Suspended marine particulate proteins in coastal and oligotrophic waters.

Maxime C. Bridoux[1,3], Jaqui Neibauer[1], Anitra E. Ingalls[1],
Brook L. Nunn[2], and Richard G. Keil[1]*

[1] School of Oceanography, Box 355351, University of Washington, Seattle, Washington 98195, USA

[2] Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA

[3] currently at: Commissariat à l'énergie atomique et aux énergies alternatives, CEA, DAM, DIF, F-91297 Arpajon, France

* corresponding author email: rickkeil@uw.edu

Abstract

Metaproteomic analyses were performed on suspended sediments collected in one coastal environment (Washington margin, Pacific Ocean, n=5) and two oligotrophic environments (Atlantic Ocean near BATS, n=5, and Pacific Ocean near HOTS, n=5). Using a database of 2,324,578 marine proteins developed using the NCBI database, 443 unique peptides were detected from which 363 unique proteins were identified. Samples from the euphotic zone contained on average 2-3x more identifiable proteins than deeper waters (150-1500 m) and these proteins were predominately from photosynthetic organisms. Diatom peptides dominate the spectra of the Washington margin while peptides from cyanobacteria, such as *Synechococcus sp.*. dominated the spectra of both oligotrophic sites. Despite differences in the exact proteins identified at each location, there is good agreement for protein function and cellular location. Proteins in surface waters code for a variety of cellular functions including photosynthesis (24% of detected proteins), energy production (10%), membrane production (9%) and genetic coding and reading (9%), and are split 60-40 between membrane proteins and intracellular cytoplasmic proteins. Sargasso Sea surface waters contain a suite of peptides consistent with proteins involved in circadian rhythms that promote both C and N fixation at night. At depth in the Sargasso Sea, both fish muscle-derived myosin protein and the muscle-hydrolyzing proteases deseasin MCP-01 and metalloprotease Mcp02 from γ-proteobacteria were observed. Deeper waters contain peptides predominately sourced from γ-proteobacteria (37% of detected proteins) and α-proteobacteria (26%), although peptides from membrane and photosynthetic proteins attributable to phytoplankton were still observed (13%). Relative to surface values, detection frequencies for bacterial membrane proteins and extracellular enzymes rose from 9 to 16 and 2 to 4% respectively below the thermocline and the overall balance between membrane proteins and intracellular proteins grows to a 75-25 split. Unlike the phytoplankton membrane proteins, which are detrital in nature, the bacterial protein suite at depth is consistent with living biomass.

Keywords: metaproteome, protein, peptide, POM, organic carbon

Regional Index Terms: Washington margin, HOT, BATS Sargasso Sea

1.0 Introduction

Marine particulate organic matter (POM) is a mixture of living biomass (e.g., microalgal cells, bacteria, archea etc.) and nonliving detritus (e.g., fecal pellets, diatom frustules, clay particles with organic coating) with a broad range of sizes, forms and diagenetic reactivities (Kaiser and Benner 2008; Lee et al. 2009b; Dong et al. 2010). This pool of organic matter (OM) plays a major role in many oceanic and global biogeochemical processes such as the regulation of air-sea carbon fluxes, vertical transport of nutrients to deep water, fueling of marine food webs in the ocean interior and benthos, as well as transformation of particulate OM to the dissolved organic matter pool (Hedges et al. 2001; Lee et al. 2009a). Detailed characterization of its sources, forms and fates is thus essential in order to access the biogeochemical information the POM pool carries.

Hydrolyzable amino acids are the largest identified molecular fraction in marine OM (Wakeham et al. 1997), and much of this is thought to be in the form of recognizable proteins as opposed to detrital peptides or humic-type organics (Tanoue et al. 1995; Long and Azam 1996; Saijo and Tanoue 2005). Although analysis of this pool has historically been via hydrolysis and amino acid quantification (Keil and Kirchman 1993; Nguyen and Harvey 1997; Kaiser and Benner 2012), the identification of proteins and peptides in POM promises to provide additional information about their origin and cellular function, including structural intent, energy production and metabolism. Because of this, additional tools for understanding amino acids in POM have long been sought (Tanoue 1992), and a variety of electrophoretic and proteomic approaches have been applied (Suzuki et al. 1997; Nunn et al. 2003; Yamada and Tanoue 2003; Morris et al. 2010; Nunn et al. 2010; Tsukasaki and Tanoue 2010; Moore et al. 2012b). Although applications to marine POM are still limited, the most common approach is the use of 'shotgun' metaproteomics (Moore et al. 2012b; Nunn et al. 2013), a discovery-based proteomics tool that facilitates a non-biased approach to protein identification from complete unknowns. Using the rapidly growing genomic and proteomic databases in conjunction with shotgun proteomics can provide information on the origin and cellular localization of proteins which can shed light on mechanisms protecting proteins from degradation (Nunn et al. 2010; Moore et al. 2012b).

The use of proteomic techniques has allowed the discovery of new features of marine phytoplankton (Jones et al. 2011; Dyhrman et al. 2012; Nunn et al. 2013) and heterotrophic

bacteria (Sowell et al. 2009; Morris et al. 2010; Sowell et al. 2011; Mattes et al. 2013; Smith et al. 2013). For example, Morris et al (2010) showed that rhodopsins and TonB-dependent transporter proteins, which are used by bacteria to generate proton gradients and assimilate nutrients, are ubiquitous in Atlantic Ocean surface waters. Moore et al. (2012b) showed that organelle-bound, transmembrane and photosynthetic proteins from diatoms are transported to depth in the Bering Sea whilst other phytoplanktonic proteins are lost from the detection window. This work echoes that of Nunn et al. (2010) who evaluated the degradation of *Thalassiosira pseudonana* under controlled lab conditions and concluded that organelle and membrane protection represent important mechanisms to enhance protein transfer to the deep sea and preservation.

We collected euphotic zone and intermediate water samples of POM from two oligotrophic environments; the Sargasso Sea and the oligotrophic north Pacific. We also sampled one coastal environment, the Washington margin in the north-east Pacific Ocean (Fig. 1). In order to evaluate potential similarities and differences between oceans and oceanic zones (e.g. depth), we performed metaproteomic analyses to characterize the suspended material. The primary purpose of the study was to confidently identify well-represented proteins in the POM fraction from these areas, determine their catalytic functionality and ascertain taxonomic information about the proteins' point of origin. Like Morris et al (2010) transporter proteins were observed at depth in both coastal and oligotrophic waters, but rhodopsins were only detected in surface waters. This suggests that while transporter proteins are utilized by bacteria at all depths, the membrane-bound light harvesting components of phototrophic bacteria are only expressed at the surface. Like Moore et al (2012), we found that only specific types of proteins from diatoms are identified after being transported to depth, and extend this finding to include similar structures from the photosynthetic cyanobacteria of the oligotrophic ocean, suggesting that selective transport of photosynthetic proteins to depth is not limited to those derived from diatoms phenomenon is dependent on dominant biological activity.

2.0 Material and Methods

2.1 Sample Collection

POM samples were collected *in situ* using large volume water transfer systems (McLane pumps) at five depths each for the Sargasso Sea (March 2011; R/V Atlantic Explorer), the

tropical north Pacific (July 2010; R/V Kilo Moana) and the Washington margin locations (September 2010; R/V Wecoma)(Table 1, Fig 1). During filtration, samples were passed through a large (4mm) mesh screen onto combusted 142mm diameter GF/F filters that were double stacked. At one depth in the Sargasso Sea (Table 1) replicate pumps were run at the same depth to compare a 0.2 μm Durapore filter (Millipore Inc, Polyvinylidene fluoride low protein binding membrane) against a single GF/A and double stacked GF/F filters. After collection, filters were folded, packed in combusted aluminum foil sleeves and kept dark and frozen (-80°C) until analysis in the laboratory, which occurred within 9 months of collection. Pumps were run at a nominal setting of 8 L m^{-1} except for the one depth in the Sargasso Sea where the flow was reduced to 4 L m^{-1} in order to facilitate the use of the Durapore filter. Volumes filtered ranged from 600-1500 L (Table 1). A single horizontal net tow (20 μm mesh) was taken along the Washington margin. A 300 mm diameter Puget Sound-type closing net (Aquatic Research Instruments, Hope ID) was lowered open to the chlorophyll max and towed 1 knot for 10 min before being closed by messenger and recovered. The material trapped in the cod-end was poured into centrifuge tubes, centrifuged (2,800x*g*), the water was then discarded and the resulting pellet was frozen at -20 °C until analysis. Ancillary data for each sample comes from CTD casts, which were performed immediately before each pump or net collection event.

2.2 Protein extraction and quantification

Filters were thawed and cut into chips (2x2 mm) using a sterilized razorblade. Proteins were extracted following a protocol adapted from (Dong et al. 2009). Filter chips were suspended in 5 mL lysis buffer consisting of 7M urea, 2M thiourea, carrier ampholytes (2% v/v, pH 3-10), dithiothreitol (DTT, 1% w/v), and protease inhibitor cocktail (1% v/v). The suspension was then placed on an ice cold bath and lysed with the help of a high-power sonicator (600 W, 20-kHz sonicator with a CV26 probe (5 min, 6 s pulse, 60 amplitude) resulting in a whole-debris lysate. The slurry was incubated at 25°C for 1 h and then centrifuged at 20,000 x *g* for 30 min at 10°C. The supernatant was transferred to a clean Teflon centrifuge tube and precipitated with ice-cold 20% trichloroacetic acid in acetone (TCA/acetone, 1:5 v/v) for at least 12 h at -20°C. The precipitated protein mixtures were centrifuged at 20,000 x *g* for 30 min at 4°C, and the resulting protein pellets were rinsed twice with ice-cold acetone and air-dried. The dry pellets were resuspended in 100 μL rehydration buffer (6M urea 10mM Tris). Proteins extracted from the

POM filter samples were quantified using the NanoOrange® assay using bovine serum albumin as the protein standard.

