2010

The Path to Preservation: Using Proteomics to Decipher the Fate of Diatom Proteins During Microbial Degradation

Brook L. Nunn

Ying S. Ting

Lars Malmström

Yihsuan S. Tsai

Angela Aquier

See next page for additional authors

Original Publication Citation

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

Limnol. Oceanogr., 55(4), 2010, 1790–1804

doi:10.4319/lo.2010.55.4.1790

The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation

Brook L. Nunn,[a,*] Ying S. Ting,[a] Lars Malmström,[a] Yihsuan S. Tsai,[a] Angela Squier,[b] David R. Goodlett,[a] and H. Rodger Harvey[b]

[a] Department of Medicinal Chemistry, University of Washington, Seattle, Washington
[b] University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory, Solomons, Maryland

Abstract

We drew upon recent advances in tandem mass spectrometry–based proteomic analyses in order to examine the proteins that remain after a diatom bloom enters the stationary phase, precipitates out of the photic zone, and is subjected to microbial degradation over a 23-d period within a controlled laboratory environment. Proteins were identified from tandem mass spectra searched against three different protein databases in order to track proteins from *Thalassiosira pseudonana* and any potential bacterial contributions. A rapid loss of diatom protein was observed over the incubation period; 75% of the proteins initially identified were not detected after 72 h of exposure to a microbial population. By the 23rd day, peptides identified with high confidence correlated with only four *T. pseudonana* proteins. Five factors may have influenced the preservation of diatom proteins: (1) protection within organelles or structures with multiple membranes, (2) the relative cellular abundance in the photosynthetic apparatus, (3) the number of transmembrane domains in the protein sequence, (4) the presence of glycan modification motifs, and (5) the capability of proteins or peptides to aggregate into supramolecules.

Marine primary production remains a crucial global conduit for long-term atmospheric CO_2 drawdown and its subsequent preservation. Proteins are the principal macromolecular contributor to the carbon biomass of phytoplankton (~ 60% of total carbon) (Wakeham et al. 1997). Although proteins have the potential to provide a wealth of information regarding biochemical health, metabolic strategies, and potential origin, our knowledge of the ways in which these macromolecules are recycled or survive degradation processes remains incomplete. Given their role as high-quality nitrogen and carbon substrates, proteins have traditionally been considered very labile in the environment. This suggests that proteins are unlikely candidates for surviving the suite of recycling reactions involved in early diagenesis. Methods for tracking protein cycling have traditionally relied on the analysis of amino-acid residues produced by the acid hydrolysis of dissolved or particulate marine samples. Through the years and across oceans, marine scientists have documented the importance of hydrolyzable amino acids to the total nitrogen pool (Lee 1988; Hedges 1991; Boski et al. 1998). Nuclear magnetic resonance (NMR) analyses of organic material prior to hydrolysis have revealed that amino acids preserved in sediments are primarily linked by amide bonds (McCarthy et al. 1997; Knicker 2000; Zang et al. 2000), suggesting the presence of peptide-like chains. Although much insight has been gained from the study of amino acids, the complete hydrolysis of organic matter irretrievably destroys information about the original protein structure and sequence. This has become a substantial roadblock to understanding the ocean's role in the geochemical processing of nitrogen. In order to consider the use of individual proteins as ''biomarkers,'' as it is done with other organic compounds in the environment (Wakeham et al. 1997; Sauer et al. 2001; Immler et al. 2006), approaches to identifying and following individual proteins through the degradation process are required. Given that any single organism may express thousands of proteins, developing a method that is applicable to ocean ecosystems is an ambitious undertaking.

While analyses of whole cells and cellular fractions are well established for the identification of protein mixtures, there are several difficulties inherent in the detection and identification of proteins from oceanic particulate material (Pantoja et al. 1997; Nguyen and Harvey 2001; Nunn and Keil 2006). Natural organic matter is a composite of multiple organisms and is subject to numerous chemical processes, which yield a concentrated mixture of dilute peptide components. It is well documented that the isolation of the protein component from marine organic matter is not straightforward (Mayer et al. 1986; Pantoja et al. 1997). Attempts to isolate proteins from the complex mixture must balance the retention of diverse proteins with the removal of humic materials, which are often characterized by similar functional groups found in proteins. In addition, there is little information about the fate of proteins or peptides soon after cellular death, making investigations of proteins preserved over longer timescales more challenging. If one can overcome these hurdles, then tracking individual proteins and their products may provide information on the origin (sequence homology) and the chemical traits (e.g., modifications, cellular location, cross-linking) that influence protein preservation over longer timescales. Given the difficulty of identifying and sequencing proteins from a large refractory pool of organic matter, the focus of this study was to examine the initial stages of algal protein degradation and survival under idealized, but realistic, laboratory conditions. We

* Corresponding author: brookh@u.washington.edu

used a proteomic strategy that required the enzymatic digestion of the degraded phytoplankton proteins into peptides prior to inline chromatographic separation and tandem mass spectrometry (MS) analysis (Nunn and Timperman 2007). Because tryptic peptides fragment in a predictable manner when using collision-induced dissociation tandem MS, a theoretical library of peptide tandem mass spectra can be generated in silico from a protein database. Observed tandem mass spectra can then be correlated with these theoretical peptide spectra, generating a list of proteins present in the sample. Since this correlative MS method is dependent in part on the completeness and accuracy of the protein database used, this study tracks proteins from a diatom with a completed genome. This method is more sensitive than top-down protein mass spectrometry methods (Tsai et al. 2009) and is more applicable in environmental systems where unpredictable degradation cleavage points may occur on a protein.

Several studies have suggested that diatoms may contribute a large fraction of organic matter to sediment systems as a consequence of their ballast-like silica frustules (Dunne et al. 2005; Ragueneau et al. 2006; Miki et al. 2009). Stationary-phase *Thalassiosira pseudonana* cells were monitored as they experienced a simulated sedimentation out of the euphotic zone followed by degradation by an oceanic bacterial assemblage. The goal of this research was to then identify the proteins in the particulate phase that were able to survive bacterial attack and oxidation so that we may gain knowledge about sedimenting proteins in the ocean. Structure, cellular location, and the functionality of each of the proteins that remained after 23 d of exposure to microbial degradation were then examined in order to identify the mechanisms that might have been responsible for their enhanced survival.

Recently, the first whole-cell proteome was completed for the diatom *T. pseudonana* (Nunn et al. 2009), providing a foundation and natural starting point from which to examine phytoplankton protein degradation through time. Importantly, this proteomic standard provides a quantitative rank order of expressed proteins in *T. pseudonana*, allowing for an assessment of whether the abundance of an individual protein possibly contributes to detection using tandem mass spectrometry. Cellular preparations, enzymatic digestions, and mass spectrometry for this study were identical to Nunn et al. (2009), which permitted direct comparison of peptide assignments to spectra from the two experiments. In addition, the well-annotated proteome provided us with ancillary data with which we could correlate the suite of proteins identified with protein structures, cellular locations, and functions. An examination of physicochemical trends observed in the remaining proteins after extensive microbial degradation provided information on the potential mechanisms responsible for sediment burial and preservation.

Methods

Algal culture and degradation experiment—The diatom *Thalassiosira pseudonana* (CCMP 1335) was cultured in a 40-liter glass vessel containing f/2 medium prepared with 0.2-μm filtered and autoclaved estuarine water from Chesapeake Bay (salinity 20). Cultures were maintained axenically under constant (19°C) temperature with a 12 : 12 h light : dark lighting regime provided by cool white fluorescent bulbs and continual aeration with sterile air. Cell growth was monitored by cell counts. Once the culture reached early stationary phase, it was placed in darkness for 5 d in order to mimic sedimentation out of the euphotic zone. After the 5 d of darkness, time point 0 was collected and the degradation experiment was started. The flow-through experimental system used to follow protein degradation was similar to the system employed previously to establish the kinetics of algal carbon decay (Harvey et al. 1995). It was designed to simulate diatoms falling through the water column while providing continual exposure to a mixed microbial community. The flow-through system provided a steady influx of seawater from Chesapeake Bay. The water was prefiltered to 3 μm to ensure that no additional macro-algae were introduced during the experimental period. The flow rate of incoming water was set to provide an equivalent of 5% daily water exchange and allow for continual exposure to the bacterial assemblage. Aerobic conditions were maintained throughout the 30-d degradation period using continual aeration. At defined time points, samples were withdrawn after resuspending diatoms by gently stirring the water in the incubation vessel. Particulate samples were analyzed for organic carbon, nitrogen, amino acids, proteins, and bacterial abundances. For this study, the dissolved organic material was not monitored.

Particulate organic carbon and nitrogen were collected by filtration through combusted (450°C for 4 h) GF/F filters and were quantified using standard methods (Bratback 1987). For bacterial counts, a protocol modified from Velji and Albright (1993) was followed, and samples were fixed with 5% (v/v) 37% formaldehyde and stored in the dark at 4°C until filtration and staining with 4′, 6-diamidino-2-phenylindole (DAPI, Sigma). Microscopic counts were conducted using an Olympus BH2 scope with a 100× objective lens and a 10 × 10 ocular grid; at least 10 randomly chosen fields of view were counted per slide.

Samples for protein characterization and amino-acid analysis relied on centrifugation of sampled waters (17,000 × *g*) in order to provide particulate material for analysis. Total protein concentrations were determined after the sonication of the particulate material in 50 mmol L^{-1} ammonium bicarbonate using the Bradford assay (Bradford 1976).

Amino-acid analysis—Individual amino acids were quantified and identified by gas chromatography (GC) and GC mass spectrometry (GCMS) using the EZ-fast method (Phenonomex) with minor modifications. Particulate samples were acid hydrolyzed for 2 h at 150°C using a liquid-phase technique with 6 mol L^{-1} sequanal-grade HCl and L-γ-methyl-leucine as the recovery standard (Grutters et al. 2001). Following hydrolysis, amino acids were quantified using an Agilent 6890 GC; samples were injected at 250°C and separated through a fused-silica GC column (Phenomenex ZB-AAA; 10 m × 0.25 mm internal diameter [i.d.])

using helium as the carrier gas. The oven was ramped from an initial temperature of 110°C, at a rate of 30°C min^{-1}, to a final temperature of 320°C. For amino-acid identifications, the GC was coupled to an Agilent 5973N mass spectrometer, operating under the same conditions using helium as the carrier gas. Spectra with masses ranging from 50 to 600 daltons (Da) were acquired. Bovine serum albumin (BSA) was analyzed in parallel to correct for different responses among individual amino acids and to calculate molar ratios.

