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

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

4   [1]School of Oceanography, University of Washington, Seattle, WA, 98105, USA
5   [2]Darling Marine Center, University of Maine, Walpole, ME, 04573, USA
6   [3]Now at Maine Department of Marine Resources, West Boothbay Harbor, ME, 04575, USA

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

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

10   Adams, Cheyenne M.

11   Homolka, Khadijah K., ORCID iD: 0000-0002-5208-8740

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

13   Mayer, Lawrence M.

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

15   Running title: Algal protein degradation

1   Manuscript submitted to *Frontiers in Marine Science*

18 **ABSTRACT**

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

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., 2012, 2014; 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 in regards to 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 detirial 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

87 marine organic matter (Duffy et al, submitted, Abdulla et al., 2018; Liu et al., 2010; Yamada
88 & Tanoue, 2003).

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

104 **MATERIAL AND METHODS**

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

108 *Algal degradation*

109   A culture of the centric diatom *Thalassiosira weissflogii* was grown to approximately
110 1×10^6 cells/mL, concentrated by centrifugation, and frozen at -80 ⁰C to render the algal
111 culture non-viable. Aliquots were thawed and resuspended to 2 g/L by dry weight using 1 µm
112 filtered and UV sterilized seawater from the Damariscotta Estuary, Gulf of Maine.
113 Unsterilized seawater filtered to 1 µm was used as an inoculum to induce bacterial
114 decomposition of the algal detritus, with 1 mL added to 15 L of algal suspension. At each
115 time point (1, 2, 5 & 12 days), 100 mL of the degradation slurry was vacuum filtered onto 25
116 mm diameter 0.3 um pore size glass fiber membranes (Sterlitech GF75) at 4 °C and stored
117 frozen prior to proteomic extraction and analysis. Further experimental details about the
118 degradation can be found in Adams et al. (2019).

119 *Total hydrolyzable amino acids*

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

126   ***Proteomic sample extraction***

127   Peptides were extracted with and without the use of the serine protease trypsin, which
128   specifically cleaves sequences on the carboxyl side (or the "C-terminal side") of lysine and
129   arginine, except when either is bound to a C-terminal proline. Peptides created via trypsin
130   digestion are usually between 700-1500 daltons (6-14 amino acids), an ideal range for mass
131   spectrometry detection (Laskay et al., 2013). Approximately 100 mg of material was scraped
132   from a filter into detergent-free 50 mM ammonium bicarbonate lysis buffer with 5 mM
133   ethylenediaminetetraacetic acid (EDTA). The chilled suspension was lysed via three cycles
134   each of mechanical disruption with silica beads (50% 100 µm diameter and 50% 400 µm),
135   freeze-thawing in a methanol-dry ice bath, and 30 seconds in a high-power water bath
136   sonicator. The supernatant was removed after centrifugation at 4800 rpm and the protein
137   concentration was determined using a modified Lowry assay using reagents from Bio-Rad
138   (Hercules, California). Extracted protein was subjected to reduction of disulfides with
139   dithiothreitol and carbamidomethylation with iodoacetamide. One half (300 uL) of the
140   protein extraction was subjected to in-solution digestion with trypsin (Promega Gold)
141   overnight at room temperature by a ratio of 1 µg trypsin: 25 µg total protein in 50 mM
142   ammonium bicarbonate buffer with 5 mM EDTA. The remaining 300 uL of protein
143   extraction was left untreated with trypsin. Both trypsin and no-trypsin treatments were
144   desalted using NestGroup macro-spin C18 columns (Southborough, Massachusetts) and
145   resuspended in 5% acetonitrile with 0.1% formic acid and Waters Hi3 E. coli peptide
146   standard mixture (100 fmol/L).

147   Reverse-phase liquid chromatography-high resolution mass spectrometry (LC-
148   HRMS) analysis was performed in duplicate with a Thermo Science (Waltham,
149   Massachusetts) EASY-nanoLC system coupled to a Thermo Orbitrap Fusion Tribrid HRMS.
150   Peptides were separated on a home-packed analytical column consisting of a 37 cm long, 75-
151   µm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100˚A,
152   5mm; Michrom) coupled to a 4 cm long, 100 m i.d. precolumn (Magic C18AQ, 200˚A, 5mm;
153   Michrom). Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100%
154   LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 120-
155   minute gradient from 5-35% solvent B. All analyses were carried out in positive mode at an
156   NSI spray voltage of 2 kV. Data-dependent acquisition (DDA) on the top 10 ions was carried
157   out using first higher energy collision dissociation (HCD) and then electron transfer
158   dissociation (ETD) fragmentation methods for duplicate injections. The MS1 (parent peptide
159   ion) scan range was 400-2000 m/z.

Proteomic data analysis

160

161   Peptides were identified from the raw mass spectra using a combination database-
162 driven and database-independent de novo sequencing approach. De novo peptide
163 sequencing, where the amino acid sequence of peptides is determined solely from the mass
164 spectra without comparison to a reference database (Allmer, 2011), was advantageous in this
165 study because we lacked a paired metagenome to thoroughly characterize the microbial
166 community from the seawater inoculum (Muth et al., 2015, 2018; O'Bryon et al., 2020). The
167 combining of database and de novo is termed de novo-directed proteomics, and was
168 performed using PeaksDB within Peaks Studio (v8.5; Bioinformatics Solutions, Waterloo,
169 Canada; Zhang et al. (2012)). The de novo-directed approach has been shown to significantly
170 improve sensitivity and accuracy in comparison to existing database search techniques
171 (Zhang et al., 2012).

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

186   Peptide identification confidences are calculated differently between the database
187 search-identified and de novo sequenced peptides. For the de novo identifications, an amino
188 acid-level confidence score is calculated based on mass deviation from spectral features and
189 expressed as a percentage value. We accepted de novo peptides >80% average residue local
190 confidence (ALC) with no single amino acid score <50%. For database searches, a false
191 discovery rate (FDR) was set <1.0% using a reversed database target-decoy strategy (i.e.,
192 searching against reversed reference protein sequences) (Elias & Gygi, 2010). De novo
193 sequencing was also incorporated into the database searches, as PeaksDB first compares de
194 novo sequences to the reference database to find approximate matches and decrease the
195 search space. Agreement between de novo sequences and database search hits are also used,
196 in part, to generate peptide-level confidence scores derived from the p-value indicating the
197 statistical significance of the peptide-spectrum match (the -10lgP score). Our threshold was a
198 -10lgP score >20. Such a score is equivalent to a p-value of 1%, signifying the probability
199 that the identification is to a false peptide sequence (Zhang et al., 2012).