2.3 Protein trypsin digestion and peptide desalting

The whole protein lysate was enzymatically digested with trypsin following a protocol adapted from Nunn et al. (2010). Protein pellets were reconstituted in 300 µl of 6 M urea in 50 mM ammonium bicarbonate with 20 µl of 1.5 mM Tris buffer (pH 8.8) and protein disulphide bonds were reduced with 7.5 µL of 200 mM tris(2-carboxyethyl)phosphine (TCEP). The resulting mixture was incubated at 37°C for 1 h and then incubated with 60 µl of 200 mM iodoacetamide (IAM; for alkylation) in the dark for 1 h at 25°C. Excess IAM was neutralized with 60 µl of 200 mM DTT for 1 h (25°C). A volume of 150 µl of each sample was aliquoted into 2 tubes, and 800 µl of 25 mM NH_4HCO_3 was added to dilute the urea prior to the addition of 200 µl of MeOH and sequence-grade trypsin (Promega) at 50:1 substrate:enzyme (w/w). Trypsin digestions were vortexed and incubated at room temperature overnight. Samples were then taken to near dryness in a rapidvap and 200 µl of Milli-Q H_2O aliquot was added to reduce the NH_4HCO_3, and evaporated; the process was repeated 3 times to lower the pH of the sample. The resulting tryptic peptides were desalted using a macro-spin C_{18} column (NestGroup) following the manufacturer's guidelines and concentrated to dryness using a centrivap. Samples were stored at -80°C until mass spectral analysis.

2.4 Liquid chromatography mass spectrometry

The dried tryptic peptides were suspended in 20 µL of 5% Acetonitrile (ACN) /0.1% formic acid and introduced via a home-packed analytical column into a linear ion trap-orbitrap (LTQ-OT) hybrid mass spectrometer (Thermo Fisher) using a NanoAcquity HPLC (Waters Corporation). The home-packed column consisted of an 11-cm-long, 75-µm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100Å, 5mm; Michrom) coupled to a 2 cm long, 100 µm i.d. precolumn (Magic C18AQ, 200Å, 5mm; Michrom). Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-ACN gradient (5–35% ACN in 60 min). Eluting peptide ions were selected for collision induced dissociation in the LTQ by data-dependent instrument control from precursor ion scans in the Orbitrap. The LTQ mass spectrometer operated with the following parameters: spray voltage (3.4 kV), transfer capillary temperature 250°C, full scan m/z range (*m/z* 400-1800), and dynamic exclusion was enabled

with a repeat count of 2 and an exclusion duration of 1.5 min on the LTQ. The LC-MS system was fully automated and controlled by Xcalibur software (Thermo Finnigan). Data-dependent acquisition strategy was utilized where MS/MS spectra were collected on the five most intense ions observed in the MS^1 precursor scan.

2.5 Bioinformatics and particulate protein identity analysis

A composite database of proteins derived from marine planktonic communities was built using the protein bank of the National Center for Biotechnology Information (NCBI) Web site (download date: 1/2012). The database consisted of marine bacteria and archaebacteria, marine phytoplankton, and zooplankton as well as those closely related environmental samples with marine bacteria or archaea (Table S1). See Supplemental Table S1 for a complete list of organisms included in the database and their number of protein entries accounting for a total of 2,324,578 protein entries.

All MS^2 spectra were searched against the database using the SEQUEST algorithm (Eng et al. 2008). SEQUEST compares acquired tandem mass spectra to theoretical peptide spectra generated *in silico* from a protein database. Data searches were completed with the following parameters: enzyme type trypsin; parent ion tolerance, 50 ppm; fragment ion mass tolerance, 1.00 Da. Modifications of cysteine residues by 57.021 Da (resulting from the iodoacetamide modification) and methionine by 15.999 Da (oxidation) were allowed. The Trans-Proteomic Pipeline (version 4.1.1, Institute for Systems Biology) was used to validate MS^2 - based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90.0% probability as specified by the Peptide Prophet algorithm (Keller et al. 2002). Protein identifications were accepted if they could be established at greater than 90.0% probability. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al. 2003).

The method of identifying the biological origin of the proteins was completed in a similar fashion to the study completed by Schulze et al (2005) and Morris et al. (2010). Peptides identified from SEQUEST database searches were further interrogated using BLASTp (http://blast.ncbi.nlm.nih.gov) in order to extract all levels of taxonomic information for each confident protein hit. BLASTp was performed against the non-redundant database and taxonomic information on each peptide was retained for further analysis and discussion.

Bacterial proteins were classified into six phyla: Proteobacteria, Bacteroidetes, Planctomycetes, Firmicutes, Actinobacteria, and Cyanobacteria. Proteins from eukaryotic algae were classified into either phyla (Prasinophyta, Cryptophyta, Heterokonta), division (Haptophyta, Chlorophyta) or class (Dinophyta, Bacillariophyta) depending on the resolution of the BLASTp results. Proteins not belonging to the above groups were either by grouped by domain (e.g. Archaea). Protein subcellular localization was predicted using Cell-PLoc 2.0 package (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/) (Chou and Shen 2008). This prediction package can be used for not only the single-location or ''singleplex'' proteins but also the multiple-location or ''multiplex'' proteins that may simultaneously reside at, or move between two or more different subcellular locations. Functional categorization of the identified proteins was performed using Clusters of Orthologous Groups of proteins (COGs) (http://www.ncbi.nlm.nih.gov/ COG/) (Tatusov et al. 2000). Protein theoretical molecular weight (MW) and isoelectric point (pI) values were generated using ExPaSy (http://web.expasy.org/compute_pi/) (Gasteiger et al. 2005).

3.0 Results

A total number of 443 unique peptides were detected, correlating to 363 proteins from the composite planktonic marine community database. (Supplemental material Table S2). The Sargasso Sea had the most proteins identified (230) followed by the Washington Margin (96) and the oligotrophic north Pacific (73). Only three specific proteins (ABC transporter, photosystem II protein and histone H4) were common to all the three areas sampled (Table S2). Twenty two proteins were in common between the oligotrophic Pacific and Atlantic environments (Table S2). In addition to the three ubiquitous proteins, six proteins were common to the Washington margin and the Sargasso Sea and one was common to the tropical North Pacific and the Washington margin (Table S2).

The 155m sample from the Washington margin was run on the LC-MS twice in order to evaluate the consistency of the chromatography and protein identification protocols (Table S2). Eleven proteins were identified in the first run, and all eleven plus one additional protein were identified in the second run. This additional protein was observed in the first run but not with enough confidence to pass our 'double' 0.9 cutoff for quantitation using Peptide Prophet and Protein Prophet.

211 Particle capture was evaluated in the Sargasso Sea by deploying three McLane pumps at the
212 same depth. One contained two stacked GF/F filters (effective filtration cutoff of ~0.7 μm), one
213 was loaded with a single GF/A (1.6 μm) and one contained a 0.2 μm Durapore membrane. This
214 comparison was intended to determine whether the effective pore size and/or composition of the
215 filters had an impact on the number of and types of proteins identified. The double-stacked GF/F
216 and Durapore filters (nominal pore sizes of 0.7 and 0.2 μm respectively) performed very
217 similarly in this analysis, resulting in 36 and 32 proteins identified. Sixteen of these proteins
218 were in common. Of the 26 peptides identified on only the GFF or the Durapore, 20 peptide
219 spectral matches were made from samples of other filters, but the PeptideProphet threshold of
220 0.9 was not met. The greatest number of proteins confidently identified in this comparison (45)
221 resulted from the analysis of the single GFA filter with nominally the largest pore size (1.6 μm).
222 In addition to identifying 23 proteins in common with the other filters (50% overlap), an
223 additional 22 peptides were uniquely identified on the GF/A in this comparison (but only 3 were
224 unique within the entire data set - three membrane proteins from γ-proteobacteria; Table S2).
225 Again, more than half (13) of the peptides unique to the GF/A filter sample were also observed
226 in the mass spectra of the other samples but could not be identified with enough confidence to
227 pass our PeptideProphet threshold. The random nature of peptide sampling that occurs during
228 data dependent acquisition (DDA) proteomics is well known and documented (Liu et al. 2004).
229 In addition to the fact that DDA does not sample all peptides, biological replicates from an
230 environmental sample have inherent variability. Therefore, the agreement between filters was
231 probably much better than the numbers suggest. These results are in line with those of Nunn et
232 al (2009) who analyzed triplicate diatom cultures and observed only ~40% overlap in the protein
233 identifications after filtering data through PeptideProphet and ProteinProphet.
234 The diversity of proteins identified in the euphotic zone samples was greater than from
235 samples collected deeper in the water column, and the quantity of total extracted protein also
236 decreases with depth (Fig 2). For example, in the Sargasso Sea, the highest number of proteins
237 (212) were identified on filters collected at 150 m deep and the number of confident
238 identifications decreased significantly with sampling depth (e.g. 103 and 40 proteins identified at
239 300 and 500 m, respectively). Few proteins (<20) were identified below 500 m. These types of
240 numbers are consistent with Moore et al. (2012b) and Dong et al. (2010) who observed similar
241 numbers of total proteins and similar decreases with depth. There is a significant correlation (r^2

