Linking amino acid sequences to sources and fates of organic matter in aquatic systems

Megan Duffy

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Richard Keil, Chair

Jeffrey Richey

Anitra Ingalls

Program Authorized to Offer Degree:

Oceanography

University of Washington

Abstract

Linking amino acid sequences to sources and fates of organic matter in aquatic systems

Megan Duffy

Chair of the Supervisory Committee:
Dr. Richard Keil
Oceanography

Proteins enact life's intent. These macromolecules, directed by genes and informed by the environment, are the engines that power the cells of all biological entities on Earth. From viruses and bacteria to humans and blue whales, proteins serve a vast range of metabolic, transport, communication, and structural purposes. Along with geological and chemical drivers, protein-driven biological processes modulate planetary cycles of carbon, oxygen, and nitrogen. By unlocking the information stored in peptide and protein sequences, we can learn the biological origins and functions of cells within a community. For organic geochemists, there is useful information here as well: proteins make up a large proportion of organic carbon and nitrogen in aquatic systems. Thus, the cycling and degradation dynamic of proteins is of great importance when thinking about global organic carbon preservation and sequestration. The dissertation I present here addresses the cycling of proteins and protein-derived organic matter in both laboratory settings and environmental systems. Chapter 2, my M.S. project, describes the usefulness of integrating a different kind of peptide sequencing, *de novo* sequencing, into the traditional

environmental proteomics workflow to access degraded and unanticipated sequences. This is described in “Protein cycling in the eastern tropical North Pacific oxygen deficient zone: a *de novo*-assisted peptidomic approach”, (Duffy et al., *in press*). The three subsequent projects utilize this technique to ask questions about proteins and peptide cycling in complex environmental systems. Chapter 3 uses the *de novo*-assisted peptidomic technique to follow the peptides of a diatom through a simulated bloom and subsequent degradation by a natural microbial community (Duffy et al., *in press*). Chapter 4 is a comparison of organic matter flux across time and space in an ODZ. Chapter 5 moves out of the purely marine realm and bridges the span between terrestrial and marine systems in the Amazon River tidal reach, probing how proteinaceous organic matter is transformed in the final stretch of the Amazon River mouth before it empties into the Atlantic and characterizing the community composition and function of microbes working to alter its reactivity and character.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not exist without the network of mentors, students, friends, and family who supported me during all stages of graduate school.

I am very grateful to my advisor, Rick Keil, who not only pushed me in new directions to learn about the ocean in a multiplicity of ways, but supported me as a whole person. I always knew my happiness and wellbeing would be prioritized by you above all. You modeled the type of care and compassion I wish to cultivate in myself as a mentor. Thank you.

I have had wonderful committee members to guide me through what turned out to be a dissertation that spanned multiple subdisciplines and topics. Thank you, Jeff, for inviting me to the Amazon and supporting me in my foray into river geochemistry – it's been a blast. Thank you, Anitra, for your insightful questions during committee meetings and in my Chapter 2 manuscript, as well as for modeling excellent pedagogy when I was your TA. Thank you, Gabrielle, for your excellent writing and laboratory support, and always useful and detailed questions and advice. Thanks, too, to Sarah for serving as my GSR and for consistent support throughout this process.

I was lucky enough to have an 'additional' committee of mentors who were invaluable for the conception, execution, and writing of all chapters of this dissertation. Rachel Lundeen was a terrific mentor in proteomics data analysis and interpretation. Larry Mayer helped me a fantastic amount to understand and write about protein degradation and geochemistry. Al Devol was helpful in interpreting rate data. Nick Ward was an excellent mentor both formally at PNNL and informally

in interpreting Amazon River proteomics. Finally, Clara Fuchsman has been an unwavering source of ideas, data, and encouragement during my entire time at UW. Thank you, all.

Members of the greater Keil group share in this achievement, and none of these chapters could have been happened without them. Thank you to Jaqui Neibauer for your help and expertise in all things, for teaching me about electronics, proteomics, and mass spectrometry, and for being invariably positive and supportive. I am honored to have worked and learned alongside a great group of undergraduates and post-bacs, namely Khadijah Homolka, Jamee Adams, Emmet Bush, Anna Boyar, Garrett Raehild, Marlena Wied, and Rebecca Gould. Working with you was a highlight of my time at UW.

My fellow graduate students and post-docs in UW Oceanography were a constant source of encouragement, diversion, and assistance. I learned the most from you. Thank you to my cohort – Andrew, Isaiah, Rosalind, Christina, Charles, Sasha, Hilary, Caroline, Katie, and Dylan – I've so valued moving through the years with you. To the geochemists of the 5th floor – Angie, Katherine, Ashley, and Jiwoon – thank for the ideas, opera outings, gecko-sitting, bike rides and backpacking trips.

Friends in Washington, Oregon, Montana, Vermont and beyond have been rooting for me from the very beginning. Talking with you has helped me to understand my science and its purpose better, and my work is the better for it. Thanks to my family for helping me to make clear-headed decisions and to carry on when I was unable to see the joy in my work.

Finally, thank you to George Kiely, my partner in life. Your contributions to this dissertation span the logistical (when I was caught trying to write while living off-grid during a pandemic) to the fundamental, when I've questioned the whole point. It was worth it.

Chapter 1. INTRODUCTION

1.1 PROTEIN IN ENVIRONMENTAL AQUATIC SYSTEMS

Organic matter originates in the sunlit surface ocean where photosynthetic organisms fix carbon from the atmosphere. Some of this organic matter becomes detrital and sinks, forming a critical component of the biological pump (Sanders et al., 2014; Sigman & Boyle, 2000). This process removes carbon and nitrogen from the surface ocean and fuels a cascade of heterotrophic and other metabolic lifestyles throughout the water column and sediments. Only about 0.5% of gross primary production in the ocean escapes remineralization, but over geologic time this sink has important implications for global carbon and oxygen cycling (Burdige, 2007). A significant portion of the organic matter produced in the ocean surface is composed of proteins – macromolecules composed of nitrogen-rich amino acids. On average, proteins make up 60% of phytoplankton biomass (Finkel et al., 2016) and proteinaceous material (i.e., amino acid-containing) is the largest identifiable component of sinking organic matter (Wakeham et al., 1997).

Most of the organic matter in the ocean is detrital (Kawasaki et al., 2011). While the fate of most nitrogen-rich organic matter is to ultimately be remineralized back into inorganic form, the processes by which this occurs generally remain unclear for protein. Proteins are presumably labile and subject to degradation by numerous microbial peptidases (Griffith & Fletcher, 1991; Mulholland & Lee, 2009; Nunn et al., 2003). However, intact protein has been identified as dissolved in seawater (Yamada & Tanoue, 2003, 2003). Aging experiments (Keil & Kirchman, 1993) and the abundance of amide-like dissolved organic nitrogen in seawater (Aluwihare et al., 1997; McCarthy et al., 1997) further signal that proteins may be preserved over long periods of time in modified forms (e.g., Keil & Kirchman, 1994; Knicker et al., 1996). Indeed, recent work reveals compounds presumably derived from the modification of proteins persisting in diverse diagenetically relevant environments including sediments (Estes et al., 2019) and sediment pore water (Abdulla et al., 2018; Schmidt et al., 2011). Understanding the origins, molecular character, and derivation of preserved proteinaceous material could significantly untangle the protein recalcitrance paradox as well as help predict the remineralization capacity of a warmer, less oxygenated ocean of the near future.

1.2 LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY BASED PROTEOMICS

Understanding the origins and degradation patterns of detrital proteinaceous material is fraught with challenges: simply isolating and characterizing diverse proteins in complicated matrices is methodologically difficult (Moore, Nunn, Faux, et al., 2012; Nunn & Keil, 2006; Timmins-Schiffman et al., 2017). A further complication is the intermixing of living biomass (heterotrophic and chemoautotrophic bacteria and heterotrophic protists, fungi, and animals) with detrital material from which it is difficult to methodologically distinguish, and living biomass is likely to be present in much lower quantities (Kawasaki et al., 2011). Most dissolved organic matter (DOM) and POM protein determination has thus relied on hydrolyzing all the peptide bonds of proteinaceous material to constituent amino acids. Such investigations identify patterns in bulk amino acid composition and quantity in sediments and particles (Cowie & Hedges, 1992; Dauwe et al., 1999; Dauwe & Middelburg, 1998; Lee et al., 2000; Salas et al., 2018). Yet with such an approach, characterization of the original proteins or peptides is lost along with any taxonomic, structural, or functional information.

Developments in high resolution mass spectrometry and computing have facilitated metaproteomic investigations of proteins in suspended (Bergauer et al., 2017; Bridoux et al., 2015; Dong et al., 2010; Mikan et al., 2020) and sinking (Moore et al., 2012) organic matter. Metaproteomic tools could help resolve questions about protein and peptide degradation in the ocean, but they are currently limited in that they are designed to identify proteins from living biomass. By far the most common strategy for protein identification in metaproteomics is peptide-spectrum matching (PSM) (Muth et al., 2015; Saito et al., 2019). Observed peptide mass spectra are matched to theoretical spectra predicted from a protein reference database. When employed using an accurate protein database that is neither too small nor bloated with excess proteins, the PSM approach is the most sensitive and accurate protein identification method (Muth et al., 2015). Typically, such databases are derived either from complementary metagenomic or transcriptomic analyses, a compilation of annotated reference proteomes of taxa thought likely to be represented in the sample, or a combination of these two. When evaluating unknown environmental samples within a complex environmental matrix, this approach has three significant limitations: a) the taxonomically diverse and uncharacterized nature of environmental samples, b), the fact that proteins can deviate from gene product predictions due to alternative splicing, mutations,

polymorphisms, or post-translational modifications (PTM), and c) degradation, which can alter detrital protein's suitability for discovery due to mass shifts and/or the generation of small fragment peptides no longer diagnostic of a protein source. Indeed, it is typical that far fewer than half of all tandem mass spectra acquired in shotgun proteomics experiments are successfully matched to a peptide from a database, indicating that much potentially useful information is lost (Chick et al., 2015).

1.3 PROTEIN DEGRADATION CONTINUUM ACROSS AND WITHIN AQUATIC SYSTEMS

I explore a wide range of aquatic systems in this thesis, where organic carbon and nitrogen exist in variable and coexisting states of degradation and association. There is a large and deep literature on protein and amino acid degradation, preservation, and mineral interactions, but to varying degrees in the systems I explore here. These include a simulated algal bloom (Chapter 3), sinking particles in axonic seawater (Chapter 2), and fine particulate matter in a sediment-laden river (Chapter 5). I also explore the consequences of microbial processing of organic matter by measuring rates of N_2 production from sinking particles in an ODZ (Chapter 4) and the degradation rates of different size classes of particles in the lower Amazon River tidal reach (Chapter 5). While in Chapter 5 I connect the geochemical signatures of peptides (amino acid composition, post-translational modifications, and intracellular localization) to bulk organic matter degradation, I will connect these same metrics to ODZ metabolic rates in future work. That said, the groundwork for peptidomic comparisons of suspended and sinking particles in the ODZ has been laid, as is demonstrated in Chapter 2.

1.4 REFERENCES

Abdulla, H. A., Burdige, D. J., & Komada, T. (2018). Accumulation of deaminated peptides in anoxic sediments of Santa Barbara Basin. *Geochimica et Cosmochimica Acta*, *223*, 245–258. https://doi.org/10.1016/j.gca.2017.11.021

Aluwihare, L. I., Repeta, D. J., & Chen, R. F. (1997). A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature*, *387*(6629), 166–169. https://doi.org/10.1038/387166a0

Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., Jensen, O. N., & Herndl, G. J. (2017). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences*, 201708779. https://doi.org/10.1073/pnas.1708779115

Bridoux, M., Neibauer, J., Ingalls, A., Nunn, B., & Keil, R. (2015). Suspended marine particulate proteins in coastal and oligotrophic waters. *Journal of Marine Systems*, *143*. https://doi.org/10.1016/j.jmarsys.2014.10.014

Burdige, D. J. (2007). Preservation of Organic Matter in Marine Sediments: Controls, Mechanisms, and an Imbalance in Sediment Organic Carbon Budgets? *Chemical Reviews*, *107*(2), 467–485. https://doi.org/10.1021/cr050347q

Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., & Gygi, S. P. (2015). A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nature Biotechnology*, *33*(7), 743–749. https://doi.org/10.1038/nbt.3267

Cowie, G. L., & Hedges, J. I. (1992). Sources and reactivities of amino acids in a coastal marine environment. *Limnology and Oceanography*, *37*(4), 703–724. https://doi.org/10.4319/lo.1992.37.4.0703

Dauwe, B., & 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. https://doi.org/10.4319/lo.1998.43.5.0782

Dauwe, B., Middelburg, J. J., Herman, P. M. J., & Heip, C. H. R. (1999). Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology and Oceanography*, *44*(7), 1809–1814. https://doi.org/10.4319/lo.1999.44.7.1809

Dong, H.-P., Wang, D.-Z., Dai, M., & Hong, H.-S. (2010). Characterization of particulate organic matter in the water column of the South China Sea using a shotgun proteomic approach. *Limnology and Oceanography*, *55*(4), 1565–1578. https://doi.org/10.4319/lo.2010.55.4.1565

Estes, E. R., Pockalny, R., D'Hondt, S., Inagaki, F., Morono, Y., Murray, R. W., Nordlund, D., Spivack, A. J., Wankel, S. D., Xiao, N., & Hansel, C. M. (2019). Persistent organic matter in oxic subseafloor sediment. *Nature Geoscience*, *12*(2), 126–131. https://doi.org/10.1038/s41561-018-0291-5

Finkel, Z. V., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I., & Irwin, A. J. (2016). Phylogenetic Diversity in the Macromolecular Composition of Microalgae. *PLoS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0155977

Griffith, P. C., & Fletcher, M. (1991). Hydrolysis of Protein and Model Dipeptide Substrates by Attached and Nonattached Marine Pseudomonas sp. Strain NCIMB 2021. *Applied and Environmental Microbiology*, *57*(8), 2186–2191.

Kawasaki, N., Sohrin, R., Ogawa, H., Nagata, T., & Benner, R. (2011). Bacterial carbon content and the living and detrital bacterial contributions to suspended particulate organic carbon in the North Pacific Ocean. *Aquatic Microbial Ecology*, *62*, 165–176. https://doi.org/10.3354/ame01462

Keil, R. G., & Kirchman, D. L. (1993). Dissolved combined amino acids: Chemical form and utilization by marine bacteria. *Limnology and Oceanography*, *38*(6), 1256–1270. https://doi.org/10.4319/lo.1993.38.6.1256

Keil, R. G., & Kirchman, D. L. (1994). Abiotic transformation of labile protein to refractory protein in sea water. *Marine Chemistry*, *45*(3), 187–196. https://doi.org/10.1016/0304-4203(94)90002-7

Knicker, H., Scaroni, A. W., & Hatcher, P. G. (1996). 13C and 15N NMR spectroscopic investigation on the formation of fossil algal residues. *Organic Geochemistry*, *24*(6), 661–669. https://doi.org/10.1016/0146-6380(96)00057-5

Lee, C., Wakeham, S. G., & I. Hedges, J. (2000). Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(8), 1535–1568. https://doi.org/10.1016/S0967-0637(99)00116-8

McCarthy, M., Pratum, T., Hedges, J., & Benner, R. (1997). Chemical composition of dissolved organic nitrogen in the ocean. *Nature*, *390*(6656), 150–154. https://doi.org/10.1038/36535

Mikan, M. P., Harvey, H. R., Timmins-Schiffman, E., Riffle, M., May, D. H., Salter, I., Noble, W. S., & Nunn, B. L. (2020). Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. *The ISME Journal*, *14*(1), 39–52. https://doi.org/10.1038/s41396-019-0503-z

Moore, E. K., Nunn, B. L., Faux, J. F., Goodlett, D. R., & Harvey, H. R. (2012). Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments. *Limnology and Oceanography: Methods*, *10*(5), 353–366. https://doi.org/10.4319/lom.2012.10.353

Moore, E. K., Nunn, B. L., Goodlett, D. R., & Harvey, H. R. (2012). Identifying and tracking proteins through the marine water column: Insights into the inputs and preservation mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta*, *83*, 324–359. https://doi.org/10.1016/j.gca.2012.01.002

Mulholland, M., & Lee, C. (2009). Peptide Hydrolysis and the Uptake of Dipeptides by Phytoplankton. *Limnology and Oceanography*, *54*(4). https://doi.org/10.4319/lo.2009.54.3.0856

Muth, T., Kolmeder, C. A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F. J., Rensen, S. S., Reichl, U., de Vos, W. M., Rapp, E., & Martens, L. (2015). Navigating through metaproteomics data: A logbook of database searching. *Proteomics*, *15*(20), 3439–3453. https://doi.org/10.1002/pmic.201400560

Nunn, B. L., & Keil, R. G. (2006). A comparison of non-hydrolytic methods for extracting amino acids and proteins from coastal marine sediments. *Marine Chemistry*, *98*(1), 31–42. https://doi.org/10.1016/j.marchem.2005.06.005

Nunn, B. L., Norbeck, A., & Keil, R. G. (2003). Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. *Marine Chemistry*, *83*(1–2), 59–73. https://doi.org/10.1016/S0304-4203(03)00096-3

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M. R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd, A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome Research*, *18*(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761

Salas, P. M., Sujatha, C. H., Ratheesh Kumar, C. S., & Cheriyan, E. (2018). Amino acids as indicators to elucidate organic matter degradation profile in the Cochin estuarine sediments, Southwest coast of India. *Marine Pollution Bulletin*, *127*, 273–284. https://doi.org/10.1016/j.marpolbul.2017.12.010

Sanders, R., Henson, S. A., Koski, M., De La Rocha, C. L., Painter, S. C., Poulton, A. J., Riley, J., Salihoglu, B., Visser, A., Yool, A., Bellerby, R., & Martin, A. P. (2014). The Biological Carbon Pump in the North Atlantic. *Progress in Oceanography*, *129*, 200–218. https://doi.org/10.1016/j.pocean.2014.05.005

Schmidt, F., Koch, B. P., Elvert, M., Schmidt, G., Witt, M., & Hinrichs, K.-U. (2011). Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under Contrasting Sedimentary Redox Conditions in the Black Sea. *Environmental Science & Technology*, *45*(12), 5223–5229. https://doi.org/10.1021/es2003414

Sigman, D. M., & Boyle, E. A. (2000). Glacial/interglacial variations in atmospheric carbon dioxide. *Nature*, *407*(6806), 859–869. https://doi.org/10.1038/35038000

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W. S., & Nunn, B. L. (2017). Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *The ISME Journal*, *11*(2), 309–314. https://doi.org/10.1038/ismej.2016.132

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., & Peterson, M. J. (1997). Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta*, *61*(24), 5363–5369. https://doi.org/10.1016/S0016-7037(97)00312-8

Yamada, N., & Tanoue, E. (2003). Detection and partial characterization of dissolved glycoproteins in oceanic waters. *Limnology and Oceanography*, *48*(3), 1037–1048. https://doi.org/10.4319/lo.2003.48.3.1037

Chapter 2. *DE NOVO*-ASSISTED PEPTIDOMICS HELPS IN THE STUDY OF MARINE CARBON FLUX AND PROTEIN DEGRADATION[1]

ABSTRACT

Peptides are identified using a *de novo*-discovery approach in suspended and sinking particles from the eastern tropical North Pacific oxygen deficient zone (ODZ) and in a culture of a dominant autotroph from the region, the cyanobacterium *Prochlorococcus*. The benchmarking experiment with *Prochlorococcus* shows *de novo* peptides to be taxonomically specific, and thus of value in augmenting database-driven approaches. Analysis of the suspended and sinking particles using the *de novo*-discovery approach reveals the presence of fungal proteins in deep sinking particles that were not in the original search database, contributing to growing recognition that fungi may play important roles in marine organic matter cycling. Cyanobacterial peptides that have been post-translationally modified were tracked to depth, where they contribute ~1% of the phylum-level identifiable peptide pool in the sediment trap sample. Many peptides found at depth in the detrital pool are associated with membranes, indicating that cellular location is associated with early preservation within the detrital pool. Modified amino acids in sinking and suspended particles include high levels of deamidation, suggesting that partial extracellular degradation of protein could fuel observed anammox and contribute to pools of refractory organic nitrogen.

---

[1] The research contained in this chapter is currently *in press* in Limnology and Oceanography as '*Protein cycling in the eastern tropical North Pacific oxygen deficient zone: a de novo-discovery peptidomic approach*', with authors Megan E. Duffy[1], Jacquelyn A. Neibauer[1], Jamee Adams[1,2], Rachel A. Lundeen[1], Gabrielle Rocap[1], Anitra E. Ingalls[1], Clara A. Fuchsman[1,3] and Richard G. Keil[1]

[1]School of Oceanography, University of Washington, Seattle, WA, 98105, USA, [2]Scripps Institution of Oceanography, La Jolla, 92037, CA, [3]Horn Point Laboratory, University of Maryland, Cambridge, 21613, MD

2.1 INTRODUCTION

The biological pump removes carbon and nitrogen from the surface ocean, but only about 0.5% of gross primary production in the ocean escapes remineralization. None-the-less, over geologic time this sink has important implications for global carbon and oxygen cycling (Burdige, 2007). A significant portion of the organic matter produced in the ocean surface is proteinaceous; on average, proteins make up 60% of phytoplankton biomass (Finkel et al., 2016) and amino acid-containing material is the largest identifiable component of detrital sinking organic matter (Wakeham et al., 1997). If protein-derived material is assumed to represent 25% of the detrital organic matter in the ocean (Wakeham et al 1997), then the ocean contains roughly 7.5 petagrams of detrital protein.

The eastern tropical North Pacific (ETNP) is home to the Earth's largest marine oxygen deficient zone (ODZ), accounting for approximately 41% of global marine anoxic waters (Paulmier & Ruiz-Pino, 2009). Particulate organic matter (POM) fluxes through the ODZ are important in controlling N_2 production from denitrification and anammox (Fuchsman et al., 2019) and it is likely that the makeup of organic matter in the detritus controls the relative contribution of those processes (Babbin et al., 2014). Since proteinaceous matter comprises the single largest identifiable component of the sinking flux in the tropical eastern Pacific (Wakeham et al., 1997), a proteomic approach has the potential to provide powerful insight into POM dynamics and carbon and nitrogen cycling within this ODZ.

The degradation of detrital protein and peptides in the ocean remains somewhat enigmatic. Proteins are presumably labile and subject to degradation by numerous microbial peptidases (Nunn et al., 2003). However, intact detrital protein has been identified in seawater (Tanoue et al., 1995; Yamada & Tanoue, 2003), and the high abundance of amide-like organic nitrogen in seawater detritus (McCarthy et al., 1997) signals that some proteins may be preserved over long periods of time in the ocean. Indeed, compounds presumably derived from proteins persist in diverse diagenetically relevant environments including sediments (Estes et al., 2019) and sediment pore water (Abdulla et al., 2018; Schmidt et al., 2011). Thus, understanding the origins, molecular character, and preservation of proteinaceous material could significantly untangle the protein recalcitrance paradox as well as help predict the remineralization capacity of a warmer, less oxygenated ocean of the near future (Keil, 2017).

Understanding the sources and fate of detrital proteinaceous material is fraught with challenges: simply isolating and characterizing diverse proteins in complicated matrices is methodologically difficult (Moore, Nunn, Faux, et al., 2012; Nunn & Keil, 2006). A further complication is the intermixing of living biomass with detrital material from which it is difficult to methodologically distinguish, with living biomass likely to be present in much lower quantities in marine POM (Kawasaki et al., 2011). Most environmental protein determination has previously relied on hydrolyzing all the peptide bonds releasing amino acids. Such investigations identify patterns in bulk amino acid composition and quantity in sediments and particles (Dauwe et al., 1999; Lee et al., 2004; Salas et al., 2018). Yet with such an approach, characterization of the original proteins or peptides is lost along with any taxonomic, structural, or functional information. Metaproteomics can help resolve questions about protein and peptide degradation (Bergauer et al., 2017; Bridoux et al., 2015; Moore et al., 2012), but the tools are currently designed to identify proteins from living biomass, not to investigate degradation of detrital remains.

The most common strategy for protein identification in metaproteomics is peptide-spectrum matching (PSM) (Saito et al., 2019), where peptide mass spectra are matched to theoretical spectra predicted from a protein reference database. This powerful approach has several limitations including that if the database is insufficient, peptides from environmental samples will not be identified. An alternative approach is *de novo* peptide sequencing, a database-independent tool that uses the same tandem mass spectral information. Rather than matching peptides to a protein database, *de novo* approaches employ first principles-based algorithms to establish sequences directly, making them complementary to the database-driven approach (O'Bryon et al., 2020). *De novo* sequencing and database-dependent PSM have complementary advantages and disadvantages: *de novo* does not require the investigator to know what to look for from the outset, but accuracy and confidence are more difficult to assess for longer peptides. Database-dependent PSM has inherent bias but greater accuracy and sensitivity. The technique is thus adept at identifying longer peptides which provide more information about the exact source and function of proteins in a sample.

In this study I used a combination of *de novo* and database tools (PeaksDB), hereafter called the *de novo*-discovery approach (Figure 2.1). In this workflow, mass spectra are input to the database search tool Comet (Eng et al; 2013), the *de novo*-directed database search program PeaksDB which uses a combination of *de novo* and database tools (J. Zhang et al. 2012), and a

stand-alone *de novo*-only algorithm (Peaks *de novo,* Ma et al. 2003). *De novo*-only candidate sequences are input into the protein mapping tools PepExplorer and Unipept to match them to experiment databases (PepExplorer) or expansive reference databases (UniPept). A selection of *de novo*-discovered proteins is then added to the original search database and the spectra are queried again using PeaksDB.

I first evaluate the *de novo*-discovery methodology (Figure 2.1) by comparing the performance of a *de novo* algorithm, Peaks (Ma et al., 2003), to that of a widely-used database search tool (Comet; Eng et al., 2013). The control is a culture of the cyanobacteria *Prochlorococcus marinus*, the dominant autotroph in the ETNP ODZ. I then apply the *de novo*-discovery approach to suspended and sinking particle samples from the ETNP.

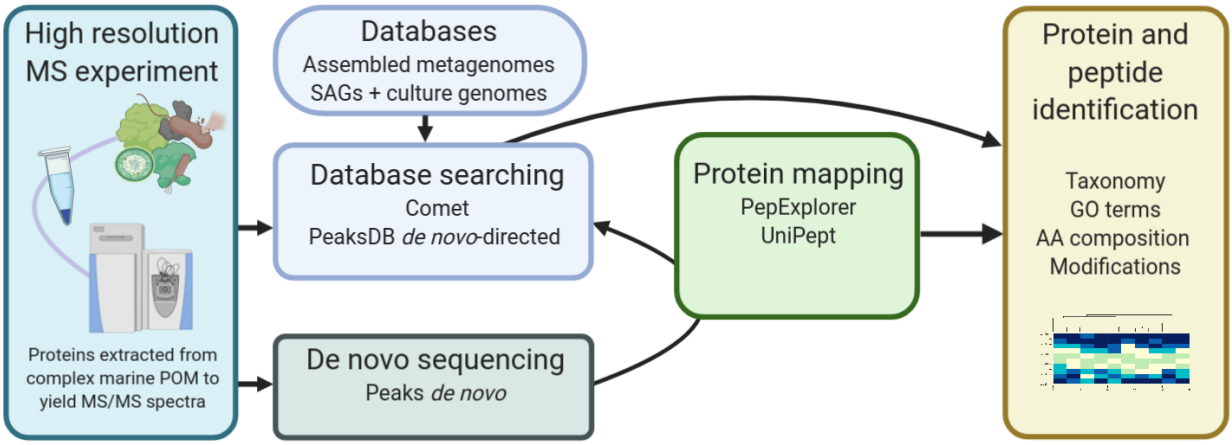

Figure 2.1. The *de novo*-discovery metaproteomic workflow. MS/MS spectra are acquired by high resolution mass spectrometry. These, along with reference databases generated from sample-specific metagenomes or from public repositories, are input to the database search tool Comet Eng et al. (2013) and the *de novo*-directed database search program PeaksDB (J. Zhang et al. 2012). Only the spectra are the input into the *de novo* algorithm (Peaks *de novo* Ma et al. 2003). *De novo* candidate sequences are input into the protein mapping tools PepExplorer and Unipept, to match them to experiment databases (PepExplorer) or expansive reference databases (UniPept). A selection of confident *de novo*-discovered proteins are then added to the original search database and the spectra are queried again. Finally, the resultant peptide and protein identifications are examined for biological and geochemical information: taxonomy, gene

ontology (GO) terms, amino acid composition, and peptide mass modifications such as deamidation or methylation).

2.2 MATERIALS AND METHODS

2.2.1 *Data collection*

The proteomics dataset from *Prochlorococcus marinus* strain MED4 (subsp. pastoris str. CCMP1986; PROMP) was obtained from a culture of *P. marinus* MED4 grown on natural seawater-amended media (see Supporting Methods for more details).

Environmental samples were collected in the ETNP on board the R/V *Sikuliaq* at an offshore ODZ station (16.58 N, 107.05 W) between January 8-13, 2017 (cruise ID: SKQ201617S) (Supporting Figure 2.4). Suspended OM samples were collected with *in situ* pumps (McLane Labs, Falmouth, MA) fitted with two sequential combusted filters of 2.7 µm and 0.3 µm pore size (142 mm diameter GF/D and GF/75 glass fiber – Whatman, Marlborough, MA) and immediately frozen at -80 ◦C; volumes filtered were between 300 and 1000 L. Only the 0.3-2.7 µm fractions were used in this study, which I presume includes both free-living microbes as well as small suspended detrital organic matter. These samples are termed suspended particles. Sinking particles were collected with free drifting, unpoisoned sediment net traps modified from Peterson et al. (2005) with deployments lasting between 24 and 96 hours. Sediment trap samples were filtered onto 47 mm GF-75 glass fiber filters (0.3 µm) and immediately frozen at -80◦ C. These samples are referred to as sinking particles.

2.2.2 *Peptide extraction and LC-HRMS*

The extraction technique used is detergent-free and results in a crude protein extract that is chromatographically challenging to interpret but also presumably not biased for nor against any particular type of protein (Bridoux et al., 2015; Nunn & Keil, 2006). For both the small (0.3 µm - 2.7 µm) suspended and sinking POM, protein was extracted from one or two 11 mm diameter punches of filter into ammonium bicarbonate buffer (adapted from Bridoux et al., 2015). The chilled suspension was lysed via three cycles each of mechanical disruption with silica beads (50% 100 µm diameter and 50% 400 µm) freeze-thawing, and 30 seconds in a high-power water bath sonicator. The resulting lysate was centrifuged at 4800 rpm to isolate protein in the supernatant.

Protein concentration in the extract was estimated using a Lowry assay (Bio-Rad). Extracted protein underwent reduction of disulfides, carbamidomethylation of free cysteine residues using iodoacetamide, and in-solution protease digestion using trypsin, following a protocol adapted from Nunn et al. (2010). Digested peptides (1 µg) were desalted using a macro-spin C18 column (NestGroup, Southborough, MA) and resuspended in 5% acetonitrile with 0.1% formic acid and Waters Hi3 E. coli peptide standard mixture (100 fmol/L).

Reverse-phase LC-HRMS analysis was performed in duplicate with a Waters nanoACQUITY UPLC coupled to a Thermo Q Exactive Plus HRMS equipped with an NSI source. Digested peptides were separated on a home-packed analytical column consisting of a 37 cm long, 75-µm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100 A, 5 µm; Michrom) coupled to a 4 cm long, 100 m i.d. precolumn (Magic C18AQ, 200 Å, 5µm; Michrom). Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100% LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 90-minute gradient from 5-35% solvent B. All analyses were carried out in positive mode at an NSI spray voltage of 2 kV, and data-dependent acquisition (DDA) on the top 10 ions.

2.2.3 *Database searching and de novo peptide sequencing*

A composite database containing 4,075,587 protein sequences was constructed consisting of assembled proteins from site- and region-specific metagenomes, single amplified genomes and culture genomes (Fuchsman et al., 2019) (see Supporting Methods, Supporting Table 2.4, and Supporting File 2.1). Database searching and validation was conducted using the open-source software contained in the Trans-Proteomic Pipeline (TPP, Deutsch et al., 2015). Raw spectra were searched using Comet (2016.01 rev. 2, Eng et al., 2013) against a FASTA protein reference database. For the benchmarking study, the reference was the GenBank *P. marinus* MED4 database (accession ID BX548174.1); for the environmental study, the ETNP assembled metagenome was used (see above and Supporting Information). Search parameters included tryptic enzymatic constraint, 2 max. missed cleavages, 8 max. modifications per peptide, 15 ppm peptide mass tolerance, and 0.5 Da fragment mass tolerance. Results from technical replicates were combined after searching and final lists of identified peptides were filtered above an XCorr (the cross-correlation of the experimental and theoretical spectra) that maintained a false discovery rate (FDR) <0.1% for the *Prochlorococcus* MED4 dataset and <1.0% for the ETNP dataset using a

reversed database target-decoy strategy (Elias & Gygi, 2010). This cutoff was preferred over performing discriminant score validation using an algorithm like PeptideProphet (Keller et al., 2002) which is less suited to metaproteomic searches of complex environmental samples. In the case of the *Prochlorococcus* MED4 dataset, I compared the peptide >XCorr 2.17 (FDR = 0.097%) to those with a >95% and >99% probability as calculated by PeptideProphet (see Supporting Figure 2.7). For protein identification, I required at minimum one unique peptide.

De novo peptide sequencing was performed in Peaks (v8.5; Bioinformatics Solutions, Waterloo, Canada; Ma et al., 2003) using a collision-induced dissociation (CID) fragmentation model. Tryptic *de novo* searches allowing up to 2 missed cleavages were run with the same fixed and variable optimized PTMs as the Comet database searches (see Supporting Methods). Parent mass error tolerance was 15.0 ppm and fragment mass error tolerance was 0.5 Da. Peptides five amino acids in length or longer were accepted if they had >80% combined average local (residue-level) confidence (ALC) and >10% local confidence. The Peaks *de novo* algorithm was selected because it performs a dynamic programming step to compensate for missing fragment ions – an expected issue in complex metaproteomic samples. For a comprehensive comparison of *de novo* algorithms, see Muth & Renard (2018) and Allmer (2011).

***De novo*-directed database PSM –** Combining an initial screening using *de novo* sequencing with a database search is termed *de novo*-discovery. This was performed using PeaksDB within Peaks Studio (v8.5; Bioinformatics Solutions, Waterloo, Canada; Zhang et al., 2012) against the same databases and parameters as the Comet searches. The PeaksDB algorithm relies on *de novo* sequencing results to 1) improve the initial peptide filtration step common to all database searching approaches and 2) improve the scoring function. The combination of PSM and *de novo* sequencing has been shown to significantly improve sensitivity and accuracy in comparison to existing database search techniques (J. Zhang et al., 2012). The PeaksDB peptide-level confidence score is a transformed P-value generated by a linear discriminative function that determines matching between fragment ions and spectrum peaks as well as the agreement between *de novo* candidates and PSM peptides. In these PeaksDB searches, *de novo* peptides were accepted at >80% ALC and PSM peptides with -10lgP scores >20. This is equivalent to a P-value of 1% (e.g., 99% confidence), signifying the probability that the identification is to a false peptide sequence.

Open modification searching – Since amino acids are frequently modified after translation, either for cell-specific purposes or during degradation, the optimization tool PeaksPTM (Han et al., 2011) was used prior to setting the modifications searched for using the database and *de novo* approaches (see Supporting Methods and Supporting Figure 2.5).

Iterative PSM – high-confidence *de novo* peptides were aligned to the UniProt Knowledgebase sequence repository (see next section). When evaluated against UniProt taxonomies, many *de novo* peptides were found to be specific to fungal and fungus-like microorganisms. Thus, fungi and fungi-like (thraustochytrids, Oomycota) protein sequences were downloaded from NCBI if *de novo* peptide alignments were as low as the phylum level (see below). These sequences (33111 protein sequences, see Supporting File 2.5) were added into the original reference database for additional, iterative PSM identification by Comet and PeaksDB. Spectral files, databases, and Comet search parameter files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD023187.

Similarity-driven identification – Because there is no reference database as there is with PSM, the output of *de novo* sequencing is only a list of candidate peptide sequences, precursor intensities, and confidence scores. To assemble information from the *de novo* results, I aligned the sequences to each dataset's respective reference database using PepExplorer (patternlabforproteomics.org, Leprevost et al., 2014). PepExplorer is a multi-sequence alignment tool designed for *de novo* peptide sequences that equates leucine and isoleucine (amino acids of identical mass and thus indistinguishable by typical *de novo* sequencing) and performs a parallel target-decoy alignment to estimate an FDR. Results are ranked using a radial basis function neural network by alignment score, *de novo* score, precursor state charge, and peptide length (see Ma & Johnson, 2012). *Prochlorococcus* MED4 *de novo* peptides were aligned to the GenBank genome-derived database and *de novo* peptides from the ETNP POM dataset were aligned to the site-specific composite database described above. Only alignments above 85% were accepted, and results were kept under 1% FDR as determined by a reverse decoy alignment. Protein identification required at least 1 specific peptide alignment.

The *de novo* peptide output was also processed using Unipept (https://unipept.ugent.be, Mesuere et al., 2016), a taxonomic annotation tool that provides access and alignment to the entire

UniProt sequence repository, or UniProt KnowledgeBase, which as of 06.19.18 contained 558 reviewed and 137,213,158 unreviewed entries (UniProt Consortium, 2018).

To assess how taxon specific the peptides were, I employed the UniPept Metaproteome Analysis (Gurdeep Singh et al., 2019) lowest common ancestor (LCA) tool to calculate the lowest common ancestor for all Comet, PeaksDB, and Peaks *de novo* peptides. In all UniPept lowest common ancestor searches, leucine and isoleucine were equated and peptide input was de-duplicated after modifications were stripped from the sequences. To evaluate the theoretical pool of tryptic *Prochlorococcus* MED4 peptides that could result from the GenBank proteome reference database, it was trypsin-digested in silico using the Protein Digestion Simulator tool (Pacific Northwest National Lab Integrative Omics).

Modification and amino acid composition analysis – PSM and *de novo* sequence outputs were filtered and evaluated for length, amino acid composition, and residue modification extent. Amino acid compositions were evaluated as mole percentages, e.g., an amino acid's percentage of the total amino acids in the identified peptides after accounting for differences in peak area for each peptide. All peptide data analysis and visualization were performed using Python, code for which can be accessed at https://github.com/Keil-Aquatic Organic-Geochemistry-Group/ETNP-denovo-peptidomics/.

2.3 RESULTS AND DISCUSSION

2.3.1 *Benchmarking with Prochlorococcus data*

To validate the *de novo*-directed workflow I compared three approaches using the *Prochlorococcus* culture dataset: 1) traditional PSM database searching with Comet, 2) *de novo*-directed database searching using PeaksDB, and 3) *de novo* sequencing using Peaks ("*de novo*-only"). Of the three methods, the *de novo*-directed PeaksDB search found the most *Prochlorococcus* peptides, nearly twice as many as the Comet search (Table 2.1). When the peptides were converted to proteins, the *de novo*-directed approach identified 1066 proteins, 127 of which were not found by Comet, meaning that adding the *de novo* step prior to database searching boosted proteome coverage by 10% compared with database searching alone. The *de novo*-only approach identified 38 proteins that were found neither by Comet nor PeaksDB (see

next paragraph). Using all three approaches together resulted in a protein identification increase of 17% compared to the Comet database search alone (see Supporting File 2.5 for peptide and protein lists). This increase in coverage is routinely observed in other studies using *de novo*-directed peptidomics (Juba et al., 2015; Tran et al., 2019).

All proteins in the *Prochlorococcus* MED4 dataset should be sourced from that single organism, barring contamination (for taxonomic peptide matching results via UniPept LCA, see Supporting Figure 2.6). Therefore, all correct *de novo*-only peptides should match to Cyanobacteria or *Prochlorococcus*, if specific enough. Peptides that match to a phylum level other than Cyanobacteria in UniPept can be used to calculate a so-called 'false mapping rate' – the number of peptides matching to an incorrect phylum over the total number of peptides matched to the phylum level. I found that *de novo*-only peptides had a higher false mapping rate (11.1%, Table 2.1 compared with *de novo*-directed PeaksDB or Comet (0.072% and 0.16%, respectively). That the *de novo*-only peptides had the highest false mapping rate could reflect a) contaminants in the culture, b) truly incorrect *de novo* sequencing, or c) a combination of both. Many of the incorrect peptide assignments are to Alpha- and Gammaproteobacteria, which are possible real contaminants and are types of bacteria known to be closely associated with diatoms within the phycosphere (Amin et al., 2012). Unfortunately, the culture was not evaluated by microscopy, flow cytometry, or amplicon sequencing, and without this information there is no good way to quantitatively distinguish if it was axenic. If I assume some heterotrophic bacteria in the culture and remove the proteobacterial proteins from consideration, the false mapping rate for the *de novo*-only sequencing falls to 4.86%.

Peptides identified solely via database matching using Comet might be expected to yield a false discovery rate of 0% because the database is derived only from a *Prochlorococcus* MED4 genome. However, the false mapping rate to the UniProt KnowledgeBase is 0.16%. This is because the specific genome used to build the Comet search database (see Supplemental Methods) is not a part of the UniProt KnowledgeBase. Proteins in the MED4 genome-derived database that are not present in UniProt map to a higher taxonomic level in UniProt and lose specificity. Were those peptides submitted to the KnowledgeBase a MED4 the Comet peptides' false mapping rate would fall to 0%.

Overall, these benchmarking results indicate that the *de novo*-discovery approach improves peptide and protein identification over PSM database searching alone and suggests that the *de*

novo-only peptides might provide additional information on peptides present in a sample but not in the reference database.

Table 2.1. Results of *Prochlorococcus* MED4 benchmarking experiment.

Approach	Total peptides[2]	False discovery rate %	*Pro.* proteins	UniPept *Pro.* IDs	False mapping rate (%)
Comet > XCorr 2.17	1.89	<0.1	1216	4172	0.16
Peaks *dn* > 80% ALC	1.94	<1.0[3]	1160	1075	11.1
PeaksDB > 20-10LgP	2.33	<0.1	1374	6522	0.78

2.3.2 *Eastern tropical North Pacific POM*

Suspended and sinking particles were collected from three depths in the ETNP using *in situ* pumps (suspended particles) and drifting sediment traps (sinking particles). Samples were from the anoxic epipelagic zone above the secondary chlorophyll maximum (100 m), in the anoxic mesopelagic (265 m), and in the oxygenated bathypelagic (965-1000 m). The reference database does not match perfectly to the sample set; it is offset both temporarily and in depth. This makes the *de novo*-directed approach more useful than it would be if a reference database from the exact same samples was available. Peptides extracted from particles are assumed to be sourced from hundreds if not thousands of different organisms and are within a matrix of detrital organic and inorganic material (Bochdansky et al., 2017), which is in stark contrast to the spectra generated from a culture such as the benchmarking study presented here. As such, I anticipated the spectra to be noisier and more prone to ionization and fragmentation suppression and PSMs were manually inspected for quality (see Figure 2.9). While in the *Prochlorococcus* benchmark study I confidently sequenced *de novo* peptides for 14% of the MS1 spectra, in the case of the ETNP POM samples this metric ranged from 1.3% to 10% (Table 2.2). The sediment trap at the top of the ODZ (100 m) had the lowest number of fragment ion spectra and the lowest number of peptides identified by PSM (Table 2.2). Though swimmers were removed, this sample was unique in that it contained numerous detrital

[2] Including modified peptides.
[3] From decoy database alignment using PepExplorer.

aragonite pteropod shells. Such mineral load could potentially cause ionization suppression, fragmentation, or spectral interference.

Table 2.2. Peptidomic results from suspended and sinking particulate organic matter from the ETNP, January 2017.

Sample	Comet peptides (%) FDR	Peaks de novo peptides > 80% ALC	PeaksDB peptides >20-lgP	Total protein IDs (all approaches)	Peptides specific to the phylum level (UniPept LCA)
100 m sus.	167 (1.0)	1140	1299	1233	443
265 m sus.	96 (1.0)	734	1023	906	513
965 m sus.	36 (0.0)	272	203	195	118
100 m sink.	21 (0.0)	465	120	189	119
265 m sink.	44 (0.0)	1602	365	378	570
965 m sink.	32 (0.0)	1086	232	235	441

Peptide modifications – Deeper samples had more post-translationally modified amino acids (PTMs), as did sinking particles compared to suspended (Figure 2.2). Methylation of arginine and lysine are PTMs that occur within living cells of both prokaryotes and eukaryotes, and it entails the substitution of a hydrogen for a methyl group residue side chain, changing the physiochemical properties of the molecule by increasing hydrophobicity, neutralizing charge, and in certain cases increasing the amount of space the moiety takes up (Zhang et al., 2018). This PTM has previously been shown to decrease that rate at which heterotrophic bacteria assimilate extracellular amino acids in seawater (Keil & Kirchman, 1992). The open modification searches indicated that arginine methylation was an important modification in the ETNP metaproteomes, and the subsequent closed modification searches showed greater extent of methylation in deep suspended POM and

sinking POM at all depths compared to epi- and mesopelagic suspended POM (30-50% of arginine residues modified vs. <10%). The methylation especially increased in the bathypelagic suspended sample, presumably the oldest and most degraded POM, indicating preferential preservation of methylated proteinaceous material that to my knowledge has not previously been observed in the water column.

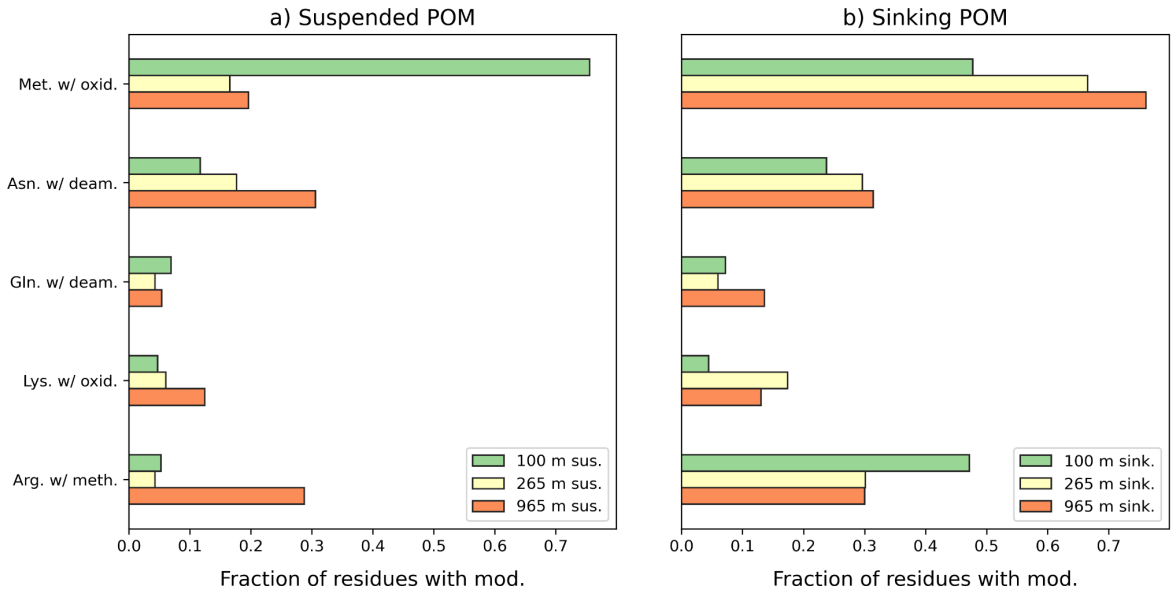

Figure 2.2. Fraction of affected residues with modifications in peptides from a) suspended particles and b) sinking particles from the eastern tropical North Pacific oxygen deficient zone. The five variable peptide modifications were methionine oxidation, asparagine deamidation, glutamine deamidation, lysine hydroxylation, and arginine methylation. All peptides were scaled by peak area before modification extent calculation. Not shown is the single fixed modification of cysteine carbamidomethylation which is induced in the protein digestion protocol with addition of iodoacetamide. Mass differences and additional details about modifications can be found in Supporting Information.

The oxidation of methionine residues is one of the most common in-cell protein degradation pathways, by which methionine sulfoxide is formed by addition of oxygen to its sulfur atom, decreasing protein function and stability (Stadtman et al., 2003). Methionine residues also can become oxidized during proteomic sample preparation (Lin et al., 2013), though the in-solution

digestions I performed are much less prone to oxidation compared to in-gel digestions. Approximately ~4% of all methionine residues are oxidized by these methods (Froelich & Reid, 2008) and the mole percentages I observe in this study greatly exceed that (Figure 2.2). In the sinking particles there is an increase in methionine oxidation from 50% at the surface to 75% at 1000 meters, consistent with degradation (Figure 2.2). The suspended sediment samples show a different profile, with the largest signal in the shallow epipelagic suspended sediment (76%) compared to 20% and 23%, respectively, in the meso and bathypelagic suspended fractions (Figure 2.2). The suspended particles collected at 100 m depth were at the bottom of the oxycline, and just above the deep chlorophyll maximum within the ODZ, an area of rapid redox change (Fuchsman et al., 2019). The large extent of methionine oxidation near the surface could be the result of oxygen stress in that unique environment (Fuchsman et al 2019). Bacterial *in vivo* methionine oxidation is reversible by methionine sulfoxide reductases (Gennaris et al., 2015), and I could be observing the use of protein methionine residues as endogenous antioxidants to combat oxidative stress at this depth.

Deamidation is a common post-translational modification, often non-enzymatic, that results in the conversion of the side chain amide group of a glutamine or asparagine residue into an acidic carboxylate group (Geiger & Clarke, 1987). In prokaryotes, deamidation has been postulated as a signal for protein degradation, a kind of 'molecular clock' indicating to a cell the time for protein turnover (Lorenzo et al., 2015). Unfortunately, for the purposes of determining environmental peptide deamidation, the usual proteomic extraction conditions (ammonium bicarbonate buffer at pH 8.5, 37 ˚C incubation) can induce *in vitro* peptide deamidation (Pace et al., 2013). Given that the same laboratory technique was applied to both the *Prochlorococcus* culture (which was at exponential growth phase) and the ETNP samples, I can evaluate whether the natural samples had more deamidation than the culture and infer process from that. *Prochlorococcus* culture asparagines and glutamines were only 13% and 2.3% deamidated (Supporting Figure 2.10), compared to 35% and 18% in the deep ETNP POM peptides (Figure 2.2), indicating the likelihood that deamidation is a common occurrence in detrital samples. Over longer degradation timescales, Abdulla et al. (2018) demonstrated the accumulation of highly deamidated peptides in anoxic sediment porewater from Santa Barbara Basin. The deamidated 'skeletal' remains of peptides may represent 25–45% of porewater DOC as determined by FTICR-MS – the deamidation I observe in early diagenesis could be greatly magnified as selective and/or

incomplete hydrolysis of proteins continues. Additionally, in an investigation of sinking organic matter and amino acid degradation in the ETNP, Van Mooy et al. (2002) suggest that preferential degradation of nitrogen-rich amino acids such as asparagine and glutamine could be a source of ammonium to support anammox in this region. By detection of differential deamidation of these residues, these results indicate a specific mechanism to fuel this chemoautotrophy, and targeted peptidomic work (e.g. Keil et al., 2016) could quantify the ammonium provided by this process.

Cyanobacterial peptides – As anticipated at this site where *Prochlorococcus* is an abundant primary producer responsible for a deep chlorophyll maximum within the ODZ, I identified many peptides that are specific to Cyanobacteria (here called 'cyanobacterial peptides') in the epipelagic suspended sample (Table 2.3). Less predictably, I also found several cyanobacterial peptides in mesopelagic suspended POM and in all sinking POM samples using *de novo* searching (Table 2.3). Identification of cyanobacterial peptides in meso and bathypelagic samples hints at export dynamics of these small (0.6 - 1.2 μm, Ting et al., 2007) cells. Several studies have pointed to the significance of picophytoplankton to global POM flux from the euphotic zone, with evidence ranging from large-scale inverse and network analysis (Richardson & Jackson, 2007) to lipid (Cavan et al., 2018; Close et al., 2013) and pigment (Lomas & Moran, 2011) measurements. These last two examples can resolve taxonomic origins of picoplanktonic material to a certain extent, but without the functional information peptides can theoretically provide. Future work could track proteinaceous material of known origins by looking for peptide markers for taxa, function, and intracellular localization using the *de novo*-discovery tool. For example, Fuchsman et al. (2019) used cyanobacterial DNA markers and peptides in POM to show the contribution of *Prochlorococcus* to *in situ* primary production in the ETNP ODZ. Peptide biomarkers have also been used in quantitative, targeted metaproteomics studies of marine processes (Keil et al., 2016; Saito et al., 2015) but the only ones to track surface proteins to depth have done so in higher latitudes, where diatom blooms can contribute to fast sinking, fresh phytodetritus (Bridoux et al., 2015; Moore et al., 2012; Nunn et al., 2010).

Table 2.3. Phylum specific peptides identified in suspended and sinking POM in the ETNP in January 2017.

Sample	Cyanobacteria (*de novo* only) [% of phylum level IDs][1]	Fungi-like[2] (*de novo* only) [% of phylum level IDs]	Other bacteria[3] (*de novo* only) [% of phylum level IDs]
100 m sus.	100 (15) [45]	42 (6) [3.7]	188 (134) [38]
265 m sus.	4 (4) [0.057]	31 (7) [1.6]	378 (103) [88]
965 m sus.	0 - -	10 (5) [1.1]	72 (55) [64]
100 m sink.	1 (1) [0.095]	24 (2) [2.2]	66 (57) [82]
265 m sink.	4 (4) [1.7]	71 (8) [11]	309 (269) [61]
965 m sink.	6 (6) [0.63]	56 (3) [7.2]	235 (222) [37]

1. Percent of phylum-level IDs per sample, after normalization of peak area abundance and weighted averaging of Comet, *de novo*, and PeaksDB by peptides total for each approach.
2. Includes Ascomycota, Basidiomycota, Chytridiomycota, Cryptomycota, Mucoromycota, Zoopagomycota, Oomycota.
3. Includes Actinobacteria, Bacteroidetes, Firmicutes, Candidatus Marinimicrobia, and Proteobacteria.

Because *de novo* peptides were mapped to the entire UniProt sequence repository during lowest common ancestor analysis, the true number of cyanobacterial peptides identified at depth is conservative; sequences above the phylum level are not counted because they are not specific-enough. Of the 100 cyanobacterial-specific peptides identified in the 100 m suspended sample, 4 had deamidated asparagine residues and 6 methylated arginines. Of the 15 cyanobacterial peptides found in the sinking POM and deep suspended POM, 3 asparagines were modified, 1 glutamine,

and 3 arginines were methylated. The number of cyanobacterial peptides identified in these samples is too small to provide a very meaningful comparison, but these modifications indicate 1) a reason surface-derived peptides are difficult to identify by traditional proteomic means and 2) that patterns of peptide deamidation may be a common degradative step in the breakdown as these small autotrophs in the ODZ.

Gene Ontology (GO) assignments for the cyanobacterial peptides provide insight on preservation in suspended or sinking POM. Of the 100 cyanobacterial peptides found in the 100 m suspended POM, where I would expect the greatest portion of living *Prochlorococcus*, 69 peptides had GO-terms, the most frequent being 10 peptides annotated as sourced from 'cytoplasm'; 9 'phycobilisome'; 8 'thylakoid membrane'; 7 'ribosome'; 7 'integral component of membrane'. Of the 15 specific cyanobacterial *de novo* peptides found in the deeper presumed detrital pools, only 6 cellular component GO-terms identified and 5 were associated with membranes (Supporting File 2.6). This is consistent with research showing the recalcitrance of bacterial cell walls and their associated amino acids in marine systems (Benner & Kaiser, 2003; Grutters et al., 2002).

Fungal peptides –One advantage of the *de novo*-directed workflow is to use *de novo*-only peptides that are not in the database to improve the database. The original database did not include sequences from any fungi or fungal-like taxa. However, mapping the *de novo*-only results to UniProt yielded peptides specific to *Ascomycota*, *Basidiomycota, Chytridiomycota, Cryptomycota, Mucoromycota*, among others, leading us to augment the original reference database with ~100,000 sequences from these and other marine fungi (Supporting File 2.4). In the ultimate searches against this *de novo*-informed database, I found many additional fungal and fungal-like peptides (Table 2.3), particularly in the deeper sinking particles. Using different techniques, fungi have been found in other low oxygen marine environments including the coastal upwelling system of Peru (hyphal chitin staining; Gutiérrez et al., 2011), and in anoxic sediments of the Arabian Sea (Stief et al., 2014). Recently Peng and Valentine (2021) evaluated the diversity of free living and particle-associated (>22 μm) fungi in the ETNP ODZ, also finding fungi to be dominated by *Basidiomycota* and *Ascomycota* at most depths, consistent with these *de novo*-discovery results. Fungi are potentially important players in marine biogeochemical cycling (Amend et al., 2019). They are osmotrophs that secrete enzymes externally and assimilate resultant metabolites. In the

microenvironment of a detrital particle, fungi may play a role transferring carbon and nutrients between the various heterotrophs and chemoautotrophs present, as modeled on laboratory particles by Roberts et al. (2020). In theory, this transfer could include substrates from mineral dissolution, as recently demonstrated by Yu et al. (2019) in fungus-mineral interface experiments showing that fungus hyphae can alter the local redox state of iron. Certain fungi have been isolated and shown to perform anaerobic dissimilatory nitrate reduction in anoxic sediments in the Arabian Sea (Stief et al., 2014) and off coastal India (Cathrine & Raghukumar, 2009), including an *Aspergillus terreus* strain. Peptides specifically identifying the filamentous genus *Aspergillus* were found in all samples of this study except the deepest suspended POM (Supporting File 2.7), as were numerous other filamentous and hyphae-producing groups of fungi and non-fungi (Actinobacteria, Oomycota). While it's unknown how important fungal nitrogen cycling may be in the water column of the ODZ, Peng & Valentine recently demonstrated using ^{15}N-tracer incubations that fungal N_2O production peaked at the upper oxic-anoxic interface in the ENTP ODZ water column.

Our findings demonstrate that fungal peptides can be extracted and identified from marine POM via metaproteomic techniques. Further targeted and directed metaproteomic approaches could reveal the abundance, diversity, and functional roles of fungi and fungi-like organisms in the processing of organic matter.

Peptide amino acid composition - The relative amino acid composition of identified peptides was converted to mole percent (mol %) the common unit used when reporting total hydrolyzable amino acid compositions (THAA), using peak intensities (Figure 2.3; I did not include arginine as it is a tryptic terminus and likely overcounted by this approach). This allows for comparison to earlier studies and provides a bridge between the peptidomic and hydrolyzable amino acid approaches. The amino acid composition of the sinking and suspended particles was, for most residues, typical for sinking particles (Lee et al., 2000; Van Mooy et al., 2002). Serine, alanine, and valine were the most abundant. The makeup of amino acids changed little with depth for either sinking or suspended particles. A similar observation was made for the hydrolyzable amino acid pool in sinking particles in the ETNP at a more coastal ODZ site (22.40 °N, 106.30 °W) by Van Mooy et al. (2002). The similarity between THAA and peptidomic-based amino acid composition values signals that the metaproteomics strategy accurately capturing this geochemical indicator of organic matter freshness without sacrificing the valuable information contained in peptide sequences. I do

note however that these current peptidomic methods are not designed to detect non-protein amino acids like ornithine, β-alanine, and γ-aminobutyric acid, which can reach high concentrations in marine sediments (Carr et al., 2016).

