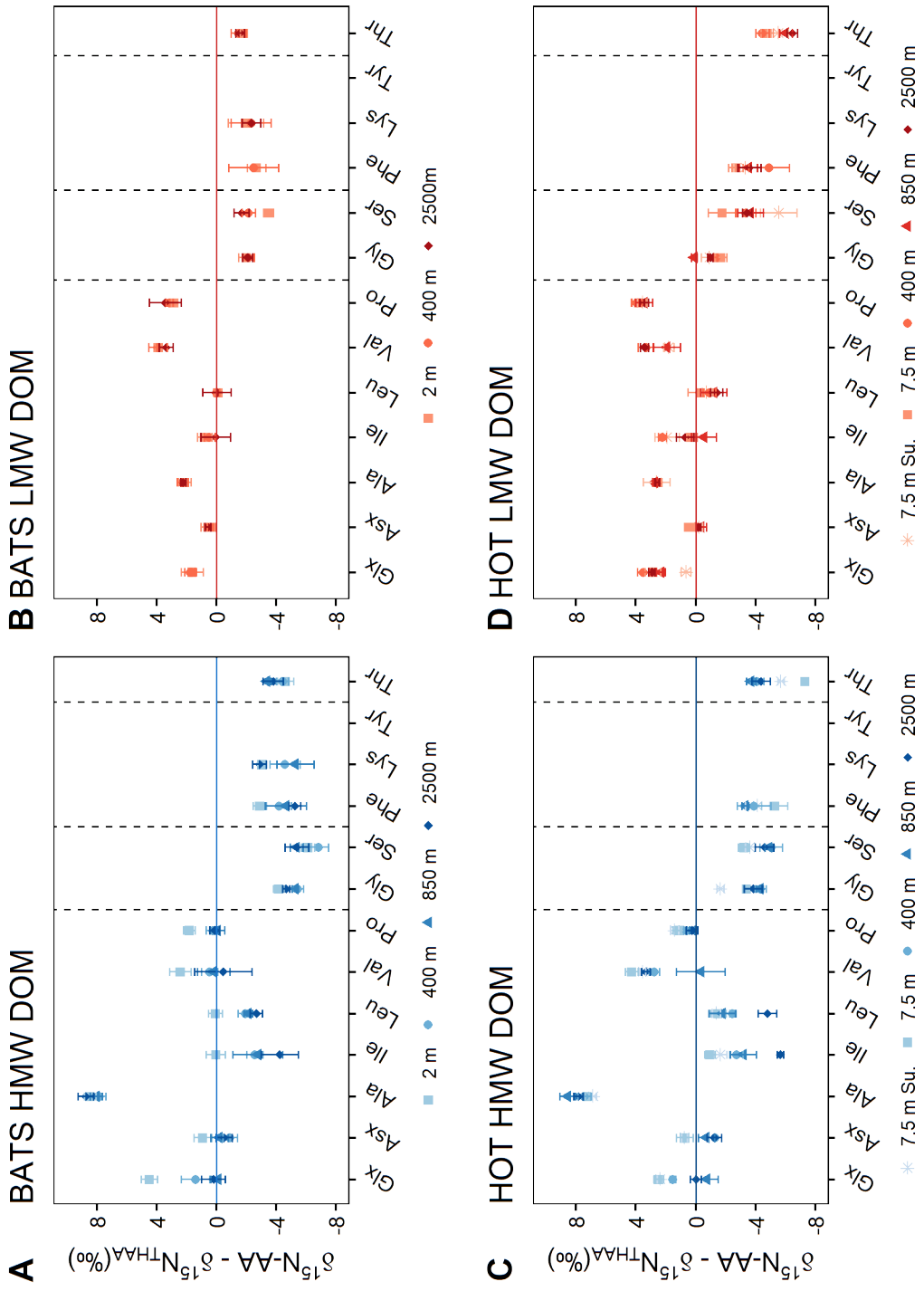
Figure S3.3: δ^{15}N-AA patterns normalized to δ^{15}N$_{THAA}$ of A) HMW DON at BATS, B) LMW SPE-DON at BATS, C) HMW DON at HOT, and D) LMW SPE-DON at HOT. Normalization to THAA does not alter the δ^{15}N-AA pattern, but shifts all patterns to the same scale. "Int." stands for Intermediate AA, while "Met." stands for Metabolic.
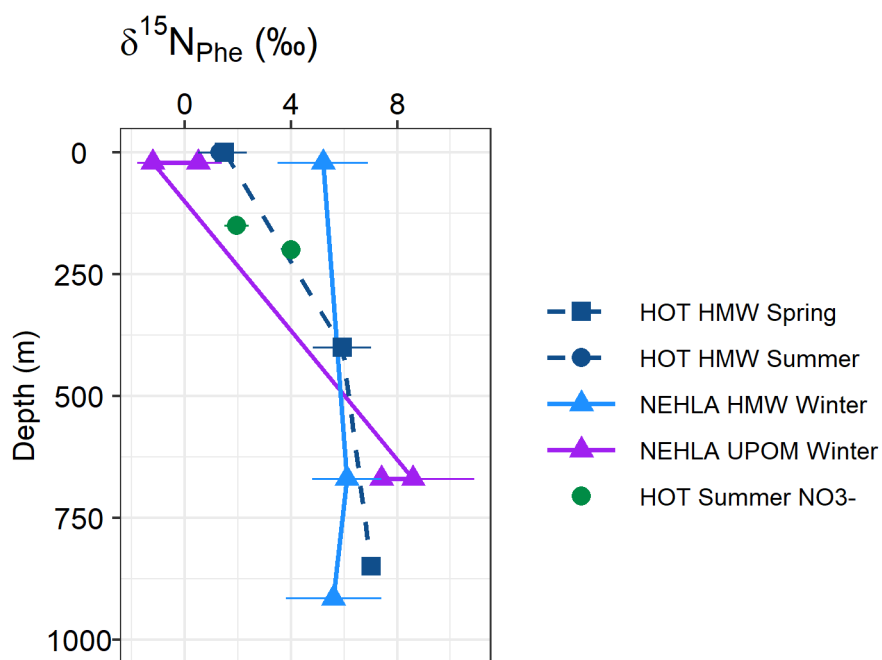
Figure S3.4: $\delta^{15}N_{Phe}$ values of HMW DON at HOT reported in this study (dark blue) compared with literature values of HMW DON (light blue) and suspended PON (purple) collected at NEHLA (Yamaguchi & McCarthy, 2018). $\delta^{15}N_{Phe}$ of surface values in spring (squares) and summer (circles) measured in this study are within error of $\delta^{15}N$ of NO_3^- measured in summer (dark green circles, Casciotti et al., 2008, Knapp et al., 2011). Subsurface $\delta^{15}N_{Phe}$ values of HMW DON in this study are within error of previous HMW DON and suspended PON measurements made in winter.
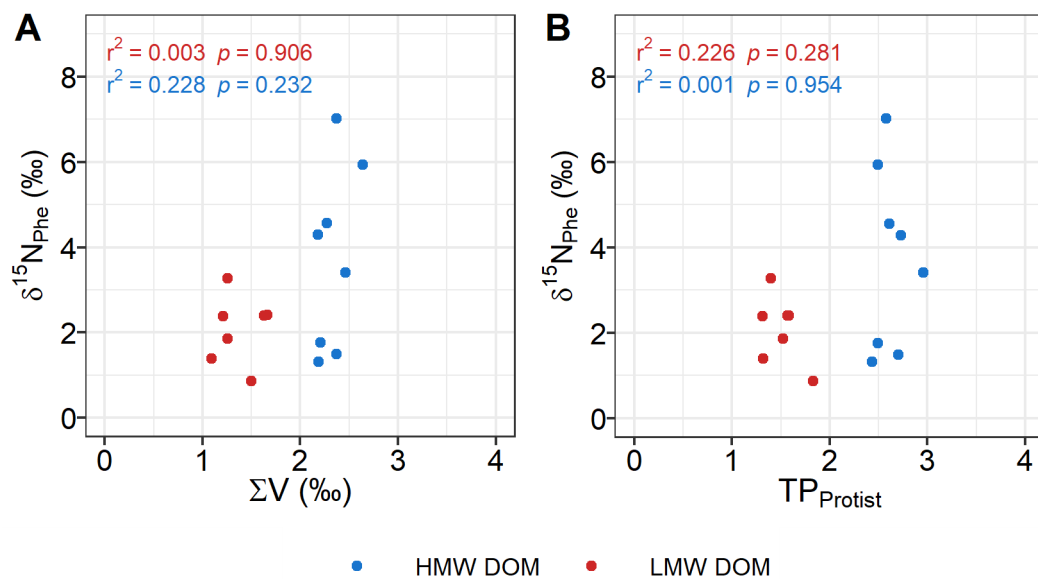
Figure S3.5: $\delta^{15}N_{Phe}$ of HMW DON (blue) and LMW SPE-DON (red) is not significantly correlated with A) ΣV or B) $TP_{Protist}$, suggesting microbial transformations do not alter $\delta^{15}N_{Phe}$. HMW and LMW SPE-DON include samples from HOT and BATS.
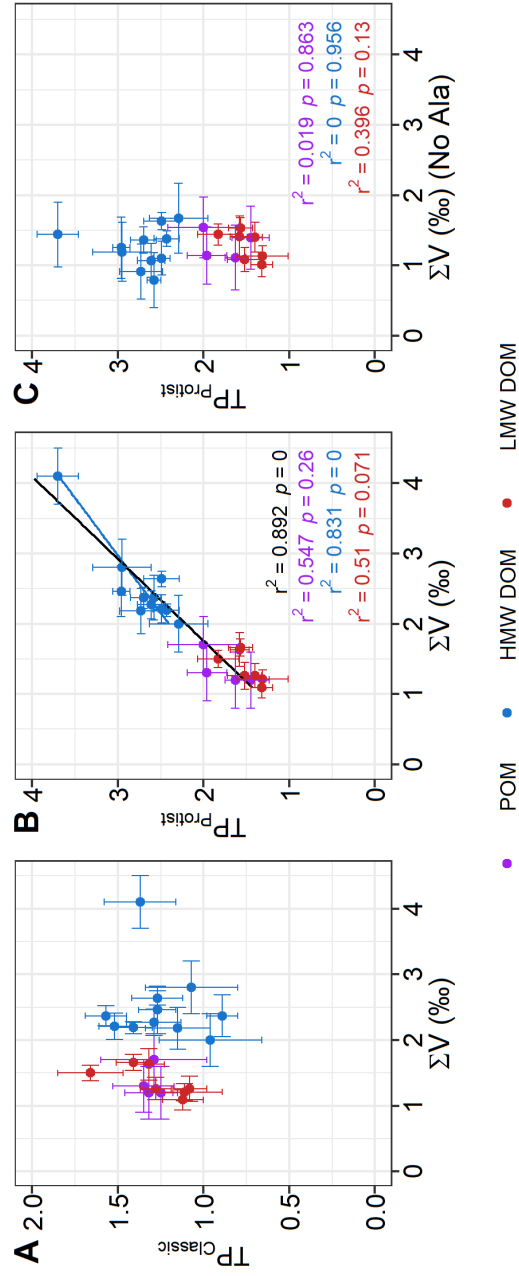
Figure S3.6: Relationships between different δ^{15}N-AA parameters in HMW DON (blue), LMW SPE-DON (red) and ultrafiltered PON (UPON, purple). A) TP$_{Classic}$ is not correlated with ΣV in HMW DON (blue), LMW SPE-DON (red) or UPON (purple). B) TP$_{Protist}$ and ΣV are significantly correlated in HMW DON as well as through the entire ON spectrum (black line), however, not are not correlated in LMW SPE-DON or UPON separately. C) If ΣV is calculated without Ala, there is no longer a significant relationship between the two variables, suggesting Ala is a key driver of HMW DON ΣV. LMW SPE-DON and UPON appear similar by all three proxies. HMW DON data includes measurements made at HOT and BATS in this study and three HMW DON samples measured at NEHLA from Yamaguchi et al. (2018). LMW SPE-DON data includes all samples measured in this study from both HOT and BATS. All UPON data is from Yamaguchi & McCarthy (2018), collected at NEHLA. Regression lines are plotted for significant linear relationships.

Table S3.1: Bulk δ^{15}N values of total, HMW, and LMW SPE-DON and NO$_3^-$ measured during cruises to BATS and HOT.

Location	N Type	Season	Year	Depth (m)	δ^{15}N	±
HOT	TDON	Summer	2014	2	3.80	0.16
HOT	TDON	Spring	2015	2	3.68	0.26
BATS	TDON	Summer	2015	2	2.50	0.21
BATS	TDON	Summer	2015	25	2.30	0.00
BATS	TDON	Summer	2015	50	2.45	0.00
BATS	TDON	Summer	2015	110	2.80	0.14
BATS	TDON	Summer	2015	150	3.15	0.07
BATS	TDON	Summer	2015	200	3.00	0.14
BATS	TDON	Summer	2015	250	2.35	0.00
BATS	TDON	Spring	2016	2	3.54	0.00
BATS	NO$_3^-$	Summer	2015	150	2.30	0.00
BATS	NO$_3^-$	Summer	2015	200	2.30	0.00
BATS	NO$_3^-$	Summer	2015	250	2.30	0.00
HOT	HMW DON	Summer	2014	7.5	6.22	0.09
HOT	HMW DON	Summer	2014	400	6.47	0.08
HOT	HMW DON	Summer	2014	850	6.85	0.11
HOT	HMW DON	Summer	2014	2500	6.82	0.21
HOT	HMW DON	Spring	2015	7.5	6.68	0.30
HOT	HMW DON	Spring	2015	400	6.59	0.11
HOT	HMW DON	Spring	2015	850	7.09	0.05
HOT	HMW DON	Spring	2015	2500	7.02	0.29
BATS	HMW DON	Summer	2015	2	3.80	0.09

Continued on next page

Location	N Type	Season	Year	Depth (m)	δ^{15}N	±
BATS	HMW DON	Summer	2015	400	5.89	0.25
BATS	HMW DON	Summer	2015	850	6.50	0.41
BATS	HMW DON	Summer	2015	2500	6.06	0.56
BATS	HMW DON	Spring	2016	2	4.06	0.09
BATS	HMW DON	Spring	2016	400	5.49	0.17
BATS	HMW DON	Spring	2016	850	6.33	0.13
BATS	HMW DON	Spring	2015	2500	5.78	0.63
HOT	LMW SPE-DON	Summer	2014	7.5	3.63	0.07
HOT	LMW SPE-DON	Summer	2014	400	3.43	0.20
HOT	LMW SPE-DON	Summer	2014	850	3.65	0.13
HOT	LMW SPE-DON	Summer	2014	2500	3.45	0.19
HOT	LMW SPE-DON	Spring	2015	7.5	4.04	0.01
HOT	LMW SPE-DON	Spring	2015	400	3.57	0.41
HOT	LMW SPE-DON	Spring	2015	850	3.13	0.21
HOT	LMW SPE-DON	Spring	2015	2500	3.43	0.16
BATS	LMW SPE-DON	Summer	2015	2	2.74	0.23
BATS	LMW SPE-DON	Summer	2015	400	3.03	0.24
BATS	LMW SPE-DON	Summer	2015	850	3.51	0.14
BATS	LMW SPE-DON	Summer	2015	2500	3.73	0.59
BATS	LMW SPE-DON	Spring	2016	2	2.79	0.21
BATS	LMW SPE-DON	Spring	2016	400	3.14	0.14
BATS	LMW SPE-DON	Spring	2016	850	3.25	0.06
BATS	LMW SPE-DON	Spring	2015	2500	3.29	0.11

Table S3.2: δ^{15}N-AA parameters of HMW and LMW SPE-DON.

Location	DON Type	Year	Depth	%N-AA	δ^{15}N$_{\text{THAA}}$	δ^{15}N$_{\text{Other}}$	TP$_{\text{Classic}}$	TP$_{\text{Protist}}$	ΣV
BATS	HMW	2016	2	7.97	4.65 ± 0.12	4.05 ± 0.54	1.52 ± 0.09	2.49 ± 0.1	2.21 ± 0.2
BATS	HMW	2016	400	5.51	8.74 ± 0.25	5.31 ± 0.67	1.29 ± 0.16	2.61 ± 0.16	2.27 ± 0.2
BATS	HMW	2016	850	5.4	8.92 ± 0.15	6.15 ± 0.84	1.15 ± 0.19	2.73 ± 0.25	2.18 ± 0.32
BATS	HMW	2016	2500	5.78	8.65 ± 0.23	5.63 ± 0.77	1.27 ± 0.11	2.96 ± 0.1	2.46 ± 0.36
BATS	LMW	2016	2	3.51	4.06 ± 0.17	2.75 ± 0.59	1.12 ± 0.12	1.32 ± 0.13	1.09 ± 0.15
BATS	LMW	2016	400	2.29	4.86 ± 0.14	3.06 ± 1.49	1.11 ± 0.22	1.31 ± 0.3	1.21 ± 0.13
BATS	LMW	2016	2500	1.81	4.96 ± 0.17	3.17 ± 1.06	ND ± ND	ND ± ND	1.21 ± 0.25
HOT	HMW	2014	7.5	10.48	6.39 ± 0.10	6.18 ± 0.31	1.41 ± 0.11	2.43 ± 0.15	2.19 ± 0.09
HOT	HMW	2015	7.5	8.52	6.77 ± 0.11	6.69 ± 0.43	1.57 ± 0.12	2.7 ± 0.15	2.37 ± 0.15
HOT	HMW	2015	400	7.11	9.83 ± 0.18	6.35 ± 0.46	1.27 ± 0.15	2.49 ± 0.21	2.64 ± 0.11
HOT	HMW	2015	850	6.51	10.31 ± 0.23	6.88 ± 0.67	0.89 ± 0.09	2.58 ± 0.08	2.37 ± 0.32
HOT	HMW	2015	2500	6.09	10.65 ± 0.21	6.76 ± 0.47	ND ± ND	ND ± ND	3.31 ± 0.13
HOT	LMW	2014	7.5	3.87	5.18 ± 0.22	3.54 ± 0.73	1.08 ± 0.1	1.52 ± 0.2	1.26 ± 0.19
HOT	LMW	2015	7.5	3.29	5.96 ± 0.16	3.93 ± 0.67	1.28 ± 0.09	1.4 ± 0.09	1.26 ± 0.17
HOT	LMW	2015	400	2.34	5.78 ± 0.12	3.55 ± 1.27	1.66 ± 0.19	1.83 ± 0.24	1.5 ± 0.12
HOT	LMW	2015	850	2.24	5.92 ± 0.14	3.04 ± 1.44	1.32 ± 0.09	1.58 ± 0.11	1.63 ± 0.24
HOT	LMW	2015	2500	1.95	6 ± 0.09	3.35 ± 0.69	1.41 ± 0.1	1.57 ± 0.14	1.66 ± 0.12

Table S3.3: Individual δ^{15}N-AA values of HMW and LMW SPE-DON.

Location	DON Type	Year	Depth	Glx	Asx	Ala	Ile	Leu	Pro	Val
BATS	HMW	2016	2	9.14 ± 0.52	5.58 ± 0.59	13.1 ± 0.33	4.71 ± 0.62	4.73 ± 0.42	6.47 ± 0.34	7.09 ± 0.69
BATS	HMW	2016	400	10.15 ± 0.92	7.83 ± 0.42	16.55 ± 0.33	6.18 ± 0.49	6.81 ± 0.43	8.96 ± 0.4	9.21 ± 0.49
BATS	HMW	2016	850	8.81 ± 0.48	8.73 ± 0.57	16.96 ± 0.36	6.15 ± 1.66	6.8 ± 0.23	8.85 ± 0.46	9.11 ± 1.09
BATS	HMW	2016	2500	8.86 ± 0.77	8.15 ± 0.5	17.4 ± 0.47	4.44 ± 1.22	5.98 ± 0.33	8.77 ± 0.22	8.21 ± 1.91
BATS	LMW	2016	2	5.69 ± 0.72	4.66 ± 0.44	6.2 ± 0.4	4.87 ± 0.45	3.95 ± 0.11	6.89 ± 0.07	7.97 ± 0.16
BATS	LMW	2016	400	6.59 ± 0.36	5.3 ± 0.14	7.14 ± 0.34	5.32 ± 0.32	4.85 ± 0.14	8 ± 0.19	8.78 ± 0.61
BATS	LMW	2016	2500	6.63 ± 0.05	5.55 ± 0.09	7.21 ± 0.04	5.02 ± 0.98	4.95 ± 0.92	8.4 ± 1.05	8.35 ± 0.44
HOT	HMW	2014	7.5	7.82 ± 0.17	6.21 ± 0.17	12.31 ± 0.17	3.79 ± 0.4	4.05 ± 0.25	6.88 ± 0.25	9 ± 0.16
HOT	HMW	2015	7.5	9.25 ± 0.32	7.51 ± 0.56	14.04 ± 0.08	5.82 ± 0.35	5.27 ± 0.51	7.86 ± 0.41	11.05 ± 0.41
HOT	HMW	2015	400	11.37 ± 0.08	8.55 ± 0.4	17.29 ± 0.4	7.12 ± 0.4	7.36 ± 0.11	10.43 ± 0.06	12.6 ± 0.31
HOT	HMW	2015	850	9.58 ± 0.72	9.6 ± 0.49	18.9 ± 0.46	7.14 ± 0.84	8.54 ± 0.86	10.47 ± 0.11	9.98 ± 1.61
HOT	HMW	2015	2500	10.65 ± 0.31	9.41 ± 0.45	18.51 ± 0.23	4.99 ± 0.07	5.86 ± 0.6	10.92 ± 0.33	13.99 ± 0.2
HOT	LMW	2014	7.5	5.84 ± 0.25	4.98 ± 0.12	7.8 ± 0.87	7.1 ± 0.8	4.46 ± 0.27	8.75 ± 0.29	6.97 ± 0.26
HOT	LMW	2015	7.5	8.83 ± 0.49	6.39 ± 0.29	8.49 ± 0.17	6.21 ± 0.31	5.65 ± 0.8	9.78 ± 0.5	ND
HOT	LMW	2015	400	9.29 ± 0.39	5.62 ± 0.11	8.52 ± 0.08	8 ± 0.24	4.87 ± 0.4	9.71 ± 0.32	9.11 ± 0.51
HOT	LMW	2015	850	8.26 ± 0.29	5.56 ± 0.31	8.67 ± 0.03	5.39 ± 0.82	4.72 ± 0.86	9.38 ± 0.53	7.87 ± 0.91
HOT	LMW	2015	2500	8.94 ± 0.19	5.82 ± 0.12	8.58 ± 0.14	6.73 ± 0.58	4.6 ± 0.39	9.45 ± 0.28	9.46 ± 0.26

Continued on next page

Location	DON Type	Year	Depth	Gly	Ser	Lys	Tyr	Phe	Thr
BATS	HMW	2016	2	0.5 ± 0.31	-1.45 ± 0.38	1.62 ± 0.22	7.6 ± 0.79	1.76 ± 0.42	0.04 ± 0.51
BATS	HMW	2016	400	3.3 ± 0.27	1.93 ± 0.63	4.16 ± 0.98	11.57 ± 0.6	4.56 ± 0.82	4.33 ± 0.33
BATS	HMW	2016	850	3.67 ± 0.14	3.56 ± 0.4	3.63 ± 1.24	ND	4.29 ± 1.37	5.55 ± 0.2
BATS	HMW	2016	2500	4 ± 0.05	3.29 ± 0.76	5.77 ± 0.4	ND	3.41 ± 0.31	4.86 ± 0.64
BATS	LMW	2016	2	2.08 ± 0.47	0.58 ± 0.3	2.1 ± 1.19	ND	1.39 ± 0.61	2.27 ± 0.11
BATS	LMW	2016	400	2.74 ± 0.36	2.71 ± 0.42	2.55 ± 1.34	5.81 ± 4.53	2.39 ± 1.66	3.38 ± 0.49
BATS	LMW	2016	2500	2.88 ± 0.27	3.3 ± 0.49	2.64 ± 0.61	ND	ND	3.4 ± 0.24
HOT	HMW	2014	7.5	3.8 ± 0.21	1.8 ± 0.33	ND	ND	1.32 ± 0.82	-0.26 ± 0.19
HOT	HMW	2015	7.5	3.39 ± 0.15	3.54 ± 0.35	ND	ND	1.49 ± 0.86	-0.55 ± 0.2
HOT	HMW	2015	400	5.5 ± 0.33	4.78 ± 0.73	ND	ND	5.93 ± 1.1	5.72 ± 0.42
HOT	HMW	2015	850	6.02 ± 0.03	5.41 ± 0.16	4.45 ± 0.73	ND	7.02	6.6 ± 0.2
HOT	HMW	2015	2500	6.81 ± 0.58	6.05 ± 0.59	ND	ND	ND	6.29 ± 0.59
HOT	LMW	2014	7.5	4.27 ± 0.52	-0.37 ± 1.21	ND	ND	1.86 ± 0.74	-0.34 ± 0.64
HOT	LMW	2015	7.5	4.28 ± 0.36	4.19 ± 0.92	ND	ND	3.27 ± 0.46	1.18 ± 0.37
HOT	LMW	2015	400	4.46 ± 0.08	2.44 ± 0.68	2.12 ± 3.85	ND	0.87 ± 1.35	1.32 ± 0.43
HOT	LMW	2015	850	6.04 ± 0.14	2.26 ± 0.84	4.48 ± 3.75	ND	2.4 ± 0.61	-0.1 ± 0.39
HOT	LMW	2015	2500	5.04 ± 0.16	2.56 ± 0.29	ND	ND	2.41 ± 0.76	-0.44 ± 0.35

Chapter 4

Advanced solid-state NMR to characterize refractory DOC and DON in the sea

4.1 Abstract

Marine dissolved organic matter (DOM) contains as much carbon as all atmospheric CO_2 and represents the largest reservoir of fixed nitrogen in the world ocean. As a result, marine DOM plays a key role in carbon sequestration and serves as the base of marine food webs. Yet, despite its global importance, most marine DOM remains molecularly uncharacterized. Due to the highly complex nature of DOM, solid-state nuclear magnetic resonance (NMR) spectroscopy has been among the most powerful techniques to understand its overall functional composition. Past experiments on isolated DOM fractions have shaped our understanding of DOC and DON main structural distributions. ^{15}N NMR experiments indicated DON has an entirely amide composition at all depths, while ^{13}C NMR experiments found DOC is dominated by sugars in the surface and operational "CRAM" (carboxyl rich alicyclic molecules) material in the deep, with limited unsaturated or aromatic content. However, it has long been known that standard cross polarization (CP) magic angle spinning (MAS) NMR, the technique used for almost all past DOM NMR studies, is at best "semi-quantitative" and can substantially underestimate non-protonated functional groups. In addition, most past NMR data has been on selected DOM fractions, mainly high molecu-

142

lar weight (HMW) material, which is now understood to represent mostly ^{14}C - young, "semi-labile" compounds. In contrast, there is much more limited information regarding the composition of low molecular weight (LMW) DOM, which represents the vast majority of the ocean's accumulated DON pool.

Here, we applied for the first time advanced solid-state ^{15}N and ^{13}C NMR methods optimized for resolving non-protonated functionalities to HMW DOM and LMW SPE-DOM from the surface and deep North Pacific Subtropical Gyre. We first describe the optimization and results of multiCP/MAS NMR experiments, synoptically applied to both HMW and LMW SPE-DOC and DON. For ^{15}N, this approach has never previously been applied to any natural organic nitrogen, but experiments with standards demonstrate that it can overcome the quantitation issues of non-protonated N moieties observed with CP/MAS NMR. For ^{13}C NMR, we demonstrate for the first time a comparison of the multi-CP approach with traditional CP/MAS, showing that improved CP/MAS approaches can in fact achieve close to quantitative results.

Using these new techniques, we find that both HMW DON and LMW SPE-DON have much more diverse structure than previously believed. We show a previously "invisible" heterocyclic component exists even in the semi-labile HMW material, and show that LMW material may be almost entirely heterocyclic N molecules. We use these data to hypothesize that inherently stable molecules in both size fractions contribute to the refractory nature of most marine DON. We also reinterpret the functional composition of marine DOC in context of these new N results. Even with these new methods optimized for unsaturated C, we find that aromatic functionalities are only a minor portion of both HMW and LMW SPE-DOC and suggest the majority of aromatic DOC may be represented by heterocyclic N molecules. Overall, these first results based on advanced and far

more quantitative solid-state NMR techniques likely represent the most accurate picture of DON and DOC functional composition and have broad implications for our understanding of marine dissolved organic matter structure and cycling.

4.2 Introduction

Marine dissolved organic matter (DOM) represents the second largest active carbon reservoir in the world (Hedges, 1992) as well as the largest reservoir of fixed nitrogen. This material is of wide importance on many fronts, including forming the basis for the microbial loop, helping to shape nutrient fields, and representing a key control on marine primary productivity. However, despite decades of research, our understanding of marine DOM source, cycling, and chemical structure remains limited (Hansell & Carlson, 2015). This is in large part due to difficulties characterizing the incredibly complex chemical composition of marine DOM, with less than 10% of marine DOM identified at the molecular level (Repeta, 2015). Only a very small fraction of DOM can be molecularly identified via wet chemical analyses, meaning alternative techniques that can provide a broad view of functionality are first necessary to guide targeted molecular-level investigations.

Perhaps the most widely applied approach to date is solid-state nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy probes the chemical composition of a sample by applying a magnetic field and measuring the response of nuclei with non-zero spins to excitation by pulses of electromagnetic radiation. The response frequency of different nuclei to these pulses is dictated by their electron distribution, which is dependent on the type of nuclei and bonds in a molecule. Fourier transformation of the time-domain responses of different nuclei to irradiation pulses results in a spectrum with chemical shifts on the x-axis and intensity on the y-axis, thus yielding information regarding overall molecu-

lar structure of a sample. ^{15}N NMR spectroscopy specifically is one of the few techniques with the ability to visual all N functionality in a sample.

While liquid-state NMR is essentially quantitative, it requires purification, can only resolve the fraction of DOM which will dissolve in NMR solvents, and can result in overlapping lines blurring most detailed structural information (Hertkorn et al., 2006; Stuermer & Payne, 1976). In contrast, solid-state NMR has the advantage of visualizing all atoms and functionalities in a sample, even if they cannot be dissolved. At the same time, the most common solid-state NMR approach, cross polarization (CP) magic angle spinning (MAS), has the disadvantage of detecting functionalities based on proton density, meaning unsaturated nuclei may be substantially underestimated. Additionally, solid-state NMR requires large purified organic samples, so large-scale isolation methods are required to recover sufficient organic material. To date, ultrafiltration, isolating the high molecular weight (HMW) fraction of the DON pool has been the most widely used isolation approach for solid-state CP/MAS experiments. However, it is now well understood that this material is the ^{14}C-younger, semi-labile component, while most DOM in the world ocean is far older low molecular weight (LMW) material. This isolation filter, together with the quantitation limitations of CP/MAS experiments, have likely contributed to several persistent conundrums regarding both dissolved organic nitrogen (DON) and dissolved organic carbon (DOC) composition.

Marine DON is much more abundant than inorganic N throughout the worlds oligotrophic surface ocean, yet primary production in these regions is generally nitrogen limited, suggesting most marine DON is nonbioavailable. Yet, ^{15}N CP/MAS NMR experiments of ultrafiltered, HMW DON indicated an essentially all amide composition (Aluwihare et al., 2005; Broek et al., Submitted; McCarthy et al., 1997), suggesting most HMW DON is amino acids (AA) with a greater

amino sugar (AS) contribution in the surface (Aluwihare et al., 2005). These results have been perplexing in the context of the apparent refractory nature of most DON: if marine DON is entirely AA and AS, supposedly labile biomolecules, why does it persist in the ocean for so long? Additionally, molecular level recoveries of amide containing molecules are much lower than those suggested by NMR, with at most 15% of HMW DON recovered as AA and AS (Benner, 2002; Kaiser & Benner, 2009). Subsequent work hypothesized that molecular-level recoveries were low due to the presence of hydrolysis resistant molecules and suggested and up to ∼65% of surface DON and ∼30% of deep DON could be accounted for by AA and AS (Aluwihare et al., 2005), but this still leaves a significant portion of HMW DON uncharacterized.