242 = 0.86) between the amount of protein in the water and the number of proteins identified, with an
243 average of two more proteins per μg L^{-1} of protein in the sample (Fig 2c).
244 Surface waters contain proteins largely sourced from phototrophs, with diatoms dominating
245 the proteins of the Washington margin and cyanobacteria proteins dominating in the oligotrophic
246 environments (Fig 3a-c). Proteins identified from samples from below the euphotic zone
247 originated from a variety of sources, with phytoplanktonic proteins remaining (13% of detected
248 proteins) but γ-proteobacteria (37%) and α-proteobacteria (26%) proteins becoming a larger
249 fraction of the total identified (Fig 3d-f). Archea, despite being numerically the most abundant
250 source of protein in our database, were conspicuously absent from the samples and only 1.5% of
251 the proteins detected can be attributed to them.
252 The functions of the proteins identified in surface waters were distributed between the 26
253 Clusters of Orthologous Groups (Fig 4a-c). Photosynthesis (24% of identified proteins), energy
254 production (10%), membrane production (9%) and genetic coding and reading proteins (9%)
255 were the best represented clusters. There is good agreement between the observed distribution of
256 euphotic zone protein functions and that expected for living diatoms and Synechococcus sp.
257 (Palenik et al. 2003; Armbrust et al. 2004), suggesting that the identified proteins likely represent
258 living phytoplankton in each respective location (data not shown).
259 Deeper waters have protein function distributions similar to the surface waters except that
260 the contribution of membrane and extracellular enzymes are both enhanced relative to other
261 functions (Fig 4d-f). Membrane proteins and extracellular enzymes rose from 9 to 16% and 2 to
262 4% respectively below the thermocline. Predicted protein locations within living cells were
263 broadly distributed, with surface waters dominated by proteins from inner membranes (33%),
264 cytoplasm (20%) and chloroplasts (15%). The overall division was roughly 65% membrane
265 proteins and 35% intracellular cytoplasmic proteins (Fig 5). Relative to the surface, deeper
266 waters are enriched in outer membrane and periplasm proteins from bacteria (24 and 11%) and
267 are depleted in chloroplast and cytoplasm proteins (Fig 5). The overall balance between
268 membrane and intracellular proteins is 75:25 in deep waters.

269 **4.0 Discussion**

270 Although the total of 363 unique proteins is considered low for a data-dependent MS
271 experiment on a cultured organism, it is typical for oceanic studies of suspended POM or other

detrital-rich samples. Moore et al (2012b) identified between 11-207 proteins in each sample from Bering Sea POC, and Nunn et al (2010) identified 121 proteins at the beginning of a diatom degradation experiment (5 days), but only 4 after extensive remineralization (23 days). Previous studies evaluating marine samples have reported greater numbers of proteins without the use of a statistically rigorous peptide spectral matching validation software, such as PeptideProphet or ProteinProphet (Dong et al. 2010; Wang et al. 2011). In addition, reported redundancy in the protein file lists of these other works (i.e., multiple protein identifications listed for individual spectral matches) does not allow for a valid comparison with data presented in this study.

Few peptide spectra were confidently correlated to proteins originating from archaea. Archaea are a dominant prokaryote of the subsurface (Karner et al. 2001) and comprise the largest source of protein sequences in our database (Table S1). The use of GF/F filters with a relatively large nominal pore size of 0.7 μm might have not captured these organisms. In our comparison of a 0.2 μm Durapore membrane with GF/F and GF/A filters, three archaeal proteins were detected on the Durapore filter, one on the GF/A and none on the GF/F (Table S2). This implies that the GF/F under-samples archaea, a result consistent with that of Wang et al (Wang et al. 2011) who report significant under-sampling of archaeal proteins even in a 0.2 – 0.7 μm fraction compared to a 5 kD – 0.2 μm fraction. In addition, Kirkland et al (2008) observed only ~25% overlap in proteins from different cultures, and Maupin-Furlow et al (2012) suggest that tools to extract and evaluate archaeal proteins are still needed. This literature implies that more work is required before we will be consistently evaluating archaeal proteomes.

Peptides from five protein homologues (based on cellular function) were observed at every location and almost at every depth, in our data as well as Dong et al (2010) and Moore et al (2012). They are ABC transporters (present in all life), TonB-dependent receptors and dicarboxylate transporters (bacterial), and phycoerythrin and photosystem II D2 proteins (phytoplankton). With the exception of cyanobacterial-derived phycoerthrin, which was not found in the Bering Sea (Moore et al 2012), these peptides can be considered ubiquitous in the ocean.

Like porins, the first proteins directly identified in the ocean (Tanoue 1995; Tanoue et al. 1995), ABC transporters and TonB-dependent receptors are membrane proteins frequently detected in oceanic samples (Morris et al. 2010; Poretsky et al. 2010; Tang et al. 2012; Williams et al. 2012). Detection of peptides derived from these proteins provides information both for

303 microbiological studies of the substrates utilized by marine bacteria (Poretsky et al. 2010; Tang
304 et al. 2012) but also to geochemical studies of carbon cycling in the ocean because they have
305 been identified as potential marker proteins to assess the contribution of bacterial necromass to
306 oceanic detrital pools (Kaiser and Benner 2008; Kawasaki et al. 2011; Kaiser and Benner 2012).
307 These and other membrane proteins such as porins comprise 7% of the proteins at the surface
308 and 24% at depth (Fig 5).
309 Peptides from the two phytoplanktonic proteins phycoerythrin and photosystem II D2, along
310 with other photosynthesis-related proteins, are among the most frequently detected peptides in
311 surface waters, which is consistent with their important role in fueling the entire upper water
312 column (Figs 4&5)(Palenik et al. 2003; Armbrust et al. 2004). They are also found at depth,
313 indicating that they might be effectively transferred to depth in an intact state (Dong et al. 2010;
314 Moore et al. 2012a). Deep mixing events in the Sargasso Sea can effectively transfer intact cells
315 to depths of 500 m or more (Conte et al. 2003; Malmstrom et al. 2010; Alonso-Gonzalez et al.
316 2013), in which case the peptides observed in this study could be derived from viable cells or
317 resting stages.
318 The hypothesis that phytoplankton peptides detected at depth mark intact cells is
319 inconsistent with the observed correlation between protein content of samples and the number of
320 proteins detected (Fig 2). This is because deep water POM is present in lower quantities per liter
321 of water and contains fewer amino acids per unit organic matter (Wakeham et al. 1997; Lee et al.
322 2004; Hannides et al. 2013). The diversity of identifiable peptides has been previously observed
323 to decrease with protein amount in a sample (Moore et al 2012) suggesting that when samples
324 have experienced increased degradation, as is the presumed case with deeper suspended POM
325 (Volkman and Tanoue 2002; Saijo and Tanoue 2005; Kaiser and Benner 2009), there are fewer
326 intact proteins amenable to tryptic digestion. Non-tryptic peptides resulting from natural
327 degradation processes can be detected by the standard proteomic method, but in order to reduce
328 the rate of false discovery peptide spectral matches to a metaproteomic database, these spectra
329 are not interrogated (i.e., all peptides must have 2 tryptic termini). This suggests that the
330 majority of detrital proteins are at least partially degraded during the sinking process, a process
331 that would prevent identification, but allow for abundant peptides and hydrolysable amino acids
332 to be found in deep water POM.

To distinguish between the likelihood that the proteins are within living cells versus detrital material, the distribution of cyanobacterial proteins from different depths within the Sargasso Sea were evaluated as a function of their predicted cellular location (i.e., membrane, cytosolic, etc.; Fig 6). At the base of the mixed layer (150m), the cyanobacterial protein functional distribution is well spread between different biological functions (Table S2) and there are a number of detected proteins located in the cytoplasm and in other cellular locations (Fig 6). At 300m there are fewer proteins, and fewer that perform cellular functions such as energy transfer, protein synthesis and cellular division. The only cyanobacterial proteins detected in the 500m samples are photosynthetic proteins in, or associated with, the thykaloid membrane. Thus, because the distribution of deep water proteins are so enriched in membrane proteins relative to the surface, this deep water protein material is likely dead biomass and detritus (Fig 6). In a study of proteins in the South China Sea, Dong et al (2010) observed many peptides in deep water POM samples that could be attributed to cyanobacterial proteins. Our evaluation of their data (Dong et al, 2010, Appendix tables 1&4) revealed only a slight relative increase in inner membrane proteins with depth. This is somewhat consistent with our data, but because similar bioinformatics workflows were not applied to both data sets a more complete comparison cannot be made.

Similarly, our Washington margin sample contains an opportunity to compare diatom-derived protein transport to that of Moore et al (2012) in the Bering Sea and the degradation experiment of Nunn et al (2009). We identified 29 diatom-specific proteins in the net tow, but only three were identified at depth: fucoxanthin-chlorophyll a/c light-harvesting protein and two proteins of unknown function. The peptides that remain detectable and are clearly from phytoplankton may hold information about the mechanisms that can transfer whole (or largely intact) proteins to the deep sea. Fucoxanthin-chlorophyll a/c light-harvesting proteins are part of the photosystem supercomplex and are associated with PS1 (Dekker and Boekema 2005). They are not membrane proteins but bound within 5 membranes and part of a large pigment-protein complex, a tertiary structure similar to that of the cyanobacterial phycoerythrins observed in the deep Sargasso Sea samples. These large protein-protein structures are typically stabilized via a combination of disulfide bonds, hydrogen bonds and ionic interactions, all of which have been proposed as mechanisms favoring protection of organic material over rapid degradation

(Canfield et al. 1998; Nguyen and Harvey 2001; Roth and Harvey 2006; Nunn et al. 2010; Moore et al. 2012b).

