5-2012

Evaluation of Electrophoretic Protein Extraction and Database-Driven Protein Identification from Marine Sediments

Eli K. Moore

Brook L. Nunn

Jessica F. Faux

David R. Goodlett

H. Rodger Harvey
Old Dominion University, hharvey@odu.edu

Original Publication Citation

Moore, E.K., Nunn, B.L., Faux, J.F., Goodlett, D.R., & Harvey, H.R. (2012). Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments. *Limnology and Oceanography: Methods, 10*, 353-366. doi: 10.4319/lom.2012.10.353

LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 10, 2012, 353–366

Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments

Eli K. Moore[1†], Brook L. Nunn[2†], Jessica F. Faux[1], David R. Goodlett[2], and H. Rodger Harvey[1,3]*
[1]University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory, Solomons, MD 20688 USA
[2]Department of Medicinal Chemistry, University of Washington, Seattle, Washington
[3]Department of Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, VA 23529, USA

Abstract

Intact proteins comprise a major component of organic carbon and nitrogen produced globally and are likely an important fraction of organic matter in sediments and soils. Extracting the protein component from sediments and soils for mass spectral characterization and identification represents a substantial challenge given the range of products and functionalities present in the complex matrix. Multiple forms of gel electrophoresis were evaluated as a means of enhancing recovery of sedimentary protein before proteomic characterization and compared with a direct enzymatic digestion of proteins in sediments. Resulting tryptic peptides were analyzed using shotgun proteomics and tandem mass spectra were evaluated with SEQUEST. Multiple databases were then evaluated to examine the ability to confidently identify proteins from environmental samples. Following evaluation of electrophoretic extraction of proteins from sediments, the recovery of an experimentally added standard protein (BSA) from older (>1 ky) sediments was optimized. Protein extraction from sediments via direct electrophoresis of a slurry mixture and the specified extraction buffer resulted in the greatest number of confident protein identifications and highest sequence coverage of the BSA standard. Searching tandem mass spectral data against larger databases with a higher diversity of proteomes did not yield a greater number of, or more confidence in, protein identifications. Regardless of the protein database used, identified peptides correlated to proteins with the same function across taxa. This suggests that while determining taxonomic-level information remains a challenge in samples with unknown mixed species, it is possible to confidently assign the function of the identified protein.

The vast majority of organic nitrogen in marine phytoplankton is represented by protein (Lourenco et al. 1998), with total hydrolysable amino acids (THAAs) accounting for up to 30% to 40% of particulate nitrogen in marine sediments (Cowie and Hedges 1992a; Grutters et al. 2001). Whereas total amino acids provide a proxy of total protein material, the functional and source information embedded in each protein's amino acid sequence is lost. To fully characterize the cellular machinery of organisms responsible for the biogeochemical cycles of nitrogen and carbon, the identification of peptides and/or proteins in marine sediment is required.

Extracting proteins from sediment for subsequent analysis has long been a challenge (Belluomini et al. 1986; Ogunseitan 1993; Craig and Collins 2000; Nunn and Keil 2006). The potential interferences present in sediment or soils include protein binding to the mineral matrix (Mayer 1994; Keil et al. 1994; Collins et al. 1995), organic matter co-extraction (Knicker and Hatcher 1997), interaction with humic acids (Zang et al. 2000) or cellular polymers (Nguyen and Harvey 2003), and protein-protein aggregation, which limits solubility (Nguyen and Harvey 2001). Although the application of strong agents to solubilize proteins can be effective, it results in the co-extraction of a suite of unknown compounds with similar physiochemical properties as protein from the sediment or soil matrix. These mixtures are inherently complex and interfere with the purification and subsequent identification of peptides and proteins (Cheng et al. 1975; Limmer and Wilson 1980; Nunn and Timperman 2007).

One important interaction between organic matter and

*Corresponding author: E-mail: rharvey@odu.edu
†These authors contributed to the work equally.

Acknowledgments

This work was supported through a collaborative research award by the Chemical Oceanography program of NSF to HRH (OCE-1128776), BLN and DRG (OCE-0825790). MOGEL contribution No. 11-004 of the Department of Ocean, Earth and Atmospheric Sciences, Old Dominion University. We thank Sonia Ying Ting and Shannon Yihsuan Tsai for bioinformatic support.

DOI 10.4319/lom.2012.10.353

sedimentary minerals appears to be surface adsorption (Mayer 1994; Mayer et al. 2002). Various mechanisms have been proposed to regulate adsorption of organic matter to mineral surfaces including van der Waals interactions (Rashid et al. 1972), ligand exchange (Davis 1982), cation bridges (Greenland 1971), cation (Wang and Lee 1993) and anion exchange (Greenland 1971), and hydrophobic effects (Nguyen and Harvey 2001). These mechanisms of interaction between protein, sedimentary minerals, and organics often include some form of charge interaction. This electrokinetic phenomenon was first observed by Reuss (1809) when the application of a constant electric field caused migration of aqueous clay particles in water. Fractionation of mineral species by electrophoresis was later demonstrated by Dunning et al. (1982). This principle of mobilization by an electric field was incorporated in the study design to assist the liberation of proteins from sediments.

Gel electrophoresis has been widely used for decades as a protein separation and visualization technique. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and related approaches separate proteins based primarily on their molecular weights (Laemmli 1970). The wide application of SDS-PAGE and its ability to solubilize and immobilize proteins have made it a standard analytical technique for protein separation and isolation across the fields of biochemistry, cell biology, and medical sciences (e.g., Reisfeld et al. 1962; Laver 1964; Shapiro et al. 1967; Fairbanks et al. 1971; Maizel 2000; Pederson 2008). This includes electrophoretic separation to purify proteins from cell cultures before analysis with mass spectrometry (Tran and Doucette 2009; Botelho et al. 2010). The focus here was to develop and validate a modified electrophoretic approach as an extraction and preparative technique for complex environmental samples before high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) analysis, allowing us to accomplish the goal of identifying proteins or peptides in marine sediments.

Gel electrophoresis as a preparatory method is founded on the visualization of protein bands by staining, followed by the excision and enzymatic digestion of the proteins bound within the bands. This method is a standard tool for protein identification using HPLC-MS/MS (Hirano et al. 1992; Shevchenko et al. 1996; Kuster, et al. 1998). Rather than using gel electrophoresis as a means for visualizing the isolated proteins, however, we employed the SDS-PAGE technique to 1) enhance protein solubilization (i.e., see Botelho et al. 2010) from the sediment matrix, 2) isolate sediment particles from soluble components, 3) stabilize and retain proteins while rinsing away unwanted contaminants, and 4) suspend denatured proteins (Fairbanks et al. 1965) for enzymatic digestion. Although the application of sediments to SDS-PAGE is unorthodox, the method provided multiple benefits in addition to being a standard technique that is frequently employed for the digestion of proteins for tandem mass spectrometry analysis.