Recently, however, the first ^{15}N CP/MAS NMR experiments of LMW SPE-DON were reported, showing that this size fraction has a completely different composition than was suggested by HMW experiments. Instead of amide, LMW SPE-DON appeared dominated by complex, heterocyclic N containing molecules (Broek et al., Submitted). While the identity of these molecules remains unknown, these results imply a dominant, heterocyclic nitrogen component to the refractory (R)DON pool that was previously unrecognized. However, these results also present an additional puzzle: how can heterocyclic functional groups dominate LMW SPE-DON, yet are completely absent from the HMW DON pool? Further, if unsaturated N forms dominate the DON pool, this suggests a major possible analytical issue, as these are exactly the functional groups that traditional CP/MAS poorly detects. This means that the quantitative functional distribution of DON likely remains an open question.

In contrast to ^{15}N NMR experiments, past ^{13}C NMR studies of marine DOC were more common. Early ^{13}C CP/MAS NMR studies indicated composition

146

varies strongly depending on isolation method, with HMW DOC dominated by complex polysaccharides with increasing concentrations of carboxyl and alkyl contributions with depth (Aluwihare et al., 1997; Benner et al., 1992; McCarthy & Benner, 1993). Since then, liquid-state ^{13}C NMR combined with high resolution mass spectrometry indicated solid phase extracted (SPE) DOM is $\sim 50\%$ carboxyl-rich alicyclic matter (CRAM), and suggested these molecules represent some of the most refractory DOC functional groups (Hertkorn et al., 2013). Only recently were ^{13}C CP/MAS NMR results of LMW SPE-DOC published, indicating a composition dominated by alkyl, alkoxy, and carboxyl carbon, supporting the theory that CRAM represents some of the most refractory DOC molecules. Still, the low aromatic contribution to refractory (R)DOC has always been somewhat surprising given the highly recalcitrant nature and old ^{14}C ages, as well as in context of the recent ^{15}N NMR data indicating a dominant heterocyclic N contribution to LMW SPE-DON pool (Broek et al., Submitted).

Importantly, almost all previous solid-state NMR experiments of DOM relied on standard CP/MAS pulse sequences, which are well understood to be only semi-quantitative. This technique enhances the signal of target nuclei (e.g., ^{13}C or ^{15}N) via cross polarization with nearby ^{1}H nuclei (Pines et al., 1972, 1973). However, the transfer of magnetization is faster for carbon and nitrogen nuclei directly bonded to protons (Metz et al., 1996). In practice, this results in CP/MAS NMR often underestimating non-protonated functional groups. For example, ^{13}C NMR experiments on standard compounds and soils found the signal intensity of CP/MAS NMR could range from 30% to 100% of the total expected signal depending on functional composition, with the most underestimation of non-protonated aromatics and carbonyl groups (Smernik & Oades, 2000a, 2000b). These problems are even more severe in ^{15}N CP/MAS due to the low abundance of ^{15}N

coupled with its lower gyromagnetic ratio. With ^{15}N CP/MAS, signal intensity can be underestimated by nearly an order of magnitude for some unsaturated N functions (Smernik & Baldock, 2005). Together, this means that essentially all marine DOM solid-state NMR spectra collected to date may have errors, perhaps very large ones, if they contain substantial proportions of functional groups that CP/MAS does not detect well. In addition, and perhaps more importantly, we do not know in what part of spectra errors may be, or what their magnitude is. This is particularly true for refractory LMW SPE-DOC and DON, which make up most of the marine DOM reservoir and appear to be dominated by traditionally underestimated functional groups, including carboxyl-rich and/or unsaturated compounds. Collectively, this suggests that while this central technique to has provided an overall "global" view of overall DOC and DON composition, we may have no idea what accurate, quantitative spectra look like.

Further, most CP/MAS experiment on ocean DOM are decades old, yet solid-state NMR approaches have developed rapidly during this time. Advances in the last 20 years have specifically targeted quantitation, addressing many of these past uncertainties. While direct polarization (DP)/MAS has been a long-standing approach which can obtain fully quantitative solid-state NMR spectra, the sensitivity is much lower, and critically the nuclei relaxation times (and so experiment times) are vastly longer than in CP experiments. This means that in practice ^{13}C DP/MAS experiments are essentially prohibitive to implement for marine DOM, requiring extraordinarily long periods of instrument time and yielding low resolution spectra. Due to the even greater limitations of the ^{15}N nuclei, ^{15}N DP/MAS experiments are unheard of. However, recently, new methods have been developed which can produce quantitative solid-state NMR spectra for both ^{15}N and ^{13}C without the long run times required for DP. Specifically, experiments using

multiple blocks of CP (multiCP) can combine the high sensitivity and speed of CP experiments while obtaining quantitative spectra (Duan & Schmidt-Rohr, 2017; Johnson & Schmidt-Rohr, 2014). While these newer techniques have never previously been applied to marine DOM, they are now the leading edge in solid-state NMR research for organics and offer a major opportunity to for the first time quantitatively probe the nature of both DOC and DON in the deep ocean LMW reservoir.

Here, we apply ^{15}N and ^{13}C multiCP/MAS NMR to HMW and LMW DOM from the surface and deep Pacific Ocean. These sampling locations represent some of the youngest and oldest DOM in the ocean, allowing for a direct comparison of DOC and DON functional group composition across a wide range of DOM age and, size, and reactivity. These novel NMR approaches allow us to address multiple major questions regarding marine DOM structure. First, by using ^{15}N NMR methods optimized for non-protonated N, we investigate the longstanding discrepancy between past ^{15}N NMR data and molecular-level analyses of HMW DON. We found that while there is an important component of HMW DON which cannot be observed using ^{15}N CP/MAS NMR, unidentified and likely novel amide functional groups may represent most refractory HMW DON compounds. Additionally, we expand recent results indicating LMW SPE-DON is dominated by heterocyclic N structures, with quantitative multiCP/MAS NMR indicating a range of heterocyclic N functionality which likely represents the large majority of the ocean's DON reservoir. Finally, we use ^{13}C multiCP/MAS NMR to provide the first wide ranging set of quantitative DOC NMR spectra throughout the water column in both HMW and LMW fractions. In contrast to ^{15}N results, this new ^{13}C data confirms a surprisingly low fraction of unsaturated and aromatic C structures throughout the water column and in all MW fractions, even in the

oldest and presumably most refractory material in the deep sea. Together with ^{15}N multiCP/MAS NMR data, these new data suggest that most aromatic DOC structures may be nitrogenous. Together, we suggest that these results are the most representative of true marine DOC and DON overall functional composition to date and imply that inherently stable chemical composition is a major control on both DOC and DON recalcitrance.

4.3 Methods

4.3.1 DOM collection and molecular weight isolation

Samples were collected at the Hawaii Ocean Time Series (HOT) Station ALOHA (A Long-term Oligotrophic Habitat Assessment, 22° 45'N, 158° 00'W) aboard the *R/V Kilo Moana* in August 2014 and May 2015. Extensive details of the sampling and sample isolation protocols are described in (Broek et al., 2017). Briefly, surface samples were collected on each vessel via underway sampling systems at approximately 7.5 m. Subsurface samples ($\sim$ 3000 L) were collected via CTD casts at 2500 m depth. Samples were prefiltered through 53 µm Nitex mesh and pumped through 0.2 µm polyether sulfone (PES) cartridge filters. HMW DOM was concentrated using large volume tangential-flow ultrafiltration (UF) with a concentration factor of 1000 and a MW cut off of 2.5 kDa. LMW DOM was collected via solid phase extraction of the UF permeate using PPL sorbent. All samples were stored in a desiccator under vacuum with Drierite desiccant and NaOH pellets to ensure complete dryness for NMR experiments.

4.3.2   Solid-state ^{15}N NMR experiments

^{15}N CP/MAS and multiCP/MAS of standards

A multiCP/MAS NMR pulse sequence was optimized for ^{15}N after Johnson and Schmidt-Rohr (2014) and Duan and Schmidt-Rohr (2017) and tested on standard compounds guanine (Sigma Aldrich) and chlorin e6 (Combi-blocks). All ^{15}N NMR experiments were collected on a Bruker Avance III spectrometer operating at 400.1 MHz for ^{1}H and 40.55 MHz for ^{15}N with a Bruker 4 mm probe. The MAS rate was 12 kHz, and the probe was kept at 295 K. All spectra were referenced to ^{15}N-labeled Glycine (Cambridge Isotope Laboratories) at 33.4 ppm (Bertani et al., 2014). The ^{1}H $\pi/2$ was reoptimized on ^{15}N-labeled Glycine prior to each experiment and ranged from 4.88 µsec 5.25 µsec. The ^{15}N $\pi/2$ was 5.75 µsec. Tppm15 and cw_13 decoupling was used. The multiCP contact time was 75 µsec and the last multiCP contact time was 200 µsec. The multiCP sequence was looped a total of 10 times. For guanine, the proton repolarization time was 2.2 sec, the pulse delay was 4.4 sec, the dwell time was 1.03 µsec, 2700 scans were collected, 2048 points were used for the Fourier transform, and 10 Hz of line broadening was applied. For chlorin e6, the proton repolarization time was 1.4 sec, the pulse delay was 2.8 sec, the dwell time was 12.0 µsec, 11,500 scans were collected, 350 points were used for the Fourier transform, and 100 Hz of line broadening was applied.

^{15}N CP/MAS experiments of guanine and chlorin e6 were also performed to compare the two methods (Fig. 4.1). All common parameters between the CP/MAS and multiCP/MAS pulse sequence (spinning speed, probe temperature, chemical shift references, ^{15}N $\pi/2$, ^{1}H $\pi/2$, pulse delay, dwell time, points for Fourier Transform, and line broadening) were kept constant. Tppm15 decoupling was used with a contact time of 2.0 ms. For guanine, 12,500 scans were collected

and for chlorin e6, 11,500 scans were collected.

^{15}N multiCP/MAS of DOM samples

^{15}N multiCP/MAS parameters for DOM samples were the same as those used for guanine and chlorin e6 with the following exceptions. For all DOM samples, the proton repolarization time was 0.3 seconds, and the pulse delay was 0.6 seconds. For the surface, HMW DOM sample, a dwell time of 1.033 µsec and 2048 points for the Fourier transform were used. For all other DOM samples, a dwell time of 12 µsec and 350 points for the Fourier transform were used. 100 Hz of line broadening was applied to all DOM samples. The total number of scans for each DON experiment ranged from 88,000 to 350,000.

Data integration

The relative distribution of N functional groups was determined using DM-Fit (Massiot et al., 2002). The number of peaks, approximate chemical shift, and Gaussian/Lorentzian lineshape ratio was manually selected, and peak width, height, and final chemical shift was calculated by DMFit to fit the original spectrum (Fig. S4.1). Peak selection was verfied by visualizing the residual (original spectrum - modeled fit) (Fig. S4.2). If the residual showed evidence of peaks above the noise, the model was deemed underfit. If the residual had less noise in the area where peaks were selected compared to areas without peaks, it was deemed overfit. For samples with high signal to noise, a denoising procedure was also applied to verify peak integration (Srivastava et al., 2021) (Fig. S4.3). Peak identification was based on literature spectra, including Aluwihare et al. (2005) and Knicker (2004), accommodating for different chemical shift referencing when necessary (Bertani et al., 2014). Error integration data was calculated using

Monte Carlo error estimations based on 500 iterations. Further data processing and analysis was completed in Microsoft Excel or R Studio version 4.0.5 (R Core Team, 2021).

4.3.3 Solid-state ^{13}C NMR experiments

^{13}C CP/MAS and multiCP/MAS of standards

A ^{13}C multiCP/MAS NMR pulse sequence was optimized after Johnson and Schmidt-Rohr (2014) and Duan and Schmidt-Rohr (2017) and tested on standard compounds guanine and chlorin e6 (Fig. 4.2). All ^{13}C NMR experiments were collected using a Bruker Ascend spectrometer operating at 301.1 MHz for ^{1}H and 75.72 MHz for ^{13}C with a Revolution 2 mm probe. The MAS rate was 15 kHz, and the probe was kept at 295 K using a temperature correction to account for spinning speed with gas flow at 1000 lph. All spectra were referenced to adamantane (Sigma Aldrich) at 38.83 ppm. The ^{1}H $\pi/2$ was 2.45 µsec and the dwell time was 5.0 µsec. Spinal64 and cw_13 decoupling was used, the multiCP contact time was 55 µsec, and the ^{13}C $\pi/2$ was 2.75 µsec. The multiCP sequence was looped a total of 10 times. For guanine, the pulse delay was 2.6 sec and the proton repolarization time was 1.3 sec. 1000 scans were collected, 1534 points were used for the Fourier transform, and 10 Hz of line broadening was applied. For chlorin e6, the pulse delay was 2.4 sec and the proton repolarization time was 1.2 sec. 3000 scans were collected, 1024 points were used for the Fourier transform, and 10 Hz of line broadening was applied.

^{13}C CP/MAS experiments of guanine and chlorin e6 were also performed to compare the two methods. All common parameters between the CP/MAS and multiCP/MAS pulse sequence (spinning speed, probe temperature, chemical shift references, ^{13}C $\pi/2$, ^{1}H $\pi/2$, pulse delay, dwell time, points for Fourier Transform,

and line broadening) were kept constant. Cross polarization was achieved with a 70-100% power ramp on the ^{1}H nucleus and a cross polarization contact time of 5 seconds. Spinal64 decoupling parameters were optimized based on peak intensity of a mixture of glycine and adamantane. For guanine, 4000 scans were collected, and for chlorin e6, 1940 scans were collected.

All ^{13}C NMR spectra were corrected for background signal from the probe. A rotor packed with NaCl was run for the same number of scans with the same parameters as each sample and the resulting spectrum was subtracted from the corresponding sample spectrum.

^{13}C CP/MAS and multiCP/MAS of DOM samples

^{13}C CP/MAS and ^{13}C multiCP/MAS parameters and data correction for DOM samples were the same as those used for guanine and chlorin e6 with the following exceptions. For all DOM samples, the proton repolarization time was 0.3 seconds, and the pulse delay was 0.6 seconds. 512 points were used for the Fourier transform and 10 Hz of line broadening was applied. For the ^{13}C CP/MAS experiments, 1940 to 20,000 scans were collected, while for the ^{13}C multiCP/MAS experiments, 15,000 to 30,000 scans were collected.

Data integration

The relative distribution of C functional groups was determined using DM-Fit (Massiot et al., 2002). The number of peaks, approximate chemical shift, and Gaussian/Lorentzian lineshape ratio was manually selected, and peak width, height, and final chemical shift was modeled to fit the original spectrum via DM-Fit (Fig. S4.4). Peak selection was verfied by visualizing the residual (original spectrum - modeled fit). Error integration data was calculated using Monte Carlo

error estimations based on 500 iterations. ^{13}C chemical shift assignments were made according to Mao et al. (2012): ketone, aldehyde, quinone (220–191 ppm); COO, NC=O (191–164 ppm); aromatic C–O (164–150 ppm); aromatics (150–117 ppm); OCO (117-94 ppm); OC (94–60 ppm); OCH$_3$, NCH (60–45 ppm); CCH$_2$C, CCHC (45–30 ppm); and CCH$_2$C, CCH3 (30–0 ppm). These regions were also summarized into more general functional groups as follows: carbonyl C (220–164 ppm), aromatic C (164–117 ppm), alkoxy C (117–60 ppm), and alkyl C (60–0 ppm). Further data processing and analysis was completed in Microsoft Excel or R Studio version 4.0.5 (R Core Team, 2021).

4.4 Results

4.4.1 Recovery and C/N of DOM size fractions

Extensive recovery and compositional data for the HMW and LMW DOM size fractions are presented in Broek et al. (2017). Briefly, for the samples investigated in this study, HMW DOC recovery from total DOC was 16.3% in the surface and 7.8% at 2500 m (Table 4.1). HMW DON recovery from total DON was 16.2% in the surface and 13.9% at 2500 m. The C/N ratio of HMW DOM was similar between the surface and deep, with ratios of 12.3 and 13.1, respectively. The Δ^{14}C of HMW DOC was -50.0 ‰ (350 years) in the surface and -379.7 ‰ (3775 years) at 2500 m.

LMW SPE-DOC recovery from total DOC was 20.4% in the surface and 32.7% at 2500 m (Table 4.1). LMW SPE-DON recovery from total DON was 9.0% in the surface and 16.7% at 2500 m. The C/N ratio of LMW SPE-DON was 27.6 in the surface and 28.5 at 2500 m. The Δ^{14}C of LMW SPE-DOC was -343.0 ‰ (3310 years) in the surface and -577.6 ‰ (6860 years) at 2500 m.

4.4.2 ^{15}N CP/MAS versus ^{15}N multiCP/MAS of test compounds

The ^{15}N multiCP/MAS method was optimized on two standard compounds, guanine and chlorin e6. Substantial differences were observed in the CP/MAS verses multiCP/MAS spectra of guanine (Fig. 4.1, Table 4.2). Because there are five N atoms in guanine, we would expect each to have a relative percent intensity of 20%. Thus, anything below 20% indicates that the N atom was underestimated, while anything above 20% indicates the N atom was overestimated. Standard ^{15}N CP/MAS significantly underestimated the signal intensity of non-protonated heterocyclic N groups N1 and N_2 (as labeled in Fig. 4.1), with the relative percent intensity of each group 5.81% $\pm$ 0.48% and 7.24% $\pm$ 0.61%, respectively. The signal intensity of the singly protonated nitrogen atoms N3 and N4 (represented by a single peak) was overestimated, with a combined signal intensity of 55.55% $\pm$ 0.73% (compared to an expected intensity of 40%). The signal intensity of the NH2 group was also overestimated, with a total signal intensity of 31.4% $\pm$ 0.63%.

In contrast, using the ^{15}N multiCP/MAS, the signal intensity of all N atoms was substantially closer to the expected intensity. The relative signal intensity of non-protonated nitrogen atoms N1 and N_2 was 16.53% $\pm$ 1.12% and 15.23% $\pm$ 2.18%, respectively. This represents a three to four-fold increase in signal compared to the standard CP/MAS method. The relative signal intensity of singly protonated N3 and N4 (represented by a single peak) was also closer to expected signal intensities, representing 46.6% $\pm$ 1.54% of the total intensity (compared to an expected intensity of 40%). The signal intensity of the NH2 group was 21.64% $\pm$ 1.92%, within error to the expected signal intensity.

Because all N atoms in chlorin e6 are in one large aromatic ring, only one peak is observed for the four nitrogen atoms. Thus, there is no different in normalized signal intensity of this peak using the CP and multiCP methods. The chemical

shift of this peak was 133.6 ppm.

4.4.3 ^{13}C CP/MAS versus ^{13}C multiCP/MAS of test compounds

In contrast to the ^{15}N multiCP results, differences between the ^{13}C CP/MAS and ^{13}C multiCP/MAS spectra were less substantial for the two compounds tested here (Fig. 4.2, Table 4.3, Table 4.4). Guanine also has five carbon atoms, meaning if the method were fully quantitative, the signal intensity of each carbon should be 20%. ^{13}C CP/MAS underestimated the signal of the carbonyl group C4 (as labeled in Fig. 4.2), with a percent signal intensity of 12.36% ± 0.07%. Aromatic carbons C3 and C1 were both overestimated as 26.02% ± 0.24% and 23.48% ± 0.06%, respectively, while aromatic carbon C5 was close to the expected intensity, 19.5% ± 0.23%. The alkoxy carbon C2 was underestimated as 18.65% ± 0.06% of the total signal.

For all C atoms except C5, the ^{13}C multiCP/MAS integration results were closer to the expected 20%. ^{13}C multiCP/MAS underestimated the signal of the carbonyl group C4 (as labeled in Fig. 4.2), by less than that of CP/MAS, with a percent signal intensity of 16.59% ± 0.05%. Aromatic carbons C3 and C1 were overestimated by less than in the CP/MAS method, with relative signal intensities of 24.30% ± 0.16% and 21.50% ± 0.04%, respectively. Aromatic carbon C5 was underestimated by more than in the CP/MAS method, with a relative signal intensity of 17.28% ± 0.15%. The alkoxy carbon C2 was only slightly greater than 20%, with a relative signal intensity of 20.33% ± 0.04%. Overall, most differences in signal intensity between the two methods were ≤ 2%.

Due to the greater complexity of the chlorin spectrum, results are presented as total signal intensity for integrated regions (Fig 4.2, Table 4.4). The average propagated error for each peak in both the ^{13}C CP/MAS and multiCP/MAS spectra

157

was $\sim 1.1\%$ of the total integrated area. Differences in the relative intensity of C functional groups using the two methods were minor, ranging between 0.29% and 3.36%. When considering the four major functional group classes (carbonyl, aromatic, alkoxy, and alkyl), the multiCP/MAS method resulted in 0.29% less carbonyl signal, 0.72% less aromatic signal, 3.36% more alkoxy signal, and 2.35% less alkyl signal compared to CP/MAS results.

4.4.4 ^{15}N multiCP/MAS of marine DOM MW fractions

For HMW DON, ^{15}N multiCP spectra of surface HMW DON demonstrated a clear heterocyclic N contribution, though, surprisingly, no heterocyclic N contribution to deep HMW DON could be distinguished (Fig. 4.3). For surface HMW DON, a heterocyclic N peak at 142 ppm in the pyrrole/indole region represented $15\% \pm 4.2\%$ of the total signal (Table 4.5). An amide N peak at 122 ppm represented the remaining $85\% \pm 4.2\%$ of the total signal. In the deep ocean, an amide peak at 121 ppm represented the entire N signal intensity. While it also appears there may be a small heterocyclic signal at 142 ppm, this peak could only represent at most 6% of the total signal and cannot be confidently integrated above the noise. The amide peak in both surface and deep HMW DON had a similar chemical shift to the amide peak in all previous HMW DON spectra (Aluwihare et al., 2005; Broek et al., Submitted; McCarthy et al., 1997). No amine signal was observed in either ^{15}N multiCP spectra.

For LMW SPE-DON, the entire ^{15}N multiCP N signal in both the surface and the deep ocean is represented by heterocyclic N (Fig. 4.3). In the surface, this signal is distributed across three broad, but separate peaks in the pyrrole/indole region at 170 ppm, 144 ppm, and 127 ppm (representing $12.5\% \pm 0.6\%$, $12.4\% \pm 3.9\%$, and $75.1\% \pm 4.3\%$ of the total signal, respectively, Table 4.5, Fig. S4.1).

While it is possible there is a small amide shoulder on the peak at 127 ppm, it cannot be reasonably distinguished from the noise. This is supported by a denoising procedure, which results in a spectrum with three clearly defined heterocyclic peaks with the chemical shifts listed above (Fig. S4.3). At 2500 m, the signal is distributed across two peaks in the pyrrole/indole regions with chemical shifts of 172.8 ppm and 132.1 ppm, representing 11.3% $\pm$ 0.2% and 88.7% $\pm$ 0.2%, respectively. The denoised data also clearly demonstrates two peaks in the pyrrole/indole region (Fig. S4.3).

4.4.5 ^{13}C multiCP/MAS and CP/MAS of marine DOM MW fractions

While differences between the ^{13}C multiCP/MAS and CP/MAS methods were minor, they should nevertheless indicate that the multiCP/MAS results represent the most accurate estimate of DOC functional composition. ^{13}C multiCP/MAS spectra of HMW DOC had the greatest total contributions from alkoxy and alkyl functional groups, followed by carbonyl functional groups, with only minor aromatic contributions to all spectra (Fig. 4.4, Table 4.6). In the surface, alkoxy functional groups represented the greatest total intensity (63.2% $\pm$ 0.4%), while at 2500 m alkoxy and alkyl groups had similar contributions ($\sim$ 40%). Carbonyl functional groups represented more of the total signal in deep HMW DOC compared to surface HMW DOC (15.7% $\pm$ 0.04% vs. 10.0% $\pm$ 0.1%, respectively). In both the surface and deep, little to no signal was observed from ketone, aldehyde, and quinone groups. Aromatic signal was also slightly greater in the deep than in the surface, representing 2.8% $\pm$ 0.03% and 0.5% $\pm$ 0.1% of the total C signal in the surface and deep, respectively. The ratio of O-alkyl (117 – 94 ppm) to acetal C (94 – 60 ppm) was 5.1 in surface HMW DOC, similar to the expected ratio of hexoses (Sannigrahi et al., 2005). In contrast, in deep HMW DOC, the O-alkyl

to acetal ratio was 8.8, substantially higher than the expected ratio for sugars.

In LMW SPE-DOC, both surface and deep ^{13}C CP/MAS spectra had the greatest total contribution from alkyl groups, followed by alkoxy C, then carbonyl C, with only a small aromatic C contribution (Fig. 4.4, Table 4.6). In contrast to HMW DOC, LMW SPE-DOC surface and deep ^{13}C multiCP/MAS spectra were almost indistinguishable. In both surface and deep spectra, alkyl functional composition represented 52% of the total signal. Alkoxy C represented 23.6% ± 0.05% of the total C signal in the surface and 22.4% ± 0.2% of the total C signal in the deep (2500 m). Slightly more carbonyl signal was observed in deep LMW SPE-DOC compared to surface LMW SPE-DOC, representing 18.8% ± 0.02% and 16.0% ± 0.02% of the total C signal, respectively. Similar to HMW DOC, most of the carbonyl signal was due to COO and NC=O groups rather than ketone, aldehydes, and quinones. Finally, aromatic groups represented slightly more of the total C signal in the surface than the deep, contributing 8.2% ± 0.02% and 6.8% ± 0.02% of the total C signal, respectively. Again similar to HMW DOC, very little aromatic C-O signal in the region of 164 – 150 ppm was observed. The O-alkyl to acetal ratio was 30.5 in the surface and 24.1 in the deep, inconsistent with expectations for sugars.

Because of the substantially shorter instrument time required to collect a ^{13}C CP spectrum compared to a ^{15}N spectrum, ^{13}C CP/MAS spectra were also acquired for all DOM samples. This allows for a direct comparison of how these two methods differ specifically for the marine DOC matrix. Differences in the relative signal from the four major functional groups (carbonyl, alkyl, alkoxy, or alkyl) between the two methods were generally minor, ranging from ∼ 0% – 7% of the total signal (Fig. 4.5). Differences were greatest for surface, HMW DOC, where multiCP mas methods demonstrated 3.3% ± 0.4% less carbonyl signal,

0.9% ± 0.1% more aromatic signal, 6.9% ± 0.5% more alkoxy signal, and 4.5% ± 0.5% less alkyl signal. Differences were smaller in deep HMW DOC, where multiCP demonstrated 0.2% ± 0.1% more carbonyl signal, 1.7% ± 0.1% less aromatic signal, similar alkoxy signal, and 1.7% ± 0.1% more alkyl signal compared to CP/MAS.

Similar to the HMW DOC pool, in LMW SPE-DOC, differences between the two methods were greater in the surface than at 2500 m. For surface LMW SPE-DOC, multiCP/MAS resulted in 3.6% ± 0.03% less carbonyl signal, 0.3% ± 0.03% less aromatic signal, 1.5% ± 0.1% more alkoxy signal, and 2.3% ± 0.3% more alkyl signal compared to CP/MAS. For deep LMW SPE-DOC, multiCP/MAS resulted in 1.2% ± 0.03% less carbonyl signal, similar aromatic signal, 1.6% ± 0.03% more alkoxy signal, and 0.4% ± 0.1% less alkyl signal compared to CP/MAS.

4.5 Discussion

4.5.1 ^{15}N multiCP/MAS reveals new, diverse functional composition of HMW DON

These first ^{15}N multiCP/MAS results for HMW DON reveal a more diverse composition than has long been assumed. Past ^{15}N CP/MAS NMR results of HMW DON all indicated almost an entirely amide contribution to this subset of the marine DON pool, with minor amine contributions as the only other functional group (Aluwihare et al., 2005; Broek et al., Submitted; McCarthy et al., 1997). Together, this created a paradigm that much or possibly most of marine DON is composed of amino acids (AA) and amino sugars (AS). However, this has long been perplexing in context of the low molecular-level recoveries of AA and AS to marine DON (Benner, 2002; Kaiser & Benner, 2009), as well as the assumed lability of these biomolecules. However, the inherent quantitative limitations of

traditional solid-state NMR have always made this interpretation uncertain. It is possible the reason for the differences in NMR integration data and molecular-level recoveries is because past experiments utilized traditional CP/MAS NMR, known to substantially underestimate heterocyclic and non-protonated N groups. While fully quantitative ^{15}N DP/MAS (single pulse) NMR is essentially impossible due to prohibitive instrument time, ^{15}N multiCP/MAS NMR represents an ideal tool to investigate these discrepancies for the first time.

Heterocyclic N in surface HMW DON

To our knowledge, this is the first application of ^{15}N multiCP/MAS NMR to any natural organic matter sample. The three to four-fold increase in the signal intensity of non-protonated heterocyclic N atoms in guanine as determined by ^{15}N multiCP/MAS compared to CP/MAS clearly demonstrated the ability of this technique for near-quantitative ^{15}N spectra of non-protonated functional groups (Fig. 4.1, Table 4.2). This far more quantitative approach thus has wide applications beyond marine DON to understanding detrital ON composition in the geosphere.

In contrast to past CP/MAS NMR data on these same samples (Broek et al., Submitted), a significant heterocyclic contribution (15%) is now observed in the surface HMW material utilizing the ^{15}N multiCP/MAS pulse sequence (Fig. 4.3, Fig. 4.6). Some earlier ^{15}N CP/MAS NMR analyses noted a heterocyclic shoulder to the amide peak in HMW DON, though the signal was never high enough to confidently distinguish from the main amide peak (Aluwihare et al., 2005; McCarthy et al., 1997). It is possible a heterocyclic shoulder was more apparent in these earlier studies because of the different size cut off used for ultrafiltration of HMW DON samples. While the most recent work (and the samples in this

162

study) used a size cut off of 2.5 kDa, past work used a size cut off of 1 kDa, thus retaining smaller molecules in the HMW DON size fraction. Considering the clear compositional difference between HMW and LMW SPE-DON (Broek, 2019), we hypothesize that the higher size cut off used here may result in less heterocyclic contribution to our HMW DON samples. Further research would be needed to determine whether there are clear differences in composition, particularly hetero-cyclic content, between these relatively closely spaced molecular weight groupings. However, if this were true, it could have important implications for the sources and cycling of the surface semi-labile DON pool.

In contrast to results in the surface ocean, the lack of quantifiable heterocyclic signal to deep HMW DON was surprising. This reinforces a previous hypothe-sis based on the LMW SPE-DON pool suggesting that there are some surface-produced semi-labile heterocyclic N groupings (Broek et al., Submitted). The results presented here specific to the HMW pool suggest that HMW heterocyclic N molecules are semi-labile, with most surface HMW heterocyclic N degraded be-fore reaching the deep ocean. The apparent relative lability of HMW heterocyclic N may explain the relatively low proportion of heterocyclic N in HMW DON compared to the LMW SPE-DON pool.

While an exact identification cannot be made based on the data presented here, it is possible these heterocyclic molecules could be purines or pyrimidines. Fang et al. (2011) and Mao et al. (2007) reported a peak with a similar ^{15}N NMR chemical shift in HMW fulvic acids, which was associated with a signal ^{13}C NMR signal at 157 ppm and accounted for about a quarter of all N in their samples. It was suggested that the purine metabolite allantoin, which is produced by most living organisms, was the most likely candidate for this peak (Fang et al., 2011). Additional advanced solid-state NMR techniques, such as methods

observing nitrogen-bonded carbon nuclei (Schmidt-Rohr & Mao, 2002), could help verify the structure of HMW DON heterocyclic functionality.

Limited amine contribution to HMW DON

Notably, the multiCP data of both surface and deep HMW DON indicates no quantifiable intensity in the amide region (Fig. 4.3). This contrasts with previous work utilizing standard CP/MAS techniques, which report between $8\% - 15\%$ of the total signal is represented by amine groups at ~ 35 ppm (Aluwihare et al., 2005; Broek et al., Submitted). This may indicate that standard CP/MAS techniques are overestimating amine N compared to amide N. Assuming amine at 35 ppm represents NH_2 groups, while amide is represented by C-NH-CO (such as in protein), standard CP/MAS may produce more signal for the amine groups than the singly protonated amide groups. Our ^{15}N NMR results of guanine strongly support this interpretation, as the signal for the NH_2 group was 50% greater than expected in the CP/MAS experiment, while much closer to the expected intensity in the multiCP/MAS experiment (Fig. 4.1, Table 4.2). Further comparisons of CP vs. multiCP ^{15}N NMR methods on a range of molecular compounds can help determine if this is the case.