200   Protein identifications are notoriously difficult in samples containing many different
201 organisms because some peptides are shared in proteins from multiple taxa. We required a
202 minimum of 1 unique peptide per protein identification. Matching of proteins from peptides
203 found only by de novo sequencing is detailed below.

204   ***Searching for Amino Acid Modifications***

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

225   *Mapping de novo peptides to proteins*

226   The de novo-directed PeaksDB workflow used here outputs peptides matched to the
227   database and de novo peptides (sequences only, no additional information). To identify
228   peptides that came from the diatom detritus or bacterial proteins not found by database
229   searching, the de novo peptides were aligned to the reference database (*T. weissflogii*
230   transcriptomes with GOS Gulf of Maine metagenomes) using PepExplorer (Leprevost et al.,
231   2014). PepExplorer considers common de novo sequencing errors and limitations (such as
232   leucine and isoleucine equivalence or other combinations of amino acids having the same
233   mass) and identifies regions of local similarity between sequences. We also performed an
234   alignment of the sequences against a reversed version of the database to estimate a false
235   discovery rate, which was kept <1%. Alignments were accepted at a 95% identify agreement
236   cutoff and protein identification required at least one unique peptide alignment.

237   Many de novo peptides did not match back to the reference database, which was not
238   unexpected given the low number of sequences in the database for heterotrophic bacteria and
239   the GOS sampling locations being more open ocean (Yooseph et al., 2007) compared to our
240   estuarine seawater inoculum. To overcome this limitation, we additionally mapped the de
241   novo peptides to proteins contained in the entire UniProt KnowledgeBase database (UniProt
242   Consortium, 2018), which contains over 200 million sequences from thousands of taxa. The
243   mapping was performed using the UniPept lowest common ancestor tool (Mesuere et al.,
244   2016), which is built specifically for metaproteomic data and determines the taxonomic
245   origins of peptides to the lowest possible phylogenetic rank (since some peptides are highly
246   conserved, they may match as to only 'Bacteria', or even 'Organism', rather than to a species
247   or genus level). The UniPept output provides the best view of the bacteria present in the
248   experiment given the lack of genes or transcripts from which to build an ideal reference
249   database.

250   *Gene Ontology Terms and Secondary Structures*

251   To identify gene ontology (GO) term annotations, the peptide sequences were aligned
252 to the UniProt protein database using UniPept's metaproteomic functional analysis tool
253 (Gurdeep Singh et al., 2019), which is built upon a lowest common ancestor peptide search
254 described above. The GO categories were condensed from the broader set in order to
255 eliminate redundancy using the REViGO tool (available at http://revigo.irb.hr/) as well as
256 manually.

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

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

266 **RESULTS & DISCUSSION**

267 ***General characteristics of the degradation***

268   As previously described (Adams et al., 2019), algal biomass decreased by a factor of
269 2.4 during the 12 day incubation. Meanwhile, bacterial abundance increased by a factor of 5.
270 Total particulate carbon decreased by 31% over the twelve days (Figure 1). In total, the algal
271 detritus was partly remineralized via respiration and partly converted to bacterial biomass,
272 but mostly remained as necromass (Adams et al., 2019). The degradation experiment started
273 with a small inoculum of bacteria added to a detrital slurry that contained some algal
274 proteases present during diatom cell harvesting. The algal detritus also likely contained small
275 amounts of bacteria (or bacterial necromass) from the diatom culture, which was not
276 confirmed to be axenic. The living bacteria then responded to the detritus and began to grow
277 exponentially (Figure 1), during which time they released enzymes in order to break down
278 and take up the diatom organic matter. The ratio of enzymatically hydrolyzable amino acids
279 (a biomimetic measurement of protein bioavailability, see Mayer et al., 1995) to the organic
280 carbon content remained constant throughout the experiment (Adams et al., 2019) indicating
281 that there was no selective remineralization or preservation of proteins relative to other
282 organic material at this early stage of degradation. However, by looking closely at peptides,
283 we show that numerous compositional changes occurred in the protein pool during the twelve
284 day period.

285   Our interpretation of protein dynamics during the degradation experiment relies on
286 the distinction between two different pool of peptides: those produced in situ during the
287 experiment (not treated with trypsin, which we will refer to as 'naturally digested peptides')
288 and those produced artificially by our laboratory digestion (referred to as 'trypsin-digested
289 peptides'). We hypothesize that the naturally digested peptides represent a mixture of
290 proteinaceous material that is being accessed by the heterotrophic bacteria and peptides that
291 have been released from larger proteins. The trypsin-digested peptides should represent the
292 total of all digestible proteinaceous material present at each time point, including that which
293 is not being actively degraded (e.g. the material that has resisted degradation thus far).

294 Differences between these two diatom protein-derived pools provides information about
295 relative labilities of the substrate proteins, in addition to information about the bacterial
296 strategies of degrading the diatom necromass.

297   Both the trypsin and naturally digested peptide pools should also contain peptides
298 sourced from the bacteria growing in the experiment. Given cell counts and estimated carbon
299 per diatom and bacterial cells, there is about 150 times more algal protein than bacterial
300 protein in the day 0 time point, and this decreases to 15 times more after 12 days (Adams et
301 al., 2019). These peptides can be used to determine the types of bacteria present during the
302 experiment, and the cellular functions that these bacteria are performing (e.g. Bergauer et al.,
303 2017; Mikan et al., 2020). In the following sections we examine the peptide results with the
304 goal of trying to better understand the degradative processes occurring over the course of
305 experiment and their implications for detrital protein cycling in the marine systems.

306 ### *Algal Peptide Identification and Characteristics*

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

319   The number of identified diatom peptides and proteins peaked at day 2 during the
320 initial phase of exponential bacterial growth, and then decreased to lowest values at day 12
321 (Table 1). We sorted the algal peptides by their gene ontology cellular compartment terms
322 (GO Terms; Figure 2; see also (Mikan et al., 2020; Riffle et al., 2017). The data were
323 aggregated into eight major cellular compartments, and the three numerically dominant
324 groups were chloroplast, cellular membrane, and cytoplasmic proteins (Figure 2). The
325 trypsin-digested day 0 fraction had a cellular component GO term distribution consistent with
326 that of living diatom cells: alongside the degradation experiment's algal GO terms we plot
327 those from peptides obtained from two *Thalassiosira pseudonana* proteomes published by
328 Dyhrman et al., 2012 (Figure 2).

329   The naturally digested peptides contained far fewer GO term identifications after day
330 2, and by day 12 the only identifications in that fraction were from chloroplast and membrane
331 peptides. This indicates that peptides and proteins from other cellular components were
332 effectively degraded during the initial phase of bacterial exponential growth (Figure 1) and
333 were not being actively degraded at day 12 because they were no longer present. Consistent
334 with this membrane recalcitrance hypothesis is that the number of identifications of trypsin-
335 digested peptides sourced from larger proteins also decreased for most cellular compartments
336 other than membranes and the chloroplast (Figure 2). However, there were still some trypsin-
337 digested peptides at day 12 for components from the mitochondria and cytoplasm, suggesting
338 that perhaps cellular location alone does not necessarily determine whether a protein or
339 peptide will survive the initial stages of environmental degradation.