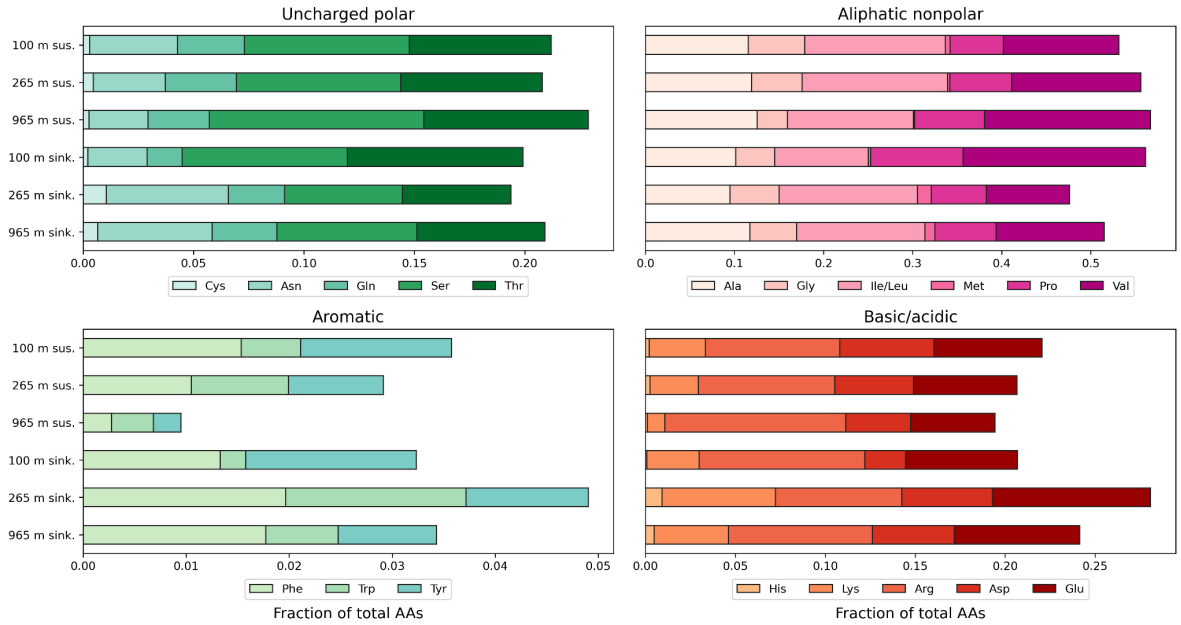

Figure 2.3. Relative amino acids composition from peptides identified in suspended and sinking particles from the eastern tropical North Pacific oxygen deficient zone. Peptide relative abundances are corrected for peak area and sequence length, or normalized area abundance factor (NAAF).

A notable exception to relative amino acid consistency in this dataset is the relative decrease of aromatic residues phenylalanine (Phe), tryptophan (Typ) and tyrosine (Tyr) in suspended particles with depth (Figure 2.3). This was not observed in the sinking particles, similar to measurements of total hydrolyzable amino acids in suspended and sinking POM from the VERTEX II expedition to the ETNP (18◦ N, 1O8 ◦ W) in October-November, 1981 (Lee & Cronin, 1984). In suspended particles from the VERTEX II site, relative molar abundance of Phe decreased from 5.18% at 100 m to 2.65% at 300 m and 0.62% at 1000 m (Lee & Cronin, 1984), while in sediment traps Phe remains similar from 100 m (0.053%) to 900 m (0.053%). The differential decrease in aromatic residues in these samples indicates that the deeper suspended particles are more degraded than sinking particles at the same depth. That aromatic residues are the only shift in amino acid composition suggests that this study is capturing an early snapshot of selective protein degradation.

For example, the enrichment of glycine in marine sediments is well documented (Cowie & Hedges, 1992; Dauwe & Middelburg, 1998; Keil et al., 2000) but not observed here. While Lee and Cronin (1984) observed the small enrichments of glycine in poisoned sediment traps after 3-week collection time, this study's results from non-poisoned, 3-day sediment trap collection show no significant changes in glycine relative abundance. The proposed mechanism for preferential accumulation of glycine is its abundance in the cell walls of diatoms (Hecky et al., 1973) and bacteria as peptidoglycan (Vollmer, 2008) in contrast to the relatively rapid degradation of residues more concentrated in the cytoplasm (phenylalanine, tyrosine, and glutamic acid) (Dauwe et al., 1999; Keil et al., 2000). The detergent-free extraction used here, and enzymatic digestion designed for proteins may then select against peptidoglycan associated residues.

In summary, adding *de novo* peptide sequencing into the processing of proteomic samples led to an increase in the number of peptides confidently identified, and allowed refinement of the database used in a traditional proteomic approach. In the ETNP the *de novo*-assisted approach allowed identification of fungal peptides in POM, with fungi and fungi-like sequences making up 11% of the phylum-level identifiable peptide pool. I identify an abundance of modified amino acids in sinking and suspended particles, including high levels of deamidation. This suggests that partial and selective degradation of protein fragments could be an important pathway for microorganisms to obtain nitrogen.

2.4 ACKNOWLEDGMENTS

I am very grateful to Drs. Penny Chisholm and Jamie Becker for use of the *P. marinus* dataset; Dr. Priska von Haller at the University of Washington Proteomics Resource Center; Cedar McKay and Ryan Groussman for database construction assistance; and the two anonymous reviewers for their helpful comments. I also thank the captain and crew of the R/V *Sikuliaq*. This work was funded by the National Science Foundation (DEB-1542240 to G.R and R.G.K) and an NSF Graduate Research Fellowship DGE-1762114 to M.E.D. and R.G.K. This work was supported by grants from the Simons Foundation (LS award ID 385428 to A.E.I. and 329108 to A.E.I.).

2.5 DATA AVAILABILITY

Environmental peptidomic raw and processed data files, search databases, and 837 search parameter files have been deposited to the ProteomeXchange Consortium via 838 the PRIDE partner repository to be made available with publication (PXD023187). *Prochlorococcus marinus* MED4 raw and processed data files have been deposited to PRIDE with Project number PXD027589.

2.6 REFERENCES

Abdulla, H. A., Burdige, D. J., & Komada, T. (2018). Accumulation of deaminated peptides in anoxic sediments of Santa Barbara Basin. *Geochimica et Cosmochimica Acta*, *223*, 245–258. https://doi.org/10.1016/j.gca.2017.11.021

Allmer, J. (2011). Algorithms for the *de novo* sequencing of peptides from tandem mass spectra. *Expert Review of Proteomics*, *8*(5), 645–657. https://doi.org/10.1586/epr.11.54

Amend, A., Burgaud, G., Cunliffe, M., Edgcomb, V. P., Ettinger, C. L., Gutiérrez, M. H., Heitman, J., Hom, E. F. Y., Ianiri, G., Jones, A. C., Kagami, M., Picard, K. T., Quandt, C. A., Raghukumar, S., Riquelme, M., Stajich, J., Vargas-Muñiz, J., Walker, A. K., Yarden, O., & Gladfelter, A. S. (2019). Fungi in the Marine Environment: Open Questions and Unsolved Problems. *MBio*, *10*(2). https://doi.org/10.1128/mBio.01189-18

Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, *76*(3), 667–684. https://doi.org/10.1128/MMBR.00007-12

Babbin, A. R., Keil, R. G., Devol, A. H., & Ward, B. B. (2014). Organic Matter Stoichiometry, Flux, and Oxygen Control Nitrogen Loss in the Ocean. *Science*, *344*(6182), 406–408. https://doi.org/10.1126/science.1248364

Benner, R., & Kaiser, K. (2003). Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter. *Limnology and Oceanography*, *48*(1), 118–128. https://doi.org/10.4319/lo.2003.48.1.0118

Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., Jensen, O. N., & Herndl, G. J. (2017). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences*, 201708779. https://doi.org/10.1073/pnas.1708779115

Bridoux, M., Neibauer, J., Ingalls, A., Nunn, B., & Keil, R. (2015). Suspended marine particulate proteins in coastal and oligotrophic waters. *Journal of Marine Systems*, *143*. https://doi.org/10.1016/j.jmarsys.2014.10.014

Burdige, D. J. (2007). Preservation of Organic Matter in Marine Sediments: Controls, Mechanisms, and an Imbalance in Sediment Organic Carbon Budgets? *Chemical Reviews*, *107*(2), 467–485. https://doi.org/10.1021/cr050347q

Carr, S. A., Mills, C. T., & Mandernack, K. W. (2016). The use of amino acid indices for assessing organic matter quality and microbial abundance in deep-sea Antarctic sediments of IODP Expedition 318. *Marine Chemistry*, *186*, 72–82.

https://doi.org/10.1016/j.marchem.2016.08.002
Cathrine, S. J., & Raghukumar, C. (2009). Anaerobic denitrification in fungi from the coastal marine sediments off Goa, India. *Mycological Research*, *113*(1), 100–109. https://doi.org/10.1016/j.mycres.2008.08.009
Cavan, E. L., Giering, S. L. C., Wolff, G. A., Trimmer, M., & Sanders, R. (2018). Alternative Particle Formation Pathways in the Eastern Tropical North Pacific's Biological Carbon Pump. *Journal of Geophysical Research: Biogeosciences*, *123*(7), 2198–2211. https://doi.org/10.1029/2018JG004392
Close, H. G., Shah, S. R., Ingalls, A. E., Diefendorf, A. F., Brodie, E. L., Hansman, R. L., Freeman, K. H., Aluwihare, L. I., & Pearson, A. (2013). Export of submicron particulate organic matter to mesopelagic depth in an oligotrophic gyre. *Proceedings of the National Academy of Sciences*, *110*(31), 12565–12570. https://doi.org/10.1073/pnas.1217514110
Cowie, G. L., & Hedges, J. I. (1992). Sources and reactivities of amino acids in a coastal marine environment. *Limnology and Oceanography*, *37*(4), 703–724. https://doi.org/10.4319/lo.1992.37.4.0703
Dauwe, B., & 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. https://doi.org/10.4319/lo.1998.43.5.0782
Dauwe, B., Middelburg, J. J., Herman, P. M. J., & Heip, C. H. R. (1999). Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology and Oceanography*, *44*(7), 1809–1814. https://doi.org/10.4319/lo.1999.44.7.1809
Deutsch, E. W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., & Moritz, R. L. (2015). Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics. Clinical Applications*, *9*(7–8), 745–754. https://doi.org/10.1002/prca.201400164
Elias, J. E., & Gygi, S. P. (2010). Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. *Methods in Molecular Biology (Clifton, N.J.)*, *604*, 55–71. https://doi.org/10.1007/978-1-60761-444-9_5
Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *PROTEOMICS*, *13*(1), 22–24. https://doi.org/10.1002/pmic.201200439
Estes, E. R., Pockalny, R., D'Hondt, S., Inagaki, F., Morono, Y., Murray, R. W., Nordlund, D., Spivack, A. J., Wankel, S. D., Xiao, N., & Hansel, C. M. (2019). Persistent organic matter in oxic subseafloor sediment. *Nature Geoscience*, *12*(2), 126–131. https://doi.org/10.1038/s41561-018-0291-5
Finkel, Z. V., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I., & Irwin, A. J. (2016). Phylogenetic Diversity in the Macromolecular Composition of Microalgae. *PLoS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0155977
Froelich, J. M., & Reid, G. E. (2008). The origin and control of ex vivo oxidative peptide modifications prior to mass spectrometry analysis. *PROTEOMICS*, *8*(7), 1334–1345. https://doi.org/10.1002/pmic.200700792
Fuchsman, C. A., Palevsky, H. I., Widner, B., Duffy, M., Carlson, M. C. G., Neibauer, J. A., Mulholland, M. R., Keil, R. G., Devol, A. H., & Rocap, G. (2019). Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. *The ISME Journal*, 1. https://doi.org/10.1038/s41396-019-0452-6
Geiger, T., & Clarke, S. (1987). Deamidation, isomerization, and racemization at asparaginyl and

aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *The Journal of Biological Chemistry*, *262*(2), 785–794.

Gennaris, A., Ezraty, B., Henry, C., Agrebi, R., Vergnes, A., Oheix, E., Bos, J., Leverrier, P., Espinosa, L., Szewczyk, J., Vertommen, D., Iranzo, O., Collet, J.-F., & Barras, F. (2015). Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature*, *528*(7582), 409–412. https://doi.org/10.1038/nature15764

Grutters, M., van Raaphorst, W., Epping, E., Helder, W., de Leeuw, J. W., Glavin, D. P., & Bada, J. (2002). Preservation of amino acids from *in situ*-produced bacterial cell wall peptidoglycans in northeastern Atlantic continental margin sediments. *Limnology and Oceanography*, *47*(5), 1521–1524. https://doi.org/10.4319/lo.2002.47.5.1521

Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615. https://doi.org/10.1021/acs.jproteome.8b00716

Gutiérrez, M. H., Pantoja, S., Tejos, E., & Quiñones, R. A. (2011). The role of fungi in processing marine organic matter in the upwelling ecosystem off Chile. *Marine Biology*, *158*(1), 205–219. https://doi.org/10.1007/s00227-010-1552-z

Han, X., He, L., Xin, L., Shan, B., & Ma, B. (2011). PeaksPTM: Mass Spectrometry-Based Identification of Peptides with Unspecified Modifications. *Journal of Proteome Research*, *10*(7), 2930–2936. https://doi.org/10.1021/pr200153k

Hecky, R. E., Mopper, K., Kilham, P., & Degens, E. T. (1973). The amino acid and sugar composition of diatom cell-walls. *Marine Biology*, *19*(4), 323–331. https://doi.org/10.1007/BF00348902

Juba, M. L., Russo, P. S., Devine, M., Barksdale, S., Rodriguez, C., Vliet, K. A., Schnur, J. M., van Hoek, M. L., & Bishop, B. M. (2015). Large Scale Discovery and *De novo*-Assisted Sequencing of Cationic Antimicrobial Peptides (CAMPs) by Microparticle Capture and Electron-Transfer Dissociation (ETD) Mass Spectrometry. *Journal of Proteome Research*, *14*(10), 4282–4295. https://doi.org/10.1021/acs.jproteome.5b00447

Kawasaki, N., Sohrin, R., Ogawa, H., Nagata, T., & Benner, R. (2011). Bacterial carbon content and the living and detrital bacterial contributions to suspended particulate organic carbon in the North Pacific Ocean. *Aquatic Microbial Ecology*, *62*, 165–176. https://doi.org/10.3354/ame01462

Keil, R. (2017). Anthropogenic Forcing of Carbonate and Organic Carbon Preservation in Marine Sediments. *Annual Review of Marine Science*, *9*(1), 151–172. https://doi.org/10.1146/annurev-marine-010816-060724

Keil, R. G., & Kirchman, D. L. (1992). Bacterial Hydrolysis of Protein and Methylated Protein and Its Implications for Studies of Protein Degradation in Aquatic Systems. *Applied and Environmental Microbiology*, *58*(4), 1374–1375.

Keil, R. G., Neibauer, J. A., Biladeau, C., van der Elst, K., & Devol, A. H. (2016). A multiproxy approach to understanding the "enhanced" flux of organic matter through the oxygen-deficient waters of the Arabian Sea. *Biogeosciences*, *13*(7), 2077–2092. https://doi.org/10.5194/bg-13-2077-2016

Keil, R. G., Tsamakis, E., & Hedges, J. (2000). *Early diagenesis of particulate amino acids in marine systems*. 69–82.

Keller, A., Nesvizhskii, A. I., Kolker, E., & Aebersold, R. (2002). Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search.

Analytical Chemistry, *74*(20), 5383–5392. https://doi.org/10.1021/ac025747h

Lee, C., & Cronin, C. (1984). Particulate amino acids in the sea: Effects of primary productivity and biological decomposition. *Journal of Marine Research*, *42*(4), 1075–1097. https://doi.org/10.1357/002224084788520710

Lee, C., Wakeham, S., & Arnosti, C. (2004). Particulate Organic Matter in the Sea: The Composition Conundrum. *AMBIO: A Journal of the Human Environment*, *33*(8), 565–575. https://doi.org/10.1579/0044-7447-33.8.565

Lee, C., Wakeham, S. G., & I. Hedges, J. (2000). Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(8), 1535–1568. https://doi.org/10.1016/S0967-0637(99)00116-8

Lin, Y.-S., Lipp, J. S., Elvert, M., Holler, T., & Hinrichs, K.-U. (2013). Assessing production of the ubiquitous archaeal diglycosyl tetraether lipids in marine subsurface sediment using intramolecular stable isotope probing. *Environmental Microbiology*, *15*(5), 1634–1646. https://doi.org/10.1111/j.1462-2920.2012.02888.x

Lomas, M. W., & Moran, S. B. (2011). Evidence for aggregation and export of cyanobacteria and nano-eukaryotes from the Sargasso Sea euphotic zone. *Biogeosciences*, *8*, 203–216. https://doi.org/10.5194/bg-8-203-2011

Lorenzo, J. R., Alonso, L. G., & Sánchez, I. E. (2015). Prediction of Spontaneous Protein Deamidation from Sequence-Derived Secondary Structure and Intrinsic Disorder. *PLoS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0145186

Ma, B., & Johnson, R. (2012). *De novo* sequencing and homology searching. *Molecular & Cellular Proteomics: MCP*, *11*(2), O111.014902. https://doi.org/10.1074/mcp.O111.014902

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., & Lajoie, G. (2003). PEAKS: Powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *17*(20), 2337–2342. https://doi.org/10.1002/rcm.1196

McCarthy, M., Pratum, T., Hedges, J., & Benner, R. (1997). Chemical composition of dissolved organic nitrogen in the ocean. *Nature*, *390*(6656), 150–154. https://doi.org/10.1038/36535

Mesuere, B., Van der Jeugt, F., Devreese, B., Vandamme, P., & Dawyndt, P. (2016). The unique peptidome: Taxon-specific tryptic peptides as biomarkers for targeted metaproteomics. *Proteomics*, *16*(17), 2313–2318. https://doi.org/10.1002/pmic.201600023

Moore, E. K., Nunn, B. L., Faux, J. F., Goodlett, D. R., & Harvey, H. R. (2012). Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments. *Limnology and Oceanography: Methods*, *10*(5), 353–366. https://doi.org/10.4319/lom.2012.10.353

Moore, E. K., Nunn, B. L., Goodlett, D. R., & Harvey, H. R. (2012). Identifying and tracking proteins through the marine water column: Insights into the inputs and preservation mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta*, *83*, 324–359. https://doi.org/10.1016/j.gca.2012.01.002

Muth, T., & Renard, B. Y. (2018). Evaluating *de novo* sequencing in proteomics: Already an accurate alternative to database-driven peptide identification? *Briefings in Bioinformatics*, *19*(5), 954–970. https://doi.org/10.1093/bib/bbx033

Nunn, B. L., & Keil, R. G. (2006). A comparison of non-hydrolytic methods for extracting

amino acids and proteins from coastal marine sediments. *Marine Chemistry*, *98*(1), 31–42. https://doi.org/10.1016/j.marchem.2005.06.005

Nunn, B. L., Norbeck, A., & Keil, R. G. (2003). Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. *Marine Chemistry*, *83*(1–2), 59–73. https://doi.org/10.1016/S0304-4203(03)00096-3

Nunn, B. L., Ting, Y. S., Malmström, L., Tsai, Y. S., Squier, A., Goodlett, D. R., & Harvey, H. R. (2010). The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation. *Limnology and Oceanography*, *55*(4), 1790–1804. https://doi.org/10.4319/lo.2010.55.4.1790

O'Bryon, I., Jenson, S. C., & Merkley, E. D. (2020). Flying blind, or just flying under the radar? The underappreciated power of *de novo* methods of mass spectrometric peptide identification. *Protein Science*, *29*(9), 1864–1878. https://doi.org/10.1002/pro.3919

Pace, A., Wong, R., Zhang, Y., Kao, Y.-H., & Wang, Y. (2013). Asparagine deamidation dependence on buffer type, pH, and temperature. *Journal of Pharmaceutical Sciences*, *102*. https://doi.org/10.1002/jps.23529

Paulmier, A., & Ruiz-Pino, D. (2009). Oxygen minimum zones (OMZs) in the modern ocean. *Progress in Oceanography*, *80*(3–4), 113–128. https://doi.org/10.1016/j.pocean.2008.08.001

Richardson, T. L., & Jackson, G. A. (2007). Small Phytoplankton and Carbon Export from the Surface Ocean. *Science*, *315*(5813), 838–840. https://doi.org/10.1126/science.1133471

Roberts, C., Allen, R., Bird, K. E., & Cunliffe, M. (2020). Chytrid fungi shape bacterial communities on model particulate organic matter. *Biology Letters*, *16*(9), 20200368. https://doi.org/10.1098/rsbl.2020.0368

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M. R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd, A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome Research*, *18*(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761

Saito, M. A., Dorsk, A., Post, A. F., McIlvin, M. R., Rappé, M. S., DiTullio, G. R., & Moran, D. M. (2015). Needles in the blue sea: Sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics*, *15*(20), 3521–3531. https://doi.org/10.1002/pmic.201400630

Salas, P. M., Sujatha, C. H., Ratheesh Kumar, C. S., & Cheriyan, E. (2018). Amino acids as indicators to elucidate organic matter degradation profile in the Cochin estuarine sediments, Southwest coast of India. *Marine Pollution Bulletin*, *127*, 273–284. https://doi.org/10.1016/j.marpolbul.2017.12.010

Schmidt, F., Koch, B. P., Elvert, M., Schmidt, G., Witt, M., & Hinrichs, K.-U. (2011). Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under Contrasting Sedimentary Redox Conditions in the Black Sea. *Environmental Science & Technology*, *45*(12), 5223–5229. https://doi.org/10.1021/es2003414

Stadtman, E. R., Moskovitz, J., & Levine, R. L. (2003). Oxidation of methionine residues of proteins: Biological consequences. *Antioxidants & Redox Signaling*, *5*(5), 577–582. https://doi.org/10.1089/152308603770310239

Stief, P., Fuchs-Ocklenburg, S., Kamp, A., Manohar, C.-S., Houbraken, J., Boekhout, T., de Beer, D., & Stoeck, T. (2014). Dissimilatory nitrate reduction by Aspergillus terreus

isolated from the seasonal oxygen minimum zone in the Arabian Sea. *BMC Microbiology*, *14*(1), 35. https://doi.org/10.1186/1471-2180-14-35

Tanoue, E., Nishiyama, S., Kamo, M., & Tsugita, A. (1995). Bacterial membranes: Possible source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta*, *59*(12), 2643–2648. https://doi.org/10.1016/0016-7037(95)00134-4

Ting, C. S., Hsieh, C., Sundararaman, S., Mannella, C., & Marko, M. (2007). Cryo-Electron Tomography Reveals the Comparative Three-Dimensional Architecture of *Prochlorococcus*, a Globally Important Marine Cyanobacterium. *Journal of Bacteriology*, *189*(12), 4485–4493. https://doi.org/10.1128/JB.01948-06

Tran, N. H., Qiao, R., Xin, L., Chen, X., Liu, C., Zhang, X., Shan, B., Ghodsi, A., & Li, M. (2019). Deep learning enables *de novo* peptide sequencing from data-independent-acquisition mass spectrometry. *Nature Methods*, *16*(1), 63–66. https://doi.org/10.1038/s41592-018-0260-3

UniProt Consortium, T. (2018). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *46*(5), 2699–2699. https://doi.org/10.1093/nar/gky092

Van Mooy, B. A. S., Keil, R. G., & Devol, A. H. (2002). Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrification. *Geochimica et Cosmochimica Acta*, *66*(3), 457–465. https://doi.org/10.1016/S0016-7037(01)00787-6

Vollmer, W. (2008). Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiology Reviews*, *32*(2), 287–306. https://doi.org/10.1111/j.1574-6976.2007.00088.x

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., & Peterson, M. J. (1997). Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta*, *61*(24), 5363–5369. https://doi.org/10.1016/S0016-7037(97)00312-8

Yamada, N., & Tanoue, E. (2003). Detection and partial characterization of dissolved glycoproteins in oceanic waters. *Limnology and Oceanography*, *48*(3), 1037–1048. https://doi.org/10.4319/lo.2003.48.3.1037

Yu, G.-H., Chi, Z.-L., Teng, H. H., Dong, H.-L., Kappler, A., Gillings, M. R., Polizzotto, M. L., Liu, C.-Q., & Zhu, Y.-G. (2019). Fungus-initiated catalytic reactions at hyphal-mineral interfaces drive iron redox cycling and biomineralization. *Geochimica et Cosmochimica Acta*, *260*, 192–203. https://doi.org/10.1016/j.gca.2019.06.029

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., & Ma, B. (2012). PEAKS DB: *De novo* sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics: MCP*, *11*(4), M111.010587. https://doi.org/10.1074/mcp.M111.010587

Zhang, M., Xu, J.-Y., Hu, H., Ye, B.-C., & Tan, M. (2018). Systematic Proteomic Analysis of Protein Methylation in Prokaryotes and Eukaryotes Revealed Distinct Substrate Specificity. *PROTEOMICS*, *18*(1). https://doi.org/10.1002/pmic.201700300

2.7 SUPPORTING FILES

Supporting datafiles (files available at https://github.com/MeganEDuffy/denovo-etnp)

1. Table S3. Supplementary NCBI genomes amended to the ETNP assembled metagenome composite protein database. Adapted from Fuchsman et al. (2019). Uploaded as Excel .xlsx file.

2. Table S4. NCBI Fungal sequences added to protein database for in iterative database search. Uploaded as Excel .xlsx file.

3. Table S5. *Prochlorococcus* MED4 proteins identified by Comet database searching, PeaksDB *de novo*-directed database searching, and Peaks *de novo* sequencing with PepExplorer protein mapping. Uploaded as Excel .xlsx file.

4. Table S6. Sequences, lowest common ancestor taxonomies, GO-terms, InterPro annotation, and Enzyme Commission names for Cyanobacteria-specific peptides identified from suspended and sinking particulate organic matter in the ETNP. Excel .xlsx file.

5. Table S7. Sequences, lowest common ancestor taxonomies, GO-terms, InterPro annotation, and Enzyme Commission names for fungi and fungi-like specific peptides identified from suspended and sinking particulate organic matter in the ETNP. Excel .xlsx file.

Table S8. Comet output of iterative searches against combined initial ETNP database and fungal and fungal-like protein sequences from NCBI suggested by *de novo* results. Excel .xlsx file.

7. Environmental peptidomic raw and processed data files have been deposited

to the ProteomeXchange Consortium via the PRIDE (project number PXD023187) partner repository to be made available with publication.

2.8 SUPPORTING TEXT

2.8.1 *Complete Prochlorococcus culture and LCMS methods*

Cell pellets from cultures of *Prochlorococcus* marinus high-light (HL) adapted strain MED4 (subsp. Pastoris str. CCMP1986; PROMP) were used as received from Chisholm lab July 2015. Cultures grown in Pro99 media (0.2 µm filtered, autoclaved coastal seawater from Vineyard Sound, MA; supplemented with 50 µM PO4, 800 µM NH4, 10 mM sodium bicarbonate, and trace metals) in acid washed polycarbonate carboys under constant light conditions at 24°C. Cells were harvested in mid/late exponential phase. Cell biomass was collected by centrifuging 4L of culture (10,000 x g; 15 minutes at 10 °C) to pellet the cell biomass.

Cell pellets were resuspended in 50 mM ammonium bicarbonate (AMBIC) buffer containing 5 mM ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA) in a microcentrifuge tube containing glass beads. In addition, samples containing only AMBIC/EDTA were processed alongside samples to serve as processing blank controls for protease digestions. MED4 cell suspension was subjected to bead-beating for 30 seconds at 6 m/s, frozen at -20 °C for 45 minutes, and thawed. The bead-beating and freeze-thaw cycles were repeated three times followed by an overnight freeze at -20°C. Frozen samples were then thawed, cells further disrupted by vortexing with beads, and then microcentrifuged at 8,000 rpm (6,010 rcf) for 3 minutes. For MED4 cell pellets, approximately 2 Ö 109 cells were extracted and used for downstream proteomics processing. Bulk protein concentration for lysed cell extracts were estimated from cell density measurements. Lysed cells were subjected to in-solution proteolytic digestion. Disulfide bonds were reduced with tris(2-carboxyethyl)-phosphine hydrochloride (TCEP HCl, Thermo Scientific; to a final concentration of 10 mM) for 1 hour at room temperature in the dark. Reduced samples were then alkylated with iodoacetamide (IAM, Sigma; to a final concentration of 30 mM) for 1 hour at room temperature in the dark. Excess IAM was quenched by the addition of dithiolthreitol (DTT, Sigma; to final concentration of 5 mM) for 1 hour at room temperature in the dark. Protein extracts were proteolytically digested with mass spectrometry-grade (Promega) trypsin at an estimated substrate-to-enzyme ratio of 25 (w/w; based on cell density conversion to

protein concentration) overnight at 37 °C. Following protease digestion, samples were dried to near-dryness using a centrivap concentrator (Thermo Savant SPD2010 SpeedVac). Digested samples were desalted using a MacroSpin C18 column (NestGroup) following an established manufacturer protocol. Samples were eluted from the MacroSpin columns using two times 200 µL of 80% LCMS-grade acetonitrile with 20% LCMS-grade water containing 0.1% LCMS-grade formic acid. Desalted samples were concentrated using the centrivap to near dryness and were immediately resuspended in 95% LCMS-grade water with 5% LCMS-grade acetonitrile containing 0.1% LCMS-grade formic acid. The resuspension solution also contained an internal standard of synthetic peptides (Hi3 Escherichia coli Standard, Waters) at 50 fmol/µL. Samples were immediately analyzed by liquid chromatography coupled to mass spectrometry (LCMS) or stored at -80 °C until analysis.

Samples were injected onto the column of a Waters ACQUITY M-class UPLC with an injection volume of 2 µL. For MED4, 2 µL injection volume corresponded to an estimated 0.6 x 108 cells on-column and samples were only injected in duplicates. Peptide separation was performed by reversed-phase chromatography using a nanoACQUITY HSS T3 C18 column (1.8 µm, 75 µm x 250 mm; 45 °C) with an ACQUITY UPLC M-class Symmetry C18 trapping column (180 µm x 20 mm). The peptides were trapped at a flow rate of 5 µL/min at 99% A for 3 minutes. A flow rate of 0.3 µL/min was used over a gradient between LCMS-grade water (A) and LCMS-grade acetonitrile (B), both modified with 0.1% LCMS-grade formic acid. The total 145-minute gradient method for the separation of the peptides started at 95% A and ramped to 60% A over the course of 120 minutes. The gradient then switched to 15% A at 122-127 minutes followed by a ramp back to starting conditions 95% A at 128-145 minutes.

The ACQUITY M-class was coupled to a Thermo QExactive HF Orbitrap high-resolution mass spectrometer (HRMS) equipped with a nano-electrospray (NSI) source made in-house following the University of Washington Proteomic Resource (UWPR) design. Using a MicroTee (PEEK, 0.025” OD), the commercial analytical column was connected to a commercial emitter (PicoTip, Waters) and the liquid path was applied a high voltage through a platinum wire (adapted from UWPR design). All analyses were carried out in positive mode and a NSI spray voltage of 2.0 kV. Data was collected using data dependent acquisition (DDA) using Xcalibur 4.0 data acquisition software (Thermo Fisher). The MS1 scan range was 400-2000 m/z at 60,000 resolution with a maximum injection time of 30 ms and automated gain control of 1e6. Following each MS1

scan, data-dependent MS2 (dd-MS2) was set to perform on the top 10 ions in a data-dependent manner at 15,000 resolution with a normalized collision energy of 27 eV. Additional selection criteria for dd-MS2 were as follows: maximum injection time of 50 ms with an automated gain control of 5e4, the isolation window was 1.5 Da and dynamic exclusion was set at 20 sec.

2.8.2 *Additional environmental POM LCMS methods*

Digested peptides were separated on a home-packed analytical column consisting of a 37 cm long, 75-µm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100Å, 5µm; Michrom) coupled to a 4 cm long, 100 m i.d. precolumn (Magic C18AQ, 200Å, 5µm; Michrom). Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100% LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 90-minute gradient from 5-35% solvent B.

2.8.3 *Open modification searches*

Since amino acids are frequently modified after translation, either for cell-specific purposes or during degradation, a computationally efficient method is needed to search for the myriad possible post translational modifications (PTMs). To efficiently search while also minimizing false discoveries, the optimized tool PeaksPTM (Han et al., 2011) was used prior to setting the modifications searched for using the database and *de novo* approaches. Both datasets were scanned using PeaksPTM in automatic precursor mass detection mode allowing for all validated modifications from the Unimod database (unimod.org). Search parameters included tryptic enzymatic constraint, 2 max. missed cleavages, 15 ppm peptide mass tolerance, 0.2 Da fragment mass tolerance, minimum A-score of 200 (a measure of modification location confidence), fixed carbamidomethylation of cysteine, and variable oxidation of methionine. Based on PeaksPTM results, the most frequently occurring, high A-score (high confidence) modifications were used to evaluate whether adding additional PTMs to the searches altered the rate of false discovery. That is, we sought the optimal balance between searching for PTMs and avoiding a vast search space that leads to decreased sensitivity. A series of Comet searches and Peaks sequencing runs with increasing numbers of variable modifications (up to 8) were evaluated for PSM-level FDR as calculated using a reversed database target-decoy search. Using these PTM ramping results,

optimal variable modification parameters (limiting searches to 6 modifications) were selected for each of the two datasets and were subsequently used for the final Comet and PeaksDB searches and Peaks *de novo* sequencing. For both datasets carbamidomethylation of cysteine was set as a fixed modification. Oxidation of methionine, deamidation of asparagine and glutamine, and methylation of arginine were set as variable modifications for both datasets. Unique to the *Prochlorococcus* MED4 dataset, methylation of lysine was set as a variable modification and unique to the ETNP dataset, hydroxylation of lysine (see Supporting Figure 2.5 for PTM-optimization results).

2.8.4 *Complete details of ETNP protein database construction*

A composite database to search ETNP POM was constructed consisting of assembled proteins from site- and region-specific metagenomes and single amplified genomes (Fuchsman et al., 2017; Glass et al., 2015; Tsementzi et al., 2016). Additional relevant culture genomes were included (for an inclusive list of database entries and metagenome station locations see Supporting Figure 2.4Table 2.4). The overall composite database contained 4,075,587 protein sequences and is the same as utilized in Fuchsman et al. (2019). Assembled metagenomes are from samples collected at our study site and a hydrographically similar station 83 km away (17.043◦ N, 106.543◦ W) aboard the R/V Thompson in April 2012 (cruise ID TN278). Metagenome samples include seawater filtered through 0.2 µm pore size polyethersulfone and 30 µm pore size nylon membranes from 60-300 m depth. Details of sampling and metagenomic library construction, sequencing, and VELVET (Zerbino, 2010) assembly of individual samples can be found in Fuchsman et al. (2017). Metagenomic reads and assembled contigs from that study can be found at GenBank Bioproject PRJNA350692. The same metagenomic samples were later individually reassembled with the MEGAHIT (Li et al., 2015) assembler and published in Widner et al. (2018). As these two assemblies differed, both were included in the proteomics database. Additionally, ETNP metagenomic assemblies (assembled with IBDA; Peng et al., 2012) from June 2013 at 18.9 °N, 104.5◦ W, and 18.8◦ N, 104.7◦ W (depth range: 30 m - 300 m) and May 2014 at 18.9 °N, 104.5 °W (68 m and 120 m) were included Glass et al. (2015); Tsementzi et al. (2016). These metagenomes can be found at NCBI BioProject ID PRJNA254808 and JGI Project IDs 1059848

and 1059863, respectively. All assembled metagenomic contigs had open reading frames called and functionally annotated with Prokka (Seemann, 2014) and the resulting proteins were deduplicated and used to form the proteomic search database. To reduce redundancy, a Python script was used to eliminate 100% identical proteins from the combined assembled metagenomic data. We note that 3 samples (100, 120 and 150 m) were filtered onto 30 μm membranes, representing large suspended and potentially sinking POM.

2.8.5 *Peptide taxonomic specificity*

De novo sequences are typically shorter than those identified by PSM (Muth & Renard, 2018). Indeed, in these results high confidence *de novo* peptides had a mean length of 10.0 amino acids while Comet and PeaksDB peptide results had lengths of 15.3 and 13.6, respectively. This length discrepancy is the reason fewer proteins are identified by mapping *de novo* sequences to the *Prochlorococcus* database proteome: there are fewer unique peptides when their length is shorter. Our PepExplorer mapping parameters required at least one specific peptide alignment for a protein to be identified, meaning that the peptide must not align to any other protein in the database. Peptide length, then, is a large factor in the usefulness of applying *de novo* peptide sequencing to a complex environmental system where proteins are sourced from numerous organisms. We evaluated how peptide length determines the taxonomic specificity between different peptidomic approaches by using a lowest common ancestor analysis after peptide mapping to the entire UniProt Knowledgebase. In theory, highly taxa-specific (typically long) peptides can match as low as the species or subspecies level, while relatively nonspecific (conserved, typically short) peptides will have lowest common ancestor results as high as the kingdom level or even across multiple kingdoms (e.g., to both Bacteria and Archaea). We first determined the maximum number of peptides in our database that are specific to *Prochlorococcus* when mapped to UniPept: an in silico tryptic digestion of the 1792 GenBank database proteins resulted in 32974 peptides. Of those, 17812 (54%) mapped specifically to the species level with UniPept lowest common ancestor analysis, the maximum number of specific peptides we could have found in this study. Peptide lowest common ancestor results showed that a greater proportion of *de novo* peptides were matched across kingdoms compared to peptides found by Comet or PeaksDB.

2.8.6 *De novo peptide accuracy*

One reason the MED4 *de novo* peptides might not match the *Prochlorococcus* sequences and that is the same as why some of the Comet searches don't—that there are expressed proteins that are not annotated and aren't in the reference proteome we used. We formally checked for this by aligning a FASTA file of all the incorrect *de novo* peptides (non phylum-level Cyanobacteria hits) versus a six frame translation of the MED4 genome using tblastn. This alignment didn't yield any significant similarities.

The Comet peptides' false mapping rate of 0.16% is the result of 7 peptides that all map to Basidiomycota (3), Bacteroidetes (1), Ascomycota (1), or Actinobacteria (2) by UniPept's lowest common ancestor analysis. They range from peptides 6-8 amino acids in length, the average being 7.3 amino acids. This is not even half of the Comet peptides' average length of 15.3 amino acids – indicating there is more potential for overlapping sequence homology with other organisms.
Using NCBI's protein BLAST search tool against the non-redundant database shows all 7 of these short peptides to be highly similar in proteins across many taxa. For example, the Comet-identified peptide LVSTCPR, which in UniPept LCA matched to the Actinobacterium *Rhodococcus erythropolis*, has 100% identity over 100% of the 7-residue query with protein sequences from 35 different taxa in the non-redundant database, including Synechococcaceae bacterium WB73xG012 and Synechococcus sp. XM-24 as well as the Rhodococcus. Another peptide, CGQVAYR, aligns perfectly with *Prochlorocccus marinus* but also a sequence of the *Rachicladosporium antarcticum.*

2.9 SUPPORTING TABLES

Table 2.4. Input genomes to ETNP POM protein search database.

Cruise, station	TN278[1] St. 136	TN278 St. BB2		NH1315[2] St. 6	NH1315 St. 10	NH1315 St. 6
	106.543°W, 17.043°N	107.148°W, 16.527°N		104.7°W, 18.9°N	104.8°W, 18.8°N	104.7°W, 18.9°N
Size fraction	>0.2 µm	>30 µm	0.2-30 µm	0.2–1.6 µm		single cell
30m				MA	MA	
60m	MA					
70m	MA					
80m					MA	
85m				MA		
90m	MA					
100m	MA	MA		MA		
110m	MA					
120m	MA	MA	MA			
125m					MA	SAG

140m	MA					
150m		MA				
160m	MA					
180m	MA					
300m	MA			MA	MA	SAG

Cruise, station, depth and size class information for samples contributing to the eastern tropical North Pacific (ETNP) assembled metagenomes (MA) and single cell genomes (SAG). Samples were sourced from: (1) Cruise TN278 R/V Thomas G. Thompson, April 17-30, 2012 (Fuchsman et al., 2017), (2) cruise NH1315 R/V New Horizon, 13–28 June 2013 (Glass et al., 2015), and (3) cruise NH1315 (Tsementzi et al., 2016) used to build our composite proteomics search database. Adapted from Fuchsman et al. (2019).

2.10 SUPPORTING FIGURES

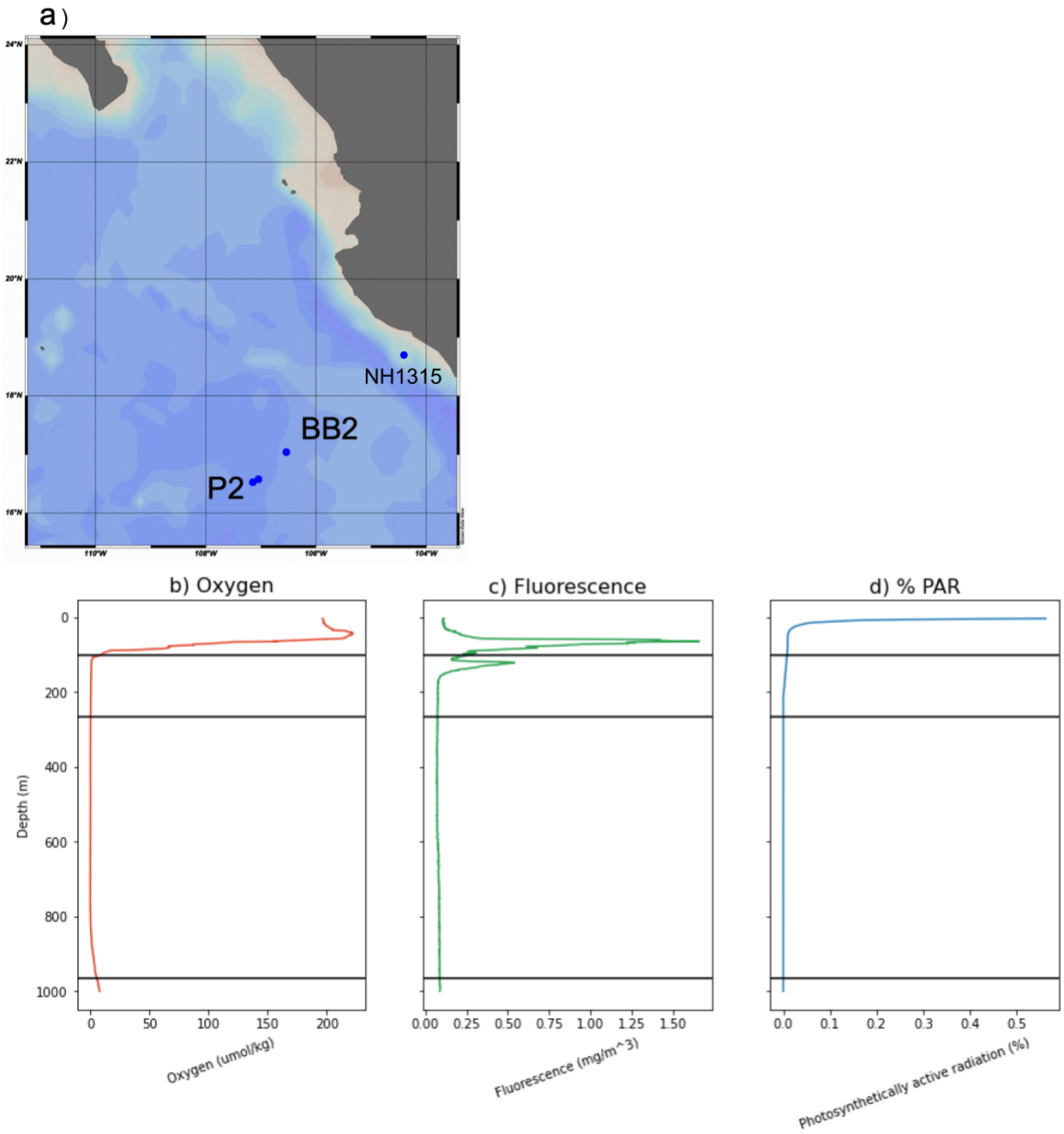

Figure 2.4. Sampling locations (a) and selected water column profiles (b-d) for sediment trap and McLane large volume pump deployments at Station P2, cruise SKQ201617S, January 8-13, 2017. Samples for assembled metagenomes and single amplified genomes used to build composite proteomics search database were collected from Stations 136 and BB2, cruise TN278

R/V Thomas G. Thompson, April 17-30 2012 (Fuchsman et al., 2017) and Stations 6 and 10, cruise NH1315 R/V New Horizon, 13–28 June 2013 (Glass et al., 2015; Tsementzi et al., 2016). Depth profiles show SKQ201617 Station P2 dissolved oxygen (b) chlorophyll a fluorescence (c), and % PAR (photosynthetically active radiation) (d). Horizontal lines indicate sediment trap and *in situ* pump deployment depths. Map produced using Ocean Data View (Schlitzer, R., Ocean Data View, odv.awi.de, 2018).

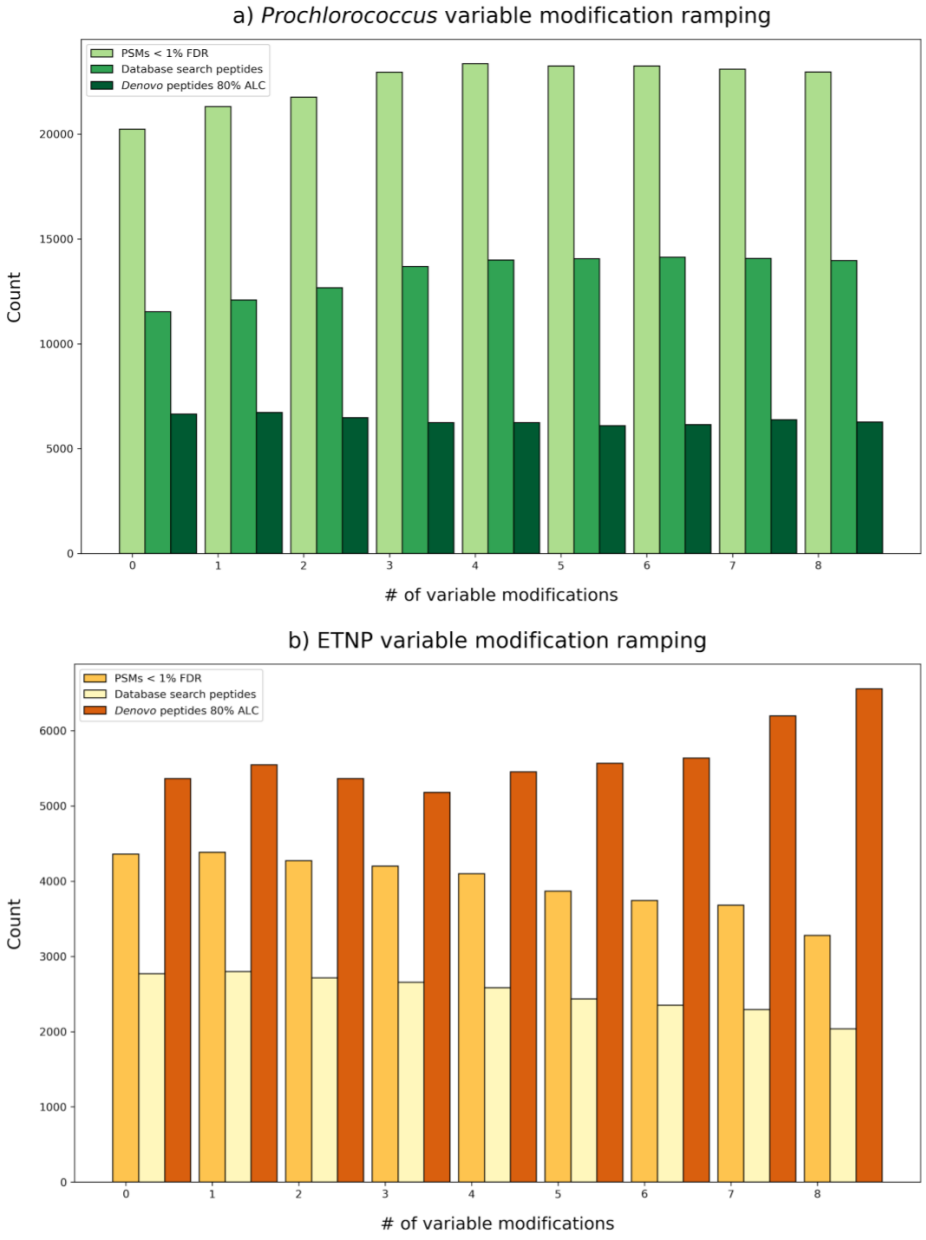

Figure 2.5. Results of variable modification ramping searches performed using PeaksDB *de novo*-directed database searching on the *Prochlorococcus* and (a) eastern tropical North Pacific particulate organic matter dataset (b). Variable PTMs to suggest were determined using the open modification search tool PeaksPTM, which considers each validated modification in the Unimod database. For the *Prochlorococcus* dataset, suggested modifications were (in order of additive introduction: 1) oxidation of methionine, 2) deamidation of asparagine, 3) deamidation of glutamine, 4) methylation of lysine, 5) methylation of arginine, 6) replacement of 2 protons on

lysine by an iron cation, 7) dehydration of aspartic acid, and 8) dimethylation of arginine. For the ETNP POM dataset, suggested modifications were (in order of additive introduction: 1) oxidation of methionine, 2) deamidation of asparagine, 3) deamidation of glutamine, 4) oxidation of lysine, 5) methylation of arginine, 6) dehydration of aspartic acid, 7) acetylation of protein N-terminal, and 8) formylation of N-terminal lysine.

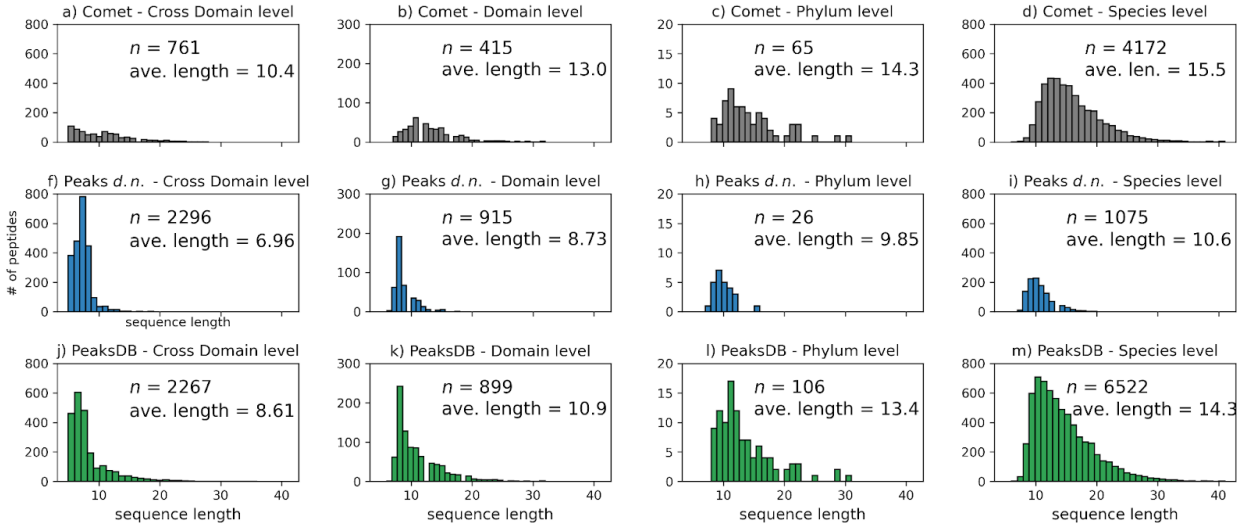

Figure 2.6. Histograms showing taxonomic specificities of peptides resulting from Comet database searches (grey), Peaks *de novo* sequencing (blue), and PeaksDB *de novo*-directed database searching (green) of cultured *Prochlorococcus* MED4 protein LC-MS/MS dataset. Of the 5929 peptides resulting from Comet database searches after deduplication and removal of modifications (FDR <0.1%, minimum peptide length = 5, mean length = 14.7), 5547 were successfully matched to the UniProt KnowledgeBase using the sequence mapping tool UniPept (Gurdeep Singh et al., 2019). Lowest common ancestor analysis of the Comet peptides showed that a) 761 were matched across multiple domains; b) 415 were matched as low as the domain level, Bacteria); c) 65 as low as the phylum level (Cyanobacteria); and d) 4172 were matched to the species (*Prochlorococcus* marinus) or subspecies level. Of the 8735 Peaks *de novo* peptides (>80% average local confidence (ALC), minimum peptide length = 5, mean length = 10.0), 4225 were matched UniProt, with f) 2296 matched across domains, g) 915 to Bacteria, h) 26 to Cyanobacteria, and i) 1075 to *Prochlorococcus* marinus or lower. Of the 11255 PeaksDB peptides (-10lgP >20, minimum peptide length = 5, mean length = 13.6), 10097 were matched UniProt, with j) 2267 matched across domains, k) 899 to Bacteria, l) 106 to Cyanobacteria, and m) 6522 to *Prochlorococcus* marinus or lower. Other taxonomic levels (order, group, etc.) are not plotted for simplicity. Note that y-axis scales are consistent for identical taxon levels between identification approaches but not across all levels.

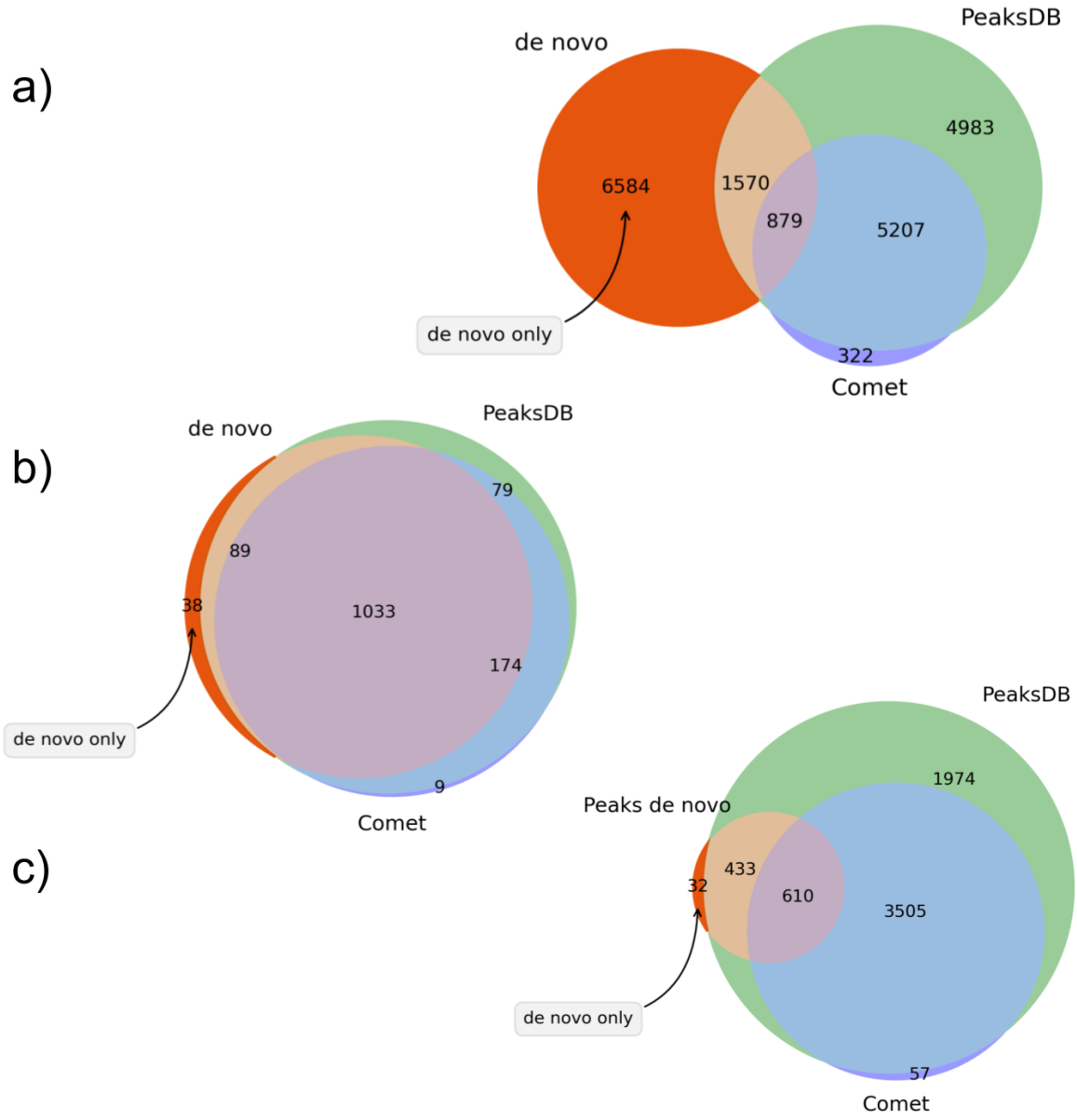

Figure 2.7. Comparison of (a) peptides, (b) proteins, and (c) *Prochlorococcus* species-level specific peptides found by Peaks *de novo* sequencing, PeaksDB *de novo*-directed database searching, and Comet database searching of tandem mass spectra from a *Prochlorococcus* culture. Peptides in this comparison (a) included modifications (such as methionine oxidation or asparagine deamidation) and were deduplicated so that replicate peptides are not counted. *Prochlorococcus* species-level specific peptides in (c) were identified by peptide lowest common

ancestor analysis of mapping to the UniProt Knowledgebase. Venn diagrams not to scale with respect to each other.

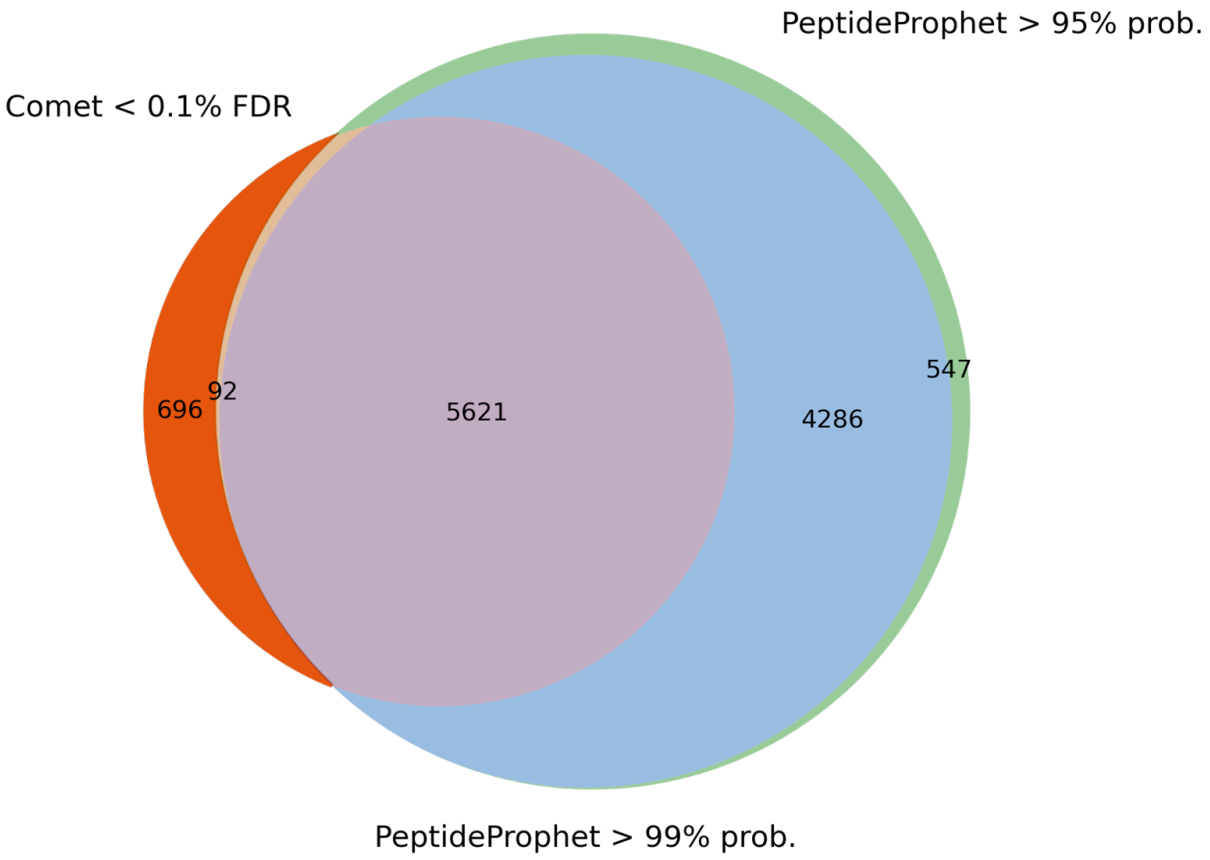

Figure 2.8. Comparison of deduplicated peptides found by Comet database searching of *Prochlorococcus* MED4 isolate MSMS spectra. Fewer Comet peptides were under the 0.1% FDR (XCorr >2.17) threshold had a PeptideProphet (Keller et al., 2002) probability >95%, 63% of all peptides were in agreement between the two metrics, and only 10% of peptides from the Comet search <0.1 % FDR (>XCorr 2.17) were not then validated by PeptideProphet <99% probability.