Overall, while there is likely some amine from AA and AS in our samples, it is quantitively small enough that it cannot be resolved from the noise. If it is assumed that most of the amide signal is proteinaceous material, then the lack of amine present in our data would suggest that persevered proteinaceous molecules may be substantially larger than implicated by past data. However, as discussed in the next section, it is possible there is additional amide N besides proteinaceous material, making this interpretation uncertain. Regardless, these ^{15}N multiCP/MAS data suggest that amine containing molecules were likely over-

estimated by past work utilizing CP/MAS NMR techniques.

Unidentified amide N

The ^{15}N multiCP/MAS technique applied here can address for the first time whether the difference in amide N contribution to HMW DON observed via ^{15}N NMR versus what is recovered using molecular level analyses is due to the biases of CP/MAS NMR. Past work measuring AS in HMW DON (Benner & Kaiser, 2003) and AA analyses of the these same samples (Ianiri et al., Submitted) indicate that $\sim 15\%$ of surface and $\sim 10\%$ of deep HMW DON can be recovered as amide N functionalities by molecular level techniques. While heterocyclic N in surface HMW DON represnted 15% of the total N signal, based on the underestimation of N in our standard, this could as high as 20%. While a similar estimate cannot be made for deep HMW DON, the lack of quantifiable heterocyclic N to this sample indicates any heterocyclic contribution is minor. Thus, non-protonated N functionalities overlooked by CP/MAS NMR clearly cannot account for the large offset in amide-N determined via molecular level techniques versus ^{15}N NMR. Instead, there must be an amide component to the HMW DON pool which is not recoverable by molecular level techniques. This unidentified amide material makes up $\sim 70\%$ of surface and $\sim 90\%$ of deep HMW DON, representing the majority of the HMW marine DON pool (Fig. 4.7). The greater contribution of this unknown amide to deep HMW DON compared to surface HMW DON indicates these molecules are likely resistant to microbial degradation and contribute to long-lived HMW DON.

There are at least two potential candidates for the unidentified amide N suggested by past literature. First, it is possible hydrolysis-resistant amide-containing molecules accumulate in HMW DON which cannot be recovered via molecular

level techniques. These could be AA and AS which are physically protected from acid hydrolysis or other unknown amide molecules. Previous work in sediments demonstrated that even after standard acid hydrolysis conditions (6N HCl), ^{15}N CP/MAS NMR detected mostly amide N, which was suggested to be physical protected proteinaceous material (Knicker & Hatcher, 1997, 2001). The presence of hydrolysis resistant amides in sediments could indicate they are also present in sinking POM and may be resistant to microbial degradation. Additionally, past work specific to the HMW DON pool suggested that a substantial portion of HMW amide N is hydrolysis resistant AA and AS, hypothesizing that 65% of surface HMW DON and 29% of deep HMW DON are represented by AA and AS (Aluwihare et al., 2005). However, this theory is not supported by the high recovery of AS from acid hydrolysis, which is generally between 72% and 82% (Kaiser & Benner, 2000). Additionally, large isotopic offsets between bulk HMW DON and AA in HMW DON suggest the "other" N in HMW DON is not AA or AS (Ianiri & McCarthy, In prep). Thus, we suggest that it is unlikely that hydrolysis-resistant AA and AS account for all of the unidentified amide N in HMW material. Still, it is possible other hydrolysis-resistant amide molecules contribute to the HMW DON pool.

An additional possibility is that HMW DON contains novel amide-N molecules besides AA and AS traditionally measured with molecular-level techniques. These molecules would not have to be hydrolysis-resistant (though they could be), but rather would fall outside the analytical window of traditional molecule-level analyses. Advanced solid-state ^{15}N NMR experiments on HMW DON derived from *Synechococcus* cultures support this theory, demonstrating a wide variety of newly identified amide-containing molecules, representing $\sim 40\%$ of *Synechococcus*-derived HMW DON (Cao et al., 2017). While the quantitative importance of these

molecules in the marine environment is unknown, the data presented here suggests it is possible these molecules are an overlooked source to the marine DON pool. Considering the importance of cyanobacterial primary production in oligotrophic gyres, in particular *Synechococcus*, it is possible similar amide-containing molecules or their degradation products could accumulate in marine HMW DON. Future work investigating these compounds in marine DON would be valuable to understand their potential importance.

Regardless of the molecular composition of the unidentified amide N, it likely represents the most refractory molecules in the HMW DON size fraction. Determining the composition and source of this subset of the HMW DON pool will be critical to understanding upper ocean marine N cycling. We suggest additional advanced solid-state NMR approaches, such as 2-D approaches or those exploring C-N bonds (Fang et al., 2011; Schmidt-Rohr & Mao, 2002), could aid in determining the chemical composition of these amide-containing molecules and identify target groupings for subsequent molecular-level analyses.

4.5.2 Heterocyclic N dominates the LMW SPE-DON pool

The results presented here using advanced solid-state NMR techniques both confirm and substantially expand initial ^{15}N CP/MAS NMR experiments which indicated LMW SPE-DON functional composition is dominated by heterocyclic N groups (Broek et al., Submitted). While CP/MAS indicated $\sim 80\%$ of LMW SPE-DON was heterocyclic N, the more quantitative ^{15}N multiCP/MAS reuslts indicate that all LMW SPE-DON is heterocyclic material (Fig. 4.3, Fig. 4.6). Combined with the decrease in LMW SPE-DON concentrations from the surface to deep ocean (0.6 µmol/L in the surface to 0.4 µmol/L at 2500 m, [Broek et al., 2017]), these new data indicate there must a direct surface source for heterocyclic

LMW DON.

The chemical shift of the major peak in both surface and deep LMW DON (127 ppm and 132 ppm, respectively) is ~ 10 ppm offset from the amide chemical shift of ~ 115 - 120 ppm seen in all past DON spectra. This clearly indicates that the major peak in LMW DON represents different N functionality than observed in past HMW DON work. Still, because solid-state ^{15}N NMR chemical shifts are dependent on sample matrix, and only one previous study has demonstrated clear heterocyclic functionality to DON (Broek et al., Submitted), there are not many past chemical shift references directly applicable to these results. Initial generalized solid-state ^{15}N NMR chemical shift measurements indicated overlapping amide and pyrrole/indole signal in this region (Knicker et al., 1996; Witanowski et al., 1993), though more recent work specific to marine DON hypothesized that the pyrrole/indole region begins at 130 ppm (Aluwihare et al., 2005). Similarly, the N atoms in chlorin measured in this work had a chemical shift of 133.6 ppm, close to the ~ 132 ppm chemical shift of the major peak in deep LMW SPE-DON. Still, it is possible the borderline chemical shift for the dominant peak we observe in LMW SPE-DON indicates there is an amide shoulder to this peak which is too small to be integrated above the noise. A small amide contribution would be consistent with past molecular-level data indicating that AA represent $\sim 3.3\%$ of surface and 2.0% of deep LMW SPE-DON investigated in this study (Ianiri et al., Submitted) while the contribution of AS is likely even lower (Kaiser & Benner, 2009). If this were the case, the slightly upfield (lower ppm) chemical shift of the main peak in surface LMW SPE-DON compared to deep LMW SPE-DON would be consistent with the slightly greater AA contribution to this sample. Still, overall, the fact that LMW SPE-DON is almost entirely composed of heterocyclic N composition supports a prior hypothesis that most of these compounds are inher-

ently stable (Broek et al., Submitted), potentially explaining the recalcitrance of most marine DON.

The three separate peaks in the LMW SPE-DON ^{15}N multiCP/MAS NMR spectra (Fig. 4.3, Fig. S4.1), which are also clearly defined in the denoised spectra (Fig. S4.3), demonstrate a range of heterocyclic N functionality to the LMW size fraction. The major peak in the deep LMW spectra (~ 132 ppm) is incredibly close to the chemical shift of chlorin measured in this study (~ 134 ppm) (Fig. 4.1), suggesting the majority of LMW heterocyclic N may be pigment molecules and/or their metabolites or degradation products. While the chemical shift of the major peak in surface LMW SPE-DON was slightly lower (~ 127 ppm), if this is due to a small amide shoulder as discussed above, the major heterocyclic signal in this sample could also be due to porphyrin-like molecules. In contrast, the peak at ~ 170 ppm in both surface and deep LMW SPE-DON is consistent with the chemical shift of one of the N atoms in the six membered ring of guanine (Fig. 4.1). This could suggest nucleic acids or other pyrimidine molecules contribute to this peak. Finally, the peak at 146 ppm observed in surface but not deep LMW SPE-DON is close to the peak at 142 ppm in surface HMW DON. The fact that this peak is only observed in the surface of either size fraction is consistent with a surface produced, semi-labile component of the DON pool. As discussed above, this chemical shift is closer to that of the protonated N atoms in guanine (Fig. 4.1) and other purine compounds such as allantoin (Fang et al., 2011). Still, most of the heterocyclic N signal of LMW SPE-DON ($\sim 90\%$) is represented by peaks which are not observed in the HMW DON spectra, suggesting that HMW and LMW SPE-DON are almost entirely compositionally distinct and may cycle independently, similar to hypotheses suggested for AA in both size fractions (Broek et al., 2019; Ianiri et al., Submitted). Further exploration of these novel N compounds via molecular

level or isotopic analyses will thus be critical in understanding long-lived, RDON in the ocean.

Finally, we note that the while the above interpretations are most applicable to the LMW SPE-DON fraction isolated via PPL extraction of the UF permeate, it is unclear if these results can be extended to the total LMW DON pool. To our knowledge, there was only one prior study to make ^{15}N measurements on any non-ultrafiltered DON fraction (apart from Broek et al. Submitted, which investigated the same samples analyzed in this study). Mao et al. (2012) collected a ^{15}N CP/MAS NMR spectrum of surface DOM collected via reverse osmosis-electrodialysis (RO/ED) from the Atlantic Coast. While this technique typically isolates between $\sim 60\%$ to 80% of the total DOM pool (Koprivnjak et al., 2009), because it does not rely principally on molecular size, it would be expected to represent a mixture of both HMW and LMW material, especially in the surface. The surface ^{15}N NMR spectrum in Mao et al. (2012) is very low resolution but appears to be mostly amide with some heterocyclic contribution. Unfortunately, neither DON recovery data nor peak integrations were reported. Considering that RO/ED DOM consists of both HMW and LMW molecules in unknown proportions, the use of non-quantitative NMR techniques, and the lack of reported numeric results, it is difficult to make any meaningful comparisons between this spectrum and our isolated size fraction.

Instead, nitrogen isotope data is likely the best current way to assess the relative representativeness of the isolated LMW fraction here compared to the entire LMW SPE-DON pool. Isotopic comparisons of LMW SPE-DOM investigated here with total LMW DOM suggest the directly recovered size fraction is similar to the total LMW DOM pool. For example, δ^{15}N values of total LMW SPE-DON measured by difference (LMW SPE-DON = Total DON - HMW DON [Broek,

2019; Knapp et al., 2012]) and LMW SPE-DON (Broek, 2019) are statistically indistinguishable. Additionally, the magnitude and depth trends of LMW SPE-DOM verses total LMW DOM $\Delta^{14}C$ values are similar, with an average offset of 45 $\pm$ 40 ‰ (Broek et al., 2020). Finally, comparisons of AA composition and degradation state from LMW SPE-DON and the UF permeate from which the SPE-DON samples were isolated are very similar (Ianiri et al., Submitted). Together, these data suggest the LMW material recovered by SPE of the UF permeate is likely generally representative of total LMW SPE-DON.

4.5.3 Potential sources of heterocyclic N in marine DON

There are multiple potential abiotic and biotic sources for heterocyclic N in marine DON. Abiotic sources include thermal heating and combustion, which produce pyrrole and indole-like structures (Knicker et al., 1996). The only known marine sources for these processes are hydrothermal vent systems or petroleum seem sites. However, the $\delta^{13}C$ and $\Delta^{14}C$ values of carbon derived from both petroleum seep sites (Pohlman et al., 2011) and off-axis hydrothermal vents (McCarthy et al., 2011) are much lower than the $\delta^{13}C$ and $\Delta^{14}C$ values of our LMW DOC (Broek et al., 2017). Still, considering the uncertainty regarding the fluxes and residence time of C from these sources, isotopic data alone cannot constrain the potential for these deep-sea sources. More importantly, however, both these sources would input DON to the deep ocean, which is inconsistent with our data which strongly supports a surface source (Section 4.5.2).

Alternatively, dissolved "black nitrogen," or N-containing condensed aromatic molecules produced by combustion such as wildfires, would also have ^{15}N NMR chemical shifts in the pyrrole/indole region (Knicker, 2010). While there is limited research specific to black nitrogen in the ocean, previous work on dissolved black

carbon (DBC) indicates the major source is riverine input (Jaffé et al., 2013), which would likely not result in the dramatic surface versus deep concentration gradient we observe in the open ocean. Additionally, concentrations of DBC measured in these same samples are nearly an order of magnitude lower than the amount of DBC we would expect if all heterocyclic N was black nitrogen (Coppola et al., In prep), meaning DBC cannot account for the majority of heterocyclic N we observe.

Alternatively, the Maillard reaction, in which amino acids and sugars react to form melanoidins, was suggested by Mao et al. (2012) as a potential source for heterocyclic N observed in RO/ED DOM. Molecular-level recovery of supposed end products of the Maillard reaction suggested this process may form heterocyclic N in sediments over thousands of years (Nguyen & Harvey, 2001) and during degradation of plant litter (Fogel & Tuross, 1999). However, there is no evidence to suggest these compounds accumulate in large enough quantities to represent the NMR signals we observe. Additionally, it is unlikely that these reactions could occur at the cold temperature of the ocean ($\sim$ 2 °C), as temperatures between 150 °C – 500 °C are usually necessary (Skog et al., 1998).

Instead, marine primary production is more consistent with the evidence indicating a surface source for both HMW and LMW heterocyclic N. Marine production results in a range of heterocyclic, pyrrolic N biomolecules such as porphyrins and chlorins (Higgins et al., 2011). Advanced ^{15}N NMR techniques of cyanobacterially-produced HMW DON indicated an entirely amide (though not proteinaceous) composition (Cao et al., 2017), but no quantitative techniques were applied nor was the LMW size fraction investigated. However, FTI-CR-MS of *Synechococcus*-produced SPE-DOM indicated a large diversity of LMW compounds containing two N atoms, many of which are heterocyclic, leading the au-

thors to suggest these compounds contribute to microbial resistant DOM (Zhao et al., 2017). Considering *Synechococcus* is the second most abundant primary producer in our sampling regions, it seems likely cyanobacterially produced compounds or their degradation products contribute to the heterocyclic N observed in our samples.

Future work making advanced ^{15}N NMR measurements of isotopically labeled cyanobacterial LMW DON and its degradation products could determine if these sources produce heterocyclic N compounds in sufficient quantities to represent the signal in our ^{15}N NMR spectra, and the role of microbial alteration in producing refractory heterocyclic DON molecules. Specifically, ^{15}N multiCP would be key to determining the quantitative contribution of these compounds, and 2D NMR methods could yield more information regarding the specific chemical composition. Considering the apparent dominance of heterocyclic N to the total DON pool, further experiments targeting the chemical composition of these compounds will be key to understanding the formation and persistence of marine RDON.

4.5.4   Reevaluating the composition of HMW and LMW DOC

Comparison of ^{13}C multiCP/MAS and CP/MAS DOC spectra

Aromatic and carboxyl functional groups are hypothesized to represent some of the most refractory DOC compounds (Hedges et al., 2000; Hertkorn et al., 2006), however, they are also the functional groups most underestimated by standard CP/MAS NMR (Smernik & Oades, 2000a, 2000b). Because almost all past ^{13}C NMR of marine DOC relied on standard CP/MAS NMR techniques, it has never been certain if substantial RDOC components were overlooked. For example, comparisons of fully quantitative DP/MAS spectra with CP/MAS spectra of soils indicated the aromatic contribution of many soils was substantially under-

estimated using CP/MAS techniques (Smernik & Oades, 2000b), suggesting the true aromatic C contribution to marine DOC may be unknown. A more significant aromatic contribution to refractory DOC could be reasonable considering the inherent stability of aromatic ring structures and the significant aromatic contribution to XAD-isolated DOC and DOC (Hedges et al., 1992; Kaiser et al., 2003; Sannigrahi et al., 2005). Unfortunately, the extremely long instrument time required for DP/MAS experiments is prohibitive for routine application to marine DOC. However, the far more rapid multiCP/MAS technique yields quantitation similar to DP/MAS for natural organic matter (Duan & Schmidt-Rohr, 2017; Johnson & Schmidt-Rohr, 2014), providing an opportunity for near-quantitative measurement of multiple samples of marine DOC.

The minor differences in ^{13}C multiCP/MAS with CP/MAS spectra of both standard compounds and natural samples (Fig. 4.2, Fig. 4.4, Fig. 4.5) seem to be difficult to reconcile with many past experiments demonstrating the limitations of standard CP/MAS. However, the explanation may lie in the specific parameters used for solid-state NMR acquisition. While there are no previous ^{13}C multiCP/MAS experiments of DOC, nor DP/MAS experiments of HMW or LMW SPE-DOC, to compare these results with, there are two past DP/MAS experiments of surface RO/ED DOC. These two studies yielded contrasting indications about the quantitative issues of standard CP/MAS. Consistent with the minor differences we report here, Mao et al. (2012) found carbonyl and aromatic signals using DP/MAS were within error of those observed using CP/MAS, with differences of $\sim 2\% \pm 2.2\%$. In contrast, Helms et al. (2015) estimated that CP/total sideband suppression (TOSS) NMR of RO/ED DOC underestimated carboxylic carbon by 15% and aromatic content by 11% compared to DP/MAS, much greater than the differences we observed here. A comparison with our own

results suggests this variation in signal intensity differences may be due to the contact time used in the CP experiments. Helms et al. (2015) used a very short contact time (0.5 ms), which is known to exacerbate the underestimation of non-protonated functional groups, while Mao et al. (2012) used a contact time of 1 ms. We hypothesize that the reason we observed little difference in the CP/MAS and multiCP/MAS experiments here is because we used an even longer contact of 5 ms. This suggests that for marine DOC specifically, contact times longer than 1 ms may provide nearly quantitative ^{13}C NMR spectra.

The similar aromatic signal intensity observed using more quantitative methods (DP/MAS, multiCP/MAS) versus CP/MAS may additionally be indicative of the most abundant aromatic compound classes in marine DOC. For example, CP/MAS of the highly condensed ring structures in charcoal only yielded $\sim 30\%$ of the total expected signal, while CP/MAS of lignin can be nearly quantitative (Smernik & Oades, 2000a). Similarly, chlorin e6, which has a chemical structure similar to hypothesized CRAM like molecules, did not demonstrate any significant differences in the aromatic region in the CP/MAS versus multiCP/MAS spectra with the acquisition parameters used in this study (Fig. 4.2, Table 4.4). This is likely due to greater proportion of C atoms in condensed aromatics which are non-protonated compared to less condensed structures which have many singly protonated C atoms. Still, past work using the multiCP/MAS technique on chars indicates this technique yields near quantitative spectra of highly condensed, non-protonated aromatics (Johnson & Schmidt-Rohr, 2014), meaning if these compounds were in our samples, we would expect to see a greater differences in the aromatic contribution as determined by the two methods. Instead, these limited differences in the aromatic signal suggest that condensed aromatics such as DBC represent only a minor portion of the total DOC pool. This is consistent with mea-

surements of DBC concentrations made on the same samples investigated here, which found condensed DBC concentrations were 2 to 15 times lower than the concentration of aromatic C in our samples (Coppola et al., In prep).

Notably, the differences we observe between the multiCP/MAS and CP/MAS spectra of HMW DOC are much smaller than inter-study variation observed when comparing past ^{13}C CP/MAS NMR experiments of HMW DOC from the Pacific Ocean (Fig. S4.5). Relative integration of the main four functional areas (carbonyl, aromatic, alkoxy, and alkyl) varies up to 20% in past work despite the same isolation method and sampling locations for these samples. It is possible some minor differences in functional composition may be expected due to differences in the recovered material between studies (Fig. S4.1). However, overall, these large differences in past CP/MAS results of similar DOM samples highlight the quantitative challenges of this technique and suggest inter-study differences could also be due to differences in CP/MAS parameters rather than true differences in sample composition. For example, the introduction of rampCP methods in the mid-90s were shown to greatly improve quantitation of natural abundance samples (Metz et al., 1994, 1996). Additionally, as noted above, differences in chosen contact time may significantly influence the relative signal of different functional groups. Still, the differences observed when comparing surface and deep samples within each study are consistent.

Taken as a whole, these new multiCP results directly address for the first time the quantitative accuracy of past solid-state NMR data for the DOC pool, particularly for the more refractory LMW material. Contrary to our expectations, we confirm that there is not a major additional aromatic or carbonyl C contribution to the LMW SPE-DOC pool which was overlooked by traditional CP/MAS NMR. Instead, our results confirm that CRAM dominate the refractory marine DOC

pool and that aromatics represent only a small proportion of marine DOC at any depth. This is in direct contrast with some of the earliest solid-state ^{13}C NMR spectra, which indicated that $\sim 20\%$ of deep HMW material alone might have been aromatic carbon. However, the correspondence between improved CP/MAS and multiCP methods seem strong evidence that the new estimates are correct. Considering that the few past DP/MAS NMR experiments of marine DOC were limited to RO/ED DOM in the surface ocean, the multiCP/MAS data presented here from two size fractions in both the surface and the deep ocean represents the most definitive picture to date of DOC functional composition.

How much CRAM is in HMW DOC?

Coupling new ^{13}C multiCP/MAS data with ^{15}N multiCP/MAS experiments on the same samples now allows for a reevaluation of carbon functional groups containing nitrogen. One example of this is the 191 ppm – 164 ppm region in ^{13}C NMR, which has overlapping signals from amide C, carboxyl C, and ester C. Our understanding of the refractory DOC pool is that it is dominated by CRAM, however the assumptions about quantitation have rested largely on carboxyl content. Past ^{13}C CP/MAS NMR studies of HMW DOM found carboxyl C to represent $\sim 8\% - 20\%$ of HMW DOM (Fig. S4.5) (Benner et al., 1992; Broek et al., 2020; Hertkorn et al., 2006; Sannigrahi et al., 2005). While these results all include a mixture of supposed refractory CRAM-like structures and other DOC, Hertkorn et al. (2006) computed a hypothesized ^{13}C NMR spectrum of solely CRAM from the difference of deep – surface HMW DOC, and suggested carboxyl content is $\sim 20\%$ of CRAM. One of the key observations the authors used to support the existence of CRAM was the proportion of the carboxyl C signal represented by non-amide C, calculated based on the assumption that all N is amide. However,

we now know non-amide N is also present in surface HMW DON. Thus, using the revised ^{15}N and ^{13}C multiCP data presented here, we revisit these calculations and their implication for CRAM abundance in HMW DOC.

Based on the C:N ratio (Table 4.1), percent of total N which is amide (Table 4.5), and percent of total C which is carboxyl (Table 4.6), we can calculate the percent of the total carboxyl signal represented by amide verses non-amide (carboxyl or ester) C to HMW DOC (Supplementary 4.7.1). We found that 74.3% of the surface HMW DOC and 42.3% of the deep HMW DOC carboxyl signal is due to amide C. This is about double what was observed by Hertkorn et al. (2006), which found only ∼ 35% of surface and ∼ 25% of the deep carboxyl signal was due to amide C. Instead, our results based on paired ^{15}N and ^{13}C solid-state multiCP/MAS analyses suggests a substantial portion of the carboxyl signal in HMW DOC is in fact not from refractory CRAM molecules but likely represents more labile amide compounds.

It is possible some of the differences between our data and past work (Hertkorn et al., 2006) are due to methodology. For example, Hertkorn et al. (2006) utilized liquid-state ^{13}C CP/MAS, which can only visualize functionalities which are soluble in the NMR solvent. Additionally, differences in MW cut off (2.5 kDa in this study, 1 kDa in past work), resulted in lower C recoveries and C/N ratios for our samples compared to those investigated by Hertkorn et al. (2006) (Table S4.1). This could influence the lower relative proportion of carboxyl C versus amide C we report here. Specifically, the lower C/N ratio in our data could indicate our MW cut off preferentially selects for amide-C (which is bonded to N) over carboxyl C. If true, this provides an informative constraint on the MW of amide versus carboxyl containing molecules, suggesting amide C is mostly in molecules greater than 2.5 kDa, while refractory CRAM molecules are smaller.

Overall, we suggest that to use carboxyl content as a proxy for the contribution of refractory CRAM molecules, amide C must first be accounted for. By coupling ^{15}N and ^{13}C multiCP/MAS NMR results, we found that CRAM cannot account for the majority of the carboxyl signal in our HMW DOC samples. Instead, most of the surface and almost half of the deep carboxyl signal was due to amide C compound classes. These results clearly indicate the presence of carboxyl functional composition should not be interpreted to represent refractory CRAM molecules. Instead, the high proportion of the carboxyl signal in deep HMW DOC which is due to amide C again highlights the importance of unknown amide material in this size fraction.

Are aromatic LMW DOC molecules dominated by N-containing heterocycles?

The low aromatic contribution to our DOC sample set was particularly surprising in the context of the dominant heterocyclic N contribution to LMW SPE-DON. If aromatic porphyrins and chlorins make up most LMW SPE-DON (as hypothesized in sections 4.5.2 and 4.5.3), then these results would suggest that most aromatic LMW SPE-DOC are N-containing molecules. Coupling ^{15}N and ^{13}C multiCP/MAS NMR with C/N ratios of LMW material, we can estimate the proportion of total LMW SPE-DOC associated with heterocyclic N. If all aromatic C does in fact contain heterocyclic N, we would expect the percent of LMW SPE-DOC associated with heterocyclic N to be roughly equivalent to the measured percent of total carbon signal from aromatic C.

A number of assumptions must be made for this calculation, which are detailed in Supplementary 4.7.1. Briefly, we assume some constraints on the structure of potential heterocyclic N molecules, which represent a minimum and maximum

C/N ratio for heterocyclic N molecules of 1.5 and 5.0, respectively. Based on this assumption, as well as the C/N ratio of DOM (Table 4.1), we calculated that between 5.4% to 18.1% of the carbon in surface LMW DOM and 5.2% to 17.5% of the carbon in deep LMW DOC is in heterocyclic N molecules (detailed calculations can be found in Supplementary 4.7.1). While these values clearly represent a wide range, the measured percent of the total carbon signal from aromatic molecules was near the minimum end of this range, 8.2% of surface and 6.8% of deep LMW DOC. This suggests that most or all aromatic LMW DOC could contain heterocyclic N. Additionally, if the percent of carbon in LMW DOC associated with heterocycles is greater than the aromatic C signal observed by ^{13}C multiCP/MAS NMR, this would imply that there is also a substantial portion of heterocyclic N which is not aromatic.

These results have significant implications for the cycling of refractory DOC and DON, suggesting that the limited aromatic C signal in LMW SPE-DOC may be almost all nitrogenous. Additionally, depending on the structure of LMW heterocycles, almost 20% of LMW SPE-DOC could be associated with heterocyclic N molecules, implying that heterocyclic N molecules may be a substantial contributor to both RDON and RDOC. Still, based on the data presented here, we cannot confirm if heterocyclic N is a substantial contributor to CRAM. CRAM is hypothesized to have a small aromatic contribution ($\sim$ 5% total C signal [Hertkorn et al., 2006]), meaning it is possible aromatic heterocyclic N contribute to CRAM. Alternatively, if there is a substantial contribution of non-aromatic heterocyclic N to LMW DON, these alicyclic molecules could also be part of the CRAM pool. Further advanced solid-state NMR characterization of marine DOM could distinguish between these two possibilities. ^{14}N-^{13}C SPIDER NMR, which selects for all ^{13}C atoms bonded to ^{14}N (Mao et al., 2007; Schmidt-Rohr & Mao, 2002), would

be particularly informative in determining the connectivity of RDOC and RDON functionality.

4.6   Conclusions

In this study, we applied for the first time ^{15}N and ^{13}C multiCP/MAS NMR to HMW and LMW SPE-DOM from the surface and deep North Pacific Subtropical Gyre. By coupling these novel NMR applications across the DOM age/reactivity spectrum, we provide the most comprehensive picture of DON and DOC functional composition to date. Using ^{15}N multiCP/MAS optimized for non-protonated N groups, we demonstrate HMW DON is in fact not composed of entirely amide N as previous ^{15}N NMR experiments had indicated. Instead, new ^{15}N multiCP/MAS data reveals a heterocyclic component which, based on DON concentration depth profiles, must be derived from surface production. However, in a surprise, no quantifiable heterocyclic N was observed to have accumulated in deep ocean HMW material, indicating HMW surface heterocyclic molecules are also likely to be relatively labile. At the same time, our results clearly demonstrate that discrepancies in amide-N contribution to HMW DON as measured by molecular-level analyses versus ^{15}N NMR cannot be explained by N functional groups overlooked by standard CP/MAS. Instead, mass balance suggests that the majority of HMW DON must be additional amide N which we suggest are novel structures not recovered by molecular-level techniques, together accounting for $\sim 73\%$ of surface and $\sim 90\%$ of deep HMW DON. The dominance of these unidentified amide structures to the HMW DON pool, particularly in the deep, indicate they are likely degradation resistant nitrogenous HMW molecules.

Our ^{15}N multiCP/MAS results of LMW SPE-DON, likely representing the vast majority of the ocean's DON pool, confirm and expand past work indicating

a dominant heterocyclic N contribution to LMW SPE-DON. Our more quantitative multiCP approach indicates that the entire ^{15}N NMR signal of LMW SPE-DON is composed of heterocyclic N functional groups. This material is represented by three main peaks in the indole/pyrrole region, further indicating a range of heterocyclic N functionality. These results suggest that the recalcitrance of most marine DON may be due almost entirely to inherently stable heterocyclic N structures.

Finally, pairing these ^{15}N measurements with ^{13}C multiCP/MAS NMR on the same samples suggested a range of new interpretations for N-containing molecules in HMW and LMW SPE-DOC. First, we found that $\sim 75\%$ of surface and almost 50% of the deep carboxyl signal in our HMW DOC samples is directly linked to amide C, not to refractory CRAM molecules. This supports past estimates that most HMW DOC is made up of relatively labile biomolecules such as polysaccharides and protein and suggests that CRAM contributions to surface HMW material is limited. Additionally, these data indicate that amide contribution to the carboxyl signal of DOC needs to be accounted for to accurately estimate the contribution of more refractory carboxyl material to any DOC fraction. Finally, contrary to our expectations, we find no evidence that aromatic contribution to HMW or LMW SPE-DOC has been substantially underestimated using previous standard CP/MAS techniques. Instead, our ^{13}C multiCP/MAS data suggest that aromaticity of HMW DOC was overestimated by early ^{13}C solid-state NMR experiments, likely due to less quantitative pulse sequences. This strongly supports previous ideas that CRAM material represents most refractory DOC, even in the deep ocean and ^{14}C-oldest LMW SPE-DOC pool. The low aromatic contributions to HMW and LMW SPE-DOC are particularly surprising in context of the substantial heterocyclic and aromatic N contributions confirmed by multiCP/MAS

and indicate that most aromatic DOC may be nitrogenous molecules.

Taken together, these data strongly support the hypothesis that DOC and DON recalcitrance is due to inherently stable chemical composition. However, at the same time, these results suggest that the molecular families leading to this recalcitrance may be different in the DOC and DON pools. Finally, these data identify multiple target N functionalities which should be targeted in future DOM research. Specifically, future work directly characterizing the chemical identity, sources, and fate of unknown HMW amide N and heterocyclic N in both size fractions will be critical to understanding RDON composition and cycling.

