Degradation of diatom protein in seawater: a peptide-level view

Duffy, Megan E.[1,*], Adams, Cheyenne M.[2,3], Homolka, Khadijah K.[1], Neibauer, Jacquelyn A.[1], Mayer, Lawrence M.[2], and Keil, Richard G.[1]

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

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

[3]now at Maine Department of Marine Resources, West Boothbay Harbor, ME, 04575, USA

*To whom correspondence should be addressed. Tel: 1.802.279.8715; Email: duffyme@uw.edu

Duffy, Megan E., ORCID iD: 0000-0002-3212-4927

Adams, Cheyenne M., ORCID iD:

Homolka, Khadijah K., ORCID iD:

Neibauer, Jacquelyn A., ORCID iD: 0000-0001-9920-2558

Mayer, Lawrence M., ORCID iD:

Keil, Richard G., ORCID iD: 0000-0001-7483-1606

Running title: Algal protein degradation

Keywords: marine organic matter, marine metaproteomics, environmental peptidomics, protein cycling, phytoplankton bloom, particulate organic matter, amino acids, early diagenesis, peptidomics, peptide degradation

ABSTRACT

Peptides and proteins were identified during a controlled laboratory degradation of the marine diatom *Thalassiosira weissflogii* by a natural assemblage of seawater microbes. Samples from each timepoint were prepared both using the protease trypsin (as is typical in proteomics experiments) and without trypsin. Such differential treatment allowed comparison of protein degradation occurring within the experiment (no trypsin added; peptides already present due to *in situ* degradation) and the protein still available for future degradation (peptides released from protein when trypsin is added during analysis). 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 of the peptides that were either degraded or remained. During the initial period of rapid algal decay and bacterial growth, intracellular components from the cytoplasm were selectively consumed, resulting in the accumulation of membrane-associated proteins and peptides in the detrital pool. Accompanying the enhanced survival of membrane protein material was an increase in the importance of α-helix motifs. Methylated arginine, a post-translationally modified amino acid that is produced within the diatom prior to senescence, is found in high amounts within the detrital peptide pool, suggesting a link between in-cell modification and resistance to immediate degradation. Another modification - asparagine deamidation - appears to occur during degradation and deamidated peptides also accumulate within the detritus. The bacterial community decomposing the algal material was rich in proteobacteria, and

employed a growth approach focusing on accumulation of solubilized material across their membranes and on DNA replication. At this early stage of diagenesis, no changes in bulk amino acids (THAA) were observed, yet a peptidomic approach allowed us to observe the differential changes in diatom protein preservation by discriminating between intracellular location, secondary structure, and modifications status.

INTRODUCTION

The application of 'omics' to environmental issues has shed strong new light on the microbial ecology and marine biology of the ocean, leading to a greater understanding of the role various organisms play in global processes (Saito et al., 2019). Some of the largest pools of carbon in the ocean are the detrital remains of living 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), suggesting that application of 'omic' approaches to understanding the cycling of carbon through the detrital pool would be similarly insightful (Moore et al., 2012, 2014; Nunn et al., 2010).

Much of the identifiable detrital organic matter in the ocean is proteinaceous (Lee et al., 2004; Wakeham et al., 1997), which follows the importance of protein in living organisms. Diatoms, one of the major primary producers in the ocean, are approximately 25% protein by dry weight (Olofsson et al., 2019), and autotrophic bacteria such as those in the genus *Prochlorococcus* can be as much as 70% protein (Finkel et al., 2016).

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 refractory dissolved organic matter pool (Tanoue et al., 1995; Yamada & Tanoue, 2003), in marine sediments (Estes et al., 2019) and in sediment pore water (Abdulla et al., 2018; Fejjar et al., 2021; Schmidt et al., 2011). This 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. The underlying reasons certain proteins or peptides may be preserved while others are degraded remain largely unexplored.

Mass spectral investigation of peptides and proteins in marine detritus indicate that heterotrophic bacteria utilize a suite of endoproteases to first break protein to smaller peptides and then attack the peptides using exopeptidases that cleave individual amino acids off the ends of the peptides (Nunn et al 2003). Peptides sourced from organelles and membranes appear to selectively resist rapid degradation (Moore et al., 2014; Nunn et al., 2010; Yamada & Tanoue, 2003), certain secondary structures such as α-helices and coils appear to be retained within the detritus while other motifs are degraded (Nunn et al 2010), and peptides containing amino acids that have been modified post-translationally accumulate in the detrital pool (Duffy et al, submitted, Abdulla et al., 2018; Yamada & Tanoue, 2003).

Previous studies that apply metaproteomics techniques to marine detritus have generally added trypsin or other proteolytic enzymes to break proteins into smaller, constituent peptides (Moore et al., 2012, 2014; Nunn et al., 2010). This 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 we complement existing work by also evaluating the peptides found natively in the system, which represent a combination of the peptides that are actively undergoing degradation and those that have been liberated from protein but are not yet degraded. In addition to observing accumulation of membrane-associated proteins and peptides in the detrital material, we observed changes in the secondary structure of the peptides with α-helix motifs becoming progressively enriched in the detrital material (see also Nunn et al 2010). We further identify methylation and deamidation to be important post-translational modifications that accumulate in the detrital material, adding to growing

knowledge that degradation of protein is accompanied by extensive side chain modification (Shen et al., 2008).

