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

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

1 1

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

37 by using amino acid sequences to infer subcellular location, secondary structures, and post-

38 translational modifications.

INTRODUCTION

42   INTRODUCTION

43   The application of 'omics' to marine systems has accelerated in recent decades and

44 contributed to a deeper understanding of the roles microorganisms play in global

45 biogeochemical cycles (Kleiner, 2019; Saito et al., 2019). Meanwhile, geochemical

46 techniques have shown that some of the largest pools of carbon in the ocean are the detrital

47 remains of these organisms, both in dissolved (Aluwihare et al., 1997; Benner & Kaiser,

48 2003; Jiao et al., 2010) and particulate organic matter (Hedges et al., 2001; Lee et al., 2004;

49 Wakeham et al., 1997). Though largely unexplored, analytical advances are increasing the

50 potential for 'omic' approaches to illuminate carbon cycling in marine detritus (Moore et al.,

51 2012, 2014; Nunn et al., 2010).

52   Much of the identifiable detrital organic matter in the ocean is proteinaceous (Lee et

53 al., 2004; Mayer et al., 1988; Wakeham et al., 1997). This composition may be unsurprising

54 given the abundance of proteins in living organisms: diatoms, one of the major primary

55 producers in the ocean, are approximately 25% protein by dry weight (Olofsson et al., 2019).

56 Photoautotrophic bacteria, such as *Prochlorococcus*, can be as much as 70% protein (Finkel

57 et al., 2016). Determining the origins and fates of protein in marine systems is therefore

58 critical to a predictive understanding of carbon transport and storage in a changing ocean.

59   Proteins are often assumed to be labile and subject to degradation (Nunn et al., 2003;

60 Pantoja et al., 2009). Yet, some proteins have been shown to survive long exposure in

63 seawater (Dong et al., 2010; Keil & Kirchman, 1994) and have been identified in the detrital

64 dissolved organic matter pool (Suzuki et al., 1997; Tanoue et al., 1995; Yamada & Tanoue,

65 2003), in coastal and open ocean marine sediments (Estes et al., 2019; Mayer et al., 1986;

66 Mayer et al., 1988) and in sediment pore water (Abdulla et al., 2018; Fejjar et al., 2021;

67 Schmidt et al., 2011). Algal proteinaceous material has been shown to survive microbial

68 degradation in 4,000 year old sediments (Knicker et al., 1996). Thus, the geochemical

69 evidence implies that while many proteins or constituents of proteins may indeed be labile,

70 some component of proteinaceous material has the potential to be preserved into the geologic

71 record. However, the underlying reasons that certain proteins or their constituents are

72 preserved while others are rapidly degraded remain largely elusive.

73   The aquatic geochemical literature is rich in regards to protein and amino acid

74 degradation. Heterotrophic bacteria first secrete free or tethered endoproteases to break larger

75 proteins to smaller peptides (Hollibaugh & Azam, 1983; Hoppe et al., 2002; Obayashi &

76 Suzuki, 2008). Exopeptidases then cleave individual amino acids off the termini, as shown

77 via mass spectral detection of marine detrital protein (Nunn et al., 2003; Roth & Harvey,

78 2006), and peptides under around 600 Da can be taken up by heterotrophic bacteria and some

79 phytoplankton (Mulholland & Lee, 2009; Weiss et al., 1991). Encapsulation and/or sorption

80 interactions with organic matter and inorganic mineral surfaces may limit bacterial enzyme

81 access to (and degradation of) protein and proteinaceous compounds (Hedges et al., 2001;

82 Knicker et al., 2001; Mayer, 1994; Nagata & Kirchman, 1996).

83   With the application of metaproteomics tools to marine detirial systems, peptides

84 sourced from algal organelles and membranes have emerged as more resistant to rapid

85 degradation in environmental and laboratory studies (Moore et al., 2014; Nunn et al., 2010;

86 Yamada & Tanoue, 2003). Certain secondary structures such as α-helices and random coils

87 appear to be retained within detritus while other motifs are degraded (Nunn et al 2010), and

88 peptides containing modified amino acids have been shown to accumulate in detrital pools of

89 marine organic matter (Duffy et al, submitted, Abdulla et al., 2018; Liu et al., 2010; Yamada

90 & Tanoue, 2003).

91   Previous studies that apply metaproteomic techniques to marine detritus have

92 generally added trypsin or other proteolytic enzymes to break proteins into smaller

93 constituent peptides (Moore et al., 2012, 2014; Nunn et al., 2010). This proteolytic approach

94 is typical in liquid chromatography-mass spectrometry (LC-MS) 'bottom-up' proteomics

95 experiments which analytically require small, ionizable peptides for detection rather than

96 bulky and complex intact proteins (Laskay et al., 2013; Saito et al., 2019). However, this

97 analytical requirement may obscure intriguing dynamics of natural protein breakdown and

98 digestion in the environment. In the present study we complement existing protein

99 degradation studies by additionally looking for small peptides produced solely by microbial

100 degradation during the experiment. This naturally-produced peptide pool represents a

101 combination of endogenous small peptides and those that have been liberated from larger

102 proteins but are not yet fully degraded. We leverage the amino acid sequences gained from

103 metaproteomic techniques to interrogate the subcellular origins, secondary structures,

104 modifications, and amino acid compositions of the degrading algal proteinaceous material

105 over time.

MATERIAL AND METHODS

107   The experimental workflow can be divided into distinct phases: the algal degradation

108   experiment itself; amino acid and peptide extraction; peptide identification via a combination

109   of a database searching and de novo peptide sequencing approaches; and data analytics.

110   *Algal degradation*

111   A culture of the centric diatom *Thalassiosira weissflogii* was grown to approximately

112   1×10^6 cells/mL, concentrated by centrifugation, and frozen at -80 °C to render the algal

113   culture non-viable. Aliquots were thawed and resuspended to 2 g/L by dry weight using 1 µm

114   filtered and UV sterilized seawater from the Damariscotta Estuary, Gulf of Maine.

115   Unsterilized seawater filtered to 1 µm was used as an inoculum to induce bacterial

116   decomposition of the algal detritus, with 1 mL added to 15 L of algal suspension. At each

117   time point (1, 2, 5 & 12-days), 100 mL of the degradation slurry was vacuum filtered onto 25

118   mm diameter 0.3 um pore size glass fiber membranes (Sterlitech GF75) at 4 °C and stored

119   frozen prior to proteomic extraction and analysis. Further experimental details about the

120   degradation can be found in Adams et al. (2019).

121

122   *Total hydrolyzable amino acids*

123   Approximately 30 mg of material scraped from a filter was hydrolyzed in 6 N HCl as

124   described in Cowie & Hedges (1992). Amino acids were derivatized as in Gray et al. (2017)

125   using the AccQ Tag Ultra derivatization kit from Waters (Milford, Massachusetts). Amino

126   acids were separated and quantified via LC-MS using a full scan method in positive ion

127   mode, with a scan range of 100 to 600 m/z and a resolution of 60,000, on a Thermo Scientific

128   Q Exactive HF Orbitrap.

129   *Proteomic sample extraction*

130   Peptides were extracted with and without the use of the serine protease trypsin, which

131   specifically cleaves sequences on the carboxyl side (or the "C-terminal side") of lysine and

132   arginine, except when either is bound to a C-terminal proline. Peptides created via trypsin

133   digestion are usually between 700-1500 daltons (6-14 amino acids), an ideal range for mass

134   spectrometry detection (Laskay et al., 2013). Approximately 100 mg of material was scraped

135   from a filter into detergent-free 50 mM ammonium bicarbonate lysis buffer with 5 mM

136   ethylenediaminetetraacetic acid (EDTA). The chilled suspension was lysed via three cycles

137   each of mechanical disruption with silica beads (50% 100 µm diameter and 50% 400 µm),

138   freeze-thawing in a methanol-dry ice bath, and 30 seconds in a high-power water bath

139   sonicator. The supernatant was removed after centrifugation at 4800 rpm and the protein

140   concentration was determined using a modified Lowry assay using reagents from Bio-Rad

141   (Hercules, California). Extracted protein was subjected to reduction of disulfides with

142   dithiothreitol and carbamidomethylation with iodoacetamide. One half (300 uL) of the

143   protein extraction was subjected to in-solution digestion with trypsin (Promega Gold)

144   overnight at room temperature by a ratio of 1 µg trypsin: 25 µg total protein in 50 mM

145   ammonium bicarbonate buffer with 5 mM EDTA. The remaining 300 uL of protein

146 extraction was left untreated with trypsin. Both trypsin and no-trypsin treatments were

147 desalted using NestGroup macro-spin C18 columns (Southborough, Massachusetts) and

148 resuspended in 5% acetonitrile with 0.1% formic acid and Waters Hi3 E. coli peptide

149 standard mixture (100 fmol/L).

150   Reverse-phase liquid chromatography-high resolution mass spectrometry (LC-

151 HRMS) analysis was performed in duplicate with a Thermo Science (Waltham,

152 Massachusetts) EASY-nanoLC system coupled to a Thermo Orbitrap Fusion Tribrid HRMS.

153 Peptides were separated on a home-packed analytical column consisting of a 37 cm long, 75-

154 µm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100˚A,

155 5mm; Michrom) coupled to a 4 cm long, 100 m i.d. precolumn (Magic C18AQ, 200˚A, 5mm;

156 Michrom). Solvents of 100% LC/MS grade water with 0.1% formic acid (A) and 100%

157 LC/MS grade acetonitrile with 0.1% formic acid (B) were used to elute peptides over a 120-

158 minute gradient from 5-35% solvent B. All analyses were carried out in positive mode at an

159 NSI spray voltage of 2 kV. Data-dependent acquisition (DDA) on the top 10 ions was carried

160 out using first higher energy collision dissociation (HCD) and then electron transfer

161 dissociation (ETD) fragmentation methods for duplicate injections. The MS1 (parent peptide

162 ion) scan range was 400-2000 m/z.

163   *Proteomic data analysis*

164   Peptides were identified from the raw mass spectra using a combination database-

165   driven and database-independent de novo sequencing approach. De novo peptide

166   sequencing, where the amino acid sequence of peptides is determined solely from the mass

167   spectra without comparison to a reference database (Allmer, 2011), was advantageous in this

168   study because we lacked a paired metagenome to thoroughly characterize the microbial

169   community from the seawater inoculum (Muth et al., 2015, 2018; O'Bryon et al., 2020). The

170   combining of database and de novo is termed de novo-directed proteomics, and was

171   performed using PeaksDB within Peaks Studio (v8.5; Bioinformatics Solutions, Waterloo,

172   Canada; Zhang et al. (2012)). The de novo-directed approach has been shown to significantly

173   improve sensitivity and accuracy in comparison to existing database search techniques

174   (Zhang et al., 2012).

175   For database searches we used a reference protein database composed of 84,000

176   sequence entries predicted from transcriptomes of 8 *T. weissflogii* strains contained in the

177   Marine Microbial Metatranscriptome Sequencing Project (NCBI BioProject PRJNA248394,

178   Keeling et al., 2014). We added to the reference database two Gulf of Maine surface seawater

179   metagenomes (Yooseph et al., 2007) from the Global Ocean Survey (GOS) as an aid in

180   identifying the seawater microbes degrading the algal detritus . Additionally, we searched

181   against a database of common mass spectral contaminants (Mellacheruvu et al., 2013).