Advancements in proteomic use of HPLC-MS/MS have increased sensitivity and detection limits, providing the user with an increased ability to identify peptides from complex mixtures (Schulze et al. 2005; Morris et al. 2010; Dong et al. 2010). The extraction, isolation, and analysis of proteins and peptides using MS/MS are only the first steps, however, that provide the researcher with spectral data. To interpret peptide mass spectral data, it must be correlated with peptides and proteins from a user-provided proteomic database. Many of the current databases are not yet mature, and bioinformatic challenges for identifying proteins retained in soils and sediments include the high diversity of unknown taxonomic contributors, the incomplete availability of protein databases, and the diverse mixture of proteins, which yield limiting concentrations of specific sequences for detection and identifications (Graves and Haystead 2002; Quince et al. 2008; Bastida et al. 2009).

The goals of this work were two-fold. The first goal was to optimize the extraction of proteins from marine sediments with a broadly applicable methodology. The second and equally important goal was to assess the effectiveness of proteomic database complexity on environmental samples. For the first goal, we evaluated two methods that employed a SDS-PAGE clean-up step: 1) a more traditional method where the buffer-solubilized material is separated from particles and loaded directly onto gels, and 2) a novel slurry extraction method where the buffer-solubilized material remains with the sediment particles and is loaded as a composite material together. Two electrophoresis gels were compared, including preparatory tube gels and standard one-dimensional flat gels with multiple combinations of extraction buffers for each. To address the second goal, mass spectra were searched against five databases of varying size to evaluate database-driven protein identifications from multiple species using probabilistic scoring. As model sediments, continental shelf surface and deeper core sediments from the Bering Sea were used as the test matrix since this area is one of the world's most productive ecosystems (Sambrotto et al. 1986; McRoy 1987; Walsh et al. 1989), and is known to be diatom-dominated during spring blooms. A high carbon export flux (Chen et al. 2003) in the spring, coupled with rapid transit times, elevates the amounts of diatom derived organic material reaching sediments. These factors make the Bering Sea a realistic system to explore sedimentary protein extraction and evaluate information from multiple database searches.

Materials and procedures

Protein extraction

Bering Sea surface sediments were extracted using a buffer followed by SDS-PAGE. The buffer consisted of 7 M urea, 2 M thiourea, 0.01 M Tris-HCl, 1 mM EDTA, 10% v/v glycerol, 2% CHAPS, 0.2% w/v ampholytes (Fluka BioChemika, high resolution pH 3-10, 40% in water), 2 mM Tributyl-phosphine (Kan et al. 2005). The mixture includes chaotropic agents, detergents, denaturants, and salts, thus proteins are solubilized and

stabilized while avoiding degradation. The use of strong chaotropic agents and subsequent trypsin digestion alleviated the use of protease inhibitors. Additional agents added to the extraction buffer including thiourea, EDTA, and CHAPS served to counteract dilution of urea and Tris when mixed with sediments. Replicate aliquots of each treatment were used for amino acid analyses to measure recoveries.

For the traditional method, approximately 1.5 g dry weight aliquots of surface sediment (~1 mL wet sediment) were combined with 5 mL extraction buffer in duplicate Falcon Tubes to yield a 5:1 buffer:sediment ratio (v/v). Tubes were sonicated on ice for 60 s using pulse sonication (Bronson microprobe, at 20 kHz). Sediment extracts were centrifuged to remove particles from the extraction liquid (5000*g*, 10 min, 4°C) and overlying liquid (approximately 5 mL) was loaded onto a Bio-Rad gel prep cell 0.5 cm diameter gel tube. The gel consisted of 10% Acrylamide/Bis, 0.125 M Tris-HCl and gel tubes were poured to a height of 3 cm, which allowed a larger sample loading volume above the tube gel than the suggested 10 cm height. THAA concentrations were used as a proxy for total protein to adjust loading volumes for gels. The gel prep cell was run at 180 volts until the ion front moved approximately 1 cm down the gel. If the ion front moved further down the gel, it was difficult to excise cleanly from the gel tube. The top 1 cm was then excised for tryptic digestion.

In the slurry method, approximately 1.5 g dry weight aliquots of surface sediment (~1 mL wet sediment) were combined with 1 mL of extraction buffer in duplicate Falcon Tubes to yield a 1:1 buffer:sediment ratio (v:v). Tubes were sonicated on ice for 60 s using pulse sonication (same conditions as above) and 500 µL of sediment + extraction buffer slurry mixture was then loaded onto a Bio-Rad gel prep cell 0.5 cm diameter gel tube poured to a height of 10 cm. Gel composition and running conditions for the slurry method were the same as the traditional method. The slurry gel was run until the ion front moved 5 cm down the gel. Sediment particles remained at the top of the gel and were easily washed away after the gel run was finished (Fig. 1). The top 5 cm of the gel was then excised for