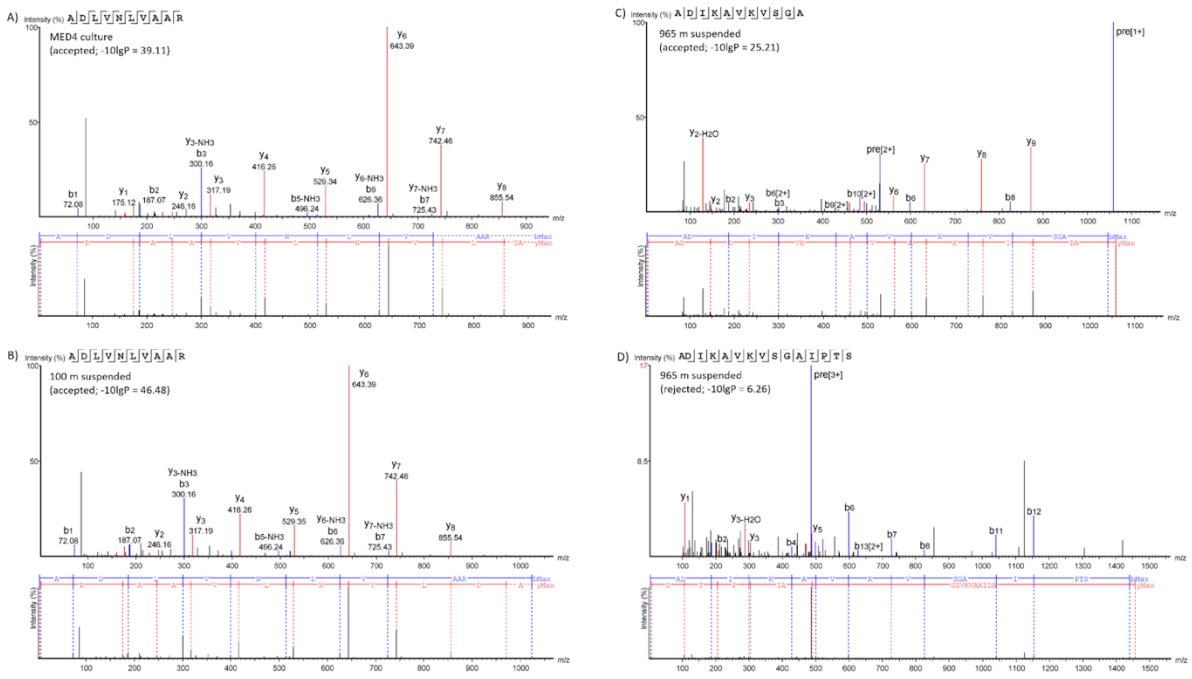

Figure 2.9. Tandem mass spectra of accepted (A, B, C) and rejected (D) peptide spectrum matches. The same peptide (ADLVNLVAAR) was found in both the A) *Prochlorococcus* and B) ETNP POM dataset in the 100 m suspended particles and is matched by UniPept to both *Prochlorococcus* and Synechococcus lineages. Peptide C (ADIKAVKVSGA) is a non-specific bacterial peptide found in numerous proteins. This spectrum match was accepted in PeaksDB *de novo*-driven database searching because it exceeded all the acceptable criteria thresholds (see methods). Visually, it can be observed that all the major ions except one (m/z 1060) are matched to either y or b ions of the corresponding database peptide. In panel D, the closest match to a peptide found using PeaksDB was rejected (ADIKAVKVSGAIPTS). This fit did not meet the stringent criteria (see Methods). Visually the fit can be seen to ignore several high intensity ions (m/z 853 and 1128) and is seeking matches to non-existent ions.

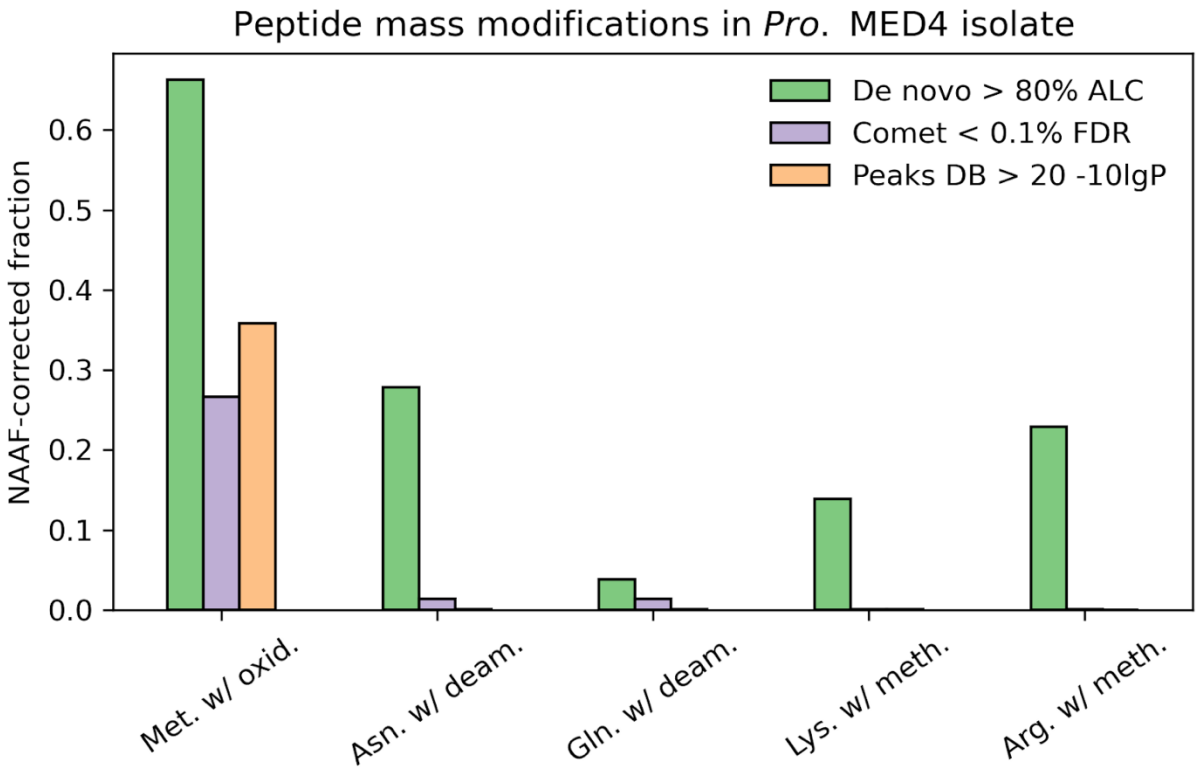

Figure 2.10. Fraction of relevant residues with variable modifications in PTM-optimized *de novo* sequencing (PEAKS), database searching (Comet), and *de novo*-directed database searching (PeaksDB) of MSMS spectra from a *Prochlorococcus* MED4 isolate. All peptides scaled by normalized area abundance factor (NAAF) before modification extent calculation. Variable modifications were selected from blind modification PTM search of the spectra using PeaksPTM (Han et al., 2011) (see Methods). Variable modifications ultimately selected were oxidation of methionine, deamidation of asparagine and glutamine, and methylation of lysine and arginine. Not shown is the single fixed modification of cysteine carbamidomethylation which is induced in the protein digestion protocol with addition of iodoacetamide. Mass differences and additional details about modifications can be found in Supporting Information Table S1.

Chapter 3. DEGRADATION OF DIATOM PROTEIN IN SEAWATER: A PEPTIDE-LEVEL VIEW[4]

ABSTRACT

Peptides and proteins were identified during a controlled laboratory degradation of the marine diatom *Thalassiosira weissflogii* by a surface seawater microbiome. Samples from each time point were processed both with and without the protease trypsin, allowing a partial differentiation between peptides produced naturally by microbial enzymatic degradation and peptides produced from the laboratory digestion of intact protein. Over the 12-day degradation experiment, 31% of the particulate organic carbon was depleted, and there was no preferential degradation of the *overall* protein pool. However, there was distinct differentiation in the cellular location, secondary structure and modifications between peptides produced by microbial vs. laboratory breakdown. During the initial period of rapid algal decay and bacterial growth, intracellular components from the cytoplasm were consumed first, resulting in the accumulation of membrane-associated proteins and peptides in the detrital pool. Accompanying the enrichment of membrane protein material was an increase in the importance of α-helix motifs. Methylated arginine, a post-translational modification common in cell senescence, was found in high amounts within the microbially-produced detrital peptide pool, suggesting a link between in-cell modification and resistance to immediate degradation. Another modification - asparagine deamidation - accumulated within the detrital peptides. Protein taxonomies showed the bacterial community decomposing the algal material was rich in Proteobacteria, and protein annotations showed abundant transportation of solubilized carbohydrates and small peptides across membranes. At this early stage of diagenesis, no changes in bulk amino acids (THAA) were observed, yet a proteomic approach allowed us to observe selective changes in diatom protein preservation by using amino acid sequences to infer subcellular location, secondary structures, and post-translational modifications.

---

[4] The research contained in this chapter is currently *in press* in Frontiers in Marine Science as '*Degradation of diatom protein in seawater: a peptide-level view*', with authors Megan E. Duffy[1], Cheyenne M. Adams[2], Khadijah K. Homolka[1], Jacquelyn A. Neibauer[1], Lawrence M. Mayer [2], and Richard G. Keil [1]

[1]School of Oceanography, University of Washington, Seattle, WA, 98105, USA

[2]Darling Marine Center, University of Maine, Walpole, ME, 04573, USA

3.1 INTRODUCTION

The application of 'omics' to marine systems has accelerated in recent decades and contributed to a deeper understanding of the roles microorganisms play in global biogeochemical cycles (Kleiner, 2019; Saito et al., 2019). Meanwhile, geochemical techniques have shown that some of the largest pools of carbon in the ocean are the detrital remains of these organisms, both in dissolved (Aluwihare et al., 1997; Benner & Kaiser, 2003; Jiao et al., 2010) and particulate organic matter (Hedges et al., 2001; Lee et al., 2004; Wakeham et al., 1997). Though largely unexplored, analytical advances are increasing the potential for 'omic' approaches to illuminate carbon cycling in marine detritus (Moore et al., 2014; Moore, Nunn, Goodlett, et al., 2012; Nunn et al., 2010).
Much of the identifiable detrital organic matter in the ocean is proteinaceous (Lee et al., 2004; Mayer et al., 1988; Wakeham et al., 1997). This composition may be unsurprising given the abundance of proteins in living organisms: diatoms, one of the major primary producers in the ocean, are approximately 25% protein by dry weight (Olofsson et al., 2019). Photoautotrophic bacteria, such as *Prochlorococcus*, can be as much as 70% protein (Finkel et al., 2016). Determining the origins and fates of protein in marine systems is therefore critical to a predictive understanding of carbon transport and storage in a changing ocean.

Proteins are often assumed to be labile and subject to degradation (Nunn et al., 2003; Pantoja et al., 2009). Yet, some proteins have been shown to survive long exposure in seawater (Dong et al., 2010; Keil & Kirchman, 1994) and have been identified in the detrital dissolved organic matter pool (Suzuki et al., 1997; Tanoue et al., 1995; Yamada & Tanoue, 2003), in coastal and open ocean marine sediments (Estes et al., 2019; Mayer et al., 1986; Mayer et al., 1988) and in sediment pore water (Abdulla et al., 2018; Fejjar et al., 2021; Schmidt et al., 2011). Algal proteinaceous material has been shown to survive microbial degradation in 4,000-year-old sediments (Knicker et al., 1996). Thus, the geochemical evidence implies that while many proteins or constituents of proteins may indeed be labile, some component of proteinaceous material has the potential to be preserved into the geologic record. However, the underlying reasons that certain proteins or their constituents are preserved while others are rapidly degraded remain largely elusive.

The aquatic geochemical literature is rich regarding protein and amino acid degradation. Heterotrophic bacteria first secrete free or tethered endoproteases to break larger proteins to

smaller peptides (Hollibaugh & Azam, 1983; Hoppe et al., 2002; Obayashi & Suzuki, 2008). Exopeptidases then cleave individual amino acids off the termini, as shown via mass spectral detection of marine detrital protein (Nunn et al., 2003; Roth & Harvey, 2006), and peptides under around 600 Da can be taken up by heterotrophic bacteria and some phytoplankton (Mulholland & Lee, 2009; Weiss et al., 1991). Encapsulation and/or sorption interactions with organic matter and inorganic mineral surfaces may limit bacterial enzyme access to (and degradation of) protein and proteinaceous compounds (Hedges et al., 2001; Knicker et al., 2001; Mayer, 1994; Nagata & Kirchman, 1996).

With the application of metaproteomics tools to marine detrital systems, peptides sourced from algal organelles and membranes have emerged as more resistant to rapid degradation in environmental and laboratory studies (Moore et al., 2014; Nunn et al., 2010; Yamada & Tanoue, 2003). Certain secondary structures such as α-helices and random coils appear to be retained within detritus while other motifs are degraded (Nunn et al 2010), and peptides containing modified amino acids have been shown to accumulate in detrital pools of marine organic matter (Duffy et al, submitted, Abdulla et al., 2018; Liu et al., 2010; Yamada & Tanoue, 2003).

Previous studies that apply metaproteomic techniques to marine detritus have generally added trypsin or other proteolytic enzymes to break proteins into smaller constituent peptides (Moore et al., 2014; Moore, Nunn, Goodlett, et al., 2012; Nunn et al., 2010). This proteolytic approach is typical in liquid chromatography-mass spectrometry (LC-MS) 'bottom-up' proteomics experiments which analytically require small, ionizable peptides for detection rather than bulky and complex intact proteins (Laskay et al., 2013; Saito et al., 2019). However, this analytical requirement may obscure intriguing dynamics of natural protein breakdown and digestion in the environment. In the present study I complement existing protein degradation studies by additionally looking for small peptides produced solely by microbial degradation during the experiment. This naturally-produced peptide pool represents a combination of endogenous small peptides and those that have been liberated from larger proteins but are not yet fully degraded. In this work, I leverage the amino acid sequences gained from metaproteomic techniques to interrogate the subcellular origins, secondary structures, modifications, and amino acid compositions of the degrading algal proteinaceous material over time.

3.2 MATERIAL AND METHODS

The experimental workflow can be divided into distinct phases: the algal degradation experiment itself; amino acid and peptide extraction; peptide identification via a combination of a database searching and *de novo* peptide sequencing approaches; and data analytics.

3.2.1 *Algal degradation*

A culture of the centric diatom *Thalassiosira weissflogii* was grown to approximately 1 x 10^6 cells/mL, concentrated by centrifugation, and frozen at -80 ºC to render the algal culture non-viable. Aliquots were thawed and resuspended to 2 g/L by dry weight using 1 µm filtered and UV sterilized seawater from the Damariscotta Estuary, Gulf of Maine. Unsterilized seawater filtered to 1 µm was used as an inoculum to induce bacterial decomposition of the algal detritus, with 1 mL added to 15 L of algal suspension. At each time point (1, 2, 5 & 12 days), 100 mL of the degradation slurry was vacuum filtered onto 25 mm diameter 0.3 µm pore size glass fiber membranes (Sterlitech GF75) at 4 °C and stored frozen prior to proteomic extraction and analysis. Further experimental details about the degradation can be found in Adams et al. (2019).

3.2.2 *Total hydrolyzable amino acids*

Approximately 30 mg of material scraped from a filter was hydrolyzed in 6 N HCl as described in Cowie & Hedges (1992). Amino acids were derivatized as in Gray et al. (2017) using the AccQ Tag Ultra derivatization kit from Waters (Milford, Massachusetts). Amino acids were separated and quantified via LC-MS using a full scan method in positive ion mode, with a scan range of 100 to 600 m/z and a resolution of 60,000, on a Thermo Scientific Q Exactive HF Orbitrap.

3.2.3 *Proteomic sample extraction*

Peptides were extracted with and without the use of the serine protease trypsin, which specifically cleaves sequences on the carboxyl side (or the "C-terminal side") of lysine and arginine, except when either is bound to a C-terminal proline. Peptides created via trypsin digestion are usually between 700-1500 Daltons (6-14 amino acids), an ideal range for mass spectrometry detection

(Laskay et al., 2013). Approximately 100 mg of material was scraped from a filter into detergent-free 50 mM ammonium bicarbonate lysis buffer with 5 mM ethylenediaminetetraacetic acid (EDTA). The chilled suspension was lysed via three cycles each of mechanical disruption with silica beads (50% 100 µm diameter and 50% 400 µm), freeze-thawing in a methanol-dry ice bath, and 30 seconds in a high-power water bath sonicator. The supernatant was removed after centrifugation at 4800 rpm and the protein concentration was determined using a modified Lowry assay using reagents from Bio-Rad (Hercules, California). Extracted protein was subjected to reduction of disulfides with dithiothreitol and carbamidomethylation with iodoacetamide. One half (300 µL) of the protein extraction was subjected to in-solution digestion with trypsin (Promega Gold) overnight at room temperature by a ratio of 1 µg trypsin: 25 µg total protein in 50 mM ammonium bicarbonate buffer with 5 mM EDTA. The remaining 300 µL of protein extraction was left untreated with trypsin. Both trypsin and no-trypsin treatments were desalted using NestGroup macro-spin C18 columns (Southborough, Massachusetts) and resuspended in 5% acetonitrile with 0.1% formic acid and Waters Hi3 E. coli peptide standard mixture (100 fmol/L).

Reverse-phase liquid chromatography-high resolution mass spectrometry (LC-HRMS) analysis was performed in duplicate with a Thermo Science (Waltham, Massachusetts) EASY-nanoLC system coupled to a Thermo Orbitrap Fusion Tribrid HRMS. Peptides were separated on a home-packed analytical column consisting of a 37 cm long, 75-µm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100˚A, 5mm; Michrom) coupled to a 4 cm long, 100 m i.d. precolumn (Magic C18AQ, 200˚A, 5mm; Michrom). Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100% LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 120-minute gradient from 5-35% solvent B. All analyses were carried out in positive mode at an NSI spray voltage of 2 kV. Data-dependent acquisition (DDA) on the top 10 ions was carried out using first higher energy collision dissociation (HCD) and then electron transfer dissociation (ETD) fragmentation methods for duplicate injections. The MS1 (parent peptide ion) scan range was 400-2000 m/z.

3.2.4 *Proteomic data analysis*

Peptides were identified from the raw mass spectra using a combination database-driven and database-independent *de novo* sequencing approach. *De novo* peptide sequencing, where the

amino acid sequence of peptides is determined solely from the mass spectra without comparison to a reference database (Allmer, 2011), was advantageous in this study because I lacked a paired metagenome to thoroughly characterize the microbial community from the seawater inoculum (Muth et al., 2015, 2018; O'Bryon et al., 2020). The combining of database and *de novo* is termed *de novo*-directed proteomics and was performed using PeaksDB within Peaks Studio (v8.5; Bioinformatics Solutions, Waterloo, Canada; Zhang et al., 2012). The *de novo*-directed approach has been shown to significantly improve sensitivity and accuracy in comparison to existing database search techniques (Zhang et al., 2012).

For database searches I used a reference protein database composed of 84,000 sequence entries predicted from transcriptomes of 8 *T. weissflogii* strains contained in the Marine Microbial Metatranscriptome Sequencing Project (NCBI BioProject PRJNA248394, Keeling et al., 2014). I added to the reference database two Gulf of Maine surface seawater metagenomes (Yooseph et al., 2007) from the Global Ocean Survey (GOS) as an aid in identifying the seawater microbes degrading the algal detritus. Additionally, I searched against a database of common mass spectral contaminants (Mellacheruvu et al., 2013). Search parameters for both database searching and *de novo* sequencing included 8 maximum modifications per peptide, 15 ppm peptide mass tolerance, and 0.5 Da fragment mass tolerance. For the trypsin-digested fractions, I performed both tryptic (maximum 2 missed cleavages) and non-enzymatic constraint searches, which means that all possible peptides up to 60 residues were considered. For the fractions not treated with trypsin, only non-enzymatic constraint searches were performed. Results from technical replicates and fragmentation strategies were combined.

Peptide identification confidences are calculated differently between the database search-identified and *de novo* sequenced peptides. For the *de novo* identifications, an amino acid-level confidence score is calculated based on mass deviation from spectral features and expressed as a percentage value. I accepted *de novo* peptides >80% average residue local confidence (ALC) with no single amino acid score <50%. For database searches, a false discovery rate (FDR) was set <1.0% using a reversed database target-decoy strategy (i.e., searching against reversed reference protein sequences) (Elias & Gygi, 2010). *De novo* sequencing was also incorporated into the database searches, as PeaksDB first compares *de novo* sequences to the reference database to find approximate matches and decrease the search space. Agreement between *de novo* sequences and database search hits are also used, in part, to generate peptide-level confidence scores derived from

the p-value indicating the statistical significance of the peptide-spectrum match (the -10lgP score). The threshold was a -10lgP score >20. Such a score is equivalent to a p-value of 1%, signifying the probability that the identification is to a false peptide sequence (J. Zhang et al., 2012). Protein identifications are notoriously difficult in samples containing many different organisms because some peptides are shared in proteins from multiple taxa. I required a minimum of 1 unique peptide per protein identification. Matching of proteins from peptides found only by *de novo* sequencing is detailed below.

3.2.5 *Searching for Amino Acid Modifications*

Since amino acids are frequently modified after translation, either for cell-specific purposes or during degradation, a computationally efficient method is needed to search for the myriad of possible post translational modifications (PTMs). I used the open modification search tool PeaksPTM (Han et al., 2011), with parameters set to tryptic or non-enzymatic constraint, 2 or fewer missed cleavages, 15 ppm peptide mass tolerance, 0.2 Da fragment mass tolerance, minimum A-score > 200 (a measure of modification location confidence), fixed carbamidomethylation of cysteine, and variable oxidation of methionine. Based on PeaksPTM results, the most frequently occurring modifications were used to evaluate whether adding additional PTMs to the overall searches altered the rate of false discovery. That is, I sought the optimal balance between searching for PTMs while avoiding a vast search space that leads to decreased sensitivity (Noble, 2015; Timmins-Schiffman et al., 2017). A series of PeaksDB searches and Peaks sequencing runs of the combined data set (using same search parameters as above) with increasing numbers of variable modifications was performed to find the optimal set of PTMs to include in searches (most peptide identifications <1% FDR). Using these PTM ramping results, 10 optimal variable PTMs were selected for the final searches. They included, in addition to methionine oxidation: deamidation of asparagine, methylation of arginine, oxidation of tyrosine, methylation of lysine, oxidation of lysine, oxidation of arginine, oxidation of proline, acetylation of lysine, and glutamine cyclization (the conversion of glutamine to pyro-glutamine).

3.2.6 *Mapping de novo peptides to proteins*

The *de novo*-directed PeaksDB workflow used here outputs peptides matched to the database and *de novo* peptides (sequences only, no additional information). To identify peptides that came from the diatom detritus or bacterial proteins not found by database searching, the *de novo* peptides were aligned to the reference database (*T. weissflogii* transcriptomes with GOS Gulf of Maine metagenomes) using PepExplorer (Leprevost et al., 2014). PepExplorer considers common *de novo* sequencing errors and limitations (such as leucine and isoleucine equivalence or other combinations of amino acids having the same mass) and identifies regions of local similarity between sequences. I also performed an alignment of the sequences against a reversed version of the database to estimate a false discovery rate, which was kept <1%. Alignments were accepted at a 95% identify agreement cutoff and protein identification required at least one unique peptide alignment.

Many *de novo* peptides did not match back to the reference database, which was not unexpected given the low number of sequences in the database for heterotrophic bacteria and the GOS sampling locations being more open ocean (Yooseph et al., 2007) compared to the estuarine seawater inoculum. To overcome this limitation, I additionally mapped the *de novo* peptides to proteins contained in the entire UniProt KnowledgeBase database (UniProt Consortium, 2018), which contains over 200 million sequences from thousands of taxa. The mapping was performed using the UniPept lowest common ancestor tool (Mesuere et al., 2016), which is built specifically for metaproteomic data and determines the taxonomic origins of peptides to the lowest possible phylogenetic rank (since some peptides are highly conserved, they may match as to only 'Bacteria', or even 'Organism', rather than to a species or genus level). The UniPept output provides the best view of the bacteria present in the experiment given the lack of genes or transcripts from which to build an ideal reference database.

3.2.7 *Gene Ontology Terms and Secondary Structures*

To identify gene ontology (GO) term annotations, the peptide sequences were aligned to the UniProt protein database using UniPept's metaproteomic functional analysis tool (Gurdeep Singh et al., 2019), which is built upon a lowest common ancestor peptide search described above. The

GO categories were condensed from the broader set in order to eliminate redundancy using the REViGO tool (available at http://revigo.irb.hr/) as well as manually.

Protein and protein secondary structures for all samples were estimated using the Proteus2 algorithm (Montgomerie et al., 2008) for the combination of proteins identified by PeaksDB database searching and *de novo* sequencing with database mapping with PepExplorer. Output is the highest likelihood of individual amino acids being part of the following structure classes: random coil, α-helices, β-strands, membrane α-helices, and membrane β-strands.

Spectral files, databases, and peptide identification (pepXML) files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD027843.

3.3 RESULTS AND DISCUSSION

3.3.1 *General characteristics of the degradation*

As previously described (Adams et al., 2019), algal biomass decreased by a factor of 2.4 during the 12-day incubation. Meanwhile, bacterial abundance increased by a factor of 5. Total particulate carbon decreased by 31% over the twelve days (Figure 3.1). In total, the algal detritus was partly remineralized via respiration and partly converted to bacterial biomass, but mostly remained as necromass (Adams et al., 2019). The degradation experiment started with a small inoculum of bacteria added to a detrital slurry that contained some algal proteases present during diatom cell harvesting. The algal detritus also likely contained small amounts of bacteria (or bacterial necromass) from the diatom culture, which was not confirmed to be axenic. The living bacteria then responded to the detritus and began to grow exponentially (Figure 3.1), during which time they released enzymes to break down and take up the diatom organic matter. The ratio of enzymatically hydrolyzable amino acids (a biomimetic measurement of protein bioavailability, see Mayer et al., 1995) to the organic carbon content remained constant throughout the experiment (Adams et al., 2019) indicating that there was no selective remineralization or preservation of proteins relative to other organic material at this early stage of degradation. However, by looking closely at peptides, I show that numerous compositional changes occurred in the protein pool during the 12-day period.

Our interpretation of protein dynamics during the degradation experiment relies on the distinction between two different pools of peptides: those produced *in situ* during the experiment (not treated with trypsin, which I will refer to as 'naturally digested peptides') and those produced artificially by the laboratory digestion (referred to as 'trypsin-digested peptides'). I hypothesize that the naturally digested peptides represent a mixture of proteinaceous material that is being accessed by the heterotrophic bacteria and peptides that have been released from larger proteins. The trypsin-digested peptides should represent the total of all digestible proteinaceous material present at each time point, including that which is not being actively degraded (e.g., the material that has resisted degradation thus far). Differences between these two different diatom protein-derived pools provides information about relative labilities of the substrate proteins, in addition to information about the bacterial strategies of degrading the diatom necromass.

Both the trypsin and naturally digested peptide pools should also contain peptides sourced from the bacteria growing in the experiment. Given cell counts and estimated carbon per diatom and bacterial cells, there is about 150 times more algal protein than bacterial protein in the day 0 time point, and this decreases to 15 times more after 12 days (Adams et al., 2019). These peptides can be used to determine the types of bacteria present during the experiment, and the cellular functions that these bacteria are performing (e.g. Bergauer et al., 2017; Mikan et al., 2020). In the following sections I examine the peptide results with the goal of trying to better understand the degradative processes occurring over the course of experiment and their implications for detrital protein cycling in the marine systems.

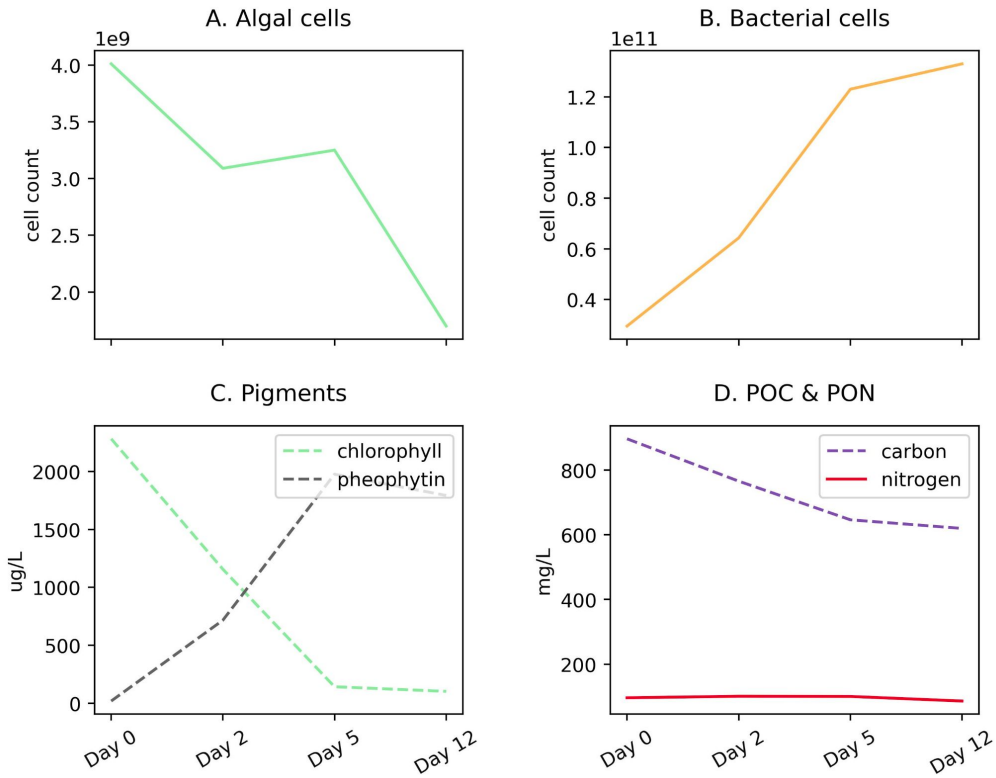

Figure 3.1. Progression of *Thalassiosira weissflogii* degradation experiment. (A) algal and (B) bacterial cell counts over the 12-day incubation period; (C) chlorophyll and its degradation product, pheophytin; and (D) particulate carbon and nitrogen (filtered on GF/F membranes). For more information about algal degradation see Adams et al., 2019.

3.3.2 *Algal peptide identification and characteristics*

The naturally digested and trypsin-digested peptides are methodologically distinct yet analytically overlapping pools. The naturally digested peptides are present in the background of the trypsin-digested fraction, since the incubation preceded proteomic sample extraction and laboratory digestion. The mass window of MS1 detection is 400-2,000 m/z. Typically, peptides produced via trypsin digestion are between 6-14 amino acids in length; in this data, the average trypsin-digested peptide length was 10.3 amino acids, and the average peptide (MS1) mass was 1114 Da. By contrast, the average length of peptides in the naturally digested fractions was 10.1 and the average mass, 1082 Da. Larger polypeptides created by microbial degradation processes, if large enough, would likely have sites available to cleavage by trypsin. Thus, there is additional overlap between these two pools because some larger naturally digested peptides are further subjected to trypsin digestion in the trypsin-digested fraction.

Table 3.5. Diatom and bacterial peptide and protein identifications.

Proteomic time point and treatment	PSM diatom peptides[a]	De novo diatom peptides[b]	Diatom proteins[c]	Bacterial peptides[d]
trypsin-digested		*laboratory trypsin digestion*		
Day 0	1217	585	528	404
Day 2	1537	720	736	494
Day 5	820	370	321	389
Day 12	641	437	265	500
naturally digested		*natural microbial digestion*		
Day 0	252	130	112	86
Day 2	2125	337	231	265
Day 5	232	116	78	117
Day 12	55	165	43	147

[a] <1% peptide false discovery rate

[b] with at least 1 unique peptide ID by database or de novo sequencing

[c] >80% average local residue confidence and protein mapping using PepExplorer at 95% similarity and 1% FDR.

[d] by >80 ALC de novo and lowest common ancestor analysis with UniPept

The number of identified diatom peptides and proteins peaked at day 2 during the initial phase of exponential bacterial growth, and then decreased to lowest values at day 12 (Table 3.5) I sorted the algal peptides by their gene ontology cellular compartment terms (GO Terms; Figure 3.2; see also (Mikan et al., 2020; Riffle et al., 2017). The data were aggregated into eight major cellular compartments, and the three numerically dominant groups were chloroplast, cellular membrane,

and cytoplasmic proteins (Figure 3.2). The trypsin-digested day 0 fraction had a cellular component GO term distribution consistent with that of living diatom cells: alongside the degradation experiment's algal GO terms I plot those from peptides obtained from two *Thalassiosira pseudonana* proteomes published by Dyhrman et al., 2012 (Figure 3.2).

The naturally digested peptides contained far fewer GO term identifications after day 2, and by day 12 the only identifications in that fraction were from chloroplast and membrane peptides. This indicates that peptides and proteins from other cellular components were effectively degraded during the initial phase of bacterial exponential growth (Figure 3.1) and were not being actively degraded at day 12 because they were no longer present. Consistent with this membrane recalcitrance hypothesis is that the number of identifications of trypsin-digested peptides sourced from larger proteins also decreased for most cellular compartments other than membranes and the chloroplast (Figure 3.2). However, there were still some trypsin-digested peptides at day 12 for components from the mitochondria and cytoplasm, suggesting that perhaps cellular location alone does not necessarily determine whether a protein or peptide will survive the initial stages of environmental degradation.

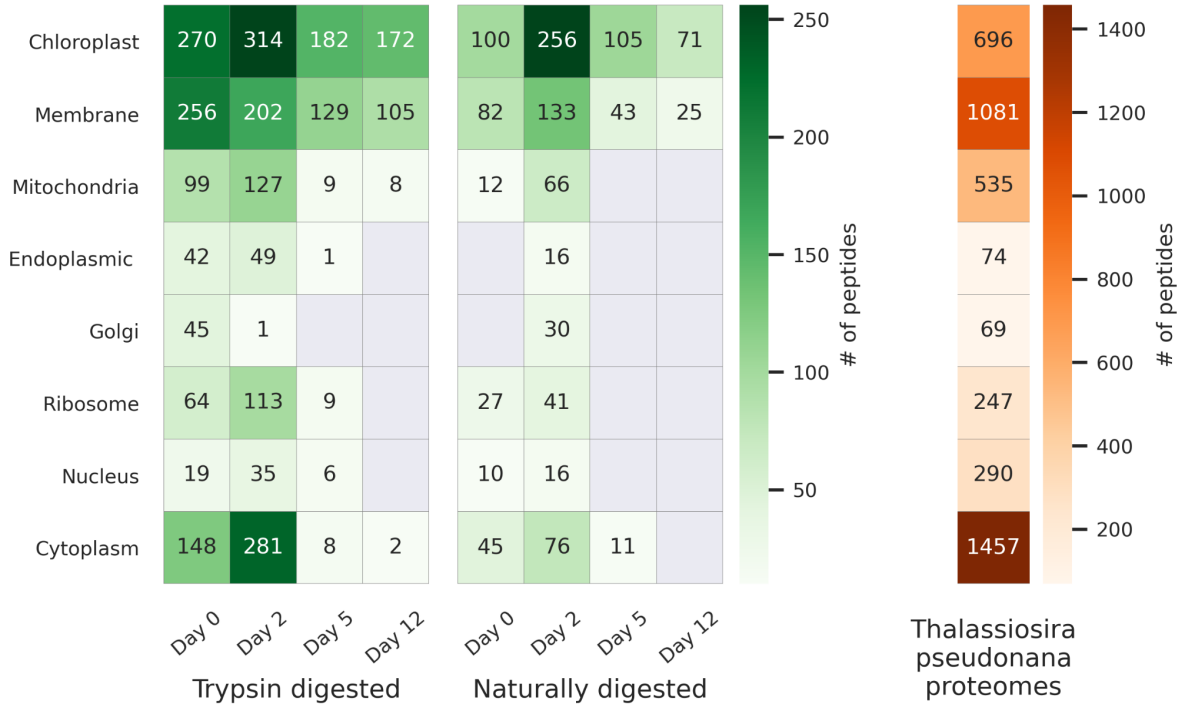

Figure 3.2. Heatmap showing the number of cellular compartment gene ontology (GO) terms specific to diatom peptides over a 12-day degradation experiment. Trypsin-digested (left panel) and naturally digested fractions (digestion only by natural microbial community, right panel)

shown separately. Shown as context are peptides identified in two proteomes of *Thalassiosira pseudonana* cultures from (Dyhrman et al., 2012) (in orange). GO terms shown were condensed from a broader set to eliminate redundancy for ease of visualization using REViGO (http://revigo.irb.hr/) and further manually organized into broad categories.

The number of trypsin-digested peptides of cytoplasmic location also dropped precipitously after day 2, while chloroplast and membrane peptides remained abundantly detectable throughout the 12-day experiment (Figure 3.2). This indicates a shift in the detrital protein pool from one that mimics an intact living cell to one that is dominated by proteins that are associated with membranes. These findings are consistent with Laursen et al. (1996), who showed membrane proteins to be the more refractory fraction of phytoplankton protein by physical separation of subcellular fractions of phytoplankton cells. Their study reported higher proteolysis rate constants for the cytoplasmic fraction (>1.2 h-1) than for the membrane fraction (0.2 to 1 h-1), which correlated negatively with the ratio of chlorophyll to pheophytin (Laursen et al., 1996).

Studies of algal nutritional value to zooplankton and other animals also suggest preferential soluble protein consumption. In an evaluation of elemental uptake from diatom detritus, Reinfelder and Fisher (1991) showed that metal assimilation efficiencies of marine copepods were directly related to the cytoplasmic content of diatoms. This relationship indicates that the copepods sourced nearly all their nutrition from the diatom cytoplasm rather than from other cellular constituents (Reinfelder & Fisher, 1991). This experiment didn't involve zooplankton grazing but does indicate that membrane protein solubility or structure are factors in differential consumption by bacteria.

3.3.3 *Preserved protein motifs*

To further interrogate the proteinaceous material that survived the degradation, five classes of peptide secondary structure were evaluated: random coil, α-helices, β-strands, transmembrane α-helices, and transmembrane β-strands. Closely set α-helices contain strong hydrogen bonding between weakly polar (Ser, Thr, Cys) and strongly polar (Asn, Gln, Glu, Asp, His, Arg, Lys) amino acid residues of neighboring α-helices (Liu et al., 2011). Helices are commonly found in transmembrane proteins (Sakai & Tsukihara, 1998; Stevens & Arkin, 2000), and their ability to bend can account for the hydrophobic mismatch of the lipid bilayer (Yeagle et al., 2007). In contrast, β-strands lay relatively flat and have been hypothesized in a marine context to adhere to

mineral surfaces - ultimately aiding to their protection and resulting in enhanced preservation (Shamblin et al. 1998; Oleschuk et al. 2000; Ovesen et al. 2011). 'Random coil' is not a true secondary structure, but rather is an aggregate term for short sequences where there is an absence of a helix or sheet character (Smith et al., 1996). Benchmarking of the Proteus2 machine-learning algorithm I use here showed per-residue prediction accuracy to be 87-91% for transmembrane α-helices, 86% for transmembrane β-strands, and 88% for non-membrane secondary structures (Montgomerie et al., 2008).

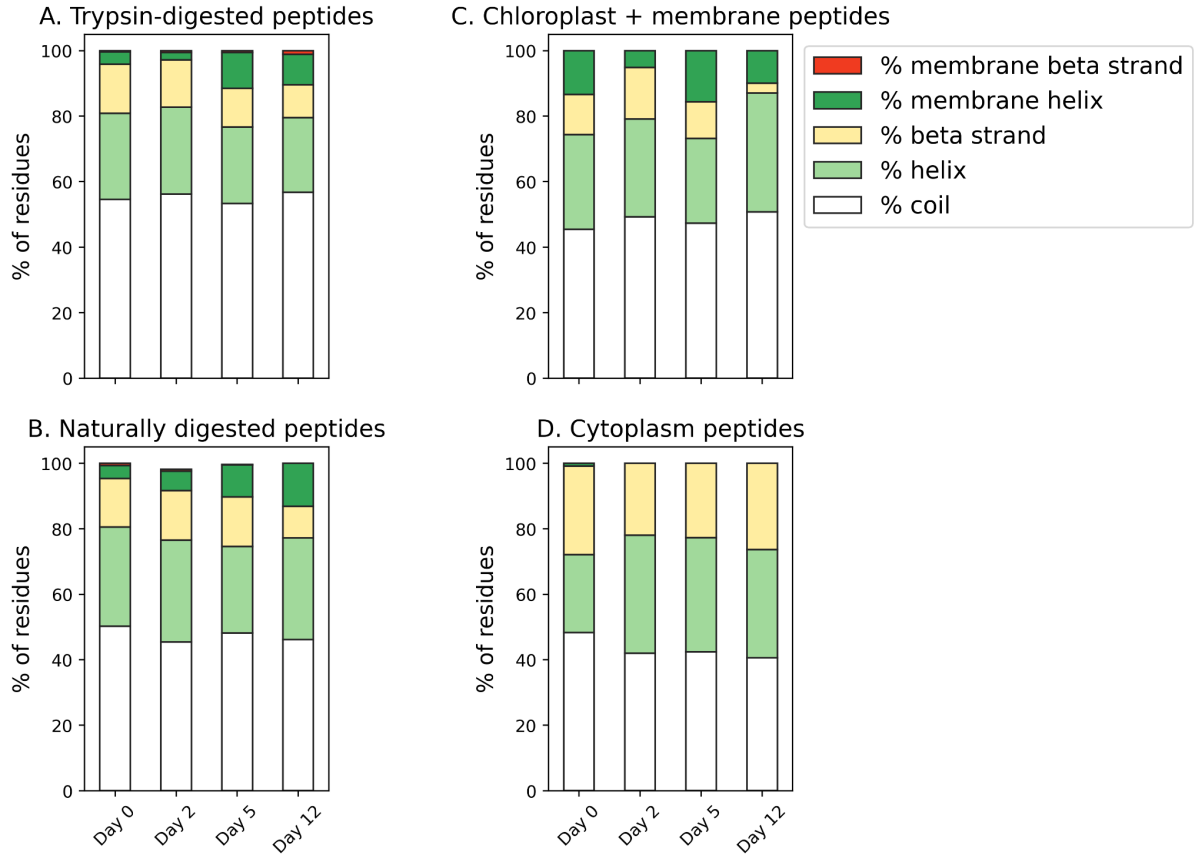

Figure 3.3. (A-B) Secondary structure predictions of algal peptides identified in trypsin-digested and naturally digested fractions at four points during the 12-day degradation experiment. Secondary structure motifs (coil, α-helix, β-strand, membrane α-helix, membrane β-strand) were predicted from full protein sequences using Proteus2 (Montgomerie et al., 2008) and the relative contribution of motifs determined by the identifying peptide's normalized area abundance factor (NAAF). (C-D) Secondary structure predictions of just algal chloroplast and membrane peptides and algal cytoplasmic peptides that are identified at each time point of the

degradation. For this comparison, proteins identified in the trypsin-digested and naturally digested fractions were combined and NAAF-adjusted.

I note a progressive change in secondary structure distribution as the degradation proceeded, with membrane α-helices becoming more important (Figure 3.3). At the same time, β-strands became a less common motif. This trend in secondary structure distributions is consistent with the GO term evidence indicating that membrane proteins are preferentially retained in the system and suggests that their tightly wrapped, difficult to denature, secondary structure could be a factor aiding in preservation. Random coils are the most common motif, and they do not change in relative importance over time, indicating that they are not particularly prone to resistance or degradation in this experiment, or that the coil category is too broad to capture any selective processes. These results are generally consistent with those of Nunn et al. (2010) in a similar degradation of a marine diatom. They identified 23 and 4 diatom peptides after 10 and 23 days of incubation with a seawater microbiome, respectively. Of the surviving 4 diatom peptides at the end of the experiment, 2 had transmembrane domains (Nunn et al., 2010).

To further investigate the theory that secondary structure is related to enhanced preservation of membrane protein components, I also compared the predicted secondary structure of just algal chloroplast/integral component of membrane peptides and algal cytoplasmic peptides that are identified at each time point of the degradation (Figure 3.3). Chloroplast and membrane peptides, the bulk of diatom peptide identifications on day 12 (from 167 individual proteins across both trypsin and naturally digested fractions), were 50.7% coil, 36.3% α-helix, 3.02% β-strand, 9.93% transmembrane α-helix, and 0% membrane β-strand. In comparison, the cytoplasmic peptides (from 98 proteins across both trypsin and naturally digested fractions) identified in the initial day 0 proteome were comparatively more enriched in β-strands and depleted in α-helices and transmembrane α-helices, with 48.3% coil, 23.8% α-helix, 27.0% β-strand, 0% transmembrane α-helix, and 0% transmembrane β-strand. Supporting the theory that α-helices are linked to preservation is that the few surviving cytoplasmic peptides on day 12 (3 proteins total, all in the trypsin-digested fraction) had 10% more α-helix character than the initial day 0 cytoplasmic peptides, at 40.6% coil, 33.1% α-helix, 26.3% β-strand, 0% transmembrane α-helix, and 0% transmembrane β-strand (Figure 3.3). These findings are consistent with the notion that

hydrophobic interactions appear to be important in preserving membrane-associated proteins and peptides during early diagenesis (Nguyen & Harvey, 2001; Nunn et al., 2010).

While cellular location and secondary structure may be significant components that allow certain proteins to resist the early stages of degradation, two other factors have been observed in the literature; an abundance of post-translationally modified amino acids in degradation-resistant material and the enrichment of certain amino acids, like glycine, in the recalcitrant proteinaceous pool (by measurement of total hydrolyzable amino acids).

3.3.4 *Post translational modifications of amino acids*

Protein post-translational modifications (PTMs) such as oxidation, phosphorylation, and methylation, play critical roles in a diverse range of biological processes like signaling, protein activity and transport, and regulation of gene expression (Cain et al., 2014; Shen et al., 2008). PTMs are also associated with in vivo protein recycling and cell senescence (Cain et al., 2014; Dhillon & Denu, 2017). For these purposes, the 'PTM' umbrella includes modifications to amino acids that could occur in detritus, including those due to protein consumption or abiotic transformations.

In the marine environment, PTMs have been linked to degradation and early diagenesis. For example, the amino acid beta alanine accumulates in the hydrolyzable phase of marine sedimentary organic matter via modification of aspartic acid (Cowie and Hedges 1994). Modification of the nitrogen-containing side chains of glutamine, asparagine and arginine can lead to the accumulation of peptides containing deaminated side chains within anoxic marine sediment pore waters (Abdulla et al., 2018). I recently observed deamidated peptides in the sinking POM from the eastern tropical North Pacific oxygen deficient zone (Duffy et al., *submitted*), where Van Mooy et al. (2002) hypothesized that amino acids could be selectively deaminated in order to provide reduced nitrogen to fuel chemoautotrophic processes (Van Mooy et al., 2002). In laboratory experiments, Keil and Kirchman showed that methylated peptides were accessed less efficiently by bacteria than non-methylated peptides (Keil & Kirchman, 1992). Glycan modifications of proteins contribute to stability, a mechanism that's been used to increase the shelf life of protein and peptide-based pharmaceuticals (Zhou & Qiu, 2019). Indeed, Keil and Kirchman showed that glycosylated ribulose-1,5-bisphosphate (RuBisCo) was degraded 100 times more slowly than its unmodified counterpart (Keil & Kirchman, 1993). Altogether, the marine literature

suggests that PTMs are highly relevant to protein degradation, as the by-products of protein degradation processes and/or as potential factors in degradation resistance.

In the present study, I first performed a non-discriminatory 'open' search of thousands of possible PTMs in the UniMod database using PeaksPTM (see above in Methods). I then selected only the 10 most frequently occurring mass modifications in the open search results for the suite of PTMs in the ultimate database and *de novo* peptide searches. My goal was to identify patterns in PTM distribution to learn more about their roles in degradation and preservation. To that end, I wanted to distinguish between PTMs present in algal protein prior to degradation and PTMs associated with the degradative process. I compared the PTMs observed in the naturally digested peptide pool, which should represent proteins being actively degraded by the bacteria, against the trypsin-digested peptides. This comparison was done for only the algal protein; no bacterial proteins were included in this differential analysis of PTMs (Figure 3.4).

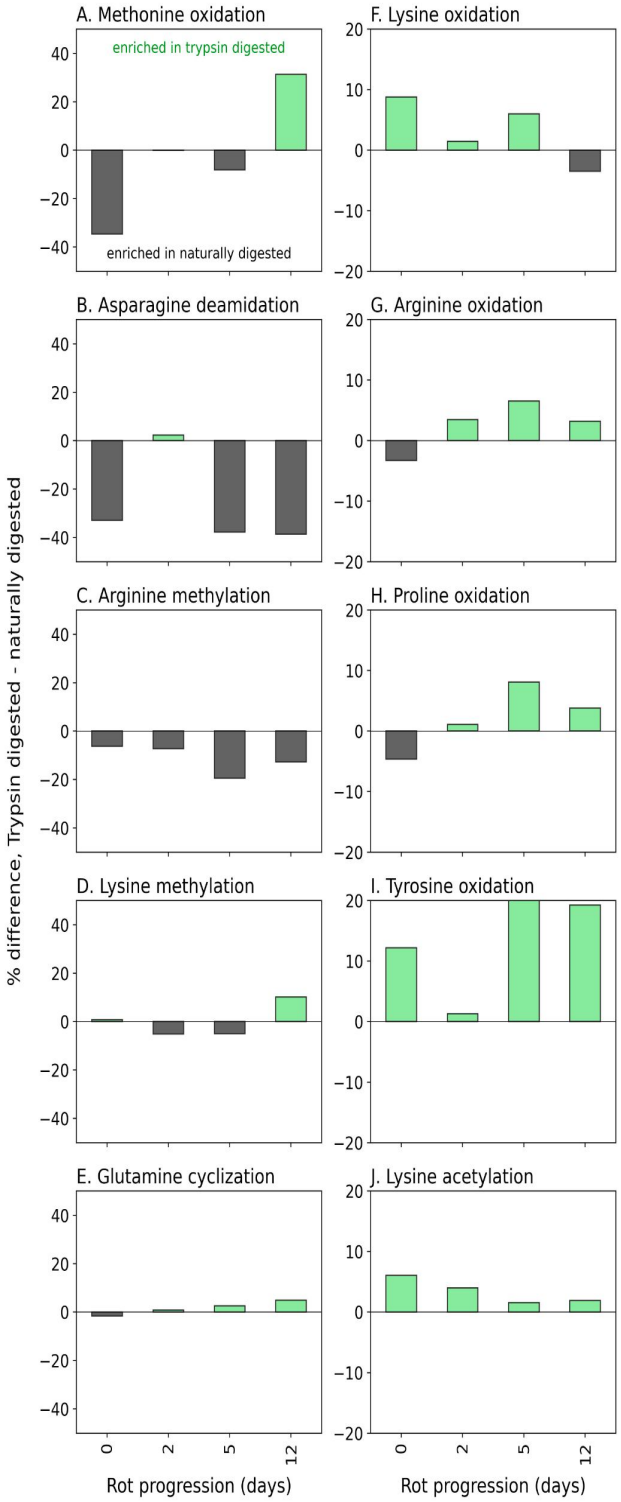

Figure 3.4. Post-translational modifications (PTMs) of algal peptides at four time points along the 12-day incubation. Shown are the difference between trypsin-digested naturally digested peptides for the 10 variable modifications included in the database searches and *de novo*

sequencing parameters, which were selected from preliminary open modification searches of the mass spectral data (see Methods): (A) methionine oxidation, (B) asparagine deamidation, (C) arginine methylation, (D) lysine methylation, (E) glutamine cyclization (pyro-glutamation), (F) lysine oxidation, (G) arginine oxidation, (H) proline oxidation, (I) tyrosine oxidation, and J) lysine acetylation. PTM distributions are expressed as the differences in percent of residue modification in the entire peptide pool (database and *de novo* sequences) as corrected by normalized area abundance factor (NAAF). Thus, positive values indicate the PTM is relatively enriched in the trypsin-digested peptides, and negative values mean the PTM is relatively enriched in the naturally digested peptides. Bacterial peptides were not included in the PTM analysis. Y-axis scales are the same between columns (A-E; F-J).

Oxidation PTMs were generally enriched in the trypsin-digested component, implying that these modifications occurred within the cell prior to heterotrophic attack. Generally, they did not increase in relative abundance over time in the trypsin-digested peptide pool, suggesting that oxidations don't meaningfully impact the lability of proteinaceous material (Figure 3.4). It's notable that during the experiment, chloroplast protein-derived peptides increasingly became dominant among identifications (Figure 3.2). The algal culture was exposed to light before the degradation experiment in the dark. Photosynthesis as an oxygenic process produces active oxygen species and radicals which can cause damage to cells. Oxidations of amino acids are frequent PTMs in photosynthesis-associated proteins (Aro et al., 2005; Galetskiy et al., 2011). Similarly, lysine acetylation of chloroplast proteins has been demonstrated in plants (Lehtimäki et al., 2015) and in diatoms (X.-H. Chen et al., 2018), though the PTM frequently occurs elsewhere in the cell (Z. Chen et al., 2018). The abundance of thylakoid membrane and chloroplast-associated proteins that accumulate due to preferential preservation (Figure 3.2) is likely why oxidations and lysine acetylations are dominant PTMs of the trypsin-digested peptides (Figure 3.4).

In contrast, two PTMs were strongly associated with the presumably more detrital naturally digested peptides: deamidation of asparagine and arginine methylation (Figure 3.4). This observation that asparagines and arginines are more modified in detrital proteins has several possible explanations. These modifications could have occurred within the living cell, resulting in a pool of protein that was easily accessible, which is why they were found more in the naturally

degraded peptide pool. Alternatively, these PTMs might have been created during the degradation process and could accumulate because once created they are further degraded slower than their unmodified counterparts. This latter explanation would account for the effective accumulation of deamidated peptides in both sediment pore waters (Abdulla et al., 2018) and in sinking particulate matter (Duffy et al., *submitted*). Thus, I hypothesize that deamination occurs during degradation, though continued research is warranted.

Unlike deamidation, methylation most likely occurs within the living cell, where it is a common PTM that is used as a control of numerous cellular functions (Ghesquière et al., 2011). Methylated peptides have been shown to be inefficiently assimilated and degraded by heterotrophic bacteria (Keil & Kirchman, 1992), suggesting that PTMs produced within living cells may also play a role in determining the rate of protein degradation.

The open PTM searches did not yield high levels of protein glycosylation, a very broad classification of modification that entails a sugar-amino acid bond. Though glycosylation is an extremely common enzymatic PTM in both eukaryotes and prokaryotes, the non-enzymatic form of the reaction (sometimes called 'glycation', or the Maillard reaction) has been posited as a mechanism of environmental protein preservation via geopolymer formation (Collins 1992, Burdige 2007). Glycosylation has been shown to decrease microbial protein degradation rates in incubations (Keil & Kirchman, 1993) and has been found in high abundance in seawater (Yamada & Tanoue, 2003). Peptides with glycosylations are difficult to ionize and detect under the mass spectral conditions I used here (Alley et al., 2013), and to evaluate them more accurately would require different proteomics preparation techniques (e.g., as performed by Moore et al., 2014). Given the findings here and in the literature, more targeted work is needed to better understand the effect that PTMs have during the early stages of protein degradation.

3.3.5 *Amino acid compositions*

One of the deepest sets of literature related to protein degradation in marine systems is that of the 'total hydrolyzable amino acid' pool (Dauwe et al., 1999; Dauwe & Middelburg, 1998; Lee et al., 2004; Wakeham et al., 1997). THAA analyses show clear trends during long-term carbon degradation and preservation including an accumulation of the amino acids glycine, serine and threonine (Dauwe and Middelburg 1998), and the creation of the non-protein amino acids beta-alanine and gamma-aminobutyric acid from aspartic and glutamic acids (Cowie and Hedges 1994).

Despite the widespread use of degradation indices derived from THAA analyses, it has been difficult to reconcile changes in bulk amino acid compositions to known protein amino acid compositions, especially during the very early stages of degradation (Keil et al., 2000). I compared the amino acid composition of the peptides identified during the degradation experiment to the THAA pool, which I measured independently. To facilitate this comparison, I plotted mole fractions of amino acids in the THAA against those in the identified algal peptides, combining the naturally and trypsin-digested amino acid compositions adjusted for relative peptide abundance (Figure 3.5). The near 1:1 agreement between the two approaches indicates two things: a) the protein amino acid compositions measured using the newer 'omic' approach can be effectively integrated into the large body of literature based on THAA analysis, which will become useful as peptidomic approaches are applied to samples further along the degradation pathway (e.g. sediment samples); and b) while the 'omic approach identifies specific ways in which protein undergoes degradation in the ocean, the early stages remain remarkably 'nonselective' at the bulk molecular level (Hedges et al., 2001).

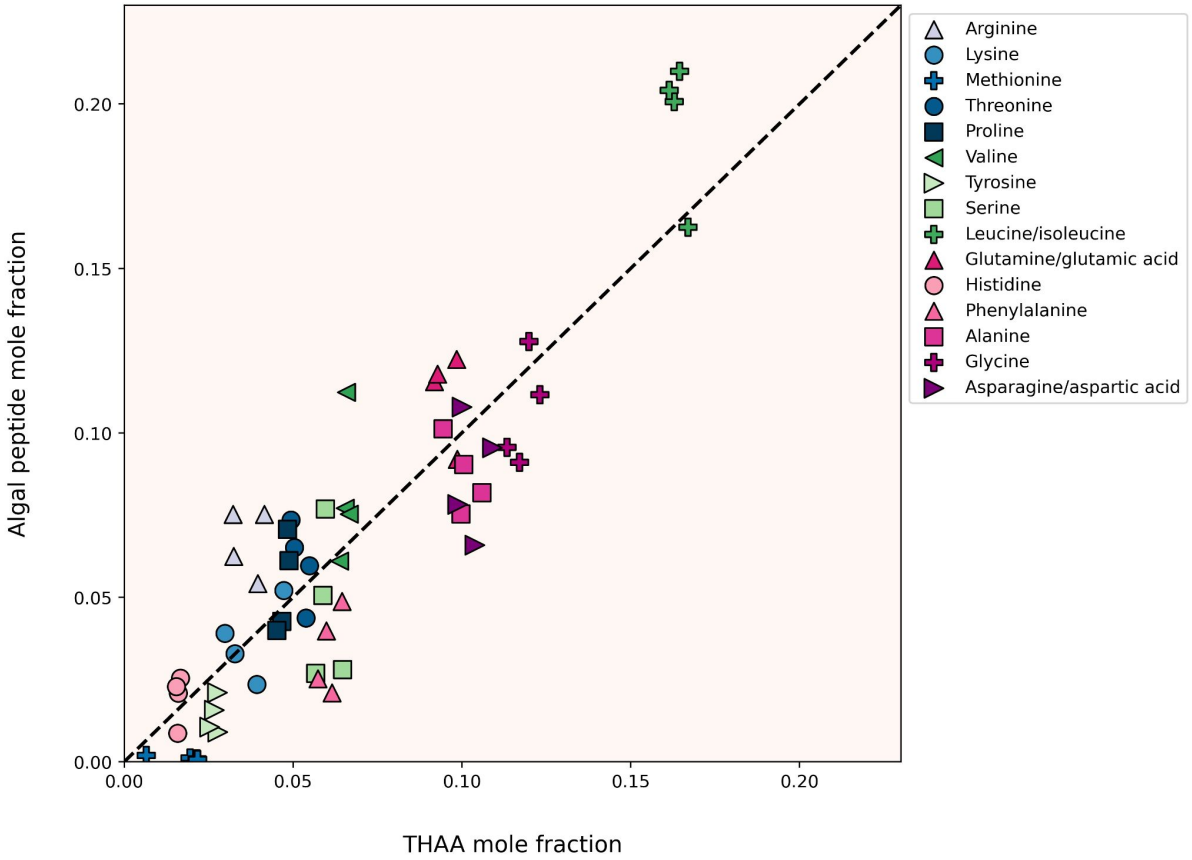

Figure 3.5. Mole fractions of individual amino acids from all four degradation time points as derived from total acid hydrolyzable amino acid analysis (THAA, x-axis) and tandem MS/MS-based proteomics (Peptide, y-axis). The dashed 1:1 line represents perfect agreement between approaches. The peptide amino acid compositions plotted here are derived from both trypsin-digested and naturally digested proteomics fractions. Label-free peptide quantitation was determined by the normalized area abundance factor (NAAF).

3.3.6 *Bacterial community*

The microbial inoculum for the degradation experiment was sourced from the Damariscotta River Estuary, whose marine waters come from the Gulf of Maine. Heterotrophic bacteria grew exponentially through the middle stages of the experiment (Figure 3.1). A peptide-based lowest common ancestor analysis was performed using UniPept (Gurdeep Singh et al., 2019) to assign bacterial taxa to the lowest phylogenetic level possible. The taxonomic hits were then adjusted to account for peptide spectral abundance and aggregated at the class level (Figure 3.6a). There is

precedence for using label-free metaproteomic data for microbial biomass determinations (Kleiner et al., 2017), and here I use the approach to learn about the community of microbes degrading the algal detritus.