Cellular preparations for mass spectrometry—For protein identifications, particulate detrital material was pelleted using centrifugation (17,000 × *g*) from 0.1-liter aliquots. Cellular debris was lysed using a 100-W, 20-KHz Measuring and Scientific Equipment (MSE) sonicater with a titanium microtip (20 × 10 s, 4°C), resulting in a whole-debris lysate. Prior to enzyme digestions, an aliquot from each sample was analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis for quality control. Two enzymes were used on sample splits to yield peptides for identifications using tandem MS. The first digestion used trypsin following the approach of Nunn et al. (2009). Pelleted material was sonicated and solubilized in urea, followed by the addition of Tris-HCl (pH 8.8) and TCEP (37°C, 1 h). Disulfide bonds were reduced with dithiothreitol (DTT), alkylated with iodoacetamide (IAM), vortexed, and stored in the dark for 1 h (25°C). The IAM was then neutralized with additional DTT for 1 h (25°C). Ammonium bicarbonate (50 mmol L^{-1}) was added to each digest in order to dilute the urea prior to the addition of methanol and sequencing-grade trypsin (Promega) at 50 : 1 substrate : enzyme (w/w). Digests were incubated overnight at 37°C. Samples were then taken to near dryness in a speedvac and stored at −80°C. A second enzyme, endoproteinase (endo) Glu-C from *Staphylococcus aureus* V8, was used on the second sample split in order to increase protein sequence coverage. Lyophilized endoproteinase Glu-C was resuspended in 25 mmol L^{-1} ammonium bicarbonate and added to protein lysates at a ratio of 20 : 1 substrate : enzyme (w/w) for 24 h at 25°C (pH 8.0). Just prior to all MS analyses, enzymatic digests were desalted using a microspin liquid chomatography (LC) C18 column (NestGroup), following the manufacturer's guidelines, dried down in a speedvac, and resuspended in 5% acetonitrile (ACN), 0.1% formic acid.

Mass spectrometry—The strategy for protein analysis used in this study is commonly referred to as "shotgun proteomics." This method involves an enzymatic digest of the mixture of proteins, followed by LC-MS/MS analysis. Peptides are eluted off the chromatography column and subsequently ionized before entering the mass spectrometer. A survey of the peptides that entered the ion cyclotron resonance cell (orbitrap) is performed, and the five most intense ions from that scan are selected for collision-induced dissociation (CID) and tandem mass spectral (MS2) detection. This standard method uses data-dependent ion selection to determine which ions to subject to CID. The advantages of this method include limited human intervention during high performance liquid chromatography (HPLC) introduction and the rapid duty cycle that can provide unbiased sampling of the most abundant molecules in mixtures (for review and applications, *see* Nunn and Timperman 2007). Each acquired tandem mass spectrum is subsequently searched against a sequence database to correlate theoretical peptide fragmentation patterns with the observed ions. These correlations are then evaluated statistically with PeptideProphet and ProteinProphet. In cases where these probability-based scores indicate high confidence, protein identifications are considered correct. In this study, MS1 scans were acquired in the orbitrap with very high mass accuracy (< 1 mmu), which provided additional confidence in the peptide and protein identifications.

Samples for proteomic analysis were introduced into the linear ion trap–orbitrap (LTQ-OT) hybrid mass spectrometer (Thermo Fisher) using a NanoAcquity HPLC (Waters Corporation) connected to an analytical column; i.e., LC-MS/MS. The home-packed analytical column consisted of an 11-cm-long, 75-μm i.d. fused-silica capillary column packed with C18 particles (Magic C18AQ, 100Å, 5μm; Michrom) coupled to a 2-cm-long, 100-μm i.d. precolumn (Magic C18AQ, 200Å, 5μm; Michrom). For each LC-MS/MS analysis, peptides from 1 μg of the protein digests were injected. Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-acetonitrile gradient (5–35% acetonitrile 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 (mass to charge [*m*/*z*] 350–2000). Tandem mass spectra were collected on the five most intense ions observed in the MS1 precursor scan. Particulate samples at time 0 provided the greatest amount of protein and allowed for additional MS analyses using gas-phase fractionation (GPF). Gas-phase fractionation is a technique that enhances protein identifications by performing repeat analyses of the sample across several narrow, but overlapping, mass to charge (*m*/*z*) ranges (e.g., 500–600), rather than one wide *m*/*z* range (e.g., 350–2000) (Nunn et al. 2006).

Searching proteome databases and data interpretation—All mass spectral results were interpreted and searched with an in-house copy of SEQUEST (PVM version 27 20070905; Eng et al. 1994, 2008). SEQUEST is correlative data-interpretation software that matches observed spectra to theoretical spectra generated from in silico digests of the provided protein sequences. All data searches were performed without specifying the enzyme; this allowed for the discovery of peptides resulting from trypsin, endo–Glu C, or other unknown proteolytic enzymes present in the system. Fixed modifications of 57 Da on cysteine (resulting from the IAM modification) and 16 Da on methionine (oxidation) were allowed. Minimum protein and peptide probability thresholds were set at 90% on ProteinProphet and PeptideProphet for searches completed in SEQUEST (Keller et al. 2002). Single peptides that were confidently identified and validated with a spectral match from midexponential growth whole-cell lysates (Nunn et al. 2009) were also assigned protein identifications.

Three different protein databases were assembled to search against and correlate with all collected mass spectra.

Each of the three databases included the latest release (version 3.0) of the nuclear *T. pseudonana* predicted protein database (11,390 proteins) from the Joint Genome Institute (Armbrust et al. 2004), and 302 *T. pseudonana* proteins from the National Center for Biotechnology Information (NCBI) Entrez Protein database, which includes predicted chloroplast and mitochondrial proteins, plus proteins of common contaminants. The results discussed herein concerning functional protein annotations were discovered using this *T. pseudonana* phytoplankton and contaminant database combined with an equal number of reverse sequences. The reverse sequences allowed the false discovery rate of spectra matches to proteins to be determined (total proteins searched: 23,384). In addition, all tandem mass spectra correlating with *T. pseudonana* peptides from samples collected at the end of the experimental period (day 23) were compared to spectra generated from peptides observed in a previously published proteomic analysis of midexponential growth *T. pseudonana* (Nunn et al. 2009). The 50,000 tandem mass spectra from our earlier study served as a *T. pseudonana* spectral library to cross-check peptide assignments made during the degradation experiment by SEQUEST. To determine whether identified proteins might have originated from the mixed bacterial assemblage present during the degradation process, a combined database was also used that included *T. pseudonana* proteins plus proteins from one heterotrophic marine bacteria (*Candidatus Pelagibacter ubique:* HTCC1062) and one autotrophic marine bacteria (*Prochlorococcus marinus*: CCMP 1986). By using two well-annotated marine bacterial proteomes, we could more confidently assign the function of bacterial proteins identified using this database. Since the assemblage of bacteria introduced to the system was unknown, a third database was used to search the observed spectra for the greatest diversity in microbial peptides. This database included the *T. pseudonana* protein database and the Global Ocean Survey (GOS) Combined Assembly Protein database, which included $>$ 6,000,000 proteins predicted from microbial environmental genomic data, though no functional annotations were available. The GOS database was provided by the Community Cyber infrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA: http://camera.calit2.net/ accessed 29 September 2008). In this large concatenated database, protein sequences from *T. pseudonana* constituted approximately 0.1% of total listed proteins, ensuring that spurious matches to diatom peptides would be very unlikely. Based on SEQUEST search results, QualScore (Nesvizhskii et al. 2006) was used to find unassigned high-quality spectra in the data set. In addition, all spectra were searched using the InsPecT program to look for chemical modifications on peptides up to 200 Da (mq-score $>$ 1 and $p < 0.05$, as described in Tanner et al. 2005).

Results

Loss of particulate carbon and nitrogen components—In the first few days after the diatoms experienced 5 d of darkness, there was a rapid decrease in measured particulate organic carbon and total amino acids (Fig. 1A,B). Over the 30-d incubation period, 92% of the total amino acids were removed from the particulate phase. The general decrease observed in total amino acids resulted from a cumulative decrease in each of the 20 amino acids measured. Only subtle changes in molar distributions of the individual amino acids were observed during the experimental period (Fig. 1C). The exception was glycine, which decreased in absolute concentration by 86%, but showed a relative enrichment from 6.8% of total amino acids on day 0 to 22.5% by day 30.

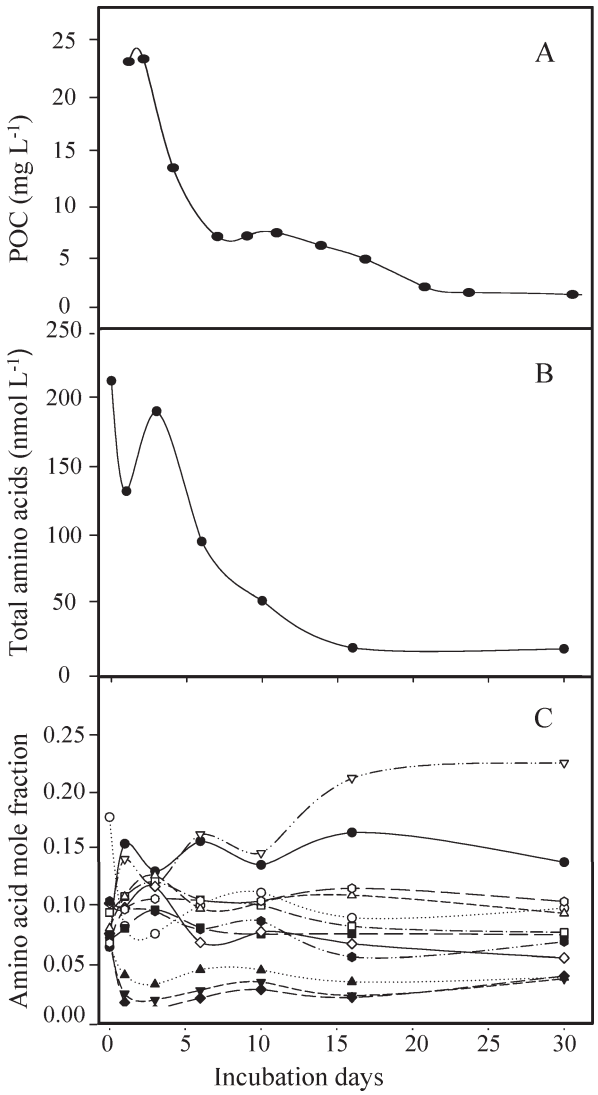

Fig. 1. Concentrations of (A) particulate organic carbon (POC) and (B) total amino acids, and (C) relative distributions of amino acids collected over the 30-d incubation of *T. pseudonana* as degradation proceeded. Solid circle, solid line: Ala; open circle, dotted line: Asp+Asn; solid upside-down triangle, dashed line: Glu+Gln; open upside-down triangle, dot–dash line: Gly; solid square, dashed line: Ile; open square, dot–dash line: Leu; closed diamond, dashed line: Lys; open diamond, solid line: Phe; closed triangle, dotted line: Pro; open triangle, dashed line: Thre; closed circle, dot–dash line: Tyr; open circle, dashed line: Val.