The remaining phytoplanktonic proteins identified at depth are Photosystem II proteins from diatoms and ferredoxin from cyanobacteria. Both are membrane bound proteins, a cellular location previously hypothesized to slow degradation in algal detritus (Nunn et al. 2010; Moore et al. 2012b). Thus, our identification of various proteins is consistent with tertiary structures and membranes both being related to peptide detection at depth, and presumably protection of protein from degradation. While the findings related to detection are clear, the evidence to support protection from degradation is not so clear cut. As noted by Nunn et al (2010), the abundance of a protein in an intact viable cell may play as much a role in determining its presence after the cell dies as anything else. The metaproteomic technique applied here simply marks presence or absence, and a protein of high abundance initially (e.g. in surface waters) might still be detectable later (e.g. at depth) because enough of its marker peptides have remained untouched. That is, the protein might not be 'special' beyond being abundant. This type of protein degradation pattern would be consistent with other work suggesting that the composition of POM in the ocean undergoes little change with depth (Hedges et al. 2001; Roland et al. 2008; Buesseler and Boyd 2009). However, based on spectral counting techniques, the PSII proteins of diatoms are not the most abundant proteins in living phytoplankton (Nunn et al. 2010; Nunn et al. 2013) and in our samples few other membrane-associated proteins (phycobilisomes, ATP synthetase, etc) are detected at depth, suggesting that there indeed may be something about PSII proteins or protein types that promote their resistance to decay and thus detection at depth.

Unlike the cyanobacterial and diatom proteins, the remaining (likely heterotrophic) bacterial proteins do not show enhancement of membrane proteins with depth (Fig 6). While the total number of detected proteins decreases with depth, the distribution of these proteins into different cellular components stays the same and remains in correct proportions to living bacteria (Christie-Oleza et al. 2012). This is interpreted as an indication that these proteins are marking living bacteria and not detritus. Given how membrane constituents including proteins are thought to mark the bacterial contribution to the detrital pools of dissolved organic matter (Nagata and Kirchman 1996; Nagata et al. 1998; Kaiser and Benner 2008; Kawasaki et al. 2011) and sedimentary organic matter (Lomstein et al. 2012), it is surprising that the suspended POM

proteins collected here do not also reflect a detrital bacterial pool. Kawasaki et al (2011) used amino acid analyses and estimated that the bacterial contribution to suspended POM in the Tropical North Pacific was split ~50-50 between detritus and living cells. Our proteomic data does not support that finding for the Sargasso Sea (Fig 6) and suggests a higher proportion for living material. Our data for the tropical North Pacific come from an area near Hawaii Ocean Time Series (HOTS) where Kawasaki et al (2011) sampled. We detect too few heterotrophic bacterial proteins (3-9) at that location to make a definitive comparison, but the 750 m sample contains 3 outer membrane proteins, 1 periplasm protein, 3 inner membrane proteins and 2 cytoplasmic proteins, a distribution more closely resembling living biomass than detritus that is rich in membrane proteins.

Several proteins not previously identified in POM were observed in the Sargasso Sea samples. Peptides from fish muscle-derived myosin and the collagen-hydrolyzing proteases deseasin MCP-01 and metalloprotease Mcp02 from g-proteobacteria were co-detected in the 300 and 500m samples. This suggests a close coupling between vertebrate detritus and bacterial activity, and marks the first time that deseasin or metalloprotease Mcp02 have been detected in the water column, having previously been found only in deep sea sedimentary γ-proteobacteria (Zhao et al. 2008).

A nitrogenase iron protein used by diazotropic marine cyanobacteria for fixing N was detected in the euphotic zone of the Sargasso Sea. The putative source organism, *Crocosphaera sp.*, fix N_2 during the night and are thought to up regulate nitrogenases accordingly (Webb et al. 2009; Sohm et al. 2011). As our samples were collected at night, the unique identification of this nitrogenase demonstrates the ability of this analytical technique to be used to monitor *in situ* biological adaptation. We also detected the carbon dioxide concentrating protein CcmK, carbonic anhydrase, and a circadian clock protein, KaiC, which is attributable to cyanobacteria and may indeed be sourced from *Crocosphaera*. Thus, both C and N fixation may be occurring at night in the Sargasso Sea and these reactions appear to be tractable using a metaproteomic approach. The two other proteins not previously detected in POM are oxygen stress relief proteins found in euphotic zone waters; iron superoxide dismutase (Tropical North Pacific) and thioredoxin peroxidase (Sargasso Sea).

Yamada and Tanoue (2009) suggested that coastal and offshore dissolved proteins were similar in character (function, source, etc). In order to assess this premise for particulate

425 proteins, this dataset was evaluated and then pooled with that of Dong et al (2012; oligotrophic
426 South China Sea) and Moore et al (coastal Bering Sea) to create a larger combined data set. In
427 this dataset, there is little overlap in the proteins detected between the coastal Washington
428 margin and the Pacific or Atlantic oligotrophic sites (only 6 and 9 proteins, respectively).
429 However, 22 proteins are shared between the Pacific and Atlantic oligotrophic sites. Eight of
430 these are from cyanobacteria and seven from heterotrophic bacteria. In the larger data set, there
431 is again better overlap between proteins of the oligotrophic sites (19 proteins shared) and the two
432 coastal sites (22 proteins shared) than there is between the oligotrohic and coastal sites (6
433 proteins shared). This is likely a reflection of the different organisms that anchor the food chains
434 in these environments; diatoms in coastal waters and cyanobacteria in the oligotrophic seas.
435 Among the oligotrophic-specific proteins is the methyl-accepting chemotaxis protein from γ-
436 proteobacteria. It was detected in all three oligotrophic waters, but not along the Washington
437 margin or in the Bering Sea. Methyl-accepting chemotaxis protein (MCP) is a transmembrane
438 sensor protein that allows bacteria to detect concentrations of molecules in the extracellular
439 matrix so that the bacteria may smooth swim or tumble accordingly. These types of proteins
440 may not be as abundant in POM-rich coastal waters.
441 Although the Sargasso Sea dataset is our largest, the best overlap of oligotrophic
442 environments was between the Tropical North Pacific and the South China Sea. Proteins
443 common to these two sites include beta actin (eukaryotic source), elongation factor Tu 1 (γ-
444 proteobacteria), GroEL chaperonin (cyanobacteria) and NAD(P)H-quinone oxidoreductase
445 (cyanobacteria). Urea ABC transporters (cyanobacterial) were found in the tropical North
446 Pacific and the Sargasso Sea, but not the South China Sea.
447 Despite the differences in the exact overlap of the proteins, the general sources
448 (phytoplankton, bacteria, heterotrophic eukaryotes, etc), functions and cellular locations of the
449 detected proteins vary little between the coastal and oligotrophic sites (e.g. Figs 3-5). Proteins in
450 surface waters at all locations are 60-80% derived from phytoplankton and this value decreased
451 by half in deep water samples. Bacterial proteins become proportionally more important with
452 depth, comprising 50-80% of the proteins at depth. There may be a difference between
453 oligotrophic and coastal sites with regard to bacterial proteins, as our Washington margin
454 samples contain proportionally more α-proteobacterial proteins than the oligotrophic sites, which
455 are dominated by γ-proteobacterial proteins at depth (Fig 4). Yin et al (2013) and Van Mooy et

al (2004) showed that α-proteobacteria can be dominant in the oligotrophic Pacific, while γ-proteobacteria such as SAR11 and SAR83 are dominant in the oligotrophic Atlantic (Morris et al. 2002; Sowell et al. 2009).

Previously, Moore et al (2012a) stated that while functional information on proteins can be attained with confidence, taxonomic information can not be gained when performing metaprotoemics on ocean samples with no corresponding metagenome. This study, like Morris et al (2010), demonstrated that while the level of specificity for taxonomic assignment to a particular species is not quite developed, due to a lack of completed genomes, certain high-level taxonomic ranks can be confidently assigned. When trying to decipher the complex global biogeochemical cycles of carbon and nitrogen in the ocean, phyla, division or class--level taxonomic information provides links for tracking the transfer of nutrients from the upper ocean to the deep. The exponential increase of genomic sequencing in the ocean will soon allow studies to reveal more specific taxonomic information. If needed, proteomics data, such as is presented here, can always be reanalyzed as genomes get added.

Although there is relatively little overlap in the exact proteomes recovered between any given location or depth, the observed consistency in finding phytoplankton-derived membrane-bound proteins at depth implies a strong role for membrane-binding and/or protein complexes in transferring phytoplankton-produced proteins to depth, regardless of whether the phytoplankton are diatoms or cyanobacteria. Interestingly, this does not mandate a role for ballast (e.g. silicate frustules, etc.) in transferring the material rapidly to depth (Engel et al. 2009; Moore et al. 2012b) because the cyanobacteria in the open ocean stations contain no natural ballast. Further, the relative increase of non-photosynthetic bacterial proteins with depth is an indication of their role in remineralizing POM, and the lack of bacterial membrane proteins over other bacterial functional proteins implies that the detected bacterial proteins mark living biomass and not detritus. These findings highlight the potential role of proteomics analyses in understanding the geochemical processes promoting carbon and nitrogen cycling in the sea.