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

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

358   ***Preserved protein motifs***

359   To further interrogate the proteinaceous material that survived the degradation, five
360   classes of peptide secondary structure were evaluated: random coil, α-helices, β-strands,
361   transmembrane α-helices, and transmembrane β-strands. Closely set α-helices contain strong
362   hydrogen bonding between weakly polar (Ser, Thr, Cys) and strongly polar (Asn, Gln, Glu,
363   Asp, His, Arg, Lys) amino acid residues of neighboring α-helices (Liu et al., 2011). Helices
364   are commonly found in transmembrane proteins (Sakai & Tsukihara, 1998; Stevens & Arkin,
365   2000), and their ability to bend can account for the hydrophobic mismatch of the lipid bilayer
366   (Yeagle et al., 2007). In contrast, β-strands lay relatively flat and have been hypothesized in a
367   marine context to adhere to mineral surfaces - ultimately aiding to their protection and
368   resulting in enhanced preservation (Shamblin et al. 1998; Oleschuk et al. 2000; Ovesen et al.
369   2011). 'Random coil' is not a true secondary structure, but rather is an aggregate term for
370   short sequences where there is an absence of a helix or sheet character (Smith et al., 1996).
371   Benchmarking of the Proteus2 machine-learning algorithm we use here showed per-residue
372   prediction accuracy to be 87-91% for transmembrane α-helices, 86% for transmembrane β-
373   strands, and 88% for non-membrane secondary structures (Montgomerie et al., 2008).

374   We note a progressive change in secondary structure distribution as the degradation
375   proceeded, with membrane α-helices becoming more important (Figure 3). At the same
376   time, β-strands became a less common motif. This trend in secondary structure distributions
377   is consistent with the GO term evidence indicating that membrane proteins are preferentially
378   retained in the system and suggests that their tightly wrapped, difficult to denature, secondary
379   structure could be a factor aiding in preservation. Random coils are the most common motif,
380   and they do not change in relative importance over time, indicating that they are not
381   particularly prone to resistance or degradation in our experiment, or that the coil category is
382   too broad to capture any selective processes. These results are generally consistent with those
383   of Nunn et al. (2010) in a similar degradation of a marine diatom. They identified 23 and 4
384   diatom peptides after 10 and 23 days of incubation with a seawater microbiome, respectively.

385 Of the surviving 4 diatom peptides at the end of the experiment, 2 had transmembrane
386 domains (Nunn et al., 2010).

387 To further investigate the theory that secondary structure is related to enhanced
388 preservation of membrane protein components, we also compared the predicted secondary
389 structure of just algal chloroplast/integral component of membrane peptides and algal
390 cytoplasmic peptides that are identified at each time point of the degradation (Figure 3).
391 Chloroplast and membrane peptides, the bulk of diatom peptide identifications on day 12
392 (from 167 individual proteins across both trypsin and naturally digested fractions), were
393 50.7% coil, 36.3% α-helix, 3.02% β-strand, 9.93% transmembrane α-helix, and 0%
394 membrane β-strand. In comparison, the cytoplasmic peptides (from 98 proteins across both
395 trypsin and naturally digested fractions) identified in the initial day 0 proteome were
396 comparatively more enriched in β-strands and depleted in α-helices and transmembrane α-
397 helices, with 48.3% coil, 23.8% α-helix, 27.0% β-strand, 0% transmembrane α-helix, and 0%
398 transmembrane β-strand. Supporting the theory that α-helices are linked to preservation is
399 that the few surviving cytoplasmic peptides on day 12 (3 proteins total, all in the trypsin-
400 digested fraction) had 10% more α-helix character than the initial day 0 cytoplasmic peptides,
401 at 40.6% coil, 33.1% α-helix, 26.3% β-strand, 0% transmembrane α-helix, and 0%
402 transmembrane β-strand (Figure 3). Our findings are consistent with the notion that
403 hydrophobic interactions appear to be important in preserving membrane-associated proteins
404 and peptides during early diagenesis (Nguyen & Harvey, 2001; Nunn et al., 2010).

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

410 *Post translational modifications of amino acids*

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

418 In the marine environment, PTMs have been linked to degradation and early
419 diagenesis. For example, the amino acid beta alanine accumulates in the hydrolyzable phase
420 of marine sedimentary organic matter via modification of aspartic acid (Cowie and Hedges
421 1994). Modification of the nitrogen-containing side chains of glutamine, asparagine and
422 arginine can lead to the accumulation of peptides containing deaminated side chains within
423 anoxic marine sediment pore waters (Abdulla et al., 2018). We recently observed deamidated
424 peptides in the sinking POM from the eastern tropical North Pacific oxygen deficient zone
425 (Duffy et al, submitted), where Van Mooy et al. (2002) hypothesized that amino acids could
426 be selectively deaminated in order to provide reduced nitrogen to fuel chemoautotrophic
427 processes (Van Mooy et al., 2002). In laboratory experiments, Keil and Kirchman showed
428 that methylated peptides were accessed less efficiently by bacteria than non-methylated
429 peptides (Keil & Kirchman, 1992). Glycan modifications of proteins contribute to stability, a
430 mechanism that's been used to increase the shelf life of protein and peptide-based

431 pharmaceuticals (Zhou & Qiu, 2019). Indeed, Keil and Kirchman showed that glycosylated
432 ribulose-1,5-bisphosphate (RuBisCo) was degraded 100 times more slowly than its
433 unmodified counterpart (Keil & Kirchman, 1993). All together, the marine literature
434 suggests that PTMs are highly relevant to protein degradation, as the by-products of protein
435 degradation processes and/or as potential factors in degradation resistance.

436   In the present study, we first performed a non-discriminatory 'open' search of
437 thousands of possible PTMs in the UniMod database using PeaksPTM (see above in
438 Methods). We then selected only the 10 most frequently occurring mass modifications in the
439 open search results for the suite of PTMs in the ultimate database and de novo peptide
440 searches. Our goal was to identify patterns in PTM distribution to learn more about their roles
441 in degradation and preservation. To that end, we wanted to distinguish between PTMs present
442 in algal protein prior to degradation and PTMs associated with the degradative process. We
443 compared the PTMs observed in the naturally digested peptide pool, which should represent
444 proteins being actively degraded by the bacteria, against the trypsin-digested peptides. This
445 comparison was done for only the algal protein; no bacterial proteins were included in this
446 differential analysis of PTMs (Figure 4).