MATERIAL AND METHODS

The experimental workflow can be divided into distinct phases; the degradation experiment itself, amino acid and peptide extraction, peptide identification via a combination of a database and a de novo approach, and data analytics.

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 ^{0}C to render the algal culture non-viable. An aliquot was 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 Damariscotta estuary seawater filtered to 1 μm was used as an inoculum to induce bacterial decomposition of algal material, with 1 mL added to 15 L of algal suspension. At each timepoint (1, 2,5 & 12 days), 100 mL of the degradation slurry was vacuum filtered onto 25 mm diameter 0.3 um pore size glass fiber membranes (Sterlitech GF75) at 4 °C and stored frozen prior to proteomic analysis. Further experimental details can be found in Adams et al 2019.

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. Glutamine and asparagine are deaminated during hydrolysis and converted into glutamic and aspartic acids, respectively.

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, or 'tryptic peptides' 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 a detergent-free 50 mM ammonium bicarbonate lysis buffer in an extraction protocol adapted from Bridoux et al., 2015 (digestion was performed differently here). 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 extraction supernatant was removed after centrifugation at 4800 rpm and protein concentration determined using modified Lowry assay using reagents from Bio-Rad (Hercules, California). Extracted protein subjected to reduction of disulfides and carbamidomethylation. One half (300 uL) 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. The remaining 300 uL 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, and data-dependent acquisition (DDA) on the top 10 ions 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.

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, was advantageous in this study because we 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 we 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). We 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 heterotrophs. Additionally, we search against a database of common mass spectral contaminants (Mellacheruvu et al., 2013). Search parameters for both database searching and de novo sequencing included 8 max. modifications per peptide, 15 ppm peptide mass tolerance, and 0.5 Da fragment mass tolerance. For the trypsin-digested fractions, we 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. We 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). Our 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. We required a minimum of 1 unique peptide per protein identification. Matching of proteins from peptides found only by de novo sequencing is detailed below.

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). We 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, high A-score (high confidence) modifications were used to evaluate whether adding additional PTMs to the overall searches altered the rate of false discovery. That is, we 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 with increasing numbers of variable modifications (up to 8 per peptide) 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 10 variable modifications) were selected for the data set. 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).

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 takes into account 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. We 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 the reference database, which was not unexpected given the low number of sequences in the database for heterotrophic bacteria and the fact that the GOS sampling locations were in more open ocean (Yooseph et al., 2007), compared to our estuarine seawater inoculum. To overcome this limitation, we 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' 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 a reference database.

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.

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

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: coils (unformed), 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 PXXXXXX.

RESULTS & DISCUSSION

General characteristics of the degradation

As previously described (Adams et al., 2019), algal biomass decreased by a factor of 2.4, bacterial abundance increased by a factor of 5 during the incubation, and total particulate carbon decreased by 31% over the twelve days (Figure 1). The algal detritus was partly remineralized via respiration and partly converted to bacterial biomass, but mostly remained as necromass (68%; Adams et al 2019). The degradation experiment started with a small inoculum of bacteria to a detrital slurry that already contained some proteases present during diatom cell harvesting, and likely also contained bacteria (or bacterial necromass) from the non-axenic culture. The living bacteria then responded to the detritus and began to grow exponentially (Figure 1), during which time they released enzymes in order to break down

and uptake the diatom organic matter. As time progressed, the algal material that remained decreased in nutritional value (Adams et al 2019). The ratio of enzymatically hydrolyzable amino acids to the organic carbon content remained constant throughout the experiment (Adams et al 2019). Thus, although there was no selective remineralization or preservation of the *overall* protein concentration relative to other organic material during the twelve day period, many compositional changes to the protein pool were observed.

Our interpretation of protein dynamics during the degradation experiment relies on the distinction between two different pool of peptides: those produced in situ during the experiment (not treated with trypsin, which we will refer to as 'naturally digested peptides') and those produced artificially by our laboratory digestion (referred to as 'trypsin-digested peptides'). We 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 (liberated from larger proteins when we add the protease) represent proteinaceous material including that which is not being actively degraded (e.g. the material that has resisted degradation thus far). Differences between these two diatom protein-derived pools provides information about relative labilities of the substrate proteins, in addition to the bacterial "methods" 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. 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 we examine the peptide results with an eye toward trying to better understand the

degradative processes occurring over the course of experiment and their implications for detrital protein cycling in the marine systems.

Algal Peptide Identification and Characteristics

The naturally digested and trypsin-digested peptides are methodologically distinct yet analytically overlapping pools.. The mass window of our MS1 detection is 400-2,000 m/z. Typical tryptic peptides are between 6-13 amino acids in length; in our data, the average tryptic 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 Da and the average mass, 1082 Da. Larger polypeptides created by microbial degradation processes, if they're large enough, would likely have tryptic sites and thus be cleaved, making them tryptic. Thus there is overlap between these two pools both since some larger naturally digested peptides are further subjected to trypsin digestion in the trypsin-digested fraction.