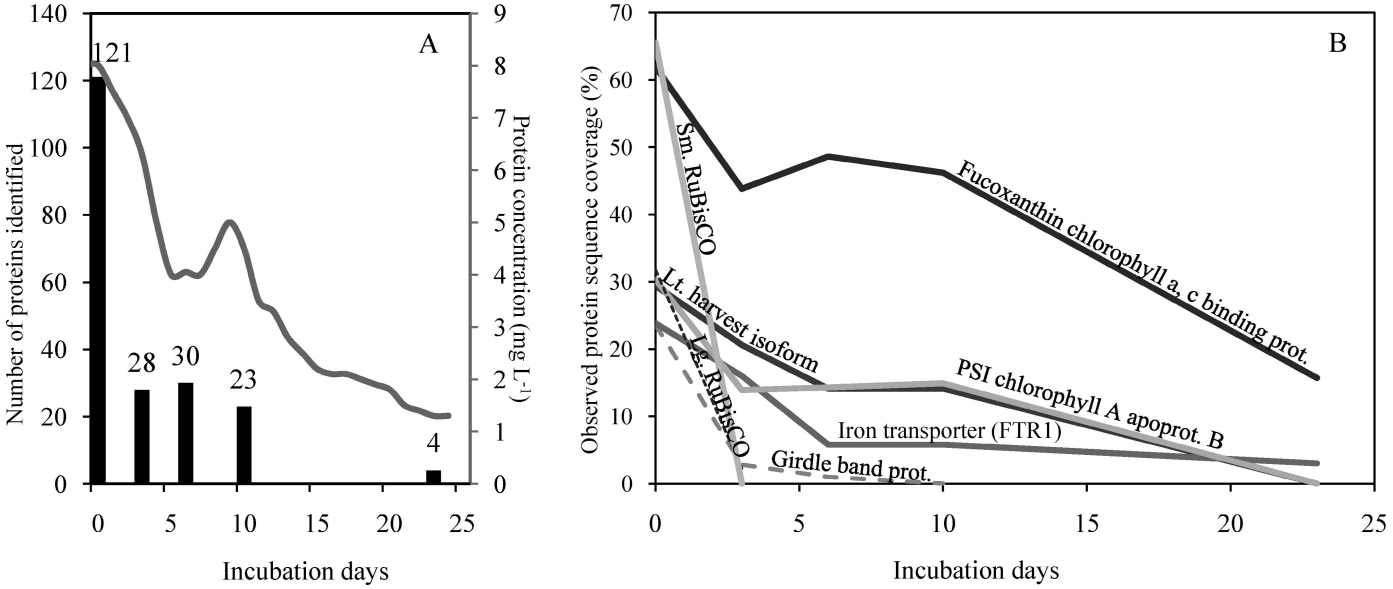

Fig. 2. (A) The time course of total particulate protein present during diatom degradation. Histogram: total number of proteins identified using LC-MS/MS; solid gray line: concentration of protein as measured by the Bradford assay. (B) Line graph tracking the percent of protein sequence detected using LC-MS/MS of seven different proteins through the degradation experiment.

Protein concentrations and identifications—Protein concentrations observed at day 0 (Fig. 2A) were the highest reported throughout the experiment. Total protein concentrations decreased from 8.05 mg L^{-1} to 1.41 mg L^{-1} (Fig. 2A). While the total number of tandem mass spectra collected per MS analysis remained constant over days 0–23, fewer spectra were confidently identified as *T. pseudonana* peptides (Table 1). In the first 3 d of bacterial degradation, 75% of the day 0 diatom protein identifications were lost, and by day 23, only four diatom proteins were detected with high confidence. This is equivalent to a 97% reduction of identified proteins over the experimental period. Gel electrophoresis was used to rapidly visualize plankton debris through the duration of the experiment. Samples collected on day 0 exhibited numerous discernible protein bands, while the samples collected later in the experiment exhibited only a few protein bands and continuous smear of stain-reactive material down the length of the gel. By the 23rd day, no distinct protein bands were apparent.

LC-MS/MS analyses identified 121 proteins from whole-cell lysates of precipitated diatom cells collected from the incubation vessel at day 0, after the 5 d of darkness. When profiled by the gene ontology (GO) classification software for biological processes, over 60% of these proteins were classified to be involved in cellular metabolism. This category includes all light-harvesting and fucoxanthin chlorophyll *a,c–binding* proteins. These proteins, and many others identified on day 0, were also affiliated with the family of proteins having chlorophyll-binding domains (PFAM). Two other well-represented protein families present at day 0 were ribosomal and adenosine triphosphate (ATP) synthase proteins.

After 3 d of exposure to the mixed bacterial assemblage, less than a quarter (28) of the proteins observed at the start of the experiment remained (Fig. 2A), and the percent sequence coverage identified decreased for those proteins that persisted from day 0. Only one new protein was detected on day 3: a glycine-rich ribonucleic acid (RNA) binding protein identified from peptides SLDIDEDIRR and LVEEDFAR. In most cases, if a peptide was the sole representative of its parent protein, it survived the degradation processes beyond day 3 and was detected on day 6. Ubiquitin, an important regulatory protein involved in targeting and tagging proteins for proteolytic degradation, was last observed on day 3 of the experiment.

Thirteen of the proteins identified on day 3 remained in the system through day 10. Most of the unique proteins identified were isoforms of, or homologous to, proteins involved in harvesting light. Although the total number of proteins from *T. pseudonana* decreased from 30 proteins at day 6 to 23 at day 10 (Fig. 2A), there was a minor increase in all measured parameters on day 10 (particulate organic carbon, amino acids, and total protein; Figs. 1A,B, 2A). The examination of all protein identifications and functions showed that only one protein, *T. pseudonana*'s photosystem II reaction center (PsbK), was uniquely identified on day 10. Over 60% of its sequence was detected. Since only one protein was unique to day 10, yet all bulk parameters increased, we examined the possibility of protein contributions by the degrading bacterial consortia (*see* following).

After 23 d of microbial degradation, only four proteins were confidently identified from a total of five unique peptides (Table 2). The four proteins identified were (1) fucoxanthin chlorophyll *a,c–binding* protein 9, (2) iron transporter protein, (3) proteophosphoglycan, and (4) an ATP-dependent Clp protease regulatory chain X.

Tracking the loss of peptides within a protein—Tracking individual peptides allowed us to examine how proteins

Table 1. Details of experimental MS data collected over the duration of the incubation.

Experimental time points	Number of MS analyses*	Number of MS2 collected		QualScore†		Annotated to *T. pseudonana* proteome		
	(*n*)	Total	Avg. per analysis	Avg. of all MS2 spectra	Avg. of annotated spectra	Total spectra	Total peptides identified	Total proteins identified
Day 0	12	43,698	3642	−1.67	0.74	807	340	121
Day 3	4	15,085	3771	−1.74	0.66	165	67	28
Day 6	4	15,192	3798	−1.66	0.78	192	81	30
Day 10	4	14,870	3718	−1.75	0.59	154	63	23
Day 23	4	16,039	4010	−1.75	0.2	10	6	4
Combined	28	104,884	3746	−1.7	0.71	1328	399	126

* Total physical injections into the LC-MS/MS for analysis.
† QualScore is a proteomic software tool that assigns a score to allow users to easily assess the quality of each spectrum collected. High QualScores are indicative of high-quality spectra.

Table 2. Diatom proteins identified from sample collected after 23 d of degradation by a natural assemblage of marine bacteria. Peptides detected using LC-MS/MS: bold peptides were identified in tryptic digest preparations, while nonbold peptides were present in endo–glu-C digest preparations.

Predicted function	Protein ID number*	Peptide sequences (day 23)	Percent sequence coverage	Number of TM helices†	Molecular weight (Da)	Number of times observed	Rank order (most abundant)‡
Fucoxanthin chlorophyll *a,c*–binding protein 9 (FCP9)	268127	E.IIPLGYDADLPIIGHLQ.- , **R.NGYIDFGWDEFDEETK.L**	15.7	0	22,628	1	1
ATP-dependent Clp proteinase regulatory chain X (Clp)	13459	R.GIVYIDEMDKIRKSGGNV…. SISRDVSGEGVQHALLKIVE.G	3.4	0	38,194	1	—
Iron transporter (FTR1)	20810	**R.SPEWQDPDKQK.A**	3	4	40,200	1	26
Proteophosphoglycan 5	3436	**R.LLEKGEAR.D**	0.7	6	129,270	2	—

* Protein ID numbers corresponds to numbers designated by the Joint Genome Institute Genome of *Thalassiosira pseudonana* (http://genome.jgi-psf.org/Thaps3/Thaps3.home.html).
† Number of transmembrane helices and domains in proteins predicted by TMHMM version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).
‡ Rank order of highly expressed proteins from *T. pseudonana* at midexponential growth (previously published in Nunn et al. 2009).

were removed from the system via degradation and illustrated the rapid loss among multiple amino-acid segments of the protein sequence. For example, both the small and large subunits of RuBisCO were identified with high confidence across large portions of the protein sequence on day 0 (32% and 65% sequence coverage, respectively). However, by day 3, no peptides from the small subunit were detected, and peptides representing only 2% of the protein sequence of the large subunit were observed. Further, by day 10, no peptides from the large subunit of RuBisCO were present. Over the course of the experiment, peptides used as diagnostic identifiers for specific proteins were observed to fall below MS detection limits. This tracking of peptides, using percent protein sequence coverage, illustrates the degradation of an individual protein (Fig. 2B). In a few cases, unique peptides remained detectable throughout the duration of the experiment (e.g., iron transporter: Fig. 2B).