5.0 Acknowledgements.

This research was funded by NSF grants 0726522, 0926395 and 1153935 to RGK, 1029281 to AEI and 1233014 to BLN. We thank Allison Myers-Pigg, Brittany Forrest, Chloe Anderson and Anna Belcher for help in the lab and at sea, Priska von Haller and the UWRP for advice on the protocols and analyses and the captains and crews of the research vessels Wecoma, Kilo Moana and Atlantic Explorer for their help collecting samples and deploying instrumentation. This work is supported in part by the University of Washington's Proteomics Resource (UWPR95794).

References

Alonso-Gonzalez IJ, Aristegui J, Lee C, Sanchez-Vidal A, Calafat A, Fabres J, Sangra P, Mason E (2013) Carbon Dynamics within Cyclonic Eddies: Insights from a Biomarker Study. Plos One **8**:10. doi:10.1371/journal.pone.0082447

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. Science **306**:79-86. doi:10.1126/science.1101156

Buesseler KO, Boyd PW (2009) Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. Limnology and Oceanography **54**:1210-1232. doi:10.4319/lo.2009.54.4.1210

Canfield DE, Boudreau BP, Mucci A, Gundersen JK (1998) The early diagenetic formation of organic sulfur in the sediments of Mangrove Lake, Bermuda. Geochimica Et Cosmochimica Acta **62**:767-781. doi:10.1016/s0016-7037(98)00032-5

Chou KC, Shen HB (2008) Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. Nature Protocols **3**:153-162. doi:10.1038/nprot.2007.494

Christie-Oleza JA, Pina-Villalonga JM, Bosch R, Nogales B, Armengaud J (2012) Comparative Proteogenomics of Twelve Roseobacter Exoproteomes Reveals Different Adaptive Strategies Among These Marine Bacteria. Molecular & Cellular Proteomics **11**:12. doi:10.1074/mcp.M111.013110

Conte MH, Dickey TD, Weber JC, Johnson RJ, Knap AH (2003) Transient physical forcing of pulsed export of bioreactive material to the deep Sargasso Sea. Deep-Sea Research Part I-Oceanographic Research Papers **50**:1157-1187. doi:10.1016/s0967-0637(03)00141-9

Dekker JP, Boekema EJ (2005) Supramolecular organization of thylakoid membrane proteins in green plants. Biochimica Et Biophysica Acta-Bioenergetics **1706**:12-39. doi:10.1016/j.bbabio.2004.09.009

Dong H-P, Wang D-Z, Dai M, Chan LL, Hong H-S (2009) Shotgun proteomics: Tools for analysis of marine particulate proteins. Limnology and Oceanography-Methods **7**:865-874

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

Dyhrman ST, Jenkins BD, Rynearson TA, Saito MA, Mercier ML, Alexander H, Whitney LP, Drzewianowski A, Bulygin VV, Bertrand EM, Wu ZJ, Benitez-Nelson C, Heithoff A (2012) The Transcriptome and Proteome of the Diatom Thalassiosira pseudonana Reveal a Diverse Phosphorus Stress Response. Plos One **7**. doi:10.1371/journal.pone.0033768

Eng JK, Fischer B, Grossmann J, MacCoss MJ (2008) A Fast SEQUEST Cross Correlation Algorithm. Journal of Proteome Research **7**:4598-4602. doi:10.1021/pr800420s

536 Engel A, Abramson L, Szlosek J, Liu ZF, Stewart G, Hirschberg D, Lee C (2009) Investigating
537 the effect of ballasting by CaCO3 in Emiliania huxleyi, II: Decomposition of particulate
538 organic matter. Deep-Sea Research Part Ii-Topical Studies in Oceanography **56**:1408-
539 1419. doi:10.1016/j.dsr2.2008.11.028
540 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005)
541 Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The
542 Proteomics Protocols Handbook. Humana Press, Totowa, New Jersey 07512, p 969
543 Hannides CCS, Popp BN, Choy CA, Drazen JC (2013) Midwater zooplankton and suspended
544 particle dynamics in the North Pacific Subtropical Gyre: A stable isotope perspective.
545 Limnology and Oceanography **58**:1931-1946. doi:10.4319/lo.2013.58.6.1931
546 Hedges JI, Baldock JA, Gelinas Y, Lee C, Peterson M, Wakeham SG (2001) Evidence for non-
547 selective preservation of organic matter in sinking marine particles. Nature **409**:801-804.
548 doi:10.1038/35057247
549 Jones BM, Edwards RJ, Skipp PJ, O'Connor CD, Iglesias-Rodriguez MD (2011) Shotgun
550 Proteomic Analysis of Emiliania huxleyi, a Marine Phytoplankton Species of Major
551 Biogeochemical Importance. Marine Biotechnology **13**:496-504. doi:10.1007/s10126-
552 010-9320-0
553 Kaiser K, Benner R (2008) Major bacterial contribution to the ocean reservoir of detrital organic
554 carbon and nitrogen. Limnology and Oceanography **53**:99-112.
555 doi:10.4319/lo.2008.53.1.0099
556 Kaiser K, Benner R (2009) Biochemical composition and size distribution of organic matter at
557 the Pacific and Atlantic time-series stations. Marine Chemistry **113**:63-77.
558 doi:10.1016/j.marchem.2008.12.004
559 Kaiser K, Benner R (2012) Organic matter transformations in the upper mesopelagic zone of the
560 North Pacific: Chemical composition and linkages to microbial community structure.
561 Journal of Geophysical Research-Oceans **117**:12. doi:10.1029/2011jc007141
562 Karner MB, DeLong EF, Karl DM (2001) Archaeal dominance in the mesopelagic zone of the
563 Pacific Ocean. Nature **409**:507-510. doi:10.1038/35054051
564 Kawasaki N, Sohrin R, Ogawa H, Nagata T, Benner R (2011) Bacterial carbon content and the
565 living and detrital bacterial contributions to suspended particulate organic carbon in the
566 North Pacific Ocean. Aquatic Microbial Ecology **62**:165-176. doi:10.3354/ame01462
567 Keil RG, Kirchman DL (1993) Dissolved combined amino acids - chemical forma and utilization
568 by marine bacteria. Limnology and Oceanography **38**:1256-1270
569 Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate
570 the accuracy of peptide identifications made by MS/MS and database search. Analytical
571 Chemistry **74**:5383-5392. doi:10.1021/ac025747h
572 Lee C, Armstrong RA, Cochran JK, Engel A, Fowler SW, Goutx M, Masque P, Miquel JC,
573 Peterson M, Tamburini C, Wakeham S (2009a) MedFlux: Investigations of particle flux
574 in the Twilight Zone. Deep-Sea Research Part Ii-Topical Studies in Oceanography
575 **56**:1363-1368. doi:10.1016/j.dsr2.2008.12.003
576 Lee C, Peterson ML, Wakeham SG, Armstrong RA, Cochran JK, Miquel JC, Fowler SW,
577 Hirschberg D, Beck A, Xue JH (2009b) Particulate organic matter and ballast fluxes
578 measured using time-series and settling velocity sediment traps in the northwestern
579 Mediterranean Sea. Deep-Sea Research Part Ii-Topical Studies in Oceanography
580 **56**:1420-1436. doi:10.1016/j.dsr2.2008.11.029

581 Lee C, Wakeham S, Arnosti C (2004) Particulate organic matter in the sea: The composition
582 conundrum. Ambio **33**:565-575. doi:10.1639/0044-7447(2004)033[0565:pomits]2.0.co;2
583 Liu HB, Sadygov RG, Yates JR (2004) A model for random sampling and estimation of relative
584 protein abundance in shotgun proteomics. Analytical Chemistry **76**:4193-4201.
585 doi:10.1021/ac0498563
586 Lomstein BA, Langerhuus AT, D'Hondt S, Jorgensen BB, Spivack AJ (2012) Endospore
587 abundance, microbial growth and necromass turnover in deep sub-seafloor sediment.
588 Nature **484**:101-104. doi:10.1038/nature10905
589 Long RA, Azam F (1996) Abundant protein-containing particles in the sea. Aquatic Microbial
590 Ecology **10**:213-221. doi:10.3354/ame010213
591 Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, Chisholm SW
592 (2010) Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific
593 oceans. Isme Journal **4**:1252-1264. doi:10.1038/ismej.2010.60
594 Mattes TE, Nunn BL, Marshall KT, Proskurowski G, Kelley DS, Kawka OE, Goodlett DR,
595 Hansell DA, Morris RM (2013) Sulfur oxidizers dominate carbon fixation at a
596 biogeochemical hot spot in the dark ocean. Isme Journal **7**:2349-2360.
597 doi:10.1038/ismej.2013.113
598 Maupin-Furlow JA, Humbard MA, Kirkland PA (2012) Extreme challenges and advances in
599 archaeal proteomics. Current Opinion in Microbiology **15**:351-356.
600 doi:10.1016/j.mib.2012.02.002
601 Moore EK, Nunn BL, Faux JF, Goodlett DR, Harvey HR (2012a) Evaluation of electrophoretic
602 protein extraction and database-driven protein identification from marine sediments.
603 Limnology and Oceanography-Methods **10**:353-366. doi:10.4319/lom.2012.10.353
604 Moore EK, Nunn BL, Goodlett DR, Harvey HR (2012b) Identifying and tracking proteins
605 through the marine water column: Insights into the inputs and preservation mechanisms
606 of protein in sediments. Geochimica Et Cosmochimica Acta **83**:324-359.
607 doi:10.1016/j.gca.2012.01.002
608 Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G (2010) Comparative
609 metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy
610 transduction. Isme Journal **4**:673-685. doi:10.1038/ismej.2010.4
611 Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ
612 (2002) SAR11 clade dominates ocean surface bacterioplankton communities. Nature
613 **420**:806-810. doi:10.1038/nature01240
614 Nagata T, Fukuda R, Koike I, Kogure K, Kirchman DL (1998) Degradation by bacteria of
615 membrane and soluble protein in seawater. Aquatic Microbial Ecology **14**:29-37.
616 doi:10.3354/ame014029
617 Nagata T, Kirchman DL (1996) Bacterial degradation of protein adsorbed to model submicron
618 particles in seawater. Marine Ecology Progress Series **132**:241-248.
619 doi:10.3354/meps132241
620 Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying
621 proteins by tandem mass spectrometry. Analytical Chemistry **75**:4646-4658.
622 doi:10.1021/ac0341261
623 Nguyen RT, Harvey HR (1997) Protein and amino acid cycling during phytoplankton
624 decomposition in oxic and anoxic waters. Organic Geochemistry **27**:115-128.
625 doi:10.1016/s0146-6380(97)00076-4