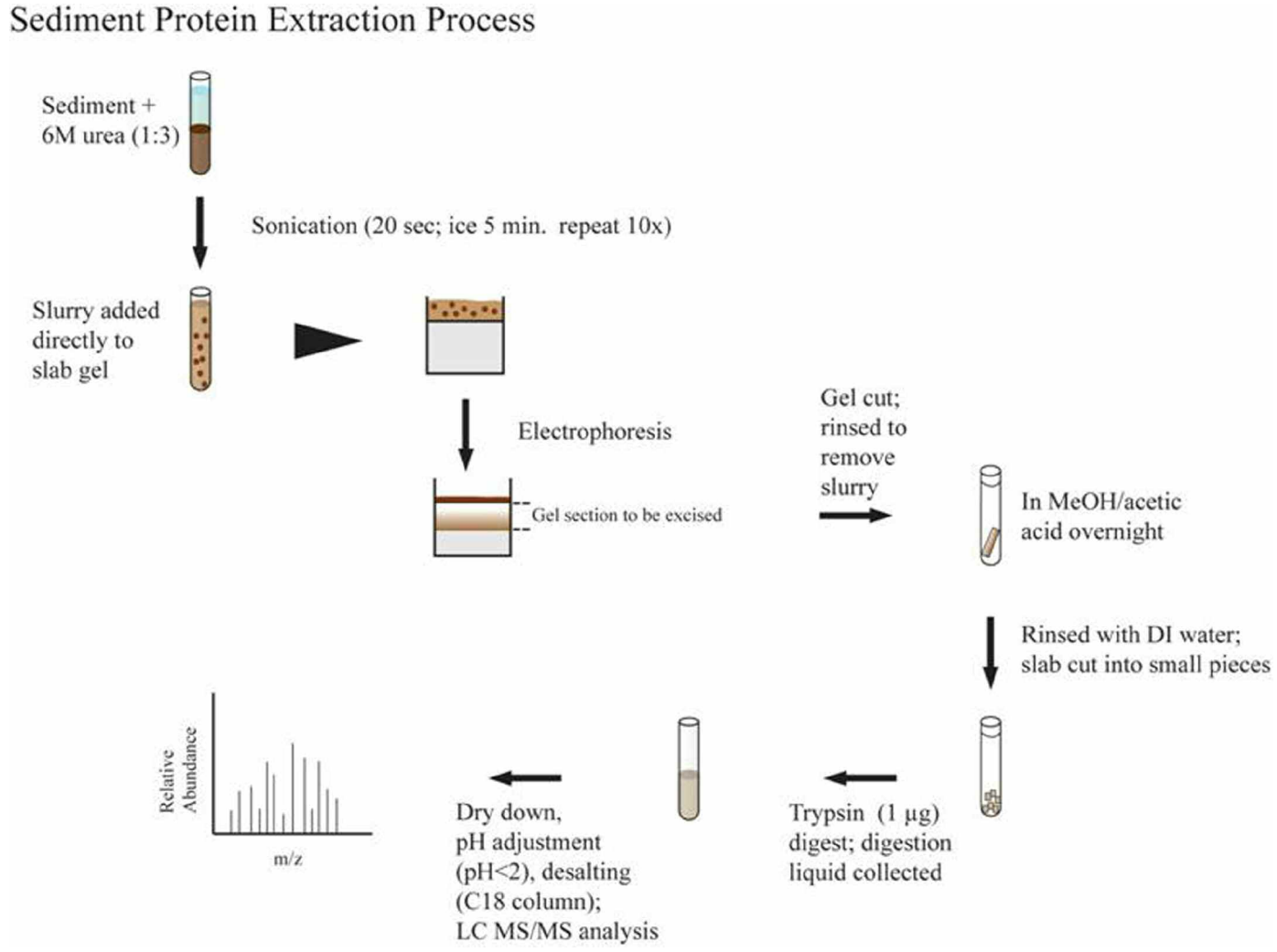

Fig. 1. Schematic workflow for the slurry approach for extraction, purification, and digestion of sedimentary samples prior to LC/MS analysis.

digestion. Slurry mixture was also loaded onto a pre-cast 12% Bis-Tris Bio-Rad 1-dimension gel (referred to as "flat gel") and run until the ion front moved 5 cm down the gel. The loading volume of the flat gel was not large enough to load equivalent amounts of protein material in liquid extract compared to the sediment + buffer slurry mixture, so only the sediment + buffer slurry was tested for the flat gel. The top 5 cm of the slurry flat gel was then excised for tryptic digestion.

Optimization of protein recovery

Several methods for the optimization of electrophoretic extraction of protein were tested using BSA as a model protein. Deeper sediments, the 20 to 22 cm horizon of a second sediment core from the Bering Sea, were used for all the permutations of the optimization investigation. Treatments included varying the type of extraction buffer and gel type used, as well as the type of preparation loaded onto the gel. Four different extraction buffers were tested including 1) EDTA extraction buffer, 2) $CaCl_2$ extraction buffer, 3) SDS extraction buffer, and 4) urea alone. Two types of gels were used in the extraction: preparative tube gels (run in BioRad Mini-Prep Cell) and precast 12% Bis-Tris 1-dimension, flat gels (Invitrogen NuPAGE Novex). Due to the limited amount of sediment available, the urea buffer was tested only on flat gels. Two types of sample preparations were tested: a traditional method where extraction buffer was loaded onto the gel and a slurry method in which a mixture of extraction buffer and sediment were loaded directly onto the gel. The various combinations of buffers and gels resulted in the testing of fourteen extraction permutations.

For most treatments, 750 µL extraction buffer was mixed with 250 µL sediment, the exception being those samples that employed only urea for extraction. To those samples, 300 µL urea was added. All extraction buffers contained the following (with the exception of the urea only samples): urea (7M), thiourea (2M), tris-HCl (0.01M), glycerol (10% v/v), ampholytes (pH - 3 to 10) (0.2% v/v), and tributylphosphine (0.002M). In addition, the EDTA extraction buffer contained CHAPS (2% w/v) and EDTA (1mM); the $CaCl_2$ extraction buffer contained CHAPS (2% w/v) and $CaCl_2$ (0.1 M); and the SDS extraction buffer contained SDS (1% w/v) and EDTA (1 mM). Those samples which used urea only were extracted with 6 M urea. To all samples, immediately following the addition of buffer, 1 µg of BSA (2mg/mL in ultrapure 0.9% sodium chloride; Thermo-Fisher) was added and samples were sonicated. Samples were sonicated for 20 s with a titanium microtip and placed on ice for 5 min. This was repeated for a total of ten sonication treatments and samples were placed in the –20°C freezer overnight. Slurry samples were thawed and the entire sample was placed on top of either the tube gel or the flat gel. Samples that did not include sediment particles were centrifuged, the supernatant was filtered, and 500 µL of extract was placed on either the tube or flat gels. Tube gels were prepared according to BioRad Mini-Prep Cell specifications for discontinuous gels (12% resolving gel with 200 µL of 12% stacking gel on top) and poured to a height of 10 cm. Tube gels were run until the ion front had migrated approximately 2 cm from the top of the gel, and flat gels were run until the ion front had migrated approximately 2.5 cm from the top of the gel, as this provided a cleanly excisable gel fraction. Gels were cut just below the ion front, rinsed three times with DI water to rinse away excess SDS, and placed in a fixing solution of 50% methanol, 10% acetic acid overnight.

Trypsin digestion of SDS-PAGE slices

All samples purified via SDS-PAGE gels were digested using the same protocol. Before digestion the excised tube gels and flat gels were cut into 2 mm sized cubes to increase the exposed surface area. Pieces were covered with 100 mM ammonium bicarbonate and rinsed for 15 min followed by a 15 min rinse in acetonitrile. Volumes used to rinse flat and tube gels with ammonium bicarbonate and acetonitrile were 18 mL and 8 mL, respectively. The rinse cycle was repeated twice more and gel pieces were dried via Speedvac for 45 min. To the gel pieces, 1 µg trypsin was added in solution and the sample was placed on ice for 45 min. Reduction, alkylation, and digestion for surface sediment samples generally followed procedure by Shevchenko et al. (1996). For optimization testing, samples were removed from the ice covered with ammonium bicarbonate and digested overnight without reduction or alkylation. The pH of the sample was adjusted with 5% formic acid to a pH ≤_2, and the sample was run through a C18 desalting column (Nest Group), following which samples were dried and volumes were adjusted in preparation for analysis.

Direct digestion of sediment

Prior to shotgun proteomic analysis, 100 mg sediment was mixed with 300 µL 6 M urea and 50 mM ammonium bicarbonate. The sediments were then sonicated using a Bronson sonicating microprobe, at 20 kHz for 60 s on ice. The pH was raised by adding 18 µL 1.5 M Tris-HCl (pH 8.8). To reduce sulfhydryl linkages in proteins, 7.5 µL TCEP (2,2′,2′(-phosphanetriyltripropanoic acid) was added to the sediments, vortexed and incubated for 1 h (37°C). Proteins were then alkylated by adding 60 µL of 200 mM iodoacetic acid and incubated in the dark for 1 h. After the addition and incubation of 60 µL of dithiothreitol (1 h room temp), the urea was diluted with the addition of 2.4 mL 25 mM ammonium bicarbonate, 600 µL HPLC-grade methanol, and 1 µg of sequencing grade trypsin. The trypsin incubation was completed overnight at room temperature. Samples were centrifuged (14,000*g*, 20 min), and the digest with buffer removed. The sediments were then washed 3 times with 1 mL 25 mM ammonium bicarbonate, centrifuged, and extracts combined. The volume was reduced to ~10 µL, and 200 µL of 5% ACN, 0.1 % trifluoroacetic acid was added before desalting the peptides using a C18 desalting centrifuge column (NEST group). Sample pH was adjusted to < 2 using small additions of 10% TFA and was desalted using the protocol provided by manufacturer.