Structural analyses of remaining algal proteins—To better understand mechanisms controlling the survival of these unique peptides, the secondary structures, cellular functions, and predicted locations of the proteins from which these peptides were derived were examined. Fucoxanthin chlorophyll *a,c–binding* protein 9 (FCP9) was identified in every sample collected during the degradation. These have an important function in light-harvesting complexes in diatoms since they bind different carotenoids that harvest and transfer light energy to the reaction center. FCPs are integral to thylakoid membranes and were previously revealed to be the most abundant protein in *T. pseudonana* at midexponential growth (Nunn et al. 2009). Peptides identified from FCP9 were overlaid onto a secondary structure prediction outline (Cole et al. 2008) in order to explore physicochemical property trends of proteins detected at the end of the experiment (Fig. 3). On day 0, 53 different peptides, with various cleavage points and modifications from FCP9, were observed an average of four times each, covering 62% of the 22-kDa protein sequence. By day 23, only two peptides were identified (once each), comprising 15.7% of the protein sequence. As degradation proceeded, random coil domains were more likely to be preserved than helical structures (Fig. 3). Most (60%) of the amino acids in observed peptides, via LC-MS/MS at day 0, were involved in random coil peptide structures. By day 23, this proportion had increased to 90%.

One peptide was matched to the membrane-bound iron transporter protein FTR1. At the beginning of the degradation experiment, four peptides were observed, and by days 10 and 23, only one peptide was identified. Both QualScore and manual investigation of the tandem mass spectrum for this FTR1 peptide verified that the sequence was an excellent match to the tandem mass spectrum (Fig. 4A). Mapping of the protein sequence with transmembrane (TM) prediction software (TMHMM 2.0) revealed that the only peptide remaining after 23 days of degradation was adjacent to a section of the protein that spanned the membrane (Fig. 4B).

Another membrane-bound protein identified at the end of the degradation experiment was a predicted proteophosphoglycan, which has six transmembrane (TM) domains on the C-terminus anchoring it to the membrane. As with the FTR1 protein, the peptide identified from the proteophosphoglycan was adjacent to a membrane-spanning domain (Fig. 4C).

The fourth protein identified on day 23 was an ATP-dependent Clp protease regulatory chain X. All Clp proteases require an ATP substrate, which is thought to help regulate and control protein selection for proteolysis (Gottesman et al. 1993). This protein is important in intracellular cytoplasmic proteolysis for the purpose of regulating proteins with short half-lives and removing damaged and abnormal proteins. The identified ClpX subunit is known to direct the complex to specific substrates and regulate degradation via ATP hydrolysis (Wojtkowiak et al. 1993). In order to function, two or more types of Clp protease subunits are thought to form large multimeric complexes. Portions of the Clp protease complex are known to be encoded in the nucleus, and yet the complex resides in the chloroplast. Unfortunately, attempts to predict the cellular location of *T. pseudonana*'s ClpX protein were unsuccessful. According to hydropathic models, the protein has no membrane-spanning domains and is hydrophilic.

Bacterial contributions—Over the course of the incubation, bacterial populations ranged from 1.48×10^5 cells mL^{-1} (day 0) to a maximum of 8.86×10^6 cells mL^{-1} (day 10). Based on diatom and bacterial cell counts at the maximal bacterial abundance and estimated protein concentrations (Simon and Azam 1989; Montagnes et al. 1994; Fukuda et al. 1998; Diekmann et al. 2009), up to 1000 times more protein results from diatom biomass rather than bacteria cells. While bacteria provided the active catalyst for protein hydrolysis and recycling, the focus here was on the preservation of particulate diatom proteins rather than on following bacterial protein input. To reduce the chance of bacterial debris or cells contributing to the proteomic analysis further, phytoplankton cells were isolated with a low-speed centrifugation. To cross-check the observed results for contaminating bacterial protein input, all tandem mass spectra were subsequently searched against two additional databases that included microbial proteomes. Tandem mass spectra that correlated with bacterial peptides by SEQUEST were then manually investigated. The first search was performed using a concatenated database of *Thalassiosira pseudonana*, *Candidatus Pelagibacter ubique* (HTCC1062), and *Prochlorococcus marinus* (CCMP 1986). On the first day of the degradation (day 0), only two tandem mass spectra correlated with peptides from *P. marinus* proteins: a DNA-directed RNA polymerase and a photosystem II PsbD protein. In the SEQUEST searches conducted with only the *T. pseudonana* database, the spectrum that correlated with a peptide from the *P. marinus* PsbD protein was also identified when searched against the *T. pseudonana* database. The *T. pseudonana* search also resulted in three more unique peptide identifications from the same PsbD protein. The spectrum that identified the bacterial RNA polymerase was not correlated with any *T. pseudo-*

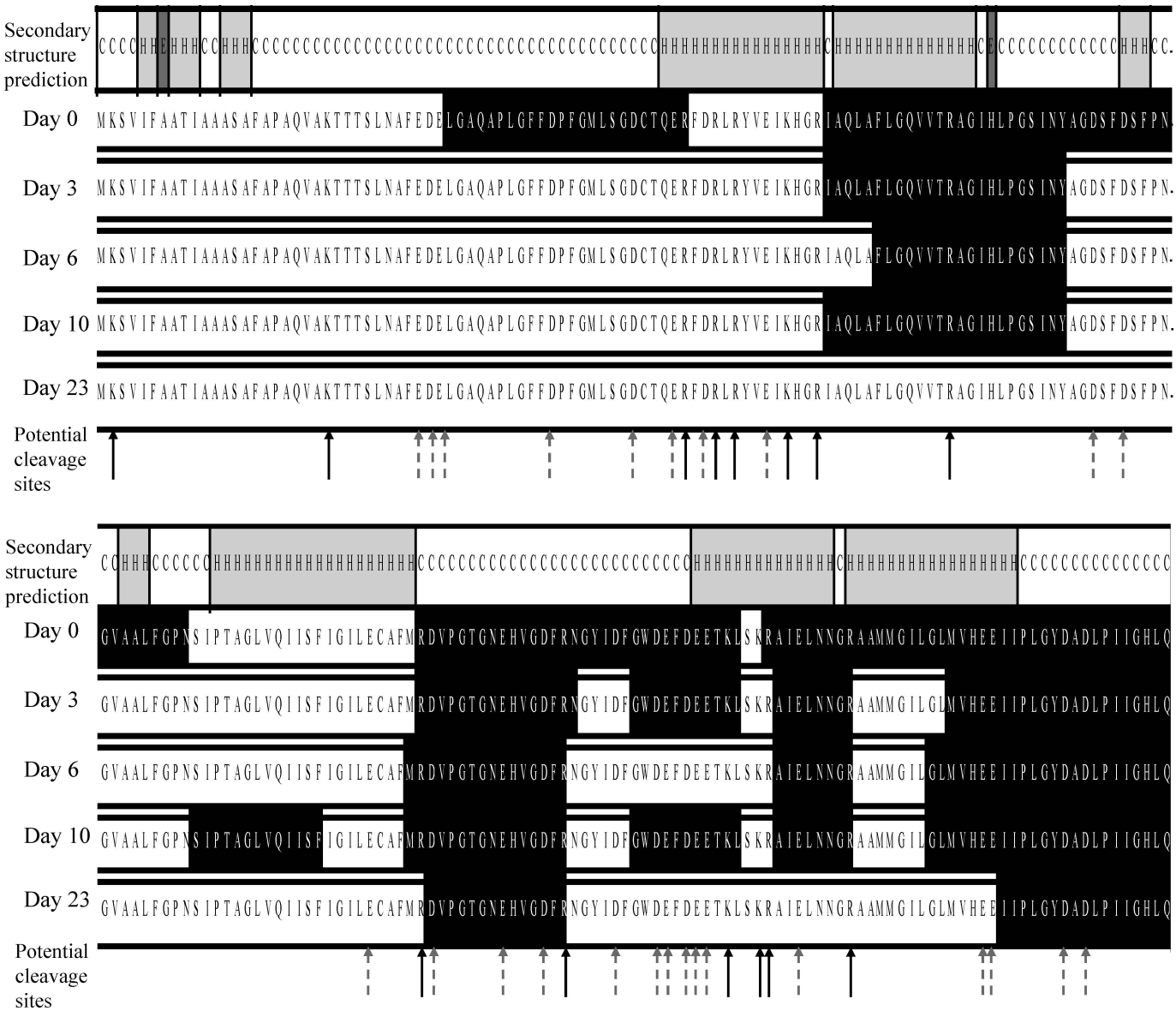

Fig. 3. Graphical representation of the relative location of secondary structural features (Cole et al. 2008) in the fucoxanthin chlorophyll *a,c–binding* protein 9 detected with respect to peptides observed on the 5 d sampled through the degradation experiment. For structural predictions, amino acids have been designated into secondary structures: white, C: random coil; gray, H: α-helix; dark gray, E: *β*-sheet. Days 0–23 display a horizontal amino-acid sequence with the detected peptides blocked off in black with white lettering. Solid black arrows represent cleavage points accessible by trypsin: C-terminal to arginine (R) and lysine (K), while dashed gray arrows indicate the most commonly observed cleavage points accessible by the serine protease endo–Glu C: C-terminal to aspartic acid (D) and glutamic acid (E).

nana peptide, and no diatom-derived RNA polymerase proteins were discovered. No other bacterial proteins were identified in the remainder of the experiment using this concatenated database.