Main text figures and tables

A Guanine
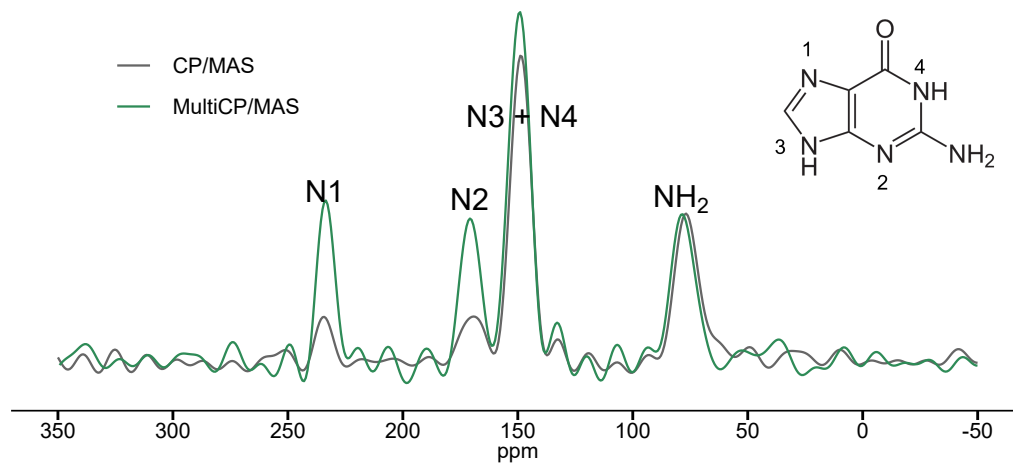
B Chlorin e6
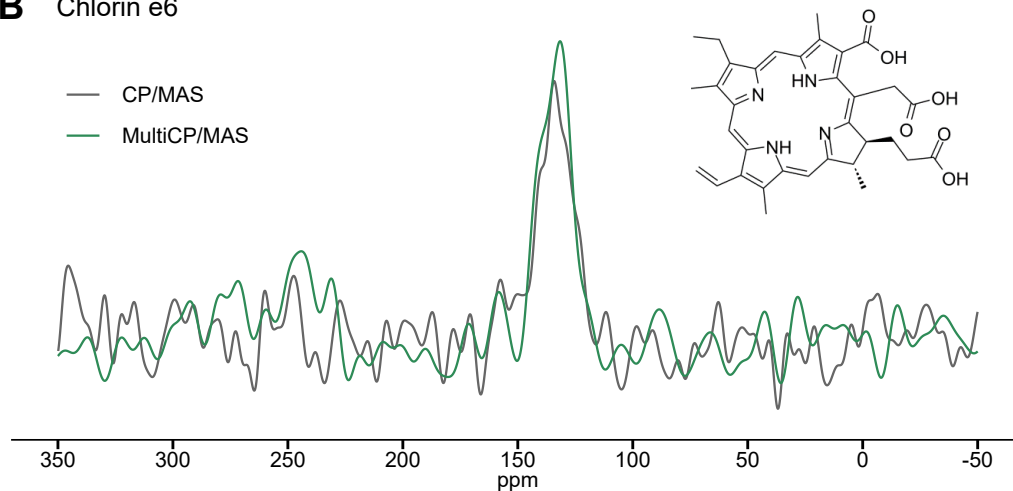
Figure 4.1: ^{15}N CP/MAS (grey) and ^{15}N multiCP/MAS (green) spectra of A) guanine and B) chlorin e6. For guanine, the signal intensity of non-protonated heterocyclic N1 and N2 are three to four times greater in the multiCP/MAS spectrum compared to the CP/MAS spectrum (integration data in Table 4.1). For chlorin e6, all N are represented by a single peak at 133.6 ppm. Spectra are normalized to the signal of the NH$_2$ group of guanine and to the single N peak for chlorin e6.
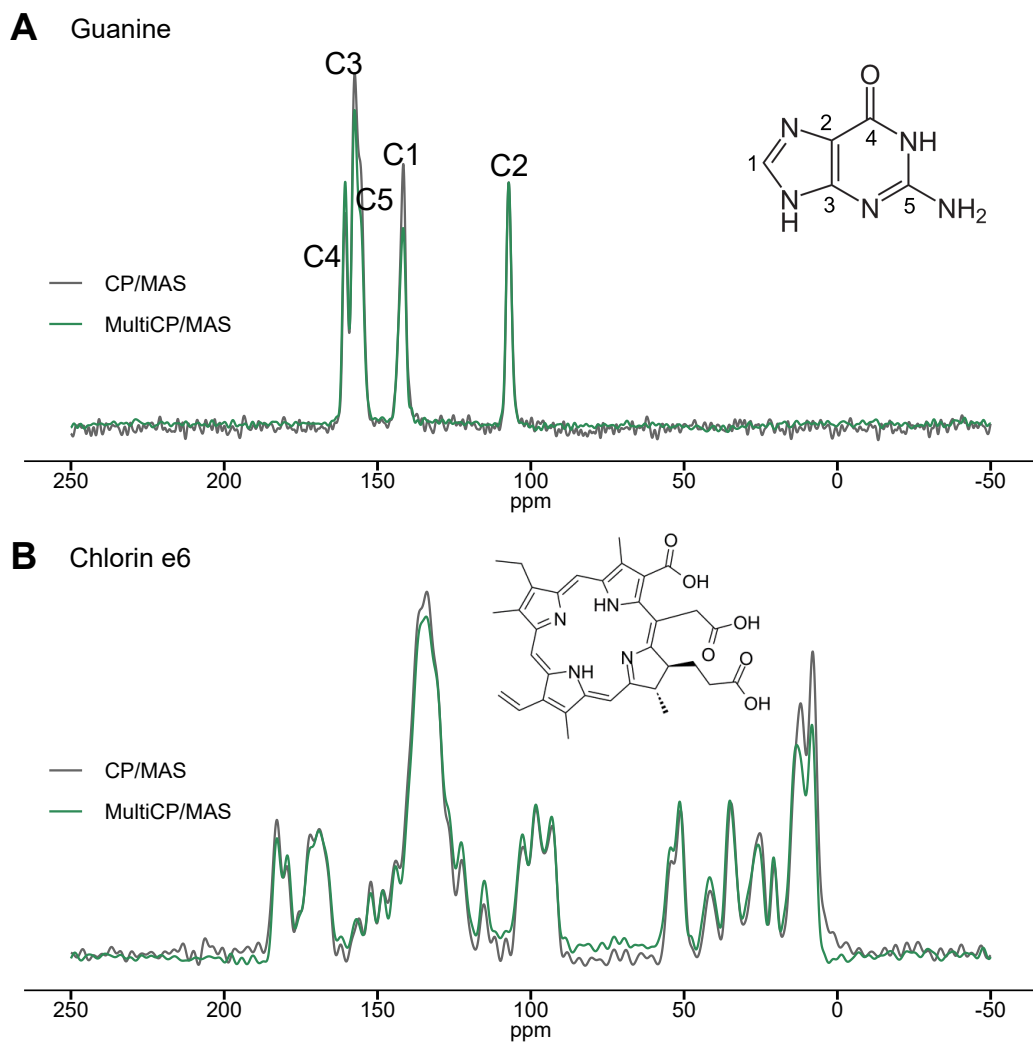
Figure 4.2: ^{13}C CP/MAS (grey) and ^{13}C multiCP/MAS (green) spectra of A) guanine and B) chlorin e6. ^{13}C CP/MAS and multiCP/MAS spectra of both compounds are similar (integration data in Table 4.2 and Table 4.3). Spectra are normalized to the signal intensity of C2 for guanine, and to the same total area for chlroin e6.
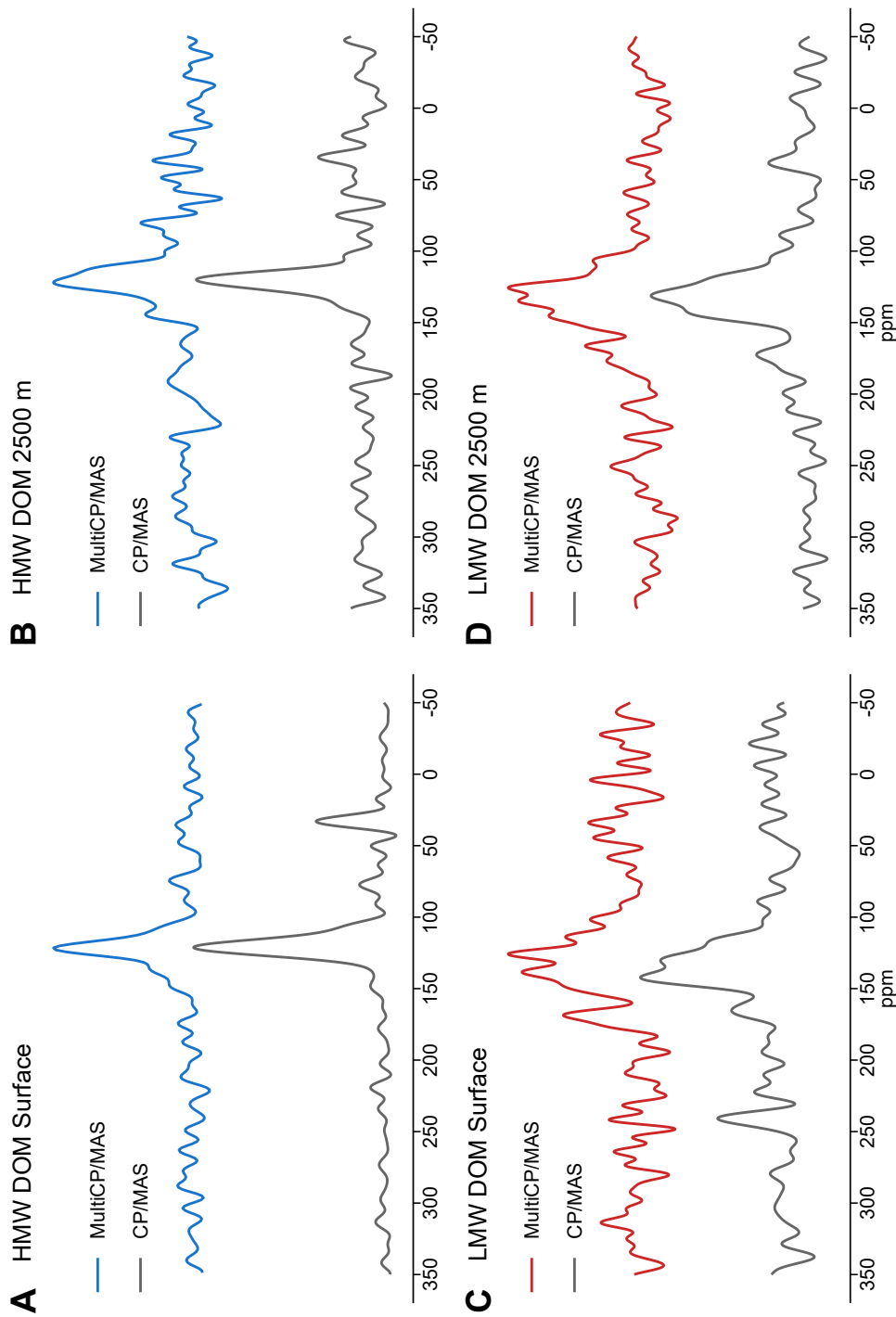
Figure 4.3: [15]N multiCP/MAS spectra (blue/red, top spectrum in each panel) versus ^{15}N CP/MAS spectra (grey, bottom spectrum in each panel) of A) surface HMW DON, B) HMW DON at 2500 m, C) surface LMW SPE-DON, and D) LMW SPE-DON at 2500 m. [15]N CP/MAS spectra are from Broek et al. (Submitted). Chemical shifts between 240 - 128 ppm were identified as heterocyclic N, 122 - 120 ppm as amide N, and 35 ppm as amine N. Individual peak integrations are visualized in Fig. S4.1 and chemical shifts and integration data are in Table 4.4.
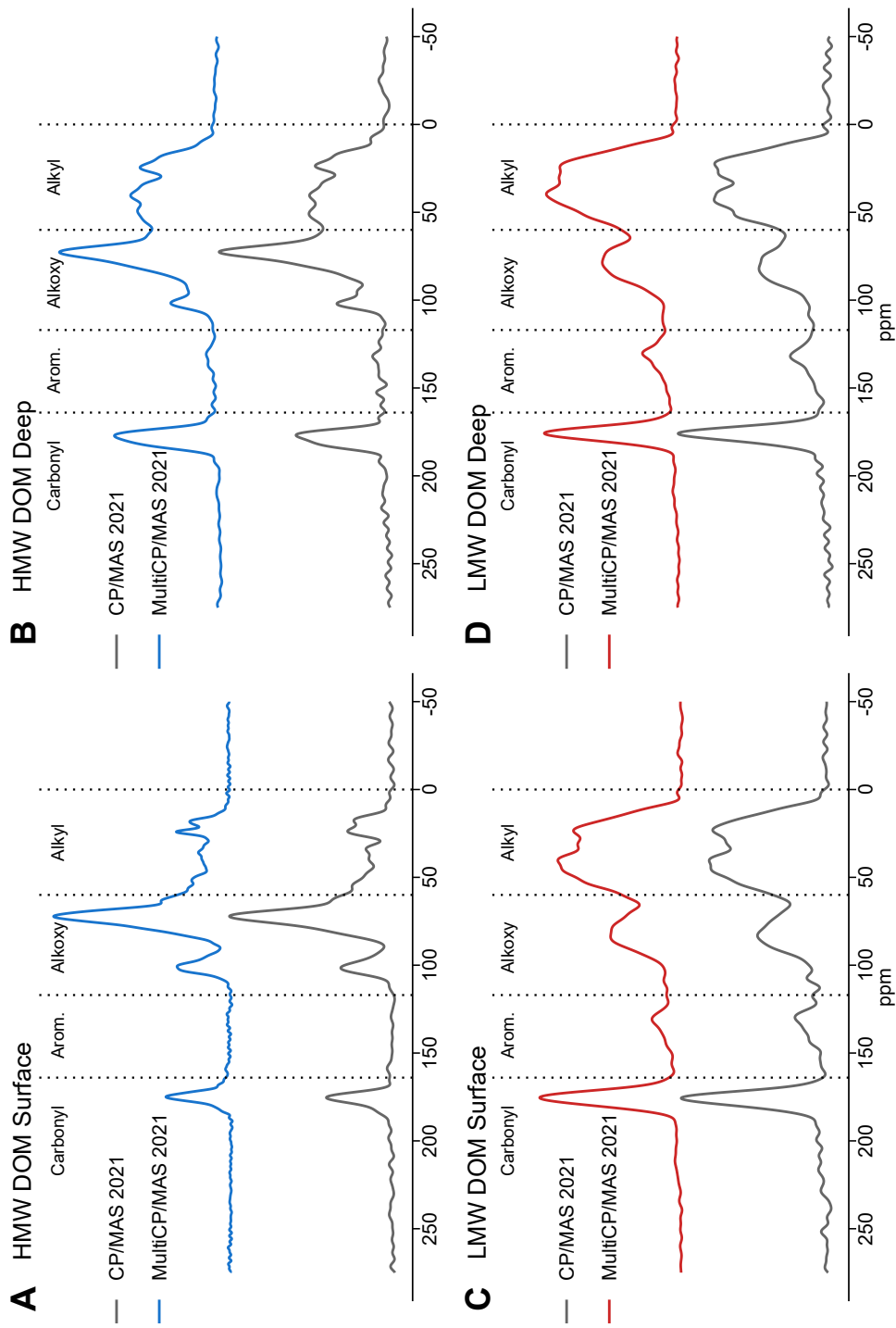
Figure 4.4: ^{13}C multiCP/MAS spectra (red/blue, top spectrum in each panel) versus ^{13}C CP/MAS spectra (grey, bottom spectrum in each panel) of A) surface HMW DOC, B) HMW DOC at 2500 m, C) surface LMW SPE-DOC, and D) LMW SPE-DOC at 2500 m. Individual peak integrations are in Fig. S4.4 in chemical shifts and integration data are in Table 4.5.
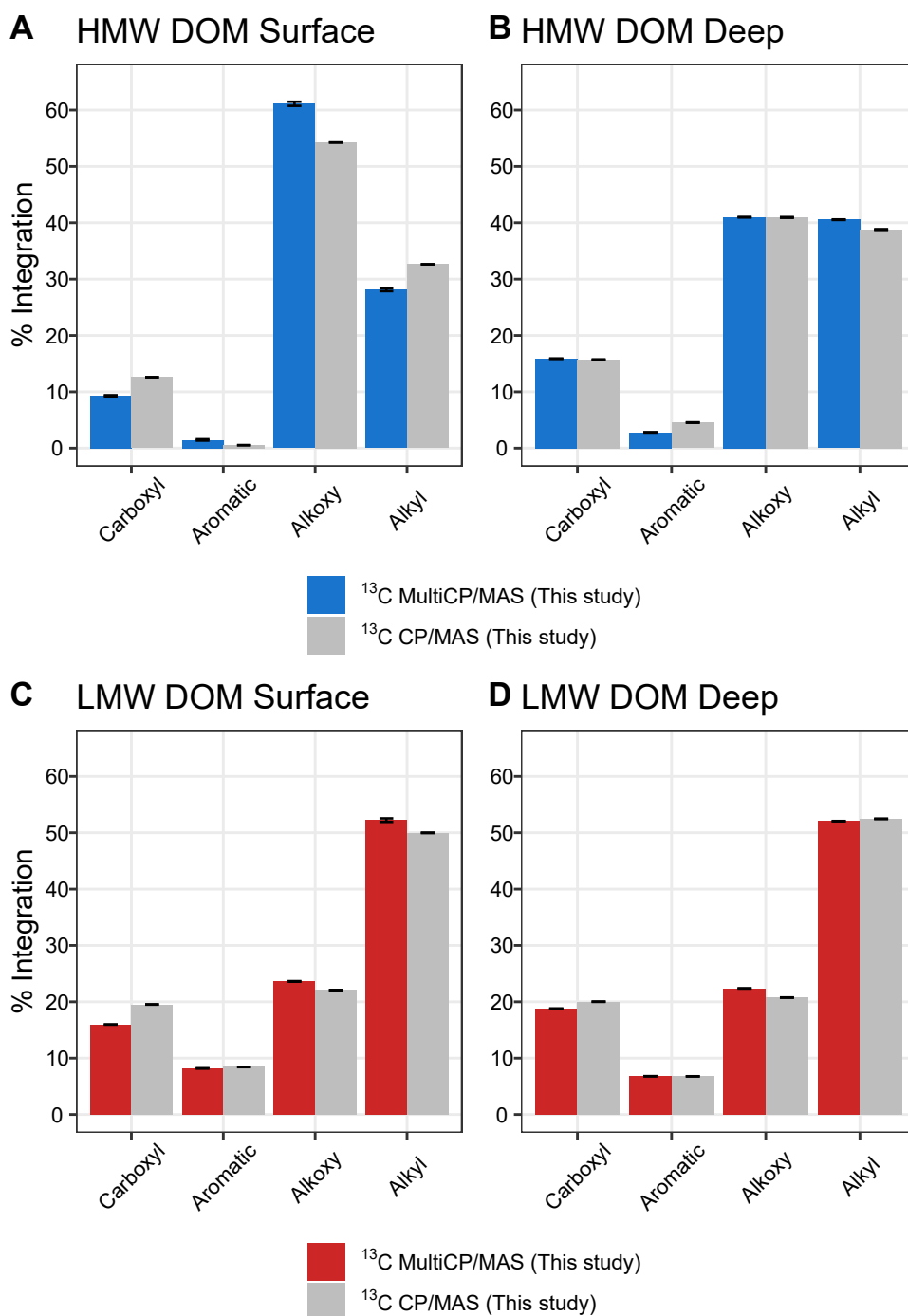
Figure 4.5: [13]C multiCP/MAS (blue, red) and CP/MAS (grey) integration results of A) surface HMW DOC, B) HMW DOC at 2500 m, C) surface LMW SPE-DOC, and D) LMW SPE-DOC at 2500 m. [13]C multiCP/MAS and CP/MAS spectra of all DOC samples are very similar. Error bars represent the propagated Monte Carlo error.
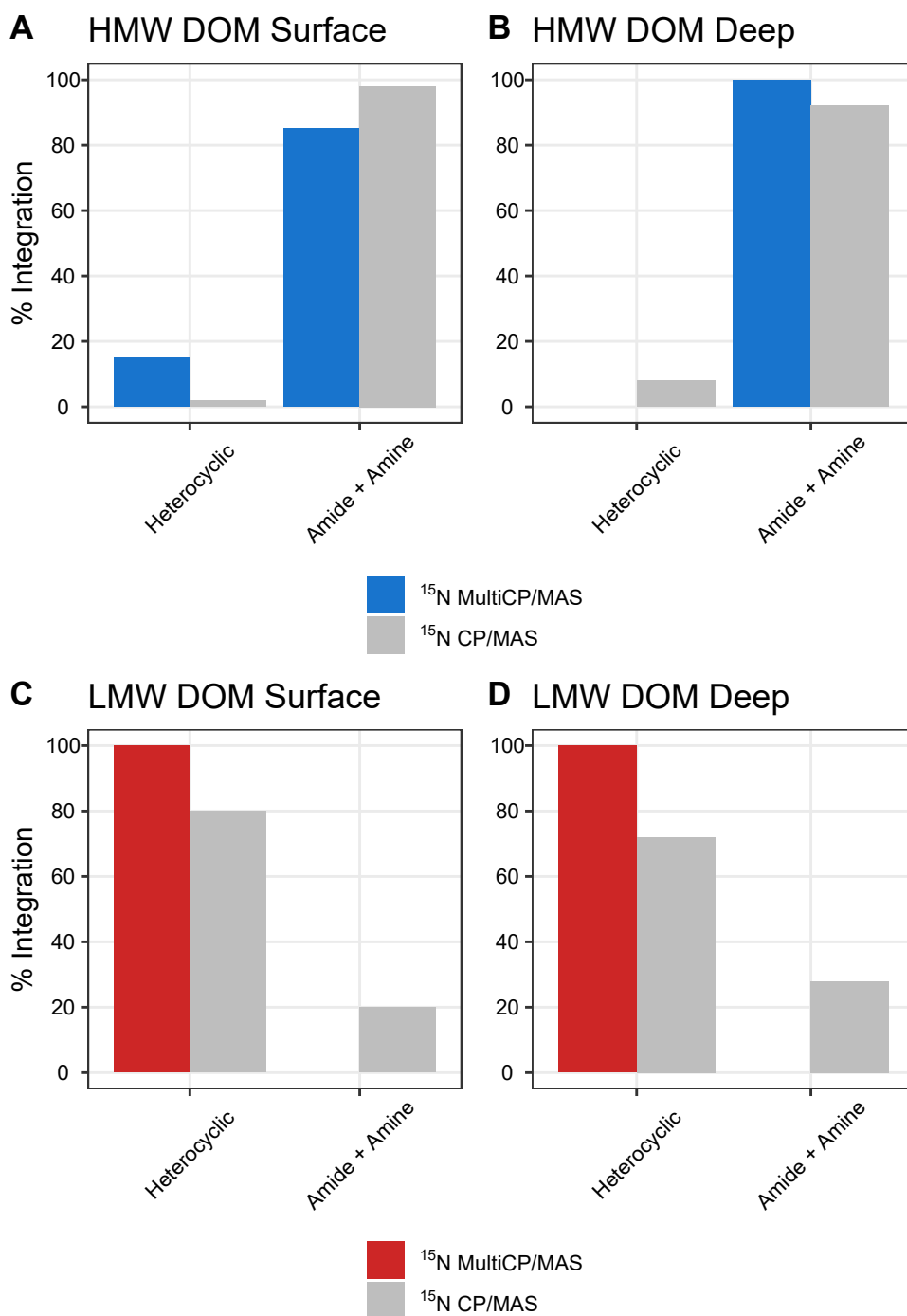
Figure 4.6: ^{15}N multiCP/MAS (blue, red) and CP/MAS (grey) integration results of A) surface HMW DON, B) HMW DON at 2500 m, C) surface LMW SPE-DON, and D) LMW SPE-DON at 2500 m. ^{15}N CP/MAS data is from Broek et al. (Submitted), which did not provide individual integration results for amide vs. amine contribution. In this study no quantifiable amine signal was observed in any sample.

HMW DOM Surface   HMW DOM Deep

Legend:
- Amino acid
- Amino sugar
- Heterocyclic
- Unidentified amide
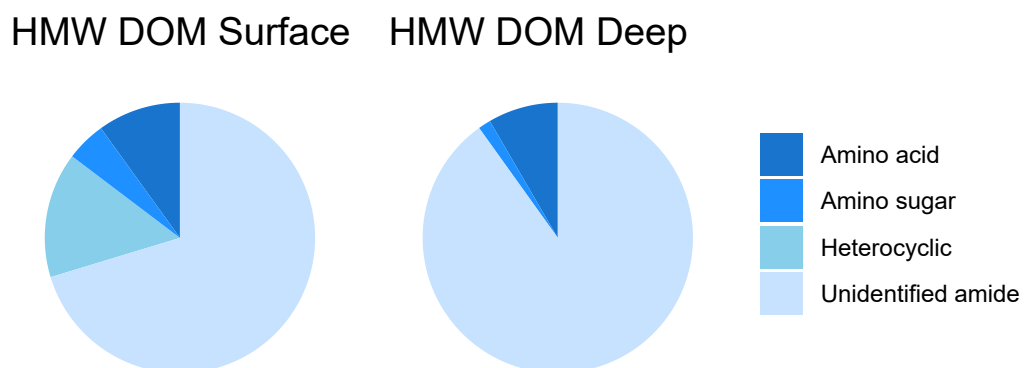
Figure 4.7: Combined data from molecular level analyses for AA (Ianiri et al., Submitted) and AS (Benner and Kaiser, 2003) and heterocyclic N recovery as determined by ^{15}N multiCP/MAS NMR (this study) indicates amide N which cannot be recovered by molecular level techniques ("Unidentified amide") makes up $\sim 70\%$ of surface and $\sim 90\%$ of deep HMW DON.

Table 4.1: Percent of total DOC and DON recovered, C/N ratios, and Δ^{14}C data for HMW and LMW SPE-DOM samples investigated in this study. Data is from Broek et al. (2017).

Size fraction	Depth (m)	DOC (%)	DON (%)	C/N	Δ^{14}C (‰)
HMW DOM	7.5	16.3	16.2	12.3	-50.0
HMW DOM	2500	7.8	13.9	13.1	-379.7
LMW SPE-DOM	7.5	20.4	9.0	27.6	-343.0
LMW SPE-DOM	2500	32.7	16.7	28.5	-577.6

Table 4.2: Relative percent integration data for ^{15}N CP/MAS and multiCP/MAS spectra of guanine. For each N atom, a relative intensity less than 20% indicates the signal is underestimated, while greater than 20% indicates the signal is overestimated. The signal intensity of all heterocyclic N groups are closer to the expected 20% using the multiCP method, representing a 3 - 4 times signal increase in non-protonated N groups. Integration error represents the propagated Monte Carlo error of 500 fit iterations. Nitrogen atoms are numbered in Fig. 4.1.

	Functional Group	Chemical shift (ppm)	Relative Intensity	±
CP/MAS				
N1	Heterocyclic N	234.64	5.81	0.48
N2	Heterocyclic N	170.5	7.24	0.61
N3 + N4	Heterocyclic NH	149.05	55.55	0.73
NH$_2$	NH$_2$	76.85	31.4	0.63
MultiCP/MAS				
N1	Heterocyclic N	233.83	16.53	1.12
N2	Heterocyclic N	171.22	15.23	2.18
N3 + N4	Heterocyclic NH	149.64	46.6	1.54
NH$_2$	NH$_2$	78.55	21.64	1.92
Difference (multiCP - CP)				
N1	Heterocyclic N		10.72	2.24
N2	Heterocyclic N		7.99	1.65
N3 + N4	Heterocyclic NH		-8.95	2.06
NH$_2$	NH$_2$		-9.76	0.63

Table 4.3: Relative percent integration data for ^{13}C CP/MAS and multiCP/MAS spectra of guanine. For each C atom, a relative intensity less than 20% indicates the signal is underestimated, while greater than 20% indicates the signal is overestimated. The signal intensity of most carbon atoms as determined by multiCP/MAS is closer to 20% than by CP/MAS. Integration error represents the propagated Monte Carlo error of 500 fit iterations. Carbon atoms are numbered in Fig. 4.2.

	Functional Group	Chemical shift (ppm)	Relative Intensity	±
CP/MAS				
C4	Carboxyl	160.6	12.36	0.07
C3	Aromatics	157.49	26.02	0.24
C5	Aromatics	155.45	19.50	0.23
C1	Aromatics	141.74	23.48	0.06
C2	OCO	107.14	18.65	0.06
MultiCP/MAS				
C4	Carboxyl	160.59	16.59	0.05
C3	Aromatics	157.52	24.30	0.16
C5	Aromatics	155.45	17.28	0.15
C1	Aromatics	141.84	21.50	0.04
C2	OCO	107.17	20.33	0.04
Difference (multiCP - CP)				
C4	Carboxyl		4.23	0.08
C3	Aromatics		-1.72	0.29
C5	Aromatics		-2.22	0.28
C1	Aromatics		-1.98	0.08
C2	OCO		1.68	0.07

Table 4.4: Relative percent integration data for ^{13}C CP/MAS and multiCP/MAS spectra of chlorin e6. The average propoagated error for each peak is 1.1% of the total integrated area.

Method	Ketone, Aldehyde, Quinone	COO, NC=O	Aromatic C-O	Aromatics	OCO	OC	OCH$_3$, NCH	CCH$_2$C, CCHC	CCH$_2$C, CCH$_3$
	220-191	191-164	164-150	150-117	117-94	94-60	60-45	45-30	30-0
^{13}C MultiCP/MAS	NA	12.76	1.83	36.83	10.36	7.29	5.51	9.58	15.83
^{13}C CP/MAS	NA	13.05	1.79	37.59	9.14	5.15	5.21	9.29	18.77
Difference (MultiCP - CP)	NA	-0.29	0.04	-0.76	1.22	2.14	0.30	0.29	-2.94

194

Table 4.5: Relative percent integration results for ^{15}N multiCP/MAS spectra of HMW and LMW SPE-DON samples. Integration error represents the propagated Monte Carlo error of 500 fit iterations.

Size Fraction	Depth (m)	Functional Group	Chemical shift (ppm)	Relative Intensity	±
HMW DON	7	Pyrrole/Indole	141.9	15.0	4.2
		Amide	122.1	85.0	4.2
HMW DON	2500	Amide	120.3	100.0	0.0
LMW SPE-DON	7	Pyrrole/Indole	169.8	12.5	0.6
		Pyrrole/Indole	143.5	12.4	3.9
		Pyrrole/Indole	126.7	75.1	4.3
LMW SPE-DON	2500	Pyrrole/Indole	172.8	11.3	0.2
		Pyrrole/Indole	132.1	88.7	0.2

Table 4.6: Relative percent integration data for ^{13}C CP/MAS and multiCP/MAS spectra of HMW and LMW SPE-DOC. The average propoagated error for each peak is 1.2% of the total integrated area. Difference data represent the signal intensity using the ^{13}C multiCP/MAS method - ^{13}C CP/MAS method.