447   Oxidation PTMs were generally enriched in the trypsin-digested component, implying
448 that these modifications occurred within the cell prior to heterotrophic attack. Generally,
449 they did not increase in relative abundance over time in the trypsin-digested peptide pool,
450 suggesting that oxidations don't meaningfully impact the lability of proteinaceous material
451 (Figure 4). It's notable that during the experiment, chloroplast protein-derived peptides
452 increasingly became dominant among identifications (Figure 2). The algal culture was
453 exposed to light before the degradation experiment in the dark. Photosynthesis as an
454 oxygenic process produces active oxygen species and radicals which can cause damage to
455 cells. Oxidations of amino acids are frequent PTMs in photosynthesis-associated proteins
456 (Aro et al., 2005; Galetskiy et al., 2011). Similarly, lysine acetylation of chloroplast proteins
457 has been demonstrated in plants (Lehtimäki et al., 2015) and in diatoms (X.-H. Chen et al.,
458 2018), though the PTM frequently occurs elsewhere in the cell (Z. Chen et al., 2018). The
459 abundance of thylakoid membrane and chloroplast-associated proteins that accumulate due to
460 preferential preservation (Figure 2) is likely why oxidations and lysine acetylations are
461 dominant PTMs of the trypsin-digested peptides (Figure 4).

462   In contrast, two PTMs were strongly associated with the presumably more detrital
463 naturally digested peptides: deamidation of asparagine and arginine methylation (Figure 4).
464 Our observation that asparagines and arginines are more modified in detrital proteins has
465 several possible explanations. These modifications could have occurred within the living
466 cell, resulting in a pool of protein that was easily accessible, which is why they were found
467 more in the naturally degraded peptide pool. Alternatively, these PTMs might have been
468 created during the degradation process and could accumulate because once created they are
469 further degraded slower than their unmodified counterparts. This latter explanation would
470 account for the effective accumulation of deamidated peptides in both sediment pore waters
471 (Abdulla et al., 2018) and in sinking particulate matter (Duffy et al, submitted). Thus, we
472 hypothesize that deamination occurs during degradation, though continued research is
473 warranted.

474   Unlike deamidation, methylation most likely occurs within the living cell, where it is
475 a common PTM that is used as a control of numerous cellular functions (Ghesquière et al.,
476 2011). Methylated peptides have been shown to be inefficiently assimilated and degraded by

477 heterotrophic bacteria (Keil & Kirchman, 1992), suggesting that PTMs produced within
478 living cells may also play a role in determining the rate of protein degradation.

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

Amino acid compositions

493   One of the deepest sets of literature related to protein degradation in marine systems is
494 that of the 'total hydrolyzable amino acid' pool (Dauwe et al., 1999; Dauwe & Middelburg,
495 1998; Lee et al., 2004; Wakeham et al., 1997). THAA analyses show clear trends during
496 long-term carbon degradation and preservation including an accumulation of the amino acids
497 glycine, serine and threonine (Dauwe and Middelburg 1998), and the creation of the non-
498 protein amino acids beta-alanine and gamma-aminobutyric acid from aspartic and glutamic
499 acids (Cowie and Hedges 1994). Despite the widespread use of degradation indices derived
500 from THAA analyses, it has been difficult to reconcile changes in bulk amino acid
501 compositions to known protein amino acid compositions, especially during the very early
502 stages of degradation (Keil et al., 2000). We compared the amino acid composition of the
503 peptides identified during the degradation experiment to the THAA pool, which we measured
504 independently. To facilitate this comparison, we plotted mole fractions of amino acids in the
505 THAA against those in the identified algal peptides, combining the naturally and trypsin-
506 digested amino acid compositions adjusted for relative peptide abundance (Figure 5). The
507 near 1:1 agreement between the two approaches indicates two things: a) the protein amino
508 acid compositions measured using the newer 'omic' approach can be effectively integrated
509 into the large body of literature based on THAA analysis, which will become useful as
510 peptidomic approaches are applied to samples further along the degradation pathway (e.g.
511 sediment samples); and b) while the omic approach identifies specific ways in which protein
512 undergoes degradation in the ocean, the early stages remain remarkably 'nonselective' at the
513 bulk molecular level (Hedges et al., 2001).

Bacterial Community

515   The microbial inoculum for the degradation experiment was sourced from the
516 Damariscotta River Estuary, whose marine waters come from the Gulf of Maine.
517 Heterotrophic bacteria grew exponentially through the middle stages of the experiment
518 (Figure 1). A peptide-based lowest common ancestor analysis was performed using UniPept
519 (Gurdeep Singh et al., 2019) to assign bacterial taxa to the lowest phylogenetic level possible.
520 The taxonomic hits were then adjusted to account for peptide spectral abundance and
521 aggregated at the class level (Figure 6a). There is precedence for using label-free

522 metaproteomic data for microbial biomass determinations (Kleiner et al., 2017), and here we
523 use the approach to learn about the community of microbes degrading the algal detritus.

524   There is taxonomic overlap of the initial composition in our study to that of a
525 pyrosequencing survey of planktonic microbes at three stations in the Gulf of Maine (Li et
526 al., 2011). We found the microbial community at the initial time point was dominated by
527 Gammaproteobacteria (~60% of the peptides), with notable contributions of Cytophaga
528 (~5%), Bacteroidia (~2%) and Alphaproteobacteria (~2%) (Figure 6a). During the 12 day
529 experiment, the microbial community composition changed minimally, with two exceptions:
530 Firstly, there was a ten-fold increase of peptides that were bacterial but could not be uniquely
531 identified at the class level or below, from 2.2% of peptides on day 0 to 22% of peptides by
532 day 12 (Figure 6a). At the same time, the contribution of low abundance (<0.5% of all
533 peptides) classes increased from 0.53% on day 0 to 5.4% on day 12 (Figure 6a). Both
534 increases, in non-specific bacterial peptides and in low-abundance class peptides, are
535 potential indications that the diversity of the bacterial community was increasing during the
536 experiment.

537   The increase of non-specific bacterial peptides can be explained in two ways. Most
538 likely, a true increase in diversity and richness resulted from a broader array of shared,
539 conserved proteins being produced in the system by related bacteria. An alternative
540 hypothesis is that there was an ingrowth of bacterial necromass generating degraded bacterial
541 peptides that are no longer very taxonomically specific. Bacterial necromass in this case
542 could be from organisms in the seawater inoculum or bacteria present during the diatom
543 culture growth. For instance, Gammaproteobacteria are known to thrive within the
544 phycosphere of diatoms (Amin et al., 2012), and their peptide abundance peaked at day 2
545 when the active bacteria were growing exponentially. It is difficult to discern which of these
546 hypotheses is correct and further work will need to be done to evaluate the processing of
547 bacterial detritus by other bacteria. Nonetheless, the overall taxonomic changes observed in
548 the bacterial peptidome during the twelve days were minimal.