182   Search parameters for both database searching and de novo sequencing included 8 maximum

183   modifications per peptide, 15 ppm peptide mass tolerance, and 0.5 Da fragment mass

184   tolerance. For the trypsin-digested fractions, we performed both tryptic (maximum 2 missed

185 cleavages) and non-enzymatic constraint searches, which means that all possible peptides up

186 to 60 residues were considered. For the fractions not treated with trypsin, only non-enzymatic

187 constraint searches were performed. Results from technical replicates and fragmentation

188 strategies were combined.

189   Peptide identification confidences are calculated differently between the database

190 search-identified and de novo sequenced peptides. For the de novo identifications, an amino

191 acid-level confidence score is calculated based on mass deviation from spectral features and

192 expressed as a percentage value. We accepted de novo peptides >80% average residue local

193 confidence (ALC) with no single amino acid score <50%. For database searches, a false

194 discovery rate (FDR) was set <1.0% using a reversed database target-decoy strategy (i.e.,

195 searching against reversed reference protein sequences) (Elias & Gygi, 2010). De novo

196 sequencing was also incorporated into the database searches, as PeaksDB first compares de

197 novo sequences to the reference database to find approximate matches and decrease the

198 search space. Agreement between de novo sequences and database search hits are also used,

199 in part, to generate peptide-level confidence scores derived from the p-value indicating the

200 statistical significance of the peptide-spectrum match (the -10lgP score). The threshold was a

201 -10lgP score >20. Such a score is equivalent to a p-value of 1%, signifying the probability

202 that the identification is to a false peptide sequence (Zhang et al., 2012).

203   Protein identifications are notoriously difficult in samples containing many different

204 organisms because some peptides are shared in proteins from multiple taxa. We required a

205 minimum of 1 unique peptide per protein identification. Matching of proteins from peptides

206 found only by de novo sequencing is detailed below.

208   Since amino acids are frequently modified after translation, either for cell-specific

209   purposes or during degradation, a computationally efficient method is needed to search for

210   the myriad possible post translational modifications (PTMs). We used the open modification

211   search tool PeaksPTM (Han et al., 2011), with parameters set to tryptic or non-enzymatic

212   constraint, 2 or fewer missed cleavages, 15 ppm peptide mass tolerance, 0.2 Da fragment

213   mass tolerance, minimum A-score > 200 (a measure of modification location confidence),

214   fixed carbamidomethylation of cysteine, and variable oxidation of methionine. Based on

215   PeaksPTM results, the most frequently occurring modifications were used to evaluate

216   whether adding additional PTMs to the overall searches altered the rate of false discovery.

217   That is, we sought the optimal balance between searching for PTMs while avoiding a vast

218   search space that leads to decreased sensitivity (Noble, 2015; Timmins-Schiffman et al.,

219   2017). A series of PeaksDB searches and Peaks sequencing runs of the combined data set

220   (using same search parameters as above) with increasing numbers of variable modifications

221   was performed to find the optimal set of PTMs to include in searches (most peptide

222   identifications <1% FDR). Using these PTM ramping results, 10 optimal variable PTMs were

223   selected for the final searches. They included, in addition to methionine oxidation:

224   deamidation of asparagine, methylation of arginine, oxidation of tyrosine, methylation of

225   lysine, oxidation of lysine, oxidation of arginine, oxidation of proline, acetylation of lysine,

226   and glutamine cyclization (the conversion of glutamine to pyro-glutamine).

227   *Mapping de novo peptides to proteins*

228   The de novo-directed PeaksDB workflow used here outputs peptides matched to the

229   database and de novo peptides (sequences only, no additional information). To identify

230   peptides that came from the diatom detritus or bacterial proteins not found by database

231   searching, the de novo peptides were aligned to the reference database (*T. weissflogii*

232   transcriptomes with GOS Gulf of Maine metagenomes) using PepExplorer (Leprevost et al.,

233   2014). PepExplorer considers common de novo sequencing errors and limitations (such as

234   leucine and isoleucine equivalence or other combinations of amino acids having the same

235   mass) and identifies regions of local similarity between sequences. We also performed an

236   alignment of the sequences against a reversed version of the database to estimate a false

237   discovery rate, which was kept <1%. Alignments were accepted at a 95% identify agreement

238   cutoff and protein identification required at least one unique peptide alignment.

239   Many de novo peptides did not match back to the reference database, which was not

240   unexpected given the low number of sequences in the database for heterotrophic bacteria and

241   the GOS sampling locations being more open ocean (Yooseph et al., 2007) compared to the

242   estuarine seawater inoculum. To overcome this limitation, I additionally mapped the de novo

243   peptides to proteins contained in the entire UniProt KnowledgeBase database (UniProt

244   Consortium, 2018), which contains over 200 million sequences from thousands of taxa. The

245   mapping was performed using the UniPept lowest common ancestor tool (Mesuere et al.,

246   2016), which is built specifically for metaproteomic data and determines the taxonomic

247   origins of peptides to the lowest possible phylogenetic rank (since some peptides are highly

248   conserved, they may match as to only 'Bacteria', or even 'Organism', rather than to a species

249 or genus level). The UniPept output provides the best view of the bacteria present in the

250 experiment given the lack of genes or transcripts from which to build an ideal reference

251 database.

252   *Gene Ontology Terms and Secondary Structures*

253   To identify gene ontology (GO) term annotations, the peptide sequences were aligned

254 to the UniProt protein database using UniPept's metaproteomic functional analysis tool

255 (Gurdeep Singh et al., 2019), which is built upon a lowest common ancestor peptide search

256 described above. The GO categories were condensed from the broader set in order to

257 eliminate redundancy using the REViGO tool (available at http://revigo.irb.hr/) as well as

258 manually.

259   Protein and protein secondary structures for all samples were estimated using the

260 Proteus2 algorithm (Montgomerie et al., 2008) for the combination of proteins identified by

261 PeaksDB database searching and de novo sequencing with database mapping with

262 PepExplorer. Output is the highest likelihood of individual amino acids being part of the

263 following structure classes: random coil, α-helices,   -strands, membrane α-helices, and

264 membrane   -strands.

265   Spectral files, databases, and peptide identification (pepXML) files have been

266 deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the

267 data set identifier PXD027843.

268   RESULTS AND DISCUSSION

269   *General characteristics of the degradation*

270   As previously described (Adams et al., 2019), algal biomass decreased by a factor of

271   2.4 during the 12-day incubation. Meanwhile, bacterial abundance increased by a factor of 5.

272   Total particulate carbon decreased by 31% over the twelve days (Figure 1). In total, the algal

273   detritus was partly remineralized via respiration and partly converted to bacterial biomass,

274   but mostly remained as necromass (Adams et al., 2019). The degradation experiment started

275   with a small inoculum of bacteria added to a detrital slurry that contained some algal

276   proteases present during diatom cell harvesting. The algal detritus also likely contained small

277   amounts of bacteria (or bacterial necromass) from the diatom culture, which was not

278   confirmed to be axenic. The living bacteria then responded to the detritus and began to grow

279   exponentially (Figure 1), during which time they released enzymes in order to break down

280   and take up the diatom organic matter. The ratio of enzymatically hydrolyzable amino acids

281   (a biomimetic measurement of protein bioavailability, see Mayer et al., 1995) to the organic

282   carbon content remained constant throughout the experiment (Adams et al., 2019) indicating

283   that there was no selective remineralization or preservation of proteins relative to other

284   organic material at this early stage of degradation. However, by looking closely at peptides,

285   we show that numerous compositional changes occurred in the protein pool during the twelve

286   day period.

287   Our interpretation of protein dynamics during the degradation experiment relies on

288   the distinction between two different pool of peptides: those produced in situ during the

289   experiment (not treated with trypsin, which I will refer to as 'naturally digested peptides')

290   and those produced artificially by the laboratory digestion (referred to as 'trypsin-digested

291   peptides'). We hypothesize that the naturally digested peptides represent a mixture of

292   proteinaceous material that is being accessed by the heterotrophic bacteria and peptides that

293   have been released from larger proteins. The trypsin-digested peptides should represent the

294   total of all digestible proteinaceous material present at each time point, including that which

295   is not being actively degraded (e.g. the material that has resisted degradation thus far).

296   Differences between these two diatom protein-derived pools provides information about

297   relative labilities of the substrate proteins, in addition to information about the bacterial

298   strategies of degrading the diatom necromass.

299   Both the trypsin and naturally digested peptide pools should also contain peptides

300   sourced from the bacteria growing in the experiment. Given cell counts and estimated carbon

301   per diatom and bacterial cells, there is about 150 times more algal protein than bacterial

302   protein in the day 0 time point, and this decreases to 15 times more after 12-days (Adams et

303   al., 2019). These peptides can be used to determine the types of bacteria present during the

304   experiment, and the cellular functions that these bacteria are performing (e.g. Bergauer et al.,

305   2017; Mikan et al., 2020). In the following sections I examine the peptide results with the

306   goal of trying to better understand the degradative processes occurring over the course of

307   experiment and their implications for detrital protein cycling in the marine systems.
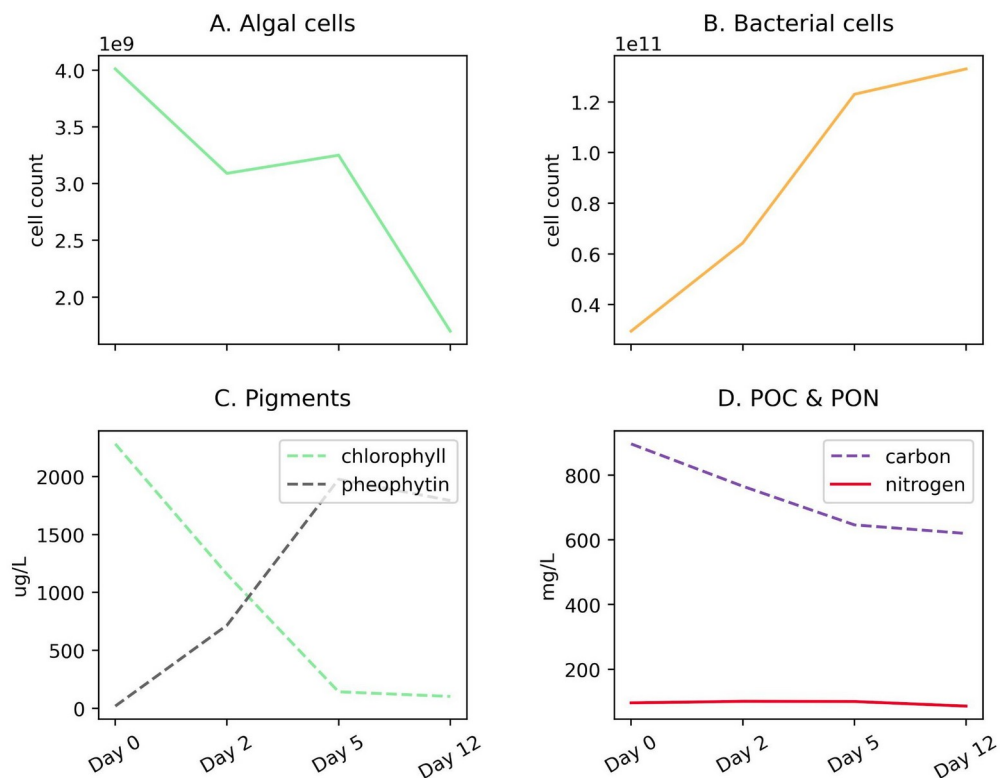
Figure 1. Progression of *Thalassiosira weissflogii* degradation experiment. (A) algal and (B) bacterial cell counts over the 12-day incubation period; (C) chlorophyll and its degradation product, pheophytin; and (D) particulate carbon and nitrogen (filtered on GF/F membranes). While algal cellular material is degrading between days 2 and 5, the silica-based cell walls remain and are identifiable via microscopy, leading to the static or slightly increasing diatom cell counts (A). For more information about algal degradation see Adams et al., 2019.