Size Fraction	Depth (m)	Method	Ketone, Aldehyde, Quinone 220-191	COO, NC=O 191-164	Aromatic C-O 164-150	Aromatics 150-117	OCO 117-94	OC 94-60	OCH$_3$, NCH 60-45	CCH$_2$C,CCH$_2$C, CCHC CCH$_3$ 45-30	30-0
HMW	7	^{13}C MultiCP/MAS	0.0	10.0	0.5	0.0	10.5	52.7	8.7	8.8	8.9
HMW	2500	^{13}C MultiCP/MAS	0.2	15.7	0.5	2.3	4.2	36.8	15.8	10.3	14.5
LMW	7	^{13}C MultiCP/MAS	1.2	14.7	1.0	7.2	0.8	22.9	2.5	34.7	15.1
LMW	2500	^{13}C MultiCP/MAS	1.5	17.3	0.7	6.1	0.9	21.5	15.4	20.3	16.3
HMW	7	^{13}C CP/MAS	0.0	12.6	0.5	0.0	10.4	43.9	16.6	5.2	10.8
HMW	2500	^{13}C CP/MAS	2.7	13.0	1.4	3.1	4.9	36.0	15.9	9.0	13.9
LMW	7	^{13}C CP/MAS	2.1	17.4	0.6	7.9	1.0	21.1	13.2	18.4	18.4
LMW	2500	^{13}C CP/MAS	2.7	17.3	0.0	6.8	0.6	20.2	14.8	19.0	18.8
HMW	7	Difference	0.0	-2.6	-0.1	0.0	0.1	8.8	-7.9	3.6	-1.9
HMW	2500	Difference	-2.5	2.7	-1.0	-0.8	-0.7	0.8	-0.1	1.3	0.5
LMW	7	Difference	-0.9	-2.7	0.4	-0.7	-0.2	1.8	-10.8	16.3	-3.3
LMW	2500	Difference	-1.2	-0.1	0.7	-0.7	0.3	1.3	0.7	1.4	-2.5

4.7 Supplementary

4.7.1 Calculations for amide-C and aromatic-C

By coupling ^{15}N and ^{13}C multiCP/MAS NMR data, we can calculate an approximate proportion of the total carbon signal attributed to amide C and aromatic C. To determine the proportion of carbon signal in the 191 ppm – 164 ppm region attributed to amide C, we first calculate the percent of C atoms bonded to amid -N according to: $[1/(C/N_{DOM}) * (P_{Amide})] * 100$, where C/N_{DOM} is the C/N ratio of each sample and P_{Amide} is the proportion of total N signal from amide N. This calculation assumes only one carbonyl C atom is bonded to each amide N. Based on this calculation, 6.91% of surface HMW DOC and 7.63% of deep HMW DOC is bonded to amide N. Thus, the proportion of C signal in the 191 ppm – 164 ppm region which is attributed to non-amide C is the total signal in this region minus the percent of HMW DOC bonded to amide N. This yields 74.31% of HMW DOC (6.91 / 9.3 * 100) and 49.3% of deep HMW DOC (7.63 / 15.48) of the signal in the 191 ppm – 164 ppm region is from amide C.

Calculating the percent of C associated with heterocyclic N is somewhat more complicated. In contrast to an amide molecule, where there is a known ratio of carbonyl C: amide N of 1:1, heterocyclic N molecules can have a range of structures with various C/N ratios. Thus, for these calculations, we assume heterocyclic molecules are 5 or 6 membered rings, with 1 or 2 N atoms and 3 to 5 carbon atoms. Based on these constraints, the minimum and maximum C/N ratios for heterocyclic molecules are 1.5 and 5.0, respectively. Additionally, the number of carbon atoms bonded to each N in heterocyclic molecules ranges from 1.5 (if there are 2 N atoms and 3 C atoms) to 2.

With these constraints in mind, we calculate an approximate value for the percent of total C which is associated with heterocyclic N according to: $[C_N$

/ (C/N$_{DOM}$) * P$_{Het}$] * (C/N$_{Het}$/C$_N$), where C$_N$ is the minimum or maximum number of C atoms bonded to N, C/N$_{DOM}$ is the C/N ratio of each DOM sample, P$_{Het}$ is the percent of the total N signal which is heterocyclic, and C/N$_{Het}$ is the minimum or maximum C/N ratio of heterocyclic molecules. Based on this equation, we found that between 5.4% to 18.1% of the carbon in surface LMW DOM and 5.2% to 17.5% of the carbon in deep LMW DOC is associated with heterocyclic N.

Supplementary Figures and Tables
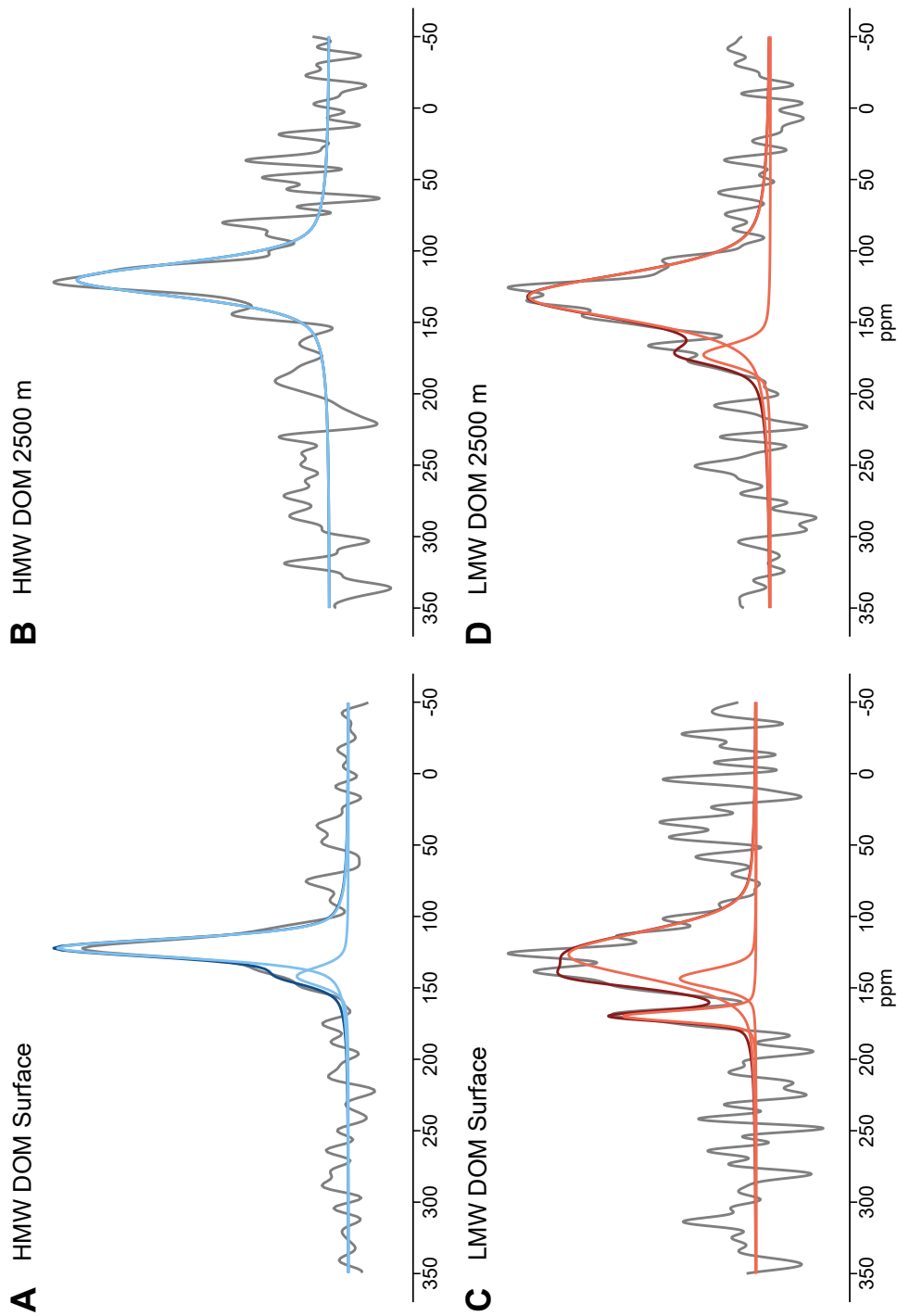
Figure S4.1: Modeled fit and peak integrations of ^{15}N multiCP/MAS spectra for A) surface HMW DON, B) HMW DON at 2500 m, C) surface LMW SPE-DON, and D) LMW SPE-DON at 2500 m. The original spectrum is in grey, the total modeled fit is in dark blue/red, and the individually integrated peak areas are in lighter blue/red. Individually integrated peak areas overlap with total modeled fit where darker line is not visible.
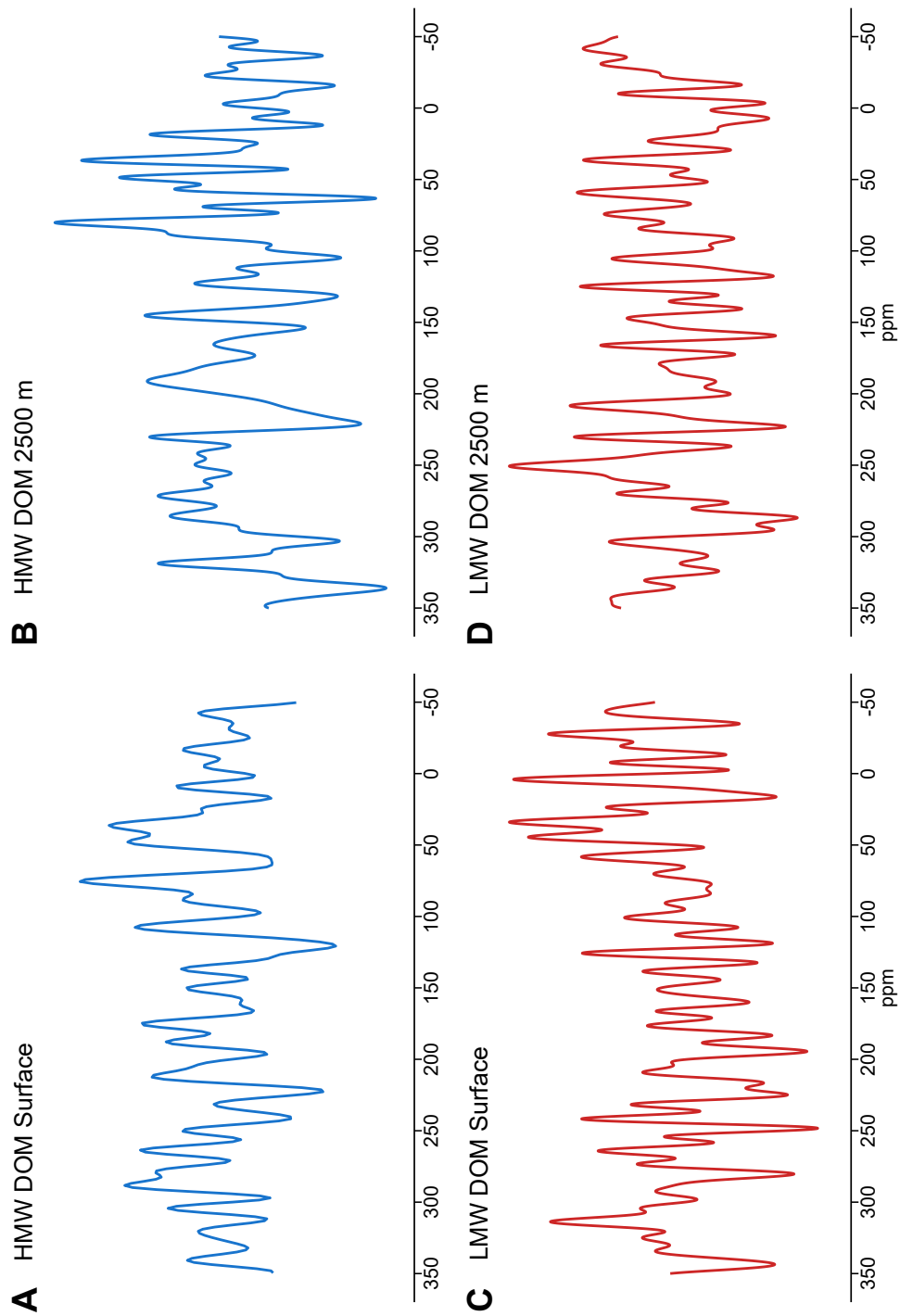
Figure S4.2: Residual (spectrum - model) plots of ^{15}N multiCP/MAS spectra for A) surface HMW DON, B) HMW DON at 2500 m, C) surface LMW SPE-DON, and D) LMW SPE-DON at 2500 m. Peaks were chosen so noise in residuals is similar to noise in each spectrum.
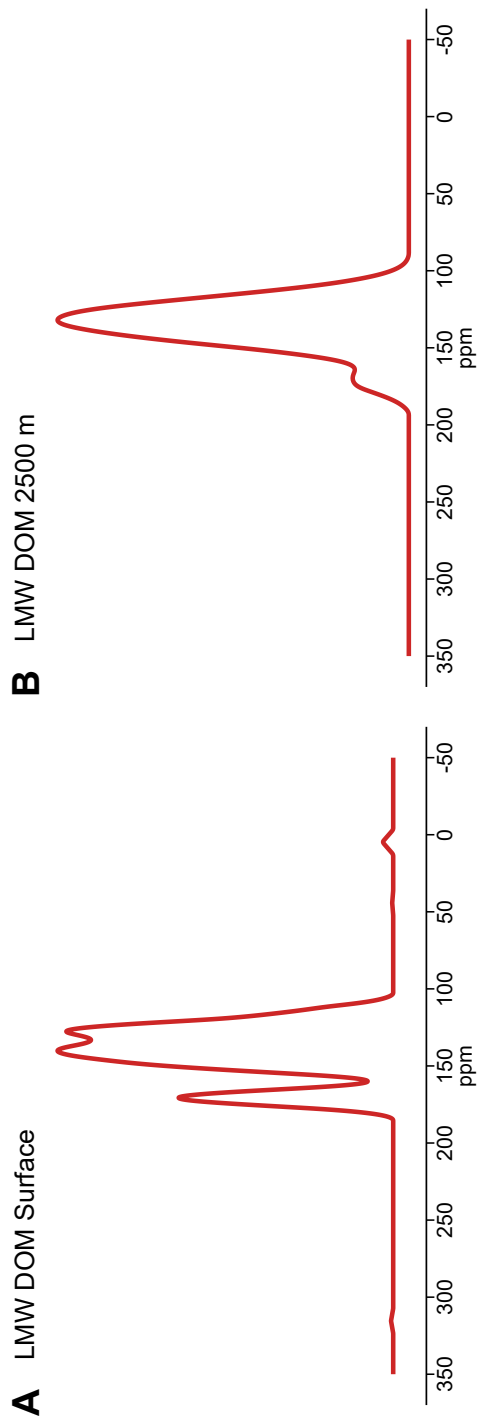
Figure S4.3: [15]N multiCP/MAS spectra of A) surface LMW SPE-DON, and B) LMW SPE-DON at 2500 m after a denoising procedure was applied. The denoising procedure verifies that in the surface LMW SPE-DON spectrum there are three separate peaks in the pyrrole/indole region, while the deep LMW SPE-DON spectrum has two separate peaks in the pyrrole/indole region.
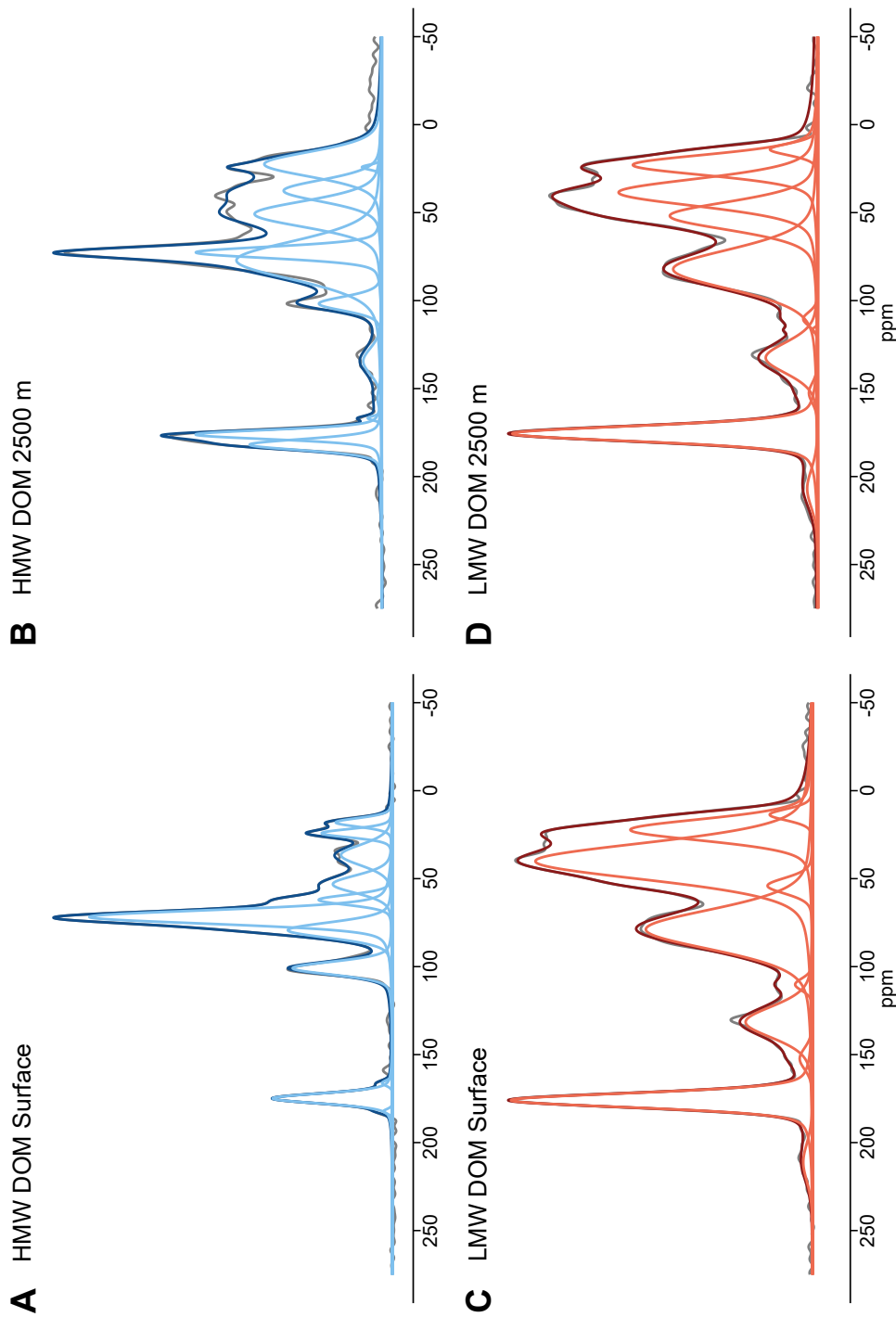
Figure S4.4: Modeled fit and peak integrations of ^{13}N multiCP/MAS spectra for A) surface HMW DON, B) HMW DON at 2500 m, C) surface LMW SPE-DON, and D) LMW SPE-DON at 2500 m. The original spectrum is in grey, the total modeled fit is in dark blue/red, and the individually integrated peak areas are in lighter blue/red. Individually integrated peak areas overlap with total modeled fit where darker line is not visible. The same peak integrations were used to fit ^{13}C multiCP and CP spectra.
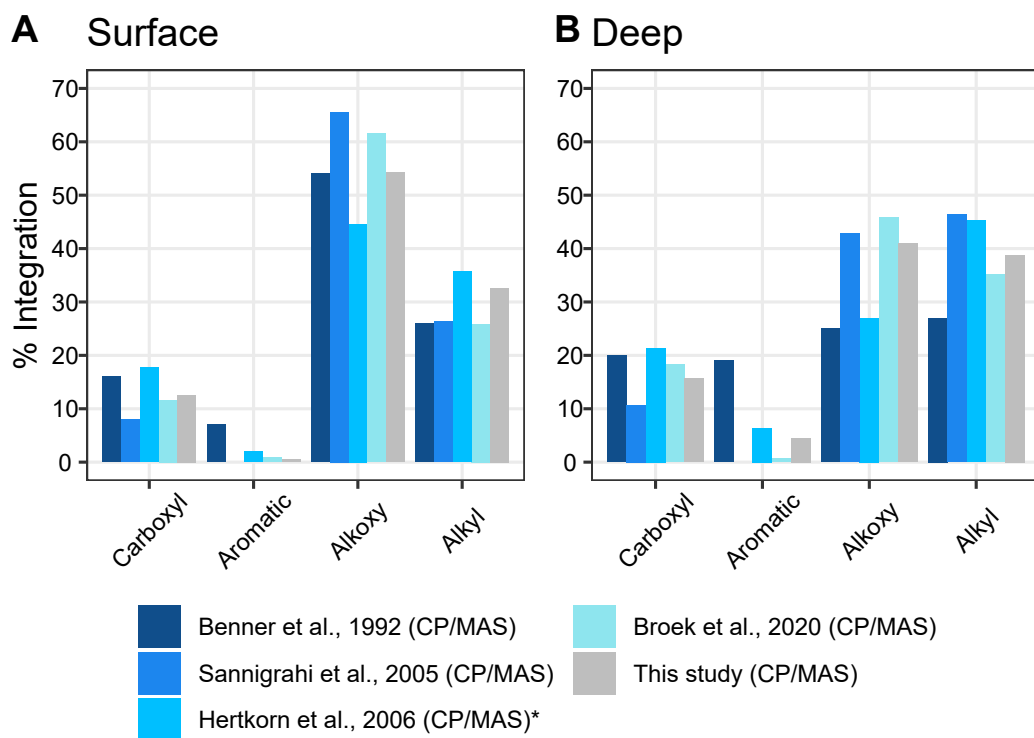
Figure S4.5: Integration data for the four major functional groups of HMW DOC from the central Pacific Ocean measured via ^{13}C CP/MAS in this study (grey) compared with previous work (shades of blue) A) in the surface ocean and B) deep ocean (2500 m to 4000 m). Despite similar sample types and sampling locations, substantial variation in CP/MAS results is observed. Data from Hertkorn et al. (2006) utilized liquid-state ^{13}C CP/MAS NMR.

Table S4.1: Comparison of NMR ^{13}C methods, DOC recovery, and C/N ratios of samples investigated in this study and past work referenced in the figures and text. Samples used in this study were the same samples investigated in Broek et al. (2020). "% Recovery" is recovery from total DOC, "LS" stands for "liquid-state", while "SS" stands for "solid-state".

Size Fraction	Location	Depth (m)	Reference	NMR Method	% Recovery	C/N Ratio
HMW DOM	Pacific Ocean	2	Hertkorn et al., 2006	LS ^{13}C CP/MAS	26	16.1
HMW DOM	Pacific Ocean	4000	Hertkorn et al., 2006	LS ^{13}C CP/MAS	20	18.4
RO/ED DOM	Gulf Stream	20	Mao et al., 2012	SS ^{13}C DP/MAS	68	17.7
RO/ED DOM	Gulf Stream	20	Mao et al., 2012	SS ^{13}C CP/MAS	68	17.7
RO/ED DOM	Atlantic Ocean	2	Helms et al., 2015	SS ^{13}C CP/TOSS	68	NR
SPE-DOM	Atlantic Ocean	5	Hertkorn et al., 2013	LS ^{13}C CP/MAS	40	25
SPE-DOM	Atlantic Ocean	5446	Hertkorn et al., 2013	LS ^{13}C CP/MAS	40	27
HMW DOM	Pacific Ocean	10	Benner et al., 1992	SS ^{13}C CP/MAS	33	15.3
HMW DOM	Pacific Ocean	4000	Benner et al., 1992	SS ^{13}C CP/MAS	22	19.6
HMW DOM	Pacific Ocean	20	Sannigrahi et al., 2005	SS ^{13}C CP/MAS	NR	15.9
HMW DOM	Pacific Ocean	4000	Sannigrahi et al., 2005	SS ^{13}C CP/MAS	NR	16.3
RO/ED DOM	Atlantic Ocean	20	Koprivnjak et al., 2009	SS ^{13}C CP/MAS	76	15.6
HMW DOM	Pacific Ocean	7	This study	SS ^{13}C CP/MAS, SS ^{13}C multiCP/MAS	16.3	12.3
HMW DOM	Pacific Ocean	2500	This study	SS ^{13}C CP/MAS, SS ^{13}C multiCP/MAS	6.9	12.5
LMW SPE-DOM	Pacific Ocean	7	This study	SS ^{13}C CP/MAS, SS ^{13}C multiCP/MAS	20.4	27.6
LMW SPE-DOM	Pacific Ocean	2500	This study	SS ^{13}C CP/MAS, SS ^{13}C multiCP/MAS	32.7	28.5

Chapter 5

Conclusion

Throughout most of the world's oceans, the bioavailability of marine DON acts to limit primary production and is a key control on marine biogeochemical cycles and carbon sequestration. Understanding the processes that control marine DON cycling and availability, and by extension long-term carbon storage in the ocean, are thus of vital importance. The three chapters presented here address the complex mystery of marine DON recalcitrance by applying cutting-edge, nitrogen specific techniques to a unique DON sample set allowing targeted investigation and comparison of semi-labile and refractory DON. Collectively, these data suggest a new understanding for marine DON source, cycling, and chemical composition.

In chapter two, I paired molecular-level proxies for bacterial DON with a suite of AA-based degradation proxies and radiocarbon age. This dataset represents one of the most comprehensive D-AA investigations of marine DON to date and suggested a new understanding for HMW and LMW SPE-DON cycling and degradation processes. These results also provided evidence that AA-containing molecules in LMW SPE-DON represent diverse and refractory N compound classes, despite previous assumptions that AA are generally labile proteinaceous molecules.

In chapter three, I utilized a novel tool, CSI-AA, to probe specific formation and degradation mechanisms of AA in HMW DON and LMW SPE-DON. This work represents the first CSI-AA analysis of LMW DON, as well as DON from the Atlantic Ocean. I proposed specific mechanisms for both production and degradation of AA-containing molecules in HMW and LMW DON, which together

address the offsets in bulk HMW and LMW DON δ^{15}N values observed in previous work as well the elevated bulk δ^{15}N values of HMW DON compared to PON and DIN. Together with the results of chapter one, these data suggest a novel theory for the bacterial role in formation of AA-containing molecules in both the HMW and LMW SPE-DON size fractions suggest a new interpretation of the formation of AA in RDON.

Finally, in chapter four, I optimized advanced solid-state NMR techniques, ^{15}N and ^{13}C multiCP/MAS, for investigation of refractory DOC and DON functional composition. For the HMW DON pool, ^{15}N multiCP/MAS NMR addresses the long-standing discrepancy between low molecular-level AA and AS recoveries despite ^{15}N CP/MAS NMR indicating an entirely amide N signal. We find that while there is a heterocyclic N component to the HMW DON pool overlooked by traditional NMR methods, HMW DON is dominated by amide N besides AA and AS which is not recovered by molecular-level techniques and likely represents the most refractory HMW nitrogenous molecules. In contrast, LMW DON is dominated by a range of heterocyclic N functionalities, representing a diverse group of RDON structures. Together, these data provide important compound classes to target in future molecular-level research. By combing these results with ^{13}C multiCP/MAS of DOC, these data suggest that while molecular composition is a key control on DOM recalcitrance, different biomolecular groupings are responsible for RDOC and RDON.

Taken together, this dissertation provides important new constraints on the source, cycling, and composition of marine DON. A novel aspect of each chapter is the targeted investigation of LMW material, which has historically been impossible to investigate directly. Additionally, the analyses applied here, especially in chapters three and four, represent cutting-edge techniques which have rarely

or never been applied to the marine DON pool. Collectively, this work suggests novel production mechanisms and chemical composition of both HMW and LMW DON and imply a paradigm shift in our understanding of marine DON cycling and recalcitrance.

Appendix

δ^{13}C compound specific isotope analysis of amino acids

A.1   Introduction

Stable carbon isotope analysis (δ^{13}C) represents one of the pioneering measurements of marine dissolved organic matter (DOM) (Williams & Druffel, 1987; Williams & Gordon, 1970). However, compared to traditional bulk δ^{13}C measurements (δ^{13}C$_{Bulk}$), which is an average value of all the carbon in a sample, δ^{13}C-amino acid (δ^{13}C-AA) values and patterns provide more specific information regarding metabolic source and transformation processes of organic material. Additionally, because all AA contain nitrogen, δ^{13}C-AA values represent an N-specific tool invaluable for the study of marine dissolved organic nitrogen (DON).

δ^{13}C-AA patterns are directly linked to metabolic origin (Scott et al., 2006), meaning diagnostic fractionation of specific AA can distinguish between autotrophic production and heterotrophic bacterial resynthesis (Keil & Fogel, 2001; Macko et al., 1987). δ^{13}C values of essential amino acids (EAA), or those which cannot be synthesized by heterotrophs *de novo*, are particularly useful tracers of metabolic source. The δ^{13}C signature of these AA acts as a "fingerprint" and can distinguish between plants, fungi, heterotrophic bacteria, and autotrophic bacteria (Larsen et al., 2009, 2013; McMahon et al., 2016). Multivariate analysis of δ^{13}C-AA values can quantitatively determine the relative source contribution of different organisms to a complex mixture (McMahon et al., 2015). Finally, δ^{13}C$_{EAA}$ also provides a proxy for "baseline" δ^{13}C of primary production (Schiff et al., 2014; Shen et al., 2021; Vokhshoori et al., 2014). Thus, while many of these tools were developed

for ecological purposes, δ^{13}C-AA analysis has potential to yield novel information regarding the source and processing history of marine DON.

Here, I present the first high-resolution δ^{13}C-AA data of marine DOM. I interpret δ^{13}C-AA patterns of HMW and LMW SPE-DOM from the Atlantic and Pacific Oceans in the context of previously published δ^{13}C$_{\text{Bulk}}$ data of total, HMW, and LMW SPE-DOM, as well as D/L-AA and δ^{15}N-AA data presented in chapters two and three of this thesis. By combining the information potential of all these tools, I aim to investigate metabolic source to open ocean HMW and LMW SPE-DOM. Additionally, I evaluate the application of the ecological tools described above for studying the more complex, millennial-scale cycling of marine DOM.

The results presented below are consistent with D/L-AA and δ^{15}N-AA data suggesting that HMW and LMW SPE-DON have distinct molecular sources. However, a comparison of δ^{13}C-AA and δ^{15}N-AA patterns also suggests a decoupling of carbon and nitrogen cycling within HMW and LMW SPE-DOM. Finally, these results suggest promise for multivariate analyses tool for predicting metabolic source to marine DOM. However, an expanded endmember dataset may be required for quantitative results can be acquired.

A.2   Methods

Materials and methods are outline in Ianiri et al., In prep (Chapter 3). Exceptions are outlined below.

A.2.1   Total DOC δ^{13}C

Stable carbon isotope ratios (δ^{13}C) of total DOC were measured at the University of California, Santa Cruz, Stable Isotope Laboratory (SIL) via elemental analyzer isotope ratio mass spectrometry using a Carlo Erba CHNS-O EA1108-

elemental analyzer interfaced with a ConFlo III device and a Thermo Finnigan Delta Plus XP isotope ratio mass spectrometer (Thermo Fisher Scientific). Standard UCSC-SIL protocols were followed for standard analysis and isotope ratio corrections. Analytical uncertainty of triplicate measurements ranged from $\pm$ 0.5 to 0.1‰.

A.2.2 Compound specific isotope analyses of δ^{13}C-AA

All isotopic analyses were completed at the UCSC Stable Isotope Laboratory. A Thermo Trace Ultra gas chromatograph coupled with a Finnigan MAT Delta-Plus XL IRMS at UCSC SIL was used for GC-IRMS analysis. AA were separated on an Agilent DB-5 column (50m x 0.32 mm, 0.52 µm film thickness) for δ^{13}C-AA analysis. Samples were injected in triplicates. A total of twelve AA were measured, including Ala, Gly, Thr, Ser, Val, Leu, Ile, Pro, Asx, Glx, Phe, and Lys (Fig. A.1). These AA were further assigned to two groups based on established classifications: essential amino acids (EAA) are defined as Phe, Thr, Ile, Leu, Val, and Lys, and nonessential amino acids (NEAA) as Asx, Glx, Pro, Ala, Ser, and Gly.

A.2.3 Statistical analyses

All data analysis was done in Microsoft Excel or R Studio version 4.0.5 (R Core Team, 2021). Data was tested for normality visually via QQ plots and using the Shapiro Wilk Test. Parametric statistical tests with a 95% confidence interval were used unless otherwise noted. For most statistical tests, samples from both ocean basins were grouped by size fraction for better statistical power due to the small sample size of this dataset.

Principal component analysis (PCA) and linear discriminant analysis (LDA)

was used to compare HMW and LMW SPE-DOM δ^{13}C-AA values with previously published source endmembers. Endmembers included cyanobacteria, heterotrophic bacteria, and eukaryotic microalgae selected from published training datasets and a previous graduate student thesis to only include environmentally relevant organisms to the central Pacific and Atlantic Oceans (Larsen et al., 2013, 2009, Lehman 2009). PCA of normalized δ^{13}C-AA values was tested using all AA and only EAA to determine which yielded the best separation of endmembers and DOM. δ^{13}C-AA values were normalized to the average of all AA included in that analysis (EAA or all AA). Only AA which were measured for every sample and endmember were included.