Mass spectrometry

Protein analyses for all samples were conducted using standard shotgun proteomic techniques employing nanocapillary HPLC-MS/MS as described previously (Washburn et al. 2001; Aebersold and Goodlett 2001). Samples were introduced into a hybrid linear ion Orbitrap (LTQ-OT) mass spectrometer (Thermo Fisher) via a NanoAcquity high performance liquid chromatography (HPLC) system (Waters). Trapping and analytical capillary columns were packed in-house using a pressurized cylinder (Brechbuhler AG). Magic C18 (5 µm diameter, 100 Å pore size) particles (Michrom Bioresource) were slurried with analytical grade MeOH and placed in the pressurized cylinder to pack columns using 1000 psi nitrogen. Trapping column capillaries were 20 mm × 100 µm i.d., while the analytical column dimensions were 150 mm × 75 µm i.d. The trapping column was prepared with a sintered glass frit on one end, and the analytical column was tapered in a flame by gravity that allowed it to serve as a frit and electrospray ionization needle.

Chromatography was performed using acidified mobile phases: A) water, 0.1% (v/v) formic acid and B) acetonitrile, 0.1% (v/v) formic acid. Chromatography was followed as in Nunn et al. (2010) using a Waters NanoAcquity system. Based on parallel amino acid measurements of each sample, 1 µg protein-equivalent material was injected onto the nanocapillary HPLC column for MS/MS analysis, which produced average total ion current (TIC) signal intensities of $> 1 \times 10^7$. The LTQ-OT XL (ThermoFisher) was operated in positive ion mode using data-dependent acquisition (DDA). Sheath, aux, and sweep gases were not used for any of these analyses. Both the linear ion trap and Orbitrap were calibrated and tuned using a standard tetrapeptide MRFA (Met-Arg-Phe-Ala: *m/z = 524*). Mass accuracy for the Orbitrap ranged from 2-5 ppm. For analyses, capillary spray voltage ranged from 1.4-1.6 kV, collision energy was set to 35%, collision activation was 30 ms, dynamic exclusion was 45 s, and only ions with a charge state + 1 were rejected. From the precursor ion scan (MS1) in the instrument, five of the most intense ions were selected for collision-induced dissociation (CID) and tandem mass spectral (MS2) detection in the LTQ (Nunn and Timperman 2007). All sample digests were analyzed first using a standard full scan, where the MS2 ion selection was chosen from the top 5 most intense ions in the *m/z* range of 350-2000. Then each tube gel surface sediment sample received further analyses on the mass spectrometer using 4 gas phase fractionations where the top five most intense ions were selected for CID by DDA from the following *m/z* ranges: *350-444, 444-583, 583-825, 825-1600* (Spahr et al. 2001; Davis et al. 2001; Yi et al. 2002; Scherl et al. 2008).

Database searching

The search engine SEQUEST was used to match tandem mass spectra to peptide sequences found in protein databases (Eng et al. 1994, 2008). Searches did not explore post translational modifications (PTM), given the high variability of PTMs, which make confident identification difficult. Four protein sequence databases were generated using a publically available FASTA database builder (http://phenyx.proteomics.washington.edu/FASTAcreator/index.cgi) and then evaluated for mass spectra collected from Bering Sea sediment gel digests:

1) Thaps database contains the proteome of *Thalassiosira pseudonana* (marine diatom, well annotated proteome; Armbrust et al. 2004), and expanded with the proteomes of *Prochlorococcus marinus* (marine cyanobacterium; Dufresne et al. 2003), and *Candidatus Pelagibacter ubique* (marine bacteria; Giovannoni et al. 2005). These proteomes provide representation of algae, autotrophic bacteria, and heterotrophic bacteria respectively (14,795 proteins, 15 megabytes).

2) GOS/Thaps database contains the proteome of *T. pseudonana* and the Global Ocean Survey Combined Assembly Protein (GOS) database (Yooseph et al. 2007). This database was used in an attempt to correlate resulting tandem mass spectra from sediments to a variety of possible bacterial proteomes (6,121,580 million protein sequences, 2.3 gigabytes). Available protein names and source organisms were acquired from the CAMERA online portal (Community Cyberinfrastructure for Advance Microbial Ecology Research & Analysis: http://camera.calit2.net/index.shtm).

3) NCBI-NR database (National Center for Biotechnology Information Reference Sequence) consists of a nonredundant collection of highly annotated DNA, RNA, and protein sequences from diverse taxa, including marine organisms (Pruitt et al. 2002). While the NCBI-NR has fewer marine proteomes compared with the GOS database, it has greater functional information, diversity, and contains a variety of eukaryotic marine organisms not found in the GOS (11,934,213 proteins, 4.9 gigabytes);

4) NCBI-Refined database was generated to include the whole proteomes of each of the species identified from the NCBI-NR search on the slurry surface sediment sample. This database contained the proteomes from 107 organisms (417,199 proteins, 187 megabytes).

Mass spectra from the digests of the traditional, slurry, and direct digest extraction methods on Bering Sea surface sediment were searched against all four databases. Mass spectra from the 1-D surface sediment slurry flat gel digest were only searched against the Thaps database. Two fixed modifications were set in the SEQUEST parameter file to replicate analytical modifications completed in the lab: 57 Da on cysteine (resulting from IAM alkylation) and 16 Da on methionine from oxidation. Predicted fragmentation versus observed tandem mass spectra were statistically evaluated with PeptideProphet software to assess the MS/MS spectra quality, and ProteinProphet software was used to assign and group peptides into proteins (Keller et al. 2002; Nesvizhskii et al. 2003). Both PeptideProphet and ProteinProphet were set to a *P* value of 0.9, which corresponds to a predicted 0.1% error rate. Ignoring these confidence limits or false discovery rates will yield inaccurate and spurious protein identifications. To quantify model protein recoveries, a specific database containing BSA and 50 common

contaminant proteins was used (51 proteins, 32.9 kilobytes). Spectra generated from the extraction optimization investigations were searched against this database with sequence coverage (%) of BSA used to determine the most successful extraction procedures. As stated earlier, for all database evaluations only proteins reported with high confidence ($P > 0.9$) were accepted and discussed in this study.