The number of identified 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 1). We sorted the peptides by their gene ontology cellular compartment terms (GO Terms; Figure 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 chloroplasts, cellular membranes, and cytoplasmic proteins (Figure 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 we plot those from peptides obtained from two *Thalassiosira pseudonana* proteomes published by Dyhrman et al., 2012 (Figure 1).

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 may indicate that peptides and proteins from other cellular components were effectively degraded during the initial phase of bacterial exponential growth 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 dropped for most cellular compartments other than membranes and chloroplasts (Figure 2). However, there were still unique identifications at day 12 for components from the mitochondria and cytoplasm, suggesting that cellular location alone does not necessarily denote whether a protein or peptide will survive the initial stages of environmental degradation.

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 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). Our data also show the same synergistic degradation of pigments (Figure 1) concurrent with retention of membrane bound proteins, suggesting that

pigment degradation and membrane protein degradation may be linked, which is consistent with their cellular co-location.

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 of their nutrition from the diatom cytoplasm rather than from other cellular constituents (Reinfelder & Fisher, 1991). Our experiment didn't involve zooplankton grazing, but does indicate that membrane protein solubility or structure are factors in differential consumption by bacteria.

Preserved protein motifs

To further interrogate the proteinaceous material that survived the degradation, five classes of protein secondary structure were evaluated: coils (unformed), α-helices, β-strands, transmembrane α-helices, and transmembrane β-strands. Closely set α-helices contain strong hydrogen bonding between of 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). 'Coil' is not a true secondary structure, but rather is an aggregate term for short sequences where there is an absence of a α-helix or sheet (Smith et al., 1996). Benchmarking

of the Proteus2 machine-learning algorithm 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).

We note a progressive change in secondary structure distribution as the degradation proceeded, with membrane α-helices becoming more important (Figure 3). At the same time, β-strands became a less common motif. This 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. Unformed 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 our experiment, or that the coil category is too broad to capture any selective processes.

To further investigate the hypothesis that secondary structure is related to enhanced preservation of membrane protein components, we also compared the predicted secondary structure of just algal chloroplast/integral component of membrane proteins and algal cytoplasmic proteins that are identified at each timepoint of the degradation (Figure 3). Chloroplast and membrane proteins, the bulk of diatom protein identifications on day 12 (167 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 proteins (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 proteins on day 12 (3 total, all in the trypsin-digested fraction) had 10% more α-helix character than the initial day 0 cytoplasmic proteins, at 40.6% coil, 33.1% α-helix, 26.3% β-strand, 0% transmembrane α-helix, and 0% transmembrane β-strand (Figure 3).

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 (e.g. Duffy et al submitted, Abdulla et al., 2018) and enrichment of certain amino acids including glycine in the preserved 'protein' pool (e.g. total hydrolyzable amino acids).

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 degradation, and regulation of gene expression (Cain et al., 2014; Shen et al., 2008). PTMs are also associated with cell senescence (Cain et al., 2014; Dhillon & Denu, 2017). For our purposes, we're including as PTMs the 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 the amino acids glutamine, asparagine and arginine can lead to the accumulation of peptides containing deaminated

amino acid side chains within anoxic marine sediment pore waters (Abdulla et al., 2018). We 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 (1992) showed that methylated peptides were accessed less efficiently by bacteria than non-methylated peptides (Keil & Kirchman, 1992). All together, the marine literature suggests that PTMs are relevant to protein degradation, as the by-products of protein degradation processes and/or as potential factors in degradation resistance.

In the present study, we first performed a non-discriminatory, or 'open' search of thousands of possible PTMs in the UniMod database using PeaksPTM (see above in Methods). We then selected just 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. Our goal was to identify patterns in PTM distribution to learn more about their roles in degradation and preservation. To that end, we wanted to distinguish between PTMs present in algal protein prior to degradation and PTMs associated with the degradative process. We compared the PTMs observed in the naturally digested peptide pool against the trypsin-digested peptides for algal protein only (no bacterial proteins were included in this differential analysis of PTMs) (Figure 4).

Oxidation PTMs were generally enriched in the trypsin with 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 4). It's notable that during the experiment, chloroplast protein-derived peptides were increasingly the bulk of identifications (Figure 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. Therefore oxidations of amino acids are frequency PTMs in photosynthesis associated proteins (Aro et al., 2005; Galetskiy et al., 2011). Similarly, lysine acetylation in chloroplast proteins has been demonstrated in plants (Lehtimäki et al., 2015), though the PTM frequently occurs elsewhere in the cell Chen et al (2018) recently showed in a model diatom (Chen et al., 2018). The abundance of thylakoid membrane and chloroplast associated proteins that accumulate due to preferential preservation (Figure 2) is likely why oxidations are dominant PTMs in the trypsin-digested peptides (Figure 4).