Using the endmember dataset and EAA optimized via PCA, LDA was performed on non-normalized data according to the literature (Larsen et al., 2013, 2009). A training dataset of 38 samples was used to create the linear discriminant model. The model was then used to predict the metabolic source of 14 known endmembers and all HMW and LMW SPE-DOM samples.

A.3 Results

A.3.1 δ^{13}C$_{\mathbf{Bulk}}$, δ^{13}C$_{\mathbf{THAA}}$, δ^{13}C$_{\mathbf{EAA}}$, and δ^{13}C$_{\mathbf{NEAA}}$ of HMW DOM and LMW SPE-DOM

δ^{13}C$_{\mathrm{Bulk}}$ values of total, HMW and LMW SPE-DOM were reported previously (Broek et al., 2020) and the relevant data is summarized here (Table A.1). δ^{13}C$_{\mathrm{Bulk}}$ of total DOM ranged from -22.7‰ $\pm$ 0.2‰ to -20.8‰ $\pm$ 0.2‰, increasing between the surface to 850 m then decreasing to 2500 m (Fig. A.2). δ^{13}C$_{\mathrm{Bulk}}$ of HMW and LMW SPE-DOM ranged from -23.3‰ $\pm$ 0.02‰ to -21.1‰ $\pm$ 0.1‰ and increased modestly between the surface and 2500 m (Fig. A.2). δ^{13}C$_{\mathrm{Bulk}}$ values of LMW SPE-DOM were significantly depleted by an average of 1.0‰ compared to HMW

DOM (Welch's two-sample t-test, p < 0.001). These values are within the range expected for marine DOM (Bauer et al., 1992; Benner et al., 1997; Williams & Druffel, 1987).

$\delta^{13}C_{THAA}$ values of HMW and LMW SPE-DON were generally greater than $\delta^{13}C_{Bulk}$ values and showed more variation, ranging from -23.9‰ ± 0.3‰ to -15.7‰ ± 0.2‰ (Fig. A.3, Table A.2). On average, $\delta^{13}C_{THAA}$ was significantly greater in LMW SPE-DON than HMW DON by 2.2‰ (Welch's two-sample t-test, p = 0.019) (Table 2). Depth trends of $\delta^{13}C_{THAA}$ values varied between ocean basins and size fractions. In HMW DON, $\delta^{13}C_{THAA}$ values were lower at 400 m compared to the surface at both BATS and HOT. HMW DON at HOT demonstrated particularly depleted $\delta^{13}C_{THAA}$ values at 2500 m compared to all other depths. LMW SPE-DON $\delta^{13}C_{THAA}$ values demonstrated opposite depth structure, with greatest $\delta^{13}C_{THAA}$ values at 400 m, decreasing to 850 m, then increasing again at 2500 m.

For both HMW and LMW SPE-DON, the average $\delta^{13}C$-AA of EAA ($\delta^{13}C_{EAA}$) was significantly depleted compared to the average $\delta^{13}C$-AA of NEAA ($\delta^{13}C_{NEAA}$) (Welch's two-sample t-test, p < 0.001) (Table A.2). $\delta^{13}C_{EAA}$ and $\delta^{13}C_{NEAA}$ were both significantly correlated with $\delta^{13}C_{THAA}$ in HMW DON, LMW SPE-DON, and throughout all DON samples (Fig. A.4). These correlations were strongest in HMW DON (p < 0.001) and HMW and LMW SPE-DON combined (p <0.001) than only LMW SPE-DON samples (p = 0.01 to 0.03).

A.3.2 $\delta^{13}C$-AA patterns

In contrast with $\delta^{15}N$-AA patterns, there is no clear distinction in $\delta^{13}C$-AA patterns between ocean basins or size fractions (Fig. A.5, Table A.2). $\delta^{13}C$-AA patterns of both HMW and LMW SPE-DON most closely resemble $\delta^{13}C$-AA

patterns of autotrophs (McCarthy et al., 2004). In addition, difference plots of deep – surface δ^{13}C-AA values are generally similar between size fractions (Fig. A.6). One exception is deep HMW DON at HOT, which exhibits particularly depleted δ^{13}C values for many AA, including Ile, Leu, Val, Glx, and Pro.

A.3.3 δ^{13}C-AA Fingerprinting

PCA of relevant environmental endmembers to open ocean DOM (Table A.3) using only EAA (Thr, Ile, Val, Leu) resulted in the best separation while still retaining DOM samples within the mixing hull of endmembers, though separation of autotrophs (cyanobacteria and eukaryotic algae) was poor. PC1 accounted for 65% of the variation, while PC2 accounted for 27% of the variation.

Like PCA, LDA resulted in good separation of autotrophs and heterotrophs, but separation of cyanobacteria and eukaryotic algae was poor (Fig. A.7). The model correctly predicted 86% of the 14 endmember samples (Table A.4). Of these samples, all heterotrophic bacteria samples were all correctly predicted, but the model did not correctly distinguish between eukaryotic algae and cyanobacteria 25% of the time. Of the 9 HMW DON samples, the model predicted 5 to be heterotrophic bacteria, 3 to be eukaryotic algae, and one to be cyanobacteria (Table A.4). The model predicted 7 LMW SPE-DON samples to be eukaryotic algae, and one to be cyanobacteria.

A.4 Discussion

A.4.1 δ^{13}C$_{\text{Bulk}}$ and δ^{13}C$_{\text{THAA}}$ of total DOM, HMW DOM, and LMW SPE-DOM

δ^{13}C$_{\text{Bulk}}$ values of total DOM, HMW and LMW SPE-DOM were published previously (Broek et al., 2020) and are discussed here briefly as context for δ^{13}C-

AA data. Initial reports of total DOC δ^{13}C values found constant values with depth throughout the ocean within the range of marine primary production ($\sim$ -21.5‰) (Druffel et al., 1992; Williams, 1968; Williams & Gordon, 1970). A more recent study reports more variable δ^{13}C values of total DOC, demonstrating a decrease with depth throughout the mesopelagic and increasing again in the deep ocean (Takano et al., 2010). The values reported here are consistent with this more recent study (Fig. A.2) (Broek et al., 2020), and could potentially indicate a loss of isotopically heavy AA and persistence of isotopically light lipids or other refractory molecules (Hayes, 2001).

The increasing δ^{13}C$_{\text{Bulk}}$ values of HMW DOM reported here (Fig. A.2) (Broek et al., 2020) are in contrast to initial ultrafiltration (UF) studies which found δ^{13}C$_{\text{Bulk}}$ values of HMW DOM to be fairly constant with depth (Benner et al., 1997). However, again, our data is consistent with a more recent study reporting HMW DOM δ^{13}C data (Sannigrahi et al., 2005). Based on limited changes to C/N ratio with depth, Broek et al. 2020 suggested the lower surface δ^{13}C values of HMW DOM was due to either a small surface contribution of labile, very isotopically light lipids or isotopically light N-containing molecules, both of which would be progressively removed with depth. δ^{13}C$_{\text{Bulk}}$ of LMW SPE-DOM also decreased slightly with depth, and a strong negative relationship between δ^{13}C$_{\text{Bulk}}$ and C/N ratio suggested removal of labile, C-rich material and persistence of more refractory N material (Broek et al., 2020).

The δ^{13}C-AA data presented here is more novel, representing the first high-precision δ^{13}C-AA data of marine DON. δ^{13}C-AA values have only been published previously for four HMW DON at HOT (McCarthy et al., 2004), and these results had very large error bars and were missing key amino acids due to analytical limitations at that time. The distinct depth structure of HMW and LMW SPE-

DON $\delta^{13}C_{THAA}$ compared to $\delta^{13}C_{Bulk}$ indicates resynthesis or recycling of carbon atoms in THAA which is not occurring within the greater DOC pool (Fig. A.3). Additionally, strong relationships between $\delta^{13}C_{EAA}$ and $\delta^{13}C_{NEAA}$ with $\delta^{13}C_{THAA}$ within both size fractions (Fig. A.4) imply resynthesis or degradation mechanisms of marine DOM are similar for all AA, rather than distinct for EAA versus NEAA. Additionally, the very similar trends in HMW DON and LMW SPE-DON suggest that the difference in $\delta^{13}C_{THAA}$ values of these two size fractions is a result of differing molecular sources rather than selective bacterial resynthesis, which would likely only fractionate NEAA (Keil & Fogel, 2001; McCarthy et al., 2004). This is consistent with D/L-AA data and $\delta^{15}N$-AA data, which both indicated distinct bacterial molecular sources to HMW versus LMW SPE-DON.

In HMW DON, surface $\delta^{13}C_{THAA}$ values are slightly heavier than $\delta^{13}C_{Bulk}$ values and the average $\delta^{13}C$ values of marine primary production (Fig. A.3), consistent with previous work indicating AA are enriched in ^{13}C compared to other biomolecules (Macko et al., 1987, Degens et al., 1968). At 400 m, the decrease in $\delta^{13}C_{THAA}$ corresponds with an increase in ΣV, a $\delta^{15}N$-AA based proxy for heterotrophic bacterial resynthesis (Ianiri et al., In prep), and the D/L ratios of some AA (Ianiri et al., Submitted). This could indicate that the decrease in $\delta^{13}C_{THAA}$ may be a local signal of mesopelagic resynthesis, which woul d be consistent with $\delta^{15}N$-AA data. However, heterotrophic activity is expected to utilize isotopically light molecules and leave behind heavier HMW DOM, which is opposite to the decrease in $\delta^{13}C_{THAA}$ we observe.

Alternatively, the lower $\delta^{13}C_{THAA}$ values in the mesopelagic could be a result of bacterial or archaeal production utilizing isotopically light DIC to produce DOM. Indeed, radiocarbon measurements of mesopelagic microbes and DOM indicated production of DOM from *in situ* chemoautotrophy (Hansman et al., 2009). Re-

216

gardless of the exact mechanism, it appears the decrease in $\delta^{13}C_{THAA}$ at 400 m may be a local mesopelagic signal.

Perhaps most intriguing in the HMW $\delta^{13}C_{THAA}$ profile is the large difference in deep ocean (2500 m) $\delta^{13}C_{THAA}$ values at HOT vs. BATS, with values at HOT $\sim$ 6‰ lighter than at BATS. Additionally, the low 2500 m $\delta^{13}C_{THAA}$ value at HOT is $\sim$ 2.5‰ lighter than $\delta^{13}C_{Bulk}$ of HMW DOM at the same depth (Fig. A.3), in contrast to the nearly universal observations that $\delta^{13}C_{THAA}$ is heavier than $\delta^{13}C_{Bulk}$ (Abelson and Hoering 1961, Keil and Fogel 2001, Macko et al., 1987). While these observations are based only on one point, they still warrant some attention.

Considering deep waters at HOT are some of the oldest in the ocean, it is possible these lighter values represent a heterotrophic bacterial degradation or resynthesis signal. This would be consistent with the high ΣV values observed in deep waters at HOT (Ianiri et al., Submitted), however, as discussed above, resynthesis or degradation would generally be expected to increase $\delta^{13}C$-AA values.

An alternative possibility is a deep-water source of isotopically light AA at HOT. $\delta^{13}C$ values of suspended PON decrease with depth in the Central Pacific, with deep water values ranging from $\sim$ -24‰ to -26‰ (Benner et al., 1997; Sannigrahi et al., 2005). Thus, a deep particle source could decrease $\delta^{13}C_{THAA}$ values in this ocean basin. If true, the large offset between $\delta^{13}C_{THAA}$ and $\delta^{13}C_{Bulk}$ suggests deep ocean production utilizing an isotopically light carbon source is preferentially used to produce AA, while the $\delta^{13}C$ of the majority of the HMW DOC pool remains unchanged. In contrast to the Pacific Ocean, deep-water (450 m to 3400 m) $\delta^{13}C$ values of suspended POM in the Central Atlantic are heavier, averaging -22.2‰ (Pedrosa-Pàmies et al., 2018). Still, these values are depleted compared to

$\delta^{13}C_{THAA}$ values of HMW DOM we observe at BATS, meaning there is no support for a similar mechanism in this ocean basin.

Finally, *in situ* chemoautotrophy utilizing isotopically light DIC could also be a source of light $\delta^{13}C$-AA values in the deep ocean at HOT. However, previous work indicated *in situ* chemoautotrophy was more common in the mesopelagic (670 m), while microbial communities at 915 m were more likely to rely on sinking POM (Hansman et al., 2009). Additional CSI-AA analyses on an expanded DOM sample set will be required to verify if this low $\delta^{13}C$-AA signal is ubiquitous throughout the deep Pacific and what are the most likely mechanisms causing this depletion.

$\delta^{13}C_{THAA}$ depth profiles of LMW SPE-DON were distinct compared to HMW DON (Fig. A.3), consistent with observations from $\delta^{15}N$-AA data which suggested different molecular sources to the two size fractions (Ianiri et al., Submitted). The greater surface offset between $\delta^{13}C_{THAA}$ and $\delta^{13}C_{Bulk}$ of LMW SPE-DON compared to HMW DON ($\sim$ 5‰ vs. $\sim$ 2.5‰, Fig. A.3) is consistent with expectations that most LMW SPE-DOC is isotopically light molecules, such as lipids or carboxyl-rich alicyclic molecules (CRAM).

The increase in LMW SPE-DON $\delta^{13}C_{THAA}$ values at 400 m is opposite to trends observed in HMW DON. It is possible this increase could indicate selective removal of isotopically light AA throughout the mesopelagic (Fig. A.3). This is consistent with D and L-AA yield data for LMW SPE-DON, which indicates selective removal of L-AA in the mesopelagic (Ianiri et al., Submitted). Below 400 m, $\delta^{13}C_{THAA}$ values are relatively similar to values observed at the surface, and likely represent a global, background average value. The statistically similar $\delta^{13}C_{THAA}$ values at 2500 m at BATS and HOT support an averaged, background pool of LMW THAA, which is $\sim$ 6‰ heavier than total LMW SPE-DOM (Fig. A.3)

A.4.2 Autotrophic vs. heterotrophic source to HMW and LMW SPE-DOM

Previous work measuring this same sample set found δ^{15}N-AA patterns of HMW and LMW SPE-DON were starkly different, with LMW SPE-DON resembling autotrophic organisms and HMW DON exhibiting signs of heterotrophic bacterial resynthesis (Ianiri et al., In prep). In contrast, δ^{13}C-AA patterns of both HMW and LMW SPE-DON appear similar at most depths and in both ocean basins (Fig. A.5). However, there are some differences which may have potential to distinguish between an autotrophic versus heterotrophic source.

While δ^{13}C-AA *values* are generally not comparable between studies, specific trends in δ^{13}C-AA patterns can be indicative of autotrophic or heterotrophic bacterial production. For example, many heterotrophic prokaryotes appear to enrich δ^{13}C values of Gly and Ala while depleting Ile, Val, and Leu (Keil & Fogel, 2001; McCarthy et al., 2004; Ziegler & Fogel, 2002). If heterotrophic resynthesis were altering HMW or LMW δ^{13}C-AA patterns, we may expect to see these trends in deep – surface δ^{13}C-AA values. However, we do not see these patterns for most HMW or LMW SPE-DON samples (Fig. A.6). While some HMW and LMW samples at BATS demonstrated enrichment in Gly, almost no samples showed enrichment of Ala or depletion of Ile, Val, and Leu. The only exception is deep (2500 m) HMW DON at HOT, which was uniformly depleted for most AA. Instead, the δ^{13}C-AA patterns seem to more closely resemble those of autotrophs (Larsen et al., 2009, 2013; McCarthy et al., 2004). These results are consistent with preliminary data reported by McCarthy et al. (2004), who found HMW DON δ^{13}C-AA signatures resembled those of autotrophs and were similar throughout the water column in the Pacific Ocean.

Multivariate analysis of δ^{13}C$_{\text{EAA}}$ values provides an additional tool to inves-

tigate metabolic source. Because $\delta^{13}C_{EAA}$ values do not change with trophic transfer, they can be used to distinguish between different primary producers, including algae, terrestrial plants, fungi, and heterotrophic bacteria (Larsen et al., 2009, 2013). However, of the previously published $\delta^{13}C$-AA endmember data, only open ocean marine producers and heterotrophic bacteria are environmentally relevant to DOM in oligotrophic gyres. Previous work applying $\delta^{13}C$-AA multi-variate analyses to open ocean POM found that LDA had difficulty distinguishing heterotrophic bacteria when data from multiple studies was included, likely due to the large diversity of bacterial metabolism and degradation mechanisms (Hannides et al., 2013). This suggests there may be complications in quantitative source estimates where heterotrophic bacteria could be an important contributor. Additionally, past work has found $\delta^{13}C$-AA patterns of autotrophic eukaryotes and prokaryotes are very similar (McCarthy et al., 2004). These complications question whether currently published endmember data accurately represent and can distinguish between all potential sources and production mechanisms of open ocean DOM.

Indeed, LDA analysis of previously published cyanobacteria, eukaryotic algae, and heterotrophic bacteria data (Table A.3) shows clear separation of autotrophs and heterotrophs but does not distinguish well between different groups of autotrophs (Fig. A.7). Additionally, it is unclear how well cultured heterotrophic bacterial endmembers from two studies (Larsen et al., 2009, 2013) represent open ocean heterotrophic bacteria. Finally, it is important to note LDA will classify each sample as a single source, while it is likely that all DOM samples derive from a range of sources. Still, with these caveats in mind, some preliminary interpretations can be made with the currently available data.

The prediction that all LMW SPE-DON are autotrophic-derived is consistent

220

with both δ^{13}C-AA and δ^{15}N-AA patterns of LMW SPE-DON (Table A.4, Ianiri et al., In prep). In contrast, the HMW DON results were less clear. The fact that all HMW DON samples from BATS were predicted as heterotrophic bacteria, while all but one HMW DON sample from HOT was predicted as autotrophs, suggests some differences in δ^{13}C-AA patterns, and thus metabolic source, of HMW DON from the Pacific and Atlantic Oceans. This is in stark contrast to δ^{15}N-AA data, which was nearly identical at both locations (discussed in further detail below). The prediction of a heterotrophic bacterial source to HMW DON is more consistent with δ^{15}N-AA patterns, which suggested heterotrophic resynthesis was had altered HMW DOM CSI-AA values. Still, as noted above, these results should be interpreted only as preliminary data. I suggest expanding currently available δ^{13}C-AA endmembers, particularly of microbial species, will assist in applying this tool to marine DON. A robust and complete endmember dataset is also necessary to apply more complex mixing models, such as MixSIAR.

A.4.3   Decoupling of δ^{13}C-AA and δ^{15}N-AA patterns

A notable trend in this dataset is an apparent decoupling in trends of δ^{13}C-AA and δ^{15}N-AA data. δ^{15}N-AA patterns were incredibly similar within each size fraction, instead demonstrating major differences between HMW and LMW SPE-DON (Ianiri et al., In prep). Additionally, HMW DON uniformly appeared resynthesized by bacteria, while LMW SPE-DON resembled autotrophic organisms. In contrast, δ^{13}C-AA patterns appear to show more differences between ocean basins, and LDA analysis only indicated a heterotrophic source to about half of all HMW DOM samples. This decoupling of δ^{13}C-AA and δ^{15}N-AA trends could imply microbial resynthesis of HMW DOM is fractionating δ^{15}N-AA values but not δ^{13}C-AA values. In this case, bacterial resynthesis would only recycle

the amino N, causing fractionation of δ^{15}N-AA patterns, while leaving the C-containing backbone and side chains intact. This would result in similar δ^{13}C-AA patterns of HMW and LMW SPE-DOM, which both resemble δ^{13}C-AA patterns of autotrophs, despite heterotrophic resynthesis evident in the N isotope data. Again, further CSI-AA analysis of a more extensive data set will be required to verify these trends.

A.5   Conclusions

The data presented here represents the first δ^{13}C-AA analyses of LMW DON, as well as DON from the Atlantic Ocean. These novel results can be summarized in three main conclusions. First, there is no distinction between fractionation of EAA and NEAA in HMW or LMW SPE-DON, suggesting the distinct δ^{13}C$_{\text{THAA}}$ values and depth trends between these two size fractions are a result of unique molecular sources, not bacterial resynthesis. This conclusion is consistent with D/L-AA and δ^{15}N-AA data, which suggests the two size fractions derive from different bacterial sources. Second, in contrast to δ^{15}N-AA data, δ^{13}C-AA patterns do not demonstrate significant differences between HMW and LMW SPE-DON. Instead, δ^{13}C-AA patterns and deep – surface δ^{13}C-AA offsets show greater similarities between ocean basin than within size fractions. LDA predictions additionally support differences in δ^{13}C-AA patterns between ocean basins. These differences suggest a decoupling between δ^{15}N-AA fractionation due to heterotrophic resynthesis and δ^{13}C-AA patterns which appear more representative of a local metabolic source. Finally, application of multivariate analyses yields promising results for this technique to study of marine DON. However, these results should be interpreted cautiously, and I suggest a more substantial open ocean endmember dataset is necessary to apply more complex techniques such as mixing models.

Together, these results provide new information regarding the cycling of HMW and LMW AA-containing molecules. Additionally, these results provide strong support for the use of δ^{13}C-AA as an invaluable tool for the study of marine DON source and cycling.

Figures and tables
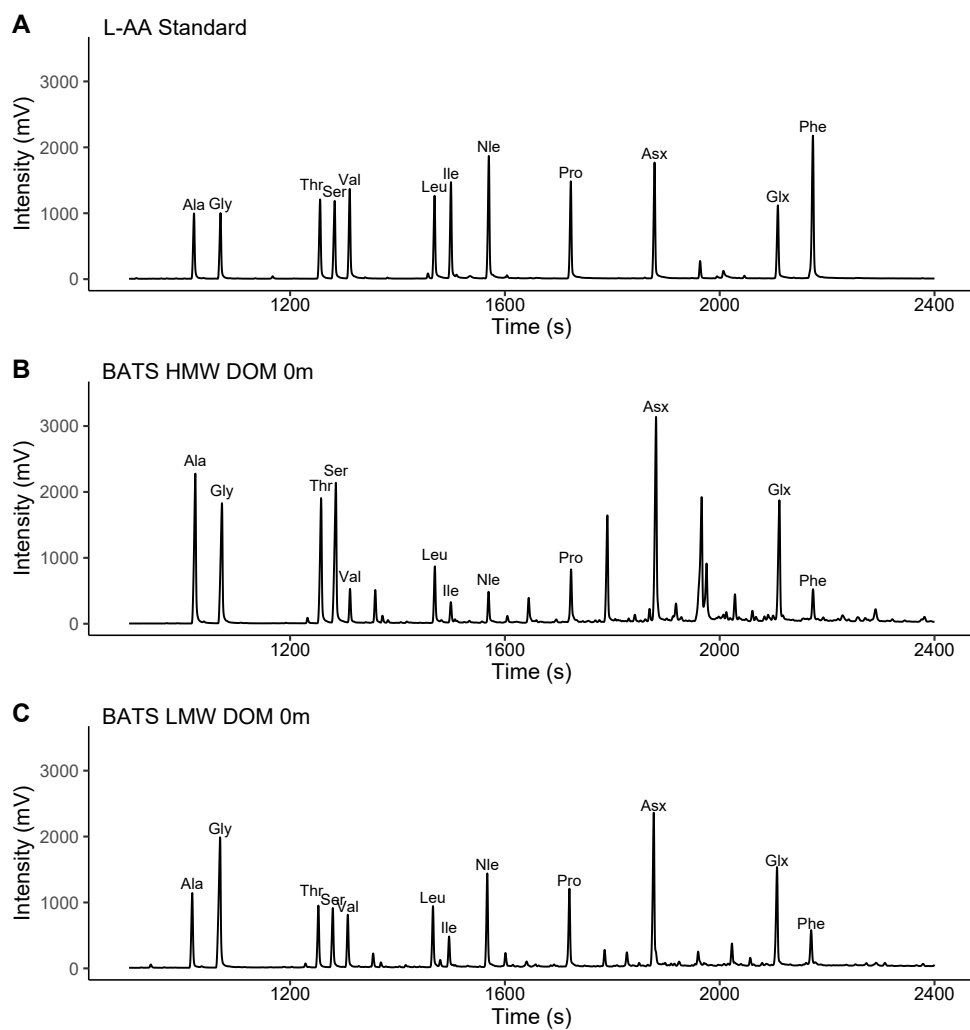
Figure A.1: δ^{13}C GC-IRMS chromatograms of A) an L-AA standard, B) HMW DON from the surface ocean at BATS, and C) LMW SPE-DON from the surface ocean at BATS. Substantial upstream purification of DOM samples (Section 3.3.3) resulted in few C-containing molecules besides target AA in either size fraction.
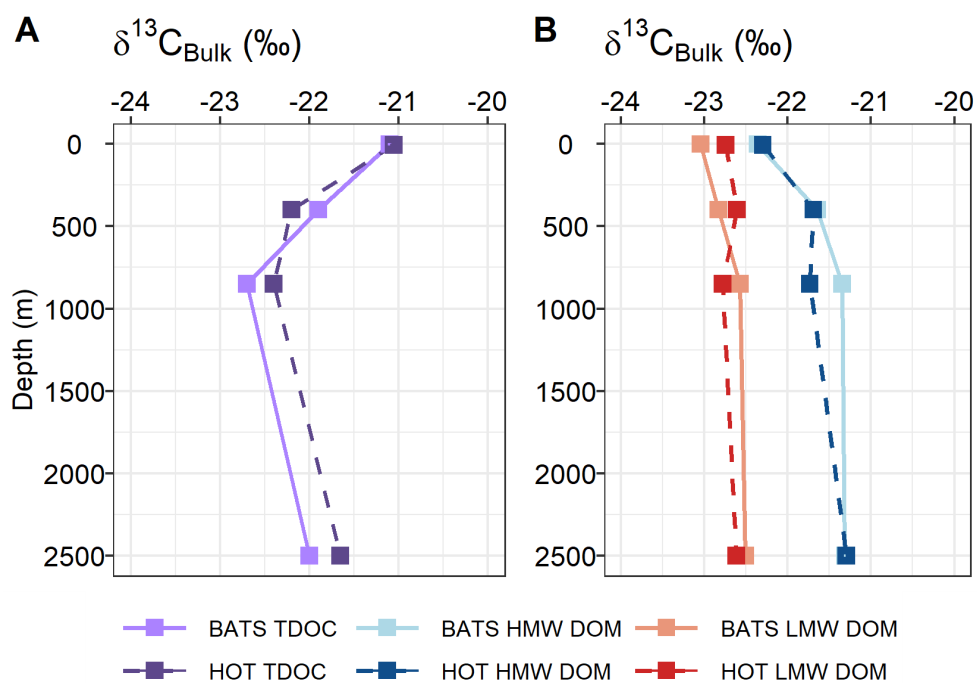
Figure A.2: Depth profiles of $\delta^{13}C_{Bulk}$ for A) total DON (purple) and B) HMW DOM (blue) and LMW SPE-DOM (red) from HOT (dashed lines, darker shade) and BATS (solid lines, lighter shade). Values represent averages of spring and summer cruises and error bars represent the propagated analytical error of triplicate measurements except for total DOC at BATS, for which $\delta^{13}C$ values were only measured on material collected during the May sampling cruise and error bars represent the instrument uncertainty. Error bars are smaller than symbol where not visible.
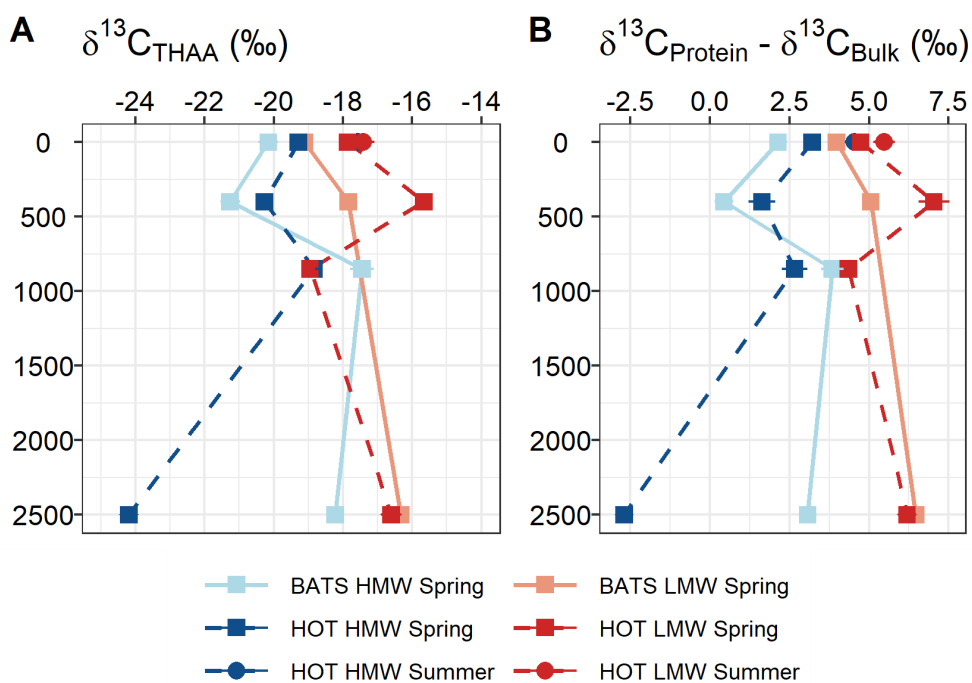
Figure A.3: Depth profiles of A) $\delta^{13}C_{THAA}$ and B) the offset of $\delta^{13}C_{Protein\text{-}Bulk}$ of HMW DON (blue) and LMW SPE-DON (red) at HOT (dashed lines, darker shade) and BATS (solid lines, lighter shade). HMW and LMW SPE-DON were measured throughout the water column in spring (squares) and at the surface in the summer at HOT (circles). Error bars for $\delta^{13}C_{THAA}$ represent the propagated error of the standard deviation of duplicate or triplicate δ^{13}C-AA and mol% measurements. $\delta^{13}C_{Protein\text{-}Bulk}$ error bars represent the propagated error of the two measurements. Error bars are smaller than symbol where not visible.
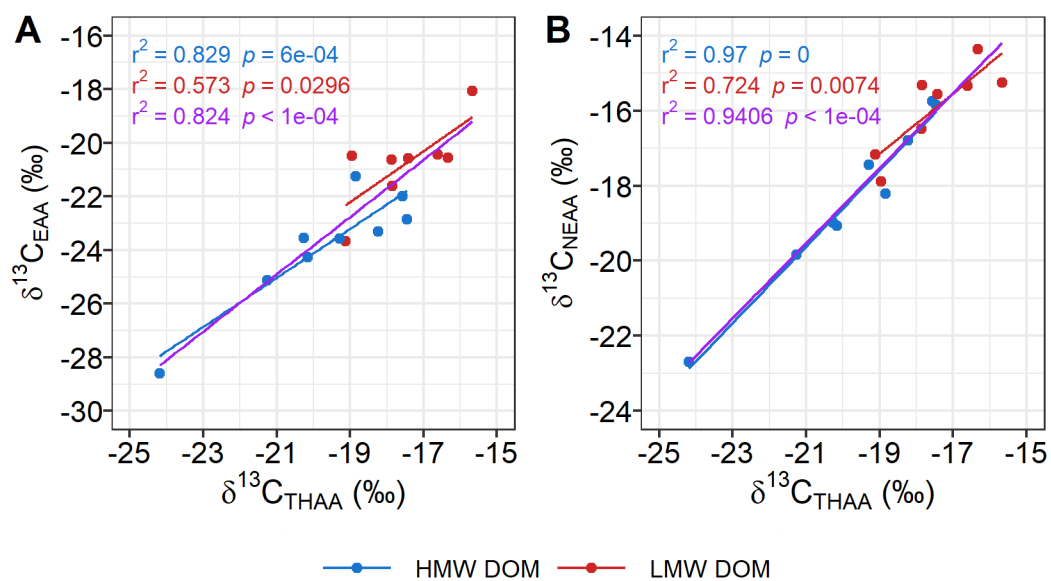
Figure A.4: A) $\delta^{13}C_{EAA}$ and B) $\delta^{13}C_{NEAA}$ are both significantly correlated with $\delta^{13}C_{THAA}$ for HMW DON (blue), LMW SPE-DON (red), and HMW and LMW SPE-DON (purple).
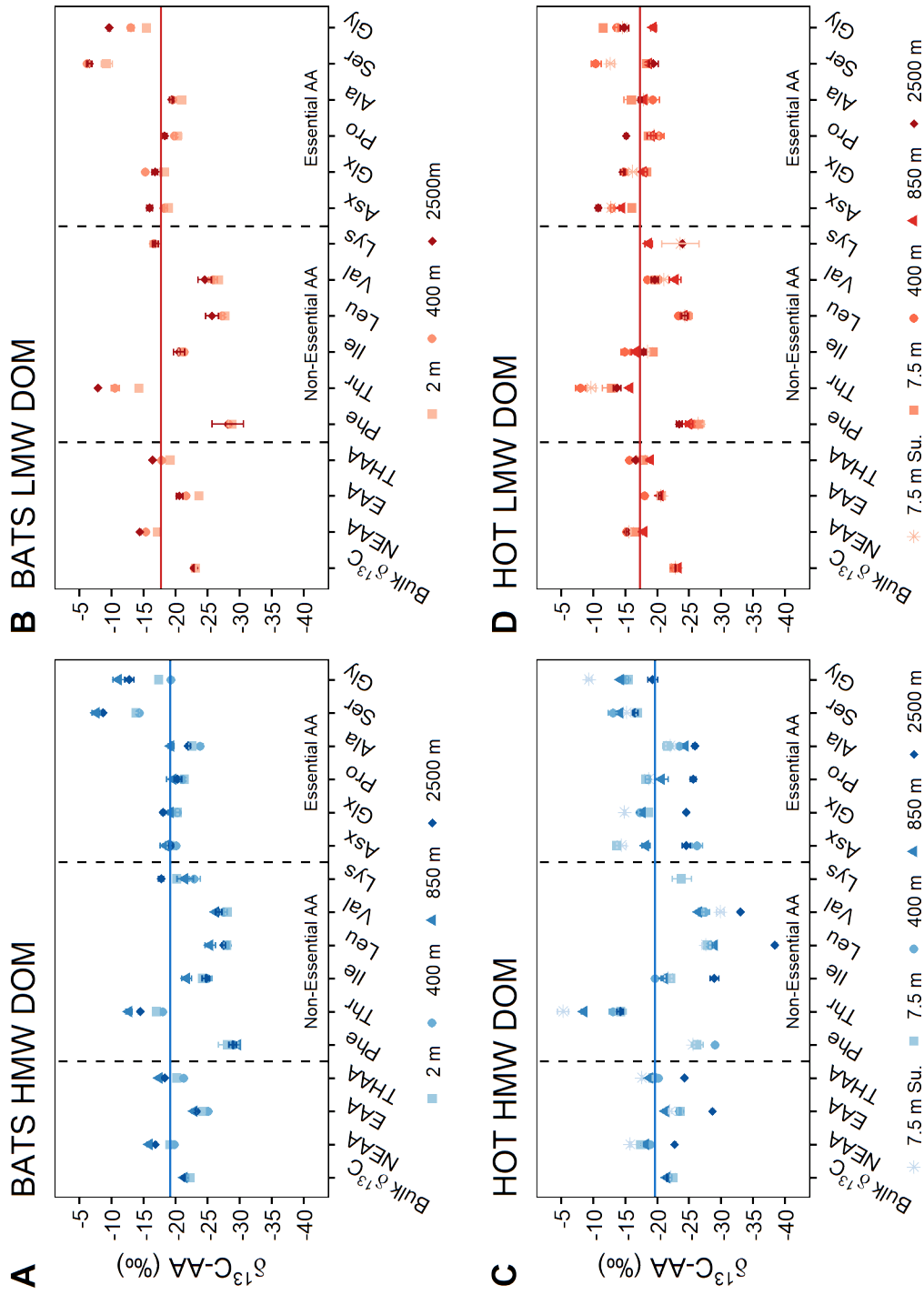
Figure A.5: δ^{13}C-AA patterns of A) HMW DON at BATS, B) LMW SPE-DON at BATS, C) HMW DON at HOT, D) LMW SPE-DON at HOT. Error bars represent the standard deviation of triplicate measurements unless otherwise noted in the text. Solid horizontal line indicates average $\delta^{13}C_{THAA}$ for all samples in that size fraction/ocean basin. AA are grouped as described in text. Samples labeled "Su." were collected during summer cruises, while all others were collected on spring cruises.
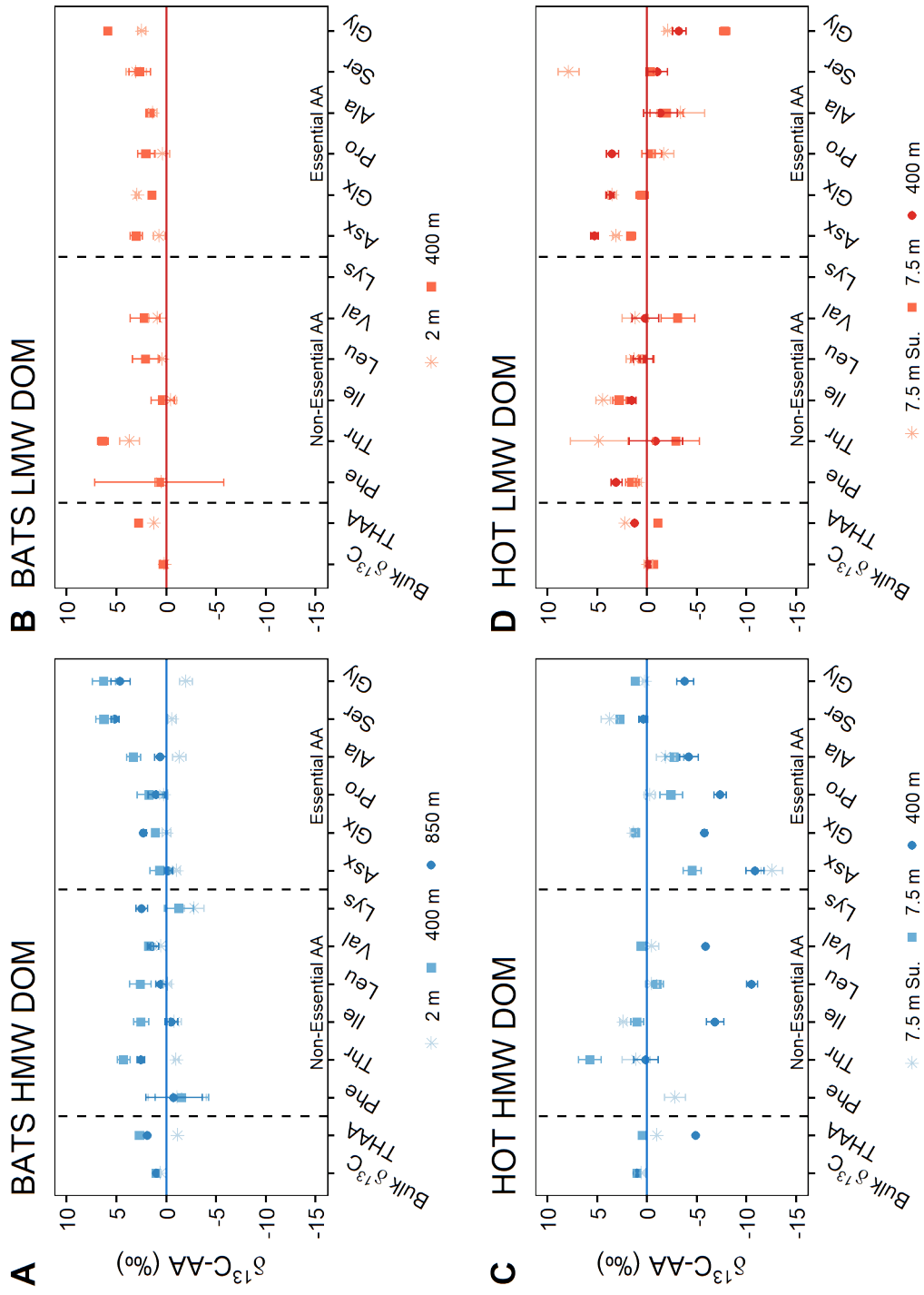
Figure A.6: Difference in deep - surface δ^{13}C-AA values of A) HMW DON at BATS, B) LMW SPE-DON at BATS, C) HMW DON at HOT, D) LMW SPE-DON at HOT. Error bars represent the propagated error of standard deviation of triplicate measurements unless otherwise noted in the text. AA are grouped as described in the text. Samples labeled "Su." were collected during summer cruises, while all others were collected on spring cruises.
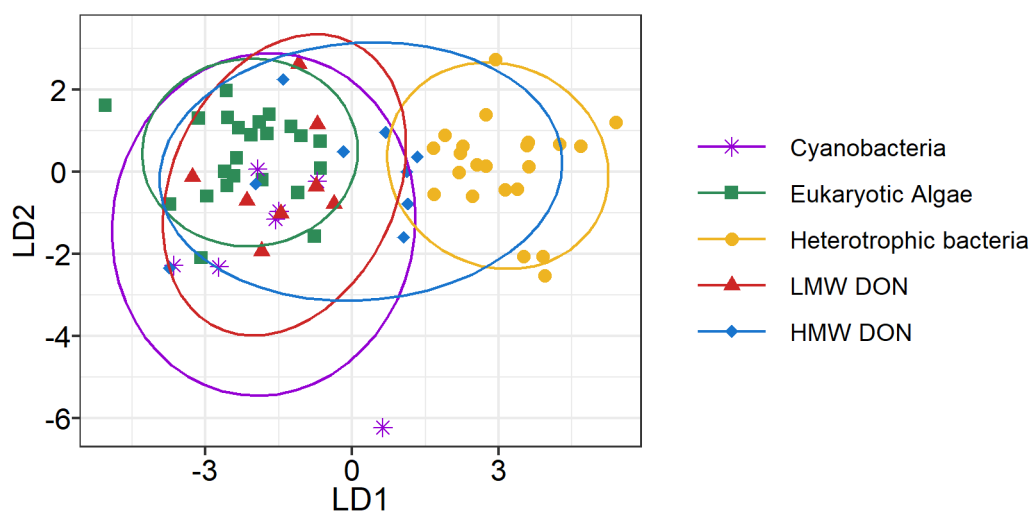
Figure A.7: LDA analysis of HMW DON, LMW SPE-DON, cyanobacteria, eukaryotic algae, and heterotrophic bacteria. Only EAA measured in every sample were used (Thr, Ile, Leu, Val). Source endmember data is from Larsen et al., 2009, 2013 and Lehman 2009 (Table A.3). HMW and LMW SPE-DON overlap with autotrophs, while heterotrophic bacteria are more clearly separated. Ellipses represent 95% confidence intervals.