Algal peptide identification and characteristics

The naturally digested and trypsin-digested peptides are methodologically distinct yet analytically overlapping pools. The naturally digested peptides are present in the background of the trypsin-digested fraction, since the incubation preceded proteomic sample extraction

318 and laboratory digestion. The mass window of MS1 detection is 400-2,000 m/z. Typically,

319 peptides produced via trypsin digestion are between 6-14 amino acids in length; in this data,

320 the average trypsin-digested peptide length was 10.3 amino acids and the average peptide

321 (MS1) mass was 1114 Da. By contrast, the average length of peptides in the naturally

322 digested fractions was 10.1 and the average mass, 1082 Da. Larger polypeptides created by

323 microbial degradation processes, if large enough, would likely have sites available to

324 cleavage by trypsin. Thus there is additional overlap between these two pools because some

325 larger naturally digested peptides are further subjected to trypsin digestion in the trypsin-

326 digested fraction.

327   The number of identified diatom peptides and proteins peaked at day 2 during the

328 initial phase of exponential bacterial growth, and then decreased to lowest values at day 12

329 (Table 1). We sorted the algal peptides by their gene ontology cellular compartment terms

330 (GO Terms; Figure 2; see also (Mikan et al., 2020; Riffle et al., 2017). The data were

331 aggregated into eight major cellular compartments, and the three numerically dominant

332 groups were chloroplast, cell membrane, and cytoplasmic proteins (Figure 2). In the case of

333 membrane-bound subcellular compartments, proteins associated with organellar membranes

334 were counted as from that organelles. For example, thylakoid membrane GO terms were

335 classified under 'Chloroplast', for simplicity. The trypsin-digested day 0 fraction had a

336 cellular component GO term distribution consistent with that of living diatom cells: alongside

337 the degradation experiment's algal GO terms we plot those from peptides obtained from two

338 *Thalassiosira pseudonana* proteomes published by Dyhrman et al., 2012 (Figure 2).

339 Table 1. Diatom and bacterial peptide and protein identifications.

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

340 [a] <1% peptide false discovery rate

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

342 [c] >80% average local residue confidence and protein mapping using PepExplorer at

343 95% similarity and 1% FDR.

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

345   The naturally digested peptides contained far fewer GO term identifications after day

346 2, and by day 12 the only identifications in that fraction were from chloroplast and cell

347 membrane peptides. This indicates that peptides and proteins from other cellular components

348 were effectively degraded during the initial phase of bacterial exponential growth (Figure 1)

349 and were not being actively degraded at day 12 because they were no longer present.

350 Consistent with this membrane recalcitrance hypothesis is that the number of identifications

351 of trypsin-digested peptides sourced from larger proteins also decreased for most cellular

352 compartments other than membranes and the chloroplast (Figure 2). However, there were

353 still some trypsin-digested peptides at day 12 for components from the mitochondria and

354 cytoplasm, suggesting that perhaps cellular location alone does not necessarily determine

355 whether a protein or peptide will survive the initial stages of environmental degradation.
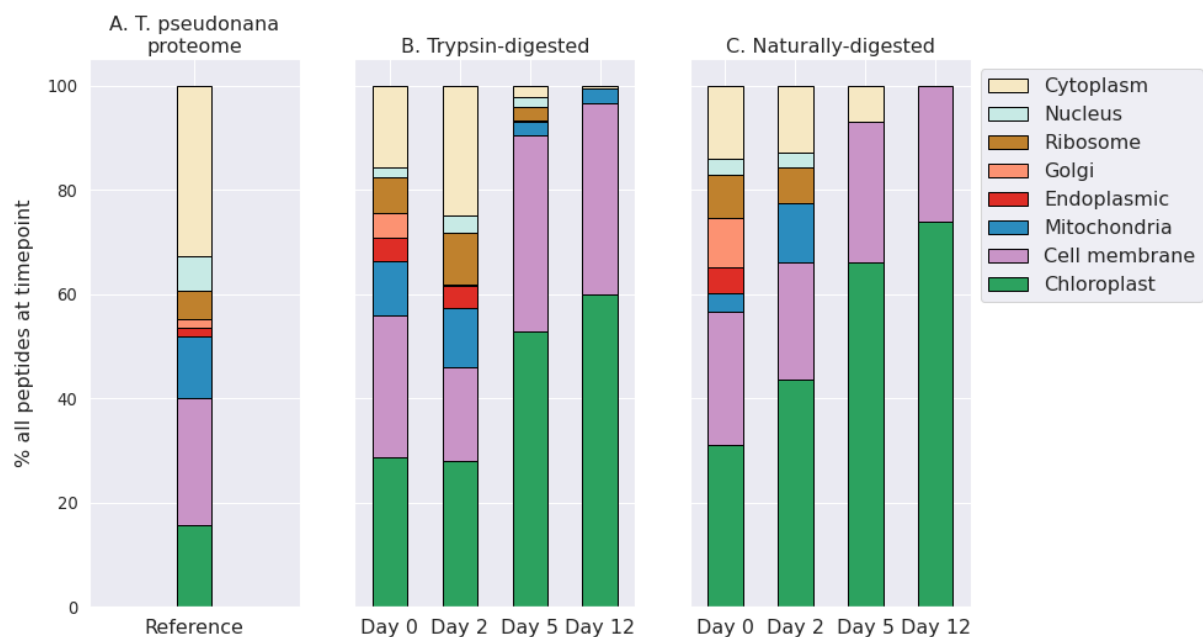
Figure 2. Bar plots showing the distribution of diatom peptides from different cellular compartments over a 12-day degradation experiment, as determined by peptide gene ontology (GO) terms annotations. Peptide distributions are shown as overall percentages of peptides for each timepoint and for the reference proteome. Shown as context are peptides identified in a proteome of *Thalassiosira pseudonana* cultures from (Dyhrman et al., 2012) A). Trypsin-digested (B) and naturally-digested fractions (digestion only by natural microbial community, C) are shown separately. Numbers of peptides at each timepoint and treatment are shown in Table 1. GO terms shown were condensed from a broader set to eliminate redundancy for ease of visualization using REViGO (http://revigo.irb.hr/) and further manually organized into broad categories.

The number of trypsin-digested peptides of cytoplasmic location also dropped precipitously after day 2, while chloroplast and membrane peptides remained abundantly detectable

Manuscript submitted to *Frontiers in Marine Science*

368 throughout the 12-day experiment (Figure 2). This indicates a shift in the detrital protein

369 pool from one that mimics an intact living cell to one that is dominated by proteins that are

370 associated with membranes. These findings are consistent with Laursen et al. (1996), who

371 showed membrane proteins to be the more refractory fraction of phytoplankton protein by

372 physical separation of subcellular fractions of phytoplankton cells. Their study reported

373 higher proteolysis rate constants for the cytoplasmic fraction (>1.2 h-1) than for the

374 membrane fraction (0.2 to 1 h-1), which correlated negatively with the ratio of chlorophyll to

375 pheophytin (Laursen et al., 1996).

376 Studies of algal nutritional value to zooplankton and other animals also suggest

377 preferential soluble protein consumption. In an evaluation of elemental uptake from diatom

378 detritus, Reinfelder and Fisher (1991) showed that metal assimilation efficiencies of marine

379 copepods were directly related to the cytoplasmic content of diatoms. This relationship

380 indicates that the copepods sourced nearly all of their nutrition from the diatom cytoplasm

381 rather than from other cellular constituents (Reinfelder & Fisher, 1991). This experiment

382 didn't involve zooplankton grazing, so these results cannot easily be extended to those

383 settings since larger organisms' gut enzymes and mechanical maceration strategies were not

384 part of our experimental system. However, the selective consumption of cytoplasmic protein

385 components does indicate that membrane protein solubility or structure are factors in

386 differential consumption by bacteria.

387   *Preserved protein motifs*

388   To further interrogate the proteinaceous material that survived the degradation, five

389   classes of peptide secondary structure were evaluated: random coil, α-helices,   -strands,

390   transmembrane α-helices, and transmembrane   -strands. Closely set α-helices contain strong

391   hydrogen bonding between weakly polar (Ser, Thr, Cys) and strongly polar (Asn, Gln, Glu,

392   Asp, His, Arg, Lys) amino acid residues of neighboring α-helices (Liu et al., 2011). Helices

393   are commonly found in transmembrane proteins (Sakai & Tsukihara, 1998; Stevens & Arkin,

394   2000), and their ability to bend can account for the hydrophobic mismatch of the lipid bilayer

395   (Yeagle et al., 2007). In contrast,   -strands lay relatively flat and have been hypothesized in

396   a marine context to adhere to mineral surfaces - ultimately aiding to their protection and

397   resulting in enhanced preservation (Shamblin et al. 1998; Oleschuk et al. 2000; Ovesen et al.

398   2011). ‘Random coil’ is not a true secondary structure, but rather is an aggregate term for

399   short sequences where there is an absence of a helix or sheet character (Smith et al., 1996).

400   Benchmarking of the Proteus2 machine-learning algorithm we use here showed per-residue

401   prediction accuracy to be 87-91% for transmembrane α-helices, 86% for transmembrane   -