626 Nguyen RT, Harvey HR (2001) Preservation of protein in marine systems: Hydrophobic and
627 other noncovalent associations as major stabilizing forces. Geochimica Et Cosmochimica
628 Acta **65**:1467-1480. doi:10.1016/s0016-7037(00)00621-9
629 Nunn BL, Aker JR, Shaffer SA, Tsai Y, Strzepek RF, Boyd PW, Freeman TL, Brittnacher M,
630 Malmstroem L, Goodlett DR (2009) Deciphering diatom biochemical pathways via
631 whole-cell proteomics. Aquatic Microbial Ecology **55**:241-253. doi:10.3354/ame01284
632 Nunn BL, Faux JF, Hippmann AA, Maldonado MT, Harvey HR, Goodlett DR, Boyd PW,
633 Strzepek RF (2013) Diatom Proteomics Reveals Unique Acclimation Strategies to
634 Mitigate Fe Limitation. Plos One **8**:16. doi:10.1371/journal.pone.0075653
635 Nunn BL, Norbeck A, Keil RG (2003) Hydrolysis patterns and the production of peptide
636 intermediates during protein degradation in marine systems. Marine Chemistry **83**:59-73.
637 doi:10.1016/s0304(03)00096-3
638 Nunn BL, Ting YS, Malmstrom L, Tsai YS, Squier A, Goodlett DR, Harvey HR (2010) The path
639 to preservation: Using proteomics to decipher the fate of diatom proteins during
640 microbial degradation. Limnology and Oceanography **55**:1790-1804.
641 doi:10.4319/lo.2010.55.4.1790
642 Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W,
643 Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterbury J
644 (2003) The genome of a motile marine Synechococcus. Nature **424**:1037-1042.
645 doi:10.1038/nature01943
646 Poretsky RS, Sun SL, Mou XZ, Moran MA (2010) Transporter genes expressed by coastal
647 bacterioplankton in response to dissolved organic carbon. Environmental Microbiology
648 **12**:616-627. doi:10.1111/j.1462-2920.2009.02102.x
649 Roland LA, McCarthy MD, Guilderson T (2008) Sources of molecularly uncharacterized organic
650 carbon in sinking particles from three ocean basins: A coupled Delta C-14 and delta C-13
651 approach. Marine Chemistry **111**:199-213. doi:10.1016/j.marchem.2008.05.010
652 Roth LC, Harvey HR (2006) Intact protein modification and degradation in estuarine
653 environments. Marine Chemistry **102**:33-45. doi:10.1016/j.marchem.2005.10.025
654 Saijo S, Tanoue E (2005) Chemical forms and dynamics of amino acid-containing particulate
655 organic matter in Pacific surface waters. Deep-Sea Research Part I-Oceanographic
656 Research Papers **52**:1865-1884. doi:10.1016/j.dsr.2005.05.001
657 Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M, Schulze ED (2005) A proteomic
658 fingerprint of dissolved organic carbon and of soil particles. Oecologia **142**:335-343.
659 doi:10.1007/s00442-004-1698-9
660 Smith DP, Thrash JC, Nicora CD, Lipton MS, Burnum-Johnson KE, Carini P, Smith RD,
661 Giovannoni SJ (2013) Proteomic and Transcriptomic Analyses of "Candidatus
662 Pelagibacter ubique" Describe the First P-II-Independent Response to Nitrogen
663 Limitation in a Free-Living Alphaproteobacterium. Mbio **4**:14. doi:10.1128/mBio.00133-
664 12
665 Sohm JA, Webb EA, Capone DG (2011) Emerging patterns of marine nitrogen fixation. Nature
666 Reviews Microbiology **9**:499-508. doi:10.1038/nrmicro2594
667 Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF, Giovannoni SJ
668 (2011) Environmental proteomics of microbial plankton in a highly productive coastal
669 upwelling system. Isme Journal **5**:856-865. doi:10.1038/ismej.2010.168
670 Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, Carlson CA, Smith
671 RD, Giovanonni SJ (2009) Transport functions dominate the SAR11 metaproteome at

672 low-nutrient extremes in the Sargasso Sea. Isme Journal **3**:93-105.
673 doi:10.1038/ismej.2008.83
674 Suzuki S, Kogure K, Tanoue E (1997) Immunochemical detection of dissolved proteins and their
675 source bacteria in marine environments. Marine Ecology Progress Series **158**:1-9.
676 doi:10.3354/meps158001
677 Tang K, Jiao NZ, Liu KS, Zhang Y, Li SH (2012) Distribution and Functions of TonB-
678 Dependent Transporters in Marine Bacteria and Environments: Implications for
679 Dissolved Organic Matter Utilization. Plos One **7**:16. doi:10.1371/journal.pone.0041204
680 Tanoue E (1992) Occurrence and characterization of partculate proteins in the Pacific Ocean.
681 Deep-Sea Research Part a-Oceanographic Research Papers **39**:743-761.
682 doi:10.1016/0198-0149(92)90118-d
683 Tanoue E (1995) Detection of dissolved protein molecules in oceanic waters. Marine Chemistry
684 **51**:239-252. doi:10.1016/0304-4203(95)00061-5
685 Tanoue E, Nishiyama S, Kamo M, Tsugita A (1995) BACTERIAL-MEMBRANES - POSSIBLE
686 SOURCE OF A MAJOR DISSOLVED PROTEIN IN SEAWATER. Geochimica Et
687 Cosmochimica Acta **59**:2643-2648. doi:10.1016/0016-7037(95)00134-4
688 Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for
689 genome-scale analysis of protein functions and evolution. Nucleic Acids Research **28**:33-
690 36. doi:10.1093/nar/28.1.33
691 Tsukasaki A, Tanoue E (2010) Chemical qualification of electrophoretically detectable peptides
692 and sugar chains in oceanic surface particulate organic matter. Marine Chemistry **119**:33-
693 43. doi:10.1016/j.marchem.2009.12.004
694 Van Mooy BAS, Devol AH, Keil RG (2004) Relationship between bacterial community
695 structure, light, and carbon cycling in the eastern subarctic North Pacific. Limnology and
696 Oceanography **49**:1056-1062
697 Volkman JK, Tanoue E (2002) Chemical and biological studies of particulate organic matter in
698 the ocean. Journal of Oceanography **58**:265-279. doi:10.1023/a:1015809708632
699 Wakeham SG, Lee C, Hedges JI, Hernes PJ, Peterson ML (1997) Molecular indicators of
700 diagenetic status in marine organic matter. Geochimica Et Cosmochimica Acta **61**:5363-
701 5369. doi:10.1016/s0016-7037(97)00312-8
702 Wang DZ, Dong HP, Xie ZX, Dai MH, Hong HS (2011) Metaproteomic characterization of
703 dissolved organic matter in the water column of the South China Sea. Limnology and
704 Oceanography **56**:1641-1652. doi:10.4319/lo.2011.56.5.1641
705 Webb EA, Ehrenreich IM, Brown SL, Valois FW, Waterbury JB (2009) Phenotypic and
706 genotypic characterization of multiple strains of the diazotrophic cyanobacterium,
707 Crocosphaera watsonii, isolated from the open ocean. Environmental Microbiology
708 **11**:338-348. doi:10.1111/j.1462-2920.2008.01771.x
709 Williams TJ, Long E, Evans F, DeMaere MZ, Lauro FM, Raftery MJ, Ducklow H, Grzymski JJ,
710 Murray AE, Cavicchioli R (2012) A metaproteomic assessment of winter and summer
711 bacterioplankton from Antarctic Peninsula coastal surface waters. Isme Journal **6**:1883-
712 1900. doi:10.1038/ismej.2012.28
713 Yamada N, Tanoue E (2003) Detection and partial characterization of dissolved glycoproteins in
714 oceanic waters. Limnology and Oceanography **48**:1037-1048
715 Yamada N, Tanoue E (2009) Similarity of electrophoretic dissolved protein spectra from coastal
716 to pelagic seawaters. Journal of Oceanography **65**:223-233. doi:10.1007/s10872-009-
717 0021-0

718 Yin Q, Fu BB, Li BY, Shi XC, Inagaki F, Zhang XH (2013) Spatial Variations in Microbial
719 Community Composition in Surface Seawater from the Ultra-Oligotrophic Center to Rim
720 of the South Pacific Gyre. Plos One **8**:12. doi:10.1371/journal.pone.0055148
721 Zhao, G.Y., Chen, X.L., Zhao, H.L., Xie, B.B., Zhou, B.C., Zhang, Y.Z., 2008. Hydrolysis of
722 Insoluble Collagen by Deseasin MCP-01 from Deep-sea Pseudoalteromonas sp SM9913
723 collagenolytic characters, collagen binding ability of C-terminal polycystic kidney
724 disease domain, and implication for its novel role in deep sea sedimentary particulate
725 orgnaic nitrogen degradation. Journal of Biological Chemistry 283, 36100-36107.