Table A.1: Bulk δ^{13}C values of total, HMW, and LMW SPE-DOM measured during cruises to BATS and HOT.

Location	Type	Season	Depth (m)	δ^{13}C‰	$\pm$
HOT	HMW DOM	Summer	7.5	-22.12	0.05
HOT	HMW DOM	Summer	400	-21.48	0.01
HOT	HMW DOM	Summer	850	-21.92	0.05
HOT	HMW DOM	Summer	2500	-21.14	0.04
HOT	HMW DOM	Spring	7.5	-22.48	0.01
HOT	HMW DOM	Spring	400	-21.90	0.33
HOT	HMW DOM	Spring	850	-21.54	0.31
HOT	HMW DOM	Spring	2500	-21.46	0.06
BATS	HMW DOM	Summer	2	-22.41	0.07
BATS	HMW DOM	Summer	400	-21.62	0.05
BATS	HMW DOM	Summer	850	-21.38	0.08
BATS	HMW DOM	Summer	2500	-21.28	0.06
BATS	HMW DOM	Summer	2500	-21.36	0.04
BATS	HMW DOM	Spring	2	-22.31	0.01
BATS	HMW DOM	Spring	400	-21.67	0.04
BATS	HMW DOM	Spring	850	-21.31	0.09
BATS	HMW DOM	Spring	2500	-21.35	0.11
HOT	LMW SPE-DOM	Summer	7.5	-22.85	0.06
HOT	LMW SPE-DOM	Summer	400	-22.50	0.01
HOT	LMW SPE-DOM	Summer	850	-22.23	0.04
HOT	LMW SPE-DOM	Summer	2500	-22.41	0.09
HOT	LMW SPE-DOM	Spring	7.5	-22.64	0.03

Location	Type	Season	Depth (m)	δ^{13}C‰	±
HOT	LMW SPE-DOM	Spring	400	-22.72	0.42
HOT	LMW SPE-DOM	Spring	850	-23.32	0.02
HOT	LMW SPE-DOM	Spring	2500	-22.82	0.06
BATS	LMW SPE-DOM	Summer	2	-22.99	0.04
BATS	LMW SPE-DOM	Summer	400	-22.74	0.05
BATS	LMW SPE-DOM	Summer	850	-22.26	0.09
BATS	LMW SPE-DOM	Summer	2500	-22.18	0.04
BATS	LMW SPE-DOM	Summer	2500	-22.14	0.02
BATS	LMW SPE-DOM	Spring	2	-23.09	0.08
BATS	LMW SPE-DOM	Spring	400	-22.92	0.00
BATS	LMW SPE-DOM	Spring	850	-22.89	0.03
BATS	LMW SPE-DOM	Spring	2500	-22.83	0.02
HOT	TDOC	Summer	7.5	-20.80	0.20
HOT	TDOC	Summer	400	-22.70	0.20
HOT	TDOC	Summer	850	-22.40	0.20
HOT	TDOC	Summer	2500	-21.80	0.20
HOT	TDOC	Spring	7.5	-21.30	0.20
HOT	TDOC	Spring	400	-21.70	0.20
HOT	TDOC	Spring	2500	-21.50	0.20
BATS	TDOC	Spring	2	-21.10	0.20
BATS	TDOC	Spring	400	-21.90	0.20
BATS	TDOC	Spring	850	-22.70	0.20
BATS	TDOC	Spring	2500	-22.00	0.20

Table A.2: δ^{13}C-AA values HMW and LMW SPE-DON.

Location	Size	Year	Depth	δ^{13}C NEAA	±	δ^{13}C EAA	±	δ^{13}C THAA	±	Phe	±	Thr	±	Ile	±	Leu	±
BATS	HMW	2016	2	-19.07	0.17	-24.27	0.28	-20.16	0.28	-28.16	1.50	-16.99	0.39	-24.23	0.18	-28.00	0.48
BATS	HMW	2016	400	-19.83	0.13	-25.13	0.23	-21.26	0.29	-29.24	0.65	-17.97	0.09	-25.02	0.64	-28.11	0.15
BATS	HMW	2016	850	-15.85	0.31	-22.86	0.31	-17.45	0.34	-29.70	0.43	-12.69	0.50	-21.68	0.75	-25.38	0.85
BATS	HMW	2016	2500	-16.80	0.18	-23.30	0.19	-18.23	0.23	-28.89	0.56	-14.42	0.24	-24.71	0.61	-27.46	0.31
BATS	LMW	2016	2	-17.17	0.20	-23.67	0.20	-19.12	0.17	-28.79	0.50	-14.23	0.53	-20.88	0.43	-27.73	0.36
BATS	LMW	2016	400	-15.33	0.15	-21.62	0.20	-17.84	0.20	-28.26	0.21	-10.52	0.67	-21.33	0.38	-27.30	0.27
BATS	LMW	2016	2500	-14.37	0.16	-20.56	0.50	-16.33	0.17	-28.08	2.45	-7.83	0.11	-20.49	0.85	-25.66	1.00
HOT	HMW	2014	7.5	-15.76	0.22	-21.99	0.23	-17.57	0.30	-25.56	0.19	-5.27	0.87	-21.67	0.51	-27.60	0.67
HOT	HMW	2015	7.5	-17.44	0.18	-23.56	0.35	-19.28	0.25	-26.26	0.92	-14.30	0.86	-22.09	0.44	-27.86	0.57
HOT	HMW	2015	400	-18.98	0.22	-23.55	0.21	-20.26	0.25	-29.04	0.20	-13.16	0.66	-19.69	0.15	-28.34	0.33
HOT	HMW	2015	850	-18.22	0.21	-21.26	0.18	-18.84	0.24	NA	NA	-8.54	0.42	-21.08	0.46	-28.91	0.27
HOT	HMW	2015	2500	-22.70	0.21	-28.61	0.22	-24.19	0.26	NA	NA	-14.17	0.48	-28.91	0.67	-38.39	0.25
HOT	LMW	2014	7.5	-15.56	0.29	-20.59	0.54	-17.42	0.33	-26.45	0.85	-9.67	0.70	-18.53	0.25	-24.30	0.24
HOT	LMW	2015	7.5	-16.49	0.24	-20.62	0.38	-17.86	0.26	-26.48	0.34	-12.84	1.45	-19.38	0.19	-24.68	0.74
HOT	LMW	2015	400	-15.25	0.26	-18.07	0.24	-15.66	0.24	-25.53	0.23	-7.99	0.80	-14.96	0.69	-23.36	0.25
HOT	LMW	2015	850	-17.89	0.19	-20.49	0.26	-18.95	0.15	-24.97	0.55	-15.75	0.23	-16.59	0.65	-24.20	0.64
HOT	LMW	2015	2500	-15.33	0.19	-20.45	0.18	-16.60	0.29	-23.40	0.43	-13.70	0.61	-17.85	0.37	-24.26	0.47

Continued on next page

233

Location	Size	Year	Depth	Val	±	Lys	±	Asx	±	Glx	±	Pro	±	Ala	±	Ser	±	Gly	±
BATS	HMW	2016	2	-28.06	0.27	-20.19	0.39	-19.04	0.35	-20.29	0.14	-21.36	0.44	-22.57	0.50	-13.83	0.42	-17.34	0.53
BATS	HMW	2016	400	-27.49	0.49	-22.95	0.86	-20.06	0.20	-20.28	0.40	-21.10	0.23	-23.85	0.41	-14.39	0.22	-19.30	0.36
BATS	HMW	2016	850	-26.28	0.25	-21.45	1.29	-18.39	0.88	-19.18	0.34	-19.64	1.07	-19.26	0.44	-7.57	0.65	-11.05	0.89
BATS	HMW	2016	2500	-26.68	0.48	-17.67	0.42	-19.23	0.30	-17.98	0.24	-20.31	0.62	-21.93	0.32	-8.65	0.23	-12.71	0.68
BATS	LMW	2016	2	-26.71	0.39	NA	NA	-18.93	0.38	-18.21	0.19	-20.27	0.68	-20.95	0.24	-9.18	0.88	-15.46	0.20
BATS	LMW	2016	400	-25.79	0.70	-16.53	0.51	-18.17	0.46	-15.26	0.05	-19.86	0.28	-19.56	0.41	-6.12	0.36	-12.98	0.46
BATS	LMW	2016	2500	-24.52	1.05	-16.80	0.45	-15.90	0.48	-16.75	0.45	-18.23	0.40	-19.31	0.33	-6.48	0.44	-9.57	0.22
HOT	HMW	2014	7.5	-29.85	0.60	NA	NA	-14.40	0.70	-14.87	0.16	-18.65	0.31	-22.07	0.73	-15.23	0.60	-9.30	0.45
HOT	HMW	2015	7.5	-27.07	0.20	-23.80	1.49	-13.69	0.51	-18.71	0.25	-18.25	0.40	-21.62	0.79	-16.90	0.20	-15.46	0.12
HOT	HMW	2015	400	-27.51	0.68	NA	NA	-26.22	0.82	-17.34	0.30	-18.51	0.37	-23.47	0.31	-13.14	0.81	-15.22	0.24
HOT	HMW	2015	850	-26.51	0.21	NA	NA	-18.23	0.65	-17.60	0.07	-20.66	0.98	-24.36	0.30	-14.19	0.26	-14.28	0.04
HOT	HMW	2015	2500	-32.98	0.10	NA	NA	-24.52	0.63	-24.48	0.15	-25.57	0.47	-25.83	0.31	-16.52	0.40	-19.26	0.82
HOT	LMW	2014	7.5	-21.01	0.74	-23.59	2.91	-12.74	0.23	-16.16	0.64	-19.46	0.77	-17.82	0.72	-12.64	0.66	-14.57	1.02
HOT	LMW	2015	7.5	-19.73	0.89	NA	NA	-16.06	0.14	-18.34	0.20	-18.63	0.57	-15.98	1.19	-18.31	0.51	-11.60	0.17
HOT	LMW	2015	400	-18.52	0.51	NA	NA	-12.92	0.28	-14.85	0.37	-20.31	0.70	-19.35	0.98	-10.41	0.77	-13.67	0.38
HOT	LMW	2015	850	-22.80	0.91	-18.65	0.60	-14.43	0.41	-17.84	0.57	-19.09	0.65	-17.94	0.24	-18.62	0.16	-19.40	0.48
HOT	LMW	2015	2500	-19.56	0.55	-23.91	NA	-10.77	0.37	-14.58	0.37	-15.14	0.26	-17.32	0.27	-19.37	0.73	-14.82	0.65

Table A.3: Endmember dataset used for LDA. "Het. Bac" stands for heterotrophic bacteria, "Euk. Algae" stands for eukaryotic algae, and "Cyano." stands for cyanobacteria. ID corresponds to IDs from the original datasets in the corresponding publications.

ID	Group	Reference	Thr-norm	Ile-norm	Val-norm	Phe-norm	Leu-norm
14b	Het. Bac.	Larsen et al., 2009	-4.23	1.02	1.88	-1.01	-2.34
15b	Het. Bac.	Larsen et al., 2009	-0.46	0.60	-0.20	-0.82	-0.88
16b	Het. Bac.	Larsen et al., 2009	-0.51	-0.13	-2.44	1.52	-1.56
17b	Het. Bac.	Larsen et al., 2009	5.22	-1.22	-0.77	-3.10	0.13
18b	Het. Bac.	Larsen et al., 2009	4.57	0.16	-1.63	-4.57	-1.47
19b	Het. Bac.	Larsen et al., 2009	-4.57	2.91	1.79	-1.87	-1.74
20b	Het. Bac.	Larsen et al., 2009	-8.66	4.32	1.84	0.35	-2.15
21b	Het. Bac.	Larsen et al., 2009	1.78	0.94	-0.13	-2.03	0.55
22b	Het. Bac.	Larsen et al., 2009	-1.56	-1.18	2.08	-0.75	-1.41
23b	Het. Bac.	Larsen et al., 2009	5.89	-1.48	-1.30	-3.05	0.06
B1	Het. Bac.	Larsen et al,. 2013	7.28	0.48	-1.12	-6.32	0.32
B2	Het. Bac.	Larsen et al,. 2013	4.84	0.14	-1.06	-3.96	-0.04
B3	Het. Bac.	Larsen et al,. 2013	4.84	0.94	-0.96	-4.76	0.06
B4	Het. Bac.	Larsen et al,. 2013	8.76	0.16	-2.54	-5.04	1.34
B5	Het. Bac.	Larsen et al,. 2013	7.34	-0.46	-2.16	-2.66	2.06
B6	Het. Bac.	Larsen et al,. 2013	8.62	0.12	-3.88	-2.88	1.98
B7	Het. Bac.	Larsen et al,. 2013	8.18	-0.02	-3.42	-2.92	1.82
B8	Het. Bac.	Larsen et al,. 2013	7.24	-1.06	-2.66	-2.66	0.86
B9	Het. Bac.	Larsen et al,. 2013	6.04	1.64	-2.26	-4.16	1.26
B10	Het. Bac.	Larsen et al,. 2013	5.9	-0.5	-3.3	-2.8	-0.7
B11	Het. Bac.	Larsen et al,. 2013	7.04	0.04	-3.46	-3.26	0.36
B12	Het. Bac.	Larsen et al,. 2013	6.26	0.16	-1.84	-4.54	0.04
D1	Euk. Algae	Larsen et al,. 2013	9.72	2.22	-1.38	-5.68	4.88
D2	Euk. Algae	Larsen et al,. 2013	11.72	2.22	-3.08	-5.88	4.98
D3	Euk. Algae	Larsen et al,. 2013	10.18	0.68	-2.52	-3.42	4.92

ID	Group	Reference	Thr-norm	Ile-norm	Val-norm	Phe-norm	Leu-norm
D4	Euk. Algae	Larsen et al,. 2013	9.02	3.72	-3.08	-3.98	5.68
D5	Euk. Algae	Larsen et al,. 2013	8.56	2.76	-2.34	-2.84	6.14
H1	Euk. Algae	Larsen et al,. 2013	11.44	2.24	-4.36	-2.66	6.66
H2	Euk. Algae	Larsen et al,. 2013	10.86	0.26	-1.64	-5.14	4.34
H3	Euk. Algae	Larsen et al,. 2013	13.18	3.78	-4.72	-2.92	9.32
H4	Euk. Algae	Larsen et al,. 2013	9.5	2.5	-2.5	-5.3	4.2
K1	Euk. Algae	Larsen et al,. 2013	10.32	1.22	-2.18	-2.68	6.68
K2	Euk. Algae	Larsen et al,. 2013	14.36	1.26	-4.04	-4.84	6.74
K3	Euk. Algae	Larsen et al,. 2013	4.52	2.22	-0.38	-3.28	3.08
K4	Euk. Algae	Larsen et al,. 2013	8.46	1.26	-2.64	-3.04	4.04
K5	Euk. Algae	Larsen et al,. 2013	11.3	1.4	-2.7	-4.5	5.5
K6	Euk. Algae	Larsen et al,. 2013	7.14	2.84	-1.66	-4.16	4.16
X1	Euk. Algae	Larsen et al,. 2013	13.9	0.8	-2.7	-4.4	7.6
X2	Euk. Algae	Larsen et al,. 2013	14.14	2.04	-3.16	-5.16	7.86
X3	Euk. Algae	Larsen et al,. 2013	11.1	1.4	-3.2	-4.8	4.5
X4	Euk. Algae	Larsen et al,. 2013	10.92	1.12	-2.18	-3.88	5.98
Y1	Euk. Algae	Larsen et al,. 2013	10.18	2.58	-2.72	-5.12	4.92
N1	Euk. Algae	Larsen et al,. 2013	12.02	2.02	-3.38	-3.98	6.68
N2	Euk. Algae	Larsen et al,. 2013	11.82	2.42	-4.18	-3.78	6.28
N3	Euk. Algae	Larsen et al,. 2013	12.08	2.98	-4.32	-4.12	6.62
C1	Cyano.	Larsen et al,. 2013	12.56	-4.74	-1.04	-1.94	4.84
C2	Cyano.	Larsen et al,. 2013	10.86	1.46	-2.44	-4.24	5.64
C3	Cyano.	Larsen et al,. 2013	11.46	0.06	-3.14	-2.44	5.94
C4	Cyano.	Larsen et al,. 2013	10.72	0.82	-3.38	-3.18	4.98
NA	Cyano.	Jenny Lehman	11.86	2.11	-2.42	-6.40	5.13
NA	Cyano.	Jenny Lehman	17.26	-0.30	-2.84	-7.19	6.93
NA	Cyano.	Jenny Lehman	16.47	0.69	-1.42	-8.91	6.83

Table A.4: Predicted groupings by LDA for test endmembers and DOM samples.

Predicted Group	Sample
Heterotrophic bacteria	Heterotrophic bacteria
Heterotrophic bacteria	Heterotrophic bacteria
Heterotrophic bacteria	Heterotrophic bacteria
Heterotrophic bacteria	Heterotrophic bacteria
Heterotrophic bacteria	Heterotrophic bacteria
Heterotrophic bacteria	Heterotrophic bacteria
Eukaryotic algae	Eukaryotic algae
Eukaryotic algae	Eukaryotic algae
Eukaryotic algae	Eukaryotic algae
Eukaryotic algae	Eukaryotic algae
Cyanobacteria	Eukaryotic algae
Eukaryotic algae	Eukaryotic algae
Cyanobacteria	Cyanobacteria
Eukaryotic algae	Cyanobacteria
Heterotrophic bacteria	BATS HMW DOM 2m 2016
Heterotrophic bacteria	BATS HMW DOM 400m 2016
Heterotrophic bacteria	BATS HMW DOM 850m 2016
Heterotrophic bacteria	BATS HMW DOM 2500m 2016
Eukaryotic algae	BATS LMW DOM 2m 2016
Eukaryotic algae	BATS LMW DOM 400m 2016
Eukaryotic algae	BATS LMW DOM 2500m 2016
Heterotrophic bacteria	HOT HMW DOM 7.5m 2014
Eukaryotic algae	HOT HMW DOM 7.5m 2015
Eukaryotic algae	HOT HMW DOM 400m 2015
Eukaryotic algae	HOT HMW DOM 850m 2015
Cyanobacteria	HOT HMW DOM 2500m 2015
Eukaryotic algae	HOT LMW DOM 7.5m 2014

Predicted Group	Sample
Cyanobacteria	HOT LMW DOM 7.5m 2015
Eukaryotic algae	HOT LMW DOM 400m 2015
Eukaryotic algae	HOT LMW DOM 850m 2015
Eukaryotic algae	HOT LMW DOM 2500m 2015

References

Abelson, P. and Hoering, T. C. (1961). Carbon isotope fractionation in formation of amino acids by photosynthetic organisms. *Proceedings of the National Academy of Sciences*, 47(5).

Aluwihare, L. and Meador, T. (2008). Chemical composition of marine dissolved organic nitrogen. In *Nitrogen in the Marine Environment*, pages 95–140. Academic Press, 2nd edition.

Aluwihare, L. I., Repeta, D. J., and Chen, R. F. (1997). A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature*, 387(6629):166–169.

Aluwihare, L. I., Repeta, D. J., and Pantoja, S. (2005). Two chemically distinct pools of organic nitrogen accumulate in the ocean. *Science*, 308.

Amon, R. M. W. and Benner, R. (1994). Rapid cycling of high-molecular-weight dissolved organic matter in the ocean. *Nature*, 369(6481):549–552.

Amon, R. M. W. and Benner, R. (1996). Bacterial utilization of different size classes of dissolved organic matter. *Limnology and Oceanography*, 41(1):41–51.

Amon, R. M. W., Fitznar, H.-P., and Benner, R. (2001). Linkages among the bioreactivity, chemical composition, and diagenetic state of marine dissolved organic matter. *Limnology and Oceanography*, 46(2):287–297.

Asano, Y. and Lübbehüsen, T. L. (2000). Enzymes acting on peptides containing

D-amino acid. *Journal of Bioscience and Bioengineering*, 89(4):295–306.

Azúa, I., Goiriena, I., Baña, Z., Iriberri, J., and Unanue, M. (2014). Release and consumption of D-amino acids during growth of marine prokaryotes. *Microbial Ecology*, 67(1):1–12.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, 10:257–263.

Bada, J. L. (1971). Kinetics of the nonbiological decomposition and racemization of amino acids in natural waters. In Hem, J. D., editor, *Nonequilibrium Systems in Natural Water Chemistry*, volume 106, pages 309–331. American Chemical Society.

Baltar, F., Alvarez-Salgado, X. A., Arístegui, J., Benner, R., Hansell, D. A., Herndl, G. J., and Lønborg, C. (2021). What is refractory organic matter in the ocean? *Frontiers in Marine Science*, 8.

Bao, H., Niggemann, J., Luo, L., Dittmar, T., and Kao, S.-J. (2017). Aerosols as a source of dissolved black carbon to the ocean. *Nature Communications*, 8(1):510.

Batista, F. C., Ravelo, A. C., Crusius, J., Casso, M. A., and McCarthy, M. D. (2014). Compound specific amino acid δ^{15}N in marine sediments: A new approach for studies of the marine nitrogen cycle. *Geochimica et Cosmochimica Acta*, 142:553–569.

Bauer, J. E., Williams, P. M., and Druffel, E. R. M. (1992). ^{14}C activity of dissolved organic carbon fractions in the north-central Pacific and Sargasso Sea. *Nature*, 357(6380):667–670.

Benner, R. (2002). Chemical composition and reactivity. In *Biogeochemistry of Marine Dissolved Organic Matter*, pages 59–90. Elsevier.