402   strands, and 88% for non-membrane secondary structures (Montgomerie et al., 2008).
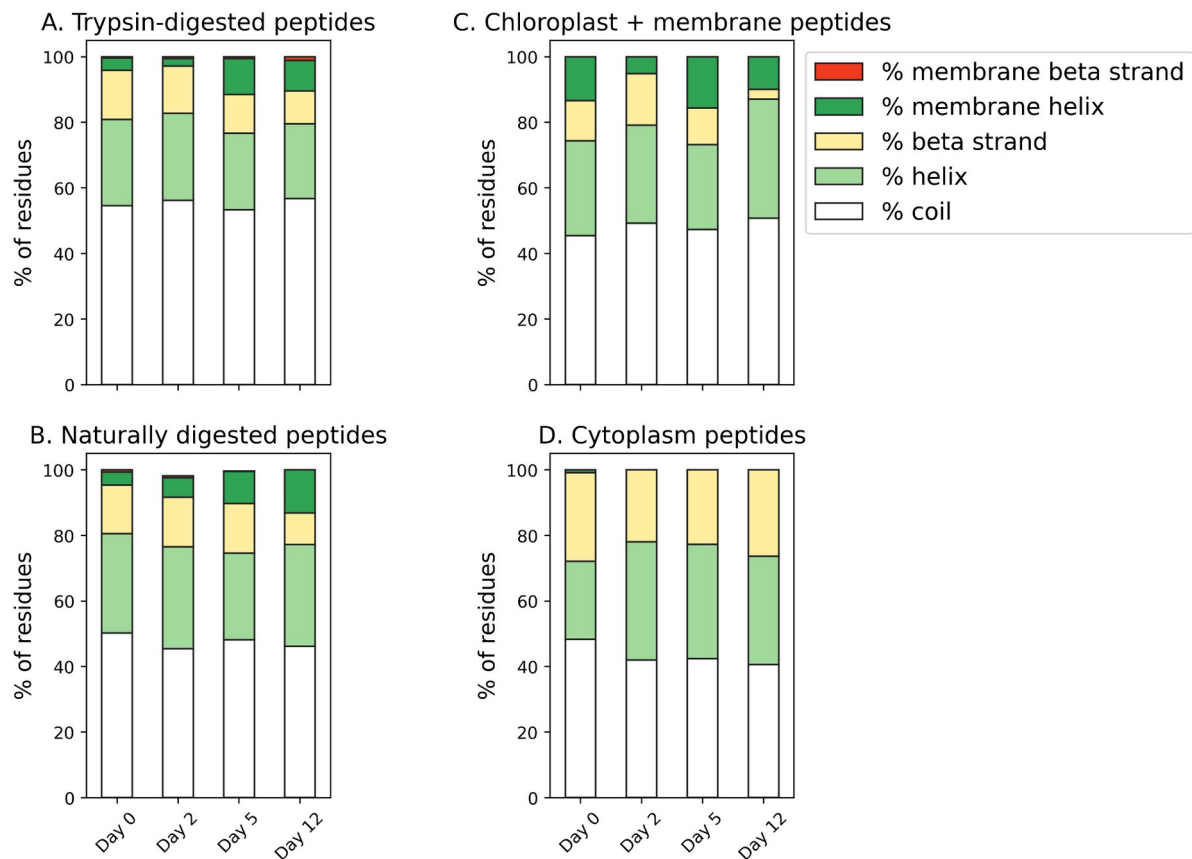
Figure 3. (A-B) Secondary structure predictions of algal peptides identified in trypsin-digested and naturally digested fractions at four points during the 12-day degradation experiment. Secondary structure motifs (coil, α-helix, -strand, membrane α-helix, membrane -strand) were predicted from full protein sequences using Proteus2 (Montgomerie et al., 2008) and the relative contribution of motifs determined by the identifying peptide's relative peptide area abundance. (C-D) Secondary structure predictions of just algal chloroplast and membrane peptides and algal cytoplasmic peptides that are identified at each time point of the degradation. For this comparison, proteins identified in the trypsin-digested and naturally digested fractions were combined and adjusted by relative peptide area abundance.

413   We note a progressive change in secondary structure distribution as the degradation

414   proceeded, with membrane α-helices becoming more important (Figure 3). At the same

415   time,   -strands became a less common motif. This trend in secondary structure distributions

416   is consistent with the GO term evidence indicating that membrane proteins are preferentially

417   retained in the system and suggests that their tightly wrapped, difficult to denature, secondary

418   structure could be a factor aiding in preservation. Random coils are the most common motif,

419   and they do not change in relative importance over time, indicating that they are not

420   particularly prone to resistance or degradation in this experiment, or that the coil category is

421   too broad to capture any selective processes. These results are generally consistent with those

422   of Nunn et al. (2010) in a similar degradation of a marine diatom. They identified 23 and 4

423   diatom peptides after 10 and 23 days of incubation with a seawater microbiome, respectively.

424   Of the surviving 4 diatom peptides at the end of the experiment, 2 had transmembrane

425   domains (Nunn et al., 2010).

426   To further investigate the theory that secondary structure is related to enhanced

427   preservation of membrane protein components, we also compared the predicted secondary

428   structure of just algal chloroplast/integral component of membrane peptides and algal

429   cytoplasmic peptides that are identified at each time point of the degradation (Figure 3).

430   Chloroplast and membrane peptides, the bulk of diatom peptide identifications on day 12

431   (from 167 individual proteins across both trypsin and naturally digested fractions), were

432   50.7% coil, 36.3% α-helix, 3.02%   -strand, 9.93% transmembrane α-helix, and 0%

433   membrane   -strand. In comparison, the cytoplasmic peptides (from 98 proteins across both

434   trypsin and naturally digested fractions) identified in the initial day 0 proteome were

435 comparatively more enriched in -strands and depleted in α-helices and transmembrane α-

436 helices, with 48.3% coil, 23.8% α-helix, 27.0% -strand, 0% transmembrane α-helix, and 0%

437 transmembrane -strand. Supporting the theory that α-helices are linked to preservation is

438 that the few surviving cytoplasmic peptides on day 12 (3 proteins total, all in the trypsin-

439 digested fraction) had 10% more α-helix character than the initial day 0 cytoplasmic peptides,

440 at 40.6% coil, 33.1% α-helix, 26.3% -strand, 0% transmembrane α-helix, and 0%

441 transmembrane -strand (Figure 3). These findings are consistent with the notion that

442 hydrophobic interactions appear to be important in preserving membrane-associated proteins

443 and peptides during early diagenesis (Nguyen & Harvey, 2001; Nunn et al., 2010). Together,

444 peptide subcellular localization annotations and secondary structure analyses connect

445 previous observations of membrane-associated protein selective preservation for bacteria

446 (Jiao et al., 2010; Kaiser & Benner, 2008; Nagata et al., 1998) and phytoplankton (Laursen et

447 al., 1996; Nguyen & Harvey, 2001; Wolfe et al., 2006) to molecular characteristics like 3-

448 dimensional shape and hydrophobicity.

449 While subcellular location and secondary structure may be significant components

450 that allow certain proteins to resist the early stages of degradation, two other factors have

451 been observed in the literature; an abundance of post-translationally modified amino acids in

452 degradation-resistant material and the enrichment of certain amino acids, like glycine, in the

453 recalcitrant proteinaceous pool (by measurement of total hydrolyzable amino acids).

454   *Post translational modifications of amino acids*

455   Protein post-translational modifications (PTMs) such as oxidation, phosphorylation,

456   and methylation, play critical roles in a diverse range of biological processes like signaling,

457   protein activity and transport, and regulation of gene expression (Cain et al., 2014; Shen et

458   al., 2008). PTMs are also associated with in vivo protein recycling and cell senescence (Cain

459   et al., 2014; Dhillon & Denu, 2017). For these purposes, the 'PTM' umbrella includes

460   modifications to amino acids that could occur in detritus, including those due to protein

461   consumption or abiotic transformations.

462   In the marine environment, PTMs have been linked to degradation and early

463   diagenesis. For example, the amino acid beta alanine accumulates in the hydrolyzable phase

464   of marine sedimentary organic matter via modification of aspartic acid (Cowie and Hedges

465   1994). Modification of the nitrogen-containing side chains of glutamine, asparagine and

466   arginine can lead to the accumulation of peptides containing deaminated side chains within

467   anoxic marine sediment pore waters (Abdulla et al., 2018). We recently observed deamidated

468   peptides in the sinking POM from the eastern tropical North Pacific oxygen deficient zone

469   (Duffy et al, submitted), where Van Mooy et al. (2002) hypothesized that amino acids could

470   be selectively deaminated in order to provide reduced nitrogen to fuel chemoautotrophic

471   processes (Van Mooy et al., 2002). In laboratory experiments, Keil and Kirchman showed

472   that methylated peptides were accessed less efficiently by bacteria than non-methylated

473   peptides (Keil & Kirchman, 1992). Glycan modifications of proteins contribute to stability, a

474   mechanism that's been used to increase the shelf life of protein and peptide-based

475   pharmaceuticals (Zhou & Qiu, 2019). Indeed, Keil and Kirchman showed that glycosylated

476   ribulose-1,5-bisphosphate (RuBisCo) was degraded 100 times more slowly than its

477   unmodified counterpart (Keil & Kirchman, 1993). All together, the marine literature

478   suggests that PTMs are highly relevant to protein degradation, as the by-products of protein

479   degradation processes and/or as potential factors in degradation resistance.

480   In the present study, we first performed a non-discriminatory 'open' search of

481   thousands of possible PTMs in the UniMod database using PeaksPTM (see above in

482   Methods). We then selected only the 10 most frequently occurring mass modifications in the

483   open search results for the suite of PTMs in the ultimate database and de novo peptide

484   searches. Our goal was to identify patterns in PTM distribution to learn more about their roles

485   in degradation and preservation. To that end, we wanted to distinguish between PTMs present

486   in algal protein prior to degradation and PTMs associated with the degradative process. We

487   compared the PTMs observed in the naturally digested peptide pool, which should represent

488   proteins being actively degraded by the bacteria, against the trypsin-digested peptides. This

489   comparison was done for only the algal protein; no bacterial proteins were included in this

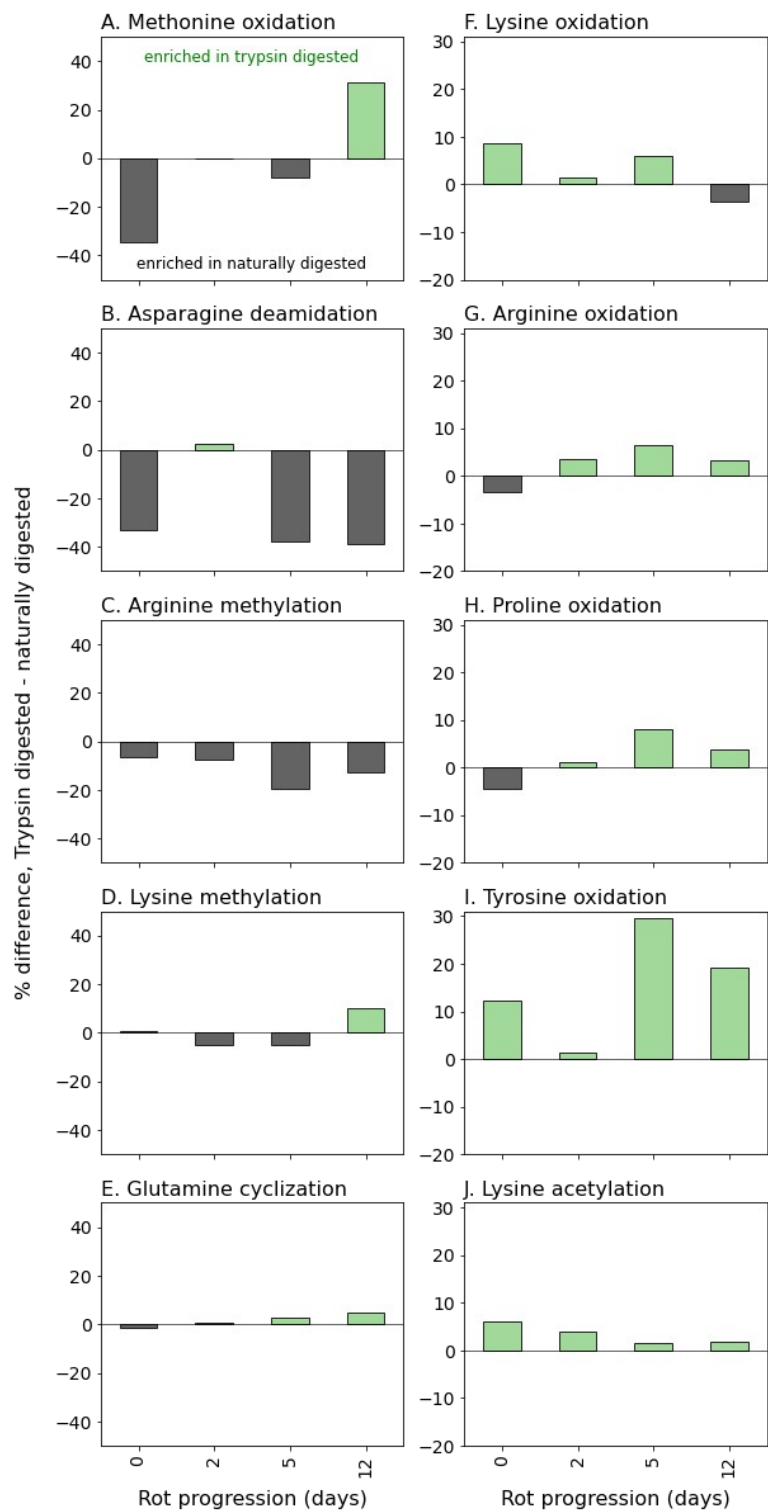490   differential analysis of PTMs (Figure 4).