There is taxonomic overlap of the initial composition in this study to that of a pyrosequencing survey of planktonic microbes at three stations in the Gulf of Maine (Li et al., 2011). I found the microbial community at the initial time point was dominated by Gammaproteobacteria (~60% of the peptides), with notable contributions of Cytophaga (~5%), Bacteroidia (~2%) and Alphaproteobacteria (~2%) (Figure 3.6a). During the 12-day experiment, the microbial community composition changed minimally, with two exceptions: Firstly, there was a ten-fold increase of peptides that were bacterial but could not be uniquely identified at the class level or below, from 2.2% of peptides on day 0 to 22% of peptides by day 12. At the same time, the contribution of low abundance (<0.5% of all peptides) classes increased from 0.53% on day 0 to 5.4% on day 12 (Figure 3.6a). Both increases, in non-specific bacterial peptides and in low-abundance class peptides, are potential indications that the diversity of the bacterial community was increasing during the experiment.

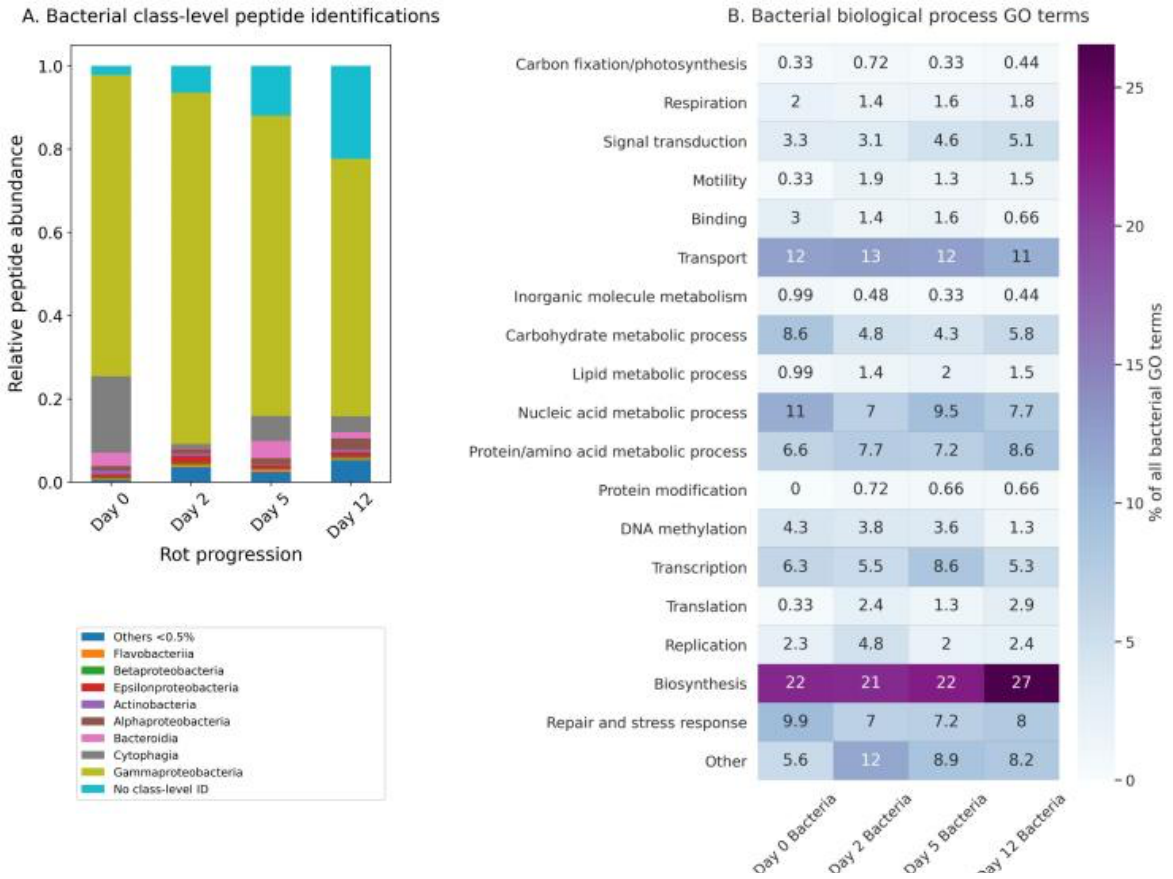

Figure 3.6. Bacterial progression of the 12-day algal degradation: (A) relative abundance contribution of the major bacterial taxonomic classes (>0.5%) from the time point proteomes adjusted by relative peptide spectral abundance (NAAF); (B) the percentage of the total number of bacterial peptide biological process gene ontology (GO) terms. GO terms shown were condensed from a broader set to eliminate redundancy for ease of visualization using the REViGO (available http://revigo.irb.hr/) and further manually organized into broader categories.

The increase of non-specific bacterial peptides can be explained in two ways. Most likely, a true increase in diversity and richness resulted from a broader array of shared, conserved proteins being produced in the system by related bacteria. An alternative hypothesis is that there was an ingrowth of bacterial necromass generating degraded bacterial peptides that are no longer very taxonomically specific. Bacterial necromass in this case could be from organisms in the seawater inoculum or bacteria present during the diatom culture growth. For instance, Gammaproteobacteria are known to thrive within the phycosphere of diatoms (Amin et al., 2012), and their peptide

abundance peaked at day 2 when the active bacteria were growing exponentially. It is difficult to discern which of these hypotheses is correct and further work will need to be done to evaluate the processing of bacterial detritus by other bacteria. Nonetheless, the overall taxonomic changes observed in the bacterial peptidome during the twelve days were minimal.

Unlike the somewhat ambiguous taxonomic information within the bacterial peptidome, biological process GO terms associated with the bacterial peptides provide a clearer view of how the bacterial community responded to the algal detritus (Figure 3.6b). The most detected GO terms are associated with transmembrane transport, carbohydrate metabolism, and DNA replication and transcription (Figure 3.6b). These are the terms most strongly associated with bacterial growth (Mikan et al 2020) and suggest that most of the bacterial peptides detected are from living bacteria. This strengthens support for the hypothesis that the Gammaproteobacteria in the samples are for the most part living and not detrital.

The bacterial functional data are generally consistent with recent work by Mikan et al. (2020), who used metaproteomic tools and a GO-term based functional analysis to evaluate the heterotrophic bacterial response to a pulse of detrital organic matter in two Arctic microbiomes during 10-day shipboard incubations. In that study, the bacterial community increased protein synthesis, carbohydrate degradation, and cellular redox processes while simultaneously decreasing C1 metabolism (Mikan et al., 2020). Throughout this experiment I observe steady levels of transmembrane transport and protein metabolism terms. Carbohydrate metabolic process GO terms maximize in the first two days of the degradation, with increasing biosynthesis GO terms by day 12 (Figure 3.6b). Mikan et al (2020) suggested that the bacterial community shifted their carbon acquisition strategies intracellularly before there were large shifts in the taxonomic structure of the community. Without a paired metagenome or metatranscriptome with which to perform proteomic database searches, these bacterial peptide data and paired GO term data are not as complete, but show the same general trends, lending further support to the notion of Mikan et al. that functional composition and redundancy, not taxonomy, may be the most relevant factor when evaluating how effectively organic matter is or will be processed by bacteria in the ocean.

3.4 Conclusions and future directions

In this study I evaluated proteinaceous material in a marine system both with and without the use of trypsin as an extraction and identification tool. To my knowledge, this is the first such attempt

to disentangle proteins that are being degraded by microbes from those that are resistant to degradation. I show that proteins from specific cellular locations are preferentially preserved during the initial stages of degradation. As has been hypothesized and demonstrated for bacterial membrane proteins (Jiao et al., 2010; Kaiser & Benner, 2008; Nagata et al., 1998) and algal membrane proteins (Laursen et al., 1996; Nguyen & Harvey, 2001; Wolfe et al., 2006), I conclusively illustrate that proteins associated with diatom chloroplasts and membranes resist initial degradation better than those without such association. The many membrane-associated and few cytoplasm peptides that resist degradation also are relatively enriched in α-helices and depleted in β-strands, consistent with the cellular location data since α-helices are enriched in membrane proteins. However, the extent to which α-helices themselves lead to degradation resistance remains to be more thoroughly evaluated, as 1) α-helices are not exclusive to membrane proteins and 2) there could be other reasons for membrane protein survival over time that causes membrane proteins' α-helix rich motifs to become enriched in detrital material.

The novel application of proteomics without the use of trypsin also allowed for the evaluation of how post translational modifications (PTMs) relate to protein degradation. I found that the oxidation and acetylation PTMs observed likely originated within the living cell, with only asparagine deamidation and arginine methylation being predominantly associated with the degraded peptide pool. I hypothesize that PTMs have an impact on the bioavailability of peptides during early diagenesis, but again more work is needed to evaluate the extent to which PTMs provide protection. In all, I add to earlier evidence of selective protein degradation mechanisms enabled by proteomic techniques (Bridoux et al., 2015; Dong et al., 2010; Moore et al., 2014; Moore, Nunn, Goodlett, et al., 2012; Nunn et al., 2010). Continued advancements in metaproteomic instrumentation and computational capabilities have great potential to better our understanding of protein degradation and preservation dynamics in the ocean.

3.4 ACKNOWLEDGEMENTS

Thanks to my co-authors for this chapter's corresponding manuscript, Cheyenne Adams, Khadijah Homolka, Jaqui Neibauer, Rick Keil, and Larry Mayer. I'd like to thank Larry Mayer for continued conversations and help in thinking about protein preservation. I am also very grateful for the assistance of Priska von Haller at the University of Washington Proteomics Resource Center to Laura Carlson and Anitra Ingalls for THAA measurements. I thank Eli Moore for his useful points

in conversation. Funding for this research came from the National Science Foundation (OCE 1542240 to L.M.M. and R.G. K.), and the NSF Graduate Research Fellowship Program (DGE-1762114 to M.E.D. and R.G.K.).

3.5 Data availability

Peptidomic raw and processed data files, search databases, and search parameter files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD027843 to be made available with publication of this chapter's corresponding manuscript.

3.6 References

Abdulla, H. A., Burdige, D. J., & Komada, T. (2018). Accumulation of deaminated peptides in anoxic sediments of Santa Barbara Basin. Geochimica et Cosmochimica Acta, 223, 245–258. https://doi.org/10.1016/j.gca.2017.11.021Adams, C., Mayer, L., Rawson, P., Brady, D., & Newell, C. (2019). Detrital protein contributes to oyster nutrition and growth in the Damariscotta estuary, Maine, USA. *Aquaculture Environment Interactions*, *11*, 521–536. https://doi.org/10.3354/aei00330

Alley, W. R., Mann, B. F., & Novotny, M. V. (2013). High-sensitivity Analytical Approaches for the Structural Characterization of Glycoproteins. *Chemical Reviews*, *113*(4), 2668–2732. https://doi.org/10.1021/cr3003714

Aluwihare, L. I., Repeta, D. J., & Chen, R. F. (1997). A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature*, *387*(6629), 166–169. https://doi.org/10.1038/387166a0

Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, *76*(3), 667–684. https://doi.org/10.1128/MMBR.00007-12

Aro, E.-M., Suorsa, M., Rokka, A., Allahverdiyeva, Y., Paakkarinen, V., Saleem, A., Battchikova, N., & Rintamäki, E. (2005). Dynamics of photosystem II: A proteomic approach to thylakoid protein complexes. *Journal of Experimental Botany*, *56*(411), 347–356. https://doi.org/10.1093/jxb/eri041

Benner, R., & Kaiser, K. (2003). Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter. *Limnology and Oceanography*, *48*(1), 118–128. https://doi.org/10.4319/lo.2003.48.1.0118

Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., Jensen, O. N., & Herndl, G. J. (2017). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences*, 201708779. https://doi.org/10.1073/pnas.1708779115

Bridoux, M., Neibauer, J., Ingalls, A., Nunn, B., & Keil, R. (2015). Suspended marine particulate proteins in coastal and oligotrophic waters. *Journal of Marine Systems*, *143*. https://doi.org/10.1016/j.jmarsys.2014.10.014

Cain, J. A., Solis, N., & Cordwell, S. J. (2014). Beyond gene expression: The impact of protein post-translational modifications in bacteria. *Journal of Proteomics*, *97*, 265–286. https://doi.org/10.1016/j.jprot.2013.08.012

Chen, X.-H., Li, Y.-Y., Zhang, H., Liu, J.-L., Xie, Z.-X., Lin, L., & Wang, D.-Z. (2018). Quantitative Proteomics Reveals Common and Specific Responses of a Marine Diatom Thalassiosira pseudonana to Different Macronutrient Deficiencies. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02761

Chen, Z., Luo, L., Chen, R., Hu, H., Pan, Y., Jiang, H., Wan, X., Jin, H., & Gong, Y. (2018). Acetylome Profiling Reveals Extensive Lysine Acetylation of the Fatty Acid Metabolism Pathway in the Diatom Phaeodactylum tricornutum. *Molecular & Cellular Proteomics: MCP*, *17*(3), 399–412. https://doi.org/10.1074/mcp.RA117.000339

Cowie, G. L., & Hedges, J. I. (1992). Improved amino acid quantification in environmental samples: Charge-matched recovery standards and reduced analysis time. *Marine Chemistry*, *37*(3), 223–238. https://doi.org/10.1016/0304-4203(92)90079-P

Dauwe, B., & 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. https://doi.org/10.4319/lo.1998.43.5.0782

Dauwe, B., Middelburg, J. J., Herman, P. M. J., & Heip, C. H. R. (1999). Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnology and Oceanography*, *44*(7), 1809–1814. https://doi.org/10.4319/lo.1999.44.7.1809

Dhillon, R. S., & Denu, J. M. (2017). Using comparative biology to understand how aging affects mitochondrial metabolism. *Molecular and Cellular Endocrinology*, *455*, 54–61. https://doi.org/10.1016/j.mce.2016.12.020

Dong, H.-P., Wang, D.-Z., Dai, M., & Hong, H.-S. (2010). Characterization of particulate organic matter in the water column of the South China Sea using a shotgun proteomic approach. *Limnology and Oceanography*, *55*(4), 1565–1578. https://doi.org/10.4319/lo.2010.55.4.1565

Dyhrman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. I., Alexander, H., Whitney, L. P., Drzewianowski, A., Bulygin, V. V., Bertrand, E. M., Wu, Z., Benitez-Nelson, C., & Heithoff, A. (2012). The transcriptome and proteome of the diatom Thalassiosira pseudonana reveal a diverse phosphorus stress response. *Plos One*, *7*(3), e33768–e33768. https://doi.org/10.1371/journal.pone.0033768

Estes, E. R., Pockalny, R., D'Hondt, S., Inagaki, F., Morono, Y., Murray, R. W., Nordlund, D., Spivack, A. J., Wankel, S. D., Xiao, N., & Hansel, C. M. (2019). Persistent organic matter in oxic subseafloor sediment. *Nature Geoscience*, *12*(2), 126–131. https://doi.org/10.1038/s41561-018-0291-5

Fejjar, S., Melanson, A., & Tremblay, L. (2021). Pore waters as a contributor to deep-water amino acids and to deep-water dissolved organic matter concentration and composition in estuarine and marine waters. *Marine Chemistry*, *233*, 103985. https://doi.org/10.1016/j.marchem.2021.103985

Finkel, Z. V., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I., & Irwin, A. J. (2016). Phylogenetic Diversity in the Macromolecular Composition of Microalgae. *PLoS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0155977

Galetskiy, D., Lohscheider, J. N., Kononikhin, A. S., Popov, I. A., Nikolaev, E. N., & Adamska, I. (2011). Mass spectrometric characterization of photooxidative protein modifications in Arabidopsis thaliana thylakoid membranes. *Rapid Communications in Mass Spectrometry: RCM*, *25*(1), 184–190. https://doi.org/10.1002/rcm.4855

Ghesquière, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M., Schymkowitz, J., Rousseau, F., Vandekerckhove, J., & Gevaert, K. (2011). Redox Proteomics of Protein-bound Methionine Oxidation. *Molecular & Cellular Proteomics : MCP*, *10*(5). https://doi.org/10.1074/mcp.M110.006866

Gray, N., Zia, R., King, A., Patel, V. C., Wendon, J., McPhail, M. J. W., Coen, M., Plumb, R. S., Wilson, I. D., & Nicholson, J. K. (2017). High-Speed Quantitative UPLC-MS Analysis of Multiple Amines in Human Plasma and Serum via Precolumn Derivatization with 6-Aminoquinolyl-N-hydroxysuccinimidyl Carbamate: Application to Acetaminophen-Induced Liver Failure. *Analytical Chemistry*, *89*(4), 2478–2487. https://doi.org/10.1021/acs.analchem.6b04623

Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615. https://doi.org/10.1021/acs.jproteome.8b00716

Hedges, J. I., Baldock, J. A., Gélinas, Y., Lee, C., Peterson, M., & Wakeham, S. G. (2001). Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature*, *409*(6822), 801–804. https://doi.org/10.1038/35057247

Hollibaugh, J. T., & Azam, F. (1983). Microbial degradation of dissolved proteins in seawater1. *Limnology and Oceanography*, *28*(6), 1104–1116. https://doi.org/10.4319/lo.1983.28.6.1104

Hoppe, H.-G., H.-G, Arnosti, C., & Herndl, G. (2002). Ecological Significance of Bacterial Enzymes in the Marine Environment. In *Enzymes in the Environment*. https://doi.org/10.1201/9780203904039.ch3

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., & 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. https://doi.org/10.1038/nrmicro2386

Kaiser, K., & Benner, R. (2008). Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. *Limnology and Oceanography*, *53*(1), 99–112. https://doi.org/10.4319/lo.2008.53.1.0099

Keil, R. G., & Kirchman, D. L. (1992). Bacterial Hydrolysis of Protein and Methylated Protein and Its Implications for Studies of Protein Degradation in Aquatic Systems. *Applied and Environmental Microbiology*, *58*(4), 1374–1375.

Keil, R. G., & Kirchman, D. L. (1993). Dissolved combined amino acids: Chemical form and utilization by marine bacteria. *Limnology and Oceanography*, *38*(6), 1256–1270. https://doi.org/10.4319/lo.1993.38.6.1256

Keil, R. G., & Kirchman, D. L. (1994). Abiotic transformation of labile protein to refractory protein in sea water. *Marine Chemistry*, *45*(3), 187–196. https://doi.org/10.1016/0304-4203(94)90002-7

Keil, R. G., Tsamakis, E., & Hedges, J. (2000). *Early diagenesis of particulate amino acids in marine systems*. 69–82.

Kleiner, M. (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial

Communities. *MSystems*, *4*(3). https://doi.org/10.1128/mSystems.00115-19
Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., & Strous, M. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nature Communications*, *8*(1), 1–14. https://doi.org/10.1038/s41467-017-01544-x
Knicker, H., Río, J. C. del, Hatcher, P. G., & Minard, R. D. (2001). Identification of protein remnants in insoluble geopolymers using TMAH thermochemolysis/GC–MS. *Organic Geochemistry*, *32*(3), 397–409. https://doi.org/10.1016/S0146-6380(00)00186-8
Knicker, H., Scaroni, A. W., & Hatcher, P. G. (1996). 13C and 15N NMR spectroscopic investigation on the formation of fossil algal residues. *Organic Geochemistry*, *24*(6), 661–669. https://doi.org/10.1016/0146-6380(96)00057-5
Laskay, Ü. A., Lobas, A. A., Srzentić, K., Gorshkov, M. V., & Tsybin, Y. O. (2013). Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. *Journal of Proteome Research*, *12*(12), 5558–5569. https://doi.org/10.1021/pr400522h
Laursen, A. K., Mayer, L. M., & Townsend, D. W. (1996). Lability of proteinaceous material in estuarine seston and subcellular fractions of phytoplankton. *Marine Ecology Progress Series*, *136*, 227–234. https://doi.org/10.3354/meps136227
Lee, C., Wakeham, S., & Arnosti, C. (2004). Particulate Organic Matter in the Sea: The Composition Conundrum. *AMBIO: A Journal of the Human Environment*, *33*(8), 565–575. https://doi.org/10.1579/0044-7447-33.8.565
Lehtimäki, N., Koskela, M. M., & Mulo, P. (2015). Posttranslational Modifications of Chloroplast Proteins: An Emerging Field. *Plant Physiology*, *168*(3), 768–775. https://doi.org/10.1104/pp.15.00117
Leprevost, F. V., Valente, R. H., Lima, D. B., Perales, J., Melani, R., Yates, J. R., Barbosa, V. C., Junqueira, M., & Carvalho, P. C. (2014). PepExplorer: A similarity-driven tool for analyzing *de novo* sequencing results. *Molecular & Cellular Proteomics: MCP*, *13*(9), 2480–2489. https://doi.org/10.1074/mcp.M113.037002
Li, W. K. W., Andersen, R. A., Gifford, D. J., Incze, L. S., Martin, J. L., Pilskaln, C. H., Rooney-Varga, J. N., Sieracki, M. E., Wilson, W. H., & Wolff, N. H. (2011). Planktonic Microbes in the Gulf of Maine Area. *PLOS ONE*, *6*(6), e20981. https://doi.org/10.1371/journal.pone.0020981
Liu, Z., Kobiela, M. E., McKee, G. A., Tang, T., Lee, C., Mulholland, M. R., & Hatcher, P. G. (2010). The effect of chemical structure on the hydrolysis of tetrapeptides along a river-to-ocean transect: AVFA and SWGA. *Marine Chemistry*, *119*(1), 108–120. https://doi.org/10.1016/j.marchem.2010.01.005
Mayer, L. M. (1994). Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. *Chemical Geology*, *114*(3), 347–363. https://doi.org/10.1016/0009-2541(94)90063-9
Mayer, L. M., Linda L., S., Sawyer, T., Plante, C. J., Jumars, P. A., & Sel, R. L. (1995). Bioavailable amino acids in sediments: A biomimetic, kinetics based approach. *Limnology and Oceanography*, *40*(3), 511–520. https://doi.org/10.4319/lo.1995.40.3.0511
Mayer, L. M., Macko, S. A., & Cammen, L. (1988). Provenance, concentrations and nature of sedimentary organic nitrogen in the Gulf of Maine. *Marine Chemistry*, *25*(3), 291–304. https://doi.org/10.1016/0304-4203(88)90056-4
Mayer, L. M., Schick, L. L., & Setchell, F. W. (1986). Measurement of protein in nearshore

marine sediments. *Marine Ecology Progress Series*, *30*, 159–165. https://doi.org/10.3354/meps030159

Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., St-Denis, N., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., … Nesvizhskii, A. I. (2013). The CRAPome: A Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nature Methods*, *10*(8), 730–736. https://doi.org/10.1038/nmeth.2557

Mesuere, B., Willems, T., Jeugt, F. V. der, Devreese, B., Vandamme, P., & Dawyndt, P. (2016). Unipept web services for metaproteomics analysis. *Bioinformatics*, *32*(11), 1746–1748. https://doi.org/10.1093/bioinformatics/btw039

Mikan, M. P., Harvey, H. R., Timmins-Schiffman, E., Riffle, M., May, D. H., Salter, I., Noble, W. S., & Nunn, B. L. (2020). Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. *The ISME Journal*, *14*(1), 39–52. https://doi.org/10.1038/s41396-019-0503-z

Montgomerie, S., Cruz, J. A., Shrivastava, S., Arndt, D., Berjanskii, M., & Wishart, D. S. (2008). PROTEUS2: A web server for comprehensive protein structure prediction and structure-based annotation. *Nucleic Acids Research*, *36*(suppl_2), W202–W209. https://doi.org/10.1093/nar/gkn255

Moore, E. K., Harvey, H. R., Faux, J. F., Goodlett, D. R., & Nunn, B. L. (2014). Protein recycling in Bering Sea algal incubations. *Marine Ecology Progress Series*, *515*, 45–59. https://doi.org/10.3354/meps10936

Moore, E. K., Nunn, B. L., Goodlett, D. R., & Harvey, H. R. (2012). Identifying and tracking proteins through the marine water column: Insights into the inputs and preservation mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta*, *83*, 324–359. https://doi.org/10.1016/j.gca.2012.01.002

Mulholland, M., & Lee, C. (2009). Peptide Hydrolysis and the Uptake of Dipeptides by Phytoplankton. *Limnology and Oceanography*, *54*(4). https://doi.org/10.4319/lo.2009.54.3.0856

Nagata, T., Fukuda, R., Koike, I., Kogure, K., & Kirchman, D. (1998). Degradation by bacteria of membrane and soluble protein in seawater. *Aquatic Microbial Ecology*, *14*, 29–37. https://doi.org/10.3354/ame014029

Nagata, T., & Kirchman, D. (1996). Bacterial degradation of protein adsorbed to model submicron particles in seawater. *Marine Ecology Progress Series*, *132*, 241–248. https://doi.org/10.3354/meps132241

Nguyen, R. T., & 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. https://doi.org/10.1016/S0016-7037(00)00621-9

Noble, W. S. (2015). Mass spectrometrists should only search for peptides they care about. *Nature Methods*, *12*(7), 605–608. https://doi.org/10.1038/nmeth.3450

Nunn, B. L., Norbeck, A., & Keil, R. G. (2003). Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. *Marine Chemistry*, *83*(1–2), 59–73. https://doi.org/10.1016/S0304-4203(03)00096-3

Nunn, B. L., Ting, Y. S., Malmström, L., Tsai, Y. S., Squier, A., Goodlett, D. R., & Harvey, H. R. (2010). The path to preservation: Using proteomics to decipher the fate of diatom

proteins during microbial degradation. *Limnology and Oceanography*, *55*(4), 1790–1804. https://doi.org/10.4319/lo.2010.55.4.1790

Obayashi, Y., & Suzuki, S. (2008). Occurrence of exo- and endopeptidases in dissolved and particulate fractions of coastal seawater. *Aquatic Microbial Ecology - AQUAT MICROB ECOL*, *50*, 231–237. https://doi.org/10.3354/ame01169

Olofsson, M., Lindehoff, E., & Legrand, C. (2019). Production stability and biomass quality in microalgal cultivation – Contribution of community dynamics. *Engineering in Life Sciences*, *19*(5), 330–340. https://doi.org/10.1002/elsc.201900015

Pantoja, S., Rossel, P., Castro, R., Cuevas, L. A., Daneri, G., & Córdova, C. (2009). Microbial degradation rates of small peptides and amino acids in the oxygen minimum zone of Chilean coastal waters. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(16), 1055–1062. https://doi.org/10.1016/j.dsr2.2008.09.007

Reinfelder, J. R., & Fisher, N. S. (1991). The Assimilation of Elements Ingested by Marine Copepods. *Science*, *251*(4995), 794–796. https://doi.org/10.1126/science.251.4995.794

Riffle, M., May, D. H., Timmins-Schiffman, E., Mikan, M. P., Jaschob, D., Noble, W. S., & Nunn, B. L. (2017). MetaGOmics: A Web-Based Tool for Peptide-Centric Functional and Taxonomic Analysis of Metaproteomics Data. *Proteomes*, *6*(1). https://doi.org/10.3390/proteomes6010002

Roth, L., & Harvey, H. (2006). *Intact protein modification and degradation in estuarine environments*. https://doi.org/10.1016/J.MARCHEM.2005.10.025

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M. R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd, A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome Research*, *18*(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761

Sakai, H., & Tsukihara, T. (1998). Structures of Membrane Proteins Determined at Atomic Resolution. *J Biochem (Tokyo)*, *124*(6), 1051–1059.

Schmidt, F., Koch, B. P., Elvert, M., Schmidt, G., Witt, M., & Hinrichs, K.-U. (2011). Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under Contrasting Sedimentary Redox Conditions in the Black Sea. *Environmental Science & Technology*, *45*(12), 5223–5229. https://doi.org/10.1021/es2003414

Shen, Y., Hixson, K., Tolić, N., Camp, D., Purvine, S., Moore, R. J., & Smith, R. (2008). Mass spectrometry analysis of proteome-wide proteolytic post-translational degradation of proteins. *Analytical Chemistry*. https://doi.org/10.1021/ac800077w

Smith, L. J., Fiebig, K. M., Schwalbe, H., & Dobson, C. M. (1996). The concept of a random coil: Residual structure in peptides and denatured proteins. *Folding and Design*, *1*(5), R95–R106. https://doi.org/10.1016/S1359-0278(96)00046-6

Stevens, T. J., & Arkin, I. T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins: Structure, Function, and Bioinformatics*, *39*(4), 417–420. https://doi.org/10.1002/(SICI)1097-0134(20000601)39:4<417::AID-PROT140>3.0.CO;2-Y

Suzuki, S., Kogure, K., & Tanoue, E. (1997). Immunochemical detection of dissolved proteins and their source bacteria in marine environments. *Marine Ecology Progress Series*, *158*, 1–9. https://doi.org/10.3354/meps158001

Tanoue, E., Nishiyama, S., Kamo, M., & Tsugita, A. (1995). Bacterial membranes: Possible

source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta*, *59*(12), 2643–2648. https://doi.org/10.1016/0016-7037(95)00134-4

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W. S., & Nunn, B. L. (2017). Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *The ISME Journal*, *11*(2), 309–314. https://doi.org/10.1038/ismej.2016.132

UniProt Consortium, T. (2018). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *46*(5), 2699–2699. https://doi.org/10.1093/nar/gky092

Van Mooy, B. A. S., Keil, R. G., & Devol, A. H. (2002). Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrification. *Geochimica et Cosmochimica Acta*, *66*(3), 457–465. https://doi.org/10.1016/S0016-7037(01)00787-6

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., & Peterson, M. J. (1997). Molecular indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica Acta*, *61*(24), 5363–5369. https://doi.org/10.1016/S0016-7037(97)00312-8

Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E., & Schulz, G. E. (1991). Molecular architecture and electrostatic properties of a bacterial porin. *Science (New York, N.Y.)*, *254*(5038), 1627–1630. https://doi.org/10.1126/science.1721242

Wolfe, A., Edlund, M., Sweet, A., & Creighton, S. (2006). A First Account of Organelle Preservation in Eocene Nonmarine Diatoms: Observations and Paleobiological Implications. *PALAIOS*, *21*, 298–304. https://doi.org/10.2110/palo.2005.p05-14e

Yamada, N., & Tanoue, E. (2003). Detection and partial characterization of dissolved glycoproteins in oceanic waters. *Limnology and Oceanography*, *48*(3), 1037–1048. https://doi.org/10.4319/lo.2003.48.3.1037

Yeagle, P. L., Bennett, M., Lemaître, V., & Watts, A. (2007). Transmembrane helices of membrane proteins may flex to satisfy hydrophobic mismatch. *Biochimica Et Biophysica Acta*, *1768*(3), 530–537. https://doi.org/10.1016/j.bbamem.2006.11.018

Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C. S., Li, H., Mashiyama, S. T., Joachimiak, M. P., Belle, C. van, Chandonia, J.-M., Soergel, D. A., … Venter, J. C. (2007). The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. *PLOS Biology*, *5*(3), e16. https://doi.org/10.1371/journal.pbio.0050016

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., & Ma, B. (2012). PEAKS DB: *De novo* sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics: MCP*, *11*(4), M111.010587. https://doi.org/10.1074/mcp.M111.010587

Zhou, Q., & Qiu, H. (2019). The Mechanistic Impact of N-Glycosylation on Stability, Pharmacokinetics, and Immunogenicity of Therapeutic Proteins. *Journal of Pharmaceutical Sciences*, *108*(4), 1366–1377. https://doi.org/10.1016/j.xphs.2018.11.029

Chapter 4. HIGH-RESOLUTION MARINE FLUX AND IN SITU N_2 PRODUCTION RATE DETERMINATIONS IN THE EASTERN TROPICAL NORTH PACIFIC OXYGEN DEFICIENT ZONE

ABSTRACT

The eastern tropical North Pacific (ETNP) is home to the world's largest oxygen deficient zone (ODZ), where dissolved oxygen falls below the detection limits of most instruments and where organic matter from the oligotrophic surface fuels alternative metabolisms such as denitrification and anammox. In this chapter I present data collected with sediment trap-*in situ* incubators, which were used to measure fluxes of organic carbon in sinking particles and to quantify N_2 production rates from them. Sinking particle fluxes in the oligotrophic ETNP were variable between a coastal station occupation, which were conducted at two different times of year (January 2017, April 2018). At an offshore station occupied at three different times of year (January 2017, April 2018, and October 2019), sinking particle flux attenuation was consistent (b = 0.32). Both stations had profiles with 'spiked' fluxes, where fluxes of both bulk organic carbon and protein increased under the deep chlorophyll maximum within the ODZ. At a northern ODZ station without a deep chlorophyll maximum, fluxes of protein were an order of magnitude lower, indicating that deep productivity exports labile organic matter from the upper ODZ. *In situ* experiments to quantify N_2 production rates due to sinking particles performed at the same time POM collection revealed that on average, rates from sinking particles were similar to those in the water column (6.38 nM N day^{-1} and 8.19 nM N day^{-1}, respectively) but became elevated when zooplankton or fish carcasses were collected in the chamber.

4.1 INTRODUCTION

4.1.1 *Ocean carbon flux: the biological pump*

The net transfer of photosynthetically-produced organic matter from the ocean's sunlit surface to the deep is a central process of marine food webs and elemental cycles. In all, the combination of

physical and biogeochemical processes that contribute to this transfer of carbon and nutrients is called 'the biological pump' (Ducklow, 2001; Passow & Carlson, 2012; Siegel et al., 2016). Organic material is transferred to the deep as sinking particles (aggregates of dead phytoplankton and bacteria cells, fecal pellets, minerals) in suspended or dissolved form (Ploug et al., 2008; Simon et al., 2002). However, the numerous and interrelated processes that control this export– its quality, composition, and variability–are difficult to observe and measure over time and space, and thus difficult to constrain.

Understanding the mechanisms of organic matter flux in the ocean is critical to creating predictive models of Earth's food webs and carbon cycling: the degradation of surface-produced organic matter supports deep ocean ecosystems and dictates the distribution of carbon and nutrients throughout the ocean's depths. The remineralization of organic matter in the mesopelagic zone, here defined as depths between the sunlit euphotic zone and 1000 m, provides nutrients for further surface primary production (Emerson, 2014). Organic matter that sinks or is otherwise transported deeper than 1000 m can be thought of as sequestered and will not contact the atmosphere as CO_2 for centuries or more. In this way, the outcome of how much surface production gets remineralized - and when and where in the water column it gets remineralized - ultimately exerts control on the Earth's atmospheric carbon levels and climate (Kwon et al., 2009). Without life in the ocean's surface and its export via the biological pump, global estimates of which range from 5-12 Gt carbon per year (Henson et al., 2011; Laws et al., 2000), Earth's atmosphere would have approximately 400 ppm more CO_2 than at present (Boyd, 2015; Sanders et al., 2014). Consequently, projections of future warming and other climatic changes due to human activities may impact the biological pump's functioning (Bopp et al., 2013) – potentially reducing its efficiency and resulting in positive warming feedback (Bernardello et al., 2014). Therefore, it is critical to establish accurate and highly resolved measurements of carbon export from the surface ocean.

4.1.2 *The changing role of ODZs in Earth's carbon and nitrogen cycles*

Oxygen deficient zones (ODZs) naturally occur where aerobic respiration of organic matter (OM) combines with water column stabilization to form a persistent, low-oxygen layer at mid-depths. ODZs make up less than 1% by volume of the world's oceans, yet account for 30-50% of the oceanic nitrogen loss as N_2 (DeVries et al., 2013), driving nitrogen limitation of primary

productivity over vast regions of the ocean. The sizes of ODZs are sensitive to climate change and variability: a 1% reduction of the ocean's O_2 content is predicted to double the size of world ODZs (Deutsch et al., 2011). The ocean's oxygen content has significantly decreased since the 1980s (Ito et al., 2017). At the same time, repeat hydrography indicates that the vertical extent of the ETNP ODZ has increased (Horak et al 2016).

The eastern tropical North Pacific (ETNP) is home to the Earth's largest marine oxygen deficient zone, accounting for approximately 41% of global marine anoxic waters (Paulmier & Ruiz-Pino, 2009). Numerous studies have documented an 'enhanced' flux through ODZs, implying that a high fraction of the surface-derived particulate organic carbon sinks to the deep ocean (Devol & Hartnett, 2001; Keil et al., 2016; Van Mooy et al., 2002). However, there remains much uncertainty about the mechanism(s) explaining these observations. In the ODZ, shifts in zooplankton behavior (Keil et al., 2016) and *in situ* production from anammox (Ganesh et al., 2018) or deep photoautotrophy by *Prochlorococcus* populations within the anoxic deep chlorophyll maximum (Fuchsman et al., 2019) likely have roles in an enhanced flux. These POM fluxes are important in controlling N production by denitrification and anammox (Fuchsman et al., 2019) and likely the makeup of POM controls the relative contribution of those processes (Babbin et al., 2014). Since diffusion of electron acceptors like O_2 and NO_3^- can be limiting in the interior of large particles (Ploug et al., 1997, 2008), they might become niches where both aerobic and anaerobic metabolisms can exist in tight spatial arrangement.

Marine particles are hotspots of microbial biomass and activity (Simon et al., 2002). Even in relatively oxygenated surrounding seawater (<100 μm), reduced O_2 permeability means particles can be loci of microbial denitrification and sulfate reduction (Smriga et al., 2021; Raven et al., 2021; Ploug et al., 1997; Stief et al., 2016). Recent work has deconvoluted the boundaries of different water masses in the ETNP (Evans et al., 2020) and modeled injections of oxygen into the ODZ (Margolskee et al., 2019). Questions of seasonality in this ODZ have given way to questions about smaller time scale variability driven by eddies, countercurrents, and subsurface zonal jets. The impacts of this variability have yet to be linked to the organic carbon cycling of the ETNP. The work of this large field endeavor provides an extremely valuable carbon-based dataset that covers time and space in ways that can inform future modeling efforts to predict future changes of the ocean's largest ODZ. In this chapter, I ask two main questions: how variable are sinking particle flux profiles across seasons between coastal and offshore oligotrophic stations? Is there an

observable relationship between the rates of N_2 production (heterotrophic denitrification and anammox) and sinking particle fluxes? I use data from sediment trap *in situ* incubators to simultaneously measure sinking particle flux and denitrification rates, testing the hypothesis that flux directly controls N_2 production rates in the ODZ.

4.1.3 *Measuring mesopelagic carbon fluxes with sediment traps*

The biological pump's export processes and efficiency have traditionally been studied by collecting data in sediment traps, which capture sinking particles. While sinking particles are not the only vector of the biological pump – the combination of processes also includes dissolved organic matter and suspended particle diffusion and transport (Emerson, 2014), and active transport by zooplankton migration (Boyd et al., 2019) – sinking particle flux has traditionally been used to inform and validate experimental studies, satellite estimate approaches, and models. Several general types of sediment traps are used and have varying advantages and disadvantages. These include neutrally buoyant sediment traps (NBSTs, Estapa et al., 2020), bottom-moored sediment traps (e.g., Karl et al., 1996), profiling optical sediment trap floats (Bourne et al., 2018), and surface-tethered traps (e.g., Peterson et al., 2005). Most sediment trap arrays are set out for months to years, at well below the mesopelagic zone (Buesseler et al., 2007; Buesseler & Boyd, 2009). However, recent sediment trap-based and thorium isotope measurements demonstrate the importance of upper ocean organic matter remineralization process and controls in understanding the biological pump (Marsay et al., 2015; Pavia et al., 2019). In these mid-water depths below the photic zone and above 1,000 m, the attenuation of sinking particle flux is greatest, and varies the most between differing ocean regimes (Buesseler & Boyd, 2009).

High resolution, seasonal, and short time frame trap deployment programs are rare but offer some advantages, especially when trying to evaluate whether there is short-term or spatial variability. Compared to export estimations using thorium isotope tracers (Coale & Bruland, 1985; Henson et al., 2011; Pavia et al., 2019), which can confer high spatial and temporal coverage, high resolution unpoisoned sediment traps additionally collect discrete samples of sinking particles that can be evaluated for their microbial and geochemical properties. Traps with nets or large collection systems, such as those presented here, confer the advantages collecting large amounts of sinking material (Peterson et al., 2005) which is particularly advantageous in oligotrophic regions (like the

ETNP) and when experiments and downstream analyses (like rate incubations or combined proteogenomics) require substantial organic matter samples.

As the major currency and vector of the biological pump bringing carbon and nutrients from the surface, sinking particles are desirable in experiments to estimate production and respiration rate measurements in the ocean's twilight zone. Such measurements are challenging, however, because once seawater, organic matter, and microorganisms are brought to the surface, it's extremely difficult to maintain *in situ* conditions that include light, pressure, and temperature in a laboratory setting. This is especially true in low and very low oxygen systems like the ETNP ODZ.

This chapter describes the function and use of novel sediment trap *in situ* incubation systems to measure mesopelagic organic matter flux and rate determinations. A more detailed description of the system's development will be published soon (Keil et al., *in prep.*) but results from the systems have been published in Raven et al. (2021). The trap-incubators described here are designed to be deployed for short periods of time (1-3 days) and unlike longer-deployment traps, can remain unpoisoned. This allows for investigations of microbial community composition and dynamics. Chapter 2 presented a study of paired-depth sinking and suspended particle metaproteomes using trap-incubator samples from a week-long occupation of an offshore station in the ETNP in January 2017 (see Chapter 2 and Duffy et al., submitted). This chapter will evaluate the flux and remineralization rate determination results from those deployments in addition to deployments made in April 2018 and October 2019 in the ETNP at the same offshore station and two others, one coastal and one without the deep chlorophyll maximum observed in the heart of the ODZ.

4.2 MATERIALS AND METHODS

4.2.1 *Sediment trap-in situ incubator design and function*

The current version (Figure 4.1) collects sinking particles into a 1-liter chamber that is initially open on both ends. A second 500 mL collection chamber immediately above it collects additional sinking particles during the time when the lower chamber is acting as an incubation chamber. A control chamber sits off to the side in the shadow of the sediment trap. The system is capable of measuring depth (pressure) and temperature of the ambient water, and oxygen and pH in the

ambient water and the two incubation chambers. In theory, these parameters enable determinations of *in situ* respiration. Both incubation chambers can also receive injections of labeled tracer to perform *in situ* experiments.

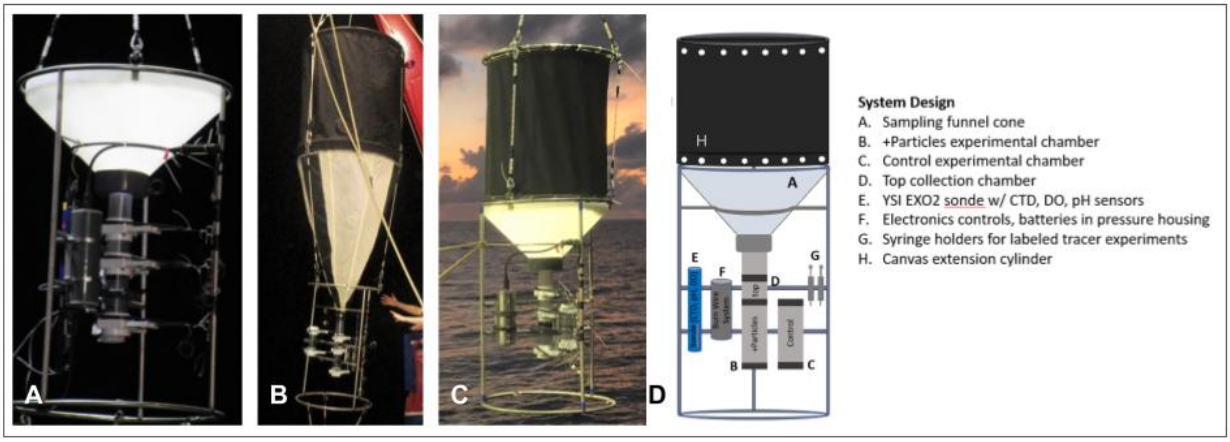

Figure 4.1. Design of sediment trap-*in situ* incubation systems. Photos show the three different particle concentration systems used during the three expeditions to the ETNP: cone traps (A), conical net traps (B), and hybrid canvas extension cylinder-cone traps (C). The *in situ* incubator design (D) includes two experimental chambers, tractor injection syringes, and an onboard CTD and ambient and in chamber pH and dissolved oxygen sensors.

The sediment-trap incubator system can be conceptually broken down into five components:

1) *Particle collection.*

   A particle collecting system - either a hard plastic funnel cone (Figure 4.1A), a conical net made from 53 µm nylon mesh similar to those described in Peterson et al. (2005) (Figure 4.1B) or a hybrid cone with a canvas extension cylinder (Figure 4.1C) - directs particles into the collection and incubation system (Figure 4.1D) which is housed in a custom welded frame. Cone traps, generally deployed in shallower water, have a collection of area 0.46 m^2. Net traps were generally deployed in deeper water (>700 m) and have a collection area of 1.24 m^2.

2) *Experimental chambers.*

The system contains three different experimental chambers for controlled *in situ* incubations. Particles concentrated in the net or cone fall into the '+ particles' collection chamber (B). A 'control' chamber of the same size is underneath but not connected to the net or cone (C). Both chambers are deployed open end-to-end, but have top and bottom gate valves that can be closed using a 'burn wire' system as described below. The + particles and control experimental chambers are both 1 L in volume, and the top collector chamber is 500 mL.

In a typical deployment, all chambers are deployed fully open end-to-end. Once at depth, the system 'soaks' for 8 hours to allow any bubbles of air from the surface to dissipate so as to not contaminate the experimental chambers. This is particularly important in low oxygen systems as described here. After soaking, the bottom gate valve of the + particles chamber closes, allowing particles to sink and collect. After particles have collected for the desired amount of time (in this study, typically 12-24-hours), the top gate valve of the + particles chamber closes, closing the chamber off for the experiment (see below in *Experimental injection system*). While the bottom of the + particles chamber closes, burn wires also close the top and bottom of the control chamber.

During the incubation period of ~23 hours, particles collect in the chamber directly above the + particles chamber, called the 'top collector' chamber (D). The top of the top collector chamber is closed just before the entire system is brought to the surface and recovered on the ship's deck.

3) *Chamber closure system*.

The + particles and control chambers are held open in deployment by rubber tubing itself tied with tension to stainless steel 'burn wires'. These wires are connected to an onboard Arduino chip computer housed in a pressure case (F) and receive current from a 6V alkaline battery as directed by the preprogrammed Arduino. The burn wire is plastic-coated but is nicked near the connection to the gate valve–at the desired time, current is sent along the wire which oxidizes and breaks at the nick given its exposure to seawater within several minutes. The

Arduino is easily programmed prior to deployment, allowing the user to design experiments of varying collection and incubation times.

4) *Tracer injection system.* The Arduino-burn wire system also controls the injection of a tracer to allow for experiments to be performed at *in situ* conditions directly after particle collection. Two syringes in stainless steel cases are loaded onboard the system (G) and can be loaded with up to 1 mL of labeled tracer, which have included ^{15}N-labeled nitrite and ammonium, ^{33}S-label sulfate (Raven et al., 2021), and ^{58}Fe-labeled $FeCl_2$ (Kenneth Bolster, *PhD dissertation*). As with the gate valve, the syringes are held open with tension upon deployment, and when their associated burn wires receive current, they break and release the syringe plunger. After the top of the + particles and the top and bottom of the control chambers are closed, the burn wire system injects the tracer via 1/8" semi-rigid plastic tubing directly into the experimental chambers.

5) *On board sensors.*
Each sediment trap-*in situ* incubation system is equipped with onboard sensors to monitor respiration rates and to detect oxygen contamination in very low oxygen conditions. The current version utilizes a Yellow Springs Instrument (YSI, Yellow Springs, OH) Exo2 sonde with pH and dissolved oxygen (DO) probes in the ambient water as well as inside the control and + particles experimental chambers. The sonde is also equipped with a sensor to measure conductivity, temperature, and depth of the ambient water.

The sediment trap-*in situ* incubator systems are designed to be deployed from an oceanographic vessel and allowed to float freely for 1-3 days tethered to a surface array comprising floating ball floats, mast and buoy, and flag and communications systems for tracking and recovery the array. Part numbers of the various components and suggestions on modifications and improvements to the system will be found in Keil et al. (*in prep.*).

4.2.2 *Sites and expedition timing*

Samples were collected in the ETNP at three different stations: a coastal ODZ station (~50 km from shore at 20.15 ºN, 106.00 ºW; water depth 2200 m); an offshore station in the heart of the ODZ (16.58 ºN, 107.05 ºW, water depth 3600 m); and another coastal station towards the northern boundary of the ODZ (21.99 ºN, 109.90 ºW, 3200 m) (Figure 4.2). They will be here called Stations P1, P2, and P3, respectively. Weeklong station occupations took place at P1 and P2 on board the *R/V Sikuliaq* between January 8-13, 2017 (cruise ID: SKQ201617S) and on board the *R/V Roger Revelle* between April 15-27, 2018 (RR1805). An 8-day occupation of P2 and a 3-day occupation of P3 took place on board the *R/V Kilo Moana* between October 3-18, 2019 (KM1920).

On all expeditions, a Seabird 911 Conductivity Temperature Density meter was deployed along with the following sensors attached to the rosette: a Seabird SBE 43 Dissolved Oxygen Sensor; a WETLabs ECO Chlorophyll Fluorometer (SeaPoint fluorometer on RR1805); a Biospherical/Licor PAR/Irradiance Sensor; and a WETLabs C- Star Transmissometer. Hydrographic and nutrient data from the three cruises can be accessed at the following: SKQ201617S: https://www.bco-dmo.org/dataset/732092, RR1805: https://www.bco-dmo.org/dataset-deployment/779204, KM1920: https://www.bco-dmo.org/dataset/849710.

Nutrient samples were filtered onto Sterivex polyethersulfone membranes (pore size 0.2 μm) before analysis of filtrate using the US-JGOFS protocols (http://usjgofs.whoi.edu/protocols_rpt_19.html) at the University of Washington Marine Chemistry Laboratory. Satellite surface chlorophyll data for 8-day averages of each station occupation were extracted from a 0.3-degree square centered on the coordinates above using the MODIS data available at http://sites.science.oregonstate.edu/ocean.productivity/.

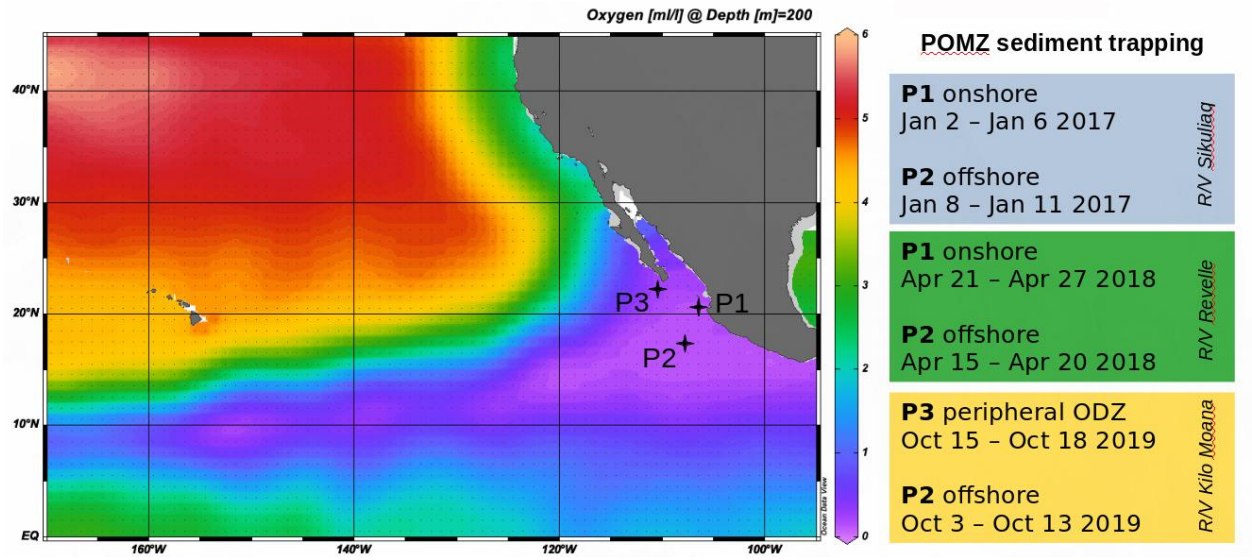

Figure 4.2. Station map of sediment trap deployments in the eastern tropical North Pacific oxygen deficient zone in January 2017, April 2018, and October 2019. Station location overlain on dissolved oxygen data at 200 m depth from the World Ocean Atlas (Garcia et al., 2013).

4.2.3 *Sediment trap flux measurements in the ETNP*

Free-floating, surface-tethered sediment traps (described in detail above) were used to quantify carbon and nitrogen fluxes from sinking particles. Arrays consisted of a surface float with 1-3 trap-incubation systems set to desired depths which were allowed to drift with surface currents, during which time they collected and incubated sinking particles. Traps deployments lasted for between 21 and 93 hours. Generally, deeper traps were left out for longer periods to collect more material. Two types of traps were deployed in 2017 and 2018 (cone and net, see above) and two types of traps were deployed in 2019 (cone-net hybrid and net).

In all cases, particles collected in the net or cone fell into one of two chambers (Figure 4.1). The bottom chamber collected particles from the net and incubated them–these samples were not used in flux analyses. The top chamber collected particles during the incubation (20-28 hours), after which the arrays were recovered. No poisons were used in any chamber, and living zooplankton, or 'swimmers', were manually removed from collection bottles, while zooplankton carcasses were retained and photographed. Sediment trap material was filtered immediately upon trap recovery onto pre-combusted glass fiber GF-75 (0.3 μm) 45 mm filters and preserved until further analysis at -80 °C.

4.2.4 *Carbon and nitrogen measurements*

Carbon content of particles in each trap was measured by elemental analyzer-isotope ratio mass spectrometry (EA–IRMS, see Keil et al., 2016; Nakatsuka et al., 1997). Elemental analyses for particulate carbon and nitrogen quantities as well as ^{13}C and ^{15}N isotopic compositions were conducted at the UC Davis Stable Isotope Facility on freeze-dried, homogenized, and acidified and non-acidified sediment trap samples in duplicate. Filtered sediment traps samples were freeze-dried and subsequently pulverized in 2 mL low-bind plastic tubes using a FastPrep Homogenizer beadbeater. Acidification of filter homogenate was performed by direct application of 1 M HCl as described in Kennedy et al. (2005). The ^{15}N and ^{13}C isotopes of sinking and suspended particles will be explored in more detail in a future manuscript (Fuchsman, Duffy and Keil et al., *in prep.*).

4.2.5 *Protein extraction and quantification*

As detailed in Chapter 2 *Materials and Methods*, protein from the tracer-free top collector chamber was extracted from one or two 11 mm diameter punches of filter into ammonium bicarbonate buffer (adapted from Bridoux et al., 2015). The chilled suspension was lysed via three cycles each of mechanical disruption with silica beads (50% 100 µm diameter and 50% 400 µm) freeze-thawing, and 30 seconds in a high-power water bath sonicator. The resulting lysate was centrifuged at 4800 rpm to isolate protein in the supernatant. Protein concentration in the extract was estimated using a modified Lowry assay (Bio-Rad).

4.2.6 *In situ N_2 production incubations*

Onboard the sediment trap-*in situ* incubator systems, a final concentration of 90 µmol L^{-1} ^{15}N-nitrite tracer with 2 µmol L^{-1} dithiothreitol to scrub potential trace O_2 contamination) was added to the + particles and control experimental chambers. Immediately after returning the sediment trap-*in situ* incubator systems shipboard, the experimental chambers were sampled using single-use needles into 12 mL glass, septum capped Exetainer vials (LabCo, UK) that had been purged with He for 5 minutes. Vials were poisoned with 50% (w/v) zinc chloride and stored in the dark at room temperature until measurement of $^{29}N_2$ and $^{30}N_2$ accumulation on a Thermo Delta V IRMS. Water from CTD casts was collected into Exetainers to measure the background isotopic

composition of N_2 gas. Duplicate samples from the + particles and control chambers were taken for alkalinity titrations and nutrient analysis at the University of Washington Marine Chemistry Laboratory (silicate, phosphate, nitrate, nitrite) from experimental chambers that will not be discussed in this chapter.

4.3 RESULTS AND DISCUSSION

4.3.1 *Sinking particle fluxes in an oxygen deficient zone*

At the coastal station P1, offshore station P2, and northern station P3, the ODZ had a sharp upper oxycline between approximately 60-100 m and a deeper oxycline at ~900-1000 m where measurable oxygen returned (Figure 4.10 and Figure 4.11). I define the ODZ as the zone where oxygen was below detection (<10 μM kg^{-1}). Except for the northern ODZ station P3, both P1 and P2 consistently exhibited depth profiles with primary and deep chlorophyll-*a* maxima. In all profiles, both maxima were below the surface mixed layer and the deep chlorophyll maximum was at the top of the ODZ (Supporting Figure 4.10 and Supporting Figure 4.11) as has been observed in earlier studies at these stations (April 2012, see Fuchsman et al., 2019). Also near the top of the ODZ at most stations and times was the location of a nitrite maximum (Figure 4.7 and Figure 4.8), indicative of N_2 production from denitrification (Buchwald et al., 2015; Codispoti & Christensen, 1985; Fuchsman et al., 2018). At the offshore Station P2, the nitrite concentrations maximum reached ~4.5 μM in both 2017 and 2018 at around 140 m and then decreased to zero by 500 m (Figure 4.7 and Figure 4.8). At the coastal stations the nitrite maxima were similar in magnitude though shallower in the water column, along with the ODZ.

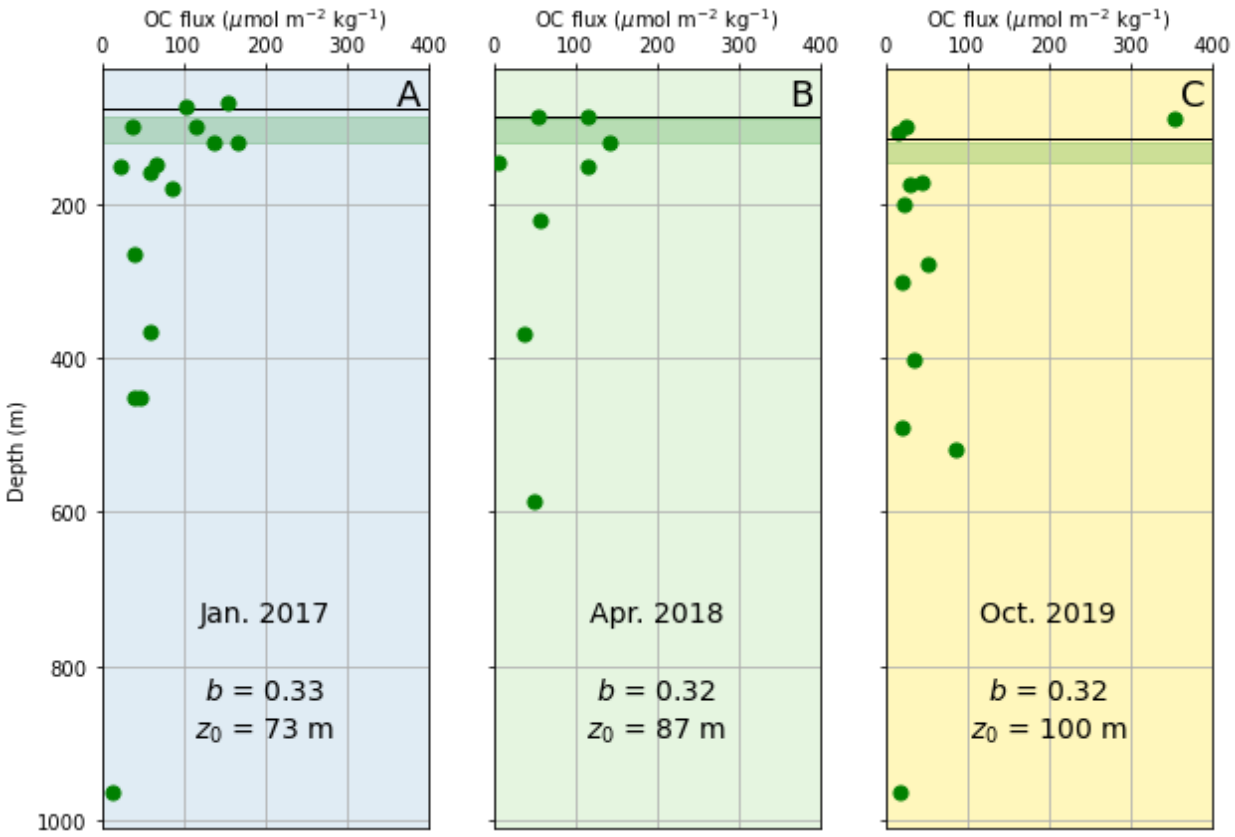

Figure 4.3. Sinking particle fluxes at offshore Station P2 in the ETNP. Fluxes of organic carbon as sinking particles at the offshore Station P2 in January 2017 (A) and April 2018 (B) and October 2019 (C). Horizontal black lines denote the onset of anoxia.