The second search for bacterial proteins utilized a much larger database that consisted of over 6 million unannotated protein sequences and that included the NCBI nonredundant database and the nonredundant GOS marine database. Because this database contains so many possible peptide sequences, it was an effective way to detect spurious identifications of diatom proteins. On day 0 of the experiment, 31 tandem mass spectra were correlated with microbial peptides, together representing 13 proteins. Closer examination of the identifications revealed that 25 of the 31 peptides had identical sequences to *T. pseudonana* peptides. The six peptides not duplicated in the *T. pseudonana* database were evaluated using Basic Local Alignment Search Tool (BLAST). Each peptide was searched with BLAST against the nonredundant database to determine the most closely related protein function. Three of the six unique microbial peptides were most closely aligned with hypothetical proteins of unknown function. Two were involved in energy metabolism, and one was involved in pyrimidine biosynthesis. These unique

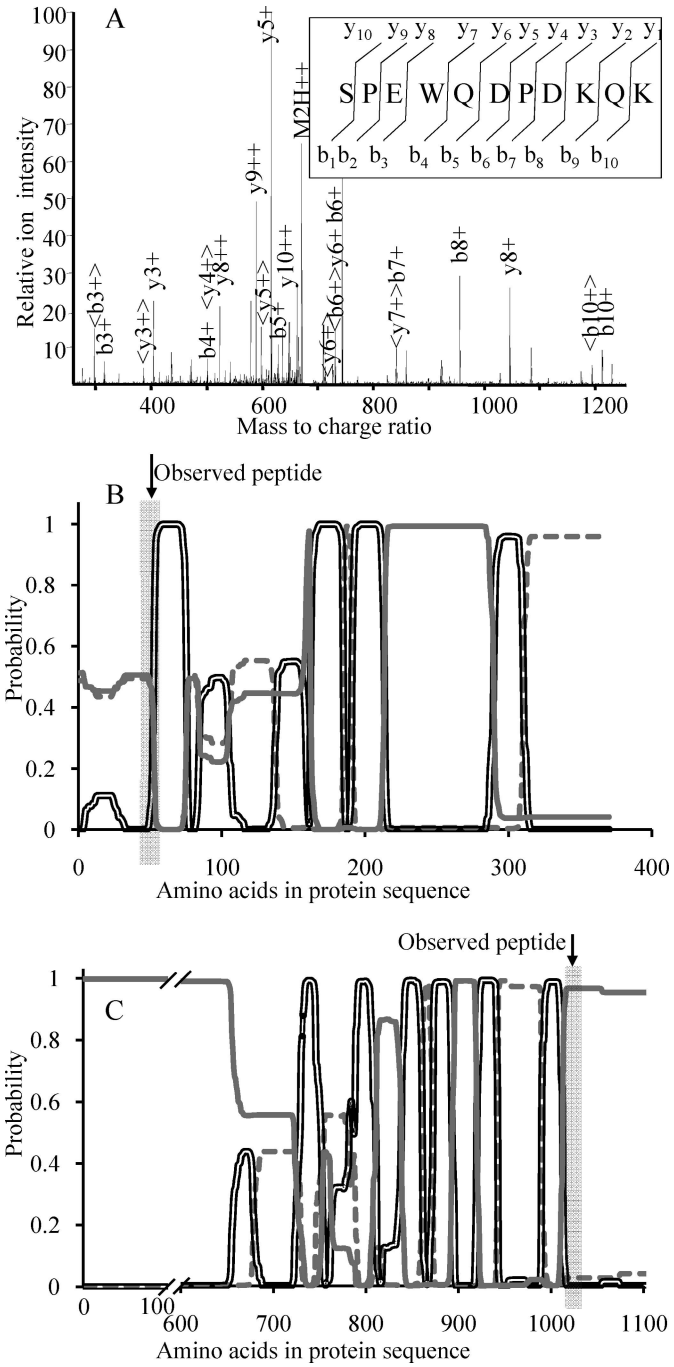

Fig. 4. (A) Tandem mass spectrum from sample day 23 illustrating corresponding b- and y-ions used to identify the peptide sequence SPEWQDPDKQK, derived from iron transporter FTR1. (B, C) Graphical representation of distribution of transmembrane spanning helices predicted (TMHMM version 2.0) for (B) FTR1 iron transporter and (C) proteophosphoglycan protein. The vertical axis represents the probability that a given amino acid is inside the membrane (dashed gray line), outside the membrane (solid gray line), or across the membrane (double black line). Shaded regions indicate the location of the peptide detected on day 23 representing the protein. (C) Note the figure has a break in the *x*-axis between amino-acids 100 and 600. The transmembrane prediction for this portion of the protein was the same as indicated in the amino acids 0–100.

peptides are therefore identifiers for the presence of some bacterial proteins and have the potential to be used as active biomarkers. A search of the GOS database using data from later time points, including day 23, yielded no confident matches to microbial peptides.

Peptide modifications—InsPecT (Tanner et al. 2005) was used to investigate the presence of chemical modifications on peptides detected during the degradation experiment. This blind search was completed to determine if there were recurring modification masses on precursor ions. Unknown modifications can disable SEQUEST's ability to accurately interpret peptide fragmentations. Once a modification mass is known, tandem mass spectrometry data can be reanalyzed with new parameters by SEQUEST. LC-MS/MS proteomic spectra collected from midexponential-phase nutrient replete phytoplankton (Nunn et al. 2009) were used as baseline data (Fig. 5A). Six modifications were frequently observed in the midexponential-phase phytoplankton: oxidation, dioxidation, acetylation, iodoacetamide derivatives, ubiquitinylation residues, and N-methylmaleimides + water. Four modifications appeared immediately on day 0 of the degradation experiment: oxidation, dimethylation, acetylation, and the iodoacetamide derivative. Iodoacetamide derivatives were introduced during sample preparation to increase the unfolding of proteins and the ultimate sequence availability to the enzymes for cleavage. The acetylation of proteins is pervasive in eukaryotic cells and has been shown to be as important as phosphorylation in cell signaling. Recently, Nielsen et al. (2008) suggested that the use of iodoacetamide as an oxidizing agent in sample preparations results in acetylation-like modifications. If all the modification data collected throughout the degradation experiment are tallied (Fig. 5B), no novel modifications present themselves nor are any previously observed modifications up-regulated in comparison to midexponential growth *T. pseudonana*.

Discussion

The rapid and almost total loss of identifiable proteins over 23 d of degradation underscores the difficulty in tracking peptides and proteins preserved in oceanic systems. In this dynamic setting, protein turnover is fast, and organic matter can be tightly bound to minerals or other particles. An additional complication expected from in situ environmental samples, such as sediments, is the presence of a vast array of proteins from numerous diverse organisms. The high number of contaminating, nonpeptide spectra collected at later time points in the degradation experiment (Table 1) suggests that additional isolation steps will be needed for many environmental matrices (Nunn et al. 2006). To date, all discovery-based proteomic methods rely first on tandem mass spectrometry to provide a survey of the peptide ions present in a sample, and second on the ability of the resulting tandem mass spectra to be correlated with a peptide sequence and protein identification. These two requirements highlight the importance of sample preparation and protein sequence database selection for assigning identifications. In addition, these

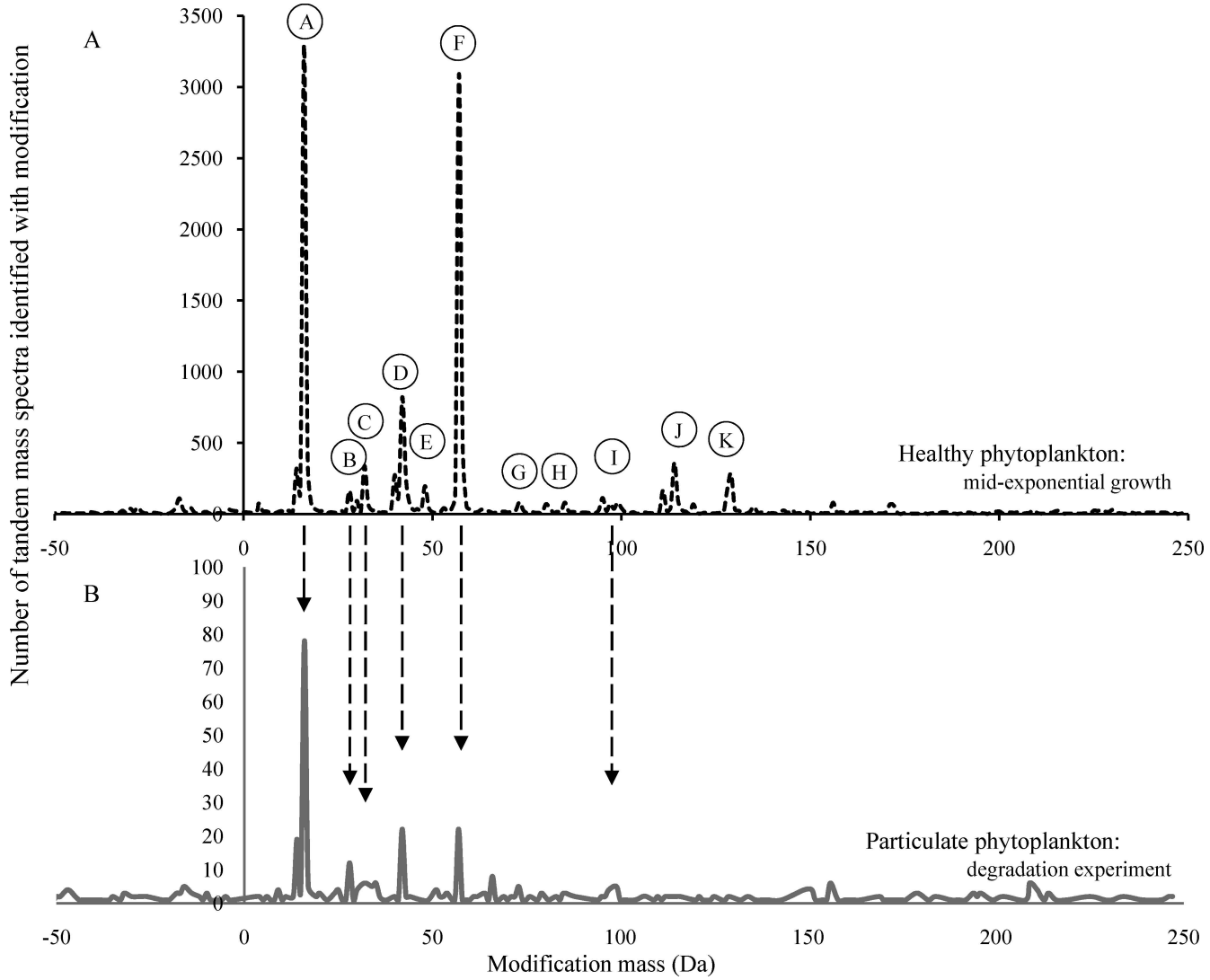

Fig. 5. Graphical representation of InsPecT modification search results on all tandem mass spectra collected from LC-MS/MS analyses on (A) midexponential growth *T. pseudonana* (black dashed line: Nunn et al. 2009) and (B) particulate phytoplankton grown to late stationary phase and exposed to a natural assemblage of marine bacteria for 23 d (gray solid line). Vertical axis represents the total number of tandem mass spectra collected in each experiment (at all time points) that exhibited the modification mass represented on the horizontal axis. (peak A) Oxidation (16 Da); (peak B) acetaldehyde or Lys → Arg substitution(28 Da); (peak C) dioxidation or dihydroxy (32 Da); (peak D) acetylation (42 Da); (peak E) amino-acid substitution Val → Phe, Asp → Tyr (48 Da); (peak F) iodoacetamide derivative (57 Da); (peak G) amino-acid substitution Leu → Trp (73 Da); (peak H) arginine replacement by nitropyrimidyl ornithine (80 Da); (peak I) thiophosphorylation (95 Da); (peak J) ubiquitinylation (114 Da); and (peak K) N-methymaleimide+water (128 Da).

prerequisites are true regardless of whether a sample is first prepared by gel electrophoresis, or if it is present as a complex mixture and analyzed using a protein profiling technique (e.g., shotgun proteomics). One bias of tandem MS instrumentation results from the nature of ion selection because it primarily selects the high-abundant ions for fragmentation. These ions are mainly derived from highly abundant proteins, and the ability of the MS to detect them depends on the physicochemical properties of the peptides (i.e., gas-phase basicity). Other factors that can influence tandem MS detection include the physicochemical traits of the protein, such as molecular weight (large proteins yield more peptides than small proteins), protease accessibility, and its solubility. Although there are these limitations with shotgun proteomics, over the past two decades, thousands of successful experiments have been completed and published using the technology to define and compare proteomes in order to understand physiology. This study demonstrates the utility of tandem MS-based proteomics to identify peptides and proteins that remain after extensive microbial degradation. These first experiments help to define the factors that might encourage a proteins' preservation potential over longer timescales.