A. Methonine oxidation

B. Asparagine deamidation

C. Arginine methylation

D. Lysine methylation

E. Glutamine cyclization

F. Lysine oxidation

G. Arginine oxidation

H. Proline oxidation

I. Tyrosine oxidation

J. Lysine acetylation

enriched in trypsin digested

enriched in naturally digested

% difference, Trypsin digested - naturally digested

Rot progression (days)

491   Figure 4. Post-translational modifications (PTMs) of algal peptides at four time points along

492   the 12-day incubation. Shown are the difference between trypsin-digested naturally digested

493   peptides for the 10 variable modifications included in the database searches and de novo

494   sequencing parameters, which were selected from preliminary open modification searches of

495   the mass spectral data (see Methods): (A) methionine oxidation, (B) asparagine deamidation,

496   (C) arginine methylation, (D) lysine methylation, (E) glutamine cyclization (pyro-

497   glutamation), (F) lysine oxidation, (G) arginine oxidation, (H) proline oxidation, (I) tyrosine

498   oxidation, and J) lysine acetylation. PTM distributions are expressed as the differences in

499   percent of residue modification in the entire peptide pool (database and de novo sequences)

500   as corrected by relative peptide area abundance. Thus, positive values indicate the PTM is

501   relatively enriched in the trypsin-digested peptides, and negative values mean the PTM is

502   relatively enriched in the naturally digested peptides. Bacterial peptides were not included in

503   the PTM analysis. Y-axis scales are the same between columns (A-E; F-J).

504   Oxidation PTMs were generally enriched in the trypsin-digested component, implying

505   that these modifications occurred within the cell prior to heterotrophic attack. Generally,

506   they did not increase in relative abundance over time in the trypsin-digested peptide pool,

507   suggesting that oxidations don't meaningfully impact the lability of proteinaceous material

508   (Figure 4). It's notable that during the experiment, chloroplast protein-derived peptides

509   increasingly became dominant among identifications (Figure 2). The algal culture was

510   exposed to light before the degradation experiment in the dark. Photosynthesis as an

511   oxygenic process produces active oxygen species and radicals which can cause damage to

512   cells. Oxidations of amino acids are frequent PTMs in photosynthesis-associated proteins

513 (Aro et al., 2005; Galetskiy et al., 2011). Similarly, lysine acetylation of chloroplast proteins

514 has been demonstrated in plants (Lehtimäki et al., 2015) and in diatoms (X.-H. Chen et al.,

515 2018), though the PTM frequently occurs elsewhere in the cell (Z. Chen et al., 2018). The

516 abundance of thylakoid membrane and chloroplast-associated proteins that accumulate due to

517 preferential preservation (Figure 2) is likely why oxidations and lysine acetylations are

518 dominant PTMs of the trypsin-digested peptides (Figure 4).

519   In contrast, two PTMs were strongly associated with the presumably more detrital

520 naturally digested peptides: deamidation of asparagine and arginine methylation (Figure 4).

521 This observation that asparagines and arginines are more modified in detrital proteins has

522 several possible explanations. These modifications could have occurred within the living

523 cell, resulting in a pool of protein that was easily accessible, which is why they were found

524 more in the naturally degraded peptide pool. Alternatively, these PTMs might have been

525 created during the degradation process and could accumulate because once created they are

526 further degraded slower than their unmodified counterparts. This latter explanation would

527 account for the effective accumulation of deamidated peptides in both sediment pore waters

528 (Abdulla et al., 2018) and in sinking particulate matter (Duffy et al, submitted). Thus, we

529 hypothesize that deamination occurs during degradation, though continued research is

530 warranted.

531   Unlike deamidation, methylation most likely occurs within the living cell, where it is

532 a common PTM that is used as a control of numerous cellular functions (Ghesquière et al.,

533 2011). Methylated peptides have been shown to be inefficiently assimilated and degraded by

534 heterotrophic bacteria (Keil & Kirchman, 1992), suggesting that PTMs produced within

535 living cells may also play a role in determining the rate of protein degradation.

536   Our open PTM searches did not yield high levels of protein glycosylation, a very

537 broad classification of modification that entails a sugar-amino acid bond. Though

538 glycosylation is an extremely common enzymatic PTM in both eukaryotes and prokaryotes,

539 the non-enzymatic form of the reaction (sometimes called 'glycation', or the Maillard

540 reaction) has been posited as a mechanism of environmental protein preservation via

541 geopolymer formation (Collins 1992, Burdige 2007). Glycosylation has been shown to

542 decrease microbial protein degradation rates in incubations (Keil & Kirchman, 1993) and has

543 been found in high abundance in seawater (Yamada & Tanoue, 2003). Peptides with

544 glycosylations are difficult to ionize and detect under the mass spectral conditions we used

545 here (Alley et al., 2013), and to evaluate them more accurately would require different

546 proteomics preparation techniques (e.g., as performed by Moore et al., 2014). Given the

547 findings here and in the literature, more targeted work is needed to better understand the

548 effect that PTMs have during the early stages of protein degradation.

549 We examined the PTM distribution across peptides from different subcellular

550 compartments (Supporting Information Figure 1), and found that 1) most modified peptides

551 were from cell membrane and cytoplasm proteins, and 2) the relative distribution of PTMs

552 across the subcellular compartments did not change significantly over the 12-day

553 degradation, even as the overall distribution of peptides from the subcellular locations did

554 change to become more membrane- and chloroplast-dominated (Figure 2). This result

555 suggests that mechanisms other than modification status may be more important to overall

556   preferential degradation pattern in this system, though the enrichment of certain PTMs in the

557   trypsin-digested peptides or naturally-digested peptides (Figure 4) points to subler

558   preferential degradation patterns amongst modified peptides.

559   *Amino acid compositions*

560   One of the deepest sets of literature related to protein degradation in marine systems is

561   that of the 'total hydrolyzable amino acid' pool (Dauwe et al., 1999; Dauwe & Middelburg,

562   1998; Lee et al., 2004; Wakeham et al., 1997). THAA analyses show clear trends during

563   long-term carbon degradation and preservation including an accumulation of the amino acids

564   glycine, serine and threonine (Dauwe and Middelburg 1998), and the creation of the non-

565   protein amino acids beta-alanine and gamma-aminobutyric acid from aspartic and glutamic

566   acids (Cowie and Hedges 1994). Despite the widespread use of degradation indices derived

567   from THAA analyses, it has been difficult to reconcile changes in bulk amino acid

568   compositions to known protein amino acid compositions, especially during the very early

569   stages of degradation (Keil et al., 2000). We compared the amino acid composition of the

570   peptides identified during the degradation experiment to the THAA pool, which I measured

571   independently. To facilitate this comparison, We plotted mole fractions of amino acids in the

572   THAA against those in the identified algal peptides, combining the naturally and trypsin-

573   digested amino acid compositions adjusted for relative peptide abundance (Figure 5). The

574   near 1:1 agreement between the two approaches indicates two things: a) the protein amino

575   acid compositions measured using the newer 'omic' approach can be effectively integrated

576   into the large body of literature based on THAA analysis, which will become useful as

577   peptidomic approaches are applied to samples further along the degradation pathway (e.g.

578   sediment samples); and b) while the omic approach identifies specific ways in which protein

579   undergoes degradation in the ocean, the early stages remain remarkably 'nonselective' at the

580   bulk molecular level (Hedges et al., 2001).
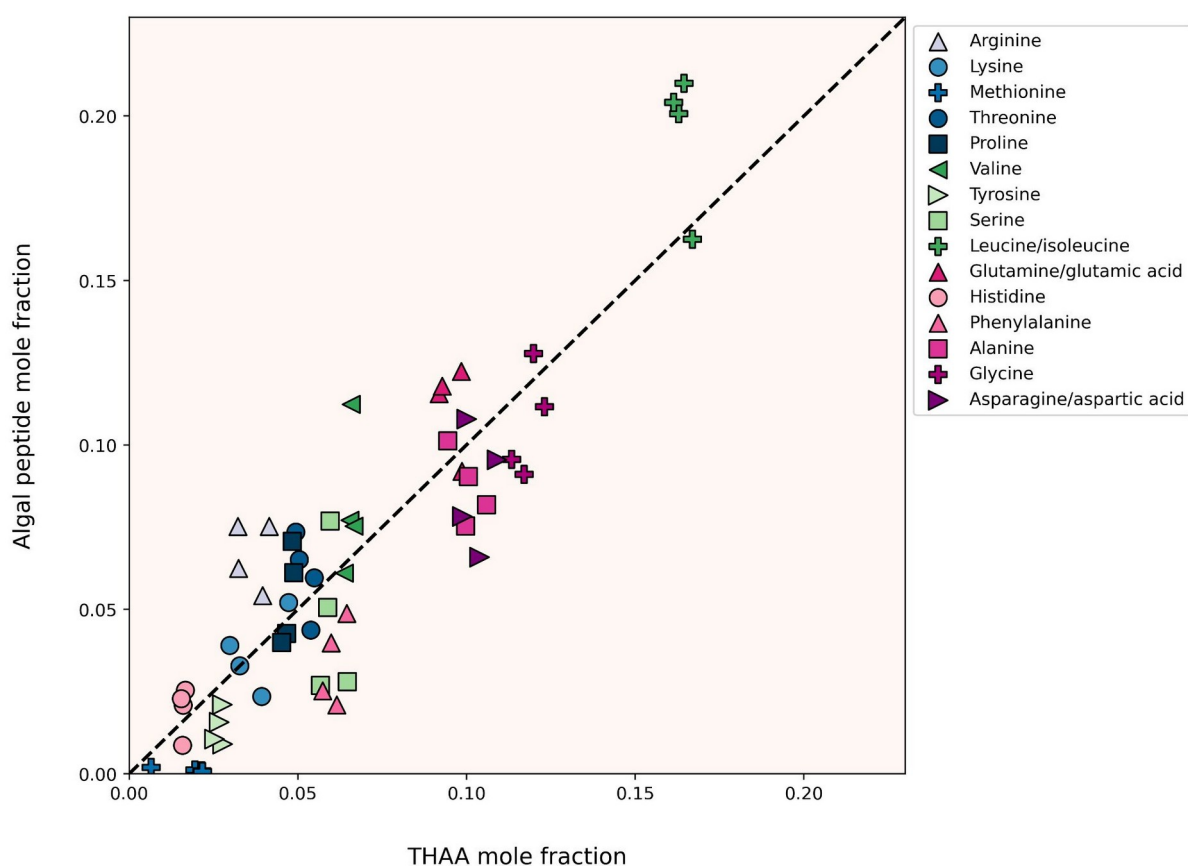
581   Figure 3.5. Mole fractions of individual amino acids from all four degradation time points as

582   derived from total acid hydrolyzable amino acid analysis (THAA, x-axis) and tandem

583   MS/MS-based proteomics (Peptide, y-axis). The dashed 1:1 line represents perfect agreement

584   between approaches. The peptide amino acid compositions plotted here are derived from both

585 trypsin-digested and naturally digested proteomics fractions. Label-free peptide quantitation

586 was determined by the relative peptide area abundance.