549   Unlike the somewhat ambiguous taxonomic information within the bacterial
550 peptidome, biological process GO terms associated with the bacterial peptides provide a
551 clearer view of how the bacterial community responded to the algal detritus (Figure 6b). The
552 most detected GO terms are associated with transmembrane transport, carbohydrate
553 metabolism, and DNA replication and transcription (Figure 6b). These are the terms most
554 strongly associated with bacterial growth (Mikan et al 2020) and suggest that most of the
555 bacterial peptides detected are from living bacteria. This strengthens support for the
556 hypothesis that the Gammaproteobacteria in the samples are for the most part living and not
557 detrital.

558   Our bacterial functional data are generally consistent with recent work by Mikan et al.
559 (2020), who used metaproteomic tools and a GO-term based functional analysis to evaluate
560 the heterotrophic bacterial response to a pulse of detrital organic matter in two Arctic
561 microbiomes during 10-day shipboard incubations. In that study, the bacterial community
562 increased protein synthesis, carbohydrate degradation, and cellular redox processes while
563 simultaneously decreasing C1 metabolism (Mikan et al., 2020). Throughout our experiment
564 we observe steady levels of transmembrane transport and protein metabolism terms.
565 Carbohydrate metabolic process GO terms maximize in the first two days of the degradation,
566 with increasing biosynthesis GO terms by day 12 (Figure 6b). Mikan et al (2020) suggested
567 that the bacterial community shifted their carbon acquisition strategies intracellularly before
568 there were large shifts in the taxonomic structure of the community. Without a paired

569 metagenome or metatranscriptome with which to perform proteomic database searches, our
570 bacterial peptide data and paired GO term data are not as complete, but show the same
571 general trends, lending further support to the notion of Mikan et al. that functional
572 composition and redundancy, not taxonomy, may be the most relevant factor when evaluating
573 how effectively organic matter is or will be processed by bacteria in the ocean.

574 **CONCLUSIONS**

575 In this study we evaluated proteinaceous material in a marine system both with and without
576 the use of trypsin as an extraction and identification tool. To our knowledge, this is the first
577 such attempt to disentangle proteins that are being degraded by microbes from those that are
578 resistant to degradation. We show that proteins from specific cellular locations are
579 preferentially preserved during the initial stages of degradation. As has been hypothesized
580 and demonstrated for bacterial membrane proteins (Jiao et al., 2010; Kaiser & Benner, 2008;
581 Nagata et al., 1998) and algal membrane proteins (Laursen et al., 1996; Nguyen & Harvey,
582 2001; Wolfe et al., 2006), we conclusively illustrate that proteins associated with diatom
583 chloroplasts and membranes resist initial degradation better than those without such
584 association. The many membrane-associated and few cytoplasm peptides that resist
585 degradation also are relatively enriched in α-helices and depleted in β-strands, consistent with
586 the cellular location data since α-helices are enriched in membrane proteins. However, the
587 extent to which α-helices themselves lead to degradation resistance remains to be more
588 thoroughly evaluated, as 1) α-helices are not exclusive to membrane proteins and 2) there
589 could be other reasons for membrane protein survival over time that causes membrane
590 proteins' α-helix rich motifs to become enriched in detrital material.

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

ABBREVIATIONS

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

GOS Global Ocean Survey

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 with the data set identifier PXD027843 to be made available with publication.

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

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

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

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

17   Manuscript submitted to *Frontiers in Marine Science*

673   samples: Charge-matched recovery standards and reduced analysis time. *Marine*
674   *Chemistry*, *37*(3), 223–238. https://doi.org/10.1016/0304-4203(92)90079-P
675 Dauwe, B., & Middelburg, J. J. (1998). Amino acids and hexosamines as indicators of
676   organic matter degradation state in North Sea sediments. *Limnology and*
677   *Oceanography*, *43*(5), 782–798. https://doi.org/10.4319/lo.1998.43.5.0782
678 Dauwe, B., Middelburg, J. J., Herman, P. M. J., & Heip, C. H. R. (1999). Linking diagenetic
679   alteration of amino acids and bulk organic matter reactivity. *Limnology and*
680   *Oceanography*, *44*(7), 1809–1814. https://doi.org/10.4319/lo.1999.44.7.1809
681 Dhillon, R. S., & Denu, J. M. (2017). Using comparative biology to understand how aging
682   affects mitochondrial metabolism. *Molecular and Cellular Endocrinology*, *455*, 54–
683   61. https://doi.org/10.1016/j.mce.2016.12.020
684 Dong, H.-P., Wang, D.-Z., Dai, M., & Hong, H.-S. (2010). Characterization of particulate
685   organic matter in the water column of the South China Sea using a shotgun proteomic
686   approach. *Limnology and Oceanography*, *55*(4), 1565–1578.
687   https://doi.org/10.4319/lo.2010.55.4.1565
688 Dyhrman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. I., Alexander,
689   H., Whitney, L. P., Drzewianowski, A., Bulygin, V. V., Bertrand, E. M., Wu, Z.,
690   Benitez-Nelson, C., & Heithoff, A. (2012). The transcriptome and proteome of the
691   diatom Thalassiosira pseudonana reveal a diverse phosphorus stress response. *Plos*
692   *One*, *7*(3), e33768–e33768. https://doi.org/10.1371/journal.pone.0033768
693 Estes, E. R., Pockalny, R., D'Hondt, S., Inagaki, F., Morono, Y., Murray, R. W., Nordlund,
694   D., Spivack, A. J., Wankel, S. D., Xiao, N., & Hansel, C. M. (2019). Persistent
695   organic matter in oxic subseafloor sediment. *Nature Geoscience*, *12*(2), 126–131.
696   https://doi.org/10.1038/s41561-018-0291-5
697 Fejjar, S., Melanson, A., & Tremblay, L. (2021). Pore waters as a contributor to deep-water
698   amino acids and to deep-water dissolved organic matter concentration and
699   composition in estuarine and marine waters. *Marine Chemistry*, *233*, 103985.
700   https://doi.org/10.1016/j.marchem.2021.103985
701 Finkel, Z. V., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I., & Irwin, A. J. (2016).
702   Phylogenetic Diversity in the Macromolecular Composition of Microalgae. *PLoS*
703   *ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0155977
704 Galetskiy, D., Lohscheider, J. N., Kononikhin, A. S., Popov, I. A., Nikolaev, E. N., &
705   Adamska, I. (2011). Mass spectrometric characterization of photooxidative protein
706   modifications in Arabidopsis thaliana thylakoid membranes. *Rapid Communications*
707   *in Mass Spectrometry: RCM*, *25*(1), 184–190. https://doi.org/10.1002/rcm.4855
708 Ghesquière, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M.,
709   Schymkowitz, J., Rousseau, F., Vandekerckhove, J., & Gevaert, K. (2011). Redox
710   Proteomics of Protein-bound Methionine Oxidation. *Molecular & Cellular*
711   *Proteomics : MCP*, *10*(5). https://doi.org/10.1074/mcp.M110.006866
712 Gray, N., Zia, R., King, A., Patel, V. C., Wendon, J., McPhail, M. J. W., Coen, M., Plumb, R.
713   S., Wilson, I. D., & Nicholson, J. K. (2017). High-Speed Quantitative UPLC-MS
714   Analysis of Multiple Amines in Human Plasma and Serum via Precolumn
715   Derivatization with 6-Aminoquinolyl-N-hydroxysuccinimidyl Carbamate:
716   Application to Acetaminophen-Induced Liver Failure. *Analytical Chemistry*, *89*(4),
717   2478–2487. https://doi.org/10.1021/acs.analchem.6b04623
718 Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S.,
719   Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of
720   Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615.
721   https://doi.org/10.1021/acs.jproteome.8b00716
722 Hedges, J. I., Baldock, J. A., Gélinas, Y., Lee, C., Peterson, M., & Wakeham, S. G. (2001).