Amino acid analysis

To compare amino acid composition of initial sediment and the treatments, individual amino acids were identified and quantified by gas chromatography mass spectrometry (GC-MS) using the EZ:faast derivatization method (Phenomenex). One gram dry weight sediment for each sample was hydrolyzed for 4 h at 110°C with analytical grade 6 M HCl (Cowie and Hedges 1992b), and L-γ -Methylleucine as the recovery standard (Waldhier et al. 2010). Following hydrolysis and derivatization, amino acids were quantified using an Agilent 6890 GC with samples injected at 250°C and separated through a DB-5MS (0.25 mm ID, 30 m) GC column with hydrogen as the carrier gas. Amino acid identification used the same GC coupled to an Agilent 5973N mass spectrometer run under the same conditions. Helium was used as the carrier gas for the amino acid analysis in the GC, and acquisition of spectra between 50-600 Da were collected. Bovine serum albumin (BSA) was analyzed in parallel to correct for responses among individual amino acids and calculation of molar ratios.

Assessment

Sediment properties and amino acids

The surface sediment examined in this study contained 0.48% organic carbon (OC) and 0.06% particulate nitrogen (PN) whereas the deeper sediments used in the optimization experiments contained 1.07% OC and 0.15% PN. Amino acids represented 1.39 ± 0.12 mg THAA/g sediment dry weight in surface sediments, similar to other northern latitude marine sediments (Mintrop and Duinker 1994; Horsfall and Wolff 1997). Deep sediments accounted for 2.59 ± 0.67 mg THAA/g sediment dry weight. THAAs contributed 28.9% of POC and THAA-N contributed 29.1% of PN for surface sediments and 24.4% of POC and 31.2% of PN for deep sediments. Extraction efficiency was based on THAA content of extracts versus THAA content of sediments as a proxy for total protein, and was calculated for surface sediment samples resulting in efficiencies of 12.5% ± 1.1 for the traditional buffer surface sediment extraction, and 100.6% ± 4.5 for buffer extraction of a surface sediment slurry mixture. Amino acid distributions in surface sediments showed only subtle differences between bulk surface sediment and the two surface sediment extraction methods (Fig. 2).

Database and method evaluation of identified proteins from surface sediments

The search against the Thaps database of the slurry tube gel method resulted in the greatest number of confident protein identifications. Using the database that contained proteomes from one diatom and two marine bacteria, 302 unique peptides were identified from the slurry tube gel method, which correlated to 126 protein identifications (Table 1, Web Appendix A). The slurry 1-D flat gel method identified 31 proteins (82 peptides). The traditional method and the direct digest retrieved 60 proteins (149 peptides) and 6 proteins (7 peptides), respectively. The majority of proteins identified from each of the four methods correlated to identifications from

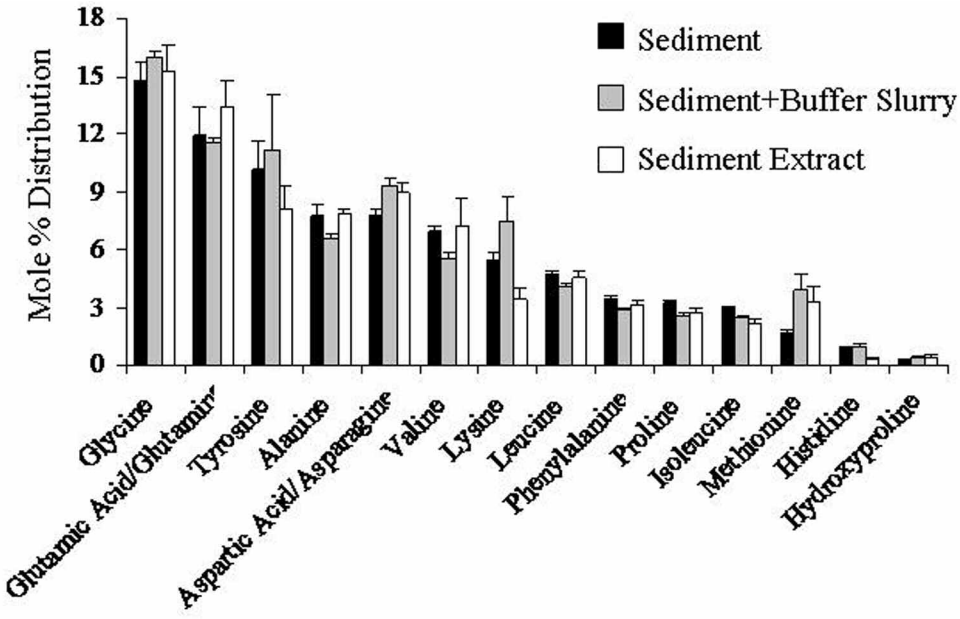

Fig. 2. The comparative distribution of amino acids observed in Bering Sea shelf sediments using the two extraction approaches verses acid-hydrolysed intact sediments.

the diatom, *T. pseudonana*. Only two *P. marinus* protein identifications from the slurry tube gel method were made with no *P. marinus* identifications in the traditional gel or direct digest. There were no proteins identified as *C. P. ubique* using any of the extraction methods. Among the three methods the slurry gel and traditional gel methods had 46 protein identifications in common (Fig. 3A). The slurry tube gel, traditional gel, and direct digest methods had five protein identifications in common.

Proteins identified from the Thaps database were grouped by Gene Ontology categories (Table 2). The majority of proteins identified were involved in metabolic processes. The biggest difference between the distribution of metabolic proteins between the slurry tube gel and traditional tube gel methods was the large contribution of translation/ribosomal proteins among slurry tube gel identifications at 20.2% (26 protein identifications), versus the small contribution from the same category identified from the traditional tube gel method at 1.7% (1 protein identification). The only unique protein identified by the direct digest method is chloroplast ribose-5-phosphate isomerase, a phosphate shunt protein.

Tandem mass spectra searched against the GOS/Thaps database yielded 114 protein identifications from the slurry tube gel method, 63 from the traditional tube gel method, and 4 from the direct digest method (Table A1, Web Appendix). There were fewer peptides identified using each extraction method for the GOS search compared with the Thaps search. The majority of proteins identified from the slurry tube gel method still primarily correlated with sequences associated with the diatom proteome, even though the database consisted almost exclusively of marine bacterial proteins. Only 20 proteins correlated uniquely to GOS protein sequences (Fig. 3B), and seven proteins were identified with peptide sequences that were identical, and therefore conserved, between diatom and marine bacterial proteins. Thirty seven of the traditional tube gel proteins correlated to *T. pseudonana*, 16 from GOS, and 10 as both *T. pseudonana* and GOS proteins (i.e., homologous sequences).

Mass spectral results searched against the NCBI non-redundant (NR) database yielded similar distributions, with the most proteins being identified from the slurry tube gel method and the least from the direct digest (Table 2). Fewer peptides and proteins were identified using the NR database compared with the Thaps and GOS databases. Proteins identified as originating from diatoms or conserved among diatoms and other organisms made up the majority of proteins identified from the slurry and traditional tube gel methods (Table A1, Web Appendix). The majority of direct digest identified proteins were bacterial in origin. The NCBI-Refined database search yielded more peptide and protein identifications than the full NCBI-NR database search for all extraction methods. Two proteins from the full NCBI-NR and NCBI-Refined searches were uniquely identified as originating from prokaryotic sources.