587 *Bacterial community*

588 The microbial inoculum for the degradation experiment was sourced from the

589 Damariscotta River Estuary, whose marine waters come from the Gulf of Maine.

590 Heterotrophic bacteria grew exponentially through the middle stages of the experiment

591 (Figure 1). A peptide-based lowest common ancestor analysis was performed using UniPept

592 (Gurdeep Singh et al., 2019) to assign bacterial taxa to the lowest phylogenetic level possible.

593 The taxonomic hits were then adjusted to account for peptide spectral abundance and

594 aggregated at the class level (Figure 6a). There is precedence for using label-free

595 metaproteomic data for microbial biomass determinations (Kleiner et al., 2017), and here we

596 use the approach to learn about the community of microbes degrading the algal detritus.

597 There is taxonomic overlap of the initial composition in this study to that of a

598 pyrosequencing survey of planktonic microbes at three stations in the Gulf of Maine (Li et

599 al., 2011), though it is worth noting that no estuarine sites were included in that work. We

600 found the microbial community at the initial time point was dominated by

601 Gammaproteobacteria (~60% of the peptides), with notable contributions of Cytophaga

602 (~5%), Bacteroidia (~2%) and Alphaproteobacteria (~2%) (Figure 6a). While the Li et al.

603 (2011) survey found clades of SAR11 to dominate surface seawater in the open waters of the

604 Gulf of Maine, some Gammaproteobacteria are symbionts of diatoms (Amin et al., 2012).

605 Since we didn't sample the seawater itself for peptidomics, just the filtered diatom rot slurry,

606 the dominance of a diatom symbiont makes sense. During the 12-day experiment, the

607 microbial community composition changed minimally, with two exceptions: Firstly, there

608 was a ten-fold increase of peptides that were bacterial but could not be uniquely identified at

609 the class level or below, from 2.2% of peptides on day 0 to 22% of peptides by day 12

610 (Figure 6a). At the same time, the contribution of low abundance (<0.5% of all peptides)

611 classes increased from 0.53% on day 0 to 5.4% on day 12 (Figure 6a). Both increases, in non-

612 specific bacterial peptides and in low-abundance class peptides, are potential indications that

613 the diversity of the bacterial community was increasing during the experiment.
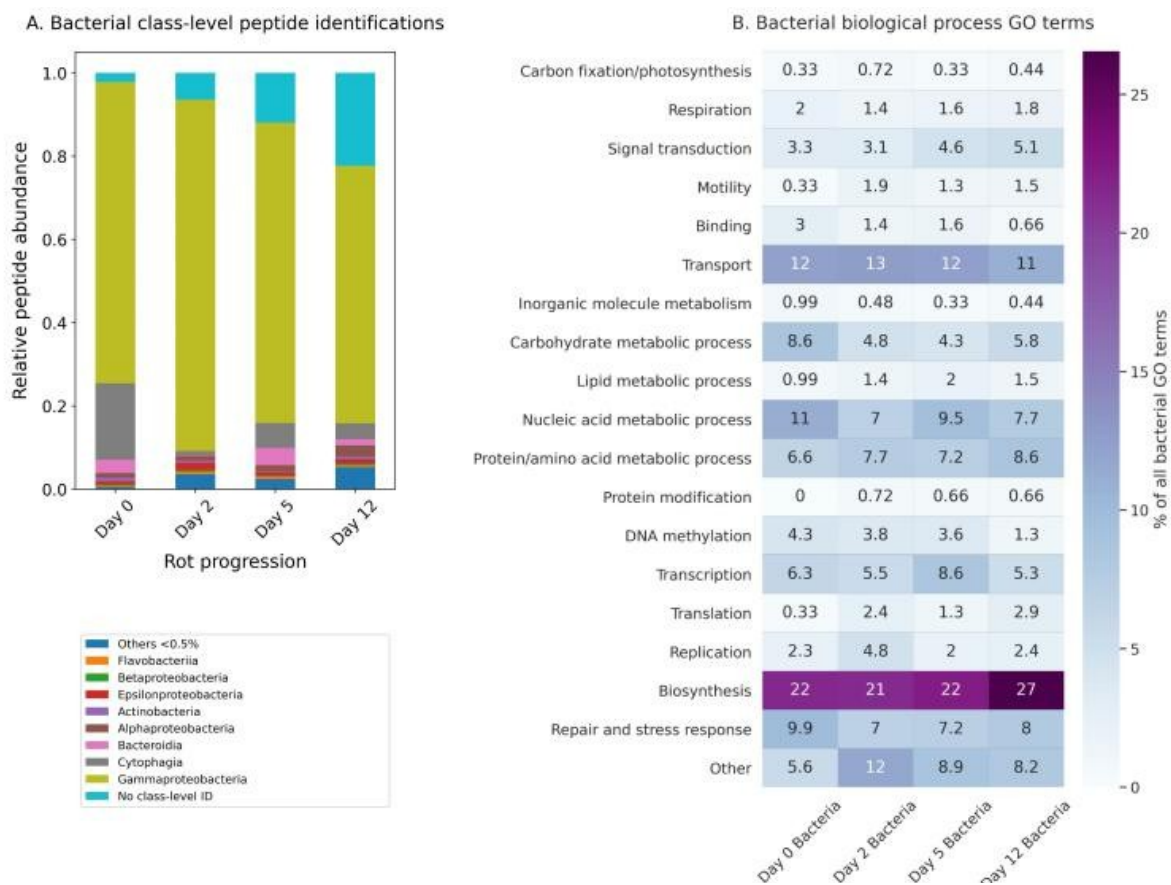
614   Figure 3.6. Bacterial progression of the 12-day algal degradation: (A) relative abundance

615   contribution of the major bacterial taxonomic classes (>0.5%) from the time point proteomes

616   adjusted by relative peptide area abundance; (B) the percentage of the total number of

617   bacterial peptide biological process gene ontology (GO) terms. GO terms shown were

618   condensed from a broader set to eliminate redundancy for ease of visualization using the

619   REViGO (available http://revigo.irb.hr/) and further manually organized into broader

620   categories.

621   The increase of non-specific bacterial peptides can be explained in two ways. Most

622   likely, a true increase in diversity and richness resulted from a broader array of shared,

623   conserved proteins being produced in the system by related bacteria. An alternative

624   hypothesis is that there was an ingrowth of bacterial necromass generating degraded bacterial

625   peptides that are no longer very taxonomically specific. Bacterial necromass in this case

626   could be from organisms in the seawater inoculum or bacteria present during the diatom

627   culture growth. For instance, Gammaproteobacteria are known to thrive within the

628   phycosphere of diatoms (Amin et al., 2012), and their peptide abundance peaked at day 2

629   when the active bacteria were growing exponentially. It is difficult to discern which of these

630   hypotheses is correct and further work will need to be done to evaluate the processing of

631   bacterial detritus by other bacteria. Nonetheless, the overall taxonomic changes observed in

632   the bacterial peptidome during the twelve days were minimal.

633   Unlike the somewhat ambiguous taxonomic information within the bacterial

634   peptidome, biological process GO terms associated with the bacterial peptides provide a

635   clearer view of how the bacterial community responded to the algal detritus (Figure 6b). The

636 most detected GO terms are associated with transmembrane transport, carbohydrate

637 metabolism, and DNA replication and transcription (Figure 6b). These are the terms most

638 strongly associated with bacterial growth (Mikan et al 2020) and suggest that most of the

639 bacterial peptides detected are from living bacteria. This strengthens support for the

640 hypothesis that the Gammaproteobacteria in the samples are for the most part living and not

641 detrital.

642   Our bacterial functional data are generally consistent with recent work by Mikan et al.

643 (2020), who used metaproteomic tools and a GO-term based functional analysis to evaluate

644 the heterotrophic bacterial response to a pulse of detrital organic matter in two Arctic

645 microbiomes during 10-day shipboard incubations. In that study, the bacterial community

646 increased protein synthesis, carbohydrate degradation, and cellular redox processes while

647 simultaneously decreasing C1 metabolism (Mikan et al., 2020). Throughout this experiment

648 we observe steady levels of transmembrane transport and protein metabolism terms.

649 Carbohydrate metabolic process GO terms maximize in the first two days of the degradation,

650 with increasing biosynthesis GO terms by day 12 (Figure 6b). Mikan et al (2020) suggested

651 that the bacterial community shifted their carbon acquisition strategies intracellularly before

652 there were large shifts in the taxonomic structure of the community. Without a paired

653 metagenome or metatranscriptome with which to perform proteomic database searches, these

654 bacterial peptide data and paired GO term data are not as complete, but show the same

655 general trends, lending further support to the notion of Mikan et al. that functional

656 composition and redundancy, not taxonomy, may be the most relevant factor when evaluating

657 how effectively organic matter is or will be processed by bacteria in the ocean.

CONCLUSIONS AND FUTURE DIRECTIONS

659   In this study we evaluated proteinaceous material in a marine system both with and

660   without the use of trypsin as an extraction and identification tool. To our knowledge, this is

661   the first such attempt to disentangle proteins that are being degraded by microbes from those

662   that are resistant to degradation. We show that proteins from specific cellular locations are

663   preferentially preserved during the initial stages of degradation. As has been hypothesized

664   and demonstrated for bacterial membrane proteins (Jiao et al., 2010; Kaiser & Benner, 2008;

665   Nagata et al., 1998) and algal membrane proteins (Laursen et al., 1996; Nguyen & Harvey,

666   2001; Wolfe et al., 2006), we conclusively illustrate that proteins associated with diatom

667   chloroplasts and membranes resist initial degradation better than those without such

668   association. The many membrane-associated and few cytoplasm peptides that resist

669   degradation also are relatively enriched in α-helices and depleted in   -strands, consistent

670   with the cellular location data since α-helices are enriched in membrane proteins. However,

671   the extent to which α-helices themselves lead to degradation resistance remains to be more

672   thoroughly evaluated, as 1) α-helices are not exclusive to membrane proteins and 2) there

673   could be other reasons for membrane protein survival over time that causes membrane

674   proteins' α-helix rich motifs to become enriched in detrital material.

675   The novel application of proteomics without the use of trypsin also allowed for the

676   evaluation of how post translational modifications (PTMs) relate to protein degradation. We

677   found that the oxidation and acetylation PTMs observed likely originated within the living

678   cell, with only asparagine deamidation and arginine methylation being predominantly

679 associated with the degraded peptide pool. We hypothesize that PTMs have an impact on the

680 bioavailability of peptides during early diagenesis, but again more work is needed to evaluate

681 the extent to which PTMs provide protection. In all, we add to earlier evidence of selective

682 protein degradation mechanisms enabled by proteomic techniques (Bridoux et al., 2015;

683 Dong et al., 2010; Moore et al., 2012, 2014; Nunn et al., 2010). Continued advancements in

684 metaproteomic instrumentation and computational capabilities have great potential to better

685 our understanding of protein degradation and preservation dynamics in the ocean.

692   DATA AVAILABILITY

693 Peptidomic raw and processed data files, search databases, and search parameter files have

694 been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with

695 the data set identifier PXD027843 to be made available with publication of this chapter's

696 corresponding manuscript (see above).