In contrast, two PTMs were strongly associated with the presumably more detrital naturally digested peptides: deamidation of asparagine and arginine methylation (Figure 4). Our 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 deaminated peptides in both sediment pore waters (Abdulla et al., 2018) and in sinking particulate matter (Duffy et al, submitted). Thus, we

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 modification that is used as a control on many cellular functions (Ghesquière et al., 2011). Methylated peptides are known 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. More work is needed to better understand the effect that post translational modifications have during the early stages of protein degradation.

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 hydrolyzable amino acid analyses, it has been difficult to reconcile changes in bulk amino acid compositions to known protein amino acid compositions, especially during the early stages of degradation (Keil et al., 2000). We compared the amino acid composition of the peptides identified during the degradation experiment to the THAA pool, which we measured independently. To facilitate this comparison, we plotted mole fractions of amino acids in the THAA against those in the identified algal peptides, combining the naturally and

trypsin-digested (Figure 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).

Bacterial Community

Heterotrophic bacteria grew exponentially through the middle stages of the experiment (Figure 1). A peptide-based lowest common ancestor analysis was performed using UniPept (Gurdeep Singh et al., 2019) to identify bacterial taxonomic assignments to as low a phylogenetic level as possible. The taxonomic hits were then adjusted to account for peptide spectral abundance and aggregated to the class level (Figure 6a). There is some precedence for using metaproteomic data with label-free quantitation such as we did here for microbial biomass determinations (Kleiner et al., 2017), though not using de novo peptide data. The microbial community degrading the algal detritus was dominated by Gammaproteobacteria (~60% of the peptides), with notable contributions of Cytophaga (~5%), Bacteroidia (~2%) and Alphaproteobacteria (~2%) (Figure 6a). The microbial inoculum was sourced from the Damariscotta River Estuary, whose marine waters come from the Gulf of Maine. There is taxonomic overlap of the initial composition in our study to that of a pyrosequencing survey of planktonic microbes at three stations in the Gulf of Maine (Li et al., 2011).

During the 12 day experiment the peptides of the microbial community changed in one major way: there was an increase in the abundance of peptides from classifications that remained less than 0.5% of the total peptides, accompanied by a significant increase in the amount of peptides that were bacterial but could not be identified at the class level or below (Figure 6a). These are both indications that the diversity of the bacterial community was increasing during the experiment. The increase of non-specific bacterial peptides can be explained in two ways. Most likely, the increase in diversity and richness results in a broader array of common proteins and peptides being produced (because while the peptides cross the class-level taxonomic assignment, they are not likely to be truly ubiquitous), which leads to a proportional increase in their total abundance. An alternative hypothesis is that the diatom culture used to start the growth experiment was not axenic and contained an abundance of bacteria, which were killed along with the algae and introduced as detritus. Over time, their degradation could result in an increase in the abundance of degraded peptides that are no longer taxonomically specific. 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 bacteria responded to the algal detritus (Figure 6b). The most detected GO terms are associated with transmembrane transport, carbohydrate metabolism, and DNA replication and transcription (Figure 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 the hypothesis that the Gammaproteobacteria in the samples are for the most part living and not detrital.

Our 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). In our experiment we observe steady levels of transmembrane transport and protein metabolism terms throughout the experiment. Carbohydrate metabolic process GO terms maximize in the first two days of the degradation, with increasing biosynthesis GO terms by day 12 (Figure 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, our 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 (2020) 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.

CONCLUSIONS

In this study, the first we are aware of to evaluate peptides in marine systems both with and without the use of trypsin as an extraction and identification tool, we show that specific

cellular locations lend themselves to promoting protein preservation during the initial stages of degradation when the first third of the organic matter in fresh detrital material is remineralized. As had been hypothesized and demonstrated for bacterial membrane proteins (Jiao et al., 2010; Kaiser & Benner, 2008) and algal membrane proteins (Laursen et al., 1996), we conclusively illustrate that proteins associated with diatom chloroplasts and membranes resist initial degradation better than those without such association. The 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 lead to degradation resistance remains to be more fully 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. We found that the oxidation PTMs observed originated within the living cell, with only asparagine deamidation and arginine methylation being predominantly associated with the degraded peptide pool. We hypothesize that PTMs have an impact on the bioavailability of proteins during early diagenesis, but again more work is needed to evaluate the extent to which PTMs provide protection. Continued advancements in metaproteomic instrumentation and computational capabilities have great potential to better our understanding of protein degradation and preservation dynamics in the ocean.

ACRONYMS

POM Particulate organic matter

THAA Total hydrolyzable amino acids

EHAA Enzymatically hydrolyzable amino acids

LC-MS Liquid chromatography-mass spectrometry

PSM Peptide-spectrum matching

ALC Average local confidence

PTM Post-translational modification

NAAF Normalized area abundance factor

FDR False discovery rate

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 to be made available with publication.