Protein recovery optimization from sediments

There was recovery of BSA standard from sediment samples with most extraction methods tested (Table 3). The exceptions

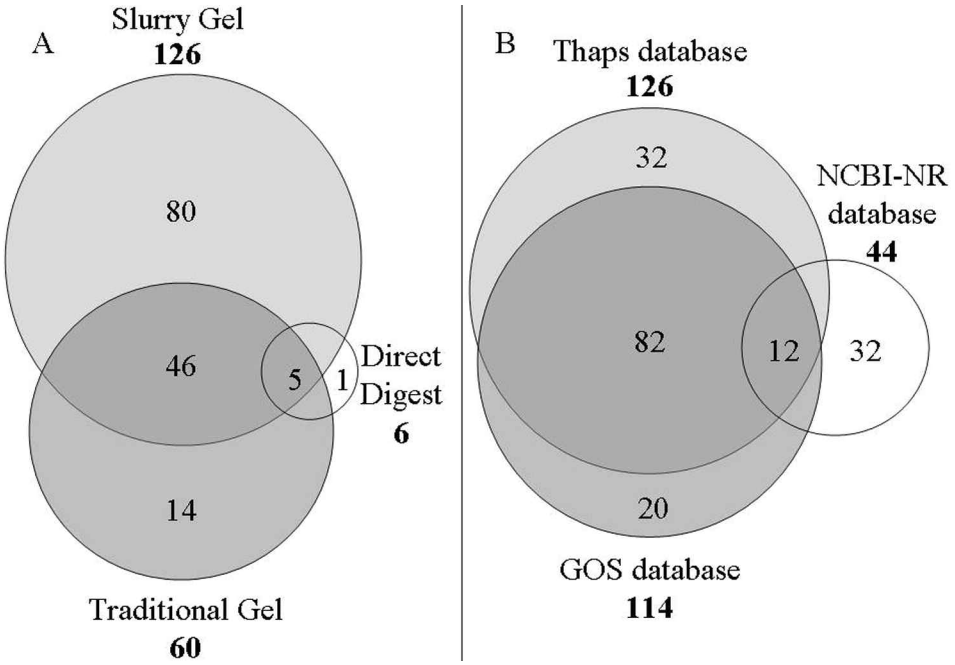

Fig. 3. Venn diagrams of (A) the number of proteins identified in common between the slurry tube gel, traditional tube gel, and direct digest methods searched against the Thaps database; (B) number of proteins in common between the Thaps, GOS/Thaps, and NCBI-NR database searches of the slurry tube gel method.

Table 1. Total proteins identified by extraction method and proteomic database.

Database	Extraction method	Proteins	Peptides	*T. pseudo*	*T. pseudo+*
Direct digest	6	7	6	0	
Thaps	Slurry flat gel	31	82	30	1
	Traditional tube gel	60	149	60	0
	Slurry tube gel	126	302	122	0
	Direct digest	4	4	0	0
GOS	Traditional tube gel	63	130	37	10
	Slurry tube gel	114	257	87	7
	Direct digest	16	16	2	0
NCBI-NR	Traditional tube gel	31	115	15	7
	Slurry tube gel	44	115	16	16
NCBI-refined	Slurry tube gel	84	205	37	44

Proteins = number of proteins identified; Peptides = number of peptides identified; *T. pseudo* = number of *Thalassiosira pseudonana* proteins identified; *T. pseudo+* = number of proteins identified conserved among *Thalassiosira pseudonana* and another source.

Table 2. The cellular functions of identified proteins found using the Thaps database and organized as subgroups of function. The numbers indicate the number of identifications with numbers in parentheses as the percentages of total proteins found in each method.

Function	Slurry tube gel	Traditional tube gel	Direct digest
Metabolism	96 (77.4%)	41 (68.3%)	5 (83.3%)
Photosynthesis	36 (29.0%)	23 (38.3%)	4 (66.7%)
Translation, Transcription	26 (21.0%)	1 (1.7%)	—
Metabolism, Recycling	8 (6.5%)	5 (8.3%)	—
Glycolysis, Respiration	8 (6.5%)	4 (6.7%)	—
Enzyme	7 (5.6%)	5 (8.3%)	—
Biosynthesis	4 (3.2%)	1 (1.7%)	—
GTPase	4 (3.2%)	—	—
Modification	3 (2.4%)	2 (3.3%)	—
Pentose-phosphate shunt	—	—	1 (16.7%)
Binding, Structure	14 (11.3%)	8 (13.3%)	0
Binding DNA, RNA	4 (3.2%)	—	—
Binding ATP, GTP	3 (2.4%)	—	—
Heat Shock	3 (2.4%)	3 (5.0%)	—
Structure	2 (1.6%)	—	—
Folding	1 (0.8%)	—	—
Binding, Zn	1 (0.8%)	—	—
Binding, Protein	—	5 (8.3%)	—
Transport	11 (8.9%)	7 (11.7%)	1 (16.7%)
Transport, Proton	5 (4.0%)	5 (8.3%)	1 (16.7%)
Transferase	2 (1.6%)	—	—
Transport	2 (1.6%)	2 (3.3%)	—
Transport, Protein	1 (0.8%)	—	—
Nucleotidyltransferase Activity	1 (0.8%)	—	—

—, Protein not detected

to this were the two $CaCl_2$/tube gel combinations in which no identifiable BSA was recovered from either the traditional or slurry methods. Sequence coverage, used as a metric to determine the most efficient extraction methods, ranged from 0% (no BSA recovered) to 22% (Table 3). Both samples extracted with the urea extraction buffer resulted in the recovery of the highest number of independent spectra, unique peptides, and sequence coverage of BSA. The slurry method yielded 22% sequence coverage while the traditional method returned 13%. Given that BSA is known to be rapidly hydrolysed by natural microbial communities (Roth and Harvey 2006), here it was used to monitor recovery by the extraction buffer and

Table 3. Results of protein search for BSA (Bovine Serum Albumin) in extraction optimization experiments given as sequence coverage, number of unique peptides, and total independent spectra identified. Sequence coverage was used to determine the most effective extraction method.

Gel type	Sequence coverage (%)	Number of unique peptides	Total independent spectra
Flat			
EDTA, traditional	9.9	4	36
EDTA, slurry	5.6	2	5
$CaCl_2$, traditional	7.4	3	29
$CaCl_2$, slurry	10.4	4	33
SDS, traditional	4.6	2	3
SDS, slurry	8.1	6	30
Urea, traditional	13.2	8	46
Urea, slurry	22.1	13	45
Tube			
EDTA, traditional	3.3	2	2
EDTA, slurry	9.1	4	4
$CaCl_2$, traditional	—	—	—
$CaCl_2$, slurry	—	—	—
SDS, traditional	9.6	3	3
SDS, slurry	5.9	2	2

—, Protein not detected; traditional, extraction buffer only added to gel; slurry, extraction buffer and sediment added to gel.

gel electrophoresis followed by tandem MS. Although extended incubations might allow secondary interactions to be assessed, it would complicate accurate measures of intact protein recovery which were the primary goal.