697   REFERENCES

698   Abdulla, H. A., Burdige, D. J., & Komada, T. (2018). Accumulation of deaminated peptides

699   in anoxic sediments of Santa Barbara Basin. Geochimica et Cosmochimica Acta, 223, 245–

700   258. https://doi.org/10.1016/j.gca.2017.11.021

701   Adams, C., Mayer, L., Rawson, P., Brady, D., & Newell, C. (2019). Detrital protein

702   contributes to oyster nutrition and growth in the Damariscotta estuary, Maine, USA.

703   *Aquaculture Environment Interactions*, *11*, 521–536. https://doi.org/10.3354/aei00330

704   Alley, W. R., Mann, B. F., & Novotny, M. V. (2013). High-sensitivity Analytical Approaches

705   for the Structural Characterization of Glycoproteins. *Chemical Reviews*, *113*(4),

706   2668–2732. https://doi.org/10.1021/cr3003714

707   Aluwihare, L. I., Repeta, D. J., & Chen, R. F. (1997). A major biopolymeric component to

708   dissolved organic carbon in surface sea water. *Nature*, *387*(6629), 166–169.

709   https://doi.org/10.1038/387166a0

710   Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and

711   Bacteria. *Microbiology and Molecular Biology Reviews*, *76*(3), 667–684.

712   https://doi.org/10.1128/MMBR.00007-12

713   Aro, E.-M., Suorsa, M., Rokka, A., Allahverdiyeva, Y., Paakkarinen, V., Saleem, A.,

714   Battchikova, N., & Rintamäki, E. (2005). Dynamics of photosystem II: A proteomic

715   approach to thylakoid protein complexes. *Journal of Experimental Botany*, *56*(411),

716   347–356. https://doi.org/10.1093/jxb/eri041

717   Benner, R., & Kaiser, K. (2003). Abundance of amino sugars and peptidoglycan in marine

718   particulate and dissolved organic matter. *Limnology and Oceanography*, *48*(1), 118–

719   128. https://doi.org/10.4319/lo.2003.48.1.0118

720 Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R.,

721   Pachiadaki, M. G., Jensen, O. N., & Herndl, G. J. (2017). Organic matter processing

722   by microbial communities throughout the Atlantic water column as revealed by

723   metaproteomics. *Proceedings of the National Academy of Sciences*, 201708779.

724   https://doi.org/10.1073/pnas.1708779115

725 Bridoux, M., Neibauer, J., Ingalls, A., Nunn, B., & Keil, R. (2015). Suspended marine

726   particulate proteins in coastal and oligotrophic waters. *Journal of Marine Systems*,

727   *143*. https://doi.org/10.1016/j.jmarsys.2014.10.014

728 Cain, J. A., Solis, N., & Cordwell, S. J. (2014). Beyond gene expression: The impact of

729   protein post-translational modifications in bacteria. *Journal of Proteomics*, *97*, 265–

730   286. https://doi.org/10.1016/j.jprot.2013.08.012

731 Chen, X.-H., Li, Y.-Y., Zhang, H., Liu, J.-L., Xie, Z.-X., Lin, L., & Wang, D.-Z. (2018).

732   Quantitative Proteomics Reveals Common and Specific Responses of a Marine

733   Diatom Thalassiosira pseudonana to Different Macronutrient Deficiencies. *Frontiers*

734   *in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02761

735 Chen, Z., Luo, L., Chen, R., Hu, H., Pan, Y., Jiang, H., Wan, X., Jin, H., & Gong, Y. (2018).

736   Acetylome Profiling Reveals Extensive Lysine Acetylation of the Fatty Acid

737   Metabolism Pathway in the Diatom Phaeodactylum tricornutum. *Molecular &*

738   *Cellular Proteomics: MCP*, *17*(3), 399–412.

739   https://doi.org/10.1074/mcp.RA117.000339

740 Cowie, G. L., & Hedges, J. I. (1992). Improved amino acid quantification in environmental

741   samples: Charge-matched recovery standards and reduced analysis time. *Marine*

742   *Chemistry*, *37*(3), 223–238. https://doi.org/10.1016/0304-4203(92)90079-P

743 Dauwe, B., & Middelburg, J. J. (1998). Amino acids and hexosamines as indicators of

744   organic matter degradation state in North Sea sediments. *Limnology and*

745   *Oceanography*, *43*(5), 782–798. https://doi.org/10.4319/lo.1998.43.5.0782

746 Dauwe, B., Middelburg, J. J., Herman, P. M. J., & Heip, C. H. R. (1999). Linking diagenetic

747   alteration of amino acids and bulk organic matter reactivity. *Limnology and*

748   *Oceanography*, *44*(7), 1809–1814. https://doi.org/10.4319/lo.1999.44.7.1809

749 Dhillon, R. S., & Denu, J. M. (2017). Using comparative biology to understand how aging

750   affects mitochondrial metabolism. *Molecular and Cellular Endocrinology*, *455*, 54–

751   61. https://doi.org/10.1016/j.mce.2016.12.020

752 Dong, H.-P., Wang, D.-Z., Dai, M., & Hong, H.-S. (2010). Characterization of particulate

753   organic matter in the water column of the South China Sea using a shotgun proteomic

754   approach. *Limnology and Oceanography*, *55*(4), 1565–1578.

755   https://doi.org/10.4319/lo.2010.55.4.1565

756 Dyhrman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. I., Alexander,

757   H., Whitney, L. P., Drzewianowski, A., Bulygin, V. V., Bertrand, E. M., Wu, Z.,

758   Benitez-Nelson, C., & Heithoff, A. (2012). The transcriptome and proteome of the

759   diatom Thalassiosira pseudonana reveal a diverse phosphorus stress response. *Plos*

760   *One*, *7*(3), e33768–e33768. https://doi.org/10.1371/journal.pone.0033768

761 Estes, E. R., Pockalny, R., D'Hondt, S., Inagaki, F., Morono, Y., Murray, R. W., Nordlund,

762   D., Spivack, A. J., Wankel, S. D., Xiao, N., & Hansel, C. M. (2019). Persistent

763   organic matter in oxic subseafloor sediment. *Nature Geoscience, 12*(2), 126–131.

764   https://doi.org/10.1038/s41561-018-0291-5

765 Fejjar, S., Melanson, A., & Tremblay, L. (2021). Pore waters as a contributor to deep-water

766   amino acids and to deep-water dissolved organic matter concentration and

767   composition in estuarine and marine waters. *Marine Chemistry, 233*, 103985.

768   https://doi.org/10.1016/j.marchem.2021.103985

769 Finkel, Z. V., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I., & Irwin, A. J. (2016).

770   Phylogenetic Diversity in the Macromolecular Composition of Microalgae. *PLoS*

771   *ONE, 11*(5). https://doi.org/10.1371/journal.pone.0155977

772 Galetskiy, D., Lohscheider, J. N., Kononikhin, A. S., Popov, I. A., Nikolaev, E. N., &

773   Adamska, I. (2011). Mass spectrometric characterization of photooxidative protein

774   modifications in Arabidopsis thaliana thylakoid membranes. *Rapid Communications*

775   *in Mass Spectrometry: RCM, 25*(1), 184–190. https://doi.org/10.1002/rcm.4855

776 Ghesquière, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M.,

777   Schymkowitz, J., Rousseau, F., Vandekerckhove, J., & Gevaert, K. (2011). Redox

778   Proteomics of Protein-bound Methionine Oxidation. *Molecular & Cellular*

779   *Proteomics : MCP, 10*(5). https://doi.org/10.1074/mcp.M110.006866

780 Gray, N., Zia, R., King, A., Patel, V. C., Wendon, J., McPhail, M. J. W., Coen, M., Plumb, R.

781   S., Wilson, I. D., & Nicholson, J. K. (2017). High-Speed Quantitative UPLC-MS

782   Analysis of Multiple Amines in Human Plasma and Serum via Precolumn

783   Derivatization with 6-Aminoquinolyl-N-hydroxysuccinimidyl Carbamate:

Manuscript submitted to *Frontiers in Marine Science*

784   Application to Acetaminophen-Induced Liver Failure. *Analytical Chemistry*, *89*(4),

785   2478–2487. https://doi.org/10.1021/acs.analchem.6b04623

786 Gurdeep Singh, R., Tanca, A., Palomba, A., Van der Jeugt, F., Verschaffelt, P., Uzzau, S.,

787   Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of

788   Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615.

789   https://doi.org/10.1021/acs.jproteome.8b00716

790 Hedges, J. I., Baldock, J. A., Gélinas, Y., Lee, C., Peterson, M., & Wakeham, S. G. (2001).

791   Evidence for non-selective preservation of organic matter in sinking marine particles.

792   *Nature*, *409*(6822), 801–804. https://doi.org/10.1038/35057247

793 Hollibaugh, J. T., & Azam, F. (1983). Microbial degradation of dissolved proteins in

794   seawater1. *Limnology and Oceanography*, *28*(6), 1104–1116.

795   https://doi.org/10.4319/lo.1983.28.6.1104

796 Hoppe, H.-G., H.-G, Arnosti, C., & Herndl, G. (2002). Ecological Significance of Bacterial

797   Enzymes in the Marine Environment. In *Enzymes in the Environment*.

798   https://doi.org/10.1201/9780203904039.ch3

799 Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., Kirchman, D.

800   L., Weinbauer, M. G., Luo, T., Chen, F., & Azam, F. (2010). Microbial production of

801   recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean.

802   *Nature Reviews Microbiology*, *8*(8), 593–599. https://doi.org/10.1038/nrmicro2386

803 Kaiser, K., & Benner, R. (2008). Major bacterial contribution to the ocean reservoir of

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

805   https://doi.org/10.4319/lo.2008.53.1.0099

806 Keil, R. G., & Kirchman, D. L. (1992). Bacterial Hydrolysis of Protein and Methylated

807   Protein and Its Implications for Studies of Protein Degradation in Aquatic Systems.

808   *Applied and Environmental Microbiology*, *58*(4), 1374–1375.

809 Keil, R. G., & Kirchman, D. L. (1993). Dissolved combined amino acids: Chemical form and

810   utilization by marine bacteria. *Limnology and Oceanography*, *38*(6), 1256–1270.

811   https://doi.org/10.4319/lo.1993.38.6.1256

812 Keil, R. G., & Kirchman, D. L. (1994). Abiotic transformation of labile protein to refractory

813   protein in sea water. *Marine Chemistry*, *45*(3), 187–196. https://doi.org/10.1016/0304-

814   4203(94)90002-7

815 Keil, R. G., Tsamakis, E., & Hedges, J. (2000). *Early diagenesis of particulate amino acids*

816   *in marine systems*. 69–82.

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

818   Microbial Communities. *MSystems*, *4*(3). https://doi.org/10.1128/mSystems.00115-19

819 Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., & Strous, M. (2017).

820   Assessing species biomass contributions in microbial communities via

821   metaproteomics. *Nature Communications*, *8*(1), 1–14. https://doi.org/10.1038/s41467-

822   017-01544-x

823 Knicker, H., Río, J. C. del, Hatcher, P. G., & Minard, R. D. (2001). Identification of protein

824   remnants in insoluble geopolymers using TMAH thermochemolysis/GC–MS.