ACKNOWLEDGEMENT

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

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

https://doi.org/10.4319/lo.2003.48.3.1037

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

Adams, 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

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, 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

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. (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., 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

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

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

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

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

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

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

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

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

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

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

TABLES AND FIGURES

Table 1. Peptide and protein identifications

Proteomic timepoint 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

Figure 1.

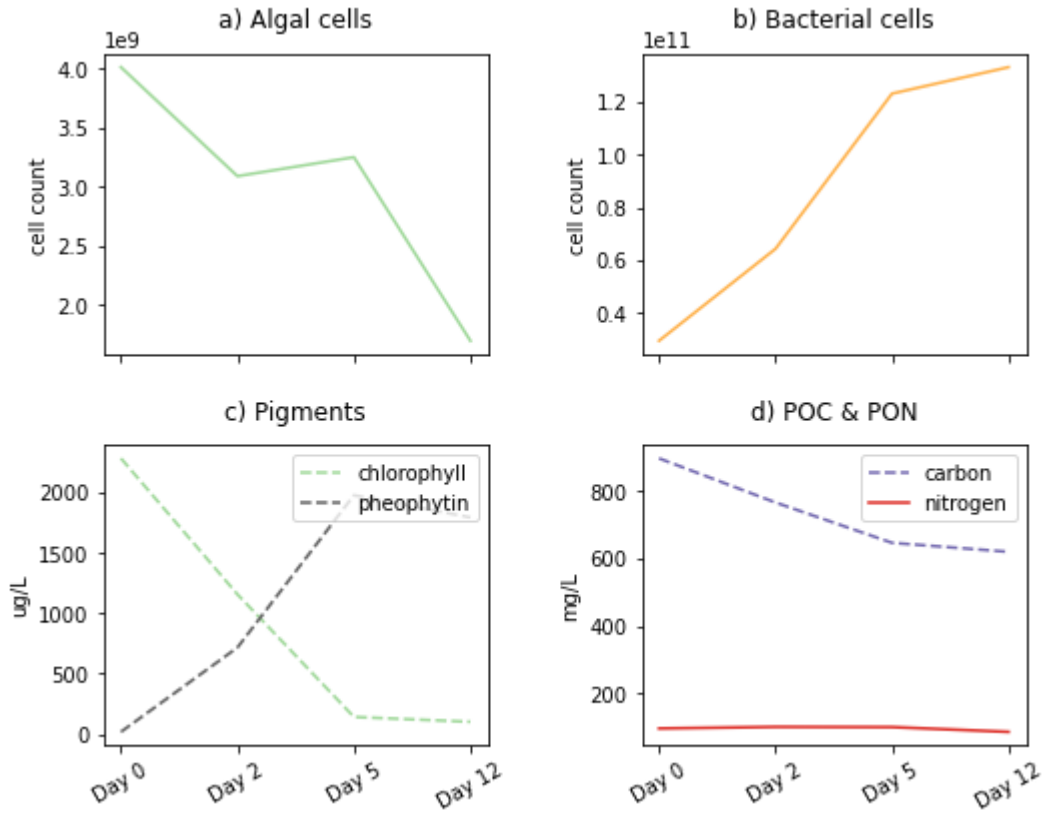

Figure 1. Progression of *Thalassiosira weissflogii* degradation experiment: **a)** algal and **b)** bacterial cell counts over the 12 day incubation period; **c)** chlorophyll and its degradation product, pheophytin; and **d)** particulate carbon and nitrogen (filtered on GF/F membranes).

Figure 2.