Discussion

The identification of a variety of proteins and the ability to recover an added standard protein demonstrate that electrophoresis can provide an effective isolation method for proteins in sediment systems. Using the slurry gel method, it was apparent that a greater number of peptides and proteins could be identified compared to the more traditional approach, supporting the hypothesis that an electric field applied directly to sediment particles enhances protein extraction. When results were examined among the varied databases, the Thaps database proved to be the most effective database at maximizing protein identifications for the Bering Sea system. It was also observed that the amount of extracted material must be considered; larger loading volumes using the tube gel can enhance the number of proteins identified but can also extract other materials that reduce confidence of the proteins identified. The low numbers of confident protein identifications using attempts at direct digest of the sediments appears reflective of the same issue, in which interactions with a complex matrix of organic materials or perhaps sorption of trypsin to the solid matrix itself interfere with digestion and extraction.

As the presence of proteins in marine sediments may be limited by degradation or masked by interactions with various matrices, it is beneficial to optimize the extraction technique in order to maximize the recovery of any available protein material. Chen et al. (2008) demonstrate the applicability of sequence coverage as a useful measure of protein expression. Though the use of sequence coverage as a measurement tool demonstrated greater intra-sample variability than other methods tested (Chen et al. 2008), it is an easily determined variable, and its use is acceptable as a metric in the optimization experiments as a determination of the effectiveness of the various protein extraction methods.

Comprehensive testing of BSA with the extraction buffers, electrophoresis gels and sample preparations found that extraction with 6M urea and the placement of the slurry directly on a flat gel was most effective (Fig. 1). The slurry method also yielded 9% greater sequence coverage of BSA than the more traditional method. This supports the observation that the application of an electric field to the slurry mixture enhanced protein extraction, but that a balance is needed to maximize recovery of proteins that can be identified with confidence.

The comparison of the Thaps, GOS/Thaps, NCBI-NR, and NCBI-Refined databases illustrates the amount of information that can be gained on protein functions and taxonomy of source organisms at different levels (i.e., kingdom, class, family). In addition, we can evaluate the utility of large databases in relation to search-time requirements and available computational resources. The distinct advantage of using the more complex GOS and NCBI-NR protein databases are the greater number of organisms from which proteins can be identified. The disadvantage of searching these databases is the amount of computational resources and dedicated time involved. Analysis of 10 tandem MS files (each file containing thou-

sands of spectra) using the GOS database consumed > 720 h, and the NCBI-NR database > 1080 h, using an 800 CPU cluster. This is a substantial time investment compared to the Thaps database searches of < 5 h. Because these searches consumed so much computational time, we focused this study on the search results from only the tandem MS analyses performed on the slurry tube gel sample treatment because it revealed the greatest number of high confidence, multiple-peptide protein hits.

Each database used, Thaps, GOS/Thaps and NCBI-NR, contained the entire forward and reversed *Thalassiosira pseudonana* proteome. Using the GOS/Thaps database, only 20 of the 114 proteins identified were not *T. pseudonana* in origin (~15%). Searches against the NCBI-NR database identified 12 proteins originating from organisms other than *T. pseudonana*. Of the 12 non-*T. pseudonana* proteins identified using the NCBI-NR database, 9 of them still correlated to marine diatom sources. Despite some minor differences in identified protein source resulting from the use of different taxonomic protein sequence databases, assigned protein functions were the same for over 95% of peptides identified from multiple databases (Fig. 4). Although we searched the same suite of tandem MS data against different databases, the larger databases (e.g., GOS and the NCBI-NR) actually yielded fewer confident protein identifications, regardless of the fact that they include an additional 6 to 11 million proteins that are not from *T. pseudonana*. These results demonstrate that larger may not be always better, as few novel non-diatom protein identifications were