Sediment trap sinking particle flux measurements were deployed mainly within the upper- and mid-ODZ. Flux rates were generally low, as was observed in April 2013, when fluxes from net traps at P2 were 390 µmol C m^{-2} day^{-1} at 105 m and 660 µmol C m^{-2} day^{-1} at 150 m. In comparison, sediment trap sinking particle fluxes at the oligotrophic Station ALOHA range from 170-330 µmol C m^{-2} day^{-1} (Church et al., 2013, 2021).

Profiles of sinking particle fluxes were fit to a normalized power function below as first described by Martin et al. (1987) (Equation 4.1):

$$F_z = F_{z_0}\left(\frac{z}{z_0}\right)^{-b} \qquad 4.1$$

In this equation, F_z is the flux at a given depth z, normalized to a reference depth z_0. The original Martin approach entailed using a universal fixed reference depth (z_0), generally 100 m, and found a composite attenuation coefficient (b value) for the ocean of 0.86 (Martin et al., 1987). Ultimately, this 'Martin curve' metric of evaluating the transfer efficiency of the biological pump is flawed for comparing different ocean regimes. Instead of using this 'idealized and empirically based flux-vs.-depth relationship', Buesseler et al. (2020) proposed using the bottom of the euphotic zone as the reference depth for better intercomparison in space and time.

For a reference depth in this study, I used the bottom of the upper oxycline as the reference depth. This was where O_2 reached below 10 μmol kg^{-1} and ranged from 73-100 m (more consistently where $\boldsymbol{\sigma_\theta}$ ~25.7). I believe this is the most biogeochemically relevant parameter in the system, with both microbial and zooplankton oxygen limitation a key hypothesized explanation for diminished flux attenuation (Devol & Hartnett, 2001; Keil et al., 2016; Van Mooy et al., 2002) in this ODZ.

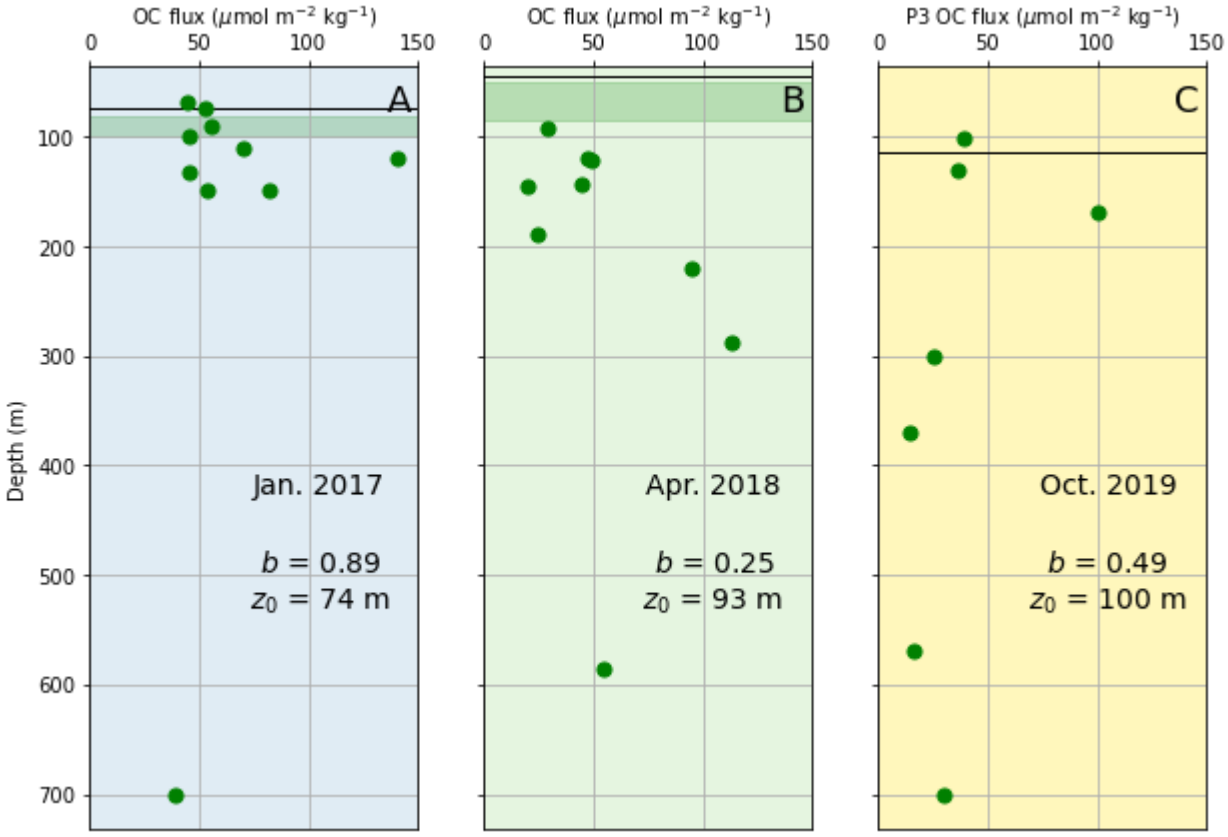

Figure 4.4. Sinking particle fluxes at coastal Station P1 and P3 in the ETNP. Fluxes of organic carbon as sinking particles at coastal Station P1 in January 2017 (A) and April 2018 (B). Fluxes at the northern coastal ODZ Station P3 in October 2019 (C). Horizontal black lines denote the onset of anoxia.

Notably, the attenuation coefficients at P2 were virtually the same ($b \sim 0.32$) for every year at different seasons (Figure 4.3). At P1, attenuation coefficients ranged from 0.25 to 0.89 between 2017 and 2018, though this is likely driven by two mid-depth flux measurements (200-300 m) where sinking particles increased with depth, and where there were no successful deployments in 2017 (Figure 4.4). In many flux profiles, a similar 'spike' was observed – where the flux of sinking particles increased between 100-400 m. Such spikes in other ocean regimes have usually been associated with transit events like phytoplankton blooms in the northeastern Atlantic even in the high-nutrient, low-chlorophyll Southern Ocean waters following experimental iron addition-induced blooms (SoFEX site, see Buesseler et al., 2020 for flux analysis). In the oligotrophic ETNP

however, there was no evidence of surface phytoplankton blooms. If I consider the spiked fluxes as increases of >20 μmol C m^{-2} day^{-1}, then 3 out of 6 of the weekly profiles show spiked fluxes: at P1 in 2018, at P2 in 2017 and 2019 (Figure 4.3 and Figure 4.4). At P3, where there was a very small or no deep chlorophyll maximum, there was no observed spiked flux at the sole station occupation in October 2019 (Figure 4.4).

At the offshore Station P2 in April 2012, there was a spiked flux in a three-point sediment trap deployment profile, where flux at 150 m was 27 μmol C m^{-2} day^{-1} greater than that at 105 m (Fuchsman et al., 2019). In that study, metagenomic analyses showed that the deep chlorophyll maximum was composed almost entirely of cyanobacteria, as has been seen in flow cytometry in 2014 (Garcia-Robledo et al., 2017). I also used metaproteomic analysis of the sinking particles caught in the sediment traps to show that cyanobacterial protein was making its way into sinking particles. Similarly, in the 2017 dataset, I demonstrated that cyanobacterial peptides are identifiable in sinking particles that have sunk through the ODZ to 1000 m (see Chapter 2). Thus, there is increasing evidence that the spiked fluxes in the ETNP are from *in situ* production in the ODZ itself by cyanobacteria. The mechanisms by which the small cyanobacterial cells become a part of sinking particles may have to do with viral infection and lysis (Fuchsman et al., 2019). Where and when in the water column this happens will be further explored by metaproteogenomics within this dataset.

With or without a spiked flux, the relatively slow particle flux attenuation in ODZs, sometimes called ‘enhanced flux’, has long been observed in the ETNP. Van Mooy et al. (2002) also observed an enhanced flux at a coastal ETNP station (on the continental slope, 22° 24′ N and 106° 18′ W with ~530 m water depth) in winter/early spring (Feb/March 1999). The authors fit their net sediment trap fluxes to a Martin curve attenuation coefficient (*b*) of 0.4. In one of the two other major ocean ODZs, in the Arabian Sea, Keil et al. (2016) deployed similar, though slightly larger net traps at two stations with permanent ODZs and one a shallow ODZ. Their more oxygenated station had a Martin attenuation coefficient of 0.98 and the ODZ stations 0.59- 0.63 (Keil et al., 2016), with normalization to the bottom of the euphotic zone (80 m).

4.3.2 *Protein flux and degradation in the ODZ*

Protein was extracted from two of the three cruises' worth of high-resolution sediment trap data, for the January 2017 (Figure 4.5) and October 2019 cruises (Figure 4.5). As a crude comparison, I first fitted the protein fluxes to the same normalized power function above (Equation 4.1) using the same or closest reference depths as I did for overall sinking particle flux calculations, finding them to be generally lower than the bulk sinking organic particles (Supporting Table 4.7). However, none of the protein flux profiles fit a Martin curve well, with R^2 values from 0.10-0.45, likely because protein fluxes peaked not at the surface in all profiles, but in the upper oxycline (Figure 4.5 and Figure 4.6). One might anticipate the flux attenuation coefficients for proteinaceous compounds to be high, as they are labile and nitrogen rich - indeed, *b* values around 0.9-1.3 have been observed for particulate hydrolyzable amino acids in oxygen minimum and deficient zones (Engel et al., 2017; Lee et al., 2000; Pantoja et al., 2004). Ratios of organic carbon to nitrogen (C_{org}:N) of the sinking particles slightly increased with depth at many stations, indicating preferential degradation of nitrogen-rich compounds within the sinking particles (Figure 4.5 and Figure 4.6). This non-Redfield regeneration has been observed in the eastern tropical North Atlantic oxygen minimum zone (Engel et al., 2017).

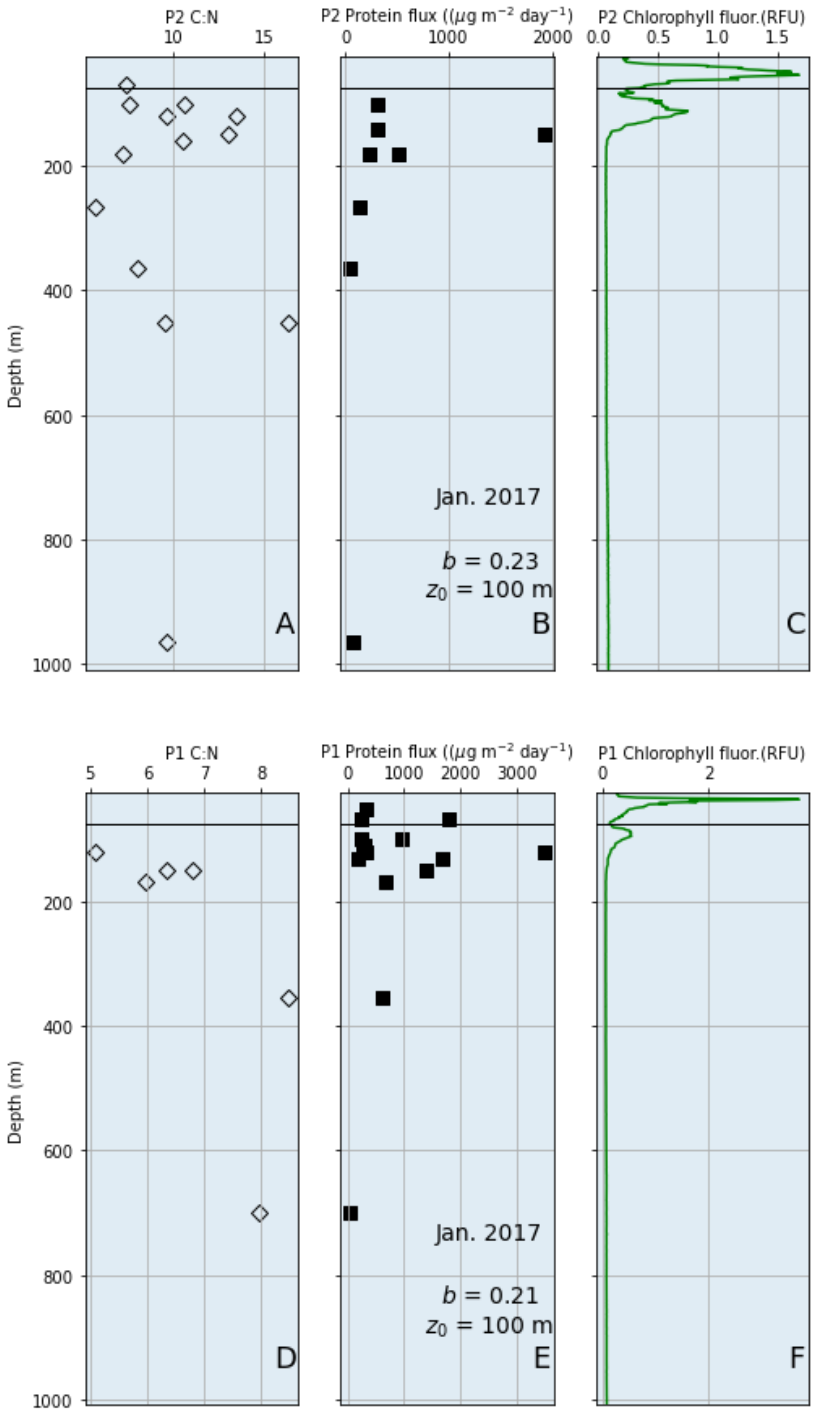

Figure 4.5. Stoichiometry and protein content in sinking particles from a coastal and offshore station in the ETNP ODZ in January 2017. Carbon to nitrogen ratios for sinking particles (A), sinking particle protein fluxes (B), and chlorophyll-a fluorescence (C) for the offshore Station P2 and for the coastal Station P1 (D, E, F). Horizontal black lines denote the onset of anoxia.

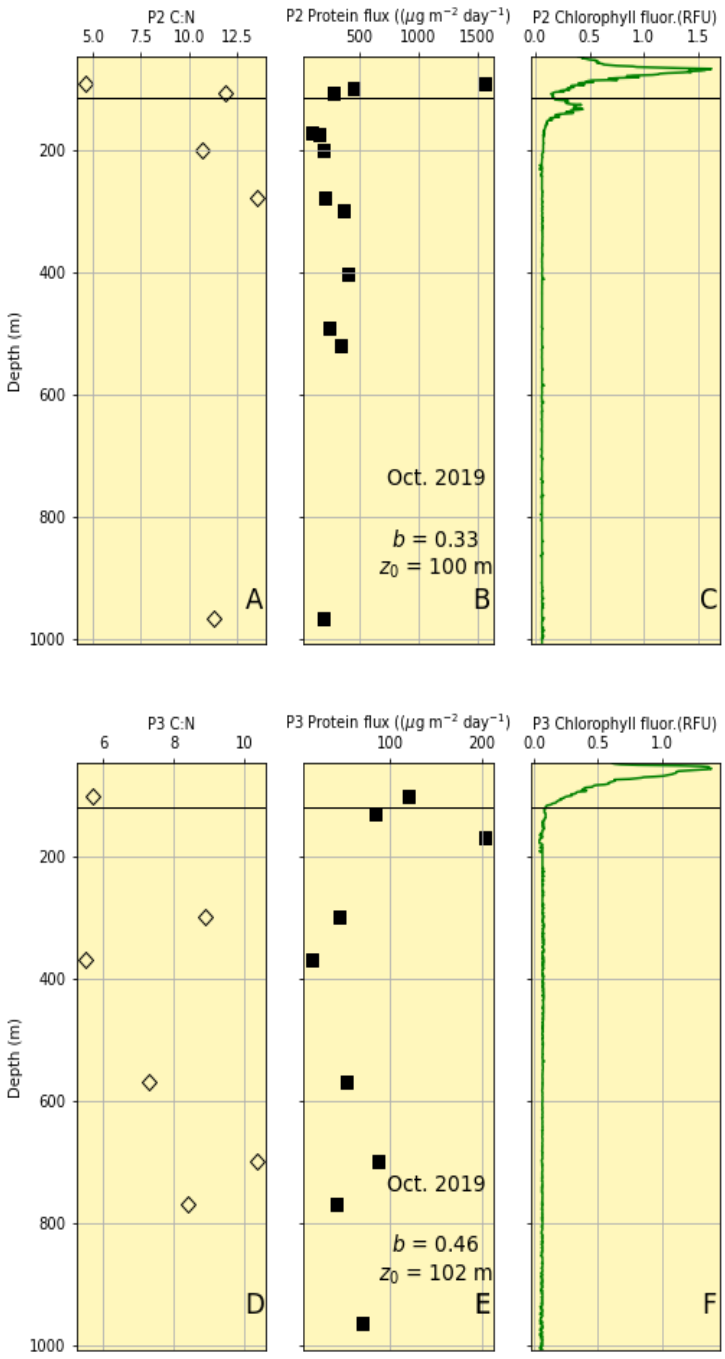

Figure 4.6. Stoichiometry and protein content in sinking particles from a northern coastal and offshore station in the ETNP ODZ in October 2019. Carbon to nitrogen ratios for sinking particles (A), sinking particle protein fluxes (B), and chlorophyll-a fluorescence (C) for the offshore Station P2 and for the northern coastal Station P3 (D, E, F). Horizontal black lines denote the onset of anoxia.

Most striking from the protein flux dataset is the difference in sinking protein fluxes between stations with a deep chlorophyll maximum, offshore Station P2 and coastal Station P1, and the single station without a deep chlorophyll maximum during these occupation dates, the northern coastal Station P3. Protein measured from sinking particles was typically highest in or directly beneath the deep chlorophyll maximum at P1 and P2, reaching ~2 mg protein m^{-2} day^{-1} at P2 and ~3.5 mg protein m^{-2} day^{-1} at P1 (Figure 4.5 and Figure 4.6). At P3 in October 2019, sinking protein flux was an order of magnitude lower, around a maximum of 0.2 mg protein m^{-2} day^{-1}, though this maximum was similarly below the upper boundary of the ODZ where there may have been a very small deep chlorophyll maximum (Figure 4.6). As discussed above, the increase in sinking protein flux under the deep chlorophyll maximum may be due to freshly detrital cyanobacterial cells, possibly through viral lysis (Fuchsman et al., 2019).

Van Mooy et al. (2002) used total amino acid measurements to show the preferential degradation of proteinaceous compounds in the sinking particles when within the ODZ, speculating that the degradation of proteins and amino acids produces ammonium to fuel anammox (Van Mooy et al., 2002). In Chapter 2, I connect the findings of that study to increased levels of asparagine and glutamine deamidation in sinking and suspended particles in the ODZ, which may indicate an important mechanism for how detrital proteinaceous compounds are degraded in particles. Protein fluxes and degradation rates were also evaluated in the coastal eastern tropical South Pacific (ETSP) ODZ using sediment traps by Pantoja et al (2004). That study found most protein is reactive and labile, with 82% of surface-produced protein degraded in the top 30 m above the oxycline and an additional 15% degraded in the ODZ, from 30 to 300 m (Pantoja et al., 2004). They therefore posit that there is no oxygen effect that lowers rates of protein degradation in the ODZ. In fact, they calculated that the remineralization of protein could support as much as 2 Tg N $year^{-1}$ in the ETSP, consistent with some estimates of denitrification in that ODZ (Pantoja et al., 2004).

As has been demonstrated in the ETNP with shipboard ^{15}N-labeled nitrite and ammonium incubations with a range of organic matter substrates including net trap sinking particles, the stoichiometry of organic matter regulates the ratio of heterotrophic denitrification and anammox (Babbin et al., 2014). The incubations performed here weren't designed to differentiate between these N_2 production metabolisms, but the differences observed in the flux of protein between

stations with and without deep chlorophyll maxima suggests that POM stoichiometry should be accounted for in future modeling and estimates of N2 production in the ETNP.

4.3.3 *In situ N_2 production rates*

Rates of N_2 production (heterotrophic denitrification and anammox) were measured in both January 2017 and April 2018 at the offshore Station P2 (Figure 4.7) and coastal station P1 (Figure 4.8). No successful rate measurements were made for the October 2019 expedition due to sample loss. The rates were adjusted from the + particles chamber to represent N_2 production due to sinking particles in the water column; unadjusted values and more details on the normalization can be found in Figure 4.13 and Figure 4.14. The rates from the control chambers represent N_2 production in the water column. Therefore, if the adjusted + particle chamber and control chamber N_2 production rates agree, all the N_2 production can be attributed to in-particle processes. Though the control chambers certainly contained suspended and dissolved organic matter, those fractions of the organic matter pool were not quantified in this study.

Rates of N_2 production from sinking particles in the offshore Station P2 ODZ ranged between 0.79-7.90 nM N day^{-1} in January 2017 and 4.07-21.09 nM N day^{-1} in April 2018. The notably higher rates in April 2018 did not correspond to overall higher sinking particle fluxes (Figure 4.7). In January 2017 at P2, the particle N_2 production rates decreased with depth in the ODZ within the 100 to 180 m range, and poorly fit a normalized power law with an attenuation coefficient (*b*) of 0.38 with an R^2 of 0.46 (Figure 4.9). In April 2018 at P2, there was an increase of particle N_2 production rates from 150 m to 220 m in the ODZ; the 3 rate measurements between 85-150 m fit a normalized power law with attenuation coefficient of 0.31 (Figure 4.9). At the coastal Station P1 in 2018, particle N_2 production rates were highest at the top of the ODZ (12.97 nM N Day^{-1} at 93 m) and decreased with depth (6.97 nM N day^{-1} at 190 m) with an attenuation coefficient of 1.0 (Figure 4.9).

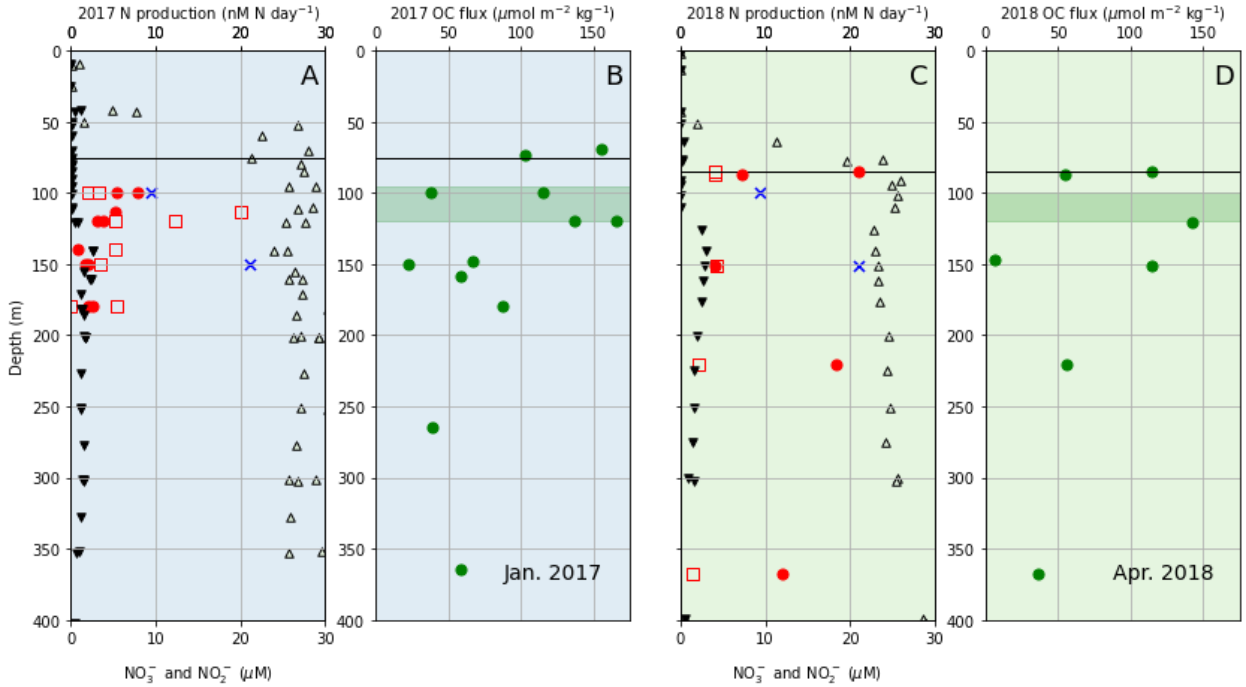

Figure 4.7. Rates of N_2 production as measured *in situ* incubations with concentrated sinking particles in the ETNP ODZ at an offshore station. Rates from sinking particles (filled red circles) and the water column (empty red boxes) with nitrate (open triangles) and nitrite (filled triangles) from deployments in January 2017 (A) and April 2018 (C). Note that a water column and particle rate overlap at 150 m in 2018 (C). Corresponding sinking particle fluxes for the two station occupations are shown for reference (B, D) with green bands denoting the approximate depth range of the deep chlorophyll maxima. Horizontal black lines denote the onset of anoxia. Total N_2 production rates from shipboard $^{15}NO_2^-$ tracer incubation with sediment trap POM in April 2012 from Babbin et al. (2014) are included as blue crosses.

At the offshore Station P1 in January 2017, *in situ* incubations were made twice at 4 depths: the bottom of the upper oxycline (100 m), just below the deep chlorophyll maximum (120 m), at the secondary nitrate maximum (150 m), and fully below the deep chlorophyll maximum (180 m). There are two inputs to the overall N_2 production rate from sinking particles, and thus two potential sources for variability: (i) the organic carbon concentration in the experimental + particles chamber and (ii) the measured accumulation of ^{15}N-labeled N_2 in the chamber. I anticipated that the sinking

particle flux, especially for collection periods as short as these (typically 12 hours) would be highly variable. For a single set of repeated depth incubations (100 m), I measured both the organic carbon in sinking particles in addition to the N_2 production rate. The two cone sediment trap deployments both collected sinking particles at the offshore station for 12 hours before incubation, two days apart, with the earlier trap collecting 37.58 μmol C m^{-2} day^{-1} and the later trap 115.68 μmol C m^{-2} day^{-1} (Table 4.6). There was more labeled N_2 accumulation in the chamber with higher organic carbon flux, and the rates adjusted for sinking particle accumulation were 6.99 and 4.88 nM N day^{-1}, respectively (Table 4.6 and Figure 4.7). Quality (nitrogen content) of the particulate organic matter could affect the N_2 production rates in the + particles chamber (Babbin et al., 2014), as this might be expected to be variable at 100 m in the ODZ. Overall, the agreement between the two incubations is close, given sources of variability and error.

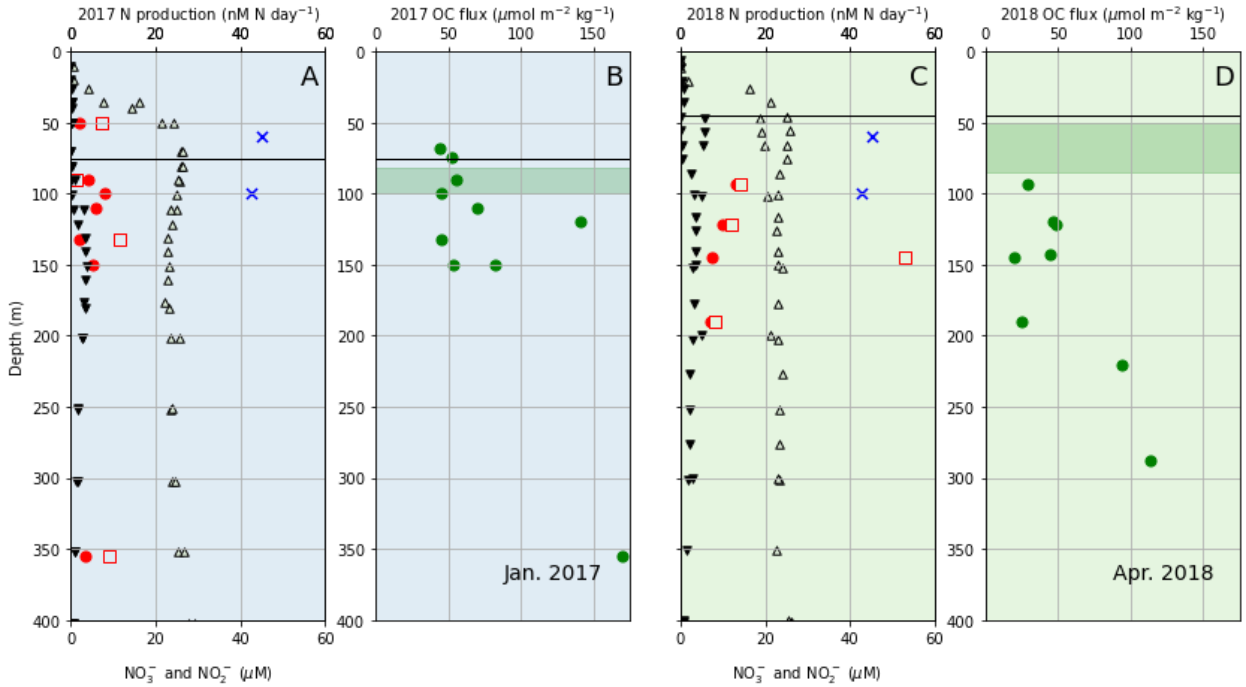

Figure 4.8. Rates of N_2 production as measured *in situ* incubations with concentrated sinking particles in the ETNP ODZ at a coastal station. Rates from sinking particles (filled red circles) and the water column (empty red boxes) with nitrate (open triangles) and nitrite (filled triangles) from deployments in January 2017 (A) and April 2018 (C). Corresponding sinking particle fluxes for the two station occupations are shown for reference (B, D) with green bands denoting the approximate depth range of the deep chlorophyll maxima. Horizontal black lines denote the onset of anoxia. Total N_2 production rates from shipboard $^{15}NO_2^-$ tracer incubation with sediment trap POM in April 2012 from Babbin et al. (2014) are included as blue crosses.

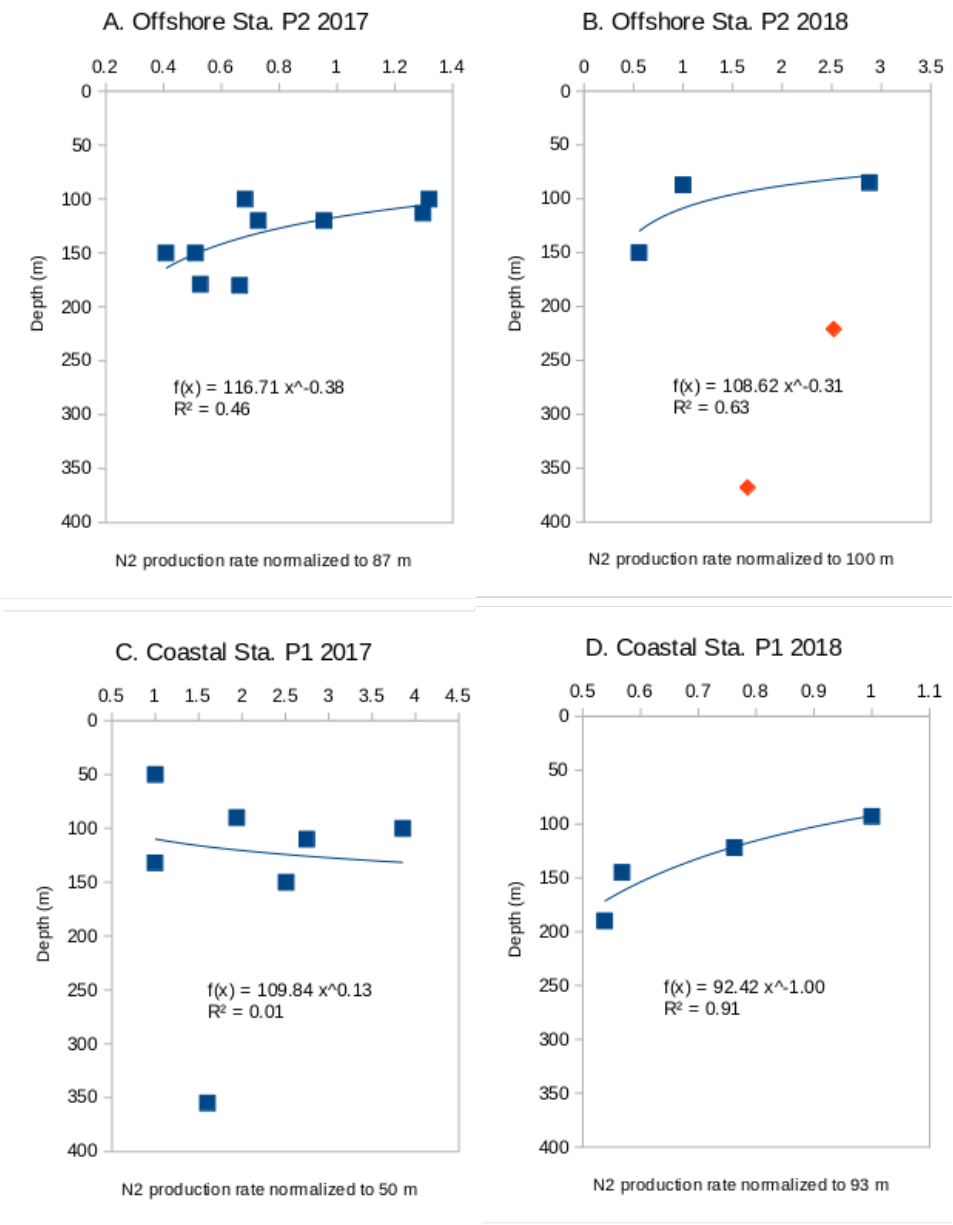

Figure 4.9. *In situ* N_2 production rates from sinking particles fit to normalized power law functions. Rates from offshore Station P2 in January 2017 normalized to the average rate at 100 m, for which there were two separate incubations (A). Offshore Station P2 in April 2018, with only the 3 rates from above the nitrite maximum fit to a power law normalized to the 87 m rate (blue boxes) and the two rates deeper not fit (red boxes) (B). Rates from coastal Station P1 in January 2017 did not fit a power law function (C). Coastal Station P1 in April 2018 normalized to the 93 m rate (D).

Table 4.6. Results of sinking particle collection and incubations at repeated depths at offshore Station P2 in January 2017.

Date	Station	Depth	Trap type	+ particles collection time (hr)	Incubation time (hrs)	Particle flux (umol C/m2 day)	N_2 production rate (nM/day)	Adjusted N_2 production rate (nM/day)
01/09/17	P2	100	cone	12	24	37.58	15	6.99
01/11/17	P2	100	cone	12	23	115.68	29	4.88
01/13/17	P2	119	cone	27	27	120 m flux used	21	3.54
01/11/17	P2	120	cone	12	23	165.80	16	2.69
01/09/17	P2	150	net	12	24	2nd 150 m flux used	30	1.89
01/11/17	P2	150	net	12	23	22.80	24	1.51
01/09/17	P2	179	net	12	23	180 m flux used	31	1.95
01/11/17	P2	180	net	12	24	86.97	39	2.46

Rates of water column N_2 production derived from the control chamber *in situ* incubations were variable, but on average across the entire dataset, were either the same or slightly greater than the particle N_2 production rates (Figure 4.7 and Figure 4.8). The notable exception is at the offshore Station P2 in April 2018, where particle N_2 production rates exceeded or equaled water column rates from the top of the ODZ until 368 m (Figure 4.7C), which was one of two deepest incubations in the dataset (see discussion of this station data below).

While these N_2 production rates are the first measured in any ODZ *in situ*, there are numerous comparisons in the literature to similar ^{15}N-tracer shipboard incubations, seawater gas and isotope derived rates, and models. Most comparable to this study, Babbin et al. (2014)

measured rates in shipboard incubations at the same coastal and offshore in April 2012. When sterilized sinking particles from an earlier version drifting net traps were added to $^{15}NO_2^-$ tracer incubations of water from the bottom of the upper oxycline (100 m) and the nitrite maximum (150 m) at P2, they measured total N_2 production rates of 9.4 and 21.1 nM N day^{-1}, respectively (plotted in Figure 4.7 as blue crosses). At P1 at the same time of year, their rate measurements in 2012 were much higher than ours in 2018, 45.2 nM N day^{-1} at the bottom of the upper oxycline (60 m) and 42.8 nM N day^{-1} at the nitrite maximum (100 m) (see Figure 4.8 and Babbin et al., 2014).

Using a one-dimensional model based on the distribution of nitrite and nitrate concentrations, Buchwald et al. (2015) predicted rates of nitrate, nitrite reduction, anammox, and nitrite oxidation in the Costa Rica Dome of the southern ETNP. Their model suggests that both nitrate and nitrite reduction maxima occur at the top of the ODZ at rates of 25 nM N d^{-1} and 15 nM N d^{-1}, respectively, with 50% of the nitrite being reoxidized back into nitrate, thus resulting in similar rates to ours (Buchwald et al., 2015). DeVries et al. (2012) produced a global, pelagic ODZ N_2 production estimate of 6 Tg N year^{-1}. Assuming that the ETNP represents 41% of the global total, that is has an average ODZ thickness of 190 m (Paulmier & Ruiz-Pino, 2009) and that ODZ volume is directly related to N_2 production activity (e.g., Paulmier and Pino-Ruiz, 2009), the overall ETNP average is 1.096 nM N day^{-1}. Across all of the stations and total of for 2017 and 2018, the *in situ* average rate due to sinking particles is 6.38 nM N day^{-1} from the heart of the ODZ. The average from the water column rates is 8.19 nM N day^{-1} (5.39 when excluding an outlier value of 53 nM N day^{-1} from 145 m at Station P1 in 2018).

4.3.4 *Potential hotspots of N_2 production: zooplankton and fish carcasses*

Sinking particle N_2 production rates were highest overall in the dataset at offshore Station P2 in April 2018 (Figure 4.7). Two incubation deployments were at a depth of great interest: within the ODZ where excess N_2 is highest (Fuchsman et al., 2018) and where some zooplankton migrate during the day to avoid predation (Cram et al., 2021; Wishner et al., 2013). Of the 5 *in situ* incubations performed at that station, 4 yielded sinking particle N_2 production rates above the dataset average of 6.38 nM N day^{-1}. In each of those 4 incubations there were zooplankton carcasses caught in the + particles collection chamber. The incubation with the highest rate (85 m, 21.09 nM N day^{-1}) had a + particles chamber with several dead pteropods, euphausid larvae, and

potentially fish fecal pellets. In the top collector chamber of the deepest trap (368 m) there was a dead lanternfish (genus *Symbolophorus*) found upon trap array recovery, which may have been in contact with the lower +particles chamber before the gate valve between then closed for the incubation (photos of all sediment trap samples were taken after filtration and just before freezing at -80 °C, see Supporting Figure 4.15). In the 150 m incubation at this station that was not above the dataset average (Figure 4.7) and did not have any zooplankton carcasses, the sinking particle rate (4.07 nM N $day^{-1)}$ was equal to the water column rate (4.02 nM N day^{-1}). Because the traps are deployed with all chambers open and collected with them closed, and because the distance from the top of the collection net or cone to the chamber is several meters, very few live zooplankton ('swimmers') were observed in the chambers.

Contributions by fish or zooplankton carcasses to N_2 production are undetectable by traditional shipboard or small vial incubation rate measurements. However, enriched ^{15}N experiments in a coastal ODZ (Golfo Dulce, Costa Rica) have shown that copepod and ostracod carcasses can host anoxic microbial hotspots in the mesopelagic where nitrogen loss metabolic processes (dissimilatory nitrate reduction to ammonium but also denitrification) can proceed even in ambient measurable dissolved oxygen (Stief et al., 2016, 2017). During the October 2019 expedition described here, the distribution, composition, and abundance of both living and dead zooplankton above and within the ODZ was investigated using net tows and relatively new staining techniques with neutral red (Elliott & Tang, 2009). Such staining revealed, surprisingly, that around 40% of copepods between 250-350 m were dead (Catherine Fitzgerald, personal communication) - the same depth range as maxima in excess N_2 (Fuchsman et al., 2018).

A recent model by Bianchi et al. (2021) showed that globally, the amount of fecal matter dropped by marine fish globally is approximately half of what it was prior to commercial fishing beginning around 1900. The authors suggested that as a result, about only half as much carbon is being sequestered by this component of the ocean's biological pump (Bianchi et al., 2021). The *in situ* incubation results I present here underline the impact of marine animal ecology on N2 production in their linkage to carbon flux and storage. Given the observed (Horak et al., 2016) and predicted shoaling and expansion of the ETNP ODZ with a warmer and less oxygenated surface ocean (Deutsch et al., 2011), understanding zooplankton distribution, behavior, and mortality in ODZs is an important factor in predicting future nitrogen cycling in the region.

4.4 CONCLUSIONS

The high-resolution dataset of sinking particle fluxes in the ETNP presented here adds considerably to the body of flux determinations from the world ocean's large permanent ODZs. High vertical and temporal resolution are critical to understanding the flow of carbon and nitrogen in ODZs (Keil et al., 2016). Particles in ODZs play host to diverse microbial processes (Raven et al., 2021) and zooplankton presence and behavior modulates their size and sinking rates (Cavan et al., 2017; Cram et al., 2021; Giering et al., 2014). Thus, there is a need for new approaches to sampling and experimenting with particles. In this chapter I demonstrate that sediment trap-incubator systems presented here and in a future methods publication (Keil et al., *in prep*.) enable controlled, *in situ* measurements of N_2 production within an ODZ that agree with previous shipboard incubation rates (Babbin et al., 2014) and models (Buchwald et al., 2015). Future work in the oxycline and above the ODZ could help in resolving questions about the seawater redox limits for particle-hosted nitrogen, sulfur, and metal cycling.

4.5 ACKNOWLEDGEMENTS

This chapter is preceding a manuscript with co-authors Clara Fuchsman, Al Devol, Gabrielle Rocap, Wendi Reuf, and Jacob Cram in addition to my fellow lab members Rick Keil, Jaqui Neibauer, Khadijah Homolka, and Jamee Adams. I thank them all for their contributions to this work and especially Clara for collaborating on the data analysis and interpretation. Many thanks to the captains and crews of the *R/V Sikuliaq, R/V Roger Revelle, and R/V Kilo Moana*. I also want to thank Paul Quay and Terry Rolfe for IRMS measurements and assistance in sample preparation. Thanks to Catherine Fitzgerald for information on ETNP zooplankton. Ben Van Mooy, Tor Bjorklund, Allison Myers-Pigg, Marlena Wied, and Miles Logsdon had parts in designing iterations of the sediment trap-incubators that will be formally published in an upcoming manuscript (Keil et al., *in prep*.). This work was funded by the National Science Foundation (DEB-1542240) and an NSF Graduate Research Fellowship DGE-1762114.

4.6 REFERENCES

Babbin, A. R., Keil, R. G., Devol, A. H., & Ward, B. B. (2014). Organic Matter Stoichiometry, Flux, and Oxygen Control Nitrogen Loss in the Ocean. *Science*, *344*(6182), 406–408.

https://doi.org/10.1126/science.1248364
Bernardello, R., Marinov, I., Palter, J. B., Sarmiento, J. L., Galbraith, E. D., & Slater, R. D. (2014). Response of the Ocean Natural Carbon Storage to Projected Twenty-First-Century Climate Change. *Journal of Climate*, *27*(5), 2033–2053. https://doi.org/10.1175/JCLI-D-13-00343.1
Bianchi, D., Carozza, D. A., Galbraith, E. D., Guiet, J., & DeVries, T. (2021). Estimating global biomass and biogeochemical cycling of marine fish with and without fishing. *Science Advances*, *7*(41). https://doi.org/10.1126/sciadv.abd7554
Bianchi, D., Weber, T. S., Kiko, R., & Deutsch, C. (2018). Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nature Geoscience*, *11*(4), 263–268. https://doi.org/10.1038/s41561-018-0081-0
Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., & Vichi, M. (2013). Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*, *10*(10), 6225–6245. https://doi.org/10.5194/bg-10-6225-2013
Bourne, H., Bishop, J., Wood, T., Loew, T., & Liu, Y. (2018). *Carbon Flux Explorer Optical Assessment of C, N and P Fluxes*. https://doi.org/10.5194/bg-2018-294
Boyd, P. W. (2015). Toward quantifying the response of the oceans' biological pump to climate change. *Frontiers in Marine Science*, *2*, 77. https://doi.org/10.3389/fmars.2015.00077
Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., & Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature*, *568*(7752), 327–335. https://doi.org/10.1038/s41586-019-1098-2
Buchwald, C., Santoro, A. E., Stanley, R. H. R., & Casciotti, K. L. (2015). Nitrogen cycling in the secondary nitrite maximum of the eastern tropical North Pacific off Costa Rica. *Global Biogeochemical Cycles*, *29*(12), 2061–2081. https://doi.org/10.1002/2015GB005187
Buesseler, K. O., & Boyd, P. W. (2009). Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. *Limnology and Oceanography*, *54*(4), 1210–1232. https://doi.org/10.4319/lo.2009.54.4.1210
Buesseler, K. O., Boyd, P. W., Black, E. E., & Siegel, D. A. (2020). Metrics that matter for assessing the ocean biological carbon pump. *Proceedings of the National Academy of Sciences*, *117*(18), 9679–9687. https://doi.org/10.1073/pnas.1918114117
Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Lam, P. J., Trull, T. W., Bidigare, R. R., Bishop, J. K. B., Casciotti, K. L., Dehairs, F., Elskens, M., Honda, M., Karl, D. M., Siegel, D. A., Silver, M. W., Steinberg, D. K., Valdes, J., Mooy, B. V., & Wilson, S. (2007). Revisiting Carbon Flux Through the Ocean's Twilight Zone. *Science*, *316*(5824), 567–570. https://doi.org/10.1126/science.1137959
Cavan, E., Henson, S., Belcher, A., & Sanders, R. (2017). Role of zooplankton in determining the efficiency of the biological carbon pump. *Biogeosciences*, *14*, 177–186. https://doi.org/10.5194/bg-14-177-2017
Church, M. J., Kyi, E., Hall Jr., R. O., Karl, D. M., Lindh, M., Nelson, A., & Wear, E. K. (2021). Production and diversity of microorganisms associated with sinking particles in the subtropical North Pacific Ocean. *Limnology and Oceanography*, *66*(9), 3255–3270. https://doi.org/10.1002/lno.11877
Church, M. J., Lomas, M. W., & Muller-Karger, F. (2013). Sea change: Charting the course for biogeochemical ocean time-series research in a new millennium. *Deep Sea Research Part*

II: Topical Studies in Oceanography, *93*, 2–15. https://doi.org/10.1016/j.dsr2.2013.01.035

Coale, K. H., & Bruland, K. W. (1985). 234Th:238U disequilibria within the California Current1. *Limnology and Oceanography*, *30*(1), 22–33. https://doi.org/10.4319/lo.1985.30.1.0022

Codispoti, L. A., & Christensen, J. P. (1985). Nitrification, denitrification and nitrous oxide cycling in the eastern tropical South Pacific ocean. *Marine Chemistry*, *16*(4), 277–300. https://doi.org/10.1016/0304-4203(85)90051-9

Cram, J., Fuchsman, C., Duffy, M., Pretty, J., Lekanoff, R., Neibauer, J., Leung, S., Huebert, K. B., Weber, T., Bianchi, D., Evans, N., Devol, A., Keil, R., & McDonnell, A. (2021, May 27). *Slow particle remineralization, rather than suppressed disaggregation, drives efficient flux transfer through the Eastern Tropical North Pacific Oxygen Deficient Zone* (world) [Preprint]. Earth and Space Science Open Archive; Earth and Space Science Open Archive. https://doi.org/10.1002/essoar.10507130.2

Deutsch, C., Brix, H., Ito, T., Frenzel, H., & Thompson, L. (2011). Climate-Forced Variability of Ocean Hypoxia. *Science*, *333*(6040), 336–339. https://doi.org/10.1126/science.1202422

Devol, A. H., & Hartnett, H. E. (2001). Role of the oxygen-deficient zone in transfer of organic carbon to the deep ocean. *Limnology and Oceanography*, *46*(7), 1684–1690. https://doi.org/10.4319/lo.2001.46.7.1684

DeVries, T., Deutsch, C., Rafter, P. A., & Primeau, F. (2013). Marine denitrification rates determined from a global 3-D inverse model. *Biogeosciences*, *10*(4), 2481–2496. https://doi.org/10.5194/bg-10-2481-2013

Ducklow, H. (2001). Upper Ocean Carbon Export and the Biological Pump. *Oceanography*, *14*, 50–54. https://doi.org/10.5670/oceanog.2001.06

Elliott, D. T., & Tang, K. W. (2009). Simple staining method for differentiating live and dead marine zooplankton in field samples. *Limnology and Oceanography: Methods*, *7*(8), 585–594. https://doi.org/10.4319/lom.2009.7.585

Emerson, S. (2014). Annual net community production and the biological carbon flux in the ocean. *Global Biogeochemical Cycles*, *28*(1), 14–28. https://doi.org/10.1002/2013GB004680

Engel, A., Wagner, H., Le Moigne, F. A. C., & Wilson, S. T. (2017). Particle export fluxes to the oxygen minimum zone of the eastern tropical North Atlantic. *Biogeosciences*, *14*(7), 1825–1838. https://doi.org/10.5194/bg-14-1825-2017

Estapa, M., Valdes, J., Tradd, K., Sugar, J., Omand, M., & Buesseler, K. (2020). The Neutrally Buoyant Sediment Trap: Two Decades of Progress. *Journal of Atmospheric and Oceanic Technology*, *37*(6), 957–973. https://doi.org/10.1175/JTECH-D-19-0118.1

Evans, N., Boles, E., Kwiecinski, J. V., Mullen, S., Wolf, M., Devol, A. H., Moriyasu, R., Nam, S., Babbin, A. R., & Moffett, J. W. (2020). The role of water masses in shaping the distribution of redox active compounds in the Eastern Tropical North Pacific oxygen deficient zone and influencing low oxygen concentrations in the eastern Pacific Ocean. *Limnology and Oceanography*, *65*(8), 1688–1705. https://doi.org/10.1002/lno.11412

Fuchsman, C. A., Devol, A. H., Casciotti, K. L., Buchwald, C., Chang, B. X., & Horak, R. E. A. (2018). An N isotopic mass balance of the Eastern Tropical North Pacific oxygen deficient zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, *156*, 137–147. https://doi.org/10.1016/j.dsr2.2017.12.013

Fuchsman, C. A., Palevsky, H. I., Widner, B., Duffy, M., Carlson, M. C. G., Neibauer, J. A.,

Mulholland, M. R., Keil, R. G., Devol, A. H., & Rocap, G. (2019). Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. *The ISME Journal*, 1. https://doi.org/10.1038/s41396-019-0452-6

Ganesh, S., Bertagnolli, A. D., Bristow, L. A., Padilla, C. C., Blackwood, N., Aldunate, M., Bourbonnais, A., Altabet, M. A., Malmstrom, R. R., Woyke, T., Ulloa, O., Konstantinidis, K. T., Thamdrup, B., & Stewart, F. J. (2018). Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria. *The ISME Journal*, *12*(11), 2706–2722. https://doi.org/10.1038/s41396-018-0223-9

Garcia, H., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., & Johnson, D. (2013). Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. In *NOAA Atlas NESDIS 70*.

Garcia-Robledo, E., Padilla, C. C., Aldunate, M., Stewart, F. J., Ulloa, O., Paulmier, A., Gregori, G., & Revsbech, N. P. (2017). Cryptic oxygen cycling in anoxic marine zones. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1619844114

Giering, S. L. C., Sanders, R., Lampitt, R. S., Anderson, T. R., Tamburini, C., Boutrif, M., Zubkov, M. V., Marsay, C. M., Henson, S. A., Saw, K., Cook, K., & Mayor, D. J. (2014). Reconciliation of the carbon budget in the ocean's twilight zone. *Nature*, *507*(7493), 480–483. https://doi.org/10.1038/nature13123

Henson, S. A., Sanders, R., Madsen, E., Morris, P. J., Le Moigne, F., & Quartly, G. D. (2011). A reduced estimate of the strength of the ocean's biological carbon pump. *Geophysical Research Letters*, *38*(4). https://doi.org/10.1029/2011GL046735

Horak, R. E. A., Ruef, W., Ward, B. B., & Devol, A. H. (2016). Expansion of denitrification and anoxia in the eastern tropical North Pacific from 1972 to 2012. *Geophysical Research Letters*, *43*(10), 5252–5260. https://doi.org/10.1002/2016GL068871

Ito, T., Minobe, S., Long, M. C., & Deutsch, C. (2017). Upper ocean O2 trends: 1958–2015. *Geophysical Research Letters*, *44*(9), 4214–4223. https://doi.org/10.1002/2017GL073613

Karl, D. M., Christian, J. R., Dore, J. E., Hebel, D. V., Letelier, R. M., Tupas, L. M., & Winn, C. D. (1996). Seasonal and interannual variability in primary production and particle flux at Station ALOHA. *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(2), 539–568. https://doi.org/10.1016/0967-0645(96)00002-1

Keil, R. G., Neibauer, J. A., Biladeau, C., van der Elst, K., & Devol, A. H. (2016). A multiproxy approach to understanding the "enhanced" flux of organic matter through the oxygen-deficient waters of the Arabian Sea. *Biogeosciences*, *13*(7), 2077–2092. https://doi.org/10.5194/bg-13-2077-2016

Kennedy, P., Kennedy, H., & Papadimitriou, S. (2005). The effect of acidification on the determination of organic carbon, total nitrogen and their stable isotopic composition in algae and marine sediment. *Rapid Communications in Mass Spectrometry: RCM*, *19*(8), 1063–1068. https://doi.org/10.1002/rcm.1889

Kwon, E. Y., Primeau, F., & Sarmiento, J. L. (2009). The impact of remineralization depth on the air–sea carbon balance. *Nature Geoscience*, *2*(9), 630–635. https://doi.org/10.1038/ngeo612

Laws, E. A., Falkowski, P. G., Smith Jr., W. O., Ducklow, H., & McCarthy, J. J. (2000). Temperature effects on export production in the open ocean. *Global Biogeochemical Cycles*, *14*(4), 1231–1246. https://doi.org/10.1029/1999GB001229

Lee, C., Wakeham, S. G., & I. Hedges, J. (2000). Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(8), 1535–1568. https://doi.org/10.1016/S0967-0637(99)00116-8

Margolskee, A., Frenzel, H., Emerson, S., & Deutsch, C. (2019). Ventilation Pathways for the North Pacific Oxygen Deficient Zone. *Global Biogeochemical Cycles*, *33*(7), 875–890. https://doi.org/10.1029/2018GB006149

Marsay, C. M., Sanders, R. J., Henson, S. A., Pabortsava, K., Achterberg, E. P., & Lampitt, R. S. (2015). Attenuation of sinking particulate organic carbon flux through the mesopelagic ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1089–1094. https://doi.org/10.1073/pnas.1415311112

Martin, J. H., Knauer, G. A., Karl, D. M., & Broenkow, W. W. (1987). VERTEX: Carbon cycling in the northeast Pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, *34*(2), 267–285. https://doi.org/10.1016/0198-0149(87)90086-0

Nakatsuka, T., Handa, N., Harada, N., Sugimoto, T., & Imaizumi, S. (1997). Origin and decomposition of sinking particulate organic matter in the deep water column inferred from the vertical distributions of its δ15N, δ13 and δ14. *Deep Sea Research Part I: Oceanographic Research*, *44*, 1957–1979. https://doi.org/10.1016/S0967-0637(97)00051-4

Pantoja, S., Sepúlveda, J., & González, H. E. (2004). Decomposition of sinking proteinaceous material during fall in the oxygen minimum zone off northern Chile. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*, 55–70. https://doi.org/10.1016/j.dsr.2003.09.005

Passow, U., & Carlson, C. A. (2012). The biological pump in a high CO2 world. *Marine Ecology Progress Series*, *470*, 249–271. https://doi.org/10.3354/meps09985

Paulmier, A., & Ruiz-Pino, D. (2009). Oxygen minimum zones (OMZs) in the modern ocean. *Progress in Oceanography*, *80*(3–4), 113–128. https://doi.org/10.1016/j.pocean.2008.08.001

Pavia, F. J., Anderson, R. F., Lam, P. J., Cael, B. B., Vivancos, S. M., Fleisher, M. Q., Lu, Y., Zhang, P., Cheng, H., & Edwards, R. L. (2019). Shallow particulate organic carbon regeneration in the South Pacific Ocean. *Proceedings of the National Academy of Sciences*, *116*(20), 9753–9758. https://doi.org/10.1073/pnas.1901863116

Peterson, M. L., Wakeham, S. G., Lee, C., Askea, M. A., & Miquel, J. C. (2005). Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology and Oceanography: Methods*, *3*(12), 520–532. https://doi.org/10.4319/lom.2005.3.520

Ploug, H., Iversen, M., & Fischer, G. (2008). Ballast, sinking velocity, and apparent diffusivity within marine snow and zooplankton fecal pellets: Implications for substrate turnover by attached bacteria. *Limnology and Oceanography*, *53*, 1878–1886. https://doi.org/10.4319/lo.2008.53.5.1878

Ploug, H., Kühl, M., & BUCHHOLZCLEVEN, B. (1997). Anoxic aggregates—An ephemeral phenomenon in the pelagic environment? *AQUATIC MICROBIAL ECOLOGY*, *13*, 285–294. https://doi.org/10.3354/ame013285

Raven, M. R., Keil, R. G., & Webb, S. M. (2021). Microbial sulfate reduction and organic sulfur formation in sinking marine particles. *Science*, *371*(6525), 178–181. https://doi.org/10.1126/science.abc6035

Sanders, R., Henson, S. A., Koski, M., De La Rocha, C. L., Painter, S. C., Poulton, A. J., Riley, J., Salihoglu, B., Visser, A., Yool, A., Bellerby, R., & Martin, A. P. (2014). The Biological Carbon Pump in the North Atlantic. *Progress in Oceanography*, *129*, 200–218. https://doi.org/10.1016/j.pocean.2014.05.005

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., Burd, A., Carlson, C. A., D'Asaro, E. A., Doney, S. C., Perry, M. J., Stanley, R. H. R., & Steinberg, D. K. (2016). Prediction of the Export and Fate of Global Ocean Net Primary Production: The EXPORTS Science Plan. *Frontiers in Marine Science*, *3*, 22. https://doi.org/10.3389/fmars.2016.00022

Simon, M., Grossart, H.-P., Schweitzer, B., & Ploug, H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquatic Microbial Ecology, 175-211 (2002)*, *28*. https://doi.org/10.3354/ame028175

Smriga, S., Ciccarese, D., & Babbin, A. R. (2021). Denitrifying bacteria respond to and shape microscale gradients within particulate matrices. *Communications Biology*, *4*(1), 1–9. https://doi.org/10.1038/s42003-021-02102-4

Stief, P., Kamp, A., Thamdrup, B., & Glud, R. N. (2016). Anaerobic Nitrogen Turnover by Sinking Diatom Aggregates at Varying Ambient Oxygen Levels. *Frontiers in Microbiology*, *7*, 98. https://doi.org/10.3389/fmicb.2016.00098

Stief, P., Lundgaard, A. S. B., Morales-Ramírez, Á., Thamdrup, B., & Glud, R. N. (2017). Fixed-Nitrogen Loss Associated with Sinking Zooplankton Carcasses in a Coastal Oxygen Minimum Zone (Golfo Dulce, Costa Rica). *Frontiers in Marine Science*, *4*, 152. https://doi.org/10.3389/fmars.2017.00152

Van Mooy, B. A. S., Keil, R. G., & Devol, A. H. (2002). Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrification. *Geochimica et Cosmochimica Acta*, *66*(3), 457–465. https://doi.org/10.1016/S0016-7037(01)00787-6

Wishner, K. F., Outram, D. M., Seibel, B. A., Daly, K. L., & Williams, R. L. (2013). Zooplankton in the eastern tropical north Pacific: Boundary effects of oxygen minimum zone expansion. *Deep Sea Research Part I: Oceanographic Research Papers*, *79*, 122–140. https://doi.org/10.1016/j.dsr.2013.05.012

4.7 SUPPORTING METHODS

The goal in the incubations was to arrive at N_2 production rates values that represented the metabolic output due to sinking particles in the water column. To achieve this, we need to adjust to the raw labeled N_2 accumulation rate in the + particles chamber by the following:

1. Time the + particles chamber was open to the surface and collecting (~24-hours, with a range from 12-91)
2. Collection area of the trap (1.23 m^2 for the net; cone is 0.46 m^2)
3. Total amount of organic carbon in the + particles chamber at the end of the incubation period (typically 0.5-2 mg C)
4. Average sinking rate of particles collected in the trap. We didn't measure this but used a value for sinking particles in the ETNP ODZ measured with a particle settling column from Cavan et al. (2017).

No or very little organic carbon was detected in any of the control chamber samples by IRMS, which as a result were not control subtracted. Cavan et al. (2017) reported the ratio of slow sinking vs. fast sinking particle flux to be 18:0.6, with sinking rates measured at 6.5 m day^{-1} and 69 m day^{-1}. I therefore used an average sinking rate value of 8.45 m day^{-1} in these adjustments.

4.8 SUPPORTING TABLES

Table 4.7. Attenuation coefficients (*b*) of bulk sinking particles and protein in sinking particles in the ETNP ODZ.