825   *Organic Geochemistry*, *32*(3), 397–409. https://doi.org/10.1016/S0146-

826   6380(00)00186-8

827 Knicker, H., Scaroni, A. W., & Hatcher, P. G. (1996). 13C and 15N NMR spectroscopic

828   investigation on the formation of fossil algal residues. *Organic Geochemistry*, *24*(6),

829   661–669. https://doi.org/10.1016/0146-6380(96)00057-5

830 Laskay, Ü. A., Lobas, A. A., Srzentić, K., Gorshkov, M. V., & Tsybin, Y. O. (2013).

831   Proteome digestion specificity analysis for rational design of extended bottom-up and

832   middle-down proteomics experiments. *Journal of Proteome Research, 12*(12), 5558–

833   5569. https://doi.org/10.1021/pr400522h

834 Laursen, A. K., Mayer, L. M., & Townsend, D. W. (1996). Lability of proteinaceous material

835   in estuarine seston and subcellular fractions of phytoplankton. *Marine Ecology*

836   *Progress Series*, *136*, 227–234. https://doi.org/10.3354/meps136227

837 Lee, C., Wakeham, S., & Arnosti, C. (2004). Particulate Organic Matter in the Sea: The

838   Composition Conundrum. *AMBIO: A Journal of the Human Environment*, *33*(8),

839   565–575. https://doi.org/10.1579/0044-7447-33.8.565

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

841   Chloroplast Proteins: An Emerging Field. *Plant Physiology*, *168*(3), 768–775.

842   https://doi.org/10.1104/pp.15.00117

843 Leprevost, F. V., Valente, R. H., Lima, D. B., Perales, J., Melani, R., Yates, J. R., Barbosa,

844   V. C., Junqueira, M., & Carvalho, P. C. (2014). PepExplorer: A similarity-driven tool

845   for analyzing de novo sequencing results. *Molecular & Cellular Proteomics: MCP,*

846   *13*(9), 2480–2489. https://doi.org/10.1074/mcp.M113.037002

847 Li, W. K. W., Andersen, R. A., Gifford, D. J., Incze, L. S., Martin, J. L., Pilskaln, C. H.,

848   Rooney-Varga, J. N., Sieracki, M. E., Wilson, W. H., & Wolff, N. H. (2011).

849   Planktonic Microbes in the Gulf of Maine Area. *PLOS ONE, 6*(6), e20981.

850   https://doi.org/10.1371/journal.pone.0020981

851   Liu, Z., Kobiela, M. E., McKee, G. A., Tang, T., Lee, C., Mulholland, M. R., & Hatcher, P.

852   G. (2010). The effect of chemical structure on the hydrolysis of tetrapeptides along a

853   river-to-ocean transect: AVFA and SWGA. *Marine Chemistry*, *119*(1), 108–120.

854   https://doi.org/10.1016/j.marchem.2010.01.005

855   Mayer, L. M. (1994). Relationships between mineral surfaces and organic carbon

856   concentrations in soils and sediments. *Chemical Geology*, *114*(3), 347–363.

857   https://doi.org/10.1016/0009-2541(94)90063-9

858   Mayer, L. M., Linda L., S., Sawyer, T., Plante, C. J., Jumars, P. A., & Sel, R. L. (1995).

859   Bioavailable amino acids in sediments: A biomimetic, kinetics based approach.

860   *Limnology and Oceanography*, *40*(3), 511–520.

861   https://doi.org/10.4319/lo.1995.40.3.0511

862   Mayer, L. M., Macko, S. A., & Cammen, L. (1988). Provenance, concentrations and nature of

863   sedimentary organic nitrogen in the gulf of Maine. *Marine Chemistry*, *25*(3), 291–

864   304. https://doi.org/10.1016/0304-4203(88)90056-4

865   Mayer, L. M., Schick, L. L., & Setchell, F. W. (1986). Measurement of protein in nearshore

866   marine sediments. *Marine Ecology Progress Series*, *30*, 159–165.

867   https://doi.org/10.3354/meps030159

868   Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., St-Denis, N., Li, T., Miteva, Y.

869   V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N.

870   C., al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault,

871   M., … Nesvizhskii, A. I. (2013). The CRAPome: A Contaminant Repository for

Manuscript submitted to *Frontiers in Marine Science*

872   Affinity Purification Mass Spectrometry Data. *Nature Methods*, *10*(8), 730–736.

873   https://doi.org/10.1038/nmeth.2557

874 Mesuere, B., Willems, T., Jeugt, F. V. der, Devreese, B., Vandamme, P., & Dawyndt, P.

875   (2016). Unipept web services for metaproteomics analysis. *Bioinformatics*, *32*(11),

876   1746–1748. https://doi.org/10.1093/bioinformatics/btw039

877 Mikan, M. P., Harvey, H. R., Timmins-Schiffman, E., Riffle, M., May, D. H., Salter, I.,

878   Noble, W. S., & Nunn, B. L. (2020). Metaproteomics reveal that rapid perturbations

879   in organic matter prioritize functional restructuring over taxonomy in western Arctic

880   Ocean microbiomes. *The ISME Journal*, *14*(1), 39–52.

881   https://doi.org/10.1038/s41396-019-0503-z

882 Montgomerie, S., Cruz, J. A., Shrivastava, S., Arndt, D., Berjanskii, M., & Wishart, D. S.

883   (2008). PROTEUS2: A web server for comprehensive protein structure prediction and

884   structure-based annotation. *Nucleic Acids Research*, *36*(suppl_2), W202–W209.

885   https://doi.org/10.1093/nar/gkn255

886 Moore, E. K., Harvey, H. R., Faux, J. F., Goodlett, D. R., & Nunn, B. L. (2014). Protein

887   recycling in Bering Sea algal incubations. *Marine Ecology Progress Series*, *515*, 45–

888   59. https://doi.org/10.3354/meps10936

889 Moore, E. K., Nunn, B. L., Goodlett, D. R., & Harvey, H. R. (2012). Identifying and tracking

890   proteins through the marine water column: Insights into the inputs and preservation

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

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

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

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

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

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

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

898   29–37. https://doi.org/10.3354/ame014029

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

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

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

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

903   Hydrophobic and other noncovalent associations as major stabilizing forces.

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

905   7037(00)00621-9

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

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

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

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

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

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

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

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

914   1790–1804. https://doi.org/10.4319/lo.2010.55.4.1790

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

930   and Taxonomic Analysis of Metaproteomics Data. *Proteomes*, *6*(1).

931   https://doi.org/10.3390/proteomes6010002

932 Roth, L., & Harvey, H. (2006). *Intact protein modification and degradation in estuarine*

933   *environments*. https://doi.org/10.1016/J.MARCHEM.2005.10.025

934 Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J.,

935   Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M.

936   R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd,

937   A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean

938   Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome*

939   *Research, 18*(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761

940 Sakai, H., & Tsukihara, T. (1998). Structures of Membrane Proteins Determined at Atomic

941   Resolution. *J Biochem (Tokyo), 124*(6), 1051–1059.

942 Schmidt, F., Koch, B. P., Elvert, M., Schmidt, G., Witt, M., & Hinrichs, K.-U. (2011).

943   Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under

944   Contrasting Sedimentary Redox Conditions in the Black Sea. *Environmental Science*

945   *& Technology, 45*(12), 5223–5229. https://doi.org/10.1021/es2003414

946 Shen, Y., Hixson, K., Tolić, N., Camp, D., Purvine, S., Moore, R. J., & Smith, R. (2008).

947   Mass spectrometry analysis of proteome-wide proteolytic post-translational

948   degradation of proteins. *Analytical Chemistry*. https://doi.org/10.1021/ac800077w

949 Smith, L. J., Fiebig, K. M., Schwalbe, H., & Dobson, C. M. (1996). The concept of a random

950   coil: Residual structure in peptides and denatured proteins. *Folding and Design, 1*(5),

951   R95–R106. https://doi.org/10.1016/S1359-0278(96)00046-6

952 Stevens, T. J., & Arkin, I. T. (2000). Do more complex organisms have a greater proportion

953   of membrane proteins in their genomes? *Proteins: Structure, Function, and*

954   *Bioinformatics, 39*(4), 417–420. https://doi.org/10.1002/(SICI)1097-

955   0134(20000601)39:4<417::AID-PROT140>3.0.CO;2-Y

956 Suzuki, S., Kogure, K., & Tanoue, E. (1997). Immunochemical detection of dissolved

957   proteins and their source bacteria in marine environments. *Marine Ecology Progress*

958   *Series, 158*, 1–9. https://doi.org/10.3354/meps158001

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

960   source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta*,

961   *59*(12), 2643–2648. https://doi.org/10.1016/0016-7037(95)00134-4

962 Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble,

963   W. S., & Nunn, B. L. (2017). Critical decisions in metaproteomics: Achieving high

964   confidence protein annotations in a sea of unknowns. *The ISME Journal*, *11*(2), 309–

965   314. https://doi.org/10.1038/ismej.2016.132

966 UniProt Consortium, T. (2018). UniProt: The universal protein knowledgebase. *Nucleic Acids*

967   *Research*, *46*(5), 2699–2699. https://doi.org/10.1093/nar/gky092

968 Van Mooy, B. A. S., Keil, R. G., & Devol, A. H. (2002). Impact of suboxia on sinking

969   particulate organic carbon: Enhanced carbon flux and preferential degradation of

970   amino acids via denitrification. *Geochimica et Cosmochimica Acta*, *66*(3), 457–465.

971   https://doi.org/10.1016/S0016-7037(01)00787-6

972 Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., & Peterson, M. J. (1997). Molecular

973   indicators of diagenetic status in marine organic matter. *Geochimica et Cosmochimica*

974   *Acta*, *61*(24), 5363–5369. https://doi.org/10.1016/S0016-7037(97)00312-8

975 Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E., & Schulz, G. E. (1991).

976   Molecular architecture and electrostatic properties of a bacterial porin. *Science (New*

977   *York, N.Y.)*, *254*(5038), 1627–1630. https://doi.org/10.1126/science.1721242

978 Wolfe, A., Edlund, M., Sweet, A., & Creighton, S. (2006). A First Account of Organelle

979   Preservation in Eocene Nonmarine Diatoms: Observations and Paleobiological

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

981 Yamada, N., & Tanoue, E. (2003). Detection and partial characterization of dissolved

982   glycoproteins in oceanic waters. *Limnology and Oceanography, 48*(3), 1037–1048.

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

984   Yeagle, P. L., Bennett, M., Lemaître, V., & Watts, A. (2007). Transmembrane helices of

985   membrane proteins may flex to satisfy hydrophobic mismatch. *Biochimica Et*

986   *Biophysica Acta, 1768*(3), 530–537. https://doi.org/10.1016/j.bbamem.2006.11.018

987   Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K.,

988   Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P.,

989   Miller, C. S., Li, H., Mashiyama, S. T., Joachimiak, M. P., Belle, C. van, Chandonia,

990   J.-M., Soergel, D. A., … Venter, J. C. (2007). The Sorcerer II Global Ocean Sampling

991   Expedition: Expanding the Universe of Protein Families. *PLOS Biology, 5*(3), e16.

992   https://doi.org/10.1371/journal.pbio.0050016

993   Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.

994   A., & Ma, B. (2012). PEAKS DB: De novo sequencing assisted database search for

995   sensitive and accurate peptide identification. *Molecular & Cellular Proteomics: MCP,*

996   *11*(4), M111.010587. https://doi.org/10.1074/mcp.M111.010587

997   Zhou, Q., & Qiu, H. (2019). The Mechanistic Impact of N-Glycosylation on Stability,

998   Pharmacokinetics, and Immunogenicity of Therapeutic Proteins. *Journal of*

999   *Pharmaceutical Sciences, 108*(4), 1366–1377.

1000   https://doi.org/10.1016/j.xphs.2018.11.029