Day 0: Rapid restructuring of the proteome—After the diatoms reached stationary phase growth and were subjected to 5 d of darkness, a sample was collected; from this sample, only 121 proteins were identified using tandem

mass spectrometry. In an earlier study, duplicate MS analyses of *T. pseudonana* collected at midexponential growth yielded over 350 diatom protein identifications (Nunn et al. 2009). This comparison suggests that the greatest loss of protein diversity occurred in the first 5 d of light limitation, prior to the introduction of bacteria. During the 5 d of darkness, diatoms likely restructure their proteome in order to adapt to low-light conditions and (eventually) disable deoxyribonucleic acid (DNA) replication pathways. This would allow the organism to invest more energy in autolytic processes induced during programmed cell death. Recently, Bidle and Bender (2008) revealed that although no visual evidence of cell lysis in *T. pseudonana* during iron starvation was observed, genomic DNA was degraded, and caspase proteins were upregulated. This suggests that under nutrient stress, diatoms begin the cascade of apoptotic, or programmed cell death, pathways. The detection of several ribosomal and ATP synthase proteins also implies a general overhaul of the proteome. Although this study cannot determine the balance of internally driven programmed cell death verses microbial activity, further culture-based degradation studies with and without bacteria could be used to quantify degradation resulting from each of these two processes. Among the proteins observed during the degradation experiment, only a few metabolic pathways appear to remain in the system. Those observed include proteins involved in photosynthesis, carbon fixation, and peptidoglycan biosynthesis. Proteins from these three pathways were previously discovered to be the most abundant proteins expressed in *T. pseudonana* at midexponential growth (Nunn et al. 2009). Since proteins involved in photosynthesis and carbon fixation are involved in the harvest of light and conversion of energy and carbon dioxide to sugars, the retention of proteins involved in these pathways after 5 d without light exposure may be a consequence of their high abundance within the cells during midexponential growth and the ability of the diatom to maintain the photosynthetic apparatus over long periods of time (Peters and Thomas 1996). Peptidoglycan synthesis in *T. pseudonana* appears to be a constitutive process with continual protein expression, and thus the presence of these proteins is expected (Nunn et al. 2009).

Day 3: Early degradation—By day 3 of the experimental period, it appears that cells were still operating programmed cell death strategies, since this was the final day that ubiquitin was observed (*see* Web Appendix, www.aslo.org/lo/toc/vol_55/issue_4/1790a.html). It has been previously demonstrated that proteins tagged with ubiquitin monomers are present immediately after a cell is stressed, and those tagged proteins are then removed from the system via degradation within minutes (Shaeffer and Kania 1995; Hochstrasser 1996). After day 3, the blind modification search by InsPecT revealed no proteins or peptides to be posttranslationally modified by the ubiquitin monomers. After 8 d of darkness (i.e., the third day of bacterial degradation), the cells would have been extremely stressed. Therefore, the lack of ubiquitin proteins, or ubiquitin modifications, after day 3 suggests that the diatoms' cellular metabolism was shut down at that point and diatom-controlled cellular degradation was completed. This would indicate that from day 3 forward, no new diatom proteins were being added to the system.

The unique identification of the glycine-rich RNA binding protein (see Web Appendix) at day 3 was intriguing because there are several amino-acid studies that report increased contributions of glycine to the total amino-acid pool as degradation proceeds (Dauwe et al. 1999*b*; Ingalls et al. 2003; Nunn and Keil 2005). No glycine-rich peptides were identified, even though the protein identified in this study on day 3 was glycine-rich. This protein was not identified later than day 3 in the experiment, although there were several glycine-rich peptides identified as resulting from human keratins.

Days 3–10: Rapid protein loss—In general, the number of identified proteins from day 3 through day 10 did not change. On day 10, a minor increase in total protein (Fig. 2A) was observed. Two potential causes of this increase are that more phytoplankton organellar membranes were being compromised (releasing their protein contents), or a detectable bacterial contribution to the total protein was measured. The bacteria-specific search performed on tandem MS data from day 10 used two very extensive protein databases, and yet it yielded no confident bacterial protein identifications. This indicates that bacterial contributions to the total protein present in the particulate samples collected did not reach the threshold required for MS detection. Therefore, on day 10, the more likely explanation is that more diatom membranes were ruptured, making intracellular proteins more available for protein measurement. PsbK, identified only on day 10, may not have been previously observed because it was encased in the large light-harvesting complex and was not adequately solvated for proteomic analysis, or it was simply missed by the data-dependent ion selection method that is biased toward high-abundance ions. Its eventual exposure due to degradation of the outside of the complex might have also contributed to the observed spike in protein concentration.

Day 23: Detrital proteins—The identification of only four *T. pseudonana* proteins after 23 d of degradation was surprisingly low considering the ability of tandem MS on-column peptide detection achieving attomole levels of peptides (Yi et al. 2003). These proteins have unique properties that could augment the oceanographic community's understanding of protein preservation in the deep ocean after large diatom blooms. Our ability to confidently identify the presence of these particular four proteins at the end of 23 d of bacterial exposure suggests that the proteins' tight association with, or enclosure in, membranes is the primary factor influencing the survival and detection of peptides after extensive microbial degradation. In addition, the number of glycan modification motifs and the ability of the protein to aggregate to a supramolecule also discouraged complete enzymatic hydrolysis. Finally, the relative abundance of a protein at the time of death may enhance the ultimate detection of these nitrogen-rich macromole-

cules from the complex matrices of detritus using tandem mass spectrometry. The four proteins retained include: fucoxanthin chlorophyll *a,c–binding* protein 9, iron transporter FTR1, proteophosphoglycan, and Clp-X proteinase. Each of these proteins has a characteristic that provides insight into possible mechanisms of preservation. Further degradation studies that duplicate the protein functionality, cellular location, or modifications proposed to control preservation need to be conducted to validate these hypotheses.

Fucoxanthin chlorophyll a,c–binding *protein 9*—Fucoxanthin chlorophyll *a,c–binding* proteins are enclosed within four, or more, membranes, since they are associated with the thylakoid disks that are located inside of the chloroplast. There are multiple FCPs in *T. pseudonana*, and thus far, it is unknown as to whether FCP9 resides within the thylakoid lumen, providing full enclosure within the disk (Westermann and Rhiel 2005), or if it is integral to the thylakoid membrane, with portions spanning the lipid bilayer (Lang and Kroth 2001). Lang and Kroth (2001) specifically designed a study to determine the mechanism responsible for FCP insertion into the thylakoid disk. In order to isolate and clean the organelles, they used a protease-based cleanup procedure to remove proteins on the outside of the membranes. During their study, they discovered that FCPs were somewhat resistant to enzymes. They attributed this find to FCPs' transmembrane confirmation. An earlier study on diatom mortality by Peters and Thomas (1996) demonstrated that after diatoms were exposed to prolonged periods of darkness (3–10 months) and then re-exposed to light, photosynthesis resumed. This suggests that the photosynthetic apparatus is likely well protected and maintained by the diatom even during periods of prolonged darkness. Although many proteins make up the photosynthetic apparatus, FCP9 was the most abundant protein expressed in the proteomic survey of midexponential growth *T. pseudonana* (Nunn et al. 2009). Therefore, the lines of evidence provided by this study, and others, suggest that both the protection within multiple membranes and relative abundance in the photosynthetic apparatus during its active life cycle likely contribute to the survival and detection of FCP9.

Iron transport protein—The iron transporter, FTR1, was also previously recognized as an abundant protein in midexponential growth *T. pseudonana* (Nunn et al. 2009). This transporter has four predicted transmembrane domains (Fig. 4B). The unique peptide that remained in the experimental system after 1 month of bacterial exposure was adjacent to a membrane-spanning domain. Findings from this study imply that peptides adjacent to, or trapped within, diatom membranes have enhanced preservation potential. Along these lines, it has been previously suggested that liposome-like structures might also withstand bacterial digestion (Nagata et al. 1998). Membrane-bound proteins are inherently difficult to isolate and solubilize, limiting our analytical ability to observe the whole protein sequence using mass spectrometry–based proteomics. As expected, the portion of the membrane-bound protein that crosses the membrane was not observed using tandem mass spectrometry (Wu and Yates 2003), but it is possible that those peptides also remain in the preserved particulate organic matter. Both FCP9 and FTR1 proteins identified after 23 d of degradation correspond to a GO category that includes all proteins enclosed in double lipid bilayers (i.e., organelles). Protection of these proteins inside an organelle may postpone bacterial intrusion and forestall protein degradation (Wolfe et al. 2006).

Proteophosphoglycan—The presence of an adhesive-type protein, proteophosphoglycan, at the end of the 23-d degradation experiment elicits another possible mechanism to enhance protein and peptide preservation in the marine environment. Several studies have proposed that modifications involving either abiotic condensation reactions or biological post-translational modifications (PTMs) increase long-term preservation potential (Keil and Kirchman 1994). More specifically, recent investigations of glycosylation in the marine environment have shown that glycosylated proteins are degraded by bacteria at a much slower rate (Squier and Harvey unpubl.). Outside of the marine environment, studies have also demonstrated that glycosylated proteins are a natural defense against bacterial infection, discouraging bacterial attack and subsequent degradation (Lasky 1992; Lengsfeld et al. 2004). These lines of evidence suggest that attached glycan residues can enhance the long-term preservation of this protein. The peptide detected in this study was not found to be modified either on the N-terminal or serine-rich portion of the protein predicted to contain glycan additions. Notably, this peptide was also adjacent to a membrane-spanning domain (Fig. 4C).