726

Table and Figure Legends

Table 1. Locations and water column depth at which POM was sampled and volume of water filtered using McLane pumps.

Fig. 1. Location of sampling sites for A) Northeastern Pacific (WA coast), B) North Pacific Subtropical gyre and C) the Sargasso Sea.

Fig 2. A) Number of detected proteins, B) amount of protein estimated in each sample using the NanoOrange assay, and C) relationship between protein quantity and protein detection. For the entire data set the correlation between protein abundance and protein detection (r^2 = 0.86) has a slope of 2 detected proteins per $\mu g\ L^{-1}$ of protein in the water.

Fig 3. Source organisms for the proteins identified at each location and depth. The three 500m samples in the Sargasso Sea are (top to bottom) GF/A 1.6 μm, Durapore 0.2 μm and double GF/F 0.7 μm filters.

Fig 4. Major functional types for proteins identified at each location and depth.

Fig 5. Predicted cellular location for proteins identified at each location and depth.

Fig 6. Number of cyanobacterial proteins identified at each depth for the Sargasso Sea station.

Supplemental Information

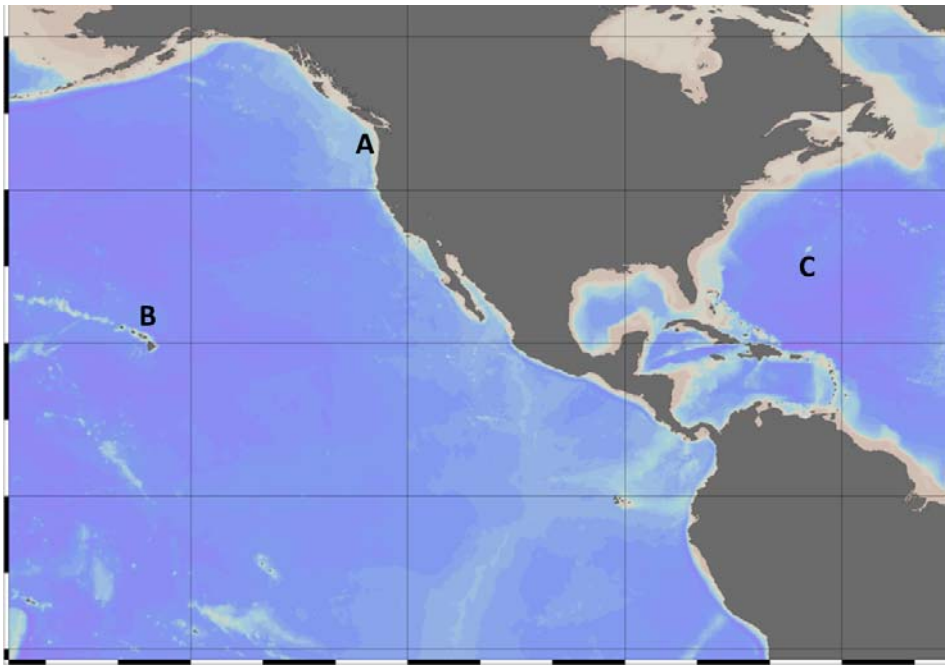
Table S1. The protein database used for searching was derived from marine planktonic communities and downloaded from the National Center for Biotechnology Information (NCBI).

Table S2. Peptide sequence, protein gi and name and % sequence coverage for all locations and depths.

754 Table 1. Locations and water column depth at which POM was sampled and volume of water
755 filtered using McLane pumps.
756

Sample No	Sampling location	Sampling date	Sampling depth (m)	Sample volume	Protein (ug L^{-1})
Northeastern Pacific (WA coast)					
P8	47º 14.973' N, 125º 22.954' W	26-Sep-10	265	661	10.1
P9	47º 14.973' N, 125º 22.954' W	30-Sep-10	500	1528	5.5
P10*	47º 12.155' N, 125º 59.538' W	23-May-10	75	2225	6.2
P11	47º 14.973' N, 125º 22.954' W	22-Sep-10	155	700	6.0
P12	47º 14.973' N, 125º 22.954' W	22-Sep-10	105	700	2.4
Subtropical gyre					
P13	22° 45.123’ N, 158° 0.2451W	17-Jul-10	50	813	8.5
P14	22° 45.123’ N, 158° 0.2451W	18-Jul-10	175	900	5.7
P15	22° 45.123’ N, 158° 0.2451W	18-Jul-10	200	900	16.9
P17	22° 45.123’ N, 158° 0.2451W	19-Jul-10	750	1188	4.7
P18	22° 45.123’ N, 158° 0.2451W	20-Jul-10	1510	1281	3.5
North Atlantic (Sargasso Sea)					
P19	33°08.744’ N, 64°21.787’ W	16-Mar-11	150	804	56.3
P20	33°07.724’ N, 64°10.426’ W	18-Mar-11	300	629	32.7
P21	33°08.744’ N, 64°21.787’ W	16-Mar-11	500	838	5.9
P22	33°08.744’ N, 64°21.787’ W	16-Mar-11	500	838	6.6
P23	33°07.724’ N, 64°10.426’ W	16-Mar-11	500	600	21.65