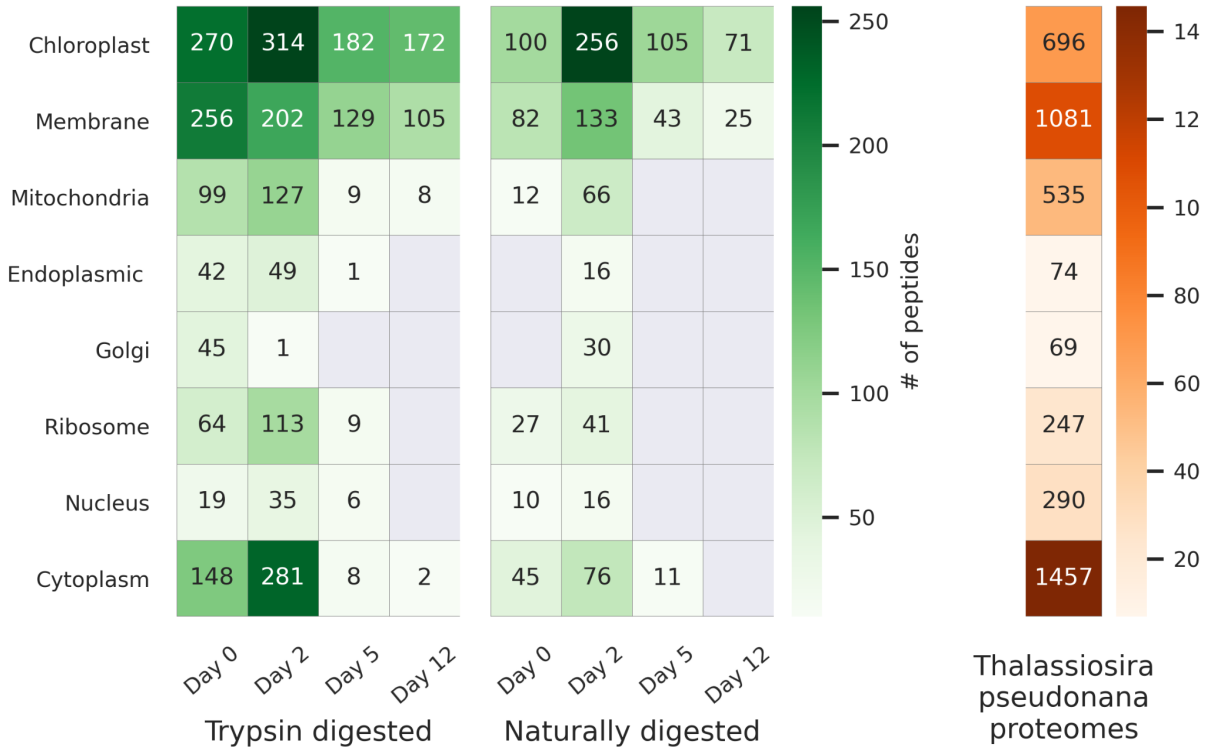

Figure 2. Heatmap showing the number of cellular compartment gene ontology (GO) terms specific to diatom peptides over a 12 day degradation experiment, both trypsin-digested (trypsin-digested) and naturally digested (digestion only by natural microbial community). Shown as context are peptides identified 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.

Figure 3.

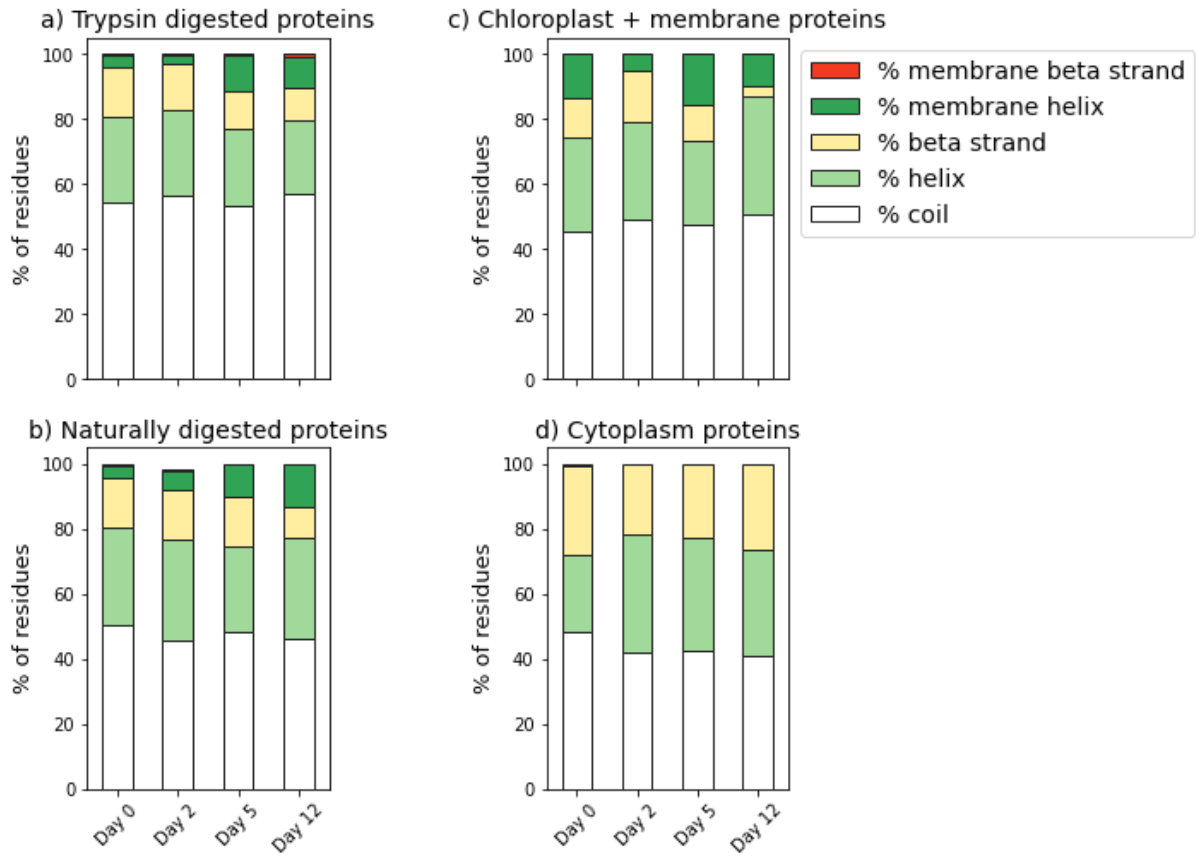

Figure 3. **a-b)** Secondary structure predictions of algal proteins identified using trypsin-digested and naturally digested peptides across a 12 day degradation experiment by a natural seawater microbial community. Secondary structure motifs (coil, α-helix, β-strand, membrane α-helix, membrane β-strand) 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/integral component of membrane proteins and algal cytoplasmic proteins that are identified at each timepoint of the degradation. For this comparison, proteins identified in the trypsin-digested and naturally digested fractions were combined and NAAF-adjusted.