Clp-X proteinase—The identification of a degradation-regulated ATP-binding ClpX diatom protease after 23 d of exposure to a natural bacterial assemblage yields clues not only to mechanisms that control preservation potential, but also to autolytic processes in the diatom. Activity of Clp proteases is initiated by formation of large multiunit heteromeric complexes, and because these complexes may get as large as 1700 kDa, portions of the supermolecule may survive degradation. Nguyen and Harvey (2001) foreshadowed the persistence of Clp protease supercomplexes in sediments when they discovered that very high-molecular-weight proteinaceous material remained in algal detritus after long-term decay. Here, we were able to confirm their postulate with an unbiased search that discovered a 38–amino-acid–long peptide from this proteinase after 23 d of degradation. Although this protein has several domains that are highly conserved between *Candidatus Pelagibacter ubique*, a marine bacteria, and *Thalassiosira pseudonana*, the peptide we identified is only 73% identical between the two species. In addition, the molecular weight of this peptide in the two different organisms differs by more than 100 Da. Since the analytical mass accuracy for MS1 ions in the orbitrap is $\pm$ 0.001 Da, and SEQUEST parameters were configured to examine peptides with a mass tolerance of $\pm$ 2.1 Da, the assignment to a *T. pseudonana*–generated protein is correct. Because

Clp protease is known to be involved in intracellular degradation, its presence at the end of the experiment reveals an important mechanism in diatom cell death. Cleavage by this protein is a processive degradation, accessing multiple cleavage sites without releasing intermediate degradation products (Thompson et al. 1994). This same type of degradation has been demonstrated to occur when BSA is degraded by marine bacteria (Nunn et al. 2003). Assuming that Clp is an abundant protein, it likely represents a primary mechanism for protein removal from the diatom cells, and its discovery explains why there are few peptide intermediates remaining in the system.

Bacterial contributions—Although the minor contribution of bacterial carbon to the system suggested that bacterial proteins were unlikely to be observed, tandem mass spectra were searched using marine microbe proteome databases, as noted previously. Only six spectra were correlated to microbial peptides on day 0. Even though the experimental system was dominated by diatom proteins, the observation of these spectra does suggest that peptides unique to bacteria were present at extremely low levels in particulate diatom material. These peptides have the potential to be used as identifiers for the presence of bacterial proteins. For example, in sedimentary systems where organic matter recycling has led to extensive degradation (and thus protein removal), proteins representing the living bacterial community might be detectable with targeted approaches. Although several tandem mass spectra from the day 10 sample did correlate with bacterial peptides, they also correlated with diatom peptides from the *T. pseudonana* database. Further validation of these individual spectral matches confirmed them to be of diatom origin.

Glycine enrichments—Glycine enrichment has been correlated with increased degradation time (i.e., with depth in a sediment core; Dauwe et al. 1999*a*). While absolute concentrations of glycine in this study were low, its concentration relative to other amino acids did increase throughout the experiment (Fig. 1C). Several hypotheses as to why the fraction of glycine increases in oceanic particulate matter with long-term exposure to bacteria have been proposed, including increases in glycine-rich cell-wall proteins from bacteria and phytoplankton after extensive degradation (Hecky et al. 1973; Grutters et al. 2002), loss of side chains from other amino acids or mycosporine-like amino acids present in the sediment, resulting in a stripped amino-acid equivalent of glycine (Ingalls et al. 2010), or higher concentrations of glycine betaine, a metabolic osmolite released by some bacteria living in saline environments (Ghoul et al. 1990). In this experiment, we did not release organic matter bound within the frustules by dissolving the silica (e.g., Ingalls et al. 2010). Using proteomic techniques, we discovered that the glycine-rich peptides correlated with human keratins. This suggests that the observed enrichment for glycine here, and possibly in other studies, may simply be due to contamination from sample handling. These samples were collected and prepared with the utmost care as to ensure that human contaminates were not introduced. That said, this is a common artifact that can only be discovered using proteomic techniques. Assuming that there is a minor, but constant, addition of human keratin to ocean particulate samples during sample handling, as the percent of total ocean-based protein decreases with depth in sediment, contaminating keratins would contribute a larger percent to the observed peptides and amino-acid distribution. Positing keratins as a significant source of glycine provides a plausible explanation for the substantial increase in the relative concentration of glycine, while overall protein content plummeted. Assuming each MS analysis was performed on similar quantities of protein (1 μg of protein), as the ratio of diatom to keratin peptides decreased, the mass spectrometer collected more data on the keratins as those peptides became the most intense ions available for data-dependent ion selection.

Since primary production from organisms like *T. pseudonana* can net $\sim 50 \times 10^{15}$ g C yr^{-1} globally, and it is estimated that up to 75% of that carbon is present as protein, examination of the turnover rate and ultimate fate of this material is essential to building accurate models of carbon transfer and burial in the ocean. Although investigations of protein from marine environments are currently under way, more genomic data and proteome annotations are needed to adequately identify all proteins from such a complex system. From this study, only a few proteins were identified to have survived after the month-long degradation experiment. More extensive experiments with mixed cultures under a variety of light and bacterial conditions may provide new mechanisms or characteristics on protein survival in oceanic environments. The knowledge gained on the potential mechanisms driving the survival of planktonic proteins can help focus future work on the degradative processes of organic nitrogen in marine systems and elsewhere. The proteins identified here provide a frame of reference for interpreting much more complex in situ environmental samples across a range of matrices.

Acknowledgments

We would like to thank J. Shaw and the two anonymous reviewers for their helpful comments and questions. B.L.N. would also like to thank Y.Y.M. and T.A.Z. for keeping the project focused. This work was funded by grants from the National Science Foundation Division of Ocean Sciences (OCE-0825790 and OCE-0825811) as well as the National Institute of Health (1S10RR017262-01, 1S10RR023044-01, and P30ES07033).

References

Armbrust, E. V., and others. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science **306:** 79–86, doi:10.1126/science.1101156

Bidle, K. D., and S. J. Bender. 2008. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom *Thalassiosira pseudonana*. Eukaryot. Cell **7:** 223–236, doi:10.1128/EC.00296-07

Boski, T., J. Pessoa, P. Pedro, J. Thorez, M. A. Dias, and I. R. Hall. 1998. Factors governing abundance of hydrolyzable amino acids in the sediments from the N. W. European continental margin (47–50°N). Prog. Oceanogr. **42:** 145–164, doi:10.1016/S0079-6611(98)00032-9

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:** 248–254, doi:10.1016/0003-2697(76)90527-3

Bratback, G. 1987. Carbon flow in an experimental microbial ecosystem. Mar. Ecol. Prog. Ser. **36:** 267–276, doi:10.3354/meps036267

Cole, C., J. D. Barber, and G. J. Barton. 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. **36:** W197–W201, doi:10.1093/nar/gkn238

Dauwe, B., J. J. Middelburg, P. M. J. Herman, and C. H. R. Heip. 1999*a*. Linking diagenetic alterations of amino acids and bulk organic matter reactivity. Limnol. Oceanogr. **44:** 1809–1814.

———, ———, P. Van-Rijswijk, J. Sinke, P. M. J. Herman, and C. H. R. Heip. 1999*b*. Enzymatically hydrolyzable amino acids in North Sea sediments and their possible implication for sediment nutritional values. J. Mar. Res. **57:** 109–134, doi:10.1357/002224099765038580

Diekmann, A. B. S., M. A. Peck, L. Holste, M. A. St. John, and R. W. Campbell. 2009. Variation in diatom biochemical composition during a simulated bloom and its effect on copepod production. J. Plankton Res. **31:** 1391–1405, doi:10.1093/plankt/fbp073

Dunne, J. P., R. A. Armstrong, A. Gnanadesikan, and J. L. Sarmiento. 2005. Empirical and mechanistic models for the particle export ratio. Glob. Biogeochem. Cy. **19:** GB4026, doi:10.1029/2004GB002390

Eng, J. K., B. Fischer, J. Grossmann, and M. J. Maccoss. 2008. A fast SEQUEST cross correlation algorithm. J. Proteome Res. **7:** 4598–4602, doi:10.1021/pr800420s

———, A. L. McCormack, and J. R. Yates. 1994. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J. Am. Soc. Mass Spec. **5:** 976–989, doi:10.1016/1044-0305(94)80016-2

Fukuda, R., H. Ogawa, T. Nagata, and I. I. Koike. 1998. Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. Appl. Environ. Microbiol. **64:** 3352–3358.

Ghoul, M., T. Bernard, and M. Cormier. 1990. Evidence that *Escherichia coli* accumulates glycine betaine from marine sediments. Appl. Environ. Microbiol. **56:** 551–554.

Gottesman, S., W. P. Clark, V. de Crecy-Lagard, and M. R. Maurizi. 1993. ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. Sequence and in vivo activities. J. Biol. Chem. **268:** 22618–22626.

Grutters, M., W. van Raaphorst, E. Epping, W. Helder, J. W. de Leeuw, D. P. Glavin, and J. Bada. 2002. Preservation of amino acids from in situ–produced bacterial cell wall peptidoglycans in northeastern Atlantic continental margin sediments. Limnol. Oceanogr. **47:** 1521–1524.

———, ———, and W. Helder. 2001. Total hydrolysable amino acid mineralisation in sediments across the northeastern Atlantic continental slope (Goban Spur). Deep-Sea Res. I **48:** 811–832, doi:10.1016/S0967-0637(00)00075-3

Harvey, H. R., J. H. Tuttle, and J. T. Bell. 1995. Kinetics of phytoplankton decay during simulated sedimentation: Changes in biochemical composition and microbial activity under oxic and anoxic conditions. Geochim. Cosmochim. Acta **59:** 3367–3377, doi:10.1016/0016-7037(95)00217-N

Hecky, R. E., K. Mopper, P. Kilham, and E. T. Degens. 1973. The amino acid and sugar composition of diatom cell-walls. Mar. Biol. **19:** 323–331, doi:10.1007/BF00348902

Hedges, J. 1991. Lignin, cutin, amino acids and carbohydrate analyses of marine particulate organic matter, p. 129–137. *In* D. C. Hurd and D. W. Spencer [eds.], Marine particles: Analysis and characterization. Amer. Geophys. Union.