757
758
759 * plankton tow
760



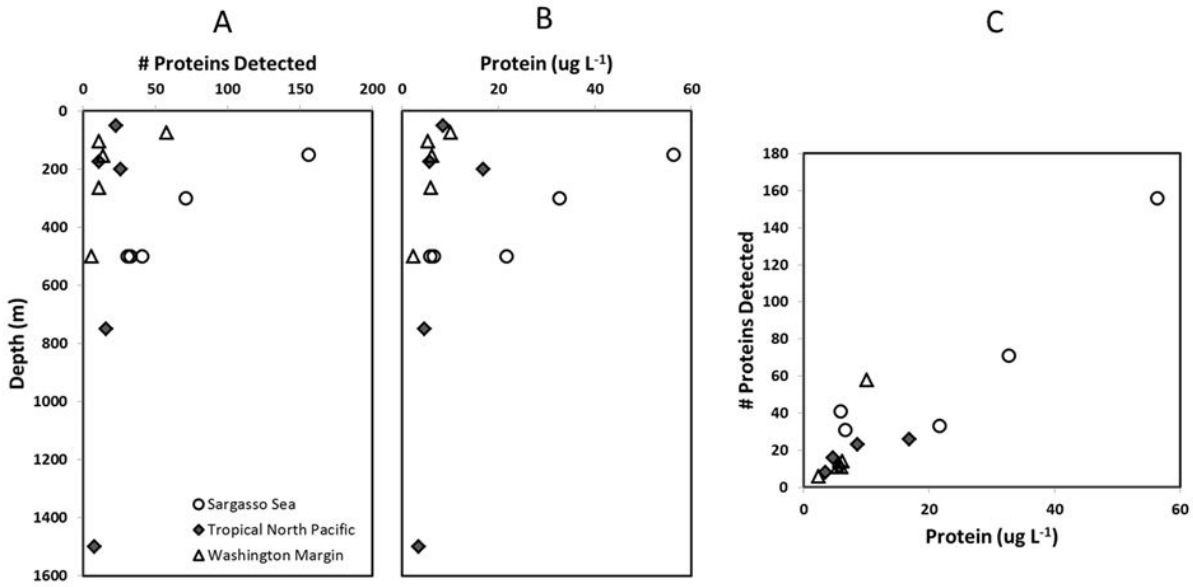
761

762

763 Fig. 1. Location of sampling sites for A) Northeastern Pacific (WA coast),

764 B) North Pacific Subtropical gyre and C) the Sargasso Sea.

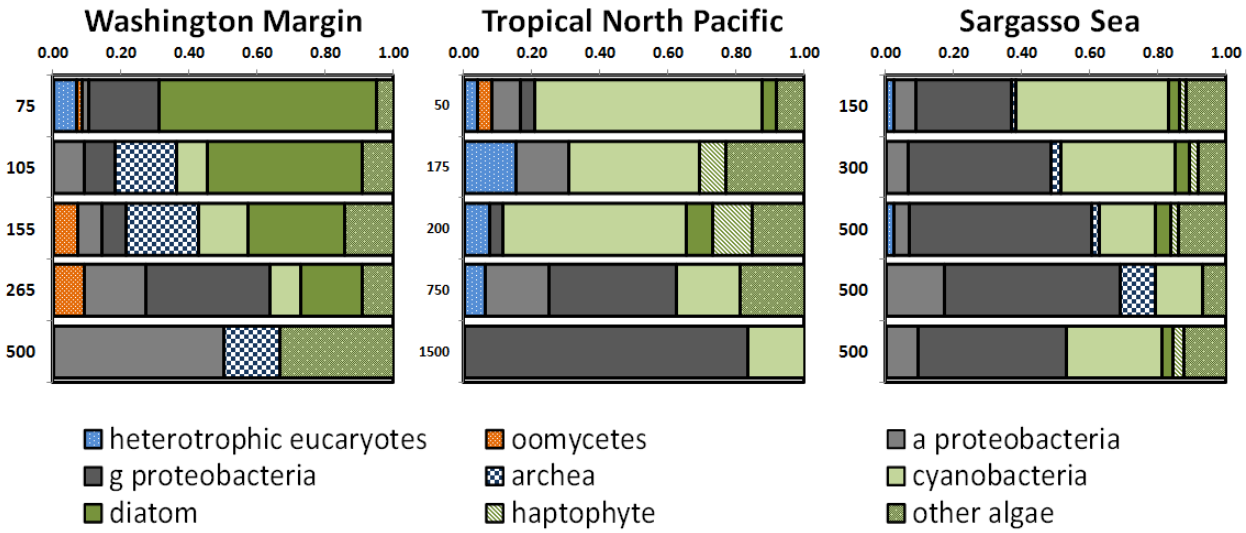
765



766
767
768

769 Fig 2. A) Number of detected proteins, B) amount of protein estimated in each sample using the
770 NanoOrange assay, and C) relationship between protein quantity and the number of distinct
771 proteins identified. For the entire data set the correlation between protein abundance and protein
772 identification ($r^2 = 0.86$) has a slope of 2 identified proteins per μg L^{-1} of protein in the water.

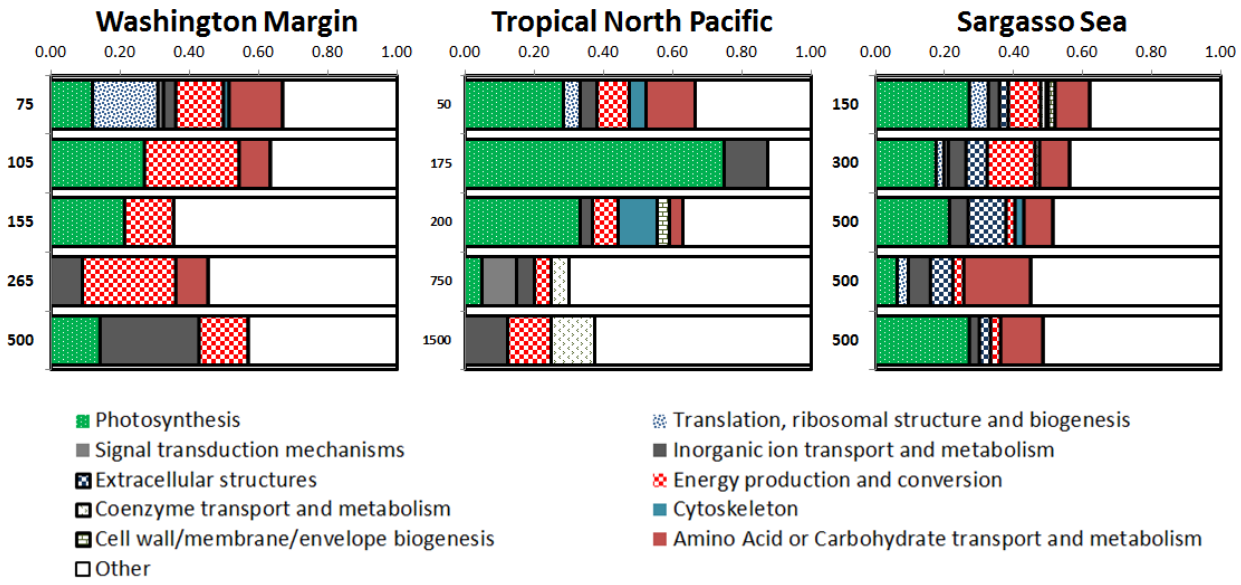
773
774



775

776 Fig 3. Source organisms for the proteins identified at each location and depth. The three 500m
777 samples in the Sargasso Sea are (top to bottom) GF/A 1.6 μm, Durapore 0.2 μm and double
778 GF/F 0.7 μm filters.

779

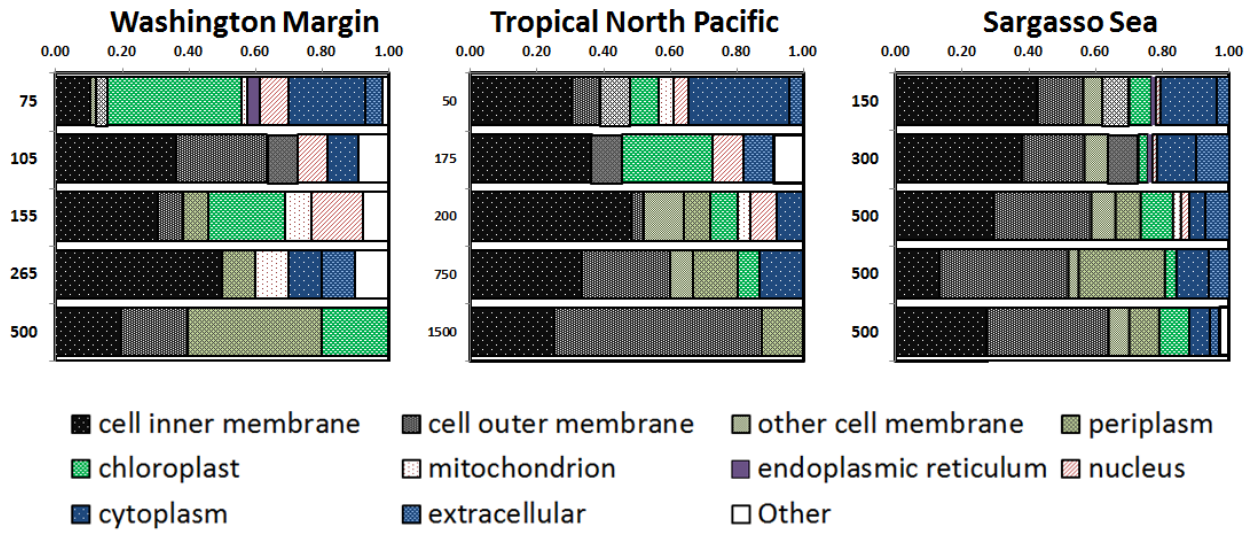


780

781

782

783 Fig 4. Major functional types for proteins identified at each location and depth.



784

785

786 Fig 5. Predicted cellular location for proteins identified at each location and depth.

787

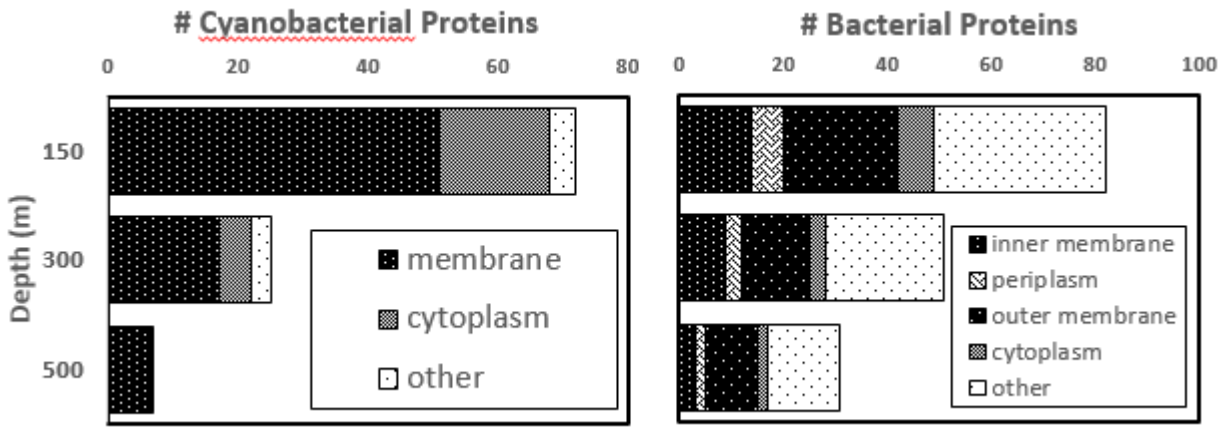


Fig 6. Number and location of cyanobacterial and heterotrophic bacterial proteins identified at each depth for the Sargasso Sea station.

Supplemental Information

Table S1. Protein database derived from marine planktonic communities and downloaded from the National Center for Biotechnology Information (NCBI).

Organism	# Proteins	% of database
Alphaproteobacteria	22365	1.0
Archaebacteria	724898	31.2
Betaproteobacteria	58267	2.5
Chlorophyta	157054	6.8
Coxiella	32973	1.4
Cryptophyta	4597	0.2
Dinoplycea	4389	0.2
Erythrobacter	18670	0.8
Flavobacteria	255479	11.0
Gammaproteobacteria	81770	3.5
Haptophycea	1271	0.1
Photobacterium	60722	2.6
Pseudomonas	711513	30.6
Stramelopiles	190610	8.2
Sum	2324578	100