Figure 4.

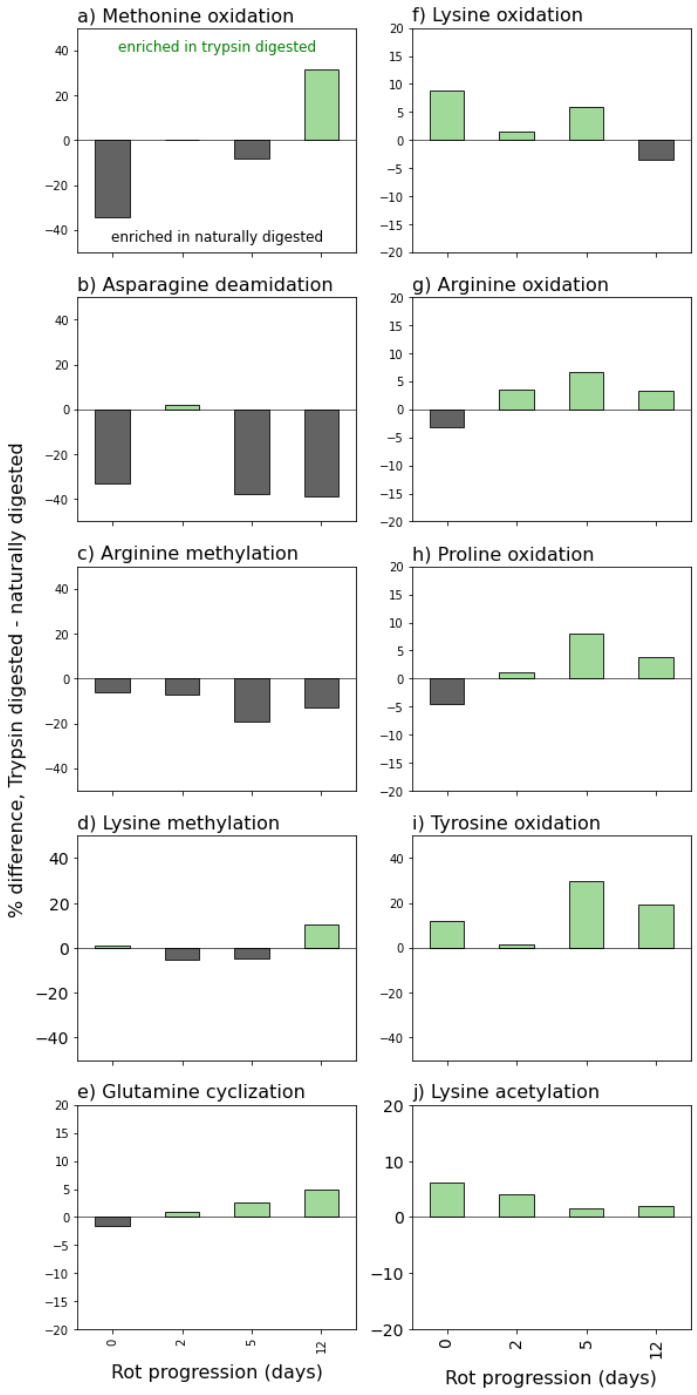

Figure 4. Mass modifications of algal peptides over the course of a 12 day incubation in a natural seawater microbial community, shown as the difference between trypsin-digested naturally digested peptides for the 10 variable modifications included in the database searches

and de novo sequencing parameters, as selected from preliminary open modification searches of the mass spectral data (see Methods): a) methionine oxidation, b) asparagine deamidation, c) arginine methylation, d) tyrosine oxidation, e) lysine methylation, f) arginine oxidation, g) proline oxidation, h) glutamine cyclization (pyro-glutamation), and i) lysine acetylation. Modification 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 modification is relatively enriched in the trypsin-digested peptides, and negative values mean the modification 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-i).

Figure 5.