723   Evidence for non-selective preservation of organic matter in sinking marine particles.
724   *Nature*, *409*(6822), 801–804. https://doi.org/10.1038/35057247

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

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

731 Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., Kirchman, D.
732   L., Weinbauer, M. G., Luo, T., Chen, F., & Azam, F. (2010). Microbial production of
733   recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean.
734   *Nature Reviews Microbiology*, *8*(8), 593–599. https://doi.org/10.1038/nrmicro2386

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

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

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

744 Keil, R. G., & Kirchman, D. L. (1994). Abiotic transformation of labile protein to refractory
745   protein in sea water. *Marine Chemistry*, *45*(3), 187–196. https://doi.org/10.1016/0304-
746   4203(94)90002-7

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

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

751 Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., & Strous, M. (2017).
752   Assessing species biomass contributions in microbial communities via
753   metaproteomics. *Nature Communications*, *8*(1), 1–14. https://doi.org/10.1038/s41467-
754   017-01544-x

755 Knicker, H., Río, J. C. del, Hatcher, P. G., & Minard, R. D. (2001). Identification of protein
756   remnants in insoluble geopolymers using TMAH thermochemolysis/GC–MS.
757   *Organic Geochemistry*, *32*(3), 397–409. https://doi.org/10.1016/S0146-
758   6380(00)00186-8

759 Knicker, H., Scaroni, A. W., & Hatcher, P. G. (1996). 13C and 15N NMR spectroscopic
760   investigation on the formation of fossil algal residues. *Organic Geochemistry*, *24*(6),
761   661–669. https://doi.org/10.1016/0146-6380(96)00057-5

762 Laskay, Ü. A., Lobas, A. A., Srzentić, K., Gorshkov, M. V., & Tsybin, Y. O. (2013).
763   Proteome digestion specificity analysis for rational design of extended bottom-up and
764   middle-down proteomics experiments. *Journal of Proteome Research*, *12*(12), 5558–
765   5569. https://doi.org/10.1021/pr400522h

766 Laursen, A. K., Mayer, L. M., & Townsend, D. W. (1996). Lability of proteinaceous material
767   in estuarine seston and subcellular fractions of phytoplankton. *Marine Ecology
768   Progress Series*, *136*, 227–234. https://doi.org/10.3354/meps136227

769 Lee, C., Wakeham, S., & Arnosti, C. (2004). Particulate Organic Matter in the Sea: The
770   Composition Conundrum. *AMBIO: A Journal of the Human Environment*, *33*(8),
771   565–575. https://doi.org/10.1579/0044-7447-33.8.565

772 Lehtimäki, N., Koskela, M. M., & Mulo, P. (2015). Posttranslational Modifications of

773   Chloroplast Proteins: An Emerging Field. *Plant Physiology, 168*(3), 768–775.
774   https://doi.org/10.1104/pp.15.00117

775 Leprevost, F. V., Valente, R. H., Lima, D. B., Perales, J., Melani, R., Yates, J. R., Barbosa,
776   V. C., Junqueira, M., & Carvalho, P. C. (2014). PepExplorer: A similarity-driven tool
777   for analyzing de novo sequencing results. *Molecular & Cellular Proteomics: MCP,
778   13*(9), 2480–2489. https://doi.org/10.1074/mcp.M113.037002

779 Li, W. K. W., Andersen, R. A., Gifford, D. J., Incze, L. S., Martin, J. L., Pilskaln, C. H.,
780   Rooney-Varga, J. N., Sieracki, M. E., Wilson, W. H., & Wolff, N. H. (2011).
781   Planktonic Microbes in the Gulf of Maine Area. *PLOS ONE, 6*(6), e20981.
782   https://doi.org/10.1371/journal.pone.0020981

783 Liu, Z., Kobiela, M. E., McKee, G. A., Tang, T., Lee, C., Mulholland, M. R., & Hatcher, P.
784   G. (2010). The effect of chemical structure on the hydrolysis of tetrapeptides along a
785   river-to-ocean transect: AVFA and SWGA. *Marine Chemistry, 119*(1), 108–120.
786   https://doi.org/10.1016/j.marchem.2010.01.005

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

790 Mayer, L. M., Linda L., S., Sawyer, T., Plante, C. J., Jumars, P. A., & Sel, R. L. (1995).
791   Bioavailable amino acids in sediments: A biomimetic, kinetics based approach.
792   *Limnology and Oceanography, 40*(3), 511–520.
793   https://doi.org/10.4319/lo.1995.40.3.0511

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

797 Mayer, L. M., Schick, L. L., & Setchell, F. W. (1986). Measurement of protein in nearshore
798   marine sediments. *Marine Ecology Progress Series, 30*, 159–165.
799   https://doi.org/10.3354/meps030159

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

806 Mesuere, B., Willems, T., Jeugt, F. V. der, Devreese, B., Vandamme, P., & Dawyndt, P.
807   (2016). Unipept web services for metaproteomics analysis. *Bioinformatics, 32*(11),
808   1746–1748. https://doi.org/10.1093/bioinformatics/btw039

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

814 Montgomerie, S., Cruz, J. A., Shrivastava, S., Arndt, D., Berjanskii, M., & Wishart, D. S.
815   (2008). PROTEUS2: A web server for comprehensive protein structure prediction and
816   structure-based annotation. *Nucleic Acids Research, 36*(suppl_2), W202–W209.
817   https://doi.org/10.1093/nar/gkn255