Station	Year	Sinking particle *b*	Sinking particle protein *b*
Offshore P2	January 2017	0.33	0.23
Offshore P2	April 2018	0.32	*na*
Offshore P2	October 2019	0.32	0.33
Coastal P1	January 2017	0.89	0.21
Coastal P1	April 2018	0.25	*na*
N. coastal P3	October 2019	0.49	0.46

4.9 SUPPORTING FIGURES

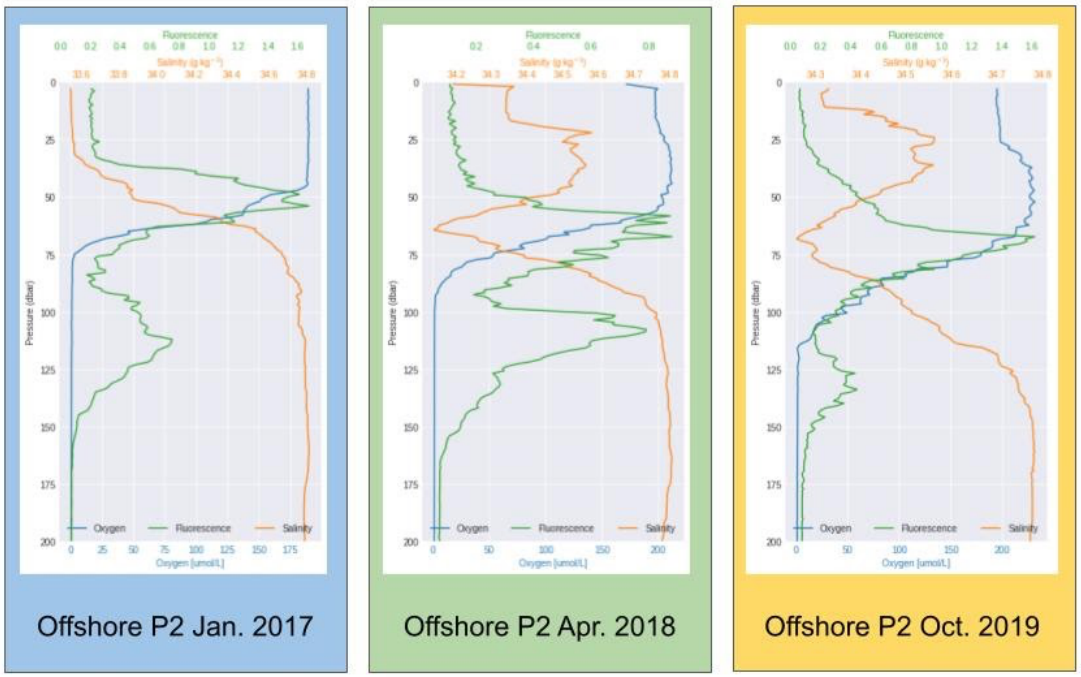

Figure 4.10. Profiles of dissolved oxygen, salinity, and chlorophyll-a fluorescence at offshore station P2 during three separation cruises: January 2017, April 2018, and October 2019.

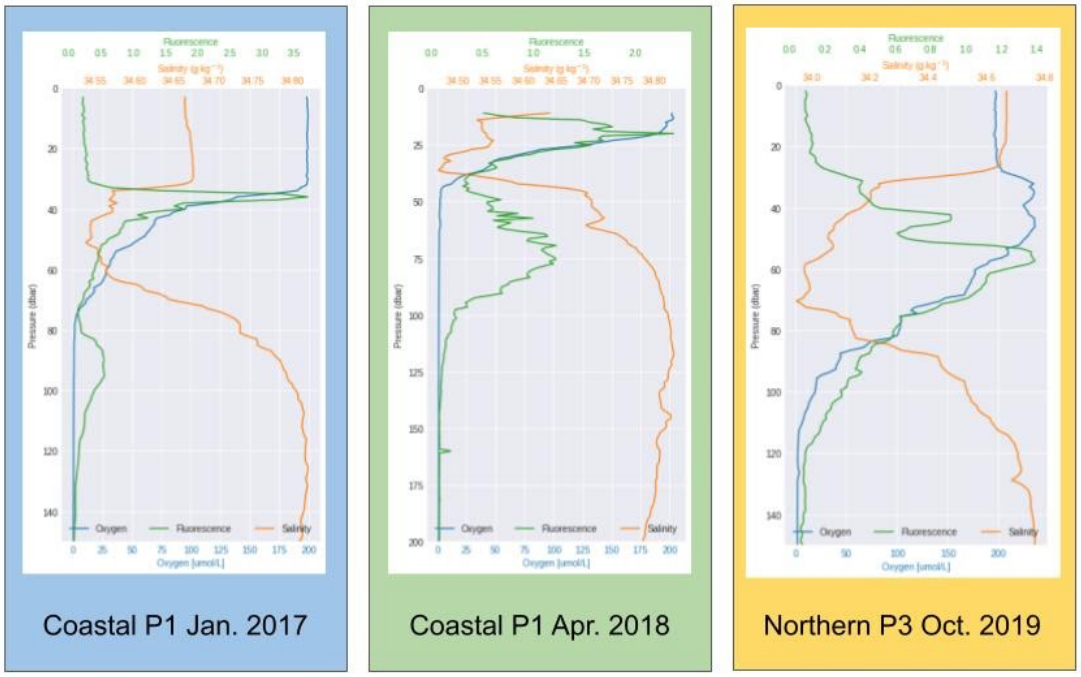

Figure 4.11. Profiles of dissolved oxygen, salinity, and chlorophyll-a fluorescence at a coastal station P1 during two separation cruises: January 2017, April 2018; and at northern ODZ station P3 in October 2019.

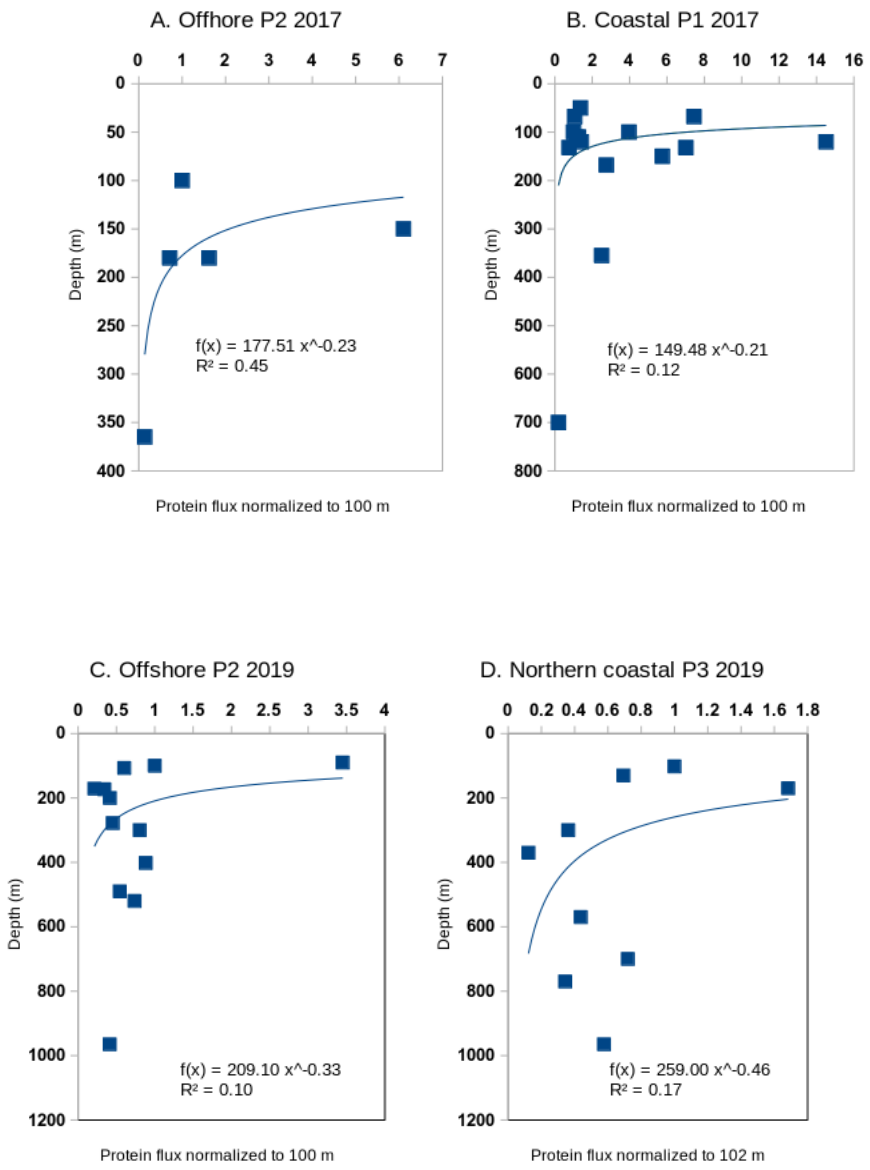

Figure 4.12. Protein in sinking particle flux attenuation constant calculations. Offshore Station P2 and coastal Station P1 in January 2017 (A, C) April 2018 (B, D). Adjusted for sinking particle flux and normalized by the rate closest to the bottom of the upper oxycline to yield a 'Martin curve' like power law fit with attenuation coefficient *b*.

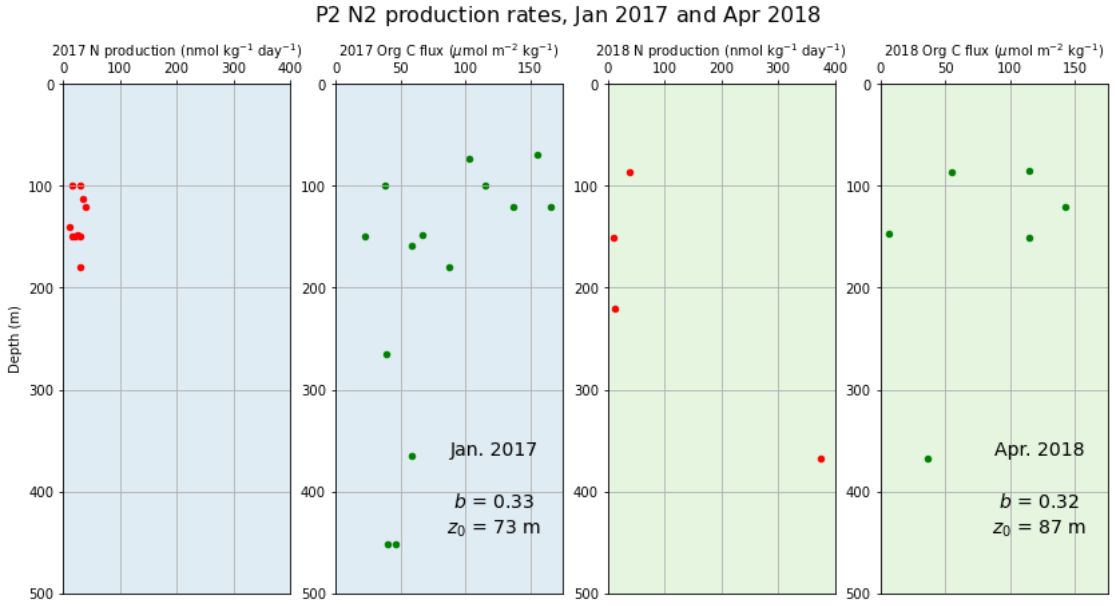

Figure 4.13. Rates of N_2 production as measured *in situ* incubations with concentrated sinking particles in the ETNP ODZ at an offshore station.

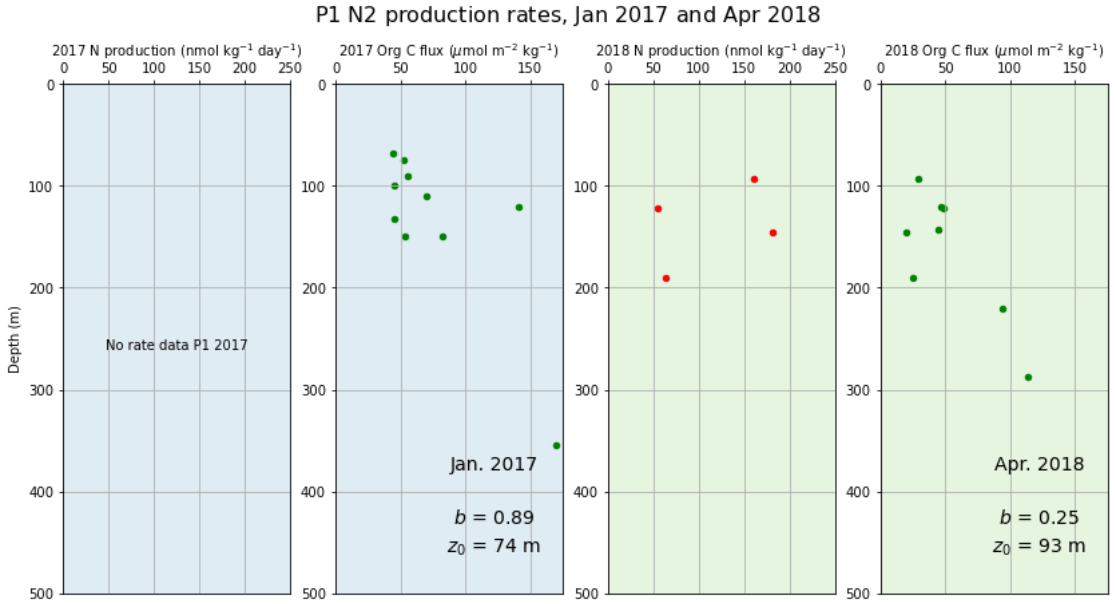

Figure 4.14. Rates of N_2 production as measured *in situ* incubations with concentrated sinking particles in the ETNP ODZ at a coastal station.

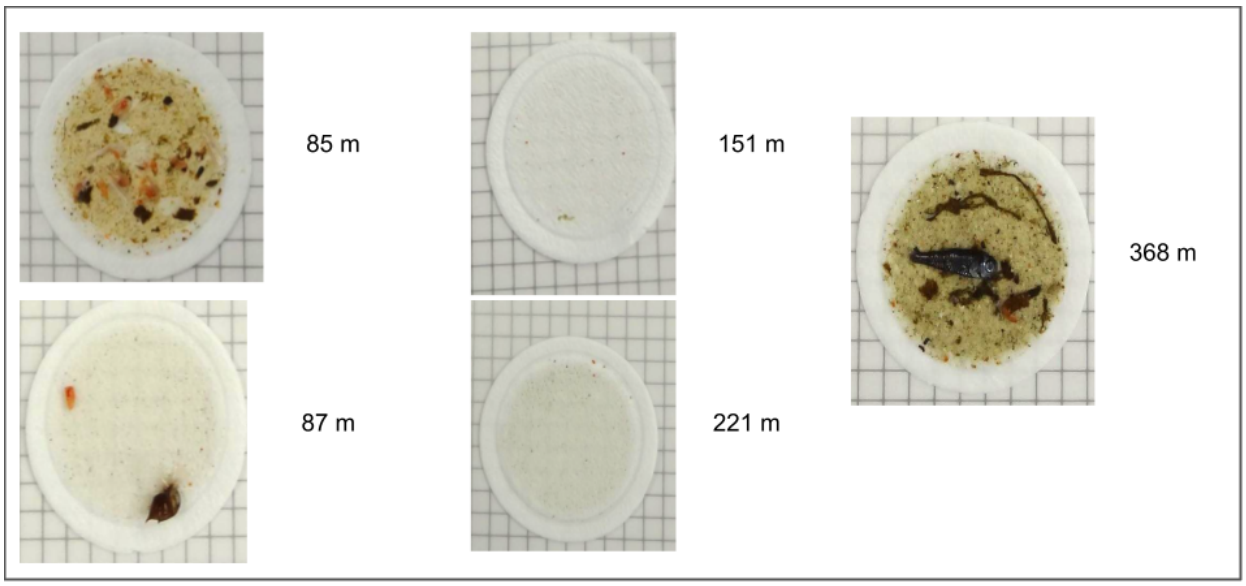

Figure 4.15. Photos of filters +particles chambers from in situ incubations at offshore Station P2 in April 2018. Zooplankton or fish carcasses were found in all chambers except at 150 m (pteropod carcass was removed from 221 m filter before photo).

Chapter 5. A PEPTIDOMIC VIEW INTO ORGANIC MATTER PROCESSING IN THE LOWER AMAZON RIVER

ABSTRACT

The proteinaceous component of particulate organic matter was studied in the lower Amazon River in April 2019 using a metapeptidomic approach, which aims to detect and identify all peptides in a sample, including those from living organisms and those that are detrital. In this study of four stations in the lower Amazon River, degradation rates of particulate proteinaceous compounds were rapid compared to bulk respiration rates and were an order of magnitude higher than previous measurements of particulate lignin degradation. Sequence-based metrics of the degradation state of peptides - relative amino acid composition, subcellular component annotations, and post translational modifications - were representative of a living microbial community with minimal detrital proteinaceous compounds. A small number of peptides that could be specifically classified as being sourced from *in situ* microbial primary producers (diatoms and dinoflagellates, primarily) did have geochemical indications of being detrital: more were associated with cell and organelle membranes than the bulk peptide pool. Algal peptides then disappeared from the proteomic window after a 24-hour incubation period, indicating that *in situ* produced organic matter is quickly consumed by heterotrophic microbes. The microbial communities derived from metapeptidomic sequencing were very similar to those observed in other studies and were consistent between stations 130 km apart in the mouth of the river. The uniformly Actinobacteria-dominated consortia were expressing enzymes mostly to reproduce and oxidize lignin, as well as bind metal ions such as zinc, a known cofactor in ligninolytic enzymes.

5.1 INTRODUCTION

The classically conceived role of rivers is that they simply export organic matter to the oceans and that once there, long-term preservation of terrestrially-derived organic matter occurs largely along continental margins. However, in recent decades rivers have come to be seen not simply as

transitory pipes, but as transformers and regulators that themselves adjust the carbon cycle of not only their watersheds but also of marine receiving waters. This paradigm shift is a result of the discovery that rivers and other inland waters outgas immense quantities of CO_2 and CH_4 to the atmosphere (Butman & Raymond, 2011; Richey et al., 2002). Two primary questions emerge from these new results – what are the geographic distributions and magnitude of aquatic CO_2 outgassing, and what are the dynamics that drive this outgassing?

The Amazon River contributes about one fifth of the total freshwater input into the world's oceans (Chong et al., 2016). As organic matter travels its length from continent to sea, respiration by heterotrophic microbes (Santos-Júnior et al., 2020; Satinsky et al., 2017) transforms organic matter fixed on land into CO_2 and CH_4 (Mayorga et al., 2005; Ward et al., 2013, 2016), resulting in a tremendous outgassing of these greenhouse gases to the atmosphere that remains poorly constrained. The quality and quantity of available organic matter likely control, to some extent, overall microbial respiration and thus outgassing (Ward et al., 2013, 2016). The dynamics of the microbial performing this heterotrophy may also modulate the carbon outputs of the Amazon (Santos-Júnior et al., 2020). However, there is little knowledge of what terrestrial organic matter compounds actually fuel this respiration or the range of their turnover rates, or the biogeography and temporal variability of the microbial communities.

Estimating how microbes and available organic matter control carbon fluxes in the Amazon is critical because depending on their actual magnitude, this important global terrestrial carbon sink may prove to be smaller than currently estimated, with the river system mobilizing and remineralizing a significant component of the organic carbon pool that is currently considered sequestered in soils. A mechanistic understanding of the drivers of these fluxes is becoming increasingly important as human activity (dams, biomass burning, and agricultural use) drastically alters the Amazon Basin, necessitating accurate models of carbon outgassing and uptake.

Geochemical evidence from ^{14}C measurements (Mayorga et al., 2005), molecular analyses of DOM (Seidel et al., 2016) and total hydrolyzable amino acid analysis (Hedges et al., 1994) points to a very small, rapidly cycling pool of labile organic matter that turns over against the background of a much larger pool of refractory material (Richey et al., 1990). In the coastal sediments under the Amazon plume, proteinaceous compounds are potentially better preserved than in the river itself, resulting in long term coastal carbon storage (Chong et al., 2016). Recently, Ward et al. (2018) developed rotating incubation systems that better simulate river flow and particle

suspension. These have been important in deciphering the source of highly reactive materials, showing that lignins, often thought to be recalcitrant (Gough et al., 1993; Hedges et al., 1988; Opsahl & Benner, 1997), are in fact remineralized in the understudied lower Amazon. Conversely, though proteins are thought to be very labile, the sediment-laden river water may slow degradation rates due to strong protein-mineral surface interactions (Aufdenkampe et al., 2001; Keil et al., 1994; Mayer, 1994; Mayer et al., 1988).

In this study of organic matter processing in the lower Amazon, I aimed to characterize the diversity and function of microbial communities, and to investigate the geochemical signatures of the entire proteinaceous pool. To evaluate the lability of proteinaceous compounds in the river, I employed rotating chamber incubations (Ward et al., 2018) without amendments to monitor protein turnover and microbial community/functional shifts in two size factions of the organic matter pool. Given the indications of a small, labile pool of organic matter in the Amazon and the body of literature on protein-mineral interactions, I hypothesized that:

H1. Proteins would represent a labile and quickly recycling pool of organic matter in the lower Amazon (Mayorga et al., 2005; Richey et al., 1990), with protein-specific degradation rates higher than those previously measured for more abundant, yet presumably more recalcitrant compounds like lignin (Ward et al., 2013).

H2. Due to high levels of sediment in the river water (Armijos et al., 2020) and the reactivity of proteins and mineral surfaces (Yu et al., 2013), there would be a signal of detrital, preserved proteinaceous compounds in the modification profiles, relative amino acid compositions, and subcellular location origins of peptides.

H3. Proteins from living organisms would represent a living community of microbial life producing enzymes to degrade and transport into their cells terrestrially-produced compounds like lignin and cellulose, and this would likely not shift during short 24-hour non-amended incubations. Though I collected the first sequence-based assessment of the microbial community in the lower Amazon tidal reach, more than 100 km oceanward from previous studies, I anticipated that due to little other inputs, the microbial community would be compositional and functionally similar to those observed upstream (Doherty et al., 2017; Satinsky et al., 2015).

I addressed these hypotheses using a metapeptidomic approach, aiming to characterize all peptides present in a system at a point in time, regardless of whether they're from living organisms or detrital. This technique presents an opportunity to evaluate both geochemical and biological processes in environmental settings. Although I was unable to measure dissolved proteins or amino acids, degradation rates of proteinaceous compounds in the particulate fraction were rapid compared to bulk respiration rates and previous measurement of particulate lignin degradation (**H1**). Contrary to my expectation (**H2**), each of these sequence-based metrics was overwhelmingly representative of the living microbial community's protein and not detrital proteinaceous compounds. However, peptides that could be specifically classified as being made by *in situ* microbial primary producers (diatoms and dinoflagellates, primarily) did have indicators of being detrital: more were associated with cell and organelle membranes than the bulk peptide pool. Many of these algal peptides disappeared from the proteomic window after the 24-hour incubation period, indicating that though they are detrital, they are quickly consumed by the microbial community. Finally, the microbial communities derived from the metapeptidomic approach were very similar to those observed by tag sequencing, metagenomics, and metatranscriptomics that overlapped with two of the stations studied here (**H3**). At every station along the tidal reach and over the course of 24-hour incubations, there were no shifts in community composition or function as evidenced by identified peptides. The uniformly Actinobacteria-dominated consortia were expressing enzymes mostly to reproduce and oxidize lignin, as predicted.

5.2 MATERIALS AND METHODS

5.2.1 *Incubations and peptidomics sampling*

Sampling was performed at 4 stations in the lower reach of the Amazon River (Figure 5.1) aboard the *B/M Mirage* between April 8-15, 2019, during the rising water period of the yearly hydrograph (Richey et al., 1989). This stretch of the river is influenced by meso-tides that drive a variation of about ~2.8 m in water column amplitude along with a reversal of river flow. Though far reaching, the flow reversal does not result in any changes in salinity due to high freshwater discharge (Sawakuchi et al., 2017). At the two locations 150 km downstream of Macapá, the channel width continues to increase and is distributed between large islands until it ultimately disconnects from

the mainland and the riparian zone (Cunha & Sternberg, 2018; Ward et al., 2015). The Amazon River water that flows into the ocean retains its freshwater characteristics on the surface approximately 60 km from the mouth (Valerio et al., 2018). Near the mouth of the Amazon River, water flow reversals increase the connectivity of the main river channel with large floodplain lakes (várzeas) and adjacent channels (Cunha and Stenberg, 2018; Santos et al., 2018) as well as its tributaries (Eom et al., 2017; Sawakuchi et al., 2017).

Amazon River discharge is strongly seasonal. At the high-water period from May-June, discharge reaches ~240,000 $m^3 s^{-1}$, and falls to ~80,000 $m^3 s^{-1}$ during the low water period from September-November (Lentz, 1995). Fine sediment suspension levels track discharge in the Amazon main stem, with highest sediment concentrations during high water compared to low water (Armijos et al., 2020; Ward et al., 2016). In contrast to some of its blackwater tributaries, high sediment load and low light penetration in the Amazon mainstem mean the contribution of photodegradation to the overall CO_2 production is exceedingly minimal (<1%) compared to depth integrated microbial respiration (Remington et al., 2011).

An important aspect of this work is to assess how microbial and biochemical evolution with time, during respiration. Respiration rates are classically determined in stationary biological oxygen demand (BOD) bottles. But Ward et al. (2018) developed spinning shipboard incubation chambers, used here, that are designed to mimic the natural movement of the river and maintain sediment suspension, which showed that the BOD substantially underestimated ambient respiration. Incubations were performed at the four stations in triplicate in spinning shipboard chambers (Ward et al., 2018). Rotating chambers were filled with river water passed through a 297 μm pore size mesh screen that was collected from the center of each channel at 50% river depth using a Shurflo submersible pump between 9 am and 12 pm. The sampling depth was chosen as it represents the rough average sediment distribution along the river water column (Filizola & Guyot, 2004). In 2 or 3 chambers per station, 2.85 L of river water was incubated for 24-hours at 60 rpm rotation with no nutritional amendments. Chambers were overfilled with river water for several minutes without bubbling and were wrapped with black tape and set in the shade to keep water temperatures close to that of the 50% river depth (river and chamber temperatures were within ~1 °C and ranged between 25 and 28 °C). Chamber and water collection equipment were acid-washed and rinsed with deionized (DI) water prior to and between stations. Samples for peptidomics (between 150-300 mL water) were filtered onto sequential, pre-combusted 2.7 and 0.3 μm

(nominal) pore size glass fiber membranes in triplicate at the beginning and end of each incubation. Proteomic samples were stored in 2 mL cryovials with 1 mL RNAlater at -20 °C in the ship's freezer until they could be shipped to the University of Washington in Seattle, whereupon they were stored at -80°C before protein extraction. Pre- and post-incubation filtrate samples (1 L in triplicate) were kept for geochemical analyses frozen in amber polypropylene bottles at -20 °C.

Rotating chamber incubations were paired with simultaneous BOD respiration measurements (Benner et al., 1995; Ellis et al., 2012; Ward et al., 2018) performed in triplicate in autoclaved, 60 mL glass bottles. BOD incubations were performed in the dark in foil-lined Styrofoam coolers in the shade, keeping them within several °C of the river temperature. Dissolved oxygen was measured before and after 24-hour incubations by Winkler titration (Carpenter, 1965) in triplicate BOD experiments for each station with water from 50% depth collected at the same time as the rotating chamber incubations (see Table 5.8 for sampling depths and parameters). Oxygen drawdown was assumed linear over the incubation period.

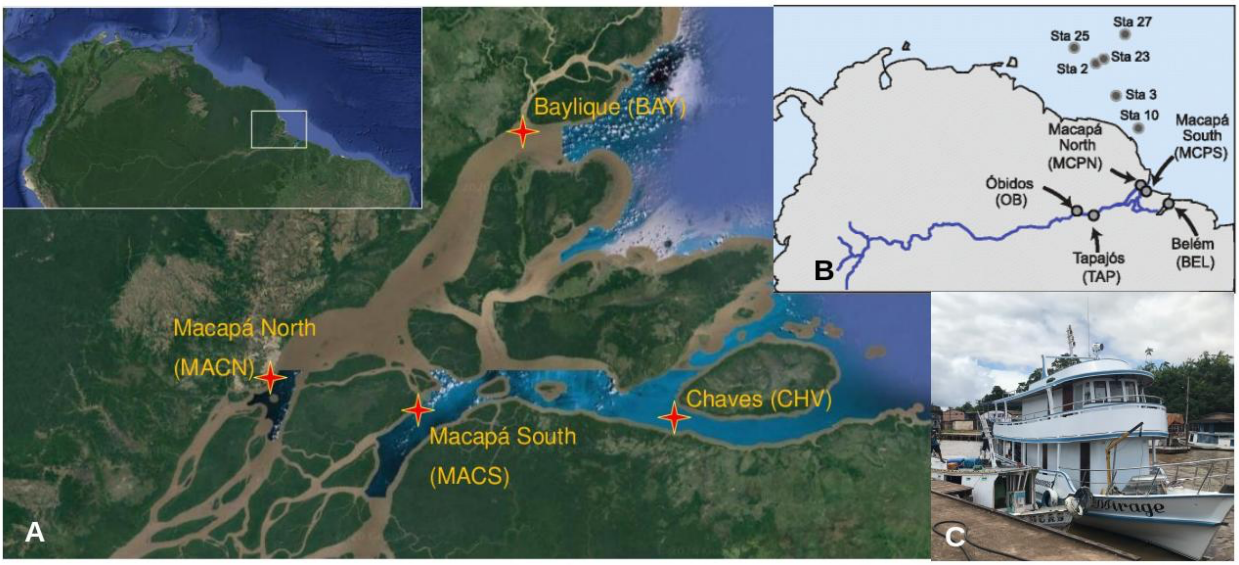

Figure 5.1. Sampling stations in the lower Amazon River tidal reach (A) in April 2019. Metagenome stations used in the creation of a protein reference database as described in Satinsky et al. (2014) (B). Sampling occurred onboard the B/M Mirage (C).

5.2.2 *Tidal reach physiochemical conditions*

Underway surface water temperature, conductivity, pH, dissolved oxygen (DO), turbidity, and chlorophyll-*a* were measured continuously at each station with a Yellow Springs Instrument Company (YSI, Yellow Springs, OH) EXO2 multi sonde. Depth profiles were collected at each station at the same time water for incubations was collected. Parameters measured included turbidity (FNU) (860 ± 15 nm excitation, 90° scattering) and chlorophyll-*a* fluorescence (Chl-*a*) (excitation at 470 nm ± 15 nm and emissions ± 685 20 nm). FNU stands for Formazin Nephelometric Units, most used in measurements of scattered light from a sample at a 90° angle from the incident light, as per the ISO 7027 (European) turbidity method. Turbidity here refers to the amount of light attenuation and backscattering due to suspended solids and dissolved load. Chl-*a* is a commonly used proxy for chlorophyll-*a* concentration, though fluorometric Chl-*a* determinations have potential interferences in systems like the lower Amazon River. For instance, terrestrial humic compounds and colored dissolved organic matter (CDOM) can also absorb in shorter wavelengths, resulting in Chl-*a* overestimation (Roesler et al., 2017). Unfortunately, higher quality measurements of chlorophyll-*a* concentration (high performance liquid chromatography (HPLC) pigment data) are available for only the Macapá stations during an earlier year (May 2011 from Satinsky et al., 2015). Underway and depth profile data were logged every 1 minute and 1 second, respectively.

5.2.3 *Protein preservation with RNAlater*

Proteomic/peptidomics experiments require immediate preservation, typically achieved in laboratory settings through flash freezing with liquid nitrogen or immediate storage of small samples at -80 °C. Such measures were impossible for these incubation experiments conducted on a boat without a -80 °C freezer or liquid nitrogen availability or storage. As an alternative, I endeavored to keep peptide bonds intact via preservation in a high ammonium salt solution, which has been shown to work for seawater proteomics samples in the context of autonomous underwater vehicle sampling (Saito et al., 2011).

To confirm that treatment with RNAlater was not significantly detrimental to protein recovery and peptide detection through high resolution liquid chromatography-mass spectrometry (HRLC-MS), I conducted a trial experiment using filtered water from Macapá South amended with a protein standard, 200 μg of bovine serum albumin (BSA). River water from 50% depth (300 mL)

was filtered onto 0.3 μm glass fiber and 0.2 μm Supor polyethersulfone membranes (Pall Laboratory, Port Washington, NY) spiked with 200 ug BSA and either 1) immediately transferred to a -80 °C freezer before shipping on dry ice to the USA (an ideal scenario not feasible on these river expeditions) or 2) submerged in RNAlater and shipped at room temperature. In a set of 12 samples (6 with RNAlater, 6 without), filters treated with RNAlater yielded slightly fewer PSMs on average, yet more overall peptide identifications (see Chapter 5 Supporting Information). I proceeded with the use of RNAlater to preserve samples before they could be frozen at -80 C in Seattle. Acetone precipitation was introduced into the extraction protocol to better avoid carrying salts through to the digested phase. An extra desalting step was also added to mitigate the high levels of salt introduced to the samples.

5.2.4 *Mineral active site swamping with lysine*

Mineral surfaces have been demonstrated to absorb and protect organic matter from degradation in a range of aquatic environments (Keil et al., 1994; Mayer, 1994; McGivern et al., 2021) including the Amazon River (Aufdenkampe et al., 2001). In temperate soils, up to 80% of organic nitrogen is proteinaceous and mineral-bound (Wang et al., 2020). Since the interest of this work is in extracting and characterizing the entire metapeptidome of a given sample, including detrital proteinaceous compounds potentially sorbed to mineral surfaces. Methodologically, the goal is to preserve the amide bonds and sequences of sorbed peptides, so strong acids or bases aren't desirable during extraction. To separate and solubilize sorbed proteinaceous compounds and prevent their reabsorption, I attempted adding minimal detergent and a free, basic amino acid (lysine) to bind and 'swamp' mineral sites (see Supporting Methods). I also used a 50 mM ammonium bicarbonate extraction buffer to increase the pH from ambient river conditions (~4.5-6) to 7.8 in order to decrease the binding strength of biomolecules to clay mineral surfaces, which are expected to be dominant in the lower Amazon fine particulate phase (Armijos et al., 2020). Detergents are increasingly used in environmental protein studies to solubilize membrane proteins (Whitelegge, 2013), and may be useful in mineral-peptide desorption. The idea of swamping activated mineral sites with small, basic free amino acids is derived from their high binding affinities for clay surfaces (Bu et al., 2019), and to my knowledge has not been investigated in MS-based proteomic/peptidomics.

I performed an experiment to test the swamping of active mineral sites with lysine. River water from 50% depth (300 mL) was filtered onto 0.3 μm glass fiber membranes and extracted as above but with prior additions of 0, 10, 50, and 100 mg lysine and 200 μg BSA protein recovery standard. The lysine treated samples had greater numbers of MS1 scans, yet ~30% fewer MS2 scans, peptide-spectrum matches, and peptide identifications than the non-lysine treated samples. The experiment had few controls, but it seems likely that the lysine decreased the chromatography of the samples and created spectral noise that lowered the quality of the actual peptide spectra. Overall, I learned this lysine swamping protocol wouldn't improve protein extraction in sediment-laden Amazon peptidomics samples, though a mineral-cleaning step could be helpful in future experiments to determine the best mineral-bound protein recovery protocol more resolutely. For instance, some soil fungi and bacteria secrete oxalic acid to liberate mineral-bound protein (Wang et al., 2020), which could inspire a biomimetic approach to proteomic/peptidomics sample preparation in the future.

5.2.5 *Protein extraction and sample preparation*

Protein was extracted from entire 27-mm glass fiber filters using an acetone precipitation protocol modified from Bridoux et al. (2015). Filters were thawed and sliced into ~7mL of Optima ultrapure water with 0.00625% Rapigest SF (Waters Corporation, Milford, Massachusetts) on ice. The chilled suspension was lysed via three cycles each of mechanical disruption with silica beads (50% 100 μm diameter and 50% 400 μm), freeze-thawing in a methanol-dry ice bath, and 2 minutes in a high-power water bath sonicator. The slurry was centrifuged at 4300 rpm at 4 °C for 5 minutes to yield a debris pellet and supernatant. The supernatant was transferred to a clean 35 mL Teflon centrifuge tube and precipitated with ice-cold acetone (supernatant/acetone, 1:4 v/v) for at least 12 hours at -20 °C. The precipitated protein mixture was then centrifuged at 20,000 x g for 30 min at 4 °C, and the resulting protein pellets were rinsed twice with ice-cold acetone and dried under N_2 gas. Dried pellets were resuspended in 600 μL of 50 mM ammonium bicarbonate buffer with 5 mM EDTA.

After resuspension and 3-minute centrifugation at 4800 rpm, an aliquot of the supernatant was removed after and the protein concentration was determined using a modified Lowry assay using reagents from Bio-Rad (Hercules, California). Extracted protein was then subjected to reduction of disulfides with dithiothreitol and carbamidomethylation with iodoacetamide. The

protein extraction was subjected to in-solution digestion with trypsin (Promega Gold) overnight at room temperature by a ratio of 1 µg trypsin: 25 µg total protein in 50 mM ammonium bicarbonate buffer with 5 mM EDTA. Samples were desalted twice (because of unusually high salt load from preservation using RNAlater) using NestGroup macro-spin C18 columns (Southborough, Massachusetts) and resuspended in 5% acetonitrile with 0.1% formic acid and Waters Hi3 E. coli peptide standard mixture (100 fmol/L).

5.2.6 *Liquid-chromatography high resolution mass spectrometry*

Reverse-phase liquid chromatography-high resolution mass spectrometry (LC-HRMS) analysis was performed in duplicate with a Thermo Science (Waltham, Massachusetts) EASY-nanoLC system coupled to a Thermo Orbitrap Fusion Tribrid HRMS. Peptides were separated on home-packed fused-silica capillary analytical columns 25- 37 cm long with 75-µm i.d. and C18 particles (Magic C18AQ, 100˚A, 5mm; Michrom). Analytical columns were coupled to a 4 cm long, 100 m i.d. precolumn (Magic C18AQ, 200˚A, 5mm; Michrom). Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100% LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 120-minute gradient from 5-35% solvent B. All analyses were carried out in positive mode at an NSI spray voltage of 2 kV. Data-dependent acquisition (DDA) on the top 10 ions was carried out using higher energy collision dissociation (HCD) for duplicate injections, when possible. The MS1 (parent peptide ion) scan range was 400-2000 m/z.

5.2.7 *Peptidomic data analysis*

Peptides were identified from the raw mass spectra using a combination database-driven and database-independent *de novo* sequencing approach as described in Chapter 2. *De novo*-directed proteomics as described above (Figure 2.1) was performed using PeaksDB within Peaks Studio (v8.5; Bioinformatics Solutions, Waterloo, Canada; Zhang et al. (2012)). For database searches, the Amazon River basin non-redundant microbial gene catalogue (AMnrGC) (Santos-Júnior et al., 2020) was used as a protein reference, which amasses 106 metagenomes taken from 30 stations between the Amazon's headwaters and plume. Additionally, spectra were searched against a database of common mass spectral contaminants (Mellacheruvu et al., 2013). Search parameters for both database searching and *de novo* sequencing included 8 maximum modifications per peptide, 15 ppm peptide mass tolerance, and 0.5 Da fragment mass tolerance. Results from

technical replicates and 'biological' replicates (different filters from the same treatments) were combined.

I accepted *de novo* peptides >80% average residue local confidence (ALC) with no single amino acid score <50%. For database searches, a false discovery rate (FDR) was set <1.0% using a reversed database target-decoy strategy (i.e., searching against reversed reference protein sequences) (Elias & Gygi, 2010). *de novo* sequencing was also incorporated into the database searches, as PeaksDB first compares *de novo* sequences to the reference database to find approximate matches and decrease the search space. Agreement between *de novo* sequences and database search hits are also used, in part, to generate peptide-level confidence scores derived from the p-value indicating the statistical significance of the peptide-spectrum match (the -10lgP score). The selected threshold was a -10lgP score >20. Such a score is equivalent to a p-value of 1%, signifying the probability that the identification is to a false peptide sequence (Zhang et al., 2012). Protein identifications are notoriously difficult in samples containing many different organisms because some peptides are shared in proteins from multiple taxa. I required a minimum of 1 unique peptide per protein identification. Matching of proteins from peptides found only by *de novo* sequencing is detailed below.

Mapping *de novo* peptides to proteins - The *de novo*-directed PeaksDB workflow used here outputs peptides matched to the database and *de novo* peptides (sequences only, no additional information). To identify peptides that came from the diatom detritus or bacterial proteins not found by database searching, the *de novo* peptides were aligned to the reference database (Amazon Gene Catalog) using PepExplorer (Leprevost et al., 2014). PepExplorer considers common *de novo* sequencing errors and limitations (such as leucine and isoleucine equivalence or other combinations of amino acids having the same mass) and identifies regions of local similarity between sequences. I also performed an alignment of the sequences against a reversed version of the database to estimate a false discovery rate, which was kept <1%. Alignments were accepted at a 95% identify agreement cutoff and protein identification required at least one unique peptide alignment.

Many *de novo* peptides did not match back to the reference database, which was not unexpected given that the Amazon Gene Catalog database only contains samples and sequences from the upriver stations at Macapá, and not the oceanward stations at Baylique and Chaves. To

overcome this limitation, I additionally mapped the *de novo* peptides to proteins contained in the entire UniProt KnowledgeBase database (UniProt Consortium, 2018), which contains over 200 million sequences from thousands of taxa. The mapping was performed using the UniPept lowest common ancestor tool (Mesuere et al., 2016), which is built specifically for metapeptidomic data and determines the taxonomic origins of peptides to the lowest possible phylogenetic rank (since some peptides are highly conserved, they may match as to only 'Bacteria', or even 'Organism', rather than to a species or genus level). The UniPept output provides the best view of the bacteria present in the experiment given the lack of genes or transcripts from which to build an ideal reference database.

Amino Acid Modifications - Since amino acids are frequently modified after translation, either for cell-specific purposes or during degradation, a computationally efficient method is needed to search for the myriad possible post translational modifications (PTMs). I used the open modification search tool PeaksPTM (Han et al., 2011), with parameters set to tryptic or non-enzymatic constraint, 2 or fewer missed cleavages, 15 ppm peptide mass tolerance, 0.2 Da fragment mass tolerance, minimum A-score > 200 (a measure of modification location confidence), fixed carbamidomethylation of cysteine, and variable oxidation of methionine. Based on PeaksPTM results, the most frequently occurring modifications were used to evaluate whether adding additional PTMs to the overall searches altered the rate of false discovery. That is, I sought the optimal balance between searching for PTMs while avoiding a vast search space that leads to decreased sensitivity (Noble, 2015; Timmins-Schiffman et al., 2017). A series of PeaksDB searches and Peaks sequencing runs of the combined data set (using same search parameters as above) with increasing numbers of variable modifications was performed to find the optimal set of PTMs to include in searches (most peptide identifications <1% FDR). Using these PTM ramping results, 10 optimal variable PTMs were selected for the final searches. They included methionine oxidation, deamidation of asparagine, and deamidation of glutamine.

Protein annotations: Gene Ontology terms - To identify gene ontology (GO) term annotations, the peptide sequences were aligned to the UniProt protein database using UniPept's metapeptidomic functional analysis tool (Gurdeep Singh et al., 2019), which is built upon a lowest common ancestor peptide search described above. The GO categories were condensed from the

broader set of results to eliminate redundancy using the REViGO tool (available at http://revigo.irb.hr/) as well as manually. Peptide GO terms were adjusted to spectral peak abundance prior to comparison between samples.

5.3 RESULTS

5.3.1 *Station properties*

In this study, I sample at the two established Macapá stations in each channel (Macapá North and South), but also venture further oceanward into the tidal reach by approximately 130 km, much closer to the true mouth of the Amazon. At the oceanward stations, Baylique in the Northern Stem and Chaves in the Southern, there is no salt intrusion; salinities were the same as the Macapá stations (0.01 PSU, Table 5.8) and remained constant throughout the depth profiles. Surface chlorophyll was slightly higher at the Macapá stations compared to their oceanward station counterparts, possibly because surface turbidity measurements were lower at Macapá and increased towards the mouth. Parameters shown in Table 5.8 were from depth profiles taken just before sampling for incubations and metapeptidomics, for which 50% depth river water was used.

Table 5.8. Lower Amazon River sampling stations and selected parameters.

Station	Average channel depth (m)	50% sampled depth (m)	50% depth average DO (%)	50% depth turbidity (FNU)	50% depth salinity (PSU)	50% depth pH	Surface turbidity (FNU)	Surface chlorophyll (µg/L)
Macapá North	24	12	89.4	67.3	0.01	6.38	79.43	1.67
Macapá South	25	12	86.5	65.3	0.01	6.31	70.84	1.99
Baylique	31	14	91.5	68.9	0.01	4.72	123.22	0.45
Chaves	28	14	93.0	43.4	0.01	4.76	111.01	0.59

5.3.2 *Protein and bulk organic matter degradation rates*

Bulk respiration rates from BOD incubations were in the range of 0.120 ± 0.012 to 0.165 ± 0.027 mg C/L/day after correction given the demonstrated underestimation of BOD compared to rotating chambers (Ward et al., 2018) (Figure 5.2a). These rates are higher than those measured at true high water by Ward et al. (2013) at Macapá stations in May 2011 (0.075 ± 0.012 mg C/L/day with the same corrections). This is unsurprising, given observed decreases in bulk respiration with discharge in the tidal reach (Ellis et al., 2012; Ward et al., 2016). At the initial timepoint of the incubation, protein concentrations measured by a colorimetric modified Lowry assay in the two size fractions were similar between the Northern and Southern channels but changed from upriver to oceanward stations. In the small particulate and free-living bacteria fractions (0.3 - 0.7 μm), protein concentrations were 48% higher at the upriver Macapá stations compared to the oceanward stations (Figure 5.2b). The large particulate fraction (>0.7 μm) had an opposite trend, with protein concentrations 144% higher at the oceanward stations than at Macapá (Figure 5.2c). Protein degradation rates followed the same pattern of magnitude difference (Figure 5.2-e).

Protein degradation rates in the small particulate/free living fraction (0.3 - 0.7 μm) ranged from 1.418 ± 0.10 and 1.487 ± 0.15 mg protein L^{-1} day^{-1} at Macapá North and South, respectively, to 1.016 ± 0.13 and 0.76 ± 0.09 mg protein L^{-1} day^{-1} at Baylique and Chaves, respectively. In the larger particulate fraction (>0.7 μm) protein was degraded at higher rates in the oceanward stations (2.666 ± 0.32 and 1.747 ± 0.25 mg protein L^{-1} day^{-1} at Baylique and Chaves, respectively) than the Macapá stations (0.338 ± 0.18 and 0.384 ± 0.21 mg protein L^{-1} day^{-1} at Macapá North and South, respectively).

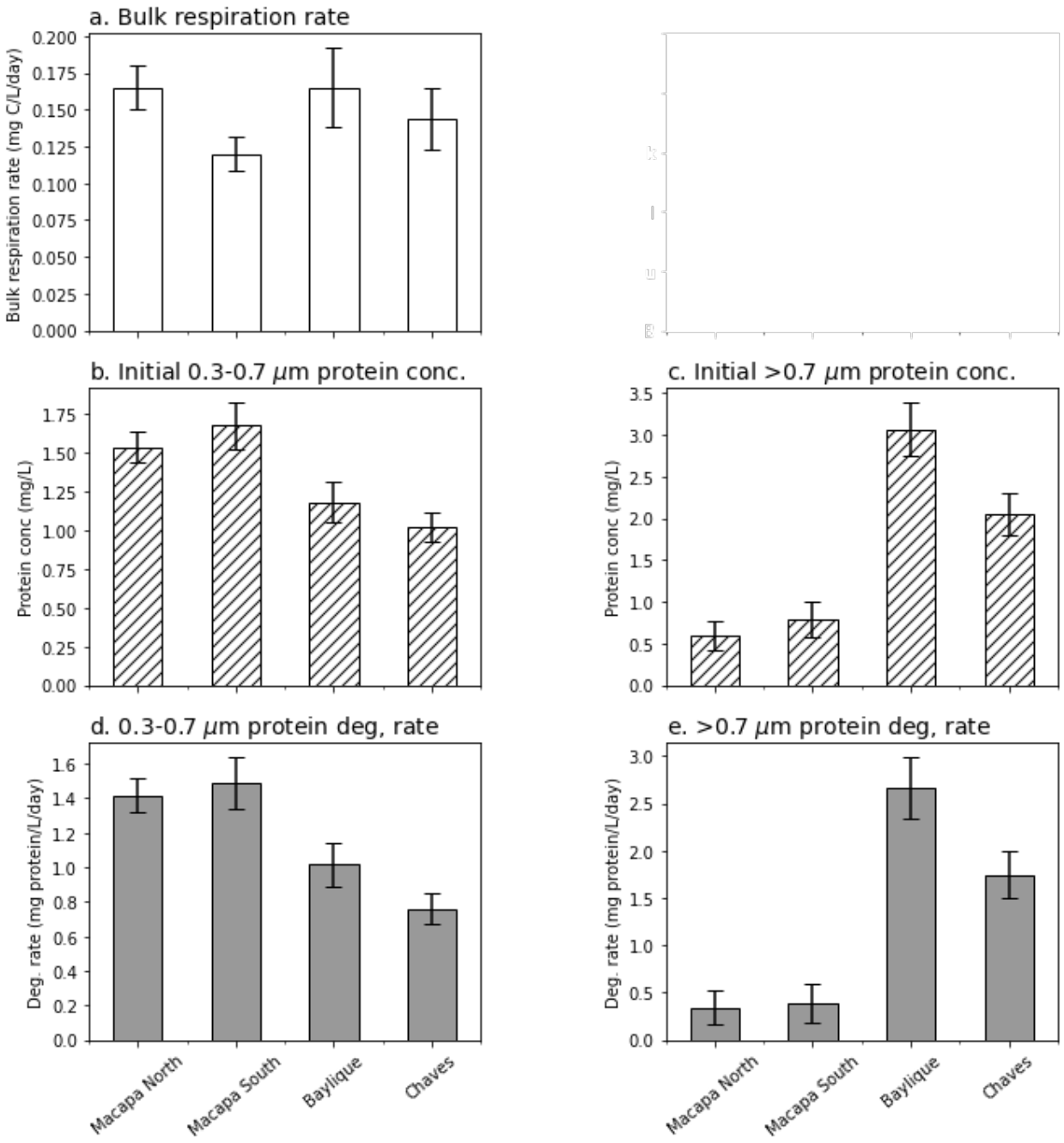

Figure 5.2. Bulk and particulate protein degradation rates in the lower Amazon River. A, Bulk respiration rates from 4 stations along the tidal reach of the lower Amazon in April, 2019. B, the degradation rate of protein in the free-living/small suspended particles (0.3-0.7 μm) in 24-hour incubations. C, the degradation rate of protein in large suspended particles (>0.7 μm) in 24-hour incubations.

5.3.3 *Protein and peptide identifications*

Peptides from the beginning and end of the rotating chamber incubations were identified from a combination of database searching against the Amazon Basin non-redundant Gene Catalog (Santos-Júnior et al., 2020) and database-independent *de novo* peptide sequencing (Table 5.9). The contribution of peptides found solely by *de novo* sequencing rather than both *de novo* and database searching was higher (average 36.3 and 39.7% in Chaves and Baylique, respectively) in the oceanward stations than at the Macapá stations (average 26.9 and 26.7% in Macapá South and North, respectively). This difference most likely reflects the origins of the reference database, which includes sequencing results from the Macapá stations but not the oceanward stations (see Methods and Santos-Júnior et al., 2020).

Table 5.9. Unique peptide identifications from 4 stations in the lower Amazon River tidal reach, April 2019.

Incubation metapeptidomes	Macapá South peptides	Macapá North peptides	Chaves peptides	Baylique peptides
0.3-0.7 μm Time 0	820	517	985	661
>0.7 μm Time 0	896	411	372	368
0.3-0.7 μm Time 24 hr	740	892	871	862
>0.7 μm Time 24 hr	835	638	649	356

5.3.4 *Protein subcellular localizations*

Peptides were sorted from each station's two metapeptidomes (time 0 and time 24-hours of the incubation) by their gene ontology cellular compartment terms (GO Terms; Figure 5.3; see also Mikan et al., 2020); Riffle et al., 2017). The GO term data were aggregated into eight major cellular compartments classifications, and the three most numerically dominant groups were

integral component of membrane (blue), cytoplasm (orange), and plasma membrane (green) proteins (Figure 5.3). Here, 'integral component of membrane' denotes a peptide sequence mapped to a parent protein that is all or partially embedded in the hydrophobic region of a membrane, i.e., a transmembrane protein. Of the bulk metapeptidomes, distribution of peptides from the classes of cellular compartment was unchanged between stations and over 24 incubation periods. The peptides assigned to *in situ* primary producers (46 peptides across the dataset from organisms like diatoms, dinoflagellates, and green algae, *see below*) had distributions that included chloroplast peptides and relatively fewer transmembrane peptides than the bulk metapeptidomes at time 0. After 24-hours, primary producers' peptides (24 remaining) were solely from transmembrane proteins or proteins in the nucleus (Figure 5.3).

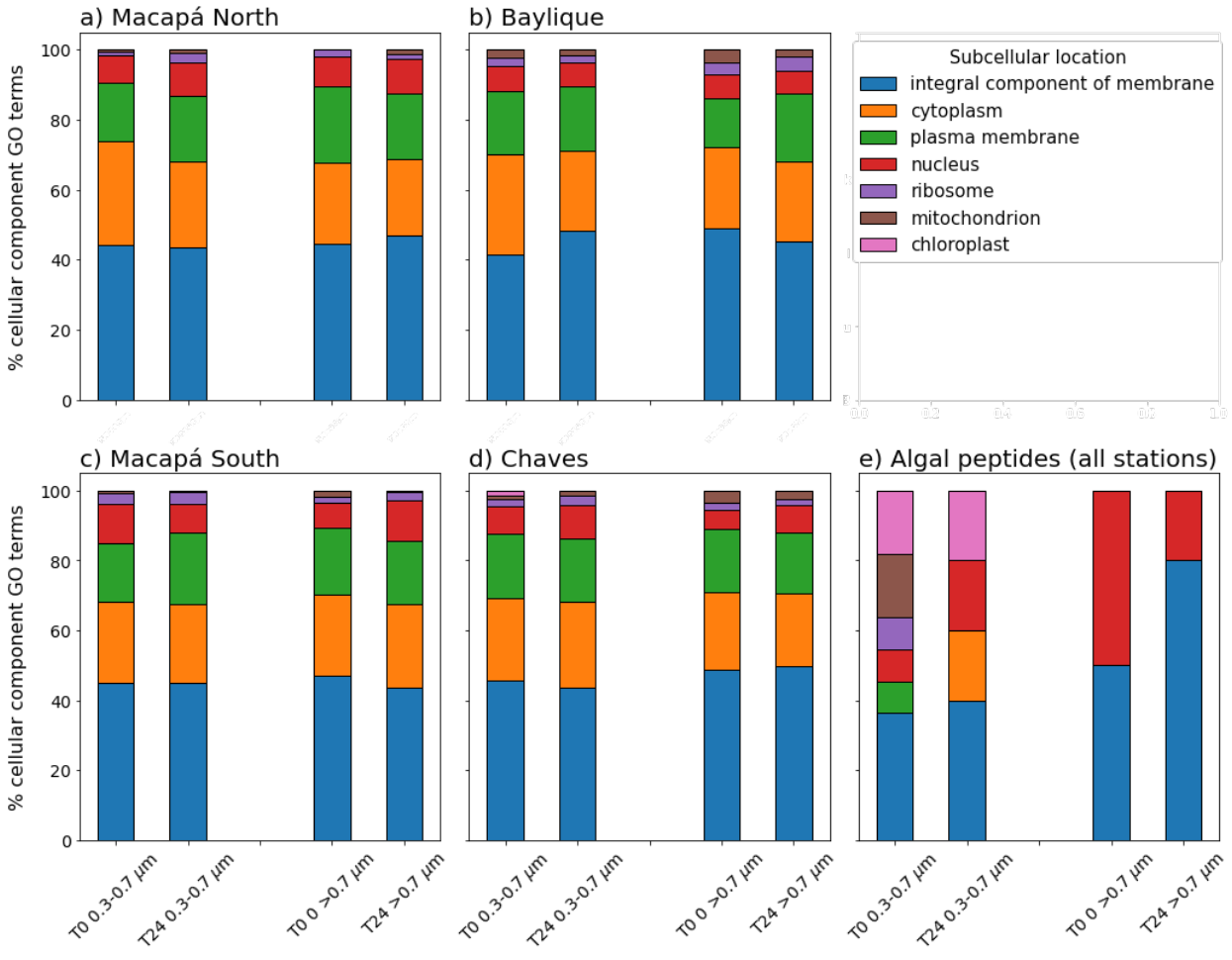

Figure 5.3. Distribution of subcellular protein locations in proteomes of incubations in lower Amazon River tidal reach, April 2019. A., Macapá North time 0 and time 24-hour proteome

subcellular protein distribution as determined by Gene Ontology (GO) peptide annotations. B., Baylique. C., Macapá South. D., Chaves. E., Subcellular protein GO terms from the 46 and 24 algal peptides in these samples. While the overall core GO terms show a subcellular distribution close to that of living bacteria, the peptides from just the primary producers (diatoms, green algae, and cyanobacteria) are dominated by the more recalcitrant nuclear and membrane proteins.

5.3.5 *Post-translational modifications*

A non-discriminatory 'open' search of thousands of possible PTMs in the UniMod database using PeaksPTM was performed (see above in Methods). Only the 3 most frequently occurring mass modifications – methionine oxidation, asparagine deamidation, and glutamine deamidation – were then used for subsequent analyses. All three of these PTMs have been implicated in protein degradation, both in the environment (Abdulla et al., 2018; Kim et al., 2014; Schmidt et al., 2011) and, especially for methionine oxidation, also in laboratory preparation of proteomics samples (Ghesquière & Gevaert, 2014). Rates of artificial methionine oxidation by the extraction and digestion methods used here result in less than 4% of all methionine oxidations (Froelich & Reid, 2008), suggesting that the values obtained in this study (10%) are environmental and not analytical. After adjusting for peptide relative abundance using peak areas, changes in PTM profiles during the 24-hour incubation were all <10% for the 3 modifications searched for (stations metapeptidomes combined) (Figure 5.4). The overall 'modifiable peptide space', that is, the percentage of amino acids in peptides that could have PTMs in the search parameters (methionine, glutamine, and asparagine) increased by <3% for both size fractions in the 24-hours incubations. Between stations, there were also no large differences in PTM profiles.

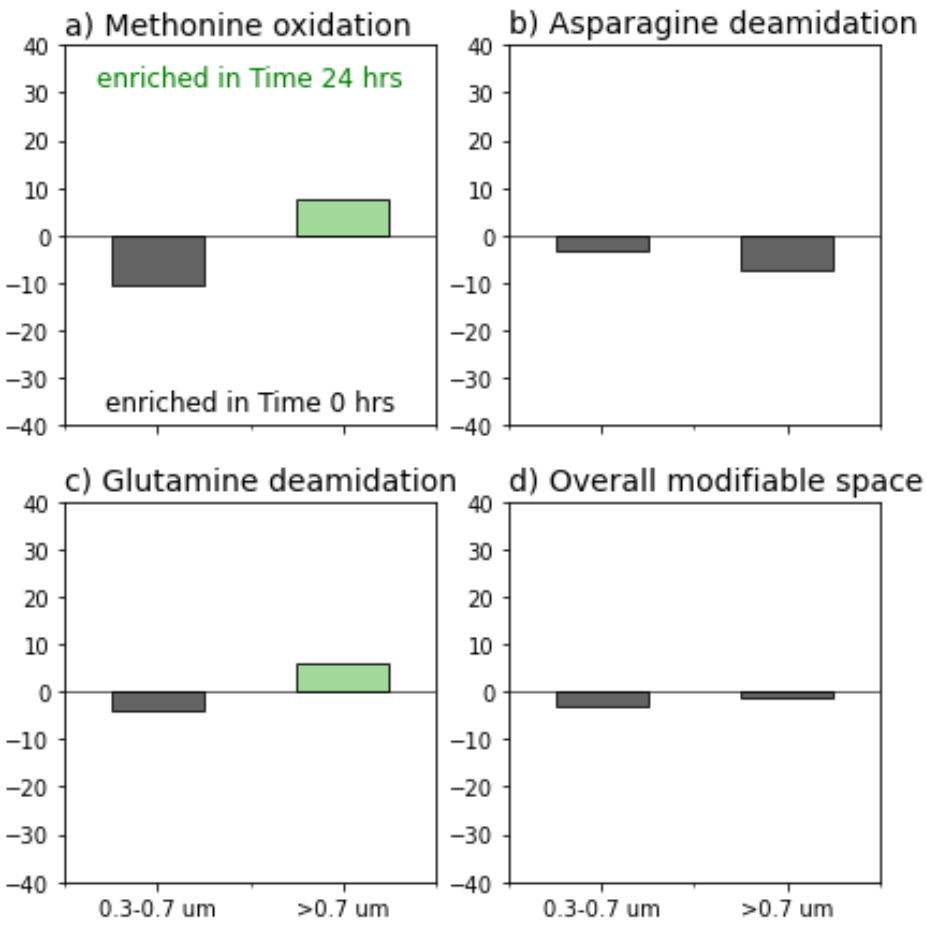

Figure 5.4. Extent of modifiable residues with post-translational modifications (PTMs) at all stations combined. Variable PTMs included A., methionine oxidation, B., asparagine deamidation, and C., glutamine deamidation. The overall modifiable space (the percent modified of all methionine, asparagine, and glutamine residues in all peptides) is shown in panel D. Peptides and their PTMs were adjusted by peak area prior to comparison.