Hochstrasser, M. 1996. Ubiquitin-dependent protein degradation. Annu. Rev. Genet. **30:** 405–439, doi:10.1146/annurev.genet.30.1.405

Immler, D., S. Greven, and P. Reinemer. 2006. Targeted proteomics in biomarker validation: Detection and quantification of proteins using a multi-dimensional peptide separation strategy. Proteomics **6:** 2947–2958, doi:10.1002/pmic.200500659

Ingalls, A. E., C. Lee, S. G. Wakeham, and J. I. Hedges. 2003. The role of biominerals in the sinking flux and preservation of amino acids in the Southern Ocean along 170°W. Deep-Sea Res. II **50:** 713–738, doi:10.1016/S0967-0645(02)00592-1

———, K. Whitehead, and M. C. Bridoux. 2010. Tinted windows: The presence of the UV absorbing compounds called mycosporine-like amino acids embedded in the frustules of marine diatoms. Geochim. Cosmochim. Acta **74:** 104–115, doi:10.1016/j.gca.2009.09.012

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

Keller, A., A. I. Nesvizhskii, E. Kolker, and R. Aebersold. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. **74:** 5383–5392, doi:10.1021/ac025747h

Knicker, H. 2000. Solid-state 2-D double cross polarization magic angle spinning ^{15}N ^{13}C NMR spectroscopy on degraded algal residues. Org. Geochem. **31:** 337–340, doi:10.1016/S0146-6380(00)00018-8

Lang, M., and P. G. Kroth. 2001. Diatom fucoxanthin chlorophyll *a*/*c*–binding protein (FCP) and land plant light-harvesting proteins use a similar pathway for thylakoid membrane insertion. J. Biol. Chem. **276:** 7985–7991, doi:10.1074/jbc.M006417200

Lasky, L. A. 1992. Selectins—interpreters of cell-specific carbohydrate information during inflammation. Science **258:** 964–969, doi:10.1126/science.1439808

Lee, C. 1988. Amino acid chemistry and amine biogeochemistry in particulate material and sediments, p. 125–141. *In* T. H. Blackburn and J. Sorensen [eds.], Nitrogen cycling in coastal marine environments. Wiley.

Lengsfeld, C., F. Titgemeyer, G. Faller, and A. Hensel. 2004. Glycosylated compounds from okra inhibit adhesion of *Helicobacter pylori* to human gastric mucosa. J. Agric. Food Chem. **52:** 1495–1503, doi:10.1021/jf030666n

Mayer, L. M., L. Schick, and F. Setchell. 1986. Measurement of protein in nearshore marine sediments. Mar. Ecol. Prog. Ser. **30:** 159–165, doi:10.3354/meps030159

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

Miki, T., L. Giuggioli, Y. Kobayashi, T. Nagata, and S. A. Levin. 2009. Vertically structured prokaryotic community can control the efficiency of the biological pump in the oceans. Theor. Ecol. **2:** 199–216, doi:10.1007/s12080-009-0042-8

Montagnes, D. J. S., J. A. Berges, P. J. Harrison, and F. J. R. Taylor. 1994. Estimating carbon, nitrogen, protein, and chlorophyll-*a* from volume in marine-phytoplankton. Limnol. Oceanogr. **39:** 1044–1060, doi:10.4319/lo.1994.39.5.1044

Nagata, T., R. Fukuda, I. Koike, K. Kogure, and D. L. Kirchman. 1998. Degradation by bacteria of membrane and soluble protein in seawater. Aquat. Microb. Ecol. **14:** 29–97, doi:10.3354/ame014029

Nesvizhskii, A. I., and others. 2006. Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data. Mol. Cell. Proteomics **5:** 652–670.

NGUYEN, R. T., AND H. R. HARVEY. 2001. Preservation of protein in marine systems: Hydrophobic and other noncovalent associations as major stabilizing forces. Geochim. Cosmochim. Acta **65:** 1460–1480, doi:10.1016/S0016-7037(00)00621-9

NIELSEN, M. L., M. VERMEULEN, T. BONALDI, J. COX, L. MORODER, AND M. MANN. 2008. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. Nature Meth. **5:** 459–460, doi:10.1038/nmeth0608-459

NUNN, B. L., AND R. G. KEIL. 2005. Size distribution and chemistry of proteins in Washington coast sediments. Biogeochemistry **75:** 177–200, doi:10.1007/s10533-004-6546-9

———, AND ———. 2006. A comparison of non-hydrolytic methods for extracting amino acids and proteins from coastal marine sediments. Mar. Chem. **98:** 31–42, doi:10.1016/j.marchem.2005.06.005

———, A. NORBECK, AND R. G. KEIL. 2003. Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. Mar. Chem. **83:** 59–73, doi:10.1016/S0304-4203(03)00096-3

———, S. A. SHAFFER, A. SCHERL, B. GALLIS, M. WU, S. I. MILLER, AND D. R. GOODLETT. 2006. Comparison of a *Salmonella typhimurium* proteome defined by shotgun proteomics directly on an LTQ-FT and by proteome prefractionation on an LCQ-DUO. Brief. Funct. Gen. Prot. **5:** 154–168, doi:10.1093/bfgp/ell024

———, AND A. T. TIMPERMAN. 2007. Marine proteomics. Mar. Ecol. Prog. Ser. **332:** 281–289, doi:10.3354/meps332281

———, AND OTHERS. 2009. Deciphering diatom biochemical pathways via whole-cell proteomics. Aquatic Microbial Ecology **55:** 241–253, doi:10.3354/ame01284

PANTOJA, S., C. LEE, AND J. F. MARECEK. 1997. Hydrolysis of peptides in seawater and sediment. Mar. Chem. **57:** 25–40, doi:10.1016/S0304-4203(97)00003-0

PETERS, E., AND D. N. THOMAS. 1996. Prolonged darkness and diatom mortality I: Marine Antarctic species. J. Exp. Mar. Biol. Ecol. **207:** 25–41, doi:10.1016/S0022-0981(96)02520-8

RAGUENEAU, O., S. SCHULTES, K. BIDLE, P. CLAQUIN, AND B. LA MORICEAU. 2006. Si and C interactions in the world ocean: Importance of ecological processes and implications for the role of diatoms in the biological pump. Glob. Biogeochem. Cy. **20:** GB4S02, doi:10.1029/2006GB002688

SAUER, P. E., T. I. EGLINTON, J. M. HAYES, A. SCHIMMELMANN, AND A. L. SESSIONS. 2001. Compound-specific D/H ratios of lipid biomarkers from sediments as a proxy for environmental and climatic conditions. Geochim. Cosmochim. Acta **65:** 213–222, doi:10.1016/S0016-7037(00)00520-2

SHAEFFER, J. R., AND M. A. KANIA. 1995. Degradation of monoubiquitinated alpha-globin by 26S proteasomes. Biochemistry **34:** 4015–4021, doi:10.1021/bi00012a020

SIMON, M., AND F. AZAM. 1989. Protein content and protein synthesis rates of planktonic marine bacteria. Mar. Ecol. Prog. Ser. **51:** 201–213, doi:10.3354/meps051201

TANNER, S., AND OTHERS. 2005. InsPecT: Fast and accurate identification of post-translationally modified peptides from tandem mass spectra. Anal. Chem. **77:** 4626–4639, doi:10.1021/ac050102d

THOMPSON, M. W., S. K. SINGH, AND M. R. MAURIZI. 1994. Processive degradation of proteins by the ATP-dependent Clp protease from *Escherichia coli*. Requirement for the multiple array of active sites in ClpP but not ATP hydrolysis. J. Biol. Chem. **269:** 18209–18215.

TSAI, Y. S., A. SCHERL, J. L. SHAW, C. L. MACKAY, S. A. SHAFFER, P. R. LANGRIDGE-SMITH, AND D. R. GOODLETT. 2009. Precursor ion independent algorithm for top-down shotgun proteomics. J. Am. Soc. Mass Spectrom. **20:** 2154–2166, doi:10.1016/j.jasms.2009.07.024

VELJI, M. I., AND L. J. ALBRIGHT. 1993. Improved sample preparation for enumeration of aggregated aquatic substrate bacteria, p. 36–37. *In* P. F. Kemp, B. F. Sherr, E. B. Sherr and J. J. Cole [eds.], Handbook of methods in aquatic microbial ecology. Lewis.

WAKEHAM, S. G., C. LEE, J. HEDGES, P. J. HERNES, AND M. L. PETERSON. 1997. Molecular indicators of diagenetic status in marine organic matter. Geochim. Cosmochim. Acta **61:** 5363–5369, doi:10.1016/S0016-7037(97)00312-8

WESTERMANN, M., AND E. RHIEL. 2005. Localisation of fucoxanthin chlorophyll *a*/*c*–binding polypeptides of the centric diatom *Cyclotella cryptica* by immuno-electron microscopy. Protoplasma **225:** 217–223, doi:10.1007/s00709-005-0083-9

WOJTKOWIAK, D., C. GEORGOPOULOS, AND M. ZYLICZ. 1993. Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of *Escherichia coli*. J. Biol. Chem. **268:** 22609–22617.

WOLFE, A. P., M. B. EDLUND, A. R. SWEET, AND S. D. CREIGHTON. 2006. A first account of organelle preservation in Eocene nonmarine diatoms: Observations and paleobiological implications. Palaios **21:** 298–304, doi:10.2110/palo.2005.p05-14e

WU, C. C., AND J. R. YATES III. 2003. The application of mass spectrometry to membrane proteomics. Nat. Biotechnol. **21:** 262–267, doi:10.1038/nbt0303-262

YI, E. C., H. LEE, R. AEBERSOLD, AND D. R. GOODLETT. 2003. A microcapillary trap cartridge-microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation. Rapid Commun. Mass Spectrom. **17:** 2093–2098, doi:10.1002/rcm.1150

ZANG, X., D. H. JASPER, K. J. DRIA, AND P. G. HATCHER. 2000. Encapsulation of protein in humic acid from a Histosol as an explanation for the occurrence of organic nitrogen in soil and sediment. Organic Geochem. **31:** 679–695, doi:10.1016/S0146-6380(00)00040-1

Associate editor: Bo Thamdrup

Received: 18 September 2009
Amended: 10 May 2010
Accepted: 12 May 2010