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

821 Moore, E. K., Nunn, B. L., Goodlett, D. R., & Harvey, H. R. (2012). Identifying and tracking
822   proteins through the marine water column: Insights into the inputs and preservation

823   mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta*, *83*, 324–

824   359. https://doi.org/10.1016/j.gca.2012.01.002

825 Mulholland, M., & Lee, C. (2009). Peptide Hydrolysis and the Uptake of Dipeptides by

826   Phytoplankton. *Limnology and Oceanography*, *54*(4).

827   https://doi.org/10.4319/lo.2009.54.3.0856

828 Muth, T., Hartkopf, F., Vaudel, M., & Renard, B. Y. (2018). A Potential Golden Age to

829   Come-Current Tools, Recent Use Cases, and Future Avenues for De Novo

830   Sequencing in Proteomics. *Proteomics*, *18*(18), e1700150.

831   https://doi.org/10.1002/pmic.201700150

832 Muth, T., Kolmeder, C. A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F. J., Rensen,

833   S. S., Reichl, U., de Vos, W. M., Rapp, E., & Martens, L. (2015). Navigating through

834   metaproteomics data: A logbook of database searching. *Proteomics*, *15*(20), 3439–

835   3453. https://doi.org/10.1002/pmic.201400560

836 Nagata, T., Fukuda, R., Koike, I., Kogure, K., & Kirchman, D. (1998). Degradation by

837   bacteria of membrane and soluble protein in seawater. *Aquatic Microbial Ecology*, *14*,

838   29–37. https://doi.org/10.3354/ame014029

839 Nagata, T., & Kirchman, D. (1996). Bacterial degradation of protein adsorbed to model

840   submicron particles in seawater. *Marine Ecology Progress Series*, *132*, 241–248.

841   https://doi.org/10.3354/meps132241

842 Nguyen, R. T., & Harvey, H. R. (2001). Preservation of protein in marine systems:

843   Hydrophobic and other noncovalent associations as major stabilizing forces.

844   *Geochimica et Cosmochimica Acta*, *65*(9), 1467–1480. https://doi.org/10.1016/S0016-

845   7037(00)00621-9

846 Noble, W. S. (2015). Mass spectrometrists should only search for peptides they care about.

847   *Nature Methods*, *12*(7), 605–608. https://doi.org/10.1038/nmeth.3450

848 Nunn, B. L., Norbeck, A., & Keil, R. G. (2003). Hydrolysis patterns and the production of

849   peptide intermediates during protein degradation in marine systems. *Marine*

850   *Chemistry*, *83*(1–2), 59–73. https://doi.org/10.1016/S0304-4203(03)00096-3

851 Nunn, B. L., Ting, Y. S., Malmström, L., Tsai, Y. S., Squier, A., Goodlett, D. R., & Harvey,

852   H. R. (2010). The path to preservation: Using proteomics to decipher the fate of

853   diatom proteins during microbial degradation. *Limnology and Oceanography*, *55*(4),

854   1790–1804. https://doi.org/10.4319/lo.2010.55.4.1790

855 Obayashi, Y., & Suzuki, S. (2008). Occurrence of exo- and endopeptidases in dissolved and

856   particulate fractions of coastal seawater. *Aquatic Microbial Ecology - AQUAT*

857   *MICROB ECOL*, *50*, 231–237. https://doi.org/10.3354/ame01169

858 O'Bryon, I., Jenson, S. C., & Merkley, E. D. (2020). Flying blind, or just flying under the

859   radar? The underappreciated power of de novo methods of mass spectrometric peptide

860   identification. *Protein Science*, *29*(9), 1864–1878. https://doi.org/10.1002/pro.3919

861 Olofsson, M., Lindehoff, E., & Legrand, C. (2019). Production stability and biomass quality

862   in microalgal cultivation – Contribution of community dynamics. *Engineering in Life*

863   *Sciences*, *19*(5), 330–340. https://doi.org/10.1002/elsc.201900015

864 Pantoja, S., Rossel, P., Castro, R., Cuevas, L. A., Daneri, G., & Córdova, C. (2009).

865   Microbial degradation rates of small peptides and amino acids in the oxygen

866   minimum zone of Chilean coastal waters. *Deep Sea Research Part II: Topical Studies*

867   *in Oceanography*, *56*(16), 1055–1062. https://doi.org/10.1016/j.dsr2.2008.09.007