Benner, R. and Amon, R. M. (2015). The size-reactivity continuum of major bioelements in the ocean. *Annual Review of Marine Science*, 7(1):185–205.

Benner, R., Biddanda, B., Black, B., and McCarthy, M. (1997). Abundance, size distribution, and stable carbon and nitrogen isotopic compositions of marine organic matter isolated by tangential-flow ultrafiltration. *Marine Chemistry*, 57(3):243–263.

Benner, R. and Kaiser, K. (2003). Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter. *Limnology and Oceanography*, 48(1):118–128.

Benner, R., Pakulski, J. D., Mccarthy, M., Hedges, J. I., and Hatcher, P. G. (1992). Bulk chemical characteristics of dissolved organic matter in the ocean. *Science*, 255(5051):1561–1564.

Bertani, P., Raya, J., and Bechinger, B. (2014). ^{15}N chemical shift referencing in solid state NMR. *Solid State Nuclear Magnetic Resonance*, 61-62:15–18.

Bourgoin, L.-H. and Tremblay, L. (2010). Bacterial reworking of terrigenous and marine organic matter in estuarine water columns and sediments. *Geochimica et Cosmochimica Acta*, 74(19):5593–5609.

Broek, T. A., Bour, A. L., Ianiri, H. L., Guilderson, T. P., and McCarthy, M. D. (2019). Amino acid enantiomers in old and young dissolved organic matter: Implications for a microbial nitrogen pump. *Geochimica et Cosmochimica Acta*, 247:207–219.

Broek, T. A., Walker, B. D., Guilderson, T. P., and McCarthy, M. D. (2017). Coupled ultrafiltration and solid phase extraction approach for the targeted study of semi-labile high molecular weight and refractory low molecular weight dissolved organic matter. *Marine Chemistry*, 194:146–157.

Broek, T. A. B. (2019). New perspectives on the composition and cycling of

dissolved organic carbon and nitrogen in the ocean.

Broek, T. A. B. and McCarthy, M. D. (2014). A new approach to δ^{15}N compound-specific amino acid trophic position measurements: preparative high pressure liquid chromatography technique for purifying underivatized amino acids for stable isotope analysis. *Limnology and Oceanography: Methods*, 12(12):840–852.

Broek, T. A. B., McCarthy, M. D., Vaughn, J. S., and Guilderson, T. P. Dominant heterocyclic composition of dissolved organic nitrogen in the ocean. *Science*, page 23.

Broek, T. A. B., Walker, B. D., Andreasen, D. H., and McCarthy, M. D. (2013). High-precision measurement of phenylalanine δ^{15}N values for environmental samples: A new approach coupling high-pressure liquid chromatography purification and elemental analyzer isotope ratio mass spectrometry. *Rapid Communications in Mass Spectrometry*, 27(21):2327–2337.

Broek, T. A. B., Walker, B. D., Guilderson, T. P., Vaughn, J. S., Mason, H. E., and McCarthy, M. D. (2020). Low molecular weight dissolved organic carbon: Aging, compositional changes, and selective utilization during global ocean circulation. *Global Biogeochemical Cycles*, 34(6).

Calleja, M., Batista, F., Peacock, M., Kudela, R., and McCarthy, M. (2013). Changes in compound specific δ^{15}N amino acid signatures and D/L ratios in marine dissolved organic matter induced by heterotrophic bacterial reworking. *Marine Chemistry*, 149:32–44.

Cao, X., Mulholland, M. R., Helms, J. R., Bernhardt, P. W., Duan, P., Mao, J., and Schmidt-Rohr, K. (2017). A major step in opening the black box of high-molecular-weight dissolved organic nitrogen by isotopic labeling of Synechococcus and multibond two-dimensional NMR. *Analytical Chemistry*,

89(22):11990–11998.

Cava, F., Lam, H., de Pedro, M. A., and Waldor, M. K. (2011). Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cellular and Molecular Life Sciences*, 68(5):817–831.

Coppola, A. I. and Druffel, E. R. M. (2016). Cycling of black carbon in the ocean. *Geophysical Research Letters*, 43(9):4477–4482.

Coppola, A. I., Druffel, E. R. M., Broek, T. A. B., Haghipour, N., McCarthy, M. D., Eglinton, T. I., and Walker, B. D. (In prep). Variable aging and storage of dissolved black carbon in the north central pacific ocean.

Cowie, G. L. and Hedges, J. I. (1994). Biochemical indicators of diagenetic alteration in natural organic matter mixtures. *Nature*, 369:304–307.

Dauwe, B. and Middelburg, J. J. (1998). Amino acids and hexosamines as indicators of organic matter degradation state in North Sea sediments. *Limnology and Oceanography*, 43(5):782–798.

Dauwe, B., Middelburg, J. J., Herman, P. M. J., and Heip, C. H. R. (1999). Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology and Oceanography*, 44(7):1809–1814.

Davis, J., Kaiser, K., and Benner, R. (2009). Amino acid and amino sugar yields and compositions as indicators of dissolved organic matter diagenesis. *Organic Geochemistry*, 40(3):343–352.

Décima, M., Landry, M. R., Bradley, C. J., and Fogel, M. L. (2017). Alanine δ^{15}N trophic fractionation in heterotrophic protists. *Limnology and Oceanography*, 62(5):2308–2322.

Degens, E. T., Behrendt, M., Gotthardt, B., and Reppmann, E. (1968). Metabolic fractionation of carbon isotopes in marine plankton—II. data on samples collected off the coasts of Peru and Ecuador. *Deep Sea Research and Oceano-*

graphic Abstracts, 15(1):11–20.

Dittmar, T. (2008). The molecular level determination of black carbon in marine dissolved organic matter. *Organic Geochemistry*, 39(4):396–407.

Dittmar, T. (2015). Reasons behind the long-term stability of dissolved organic matter. In *Biogeochemistry of Marine Dissolved Organic Matter*, pages 369–385. 2nd edition.

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

Dittmar, T. and Koch, B. P. (2006). Thermogenic organic matter dissolved in the abyssal ocean. *Marine Chemistry*, 102(3):208–217.

Druffel, E. R. M., Williams, P. M., Bauer, J. E., and Ertel, J. R. (1992). Cycling of dissolved and particulate organic matter in the open ocean. *Journal of Geophysical Research*, 97:15639.

Duan, P. and Schmidt-Rohr, K. (2017). Composite-pulse and partially dipolar dephased multiCP for improved quantitative solid-state ^{13}C NMR. *Journal of Magnetic Resonance*, 285:68–78.

Fang, X., Mao, J., Cory, R. M., McKnight, D. M., and Schmidt-Rohr, K. (2011). ^{15}N and ^{13}C{^{14}N} NMR investigation of the major nitrogen-containing segment in an aquatic fulvic acid: Evidence for a hydantoin derivative. *Magnetic Resonance in Chemistry*, 49(12):775–780.

Fogel, M. L. and Tuross, N. (1999). Transformation of plant biochemicals to geological macromolecules during early diagenesis. *Oecologia*, 120(3):336–346.

Gruber, D. F., Simjouw, J.-P., Seitzinger, S. P., and Taghon, G. L. (2006). Dy-

namics and characterization of refractory dissolved organic matter produced by a pure bacterial culture in an experimental predator-prey system. *Applied and Environmental Microbiology*, 72(6):4184–4191.

Hansell, D. and Carlson, C. (1998). Deep ocean gradients in dissolved organic carbon concentrations. *Nature*, 395:263–266.

Hansell, D. A. and Carlson, C. A. (2001). Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: control by convective overturn. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(8):1649–1667.

Hansman, R. L., Griffin, S., Watson, J. T., Druffel, E. R. M., Ingalls, A. E., Pearson, A., and Aluwihare, L. I. (2009). The radiocarbon signature of microorganisms in the mesopelagic ocean. *Proceedings of the National Academy of Sciences*, 106(16):6513–6518.

Hayes, J. M. (2001). Fractionation of the isotopes of carbon and hydrogen in biosynthetic processes. *Reviews in Mineralogy and Geochemistry*, 43(1):225–277.

Hébert, M. and Tremblay, L. (2017). Production and persistence of bacterial and labile organic matter at the hypoxic water-sediment interface of the St. Lawrence Estuary: Persistence of OM in hypoxic deep waters. *Limnology and Oceanography*, 62(5):2154–2167.

Hedges, J. I. (1992). Global biogeochemical cycles: progress and problems. *Marine Chemistry*, 39(1):67–93.

Hedges, J. I., Eglinton, G., Hatcher, P. G., Kirchman, D. L., Arnosti, C., Derenne, S., Evershed, R. P., Kögel-Knabner, I., de Leeuw, J. W., Littke, R., Michaelis, W., and Rullkötter, J. (2000). The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Organic Geochemistry*, 31(10):945–958.

Hedges, J. I., Hatcher, P. G., Ertel, J. R., and Meyers-Schulte, K. J. (1992). A comparison of dissolved humic substances from seawater with Amazon River counterparts by ^{13}C-NMR spectrometry. *Geochimica et Cosmochimica Acta*, 56(4):1753–1757.

Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., and Hedges, J. I. (2006). Characterization of a major refractory component of marine dissolved organic matter. *Geochimica et Cosmochimica Acta*, 70(12):2990–3010.

Higgins, M. B., Wolfe-Simon, F., Robinson, R. S., Qin, Y., Saito, M. A., and Pearson, A. (2011). Paleoenvironmental implications of taxonomic variation among δ^{15}N values of chloropigments. *Geochimica et Cosmochimica Acta*, 75(22):7351–7363.

Ianiri, H. L. and McCarthy, M. D. (In prep). Examining the long-term preservation of marine dissolved organic nitrogen using compound-specific isotope analysis of amino acids.

Ianiri, H. L., Shen, Y., Broek, T. A. B., and McCarthy, M. D. (Submitted). Distinct bacterial sources and cycling dynamics of HMW and LMW dissolved organic nitrogen in the ocean. *Marine Chemistry.*

Jaffé, R., Ding, Y., Niggemann, J., Vähätalo, A. V., Stubbins, A., Spencer, R. G. M., Campbell, J., and Dittmar, T. (2013). Global charcoal mobilization from soils via dissolution and riverine transport to the oceans. *Science*, 340(6130):345–347.

Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., Kirchman, D. L., Weinbauer, M. G., Luo, T., Chen, F., and Azam, F. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews Microbiology*, 8(8):593–

599.

Jiao, N. and Zheng, Q. (2011). The microbial carbon pump: from genes to ecosystems. *Applied and Environmental Microbiology*, 77(21):7439–7444.

Johnson, R. L. and Schmidt-Rohr, K. (2014). Quantitative solid-state ^{13}C NMR with signal enhancement by multiple cross polarization. *Journal of Magnetic Resonance (San Diego, Calif.: 1997)*, 239:44–49.

Jørgensen, L., Lechtenfeld, O. J., Benner, R., Middelboe, M., and Stedmon, C. A. (2014). Production and transformation of dissolved neutral sugars and amino acids by bacteria in seawater. *Biogeosciences*, 11(19):5349–5363.

Jørgensen, N. O., Stepanaukas, R., Pedersen, A.-G. U., Hansen, M., and Nybroe, O. (2003). Occurrence and degradation of peptidoglycan in aquatic environments. *FEMS Microbiology Ecology*, 46(3):269–280.

Jurado, E., Dachs, J., Duarte, C. M., and Simó, R. (2008). Atmospheric deposition of organic and black carbon to the global oceans. *Atmospheric Environment*, 42(34):7931–7939.

Kaiser, E., Simpson, A. J., Dria, K. J., Sulzberger, B., and Hatcher, P. G. (2003). Solid-state and multidimensional solution-state NMR of solid phase extracted and ultrafiltered riverine dissolved organic matter. *Environmental Science & Technology*, 37(13):2929–2935.

Kaiser, K. and Benner, R. (2000). Determination of amino sugars in environmental samples with high salt content by high-performance anion-exchange chromatography and pulsed amperometric detection. *Analytical Chemistry*, 72(11):2566–2572.

Kaiser, K. and Benner, R. (2005). Hydrolysis-induced racemization of amino acids. *Limnology and Oceanography: Methods*, 3(8):318–325.

Kaiser, K. and Benner, R. (2008). Major bacterial contribution to the ocean reser-

voir of detrital organic carbon and nitrogen. *Limnology and Oceanography*, 53(1):99–112.

Kaiser, K. and Benner, R. (2009). Biochemical composition and size distribution of organic matter at the Pacific and Atlantic time-series stations. *Marine Chemistry*, 113(1):63–77.

Kawasaki, N. and Benner, R. (2006). Bacterial release of dissolved organic matter during cell growth and decline: Molecular origin and composition. *Limnology and Oceanography*, 51(5):2170–2180.

Keil, R. G. and Fogel, M. L. (2001). Reworking of amino acid in marine sediments: Stable carbon isotopic composition of amino acids in sediments along the Washington coast. *Limnology and Oceanography*, 46(1):14–23.

Knapp, A. N., Sigman, D. M., Kustka, A. B., Sañudo-Wilhelmy, S. A., and Capone, D. G. (2012). The distinct nitrogen isotopic compositions of low and high molecular weight marine DON. *Marine Chemistry*, 136-137:24–33.

Knapp, A. N., Sigman, D. M., and Lipschultz, F. (2005). N isotopic composition of dissolved organic nitrogen and nitrate at the Bermuda Atlantic Time-series study site. *Global Biogeochemical Cycles*, 19(1):1–15.

Knapp, A. N., Sigman, D. M., Lipschultz, F., Kustka, A. B., and Capone, D. G. (2011). Interbasin isotopic correspondence between upper-ocean bulk DON and subsurface nitrate and its implications for marine nitrogen cycling. *Global Biogeochemical Cycles*, 25(4).

Knicker, H. (2004). Stabilization of N-compounds in soil and organic-matter-rich sediments—what is the difference? *Marine Chemistry*, 92(1):167–195.

Knicker, H. (2010). "Black nitrogen" – an important fraction in determining the recalcitrance of charcoal. *Organic Geochemistry*, 41(9):947–950.

Knicker, H., Almendros, G., GonzalezVila, F. J., Martin, F., and Ludemann,

H. D. (1996). δ^{13}C and δ^{15}N NMR spectroscopic examination of the transformation of organic nitrogen in plant biomass during thermal treatment. *Soil Biology & Biochemistry*, 28(8):1053–1060.

Knicker, H. and Hatcher, P. G. (1997). Survival of protein in an organic-rich sediment: Possible protection by encapsulation in organic matter. *Naturwissenschaften*, 84(6):231–234.

Knicker, H. and Hatcher, P. G. (2001). Sequestration of organic nitrogen in the sapropel from Mangrove Lake, Bermuda. *Organic Geochemistry*, 32(5):733–744.

Koprivnjak, J. F., Perdue, E. M., and Pfromm, P. H. (2006). Coupling reverse osmosis with electrodialysis to isolate natural organic matter from fresh waters. *Water Research*, 40(18):3385–3392.

Koprivnjak, J. F., Pfromm, P. H., Ingall, E., Vetter, T. A., Schmitt-Kopplin, P., Hertkorn, N., Frommberger, M., Knicker, H., and Perdue, E. M. (2009). Chemical and spectroscopic characterization of marine dissolved organic matter isolated using coupled reverse osmosis–electrodialysis. *Geochimica et Cosmochimica Acta*, 73(14):4215–4231.

Lam, H., Oh, D.-C., Cava, F., Takacs, C. N., Clardy, J., de Pedro, M. A., and Waldor, M. K. (2009). D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science*, 325(5947):1552–1555.

Larsen, T., Taylor, D. L., Leigh, M. B., and O'Brien, D. M. (2009). Stable isotope fingerprinting: a novel method for identifying plant, fungal, or bacterial origins of amino acids. *Ecology*, 90(12):3526–3535.

Larsen, T., Ventura, M., Andersen, N., O'Brien, D. M., Piatkowski, U., and McCarthy, M. D. (2013). Tracing carbon sources through aquatic and terrestrial food webs using amino acid stable isotope fingerprinting. *PLoS ONE*,

8(9):e73441.

Lechtenfeld, O. J., Hertkorn, N., Shen, Y., Witt, M., and Benner, R. (2015). Marine sequestration of carbon in bacterial metabolites. *Nature Communications*, 6(1):6711.

Lechtenfeld, O. J., Kattner, G., Flerus, R., McCallister, S. L., Schmitt-Kopplin, P., and Koch, B. P. (2014). Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochimica et Cosmochimica Acta*, 126:321–337.

Lee, C. and Bada, J. L. (1977). Dissolved amino acids in the equatorial Pacific, the Sargasso Sea, and Biscayne Bay. *Limnology and Oceanography*, 22(3):502–510.

Lee, C. and Cronin, C. (1982). The vertical flux of particulate organic nitrogen in the sea: Decomposition of amino acids in the Peru Upwelling Area and the Equatorial Atlantic. *Journal of Marine Research*, 40:227–251.

Lehman, J. (2009).

Lehmann, M., Carstens, D., Deek, A., McCarthy, M., Schubert, C., and Zopfi, J. (2020). Amino acid and amino sugar compositional changes during in vitro degradation of algal organic matter indicate rapid bacterial re-synthesis. *Geochimica et Cosmochimica Acta*, 283.

Loh, A. N., Bauer, J. E., and Druffel, E. R. M. (2004). Variable ageing and storage of dissolved organic components in the open ocean. *Nature*, 430(7002):877–881.

Macko, S. A., Fogel, M. L., Hare, P. E., and Hoering, T. C. (1987). Isotopic fractionation of nitrogen and carbon in the synthesis of amino acids by microorganisms. *Chemical Geology: Isotope Geoscience section*, 65(1):79–92.

Mao, J., Kong, X., Schmidt-Rohr, K., Pignatello, J. J., and Perdue, E. M. (2012). Advanced solid-state NMR characterization of marine dissolved organic matter isolated using the coupled reverse osmosis/electrodialysis method. *Environmental Science & Technology*, 46(11):5806–5814.

Mao, J. D., Tremblay, L., Gagné, J. P., Kohl, S., Rice, J., and Schmidt-Rohr, K. (2007). Humic acids from particulate organic matter in the Saguenay Fjord and the St. Lawrence Estuary investigated by advanced solid-state NMR. *Geochimica et Cosmochimica Acta*, 71(22):5483–5499.

Massiot, D., Fayon, F., Capron, M., King, I., Calvé, S. L., Alonso, B., Durand, J.-O., Bujoli, B., Gan, Z., and Hoatson, G. (2002). Modelling one- and two-dimensional solid-state NMR spectra. *Magnetic Resonance in Chemistry*, 40(1):70–76.

McCarthy, M. and Bronk, D. (2008). Analytical methods for the study of nitrogen. In *Nitrogen in the Marine Environment*, pages 1219–1275.

McCarthy, M., Hedges, J., and Benner, R. (1996). Major biochemical composition of dissolved high molecular weight organic matter in seawater. *Marine Chemistry*, 55(3):281–297.

McCarthy, M., Pratum, T., Hedges, J., and Benner, R. (1997). Chemical composition of dissolved organic nitrogen in the ocean. *Nature*, 390(6656):150–154.

McCarthy, M. D., Beaupré, S. R., Walker, B. D., Voparil, I., Guilderson, T. P., and Druffel, E. R. M. (2011). Chemosynthetic origin of ^{14}C-depleted dissolved organic matter in a ridge-flank hydrothermal system. *Nature Geoscience*, 4(1):32–36.

McCarthy, M. D. and Benner, R. (1993). The chemical composition of dissolved organic matter in seawater. *Chemical Geology*, (107):503–507.

McCarthy, M. D., Benner, R., Lee, C., and Fogel, M. L. (2007). Amino acid nitro-

gen isotopic fractionation patterns as indicators of heterotrophy in plankton, particulate, and dissolved organic matter. *Geochimica et Cosmochimica Acta*, 71(19):4727–4744.

McCarthy, M. D., Benner, R., Lee, C., Hedges, J. I., and Fogel, M. L. (2004). Amino acid carbon isotopic fractionation patterns in oceanic dissolved organic matter: an unaltered photoautotrophic source for dissolved organic nitrogen in the ocean? *Marine Chemistry*, 92(1):123–134.

McCarthy, M. D., Hedges, J. I., and Benner, R. (1998). Major bacterial contribution to marine dissolved organic nitrogen. *Science*, 281(5374):231–234.

McCarthy, M. D., Lehman, J., and Kudela, R. (2013). Compound-specific amino acid δ^{15}N patterns in marine algae: Tracer potential for cyanobacterial vs. eukaryotic organic nitrogen sources in the ocean. *Geochimica et Cosmochimica Acta*, 103:104–120.

McMahon, K. W., Thorrold, S. R., Houghton, L. A., and Berumen, M. L. (2016). Tracing carbon flow through coral reef food webs using a compound-specific stable isotope approach. *Oecologia*, 180(3):809–821.

Metz, G., Wu, X. L., and Smith, S. O. (1994). Ramped-amplitude cross polarization in magic-angle-spinning NMR. *Journal of Magnetic Resonance, Series A*, 110(2):219–227.

Metz, G., Ziliox, M., and Smith, S. O. (1996). Towards quantitative CP-MAS NMR. *Solid State Nuclear Magnetic Resonance*, 7(3):155–160.

Mopper, K., Stubbins, A., Ritchie, J. D., Bialk, H. M., and Hatcher, P. G. (2007). Advanced instrumental approaches for characterization of marine dissolved organic matter: Extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chemical Reviews*, 107(2):419–442.

Nagata, T., Meon, B., and L. Kirchman, D. (2003). Microbial degradation of

peptidoglycan in seawater. *Limnology and Oceanography*, 48(2):745–754.

Nguyen, R. T. and Harvey, H. (1997). Protein and amino acid cycling during phytoplankton decomposition in oxic and anoxic waters. *Organic Geochemistry*, 27(3):115–128.

Nguyen, R. T. and Harvey, H. R. (2001). Preservation of protein in marine systems: Hydrophobic and other noncovalent associations as major stabilizing forces. *Geochimica et Cosmochimica Acta*, 65(9):1467–1480.

Ogawa, H., Amagai, Y., Koike, I., Kaiser, K., and Benner, R. (2001). Production of refractory dissolved organic matter by bacteria. *Science*, 292(5518):917–920.

Pedrosa-Pàmies, R., Conte, M., Weber, J., and Johnson, R. (2018). Carbon cycling in the Sargasso Sea water column: Insights from lipid biomarkers in suspended particles. *Progress in Oceanography*, 168:248–278.

Pines, A., Gibby, M. G., and Waugh, J. S. (1972). Proton-enhanced nuclear induction spectroscopy. a method for high resolution NMR of dilute spins in solids. *The Journal of Chemical Physics*, 56(4):1776–1777.

Pines, A., Gibby, M. G., and Waugh, J. S. (1973). Proton-enhanced NMR of dilute spins in solids. *The Journal of Chemical Physics*, 59(2):569–590.

Pohlman, J. W., Bauer, J. E., Waite, W. F., Osburn, C. L., and Chapman, N. R. (2011). Methane hydrate-bearing seeps as a source of aged dissolved organic carbon to the oceans. *Nature Geoscience*, 4(1):37–41.

Pèrez, M. T., Pausz, C., and Herndl, G. J. (2003). Major shift in bacterioplankton utilization of enantiomeric amino acids between surface waters and the ocean's interior. *Limnology and Oceanography*, 48(2):755–763.

Radkov, A. D. and Moe, L. A. (2014). Bacterial synthesis of D-amino acids. *Applied Microbiology and Biotechnology*, 98(12):5363–5374.

Repeta, D. J. (2015). Chemical characterization and cycling of dissolved organic matter. In *Biogeochemistry of marine dissolved organic matter*, pages 21–63. Academic Press, 2 edition.

Sannigrahi, P., Ingall, E. D., and Benner, R. (2005). Cycling of dissolved and particulate organic matter at station Aloha: Insights from ^{13}C NMR spectroscopy coupled with elemental, isotopic and molecular analyses. *Deep Sea Research Part I: Oceanographic Research Papers*, 52(8):1429–1444.

Schiff, J. T., Batista, F. C., Sherwood, O. A., Guilderson, T. P., Hill, T. M., Ravelo, A. C., McMahon, K. W., and McCarthy, M. D. (2014). Compound specific amino acid δ^{13}C patterns in a deep-sea proteinaceous coral: Implications for reconstructing detailed δ^{13}C records of exported primary production. *Marine Chemistry*, 166:82–91.

Schleifer, K. H. and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Reviews*, 36(4):407–477.

Schmidt-Rohr, K. and Mao, J.-D. (2002). Selective observation of nitrogen-bonded carbons in solid-state NMR by saturation-pulse induced dipolar exchange with recoupling. *Chemical Physics Letters*, 359(5):403–411.

Scott, J. H., O'Brien, D. M., Emerson, D., Sun, H., McDonald, G. D., Salgado, A., and Fogel, M. L. (2006). An examination of the carbon isotope effects associated with amino acid biosynthesis. *Astrobiology*, 6(6):867–880.

Shen, Y., Benner, R., Murray, A. E., Gimpel, C., Greg Mitchell, B., Weiss, E. L., and Reiss, C. (2017). Bioavailable dissolved organic matter and biological hot spots during austral winter in Antarctic waters. *Journal of Geophysical Research: Oceans*, 122(1):508–520.

Shen, Y., Guilderson, T. P., Sherwood, O. A., Castro, C. G., Chavez, F. P., and McCarthy, M. D. (2021). Amino acid δ^{13}C and δ^{15}N patterns from sediment

trap time series and deep-sea corals: Implications for biogeochemical and ecological reconstructions in paleoarchives. *Geochimica et Cosmochimica Acta*, page S0016703720307250.

Silfer, J. A., Engel, M. H., Macko, S. A., and Jumeau, E. J. (1991). Stable carbon isotope analysis of amino acid enantiomers by conventional isotope ratio mass spectrometry and combined gas chromatography/isotope ratio mass spectrometry. *Analytical Chemistry*, 63(4):370–374.

Sipler, R. and Bronk, D. (2015). Dynamics of dissolved organic nitrogen. In *Biogeochemistry of Marine Dissolved Organic Matter*. Elsevier Academic Press, 2nd edition.

Skog, K., Johansson, M., and Jägerstad, M. (1998). Carcinogenic heterocyclic amines in model systems and cooked foods: A review on formation, occurrence and intake. *Food and Chemical Toxicology*, 36(9):879–896.

Smernik, R. J. and Baldock, J. A. (2005). Does solid-state ^{15}N NMR spectroscopy detect all soil organic nitrogen? *Biogeochemistry*, 75(3):507–528.

Smernik, R. J. and Oades, J. M. (2000a). The use of spin counting for determining quantitation in solid state ^{13}C NMR spectra of natural organic matter: 1. model systems and the effects of paramagnetic impurities. *Geoderma*, 96(1):101–129.

Smernik, R. J. and Oades, J. M. (2000b). The use of spin counting for determining quantitation in solid state ^{13}C NMR spectra of natural organic matter: 2. HF-treated soil fractions. *Geoderma*, 96(3):159–171.

Srivastava, M., Dzikovski, B., and Freed, J. H. (2021). Extraction of weak spectroscopic signals with high fidelity: Examples from ESR. *The Journal of Physical Chemistry A*, 125(20):4480–4487.

Takano, Y., Kashiyama, Y., Ogawa, N. O., Chikaraishi, Y., and Ohkouchi, N.

(2010). Isolation and desalting with cation-exchange chromatography for compound-specific nitrogen isotope analysis of amino acids: application to biogeochemical samples. *Rapid Communications in Mass Spectrometry*, 24(16):2317–2323.

Team, R. C. (2021). R: A language and environment for statistical computing.

Tremblay, L. and Benner, R. (2006). Microbial contributions to N-immobilization and organic matter preservation in decaying plant detritus. *Geochimica et Cosmochimica Acta*, 70(1):133–146.

Tremblay, L. and Benner, R. (2009). Organic matter diagenesis and bacterial contributions to detrital carbon and nitrogen in the amazon river system. *Limnology and Oceanography*, 54(3):681–691.

Ueda, K., Morgan, S. L., Fox, A., Gilbart, J., Sonesson, A., Larsson, L., and Odham, G. (1989). D-Alanine as a chemical marker for the determination of streptococcal cell wall levels in mammalian tissues by gas chromatography/negative ion chemical ionization mass spectrometry. *Analytical Chemistry*, 61(3):265–270.

Vokhshoori, N., Larsen, T., and McCarthy, M. (2014). Reconstructing δ^{13}C isoscapes of phytoplankton production in a coastal upwelling system with amino acid isotope values of littoral mussels. *Marine Ecology Progress Series*, 504:59–72.

Walker, B. D., Beaupré, S., Guilderson, T., Druffel, E., and McCarthy, M. (2011). Large-volume ultrafiltration for the study of radiocarbon signatures and size vs. age relationships in marine dissolved organic matter. *Geochimica et Cosmochimica Acta*, 75(18):5187–5202.

Walker, B. D., Beaupré, S. R., Guilderson, T. P., McCarthy, M. D., and Druffel, E. R. M. (2016a). Pacific carbon cycling constrained by organic matter size,

age and composition relationships. *Nature Geoscience*, 9(12):888–891.

Walker, B. D., Guilderson, T., Okimura, K., Peacock, M., and McCarthy, M. (2014-02). Radiocarbon signatures and size–age–composition relationships of major organic matter pools within a unique California upwelling system. *Geochimica et Cosmochimica Acta*, 126:1–17.

Walker, B. D., Primeau, F., Beaupré, S. R., Guilderson, T. P., Druffel, E. R. M., and McCarthy, M. D. (2016b). Linked changes in marine dissolved organic carbon molecular size and radiocarbon age. *Geophysical Research Letters*, 43(19):10,385–10,393.

Wang, R., Zhang, Z., Sun, J., and Jiao, N. (2020). Differences in bioavailability of canonical and non-canonical D-amino acids for marine microbes. *Science of The Total Environment*, 733:139216.

Wang, X.-C. and Druffel, E. R. M. (2001). Radiocarbon and stable carbon isotope compositions of organic compound classes in sediments from the NE Pacific and Southern Oceans. *Marine Chemistry*, 73(1):65–81.

Williams, P. M. (1968). Stable carbon isotopes in the dissolved organic matter of the sea. *Nature*, 219(5150):152–153.

Williams, P. M. and Druffel, E. R. M. (1987). Radiocarbon in dissolved organic matter in the central North Pacific Ocean. *Nature*, 330(6145):246–248.

Williams, P. M. and Gordon, L. I. (1970). Carbon-13: carbon-12 ratios in dissolved and particulate organic matter in the sea. *Deep Sea Research and Oceanographic Abstracts*, 17(1):19–27.

Witanowski, M., Stefaniak, L., and Webb, G. A. (1993). Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy*, page 9.

Yamaguchi, Y. T., Chikaraishi, Y., Takano, Y., Ogawa, N. O., Imachi, H., Yokoyama, Y., and Ohkouchi, N. (2017). Fractionation of nitrogen isotopes

during amino acid metabolism in heterotrophic and chemolithoautotrophic microbes across Eukarya, Bacteria, and Archaea: Effects of nitrogen sources and metabolic pathways. *Organic Geochemistry*, 111:101–112.

Yamaguchi, Y. T. and McCarthy, M. D. (2018). Sources and transformation of dissolved and particulate organic nitrogen in the North Pacific Subtropical Gyre indicated by compound-specific δ^{15}N analysis of amino acids. *Geochimica et Cosmochimica Acta*, 220:329–347.

Yamashita, Y. and Tanoue, E. (2008). Production of bio-refractory fluorescent dissolved organic matter in the ocean interior. *Nature Geoscience*, 1(9):579–582.

Zhang, Z., Zheng, Q., and Jiao, N. (2016). Microbial D-amino acids and marine carbon storage. *Science China Earth Sciences*, 59(1):17–24.

Zhao, Z., Gonsior, M., Luek, J., Timko, S., Ianiri, H., Hertkorn, N., Schmitt-Kopplin, P., Fang, X., Zeng, Q., Jiao, N., and Chen, F. (2017). Picocyanobacteria and deep-ocean fluorescent dissolved organic matter share similar optical properties. *Nature Communications*, 8(1):15284.

Ziegler, S. E. and Fogel, M. L. (2003). Seasonal and diel relationships between the isotopic compositions of dissolved and particulate organic matter in freshwater ecosystems. *Biogeochemistry*, 64:25–52.