5.3.6 *Relative amino acid compositions*

The changes in relative amino acid composition of proteinaceous compounds in the 0.3-0.7 µm and >0.7 µm size classes over the course of the incubations was calculated using peptide abundances adjusted by spectral area abundance (Figure 5.5). Earlier experiments have shown that total hydrolyzable amino acid analysis and my metapeptidomic approaches yield similar amino acid compositions (see Chapter 3 and Duffy et al., *submitted*). Overall, the relative amino acid compositions show little change during the 24-hour incubations which is expected given the other

evidence that this is a pool of labile and efficiently recycled proteinaceous material primarily from living biomass. Though there were some changes in the relative amino acid composition of the metapeptidomes over the course of the incubations (shown as log2 fold changes in Figure 5.5), they were not consistent across incubations for a given size class and are likely noise in datasets of several hundred peptides per sample. The four stations also had similar amino acid compositions (Supporting Figure 5.12-Figure 5.15).

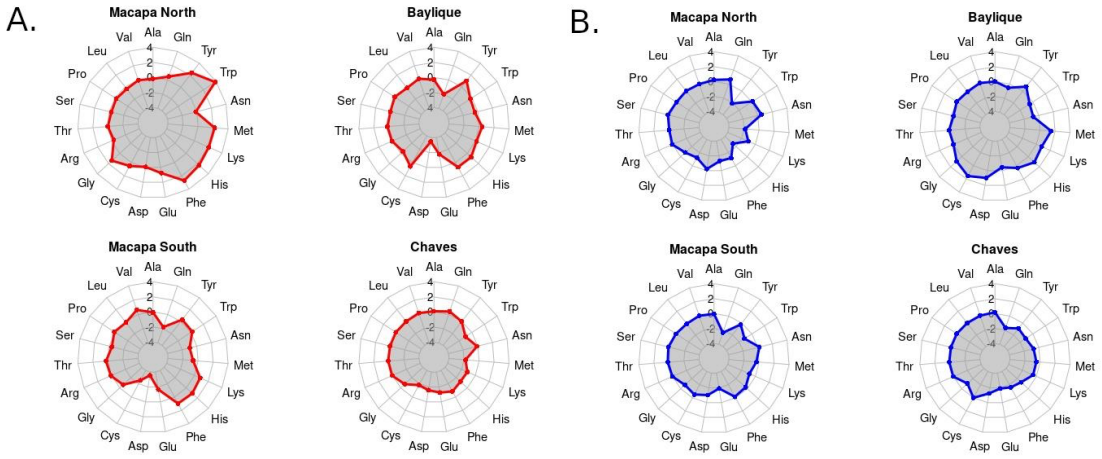

Figure 5.5. Relative amino acid composition changes during 24-hour incubations. Radar plots showing log2 fold change in individual amino acids (AA) over 24-hour incubations. Red and blue lines demarcate the log2 fold change as <0 (decrease in that AA's relative abundance with tine) or >0 decrease in that AA's relative abundance with tine). A., Amino acid composition changes at upriver (Macapá North, Macapá South) and oceanward (Baylique, Chaves) stations for small suspended and free-living (0.3-0.7 µm) and B., particle-associated (>0.7 µm) fractions organic matter fractions. Amino acid mole fractions are determined from peptides adjusted by peak abundance. Leucine and isoleucine are combined since they are indistinguishable by *de novo* peptide sequencing. Amino acid compositions for each metapeptidome can be found in Supporting Figure 5.11 - Figure 5.15.

5.3.7 *Microbial biodiversity*

Thousands of peptides across the 4 stations were identified by mass spectrometry and combined *de novo* and database searching (Table 5.9), and peptide sequences were in turn used to identify

their source taxa from the UniProt KnowledgeBase (Gurdeep Singh et al., 2019). Peptides were assigned to bacterial, archaeal, eukaryotic, and viral taxa based on uniqueness of a given sequence to each organism's known proteome. The metapeptidomes of all stations were almost entirely composed of Actinobacteria (75-85% of peptides after peak area adjustment) and Proteobacteria (15-20%) (Figure 5.6). Of the actinobacterial peptides, those assigned to the class Thermoleophilia were the most abundant at all stations, with contributions from the class Actinomycetia only at Macapá North in the 0.3 - 0.7 μm size fraction at the onset of the incubation (Figure 5.6). Mainly the representatives of Proteobacteria were from the class Gammaproteobacteria, some of which are known symbionts of fungi and larger bacterial hosts. Their role in tropical rivers has been shown to be the utilization of smaller molecular weight compounds produced by their saprophytic hosts or associates (Mai et al., 2018). However, some Gammaproteobacteria have been shown, like Actinobacteria, can also depolymerize lignin (Bugg et al., 2020).

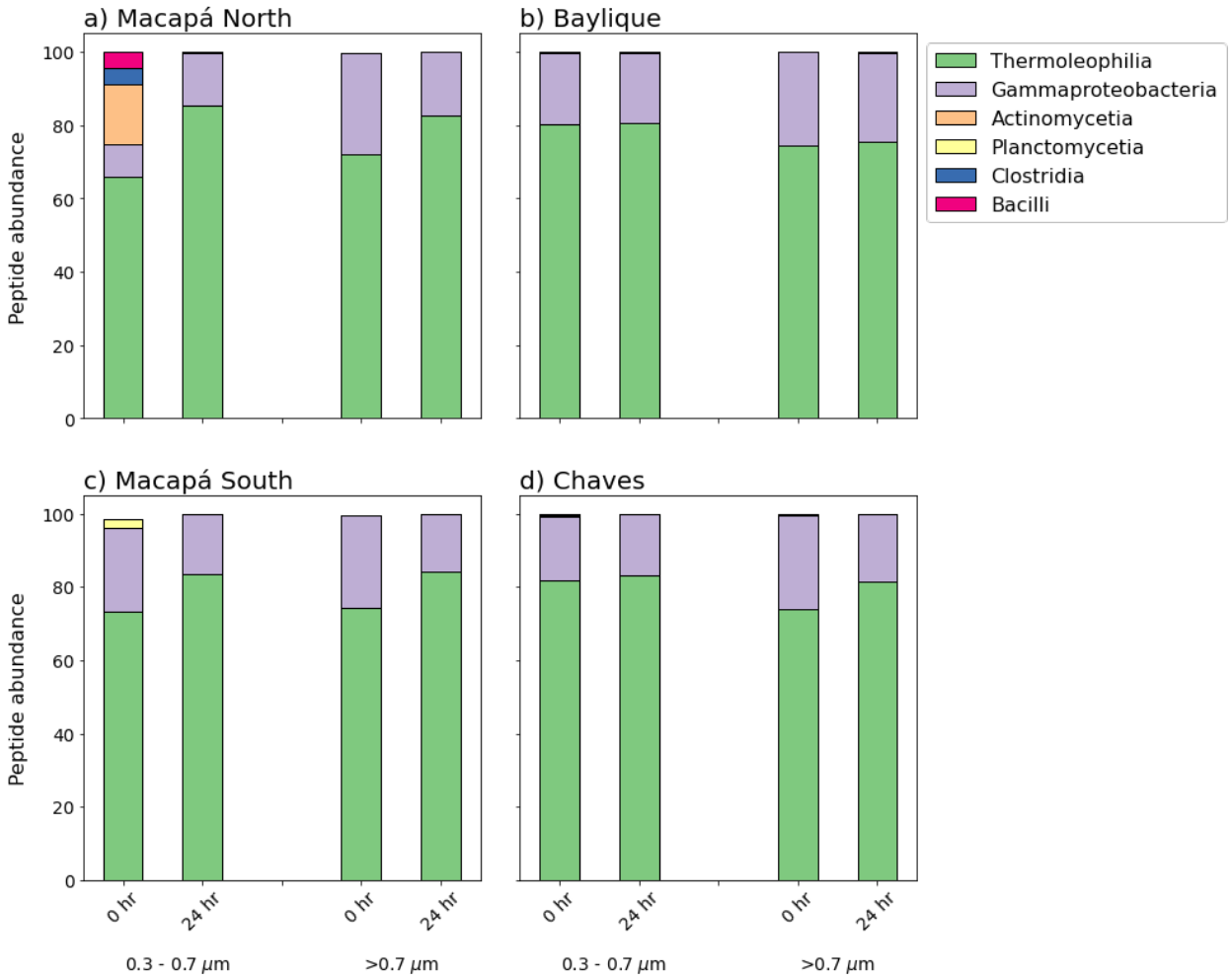

Figure 5.6. Relative abundance of peptides specific to the class level in 24-hour incubations at 4 lower Amazon River stations in April 2019. Initial and final metapeptidomes of small particle/free-living (0.3-0.7 µm) and >0.7 µm particle fractions at Macapá North (A), Baylique (B) Macapá South (C). and Chaves (D).

Less than 2% of peptides were linked to other phyla, which generally were present at the initial time points and absent after 24-hour incubations (Figure 5.11). These include 3 peptides from the bacterial phylum Calditchaeoeta, found at Macapá South, which degrade detrital protein (Marshall et al., 2017) and the bacterial phylum Balneolneota, found at Baylique, an obligate aerobe shown to utilize proteins and peptides for growth (Sorokin et al., 2018). Several viral peptides were observed at all stations, presumably from active viral infections as free virions would pass through

the filters used here unless attached to mineral surfaces or organic particles. Also identified are peptides from fungal lineages, namely Ascomycota, which were mainly found at Macapá North in both size fractions (Figure 5.7).

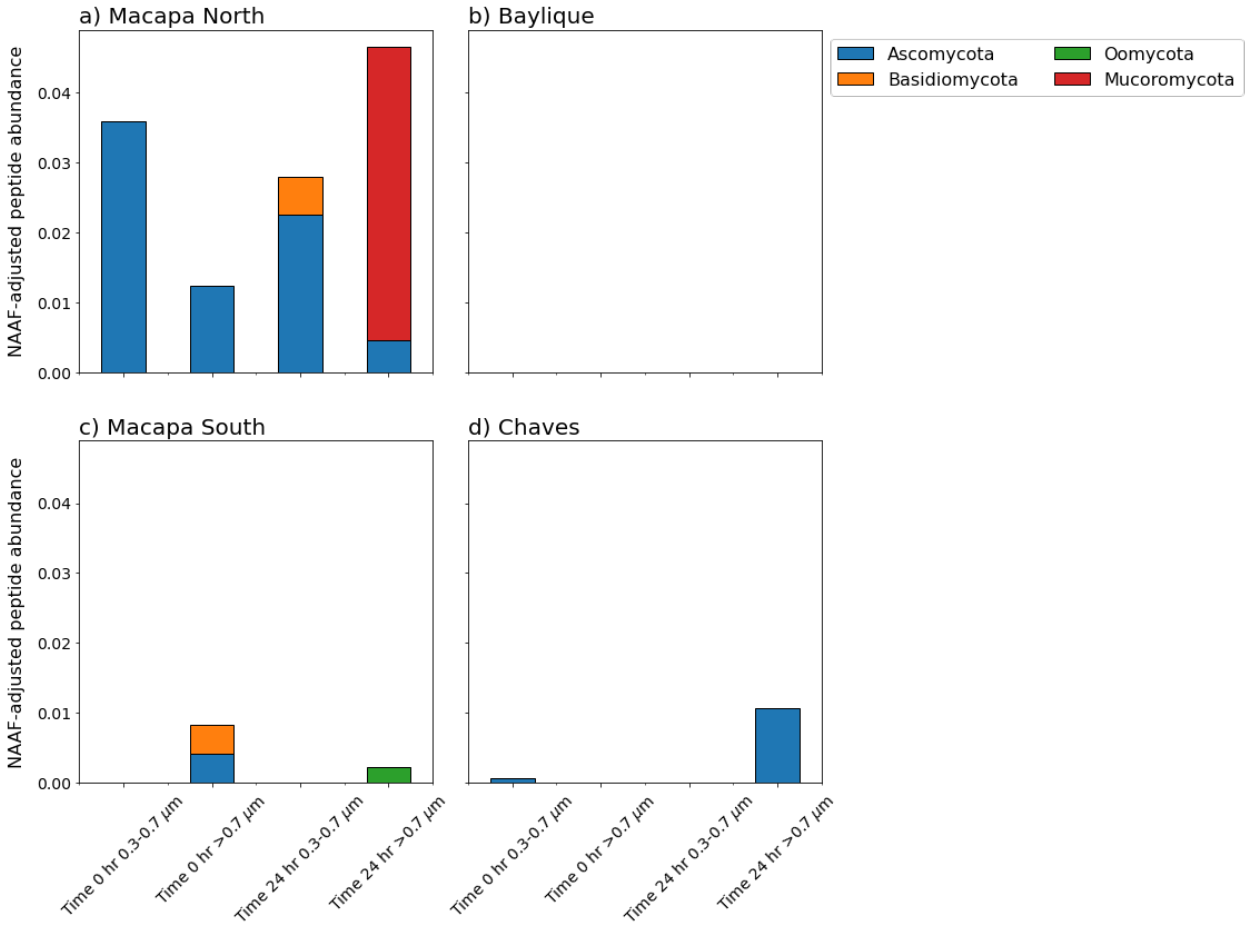

Figure 5.7. Peptides from fungi and fungal-like organisms in 24-hour incubations at 4 stations in the lower Amazon tidal reach. In total, 21 peptides specific to fungal or fungal-like lineages at the phylum level were identified at Time 0 after mapping to the UniPept Knowledgebase. After 24-hour incubation with no amendments in rotating chambers, 18 fungal peptides were identified in the metapeptidomes, mostly at the Macapá North station.

Also identified in all lower Amazon River metapeptidomes were peptides unique to primary producers, from the phyla Streptophyta (a vast clade of plants including macrophytes), Chlorophyta (green algae), Bacillariophyta (diatoms), Dinophyceae (dinoflagellates, all

photosynthetic genera), Chlorobi (green bacteria) and Cyanobacteria (small photosynthesizing bacteria). In general, more diatom proteins were identified at oceanward stations, and more dinoflagellates at the Macapá stations (Figure 5.8).

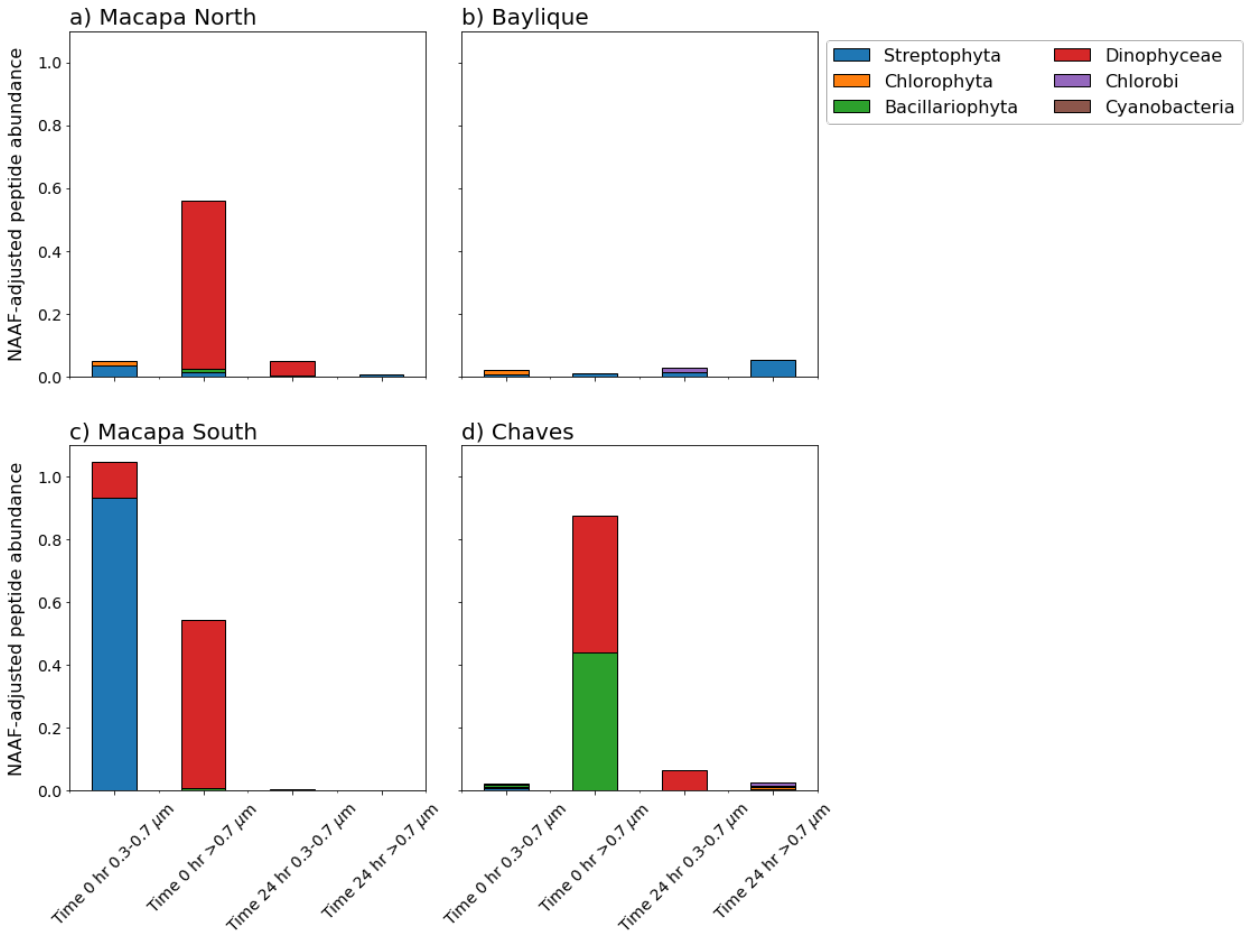

Figure 5.8. Peptides specific to classes of photosynthesizing microbes in 24-hour incubations at 4 stations in the lower Amazon tidal reach. In total, 46 peptides specific to primary producers were identified at Time 0 after mapping to the UniPept Knowledgebase. After 24-hour incubation with no amendments in rotating chambers, only 24 of these peptides were identified in the metapeptidomes.

5.3.8 *Microbial community function over incubations and stations*

Within the 16 metapeptidomes I evaluated (2 size classes at 2 timepoints from 4 stations), hundreds of database and *de novo* sequencing-identified peptides were matched to hundreds of peptide-level gene ontology (GO) term annotations (Figure 5.9). Even in the case when peptides couldn't be matched to a single organism's protein, but rather to a group of related proteins shared by multiple organisms, that peptide was useful in that it could be included in a semi-quantitative functional evaluation of each sample's metapeptidome (see Mikan et al., 2020).

The incubation period in this study was short (24-hours) and no organic matter or nutrients were added as amendments. No significant changes in community functional expression were observed over the course of 24-hours, which is not surprising given the short period of time. The majority of peptide GO term assignments were related to growth, reproduction, and energy acquisition (Figure 5.9). In the molecular function GO category, 'ATP binding' was consistently around 20% of all identified terms in each metapeptidome (Figure 5.9). Peptide GO terms also included those related to organic matter processing, including annotations for lignin and cellulose degradation (hydrolase and oxidoreductase activity, ~2.5-5.5 % each of all metapeptidome GO terms) which were not changed significantly between stations or over incubations. Organic polymer degradation (peptidase activity, ~6-10% of metapeptidome GO terms; and carbohydrate and chitin binding, ~0-2% of metapeptidome GO terms) and transmembrane transport peptides were also consistent across the 24-hour incubation period, between ~1-3 % of GO terms. Numerous peptides were also observed with metal-binding functions, including zinc ion binding (~2-5% of metapeptidome GO terms).

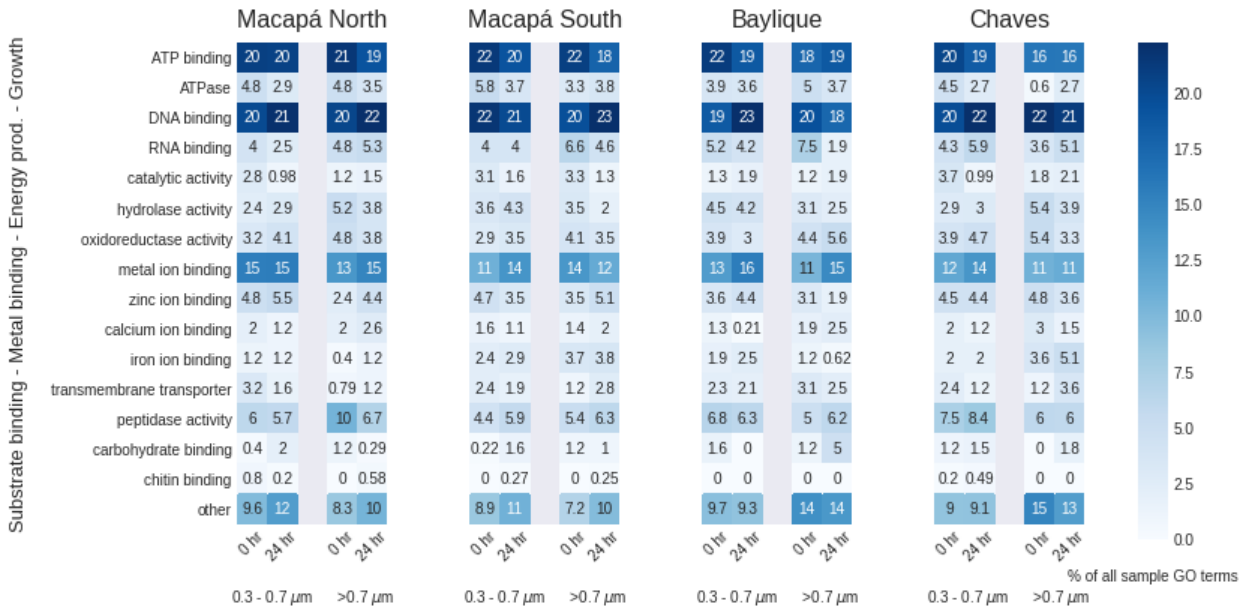

Figure 5.9. Functional annotations of identified peptides in lower Amazon River incubations. Peptide Gene Ontology (GO) term annotations found by mapping peptides to the UniProt KnowledgeBase using the UniPept algorithm (Gurdeep Singh et al., 2019) and adjusted by relative peptide peak area. Cells in the heatmap represent the percentage of each GO term category in each sample's entire suite of peptide GO terms.

5.4 Discussion

The lower Amazon River is a highly dynamic system, made difficult to access for scientific purposes due to unpredictable sandbars and wave action. Tides are measurable upriver as far as 800 km from the Atlantic; all the stations in this study experience semidiurnal flow reversals (Ward et al., 2015). Macapá stations have long been considered the 'mouth' of the Amazon in previously published work and is the last place where the Amazon is divided into two well constrained primary stems, the Northern and Southern. However, Macapá is approximately 150 km from the actual, highly channelized true mouth.

5.4.1 *Protein and bulk organic matter degradation rates*

I anticipated that proteins would represent a labile and quickly recycling pool of organic matter in the Amazon tidal reach (**H1**) (Mayorga et al., 2005; Richey et al., 1990), and this is borne out in the bulk particulate protein degradation rates (Figure 5.2). Rates of protein degradation in the small

particulate/free living fraction ranged from 0.76 ± 0.09 mg protein L^{-1} day^{-1} at Chaves to 1.487 ± 0.15 mg protein L^{-1} day^{-1} at Macapá South. Protein degradation rates in the larger particulate fraction (>0.7 μm) were between 2.666 ± 0.32 mg protein L^{-1} day^{-1} at Baylique and 0.338 ± 0.18 at Macapá North mg protein L^{-1} day^{-1}. In comparison, annual average rates of particulate (also >0.7 μm) lignin degradation as reported by Ward et al. (2013) at the Macapá stations is 0.57 ± 0.07 μg C L^{-1} day^{-1}. Converting the lowest protein degradation rates found here to carbon space (assuming proteins are on average 53% wt C), particulate protein is still degraded an order of magnitude more rapidly than particulate lignin, which is also found at much lower concentrations (22.6 ± 2.3 and 17.7 ± 1.8 μg L^{-1} at Macapá North and South in May 2011, Ward et al., 2015). However, I'm unable to make a comparison between these high protein degradation rates in large and small particulate fractions to the bulk respiration rates as there are no measurements of dissolved amino acid or protein degradation rates or estimates of bacterial protein recycling.

The larger particulate fraction (>0.7 μm) had higher protein concentrations at the oceanward stations than at the Macapá stations (Figure 5.2c), with degradation rates following the same pattern of magnitude difference (Figure 5.2d-e) This increase in larger particulate protein could suggest an input of labile organic matter in the larger size fraction as the river proceeds from Macapá to the oceanward stations, potentially *in situ*-produced algal material that would be caught on a 0.7 μm filter (*see discussion below*).

Respiration rates performed in the rotating chambers used here have been shown to account for all the measured CO_2 outgassing at Macapá South when spun at 60 rpm (Ward et al., 2018). In the same study, BOD respiration values were 30% of the measured CO2 outgassing. Bottle effects (Ionescu et al., 2015) may contribute significantly to underestimations of respiration, particularly in this system with high levels of suspended particles, highlighting the importance of particle suspension in correctly modeling the environment encountered by the microbial community. Suspended particles in the river water may act like particles in marine systems, where concentrations of carbon, nitrogen, and nutrients exceed those of the surrounding water by a factor of two or more (Blackburn et al., 1998). As such, particles are substrates for microbial enzyme activity (Catalan, 2015) and contribute 'priming' effects, when small amounts of labile organic matter stimulate the degradation of previously unreactive or recalcitrant organic matter (Ward, 2016). Previous studies have demonstrated that particle-associated bacteria (2.0 to 297 μm) are

more abundant and metabolically active than free-living bacteria (0.2 to 2.0 μm) (Satinsky et al., 2015).

Respiration rates in the lower Amazon have been shown to track the seasonality of discharge. Benner et al. (1995) showed strong seasonal variability in bacterial abundance, growth efficiency, and respiration rates in the Amazon mainstem upriver from the study area at the historical gauging station at Óbidos, which is ~900 km inland from the mouth. At Óbidos, bacterial growth rates peaked during high water, when respiration rates and overall bacterial abundance were lowest. The interpretation for enhanced growth efficiency during the low water period is an influx of labile, algae-derived organic matter from várzea, or floodplain lake, sources. Similarly, in a study of the seasonal degradation rates of lignin and related macromolecules, Ward et al. (2013) showed again that respiration rates were lowest at high water, and that the contribution of lignin degradation to the bulk rate was higher than during low water stands. This evidence suggests that at high water periods, the microbial community is mostly supported by less labile organic matter (lignin and terrestrial plant derived compounds) than during low water, when microbes can access more labile, algal organic matter produced *in situ*.

5.4.2 *Peptide and protein identifications*

Overall, the numbers of tandem mass spectra and peptide-spectrum matches (and thus, peptide identifications) was much lower than other proteomic/peptidomic samples detailed in this thesis (marine POM in Chapter 2, algal suspensions in Chapter 3), all of which were also filtered onto glass fiber filters and extracted by very similar approaches. Injections of 1 ug protein from the initial timepoint of the diatom degradation in Chapter 3, performed with the same LCMS setup and parameters as this study, yielded an average of 18222 MS spectra and 24448 MS2 spectra. In contrast, injections of 1 ug protein from Amazon River initial timepoint samples yielded between 17980-22845 and 2890-6479 MS2 spectra.

Studies of Amazon River DOM, suspended sediments, and clay mineral have shown significant sorption of DOM on mineral surfaces (Aufdenkampe et al., 2001). Results of experimental attempts to swamp active mineral sites with free lysine were inconclusive (*see earlier discussion in Methods*). The protein lysis approach used here included a very small amount of acid-soluble surfactant (Waters Rapigest SF, 0.00625%), and increased amounts of detergent may be helpful in improving the protein solubilization and liberation from mineral surfaces in addition

to increased pH (see Supporting Methods). Time constraints allowed for only the collection of three replicates of 150-300 mL filtered river water for each metapeptidome sample, because filtering took so long due to high sediment loads. By estimates of bacterial abundance at 50% river depth from a previous study at the Macapá stations in May 2011 (Satinsky et al., 2015), I then collected between 5.9×10^8 and 1.2×10^9 cells per metapeptidome for both size fractions. Assuming an average protein/cell of 0.2 pg (value for *E. coli*, which is about the same size as many Actinomycetes, the dominant bacteria found), one could maximally extract about 118-240 ug biomass protein per sample. In actuality, the sample extractions yielded between 90-615 ug protein per sample. Additionally, recovery experiments with a bovine serum albumin protein standard indicate the extraction protocol is ~80% efficient. I was extracting sufficient protein for high resolution LC-MSMS proteomics/peptidomics (0.5-2 μg protein/injection), and still got low numbers of spectra, peptide-spectrum matches, and thus peptides. These poor spectral qualities are most likely due to ionization and fragmentation issues from a complex matrix of mineral and humic compounds that could be potentially mitigated by including an additional column separation step, like ion mobility (Baker et al., 2015).

5.4.3 *Peptide subcellular location*

I showed in Chapter 3 that the cellular compartment annotations of detrital algal peptidomes become increasingly dominated by membrane and chloroplast terms as degradation proceeds in an incubation of a diatom culture with seawater microbes (Duffy et al., submitted, see Chapter 3). While the cellular compartment distributions of all metapeptidomes in this study were indicative of the living microbial community - an unchanging mixture of cytoplasm, membrane, and organelle peptides (Figure 5.3), when just the peptides specially mapped to the proteins of microbial primary producers examined, their cellular compartment distributions are more indicative of detrital peptides. The 46 peptides specific to (assumed) *in situ* primary producers across all stations at time 0 hours had cellular compartment annotations that included the chloroplast and transmembrane domains. After 24-hours however, only 24 of these peptides were identifiable by mass spectrometer and peptidomic searching and were sourced exclusively from transmembrane proteins or proteins in the nucleus (Figure 5.3e). As discussed in Chapter 2 and more thoroughly in Chapter 3, preferential preservation of both algal and bacterial membrane and cell wall proteins and peptides is thought to be a mechanism for organic matter preservation in

aquatic environments (Benner & Kaiser, 2003; Nagata et al., 1998; Tanoue et al., 1995). As I showed using a protein secondary structure model (Montgomerie et al., 2008) in Chapter 3, the α-helix motifs in transmembrane sequences are degraded more slowly in the case of detrital diatom protein, perhaps due to their tightly wrapped, difficult to denature, shape (Chapter 3 and Duffy et al., *submitted*). Transmembrane domains have relatively more acidic residues compared to bulk protein (Yeagle et al., 2007) and could also with mineral surfaces, another potential mechanism in their preferential preservation. The numbers of identified algal peptides were too low here to confidently compare secondary structure patterns in the incubations, but future experiments with more extracted protein and cleaner chromatography may enable this evaluation in the future.

5.4.4 *Post-translational modifications*

Protein post-translational modifications (PTMs) like oxidation, phosphorylation, and methylation, are critical factors at play in a wide range of biological processes such as signaling, transport, gene expression regulation (Cain et al., 2014; Shen et al., 2008). Some PTMs are also associated with protein recycling and cell senescence (Cain et al., 2014; Dhillon & Denu, 2017). For these purposes in attempting to characterize the living and detrital component of the metapeptidome, I also consider the modifications to proteinaceous compounds that could occur in detritus, including those due to protein consumption or abiotic transformations, to be PTMs (see Chapter 3).

In aquatic systems, patterns in PTM profiles have been linked to degradation and early diagenesis. One example is the amino acid β-alanine, which accumulates in the hydrolyzable phase of sedimentary organic matter via modification of aspartic acid (Cowie and Hedges 1994). Modification of the nitrogen-containing side chains of glutamine, asparagine and arginine can lead to the accumulation of peptides containing deaminated side chains within anoxic marine sediment pore waters (Abdulla et al., 2018) and in sinking POM in anoxic ocean water (Duffy et al., *submitted*). In aquatic environments, protein residues like histidine and methionine can be oxidized during photodegradative processes (Lundeen et al., 2014), though given the minimal light penetration in the lower Amazon such reactions may be insignificant in this study.

During the 24-hour incubations at all four of the stations, the PTM profiles of proteinaceous compounds show little to no indication of increasing degradation (Figure 5.4). The static nature of PTM profiles indicates that proteins are effectively consumed in 24-hours and that the bulk of the peptide signal is from proteins in living microbes.

5.4.5 *Relative amino acid compositions*

Relative compositions of amino acids in the 16 metapeptidomes showed non-systematic shifts during the 24-hour incubations likely due to random selection variability of what peptides were detected by the mass spectrometer and identified (Figure 5.5). Changes in amino acid composition that are indicative of degradation and observed in marine sediments are absent here, notably enrichment in glycine and serine (Cowie & Hedges, 1992; Dauwe & Middelburg, 1998; Keil et al., 2000). Hedges et al. (1994) measured total hydrolyzable amino acids in coarse (>63 μm), fine (0.5-63 μm), and dissolved (>1000 Da) organic matter upstream of Óbidos, 1000 km upriver from this study's sites. They found the fine particulate fraction (which corresponds to the size fractions in this study combined) to yield the greatest concentration of amino acids, which is consistent with living microbial biomass. In contrast, the DOM pool was nitrogen poor and low in total amino acids. Indicative of a degraded proteinaceous material, the dissolved amino acid was enriched in the non-protein products β-alanine and γ-aminobutyric acid as well as acidic residues, an indication of degradation (Hedges et al., 1994). Complicating this picture, though, is laboratory experimental evidence for the selective partitioning of nitrogen-rich DOM components onto mineral surfaces, leaving behind diagenetic indicator compounds like β-alanine and γ-aminobutyric acid in the dissolved phase (Aufdenkampe et al., 2001).

Unfortunately, I was unable in this study to collect sufficient water to measure dissolved peptides and proteins, and so couldn't capture what might be a more detrital peptide pool in the river. The metapeptidomic-based amino acid compositions, like those measured by Hedges et al. (1994), intimidate the amino acid signal that is dominated by microbial life.

5.4.6 *Fungi-like lignin oxidizing bacteria dominate the lower Amazon microbial community*

Few studies have explored patterns of microbial diversity in tropical rivers and their estuaries, but they are dynamic transition zones with many potential controlling factors including discharge, salinity, nutrient, and turbidity gradients and seasonality. While that evaluation of free-living and particle associated microbial communities has been performed at the upriver stations of Macapá North and South, (see Doherty et al., 2017; Satinsky et al., 2015), this is the first sequence-based study of the oceanward stations (Baylique and Chaves) that has been conducted to my knowledge. While the protein sequences predicted from previously published genomic and transcriptomic data are contained in the search database (the Amazon Basin non redundant Gene Catalog includes the

Satinsky et al. (2015) results, see Santos-Júnior et al. (2020)), there are no reference sequences to be had for the oceanward sites. This is one reason to employ the *de novo*-directed approach that uses, in addition to traditional database searching, mapping to the UniProt KnowledgeBase database in order to find unanticipated proteins and source organisms.

Peptide mapping to UniProt is biased towards finding sequences from highly studied systems like bioreactors, hydrothermal vents, wastewater, and biomedically-relevant microorganisms such as some Actinobacteria, from which many potential pharmaceuticals derive (Manivasagan et al., 2013). For instance, I identify sequences from the bacterial phylum Aminicenantes, whose known members are found in deep subsurface thermal aquifers that degrade cellulose (Kadnikov et al., 2019). However, it is likely that a microbe of a potentially related lineage in these systems is expressing similar proteins for cellulose breakdown. While imperfect, the taxonomic assignments of protein sequences in the oceanward stations are a first look at the microbial community this far into the mouth of the Amazon.

As described above, the geochemical signatures (PTMs, subcellular origins, amino acid compositions) of the proteinaceous compounds at all four stations indicate the dominance of peptide sequences from living microbes, not from detrital plant or microbial sources. This presents the opportunity to evaluate peptide sequences for taxonomic and functional insight into the microbial communities degrading (and producing) organic matter in lower Amazon. Taxonomic peptide assignments reveal that four stations are dominated by Actinobacteria (Figure 5.6), one of the prominent soil microbial lineages worldwide (Fierer, 2017) commonly found in streams and rivers, (Crump & Hobbie, 2005). Actinobacteria are some of the only recognized groups of saprotrophic microbes with the ability to oxidatively depolymerize lignin, using extracellular enzymes to produce soluble polyphenols that are subsequently available to a wider range of the community (Eisenlord & Zak, 2010). Actinobacteria form hyphae, usually considered a trait unique to fungi. In metagenomes from Macapá in May 2011, Satinsky et al. (2015) found that a single actinobacterium (UniProt reference genome SCGC AAA027-L06) recruited the greatest number of overall reads in the dataset, contributing an average of 50 billion genes and 500 million transcripts per liter of river water. In the main stem of the upper course of the Amazon, 1200 kilometers upriver of Mapacá, metagenomics also has revealed an Actinobacteria-dominated ecosystem with high genetic capacity for heterotrophic carbon processing (Ghai et al., 2011). Thus, by extending the range of sequence-based microbial community evaluation, I show that this

lineage dominates the world's largest river consistently for over a thousand kilometers from the upper river to its mouth.

This study's peptide-based inventory of the microbial communities at Macapá North and South channels stations generally agrees with previous research. Doherty et al. (2017) used 16S rRNA gene amplicon sequencing tools to assess the biogeography of the lower Amazon from Óbidos to the Amazon plume during late falling water (September 2010), early rising water (December 2010), and high water (May 2011) periods. Another study from that same May 2011 expedition generated metagenomes and metatranscriptomes using next generation paired-end Illumina sequencing from free-living (0.2 to 2.0 μm) and particle-associated (2.0 to 297 μm) (Satinsky et al., 2015). Both amplicon and metagenome sequencing approaches found the biogeography of river mainstem communities, both free-living and particle associated, to be spatially homogeneous (Doherty et al., 2017) as has been observed in other large rivers. Prokaryotic metagenomes and metatranscriptomes were dominated by members of the bacterial phyla Actinobacteria, Planctomycetes, Betaproteobacteria, Verrucomicrobia, Nitrospirae, and Acidobacteria, all of which were observed in the 16 metapeptidomes generated in this study.

Besides a bacterial lineage with fungal characteristics, I also identified actual fungal sequences at all stations except Baylique, primarily from the large and diverse group Ascomycota, a phylum common in soils as plant litter degraders (Figure 5.7). Many members of this group are animal and plant pathogens. Whether these proteins are from active and living fungal cells or from inactive fungi washed into the river from soils is an open question, as is why most of the fungal peptides (15/21) in the dataset were identified at Macapá North. Distinct lignin degrading ascomycetes have been identified in temperature rivers (Liu et al., 2015; Tanaka et al., 2005). They have also been identified in marine sinking particles (Baltar et al., 2021; Bochdansky et al., 2017; Gutiérrez et al., 2011) including in the eastern tropical North Pacific (see Chapter 2 and Peng and Valentine, 2021).

5.4.7 *Peptidomic evidence for labile, detrital algal proteins*

Though dominated by heterotrophic bacteria, 46 peptides specific to known microbial photosynthesizers were identified at time 0 across the 4 stations (Figure 5.8). Cyanobacteria, particularly *Synechococcus* at Macapá, were also observed as indicator bacterial groups during low water (Doherty et al., 2017). In that study, however, 16S sequencing does not capture

eukaryotic primary producers like the diatoms or green algae (e.g., Chlorophyta) whose peptide signatures I observe here through a metapeptidomic approach (Figure 5.8).

Photosynthetic processes in the highly turbid surface waters of the lower main stem of the Amazon have long been thought to be negligible. However, while measurements of chlorophyll-*a* are low in the Amazon River mainstem (Table 5.8), there is recent evidence from dissolved oxygen stable isotopes suggesting that primary production may actually be occurring at up to 50% the rate of respiration (Gagne-Maynard et al., 2017). There has been no thorough evaluation of primary production at or near these oceanward stations, and the only estimates of algal production are the measurements of chlorophyll-*a* fluorescence (Table 5.8) that may be complicated by high turbidity (*see Methods*). The fluorometric chlorophyll-*a* estimates here indicate there is more algal primary production in surface waters at the Macapá stations than at the oceanward Baylique and Chaves stations at the time of sampling. Underway measurements to the oceanward stations in April 2017 indicated that turbidity increases from Macapá to the oceanward Baylique and Chaves stations (Kuhn et al., 2019), and this trend was borne out in the average station measurements (Table 5.8). Increased sunlight attenuation from higher turbidity could then be a factor in this decrease in chlorophyll-*a* from Macapá further into the mouth.

All the incubations were performed with water from 50% river depth, between 12 and 14 m (Table 5.8) and far removed from the surface. I presumed that algal protein detected in metapeptidomes at the initial timepoints is detrital, and that it would be rapidly degraded in 24-hours as a labile source of carbon and nitrogen, especially compared to terrestrial plant substrates. This idea is supported by peptide cellular location annotations that become membrane-enriched during incubation (Figure 5.3). Further evidence that the primary producers' protein is detrital is also hinted at in the size fractions at which the peptides from certain taxa are identified. At the Macapá South station for instance, I identified 8 peptides specific to dinoflagellates before incubation, a group of freshwater and marine phytoplankton in the range of 1-2 μm in diameter. Five of the dinoflagellate peptides were found in the small (0.3-0.7 μm) fraction, indicating they may be detrital in nature (glass fiber membranes were used in this study, which are created from layers of silica strands and thus are nominal in pore sizes rather than strict). All but one of the dinoflagellate peptides disappeared from the detection window after 24-hours incubation, contributing to evidence that the dinoflagellate cells had already been lysed before sampling.

Primary production has been shown to be a key factor in enhancing the water column respiration rates in the Amazon mainstem and its tributaries, presumably because fresh algal material is a more labile organic carbon substrate than plant debris dominated by lignincellulose (Ellis et al., 2012), a theory supported by the particulate protein degradation rates measured here (Figure 5.2). There is also evidence that input of labile organic matter can stimulate, or 'prime', increased microbial respiration of more recalcitrant OM. Ward et al. (2016) showed that ^{13}C-labeled plant leachates degraded significantly faster in algae-rich tributaries of the Amazon compared with the sediment-laden main stem at Óbidos. Similarly, that study found increasing leachate degradation rates as the river moved from Óbidos downriver to Macapá - between which there are multiple inputs that could contain labile algal material, especially from floodplain lakes during high water stands (Abril et al., 2014; J. Richey, Mertes, et al., 1989).

5.4.8 *Peptide functional annotations*

Through peptide functional terms one can begin to capture the substrate binding and transport strategies of the microbial communities at the different stations and reflect the assemblage of organic matter available to heterotrophs in the tidal reach. The degradation of terrestrial, plant-derived organic matter is a fundamental process in the Amazon Basin as a whole. Lignincellulose comprises up to 60% of the DOM in the river during rising water (Ertel et al., 1986; Ward et al., 2015) and lignin degradation accounts for 30-50% of the bulk microbial respiration in the main stem of the river (Ward et al., 2013). The breakdown of these complex plant biopolymers (woody biomass on average being 30% lignin, 45% cellulose, and 20% hemicellulose) happens via two steps that are modulated by microbes: first, lignins are oxidized by laccase oxidoreductases; and second, cellulose is broken down via specialized glycosyl hydrolases. Unsurprisingly, hydrolase and oxidoreductase were two highly abundant annotation terms in all stations. All station metapeptidomes were dominated by Actinobacteria (Figure 5.6), and most of these terms represent actinobacterial laccases and peroxidases (Hamdi et al., 2019).

Many GO term assignments implicated metal ion binding as an important process in both 0.3-0.7 and >0.7 μm size fractions (Figure 5.9). Zinc binding terms stood out consistently in all metapeptidomes (1.9-5.5% of all metapeptidome GO terms) and were not significantly changed over the 24-hour incubation period. Lignin degrading laccases of *Pleurotus ostreatus* (oyster mushroom fungus) have been shown to be more efficient with added zinc and cadmium (Baldrian

et al., 2005). Metal cofactors in lignin degrading enzymes have yet to be explored in freshwater systems, nor has metal limitation of lignin degradation. The likelihood of bioavailable zinc or other metals limitation in the Amazon River seems slim due to the large fluxes of iron, aluminum, manganese, and other elements from soils (Aucour et al., 2003) that associate particularly with particles as metal hydroxide coatings (Gibbs, 1977).

The functional annotations also give clues as to organic substrates besides the lignin and cellulose accessed by the microbial communities. Peptidase activity terms remained at around 6% of all terms in all stations and over 24-hours, and likely represent mostly internal cellular protein turnover. Carbohydrate binding annotations remained at similar levels across the stations and over time, as did transmembrane transport terms. At all stations except Baylique, I identified 1-4 chitin binding peptides. The most chitin binding annotations in a single metapeptidome (4 peptides) was at Macapá North, where I also identified fungal (chitin-producing) protein sequences (Figure 5.7). The incubations performed here and their resulting metapeptidomes are snapshots in time of the expression of proteins by the microbial community. There is little indication in the geochemical signatures of the peptides or in their linked taxonomies of any significant generation or signal of detrital, recalcitrant proteinaceous compounds as has been the case in other degradative systems in the aquatic systems (in marine sinking and suspended particles as in Chapter 2) or in a longer laboratory incubation (Chapter 3).

5.5 CONCLUSIONS

Organic matter in the lower Amazon represents a true proteomic unknown. This thesis has so far directed peptide-level focus to organic matter with a wide range of degradation status: from the cells of *Prochlorococcus* in exponential growth (Chapter 2) and flash-frozen algal cells before degradation (Chapter 3) to deep suspended particulate organic matter in the ocean (Chapter 2) and algal proteins after 12 days of heterotrophic activity by seawater microbes (Chapter 3). The Amazon River as a system is somewhat similar to the eastern tropical North Pacific (ETNP) water column in that organic matter is traveling along a continuum of degradative reworking and biogeochemical transitions. Both systems are complicated for the organic geochemist by lateral and *in situ* organic matter inputs along these continua.

Whereas the ETNP station focused on in Chapter 2 is oligotrophic, and most organic matter is produced in the water column by microbial life, the lower Amazon has as organic matter input a vast range of sources: rainforest plants and soils, high levels of vertebrate biomass, floodplain and clearwater tributary algal material, and large amount of fine and coarse sediments. The metapeptidomes, if comprehensively and accurately characterized, can inform us about how long this significant component of organic matter from this wide range of sources is preserved over space.

Protein and peptide sequences can provide a geochemical and systems-biological window into organic matter processing. The peptidomic fingerprints at all four stations point to a proteinaceous pool that is labile and undergoing rapid turnover, consistent with older studies of bulk organic matter conducted upriver of this study's sampling sites (Hedges et al., 1994; Mayorga et al., 2005; Richey et al., 1990). Intracellular protein localization, PTMs, and relative amino acid compositions are similar both between stations and between the initial and 24-hour timepoints of the incubations. While metatranscriptomic studies of free-living and particle associated microbial communities have been performed at the upriver stations (Macapá North and South), this is the first sequence-based study of the oceanward stations (Baylique and Chaves) that has been conducted. The metapeptidomic taxonomies are similar to those found by Doherty et al. (2017) and Satinsky (2015) for the overlapping stations at Macapá, and oceanward taxonomies are similar.

In short, 24-hour incubations without organic matter, nutrient, or metal amendments, the proteins expressed by the microbial communities at all four stations revealed no significant taxonomic or functional shifts. However, I demonstrate that metapeptidomic tools have the power to identify the proteinaceous component of the river's microbial communities. Pairing this metapeptidomic profiling approach with longer incubation periods and a variety of treatments may help us to learn more about the relationships and exchanges of members of these communities over time and space. Future work could also contrast these results from this study to other points in the hydrograph. Suspended sediments in the main channels of the Amazon have been generally determined as distinct fractions with fine (clay and silt) and coarse (sand), which follow seasonal and long-term patterns (Armijos et al., 2020). How the composition of sediment may affect associated microbial communities - and the outcome of these relationships on respiration rates - could factor into more accurate and complex models of carbon dynamics in the Amazon Basin and river-to-ocean continuum.

5.6 ACKNOWLEDGEMENTS

This work is the result of a collaboration with Jeff Richey (UW) and Nick Ward (Pacific Northwest Nation Lab - PNNL) who enabled me to travel to Brazil for an expedition in the spring of 2019. Nick Ward also mentored me as a host scientist during my fellowship at PNNL as part of the US Department of Energy Office of Science Graduate Student Fellowship Program. Funding for this work came from that program, NSF (Award # 1754317), and FAPESP. Many thanks to Captain Sika and the *B/M Mirage* crew. Much laboratory effort was made by co-authors Khadijah Homolka, Jaqui Neibauer, and Jamee Adams.

5.7 REFERENCES

Abdulla, H. A., Burdige, D. J., & Komada, T. (2018). Accumulation of deaminated peptides in anoxic sediments of Santa Barbara Basin. *Geochimica et Cosmochimica Acta*, *223*, 245–258. https://doi.org/10.1016/j.gca.2017.11.021

Abril, G., Martinez, J.-M., Artigas, L. F., Moreira-Turcq, P., Benedetti, M. F., Vidal, L., Meziane, T., Kim, J.-H., Bernardes, M. C., Savoye, N., Deborde, J., Souza, E. L., Albéric, P., Landim de Souza, M. F., & Roland, F. (2014). Amazon River carbon dioxide outgassing fuelled by wetlands. *Nature*, *505*(7483), 395–398. https://doi.org/10.1038/nature12797

Armijos, E., Crave, A., Espinoza, J. C., Filizola, N., Espinoza-Villar, R., Ayes, Fonseca, P., Fraizy, P., Gutierrez, O., Vauchel, P., Camenen, B., Martiınez, J. M., Santos, A. D., Santini, W., Cochonneau, G., & Guyot, J. L. (2020). Rainfall control on Amazon sediment flux: Synthesis from 20 years of monitoring. *Environmental Research Communications*, *2*(5), 051008. https://doi.org/10.1088/2515-7620/ab9003

Aucour, A.-M., Tao, F.-X., Moreira-Turcq, P., Seyler, P., Sheppard, S., & Benedetti, M. F. (2003). The Amazon River: Behaviour of metals (Fe, Al, Mn) and dissolved organic matter in the initial mixing at the Rio Negro/Solimões confluence. *Chemical Geology*, *197*(1), 271–285. https://doi.org/10.1016/S0009-2541(02)00398-4

Aufdenkampe, A. K., Hedges, J. I., Richey, J. E., Krusche, A. V., & Llerena, C. A. (2001). Sorptive fractionation of dissolved organic nitrogen and amino acids onto fine sediments within the Amazon Basin. *Limnology and Oceanography*, *46*(8), 1921–1935. https://doi.org/10.4319/lo.2001.46.8.1921

Baker, E. S., Burnum-Johnson, K. E., Ibrahim, Y. M., Orton, D. J., Monroe, M. E., Kelly, R. T., Moore, R. J., Zhang, X., Théberge, R., Costello, C. E., & Smith, R. D. (2015). Enhancing bottom-up and top-down proteomic measurements with ion mobility separations. *Proteomics*, *15*(16), 2766–2776. https://doi.org/10.1002/pmic.201500048

Baldrian, P., Valášková, V., Merhautová, V., & Gabriel, J. (2005). Degradation of lignocellulose by Pleurotus ostreatus in the presence of copper, manganese, lead and zinc. *Research in Microbiology*, *156*(5), 670–676. https://doi.org/10.1016/j.resmic.2005.03.007

Baltar, F., Zhao, Z., & Herndl, G. J. (2021). Potential and expression of carbohydrate utilization by marine fungi in the global ocean. *Microbiome*, *9*(1), 106. https://doi.org/10.1186/s40168-021-01063-4

Benner, R., & Kaiser, K. (2003). Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter. *Limnology and Oceanography*, *48*(1), 118–128. https://doi.org/10.4319/lo.2003.48.1.0118

Benner, R., Opsahl, S., Chin-Leo, G., Richey, J. E., & Forsberg, B. R. (1995). Bacterial carbon metabolism in the Amazon River system. *Limnology and Oceanography*, *40*(7), 1262–1270. https://doi.org/10.4319/lo.1995.40.7.1262

Bochdansky, A. B., Clouse, M. A., & Herndl, G. J. (2017). Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *The ISME Journal*, *11*(2), 362–373. https://doi.org/10.1038/ismej.2016.113

Bridoux, M., Neibauer, J., Ingalls, A., Nunn, B., & Keil, R. (2015). Suspended marine particulate proteins in coastal and oligotrophic waters. *Journal of Marine Systems*, *143*. https://doi.org/10.1016/j.jmarsys.2014.10.014

Bu, H., Yuan, P., Liu, H., Liu, D., Qin, Z., Zhong, X., Song, H., & Li, Y. (2019). Formation of macromolecules with peptide bonds via the thermal evolution of amino acids in the presence of montmorillonite: Insight into prebiotic geochemistry on the early Earth. *Chemical Geology*, *510*, 72–83. https://doi.org/10.1016/j.chemgeo.2019.02.023

Bugg, T. D. H., Williamson, J. J., & Rashid, G. M. M. (2020). Bacterial enzymes for lignin depolymerisation: New biocatalysts for generation of renewable chemicals from biomass. *Current Opinion in Chemical Biology*, *55*, 26–33. https://doi.org/10.1016/j.cbpa.2019.11.007

Butman, D., & Raymond, P. A. (2011). Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geoscience*, *4*(12), 839–842. https://doi.org/10.1038/ngeo1294

Cain, J. A., Solis, N., & Cordwell, S. J. (2014). Beyond gene expression: The impact of protein post-translational modifications in bacteria. *Journal of Proteomics*, *97*, 265–286. https://doi.org/10.1016/j.jprot.2013.08.012

Carpenter, J. H. (1965). The Chesapeake Bay Institute Technique for the Winkler Dissolved Oxygen Method. *Limnology and Oceanography*, *10*(1), 141–143. https://doi.org/10.4319/lo.1965.10.1.0141

Chong, L. S., Berelson, W. M., Hammond, D. E., Fleisher, M. Q., Anderson, R. F., Rollins, N. E., & Lund, S. (2016). Biogenic sedimentation and geochemical properties of deep-sea sediments of the Demerara Slope/Abyssal Plain: Influence of the Amazon River Plume. *Marine Geology*, *379*, 124–139. https://doi.org/10.1016/j.margeo.2016.05.015

Cowie, G. L., & Hedges, J. I. (1992). Sources and reactivities of amino acids in a coastal marine environment. *Limnology and Oceanography*, *37*(4), 703–724. https://doi.org/10.4319/lo.1992.37.4.0703

Crump, B. C., & Hobbie, J. E. (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnology and Oceanography*, *50*(6), 1718–1729. https://doi.org/10.4319/lo.2005.50.6.1718

Cunha, A. C., & Sternberg, L. da S. L. (2018). Using stable isotopes 18O and 2H of lake water and biogeochemical analysis to identify factors affecting water quality in four estuarine Amazonian shallow lakes. *Hydrological Processes*, *32*(9), 1188–1201. https://doi.org/10.1002/hyp.11462

Dauwe, B., & 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. https://doi.org/10.4319/lo.1998.43.5.0782

Dhillon, R. S., & Denu, J. M. (2017). Using comparative biology to understand how aging affects mitochondrial metabolism. *Molecular and Cellular Endocrinology*, *455*, 54–61. https://doi.org/10.1016/j.mce.2016.12.020

Doherty, M., Yager, P. L., Moran, M. A., Coles, V. J., Fortunato, C. S., Krusche, A. V., Medeiros, P. M., Payet, J. P., Richey, J. E., Satinsky, B. M., Sawakuchi, H. O., Ward, N. D., & Crump, B. C. (2017). Bacterial Biogeography across the Amazon River-Ocean Continuum. *Frontiers in Microbiology*, *8*, 882. https://doi.org/10.3389/fmicb.2017.00882

Dong, F., Guo, Y., Liu, M., Zhou, L., Zhou, Q., & Li, H. (2018). Spectroscopic evidence and molecular simulation investigation of the bonding interaction between lysine and montmorillonite: Implications for the distribution of soil organic nitrogen. *Applied Clay Science*, *159*, 3–9. https://doi.org/10.1016/j.clay.2017.11.020

Eisenlord, S. D., & Zak, D. R. (2010). Simulated Atmospheric Nitrogen Deposition Alters Actinobacterial Community Composition in Forest Soils. *Soil Science Society of America Journal*, *74*(4), 1157–1166. https://doi.org/10.2136/sssaj2009.0240

Ellis, E. E., Richey, J. E., Aufdenkampe, A. K., Krusche, A. V., Quay, P. D., Salimon, C., & Cunha, H. B. da. (2012). Factors controlling water-column respiration in rivers of the central and southwestern Amazon Basin. *Limnology and Oceanography*, *57*(2), 527–540. https://doi.org/10.4319/lo.2012.57.2.0527

Eom, J., Seo, K.-W., & Ryu, D. (2017). Estimation of Amazon River discharge based on EOF analysis of GRACE gravity data. *Remote Sensing of Environment*, *191*, 55–66. https://doi.org/10.1016/j.rse.2017.01.011

Ertel, J. R., Hedges, J. I., Devol, A. H., Richey, J. E., & Ribeiro, M. de N. G. (1986). Dissolved humic substances of the Amazon River system1. *Limnology and Oceanography*, *31*(4), 739–754. https://doi.org/10.4319/lo.1986.31.4.0739

Fierer, N. (2017). Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), 579–590. https://doi.org/10.1038/nrmicro.2017.87

Filizola Jr, N., & Guyot, J.-L. (2004). The use of Doppler technology for suspended sediment discharge determinations on the River Amazon at Óbidos. *Hydrological Sciences Journal/Journal Des Sciences Hydrologiques*, *49*, 143–153. https://doi.org/10.1623/hysj.49.1.143.53990

Froelich, J. M., & Reid, G. E. (2008). The origin and control of ex vivo oxidative peptide modifications prior to mass spectrometry analysis. *PROTEOMICS*, *8*(7), 1334–1345. https://doi.org/10.1002/pmic.200700792

Gagne-Maynard, W. C., Ward, N. D., Keil, R. G., Sawakuchi, H. O., Da Cunha, A. C., Neu, V., Brito, D. C., Da Silva Less, D. F., Diniz, J. E. M., De Matos Valerio, A., Kampel, M., Krusche, A. V., & Richey, J. E. (2017). Evaluation of Primary Production in the Lower Amazon River Based on a Dissolved Oxygen Stable Isotopic Mass Balance. *Frontiers in Marine Science*, *4*, 26. https://doi.org/10.3389/fmars.2017.00026

Ghai, R., Rodriguez-Valera, F., McMahon, K. D., Toyama, D., Rinke, R., Cristina Souza de Oliveira, T., Wagner Garcia, J., Pellon de Miranda, F., & Henrique-Silva, F. (2011). Metagenomics of the water column in the pristine upper course of the Amazon river. *PloS One*, *6*(8), e23785. https://doi.org/10.1371/journal.pone.0023785

Ghesquière, B., & Gevaert, K. (2014). Proteomics methods to study methionine oxidation. *Mass Spectrometry Reviews*, *33*(2), 147–156. https://doi.org/10.1002/mas.21386

Gibbs, R. J. (1977). Transport phases of transition metals in the Amazon and Yukon Rivers. *GSA Bulletin*, *88*(6), 829–843. https://doi.org/10.1130/0016-7606(1977)88<829:TPOTMI>2.0.CO;2

Gough, M. A., Fauzi, R., Mantoura, C., & Preston, M. (1993). Terrestrial plant biopolymers in marine sediments. *Geochimica et Cosmochimica Acta*, *57*(5), 945–964. https://doi.org/10.1016/0016-7037(93)90032-R

Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615. https://doi.org/10.1021/acs.jproteome.8b00716

Gutiérrez, M. H., Pantoja, S., Tejos, E., & Quiñones, R. A. (2011). The role of fungi in processing marine organic matter in the upwelling ecosystem off Chile. *Marine Biology*, *158*(1), 205–219. https://doi.org/10.1007/s00227-010-1552-z

Hamdi, C., Arous, F., & Jaouani, A. (2019). Actinobacteria: A Promising Source of Enzymes Involved in Lignocellulosic Biomass Conversion. *Advances in Biotechnology*, *13*(5), 3.