868 Reinfelder, J. R., & Fisher, N. S. (1991). The Assimilation of Elements Ingested by Marine

869   Copepods. *Science*, *251*(4995), 794–796.

870   https://doi.org/10.1126/science.251.4995.794

871 Riffle, M., May, D. H., Timmins-Schiffman, E., Mikan, M. P., Jaschob, D., Noble, W. S., &

872   Nunn, B. L. (2017). MetaGOmics: A Web-Based Tool for Peptide-Centric Functional

873   and Taxonomic Analysis of Metaproteomics Data. *Proteomes*, *6*(1).
874   https://doi.org/10.3390/proteomes6010002
875 Roth, L., & Harvey, H. (2006). *Intact protein modification and degradation in estuarine*
876   *environments*. https://doi.org/10.1016/J.MARCHEM.2005.10.025
877 Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J.,
878   Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M.
879   R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd,
880   A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean
881   Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome*
882   *Research*, *18*(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761
883 Sakai, H., & Tsukihara, T. (1998). Structures of Membrane Proteins Determined at Atomic
884   Resolution. *J Biochem (Tokyo)*, *124*(6), 1051–1059.
885 Schmidt, F., Koch, B. P., Elvert, M., Schmidt, G., Witt, M., & Hinrichs, K.-U. (2011).
886   Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under
887   Contrasting Sedimentary Redox Conditions in the Black Sea. *Environmental Science*
888   *& Technology*, *45*(12), 5223–5229. https://doi.org/10.1021/es2003414
889 Shen, Y., Hixson, K., Tolić, N., Camp, D., Purvine, S., Moore, R. J., & Smith, R. (2008).
890   Mass spectrometry analysis of proteome-wide proteolytic post-translational
891   degradation of proteins. *Analytical Chemistry*. https://doi.org/10.1021/ac800077w
892 Smith, L. J., Fiebig, K. M., Schwalbe, H., & Dobson, C. M. (1996). The concept of a random
893   coil: Residual structure in peptides and denatured proteins. *Folding and Design*, *1*(5),
894   R95–R106. https://doi.org/10.1016/S1359-0278(96)00046-6
895 Stevens, T. J., & Arkin, I. T. (2000). Do more complex organisms have a greater proportion
896   of membrane proteins in their genomes? *Proteins: Structure, Function, and*
897   *Bioinformatics*, *39*(4), 417–420. https://doi.org/10.1002/(SICI)1097-
898   0134(20000601)39:4<417::AID-PROT140>3.0.CO;2-Y
899 Suzuki, S., Kogure, K., & Tanoue, E. (1997). Immunochemical detection of dissolved
900   proteins and their source bacteria in marine environments. *Marine Ecology Progress*
901   *Series*, *158*, 1–9. https://doi.org/10.3354/meps158001
902 Tanoue, E., Nishiyama, S., Kamo, M., & Tsugita, A. (1995). Bacterial membranes: Possible
903   source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta*,
904   *59*(12), 2643–2648. https://doi.org/10.1016/0016-7037(95)00134-4
905 Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble,
906   W. S., & Nunn, B. L. (2017). Critical decisions in metaproteomics: Achieving high
907   confidence protein annotations in a sea of unknowns. *The ISME Journal*, *11*(2), 309–
908   314. https://doi.org/10.1038/ismej.2016.132
909 UniProt Consortium, T. (2018). UniProt: The universal protein knowledgebase. *Nucleic Acids*
910   *Research*, *46*(5), 2699–2699. https://doi.org/10.1093/nar/gky092
911 Van Mooy, B. A. S., Keil, R. G., & Devol, A. H. (2002). Impact of suboxia on sinking
912   particulate organic carbon: Enhanced carbon flux and preferential degradation of
913   amino acids via denitrification. *Geochimica et Cosmochimica Acta*, *66*(3), 457–465.
914   https://doi.org/10.1016/S0016-7037(01)00787-6
915 Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., & Peterson, M. J. (1997). Molecular
916   indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica*
917   *Acta*, *61*(24), 5363–5369. https://doi.org/10.1016/S0016-7037(97)00312-8
918 Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E., & Schulz, G. E. (1991).
919   Molecular architecture and electrostatic properties of a bacterial porin. *Science (New*
920   *York, N.Y.)*, *254*(5038), 1627–1630. https://doi.org/10.1126/science.1721242
921 Wolfe, A., Edlund, M., Sweet, A., & Creighton, S. (2006). A First Account of Organelle
922   Preservation in Eocene Nonmarine Diatoms: Observations and Paleobiological

923   Implications. *PALAIOS, 21*, 298–304. https://doi.org/10.2110/palo.2005.p05-14e

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

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

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

936   Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.
937   A., & Ma, B. (2012). PEAKS DB: De novo sequencing assisted database search for
938   sensitive and accurate peptide identification. *Molecular & Cellular Proteomics: MCP*,
939   *11*(4), M111.010587. https://doi.org/10.1074/mcp.M111.010587

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

944 **TABLES AND FIGURE LEGENDS**

945 *Table 1. 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

946   [a] <1% peptide false discovery rate

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

948   [c] >80% average local residue confidence and protein mapping using PepExplorer at
949   95% similarity and 1% FDR.

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

951 Figure 1. Progression of *Thalassiosira weissflogii* degradation experiment: **(A)** algal and **(B)**
952 bacterial cell counts over the 12 day incubation period; **(C)** chlorophyll and its degradation
953 product, pheophytin; and **(D)** particulate carbon and nitrogen (filtered on GF/F membranes).
954 For more information about algal degradation see Adams et al., 2019.

955 Figure 2. Heatmap showing the number of cellular compartment gene ontology (GO) terms
956 specific to diatom peptides over a 12 day degradation experiment, both in trypsin-digested
957 and naturally digested fractions (digestion only by natural microbial community). Shown as
958 context are peptides identified in two proteomes of *Thalassiosira pseudonana* cultures from
959 (Dyhrman et al., 2012) (in orange). GO terms shown were condensed from a broader set to
960 eliminate redundancy for ease of visualization using REViGO (http://revigo.irb.hr/) and
961 further manually organized into broad categories.

962 Figure 3. **(A-B)** Secondary structure predictions of algal peptides identified in trypsin-
963 digested and naturally digested fractions at four points during the 12 day degradation
964 experiment. Secondary structure motifs (coil, α-helix, *β*-strand, membrane α-helix, membrane
965 *β*-strand) were predicted from full protein sequences using Proteus2 (Montgomerie et al.,
966 2008) and the relative contribution of motifs determined by the identifying peptide's
967 normalized area abundance factor (NAAF). **(C-D)** Secondary structure predictions of just
968 algal chloroplast and membrane peptides and algal cytoplasmic peptides that are identified at
969 each time point of the degradation. For this comparison, proteins identified in the trypsin-
970 digested and naturally digested fractions were combined and NAAF-adjusted.

971 Figure 4. Post-translational modifications (PTMs) of algal peptides at four time points along
972 the 12 day incubation. Shown are the difference between trypsin-digested naturally digested
973 peptides for the 10 variable modifications included in the database searches and de novo
974 sequencing parameters, which were selected from preliminary open modification searches of
975 the mass spectral data (see Methods): **(A)** methionine oxidation, **(B)** asparagine deamidation,
976 **(C)** arginine methylation, **(D)** lysine methylation, **(E)** glutamine cyclization (pyro-
977 glutamation), **(F)** lysine oxidation, **(G)** arginine oxidation, **(H)** proline oxidation, **(I)** tyrosine
978 oxidation, and **J)** lysine acetylation. PTM distributions are expressed as the differences in
979 percent of residue modification in the entire peptide pool (database and de novo sequences)
980 as corrected by normalized area abundance factor (NAAF). Thus, positive values indicate the
981 PTM is relatively enriched in the trypsin-digested peptides, and negative values mean the
982 PTM is relatively enriched in the naturally digested peptides. Bacterial peptides were not
983 included in the PTM analysis. Y-axis scales are the same between columns **(A-E; F-J)**.

984 Figure 5. Mole fractions of individual amino acids from all four degradation time points as
985 derived from total acid hydrolyzable amino acid analysis (THAA, x-axis) and tandem
986 MS/MS-based proteomics (Peptide, y-axis). The dashed 1:1 line represents perfect agreement
987 between approaches. The peptide amino acid compositions plotted here are derived from both
988 trypsin-digested and naturally digested proteomics fractions. Label-free peptide quantitation
989 was determined by the normalized area abundance factor (NAAF).

990 Figure 6. Bacterial progression of the 12-day algal degradation: **(A)** relative abundance
991 contribution of the major bacterial taxonomic classes (>0.5%) from the time point proteomes
992 adjusted by relative peptide spectral abundance (NAAF); **(B)** the percentage of the total
993 number of bacterial peptide biological process gene ontology (GO) terms. GO terms shown
994 were condensed from a broader set to eliminate redundancy for ease of visualization using the
995 REViGO (available http://revigo.irb.hr/) and further manually organized into broader
996 categories.