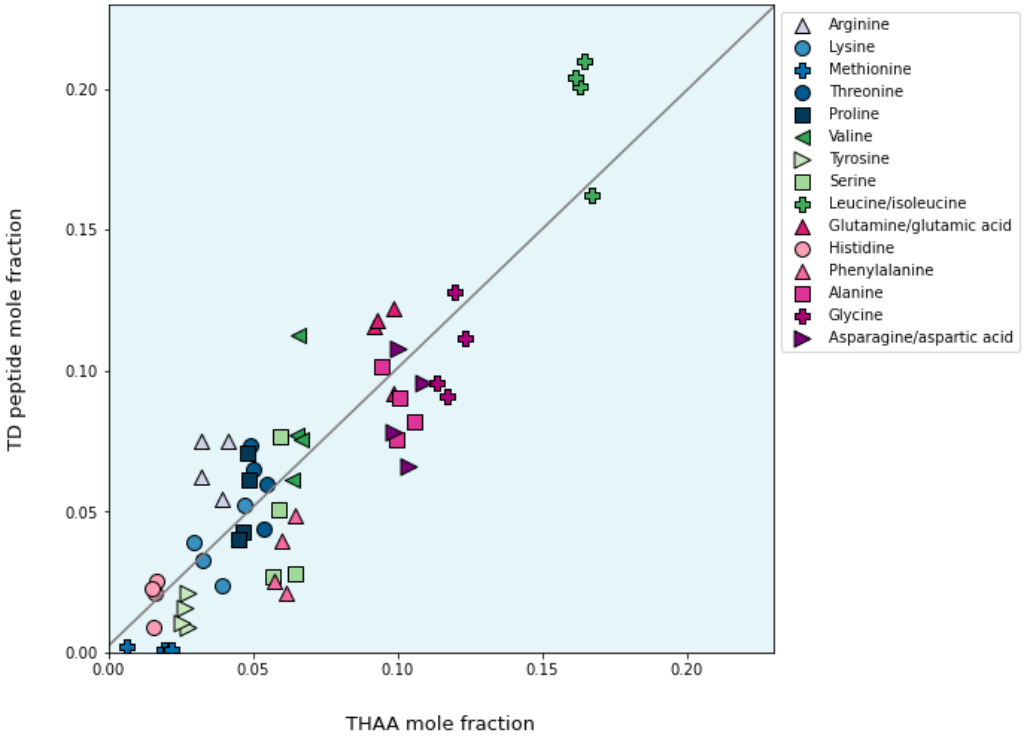

Figure 5. Mole fractions of individual amino acids from all four degradation time points as derived from total acid hydrolyzable amino acid analysis (THAA, x-axis) and tandem MS/MS-based proteomics (Peptide, y-axis). The 1:1 line represents perfect agreement between approaches. The peptide amino acid 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).

Figure 6.

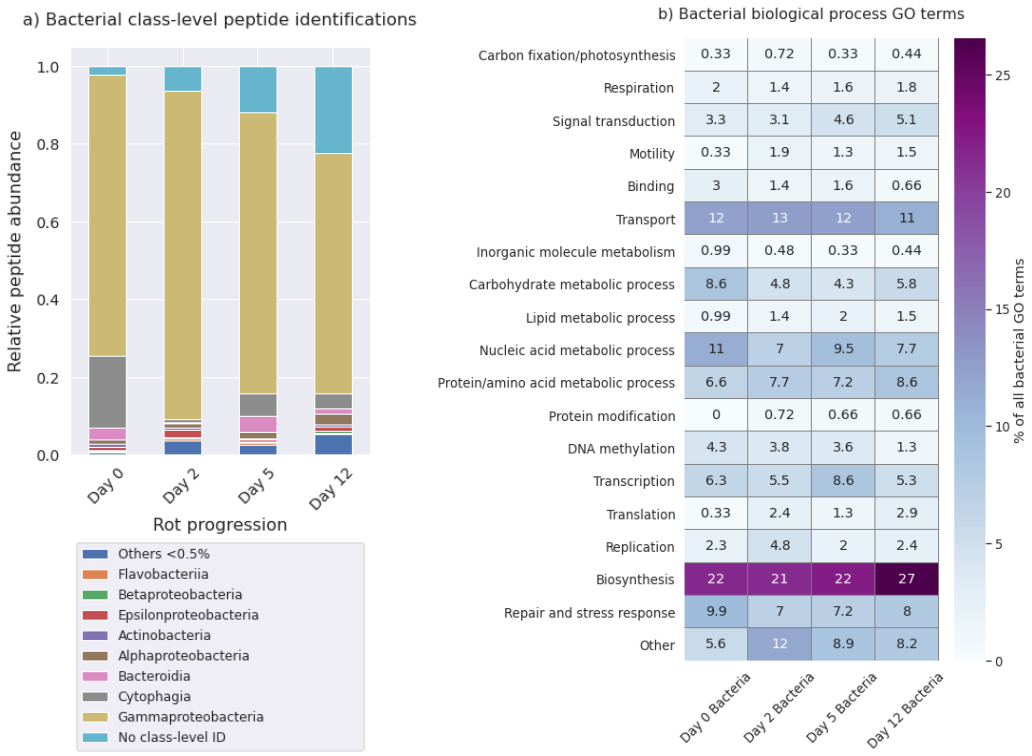

Figure 6. Bacterial progression of the algal degradation: relative abundance contribution of the major bacterial taxonomic classes (>0.5%) from the timepoint proteomes adjusted by relative peptide spectral abundance (NAAF) (a); and the percentage of the total number of bacterial peptide biological process gene ontology (GO) terms (b). 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 broad categories.