Hedges, J. I., Clark, W. A., & Come, G. L. (1988). Fluxes and reactivities of organic matter in a coastal marine bay. *Limnology and Oceanography*, *33*(5), 1137–1152. https://doi.org/10.4319/lo.1988.33.5.1137

Hedges, J. I., Cowie, G. L., Richey, J. E., Quay, P. D., Benner, R., Strom, M., & Forsberg, B. R. (1994). Origins and processing of organic matter in the Amazon River as indicated by carbohydrates and amino acids. *Limnology and Oceanography*, *39*(4), 743–761. https://doi.org/10.4319/lo.1994.39.4.0743

Ionescu, D., Bizic-Ionescu, M., Khalili, A., Malekmohammadi, R., Morad, M. R., de Beer, D., & Grossart, H.-P. (2015). A new tool for long-term studies of POM-bacteria interactions: Overcoming the century-old Bottle Effect. *Scientific Reports*, *5*, 14706. https://doi.org/10.1038/srep14706

Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Karnachuk, O. V., & Ravin, N. V. (2019). Genome of the candidate phylum Aminicenantes bacterium from a deep subsurface thermal aquifer revealed its fermentative saccharolytic lifestyle. *Extremophiles: Life Under Extreme Conditions*, *23*(2), 189–200. https://doi.org/10.1007/s00792-018-01073-5

Keil, R. G., Montluçon, D. B., Prahl, F. G., & Hedges, J. I. (1994). Sorptive preservation of labile organic matter in marine sediments. *Nature*, *370*(6490), 549–552. https://doi.org/10.1038/370549a0

Keil, R. G., Tsamakis, E., & Hedges, J. (2000). *Early diagenesis of particulate amino acids in marine systems*. 69–82.

Kim, G., Weiss, S. J., & Levine, R. L. (2014). Methionine Oxidation and Reduction in Proteins. *Biochimica et Biophysica Acta*, *1840*(2). https://doi.org/10.1016/j.bbagen.2013.04.038

Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H. O., Kampel, M., Richey, J., Stadler, P., Crawford, J., Striegl, R., Vermote, E., Pahlevan, N., & Butman, D. (2019). Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sensing of Environment*, *224*, 104–118. https://doi.org/10.1016/j.rse.2019.01.023

Lentz, S. J. (1995). Seasonal variations in the horizontal structure of the Amazon Plume inferred from historical hydrographic data. *Journal of Geophysical Research: Oceans*, *100*(C2), 2391–2400. https://doi.org/10.1029/94JC01847

Leprevost, F. V., Valente, R. H., Lima, D. B., Perales, J., Melani, R., Yates, J. R., Barbosa, V. C., Junqueira, M., & Carvalho, P. C. (2014). PepExplorer: A similarity-driven tool for analyzing *de novo* sequencing results. *Molecular & Cellular Proteomics: MCP*, *13*(9), 2480–2489. https://doi.org/10.1074/mcp.M113.037002

Liu, J., Wang, J., Gao, G., Bartlam, M. G., & Wang, Y. (2015). Distribution and diversity of fungi in freshwater sediments on a river catchment scale. *Frontiers in Microbiology*, *6*, 329. https://doi.org/10.3389/fmicb.2015.00329

Lundeen, R. A., Janssen, E. M.-L., Chu, C., & Mcneill, K. (2014). Environmental Photochemistry of Amino Acids, Peptides and Proteins. *Chimia*, *68*(11), 812–817. https://doi.org/10.2533/chimia.2014.812

Mai, Y., Lai, Z., Li, X., Peng, S., & Wang, C. (2018). Structural and functional shifts of bacterioplanktonic communities associated with spatiotemporal gradients in river outlets of the subtropical Pearl River Estuary, South China. *Marine Pollution Bulletin*, *136*, 309–321. https://doi.org/10.1016/j.marpolbul.2018.09.013

Manivasagan, P., Venkatesan, J., Sivakumar, K., & Kim, S.-K. (2013). Marine actinobacterial metabolites: Current status and future perspectives. *Microbiological Research*, *168*(6), 311–332. https://doi.org/10.1016/j.micres.2013.02.002

Marshall, I. P. G., Starnawski, P., Cupit, C., Cáceres, E. F., Ettema, T. J. G., Schramm, A., & Kjeldsen, K. U. (2017). The novel bacterial phylum Calditrichaeota is diverse, widespread and abundant in

marine sediments and has the capacity to degrade detrital proteins. *Environmental Microbiology Reports*, *9*(4), 397–403. https://doi.org/10.1111/1758-2229.12544

Mayer, L. M. (1994). Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. *Chemical Geology*, *114*(3), 347–363. https://doi.org/10.1016/0009-2541(94)90063-9

Mayer, L. M., Macko, S. A., & Cammen, L. (1988). Provenance, concentrations and nature of sedimentary organic nitrogen in the Gulf of Maine. *Marine Chemistry*, *25*(3), 291–304. https://doi.org/10.1016/0304-4203(88)90056-4

Mayorga, E., Aufdenkampe, A. K., Masiello, C. A., Krusche, A. V., Hedges, J. I., Quay, P. D., Richey, J. E., & Brown, T. A. (2005). Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers. *Nature*, *436*(7050), 538–541. https://doi.org/10.1038/nature03880

McGivern, B. B., Tfaily, M. M., Borton, M. A., Kosina, S. M., Daly, R. A., Nicora, C. D., Purvine, S. O., Wong, A. R., Lipton, M. S., Hoyt, D. W., Northen, T. R., Hagerman, A. E., & Wrighton, K. C. (2021). Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nature Communications*, *12*(1), 2466. https://doi.org/10.1038/s41467-021-22765-1

Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., St-Denis, N., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., … Nesvizhskii, A. I. (2013). The CRAPome: A Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nature Methods*, *10*(8), 730–736. https://doi.org/10.1038/nmeth.2557

Mesuere, B., Willems, T., Jeugt, F. V. der, Devreese, B., Vandamme, P., & Dawyndt, P. (2016). Unipept web services for metaproteomics analysis. *Bioinformatics*, *32*(11), 1746–1748. https://doi.org/10.1093/bioinformatics/btw039

Mikan, M. P., Harvey, H. R., Timmins-Schiffman, E., Riffle, M., May, D. H., Salter, I., Noble, W. S., & Nunn, B. L. (2020). Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. *The ISME Journal*, *14*(1), 39–52. https://doi.org/10.1038/s41396-019-0503-z

Montgomerie, S., Cruz, J. A., Shrivastava, S., Arndt, D., Berjanskii, M., & Wishart, D. S. (2008). PROTEUS2: A web server for comprehensive protein structure prediction and structure-based annotation. *Nucleic Acids Research*, *36*(suppl_2), W202–W209. https://doi.org/10.1093/nar/gkn255

Nagata, T., Fukuda, R., Koike, I., Kogure, K., & Kirchman, D. (1998). Degradation by bacteria of membrane and soluble protein in seawater. *Aquatic Microbial Ecology*, *14*, 29–37. https://doi.org/10.3354/ame014029

Ning, K., Fermin, D., & Nesvizhskii, A. I. (2012). Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *Journal of Proteome Research*, *11*(4), 2261–2271. https://doi.org/10.1021/pr201052x

Noble, W. S. (2015). Mass spectrometrists should only search for peptides they care about. *Nature Methods*, *12*(7), 605–608. https://doi.org/10.1038/nmeth.3450

Opsahl, S., & Benner, R. (1997). Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature*, *386*(6624), 480–482. https://doi.org/10.1038/386480a0

Remington, S., Krusche, A., & Richey, J. (2011). Effects of DOM photochemistry on bacterial metabolism and CO2 evasion during falling water in a humic and a whitewater river in the Brazilian Amazon. *Biogeochemistry*, *105*(1), 185–200. https://doi.org/10.1007/s10533-010-9565-8

Richey, J. E., Hedges, J. I., Devol, A. H., Quay, P. D., Victoria, R., Martinelli, L., & Forsberg, B. R. (1990). Biogeochemistry of carbon in the Amazon River. *Limnology and Oceanography*, *35*(2), 352–371. https://doi.org/10.4319/lo.1990.35.2.0352

Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M., & Hess, L. L. (2002). Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO 2. *Nature*, *416*(6881), 617–620. https://doi.org/10.1038/416617a

Richey, J., Mertes, L., Dunne, T., Victoria, R., Forsberg, B., Tancredi, A. C. F., & Oliveira, E. (1989). *Sources and routing of the Amazon River Flood Wave*. https://doi.org/10.1029/GB003I003P00191

Richey, J., Nobre, C., & Deser, C. (1989). Amazon River Discharge and Climate Variability: 1903 to 1985. *Science (New York, N.Y.)*, *246*, 101–103. https://doi.org/10.1126/science.246.4926.101

Riffle, M., May, D. H., Timmins-Schiffman, E., Mikan, M. P., Jaschob, D., Noble, W. S., & Nunn, B. L. (2017). MetaGOmics: A Web-Based Tool for Peptide-Centric Functional and Taxonomic Analysis of Metaproteomics Data. *Proteomes*, *6*(1). https://doi.org/10.3390/proteomes6010002

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmechtig, C., Poteau, A., D'Ortenzio, F., Ras, J., Drapeau, S., Haëntjens, N., & Barbieux, M. (2017). Recommendations for obtaining unbiased chlorophyll estimates from *in situ* chlorophyll fluorometers: A global analysis of WET Labs ECO sensors. *Limnology and Oceanography: Methods*, *15*(6), 572–585. https://doi.org/10.1002/lom3.10185

Saito, M. A., Bulygin, V. V., Moran, D. M., Taylor, C., & Scholin, C. (2011). Examination of Microbial Proteome Preservation Techniques Applicable to Autonomous Environmental Sample Collection. *Frontiers in Microbiology*, *2*. https://doi.org/10.3389/fmicb.2011.00215

Sampath, J., Kullman, A., Gebhart, R., Drobny, G., & Pfaendtner, J. (2020). Molecular recognition and specificity of biomolecules to titanium dioxide from molecular dynamics simulations. *Npj Computational Materials*, *6*(1), 1–8. https://doi.org/10.1038/s41524-020-0288-7

Santos-Júnior, C. D., Sarmento, H., de Miranda, F. P., Henrique-Silva, F., & Logares, R. (2020). Uncovering the genomic potential of the Amazon River microbiome to degrade rainforest organic matter. *Microbiome*, *8*(1), 151. https://doi.org/10.1186/s40168-020-00930-w

Satinsky, B. M., Fortunato, C. S., Doherty, M., Smith, C. B., Sharma, S., Ward, N. D., Krusche, A. V., Yager, P. L., Richey, J. E., Moran, M. A., & Crump, B. C. (2015). Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome*, *3*, 39. https://doi.org/10.1186/s40168-015-0099-0

Satinsky, B. M., Smith, C. B., Sharma, S., Ward, N. D., Krusche, A. V., Richey, J. E., Yager, P. L., Crump, B. C., & Moran, M. A. (2017). Patterns of Bacterial and Archaeal Gene Expression through the Lower Amazon River. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00253

Satinsky, B. M., Zielinski, B. L., Doherty, M., Smith, C. B., Sharma, S., Paul, J. H., Crump, B. C., & Moran, M. A. (2014). The Amazon continuum dataset: Quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome*, *2*(1), 17. https://doi.org/10.1186/2049-2618-2-17

Sawakuchi, H. O., Neu, V., Ward, N. D., Barros, M. de L. C., Valerio, A. M., Gagne-Maynard, W., Cunha, A. C., Less, D. F. S., Diniz, J. E. M., Brito, D. C., Krusche, A. V., & Richey, J. E. (2017). Carbon Dioxide Emissions along the Lower Amazon River. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00076

Schmidt, F., Koch, B. P., Elvert, M., Schmidt, G., Witt, M., & Hinrichs, K.-U. (2011). Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under Contrasting Sedimentary

Redox Conditions in the Black Sea. *Environmental Science & Technology*, *45*(12), 5223–5229. https://doi.org/10.1021/es2003414

Seidel, M., Dittmar, T., Ward, N. D., Krusche, A. V., Richey, J. E., Yager, P. L., & Medeiros, P. M. (2016). Seasonal and spatial variability of dissolved organic matter composition in the lower Amazon River. *Biogeochemistry*, *131*, 281–302.

Shen, Y., Hixson, K., Tolić, N., Camp, D., Purvine, S., Moore, R. J., & Smith, R. (2008). Mass spectrometry analysis of proteome-wide proteolytic post-translational degradation of proteins. *Analytical Chemistry*. https://doi.org/10.1021/ac800077w

Sorokin, D. Y., Muntyan, M. S., Toshchakov, S. V., Korzhenkov, A., & Kublanov, I. V. (2018). Phenotypic and Genomic Properties of a Novel Deep-Lineage Haloalkaliphilic Member of the Phylum Balneolaeota From Soda Lakes Possessing Na+-Translocating Proteorhodopsin. *Frontiers in Microbiology*, *9*, 2672. https://doi.org/10.3389/fmicb.2018.02672

Tanaka, K., Hatakeyama, S., & Harada, Y. (2005). Three new freshwater ascomycetes from rivers in Akkeshi, Hokkaido, northern Japan. *Mycoscience*, *46*(5), 287–293. https://doi.org/10.1007/s10267-005-0248-6

Tanoue, E., Nishiyama, S., Kamo, M., & Tsugita, A. (1995). Bacterial membranes: Possible source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta*, *59*(12), 2643–2648. https://doi.org/10.1016/0016-7037(95)00134-4

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W. S., & Nunn, B. L. (2017). Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *The ISME Journal*, *11*(2), 309–314. https://doi.org/10.1038/ismej.2016.132

UniProt Consortium, T. (2018). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *46*(5), 2699–2699. https://doi.org/10.1093/nar/gky092

Valerio, A. D. M., Kampel, M., Vantrepotte, V., Ward, N. D., Sawakuchi, H. O., Less, D. F. D. S., Neu, V., Cunha, A., & Richey, J. (2018). Using CDOM optical properties for estimating DOC concentrations and pCO(2) in the Lower Amazon River. In *Archimer, archive institutionnelle de l'Ifremer*. Archimer, archive institutionnelle de l'Ifremer. https://doi.org/10.1364/OE.26.00A657

Vlasova, N. N., & Golovkova, L. P. (2004). The adsorption of amino acids on the surface of highly dispersed silica. *Colloid Journal*, *66*(6), 657–662. https://doi.org/10.1007/s10595-005-0042-3

Wang, T., Tian, Z., Tunlid, A., & Persson, P. (2020). Nitrogen acquisition from mineral-associated proteins by an ectomycorrhizal fungus. *New Phytologist*, *228*(2), 697–711. https://doi.org/10.1111/nph.16596

Ward, N. D., Bianchi, T. S., Sawakuchi, H. O., Gagne-Maynard, W., Cunha, A. C., Brito, D. C., Neu, V., Valerio, A. de M., Silva, R. da, Krusche, A. V., Richey, J. E., & Keil, R. G. (2016). The reactivity of plant-derived organic matter and the potential importance of priming effects along the lower Amazon River. *Journal of Geophysical Research: Biogeosciences*, *121*(6), 1522–1539. https://doi.org/10.1002/2016JG003342

Ward, N. D., Keil, R. G., Medeiros, P. M., Brito, D. C., Cunha, A. C., Dittmar, T., Yager, P. L., Krusche, A. V., & Richey, J. E. (2013). Degradation of terrestrially derived macromolecules in the Amazon River. *Nature Geoscience*, *6*(7), 530–533. https://doi.org/10.1038/ngeo1817

Ward, N. D., Krusche, A. V., Sawakuchi, H. O., Brito, D. C., Cunha, A. C., Moura, J. M. S., da Silva, R., Yager, P. L., Keil, R. G., & Richey, J. E. (2015). The compositional evolution of dissolved and particulate organic matter along the lower Amazon River—Óbidos to the ocean. *Marine Chemistry*, *177*, 244–256. https://doi.org/10.1016/j.marchem.2015.06.013

Ward, N. D., Sawakuchi, H. O., Neu, V., Less, D. F. S., Valerio, A. M., Cunha, A. C., Kampel, M., Bianchi, T. S., Krusche, A. V., Richey, J. E., & Keil, R. G. (2018). Velocity-amplified microbial respiration rates in the lower Amazon River. *Limnology and Oceanography Letters*, *3*(3), 265–274. https://doi.org/10.1002/lol2.10062

Whitelegge, J. P. (2013). Integral Membrane Proteins and Bilayer Proteomics. *Analytical Chemistry*, *85*(5), 2558–2568. https://doi.org/10.1021/ac303064a

Yeagle, P. L., Bennett, M., Lemaître, V., & Watts, A. (2007). Transmembrane helices of membrane proteins may flex to satisfy hydrophobic mismatch. *Biochimica Et Biophysica Acta*, *1768*(3), 530–537. https://doi.org/10.1016/j.bbamem.2006.11.018

Yu, W. H., Li, N., Tong, D. S., Zhou, C. H., Lin, C. X. (Cynthia), & Xu, C. Y. (2013). Adsorption of proteins and nucleic acids on clay minerals and their interactions: A review. *Applied Clay Science*, *80–81*, 443–452. https://doi.org/10.1016/j.clay.2013.06.003

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., & Ma, B. (2012). PEAKS DB: *De novo* sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics: MCP*, *11*(4), M111.010587. https://doi.org/10.1074/mcp.M111.010587

5.8 SUPPORTING METHODS

5.8.1 *River parameter measurements*

Average river depth, discharge, and velocity profiles were measured across each channel at all stations using a Sontek (San Diego, CA) RiverSurveyor M9 Portable nine-beam 3.0 MHz/1.0; MHz/0.5; MHz acoustic Doppler current profiler (ADCP).

5.8.2 *Mineral active site swamping with lysine*

Matrix pH is the single most important factor controlling the adsorption of proteinaceous compounds by clay surfaces. Electrostatic interactions affect the surface charge of clay minerals and the degree of ionization of protein/peptide molecules. The isoelectric point (pI) of the protein or peptide in question matters when thinking about its binding strength to a particular mineral surface, as does the pI of the mineral and the pH of the surrounding matrix (in our case, the river water of the lower Amazon). In general, increasing the solution pH to the basic range weakens protein-clay mineral electrostatic interactions. We increased the pH of our extraction buffer to 7.8 using 50 mM ammonium bicarbonate in Optima water.

We selected lysine as the amino acid to mask mineral surfaces because it is basic and has been shown to bind tightly to a variety of mineral surfaces including clays (Dong et al., 2018), titanium

oxide (Sampath et al., 2020) and silica (Vlasova & Golovkova, 2004). The pI of lysine, our free amino acid swamping agent, is 9.74 (Figure S5.1). Therefore, since our extraction buffer is at pH 7.8, lysine will be mostly in a net positively charged form, with both amino groups protonated. As such, there may still be electrostatic interactions between lysine and clay minerals in the samples. Other factors are also at play in controlling the binding strength of mineral surfaces and proteinaceous compounds, including cation exchange, hydrophobic and hydrophilic interactions, ligand exchange, hydrogen bonding and van der Waals forces (Yu et al., 2013).

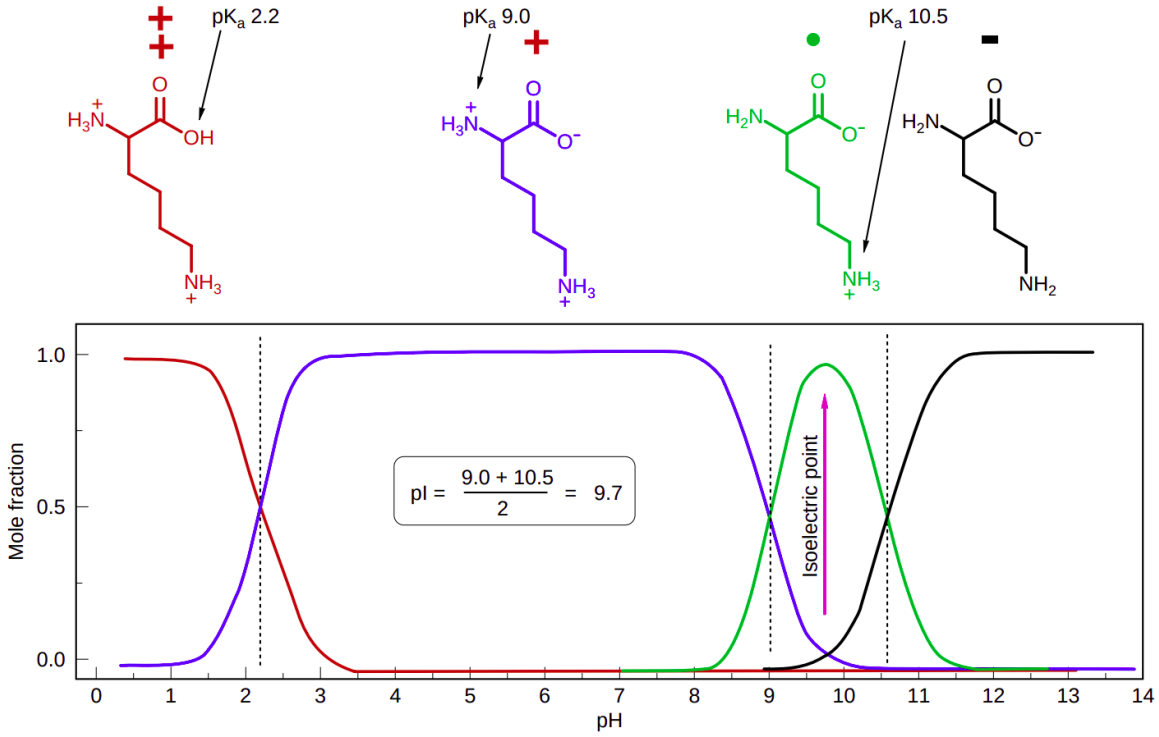

Figure 5.10. The isoelectric point (pI) of lysine. Lysine reaches a net charge of 0 at pH 9.74. Adapted from University of Wisconsin-Madison Department of Chemistry.

5.9 SUPPORTING FIGURES

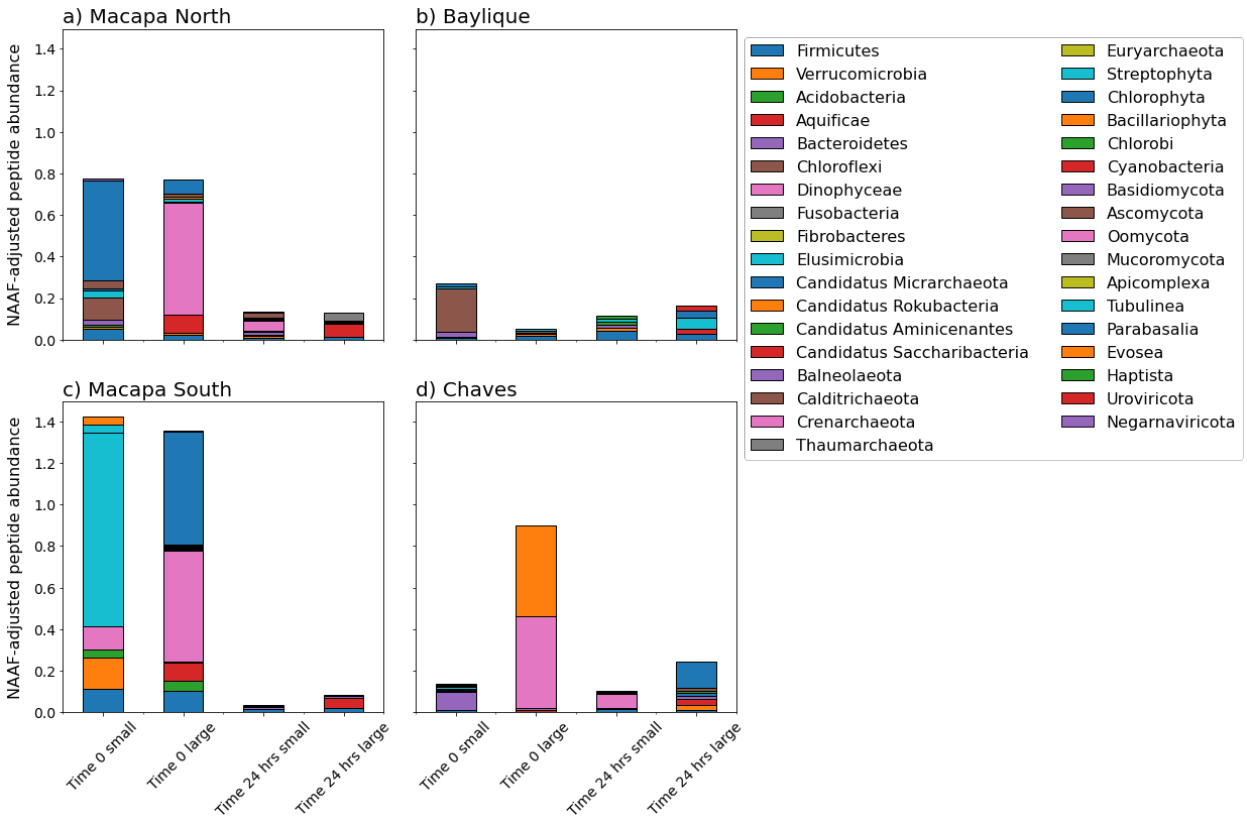

Figure 5.11. Relative abundance of peptides specific to the phylum level in 24 incubations, initial and final metapeptidomes of free-living/fine particle (0.3-0.7 µm) associated and >0.7 µm particle associated. A., Macapá North. B., Baylique. C., Macapá South. D., Chaves.

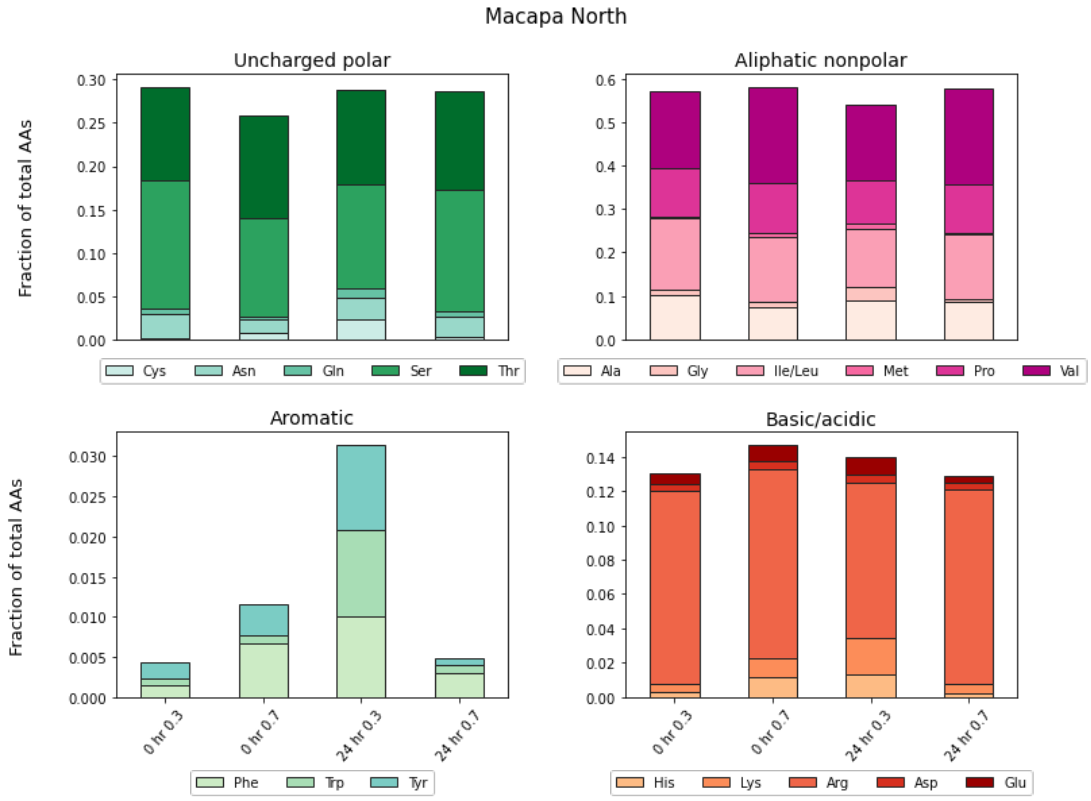

Figure 5.12. Relative amino acid abundances of metapeptidomes from 24-hour incubations at Macapá North, normalized to peptide peak area.

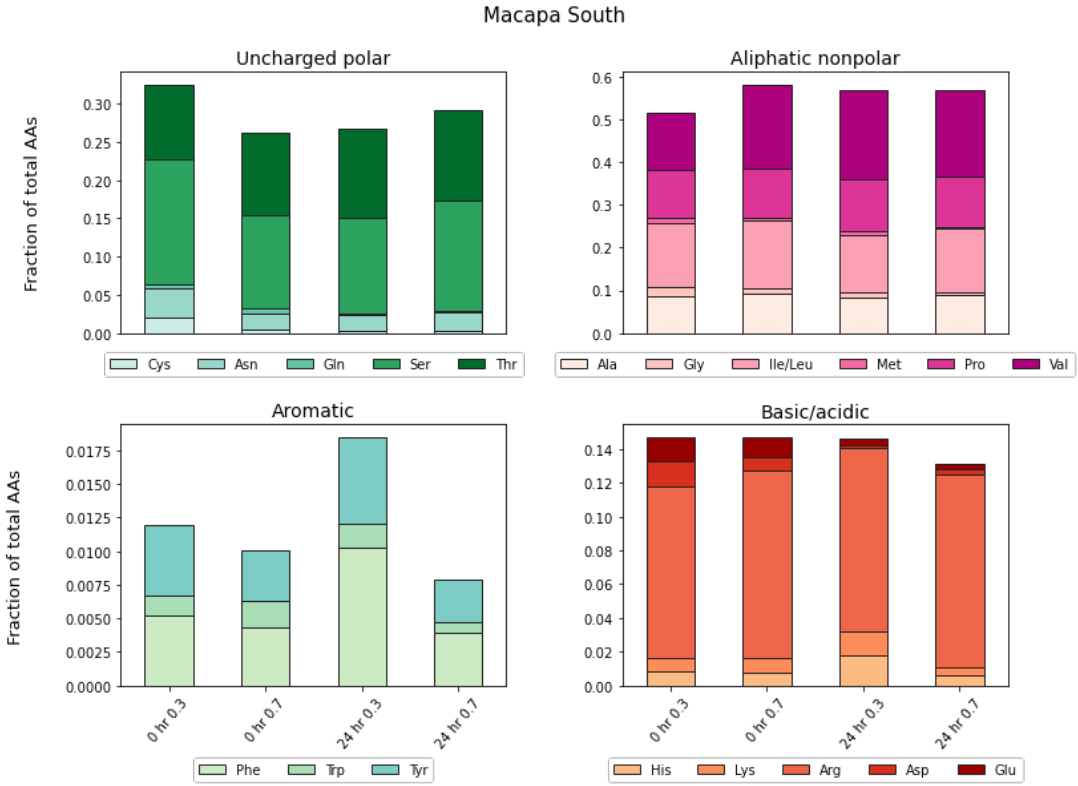

Figure 5.13. Relative amino acid abundances of metapeptidomes from 24-hour incubations at Macapá South, normalized to peptide peak area.

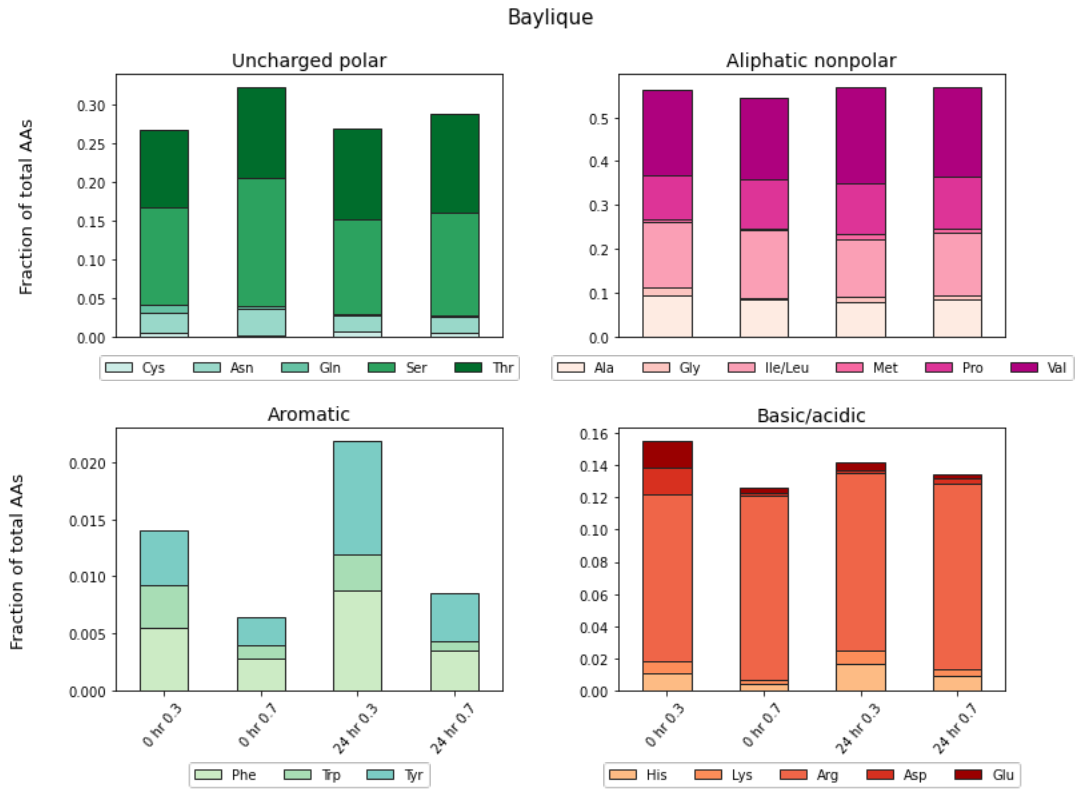

Figure 5.14. Relative amino acid abundances of metapeptidomes from 24-hour incubations at Baylique, normalized to peptide peak area.

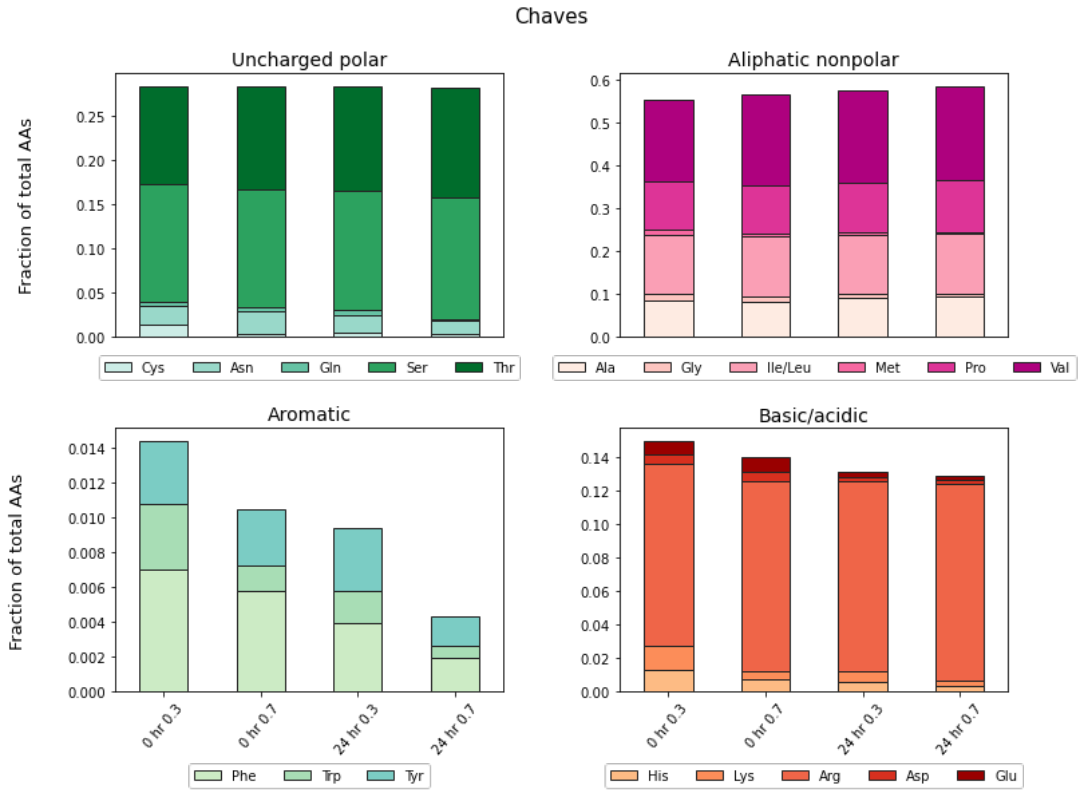

Figure 5.15. Relative amino acid abundances of metapeptidomes from 24-hour incubations at Chaves, normalized to peptide peak area.

Chapter 6. SUMMARY OF MAJOR CONCLUSIONS AND FUTURE RESEARCH

This dissertation has explored new ideas and applications for an amino acid sequence-based biogeochemical approach to the study of organic matter degradation dynamics. These endeavors are united by a goal of better understanding how, when, and what fraction of the aquatic protein pool degrades and how these processes influence carbon and nitrogen cycling in diverse and climatically important regions of the world.

In Chapter 2, I developed, tested, and demonstrated a new application of *de novo*-directed proteomic techniques for studying proteins in marine organic matter. The benchmarking experiment I carried out with *Prochlorococcus* mass spectral data revealed that *de novo* peptides found by the selected algorithm were sufficiently taxonomically specific to reference-proteins in the UniProt KnowledgeBase to be useful in identifying them amongst the thousands of other microbial, plant, and animal peptides that could be complex environmental samples. I then demonstrate that using the *de novo*-directed approach, one can indeed find specific peptide biomarkers for Cyanobacteria in the deep (1000 m) sinking particles in eastern tropical North Pacific (ETNP) oxygen deficient zone (ODZ) and later in sediment-laden lower Amazon River (Chapter 5). Not only do those sequences point to their source organism, but with peptide-specific annotations like gene ontology (GO) terms, I was able to source those Cyanobacterial peptides' subcellular origins. As I predicted given the literature on bacterial membrane preservation (Close et al., 2013; Kaiser & Benner, 2008; Tanoue et al., 1995), many of the Cyanobacterial subcellular origins were membrane-associated (Chapter 2)–a pattern I was also able to pull out of peptides as they evolved along the degradation continuum of a simulated diatom bloom (Chapter 3). As sinking particles descended through and out of the ODZ, I also found that they were more modified by processes like asparagine and glutamine deamidation (Chapter 2), potentially important mechanisms happening to proteins as a whole in anoxic waters, where proteinaceous compounds

are quickly degraded and may to some extent modulate N_2 production (Pantoja et al., 2004; Van Mooy et al., 2002) and the ratio of denitrification to anammox (Babbin et al., 2014).

Though in this thesis I don't quite arrive at linking protein concentrations, sources, and quality in sinking particles to *in situ* measurement of N_2 production in a marine ODZ, I do present the rate determination and peptidomic and proteomic tools with which research can proceed in Chapter 4, a study of fluxes and metabolic outputs of organic matter degradation in the ETNP using combination sediment trap-in situ incubation systems. In that work, I describe the variability of sinking particle fluxes at a coastal station and offshore station in the ETNP with deep chlorophyll maxima dominated by Cyanobacteria, and a coastal station still in the ODZ without a deep chlorophyll maximum. Both stations with deep chlorophyll maxima had profiles with 'spiked' fluxes, where fluxes of both bulk organic carbon and protein increased under the deep chlorophyll maximum within the ODZ. At the station without a deep chlorophyll maximum, fluxes of protein were an order of magnitude lower than those that did have this feature, indicating that deep productivity exports labile organic matter from the upper ODZ. *In situ* experiments to quantify N_2 production rates due to sinking particles performed at the same time POM collection showed that on average, rates from sinking particles were similar to those in the water column (6.38 nM N day^{-1} and 8.19 nM N day^{-1}, respectively) but increased when zooplankton or fish carcasses were collected in the chamber (Chapter 4). I don't have N_2 production rates from the station without a deep chlorophyll maximum, and I don't yet have protein measurements from the stations with high levels of zooplankton and elevated N_2 production rates. However, I hypothesize that due to their nitrogen-rich export, water columns in the ODZ with deep productivity host more N_2 production in particles than those without.

In addition to the *de novo*-directed approach to finding and identifying peptides in complex and under-characterized environmental samples, I present some other novel approaches and ideas to the field of 'detrital' metaproteomics in this thesis. In Chapter 3, I applied the *de novo*-directed technique to a study of algal protein degradation by seawater microbes. I also designed an experiment borrowed from the biomedical field of endogenous peptides that attempted to separate out 'naturally digested' small peptides created by bacterial heterotrophy and/or abiotic degradation from those created *in vitro* during proteomics trypsin digestion. Using a differential trypsin treatment, I was able to observe how heterotrophic bacteria first degraded cytoplasmic protein and then left a detrital peptide pool enriched in methylated arginines (Chapter 3), a modification shown

to slow down or prevent bacterial protein hydrolysis (Keil & Kirchman, 1992). To my knowledge, such a methodological distinction of degraded vs. intact proteinaceous compounds hasn't been done in the environmental aquatic literature. With improvements, like molecular weight cutoff spin filters and isobaric peptide termini tags (Waldbauer et al., 2017), it could prove useful in future work that aims to understand the mechanisms and timing of protein degradation.

Lastly in this thesis, I performed the first metaproteomic or metapeptidomic evaluation of organic matter in the Amazon River system (Chapter 5). I used the same metapeptidomic metrics developed in Chapter 2 and 3 to characterize both the living and detrital components of the proteinaceous pool. I found that peptides in the lower Amazon were overwhelmingly representative of a living, Actinobacteria-dominated microbial community that is processing and degrading terrestrially-sourced organic matter like lignin. While there was very little detrital signal in the metapeptidomic data (by relative amino acid compositions, modifications, and cellular compartment annotations) I was able to specifically identify a small number of peptides (~2% of the total found by combined *de novo* and database searching) that were confidently produced by microbial primary producers presumably active in the surface waters of the river. I used cellular compartment annotations to show that this small group of peptides was more detrital than the bulk microbial signal, as many disappeared quickly from the window of metapeptidomic detection after 24-hour incubations and the remaining peptides were all membrane-associated. Future work along the lines of research presented here could explore and develop peptide biomarkers for indications of preservation processes or relative status (Figure 6.1).

Muth et al. (2018) suggested recently that rapid developments in mass spectrometer resolution, machine learning-based algorithms, and computational speeds are converging to bring in a 'Golden Age' in *de novo* sequencing (Muth et al., 2018). I would expand this characterization to metaproteomics more generally, as more tools are developed that find more than the few non-canonical post-translational modifications (Guo et al., 2018) and use stable isotope labels to quantitatively track peptides in a system quantitatively (Leavitt et al., 2019; Waldbauer et al., 2017). Additionally, a greater number of genomic and transcriptomic sequences and annotations are being deposited into public repositories, enabling protein mapping and functional analysis (Gurdeep Singh et al., 2019; Riffle et al., 2017). Environmental metaproteomics may eventually even make the shift to top-down approaches, where proteins don't need to be laboratory-digested for mass spectral analysis (Kleiner, 2019), or to highly sensitive and high throughput non-mass

spectrometry methods of intact protein characterization that combine the fluorescence imaging from next-generation DNA sequencing with Edman degradation chemistry (Swaminathan et al., 2018; Timp & Timp, 2020). All these potential futures could bring aquatic geochemistry as a field closer to understanding how the dynamics of protein degradation and preservation in varied environmental contexts influence global-scale cycles of carbon and nitrogen.

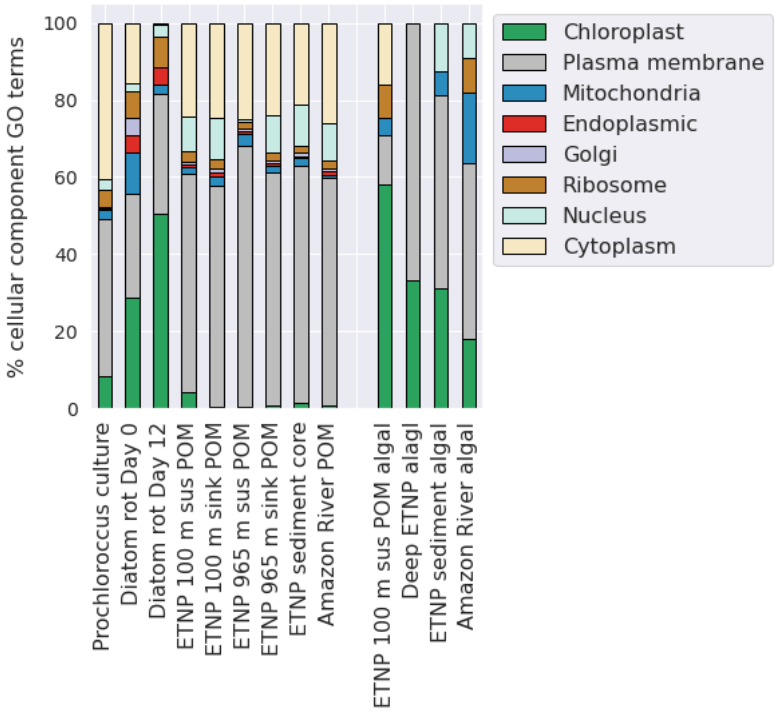

Figure 6.1. Peptide cellular compartment GO terms from the range of laboratory culture, experimental, and environmental peptidomes and metapeptidomes presented in this thesis. Peptides from a culture of *Prochlorococcus marinus* MED were evaluated in Chapter 2, as were suspended and sinking particulate organic matter (POM) samples from the ETNP. A metaproteome from ETNP sediment was evaluated but not included in this thesis. Diatom (*Thalassiosira weissflogii*) peptides over a 12-day degradation were evaluated in Chapter 3. Peptides from combined free-living/small suspended samples in the lower Amazon were evaluated in Chapter 5. Peptide specifically matched to algae (including Cyanobacteria) are plotted separately from the greater metapeptidomes from 100 m suspended POM in the ETNP, combined suspended and sinking POM from at or deeper than 265 m in the ETNP, in the ETNP sediment core, and in the lower Amazon River POM.

6.1 REFERENCES

Babbin, A. R., Keil, R. G., Devol, A. H., & Ward, B. B. (2014). Organic Matter Stoichiometry, Flux, and Oxygen Control Nitrogen Loss in the Ocean. *Science*, *344*(6182), 406–408. https://doi.org/10.1126/science.1248364

Close, H. G., Shah, S. R., Ingalls, A. E., Diefendorf, A. F., Brodie, E. L., Hansman, R. L., Freeman, K. H., Aluwihare, L. I., & Pearson, A. (2013). Export of submicron particulate organic matter to mesopelagic depth in an oligotrophic gyre. *Proceedings of the National Academy of Sciences*, *110*(31), 12565–12570. https://doi.org/10.1073/pnas.1217514110

Guo, X., Li, Z., Yao, Q., Mueller, R. S., Eng, J. K., Tabb, D. L., Hervey, W. J., & Pan, C. (2018). Sipros Ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics*, *34*(5), 795–802. https://doi.org/10.1093/bioinformatics/btx601

Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615. https://doi.org/10.1021/acs.jproteome.8b00716

Kaiser, K., & Benner, R. (2008). Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. *Limnology and Oceanography*, *53*(1), 99–112. https://doi.org/10.4319/lo.2008.53.1.0099

Keil, R. G., & Kirchman, D. L. (1992). Bacterial Hydrolysis of Protein and Methylated Protein and Its Implications for Studies of Protein Degradation in Aquatic Systems. *Applied and Environmental Microbiology*, *58*(4), 1374–1375.

Kleiner, M. (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *MSystems*, *4*(3). https://doi.org/10.1128/mSystems.00115-19

Leavitt, W. D., Venceslau, S. S., Waldbauer, J., Smith, D. A., Pereira, I. A. C., & Bradley, A. S. (2019). Proteomic and Isotopic Response of Desulfovibrio vulgaris to DsrC Perturbation. *Frontiers in Microbiology*, *10*, 658. https://doi.org/10.3389/fmicb.2019.00658

Muth, T., Hartkopf, F., Vaudel, M., & Renard, B. Y. (2018). A Potential Golden Age to Come-Current Tools, Recent Use Cases, and Future Avenues for De Novo Sequencing in Proteomics. *Proteomics*, *18*(18), e1700150. https://doi.org/10.1002/pmic.201700150

Pantoja, S., Sepúlveda, J., & González, H. E. (2004). Decomposition of sinking proteinaceous material during fall in the oxygen minimum zone off northern Chile. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*, 55–70. https://doi.org/10.1016/j.dsr.2003.09.005

Riffle, M., May, D. H., Timmins-Schiffman, E., Mikan, M. P., Jaschob, D., Noble, W. S., & Nunn, B. L. (2017). MetaGOmics: A Web-Based Tool for Peptide-Centric Functional and Taxonomic Analysis of Metaproteomics Data. *Proteomes*, *6*(1). https://doi.org/10.3390/proteomes6010002

Swaminathan, J., Boulgakov, A. A., Hernandez, E. T., Bardo, A. M., Bachman, J. L., Marotta, J., Johnson, A. M., Anslyn, E. V., & Marcotte, E. M. (2018). Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nature Biotechnology*, *36*(11), 1076–1082. https://doi.org/10.1038/nbt.4278

Tanoue, E., Nishiyama, S., Kamo, M., & Tsugita, A. (1995). Bacterial membranes: Possible

source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta*, *59*(12), 2643–2648. https://doi.org/10.1016/0016-7037(95)00134-4

Timp, W., & Timp, G. (2020). Beyond mass spectrometry, the next step in proteomics. *Science Advances*, *6*(2), eaax8978. https://doi.org/10.1126/sciadv.aax8978

Van Mooy, B. A. S., Keil, R. G., & Devol, A. H. (2002). Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrification. *Geochimica et Cosmochimica Acta*, *66*(3), 457–465. https://doi.org/10.1016/S0016-7037(01)00787-6

Waldbauer, J., Zhang, L., Rizzo, A., & Muratore, D. (2017). diDO-IPTL: A Peptide-Labeling Strategy for Precision Quantitative Proteomics. *Analytical Chemistry*, *89*(21), 11498–11504. https://doi.org/10.1021/acs.analchem.7b02752

VITA

MEGAN E DUFFY
duffyme@uw.edu
(802) 279 - 8715
github.com/MeganEDuffy
ORCID: 0000-0002-3212-4927

EDUCATION

University of Washington School of Oceanography - Seattle, WA
Ph.D anticipated September-December, 2021
Advisor: Dr. Richard Keil

University of Washington School of Oceanography - Seattle, WA
M.S., Oceanography, 2018
Advisor: Dr. Richard Keil
Thesis: *De novo-assisted protein sequencing reveals degradation patterns in marine organic matter*

Reed College - Portland, OR
B.A., Chemistry, 2012
Advisors: Drs. Arthur Glasfeld and Martina Ralle
Thesis: *Effects of Copper Exposure on Intracellular Calcium Distribution in a Wilson and Menkes Disease Human Fibroblast Cell Model*

EXPERIENCE

Graduate Research Assistant, September 2015-present
School of Oceanography, University of Washington

Investigating carbon preservation in marine sediment on a protein/peptide level, with a goal of developing quantitative protocols for extraction/identification of proteins from sediments and understanding specific preservation mechanisms through targeted and metaproteomic approaches, mineral-protein interaction studies, and bulk elemental analysis.

Research Assistant II, 2012-15
Department of Molecular and Medical Genetics, Oregon Health and Sciences University

Explored the roles of transition metals in neurodegenerative disorders using synchrotron-based X-ray fluorescence, LC-MS/MS, live-cell confocal microscopy, and molecular biological techniques. Prepared and analyzed a wide variety of biological, environmental, and industrial samples with an Agilent 7700x inductively coupled mass spectrometer (ICP-MS) as part of work for the Elemental Analysis Core.

Research Assistant, April-June 2015
Department of Chemistry, University of Tennessee

Designed iron nanoparticle-tagged antibody-based probes for live-cell fluorescent imaging of organellular targets. Developed protocols for synthesis and validation for all probe components using oxygen-free organic synthetic techniques. Mentored a summer REU student to assist with nanoparticle synthesis.

Undergraduate Research Assistant, 2011-12
Department of Biochemistry, Oregon Health and Sciences University

Used two human cells lines for copper metabolism disorder research. Developed protocols for cell fractionation and metals analysis, as well as live-cell imaging using fluorescent tags. Research part of a year-long B.A. thesis project under the guidance of Dr. Ralle and Dr. Arthur Glasfeld (Reed College).

HONORS AND AWARDS

- Department of Energy Office of Science Graduate Student Research Fellowship, 2021
- NSF Graduate Research Fellowship, 2017-2020
- University of Washington eScience Institute Student Cloud Computing grant, $ 1000
- Merit Fellowship, University of Washington School of Oceanography, 2015-2016
- Reed College Student Initiative Grant, 2011

ACTIVITIES

- Community Science Fellow, AGU Thriving Earth Exchange Program, April 2021-present
- Program coordinator, University of Washington Aquatic Organic Geochemistry High School Summer Internship, Summers 2016 - 2019
- Mentor, Earth Sciences Mentor Match program, 2020-present
- Committee Chair, John Hedges Honorary Visiting Scholar in Chemical Oceanography program, 2018-2019
- Student Mentor, Partnership for Scientific Inquiry, Oregon Health and Science University, 2015
- Volunteer Lab Instructor, Lewis and Clark College Department of Chemistry, 2013-14
- Volunteer Instructor, Reed College Chemistry Outreach, 2009-12
- Coordinator, Reed College Green Chemistry Group, 2009-12

TEACHING

- Co-instructor, Ocean 295: Chemistry of Marine Organic Carbon, Winter 2020
- Teaching assistant, Ocean 295: Chemistry of Marine Organic Carbon, Winter 2018

FIELD EXPERIENCE

- Clayoquot Sound, British Columbia - *R/V Clifford A. Barnes*, Sep. 2015
- Eastern Tropical North Pacific - *R/V Sikuliaq*, Dec. 2016 - Jan. 2017
- Eastern Tropical North Pacific - *R/V Roger Revelle*, Apr. - May, 2018
- Lower Amazon River - *B/M Mirage*, Apr., 2019
- Eastern Tropical North Pacific - *R/V Kilo Moana*, Oct. - Nov. 2019

SPECIAL SKILLS AND TRAINING

Laboratory

Extensive experience with LC-MS/MS-based proteomics and metaproteomics in environmental and biomedical contexts; ICP-MS, GC-MS. Limited experience with 1D and 2D NMR, X-ray diffraction, IR spectroscopy, Raman spectroscopy, synchrotron-based X-ray fluorescence, X-ray absorbance near-edge spectroscopy, transmission electron microscopy, and mammalian cell culture.

Computational

Python, R, Git & Github, Google Earth Engine.

Language

French, fluent written and spoken; American Field Service Student Ambassador to Belgium, 2006-2007; Spanish, intermediate written and spoken; English, native.

PEER REVIEWED PUBLICATIONS

Megan E. Duffy, Cheyenne Adams, Khadijah K. Homolka, Jacquelyn A. Neibauer, Lawrence M. Mayer, Richard G. Keil. 'Tracking peptide-level changes during microbial degradation of marine diatom protein in seawater', *in press* (Frontiers in Marine Science).

Jacob Cram, Clara A. Fuchsman, **Megan E. Duffy**, Jessica L. Pretty, Rachel M. Lekanoff, Jacquelyn A. Neibauer, Shirley W. Leung, Klaus B. Huebert, Thomas S. Weber, Daniele Bianchi, Natalya Evans, Allan H. Devol, Richard G. Keil, Andrew M.P. McDonnell. 'Efficient flux transfer throughout the Eastern Tropical North Pacific Oxygen Deficient Zone is due to slow particle remineralization rather than suppressed disaggregation', *in press* (Global Biogeochemical Cycles).

Megan E. Duffy, Jacquelyn A. Neibauer, Rachel A. Lundeen, Jamee Adams, Clara A. Fuchsman, Anitra E. Ingalls, Gabrielle Rocap, Richard G. Keil. 'Protein cycling in the eastern tropical North Pacific oxygen deficient zone: a *de novo*-discovery peptidomic approach', *in press* (Limnology and Oceanography).

Clara A. Fuchsman, Hilary I. Palevsky, Brittany Widner, **Megan E. Duffy**, Michael C.G. Carlson, Jacquelyn A. Neibauer, Margaret R. Mulholland, Richard G. Keil, Allan H. Devol, Gabrielle Rocap. 'Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone' The ISME Journal, 2019.

Mak A. Saito, Erin M. Bertrand, **Megan E. Duffy**, David A. Gaylord, Noelle A. Held, William Judson Hervey, Robert L. Hettich, Pratik Jagtap, Michael G. Janech, Danie B. Kinkade, Dasha Leary, Matthew McIlvin, Eli Moore, Robert Morris, Benjamin A. Neely, Brook Nunn, Jaclyn K. Saunders, Adam Shepherd, Nicholas Symmonds, David Walsh. 'Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing' Journal of Proteome Research, 2019.

Ashima Bhattacharjee, Haojun Yang, **Megan E. Duffy**, Emily Robinson, Arianrhod Conrad-Antoville, Ya-Wen Lu, Tony Capps, Lelita Braiterman, Michael Wolfgang, Michael P Murphy, Ling Yi, Stephen G Kaler, Svetlana Lutsenko, Martina Ralle. 'The activity of Menkes disease protein ATP7A is essential for redox balance in mitochondria' Journal of Biological Chemistry, 2016.

Mathilde L. Bonnemaison, **Megan E. Duffy**, Martina Ralle, Richard E. Mains, and Betty A. Eipper. 'Copper, Zinc and Calcium: three metals needed by one anterior pituitary secretory granule enzyme' Metallomics, 2016.

Mathilde L. Bonnemaison, Nils Bäck, **Megan E. Duffy**, Martina Ralle, Richard E. Mains, and Betty A. Eipper. 'Adaptor Protein-1 Complex Affects the Endocytic Trafficking and Function of Peptidylglycine α-Amidating Monooxygenase, a Luminal Cuproenzyme' The Journal of Biological Chemistry, 2015.

Savannah Tallino, **Megan E. Duffy**, Martina Ralle, María Paz Cortés, Mauricio Latorre, and Jason L. Burkhead. 'Nutrigenomics Analysis Reveals That Copper Deficiency and Dietary Sucrose up-Regulate Inflammation, Fibrosis and Lipogenic Pathways in a Mature Rat Model of Nonalcoholic Fatty Liver Disease' The Journal of Nutritional Biochemistry, 2015.

Kellen Voss, Christopher Harris, Martina Ralle, **Megan Duffy**, Charles Murchison, and Joseph F. Quinn. 'Modulation of Tau Phosphorylation by Environmental Copper' Translational Neurodegeneration, 2014.

PRESENTATIONS

Megan E. Duffy, Cheyenne Adams, Khadijah K. Homolka, Jacquelyn A. Neibauer, Lawrence M. Mayer, Richard G. Keil. 'Tracking peptide-level changes during microbial degradation of marine diatom protein in seawater', Virtual oral presentation at ASLO 2021 Aquatic Sciences Meeting, Palma (June 2021).

Jacquelyn A. Neibauer, **Megan E. Duffy** (presenting author), Clara A. Fuchsman, Jamee Adams, Khadijah K. Homolka, Wendi Ruef, Allan H. Devol, Richard G. Keil. 'High-resolution marine flux measurements, *in situ* respiration rate determinations, and meta-omic surveys of sinking

particulate matter in the ocean's three primary oxygen deficient zones'. Oral presentation at Ocean Sciences Meeting, San Diego, CA (February 2020).

Megan E. Duffy, Cheyenne Adams, Kathleen Thornton, Jaqui Neibauer, Jamee Adams, Lawrence Mayer, Rick Keil. *De novo*-assisted protein sequencing reveals degradation patterns in marine organic matter. Talk presented at Cascadia Proteomics Symposium in Seattle, WA (July 2018).

Megan E. Duffy, Cheyenne Adams, Kathleen Thornton, Jaqui Neibauer, Jamee Adams, Lawrence Mayer, Rick Keil. *De novo*-assisted protein sequencing reveals degradation patterns in marine organic matter Poster presented at Ocean Sciences Meeting in Portland, OR (February 2018).

Megan E. Duffy, Jacquelyn A. Neibauer, Jamee Adams, Clara A. Fuchsman, Richard G. Keil. *De novo*-assisted protein sequencing shows peptide preservation in marine systems. Poster presented at the Gordon Research Conference in New London, CT (August 2017).

Megan E. Duffy, Tony R. Capps, Amelia Munson, Charlotte S. Gleber, David Vine, Stefan Vogt,
and Martina Ralle. Characterizing Copper Resistance in Primary Astrocytes. Poster presented at the International Copper Meeting, Vico Equense, Italy (October 2014).