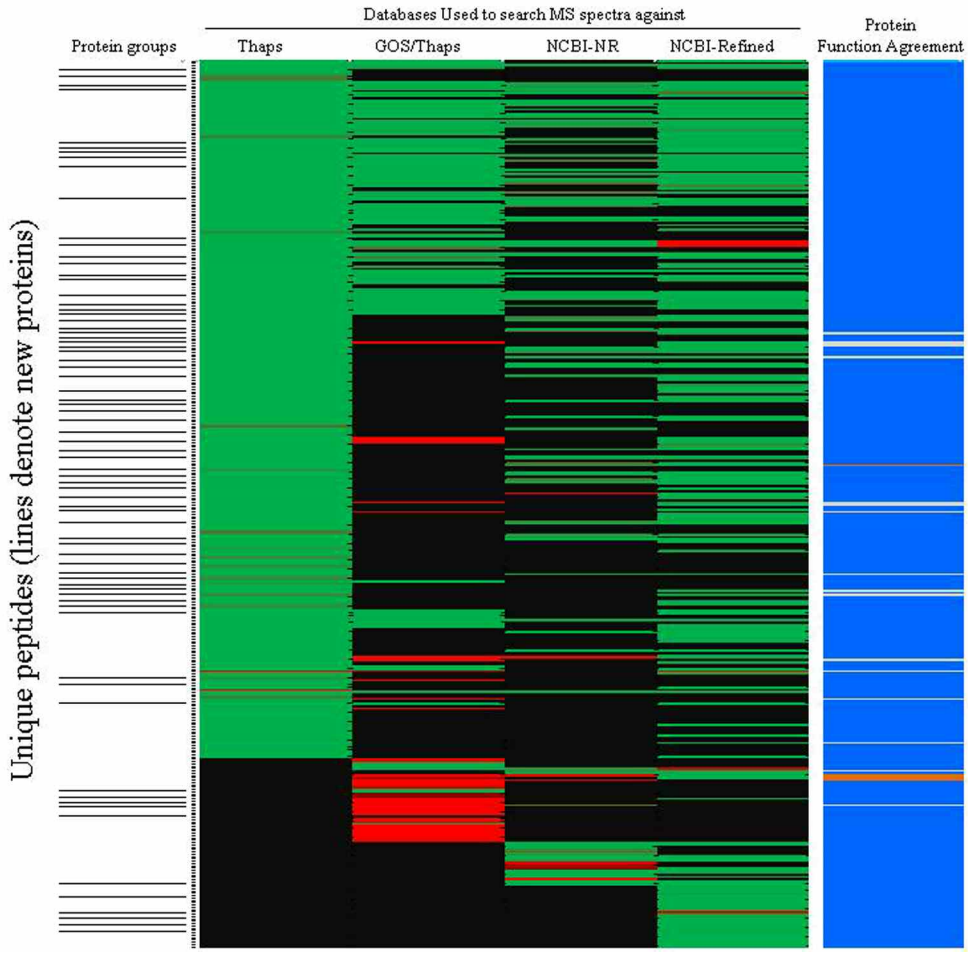

Fig. 4. Species assignment and protein function assignment comparisons from SEQUEST search of surface sediment slurry extraction of Bering Sea sediment. Mass spectra were obtained from 4 different databases: Thaps (column 1); GOS/Thaps (column 2); NCBI-NR (column 3); NCBI-Refined (column 4). 385 unique peptides are represented along the vertical axis and represented as dots. Peptides are grouped together by protein, black lines on the far left mark the beginning of a new group of peptides associated with the same protein. Species assignments are represented by shades of green or red and black. Green indicates the peptide was designated to originate from a marine or aquatic eukaryote. Red indicates the peptide was designated to originate from a marine, aquatic, or soil bacteria. For greens and reds, the brighter the color, the higher confidence PeptideProphet gave the assignment (e.g., $P > 0.95$), whereas lighter shades of red or green indicate poorer peptide correlations ($P < 0.9$). Black indicates that the peptide was not assigned to a protein using that particular database. The far right column is a color-coded chart to illustrate whether the function of the protein assigned is in agreement with the 4 databases searched. Blue indicates function is the same; gray indicates function is unknown by one or more databases; orange indicates function does not agree between database assignments. Functional agreement is present in over 95% of the peptides.

made with the large databases searched against Bering Sea surface sediment extract digests.

Interestingly, many of the peptides that were identified to originate from *T. pseudonana* when searched against the Thaps database were not confidently identified using either the GOS/Thaps or NCNI-NR databases. This results from the inability of PeptideProphet to decipher and report homologous peptides with confidence, which in this case reduced the number of protein identifications. PeptideProphet uses a correlation of two parameters: 1) how well the observed spectra matches the theoretical spectra (xcorr); 2) how different the first peptide match is from the second peptide match (fCorr). Typically, an assignment is made if the xcorr > 2.0 and fCorr > 0.1. When using a larger database, there is more peptide sequence similarity (e.g., SEVSALLGR, SEVSAILGR). As a result, PeptideProphet will assign a low fCorr to the second best peptide match. A low fCorr will decrease the overall statistical confidence and SEQUEST will not report any peptide match, even at high xcorr values.

The greatest number of identical peptide and protein assignments from different database searches was observed when searching the Thaps and GOS/Thaps databases (Fig. 3B). As mentioned earlier, all databases included *T. pseudonana* and the protein assignments that were identical between the GOS/Thaps and Thaps database searches were all *T. pseudonana* in origin. Despite fewer identified proteins, the larger databases do provide breadth to the sources of conserved proteins. Peptides from several identified proteins were conserved among *T. pseudonana* and hundreds of other organisms. Given the context of the system, seasonally diatom-dominated Bering Sea, and that other peptides are predominantly identified uniquely to *T. pseudonana* using all three databases, it is likely that these conserved proteins are also diatom in origin.

Given that the Thaps database search of the slurry method yielded the greatest number of identifications, it is not surprising that the Thaps search against the slurry method data also identified a suite of proteins with the greatest range of isoelectric points (pI). Among Thaps-identified proteins, a total of 43 slurry tube gel proteins and only 5 traditional tube gel proteins were identified with an isoelectric point above eight. The isoelectric point of a molecule is the pH at which the molecule carries no net charge, and is also the point at which the solubility of the molecule is at its lowest. The much larger proportion of high pI proteins identified illustrates the greater electrophoretic extraction achieved by the slurry tube gel method.

Basic proteins with high pI carry a more negative charge and would likely be more tightly bound to positively charged functional groups in sediments (Henrichs and Sugai 1993). The extraction buffer used in this study, containing high concentrations of protein solubilizing reagents (urea, thiourea, CHAPS, EDTA) is slightly acidic and thus not as effective at extracting basic proteins as electrophoretically assisted extraction. Of the 37 slurry tube gel identified proteins with a pI > 9, 20 are structural constituents of ribosomes. No ribosome structural constituents were identified using the traditional tube gel technique. This shows that the slurry tube gel method not only extracts a greater number of proteins than the traditional tube gel method, but a wider range of protein functionalities as well.

Recommendations

This is the first study to use gel electrophoresis of mixed matrices as an extraction method for the recovery of protein from sedimentary material. Initial results, further supported by optimization evaluations, allowed us to hypothesize that electric current disrupts the interactions between protein and sediment to mobilize protein into the electrophoresis gel. Optimization experiments have shown that the most effective extraction of peptides from sediments occurs using urea as the extraction buffer, pre-cast 1-D flat gels and the loading of the sediment and buffer slurry combination directly. By mixing the extraction buffer with the sediment before loading on the gel, proteins are solubilized and removed more efficiently from the particles, while the gel is an excellent trapping matrix for the proteins so that contaminants can be adequately washed away before enzymatic digestion and MS analysis. This method is applicable to a wide range of sediments, although actual recoveries would be expected to vary depending on input of primary production to sediment, time frame, and perhaps mineral matrix. Important for future studies is the observation that complex protein databases, while providing more potential protein sources, do not necessarily translate into a greater number or more confident protein identifications. Nevertheless, functional-level information is retained, suggesting that identifying proteins from mixed (often unknown) communities can be accomplished at the protein function-level, although determining and/or targeting the specific species the protein originated from remains difficult. The large contribution of unique algal-specific proteins (e.g., fucoxanthin chlorophyll *a*/c binding proteins) in Bering Sea sediment from searches of the simple Thaps database and the complex GOS and NCBI-NR databases indicates that core proteins associated with primary production are a major source of protein material retained in the continental shelf sediment of this system.

References

Aebersold, R., and D. R. Goodlett. 2001. Mass spectrometry in proteomics. Chem. Rev. 101(2):269-295 [doi:10.1021/cr990076h].

Armbrust, E. V., and others. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science 306(5693):79-86 [doi:10.1126/science.1101156].

Bastida, F., J. L. Moreno, C. Nicolas, T. Hernandez, and C. Carcia. 2009. Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. Eur. J. Soil Sci. 60:845-859 [doi:10.1111/j.1365-2389.2009.01184.x].

Belluomini, G., M. Branca, G. Calderoni, and M. Schnitzer. 1986. Distribution and geochemical significance of amino-acids and amino-sugars in a clay suite of the pliocene pleistocene age from central italy. Org. Geochem. 9(3):127-133 [doi:10.1016/0146-6380(86)90102-6].

Botelho, D., M. J. Wall, D. B. Vieira, S. Fitzsimmons, F. Liu, and A. Doucette. 2010. Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. J. Proteome Res. 9(6):2863-2870 [doi:10.1021/pr900949p].

Chen, J., S. Ryu, S. A. Gharib, D. R. Goodlett, and L. M. Schnapp. 2008. Exploration of the normal human bronchioalveolar lavage fluid proteome. Proteomics Clin. Appl. 2(4):585-595 [doi:10.1002/prca.200780006].

Chen, M., Y. P. Huang, P. G. Cai, and L. D. Guo. 2003. Particulate organic carbon export fluxes in the Canada Basin and Bering Sea as derived from $^{234}Th/^{238}U$ disequilibria. Arctic 56(1):32-44.

Cheng, C. N., R. C. Shufeldt, and F. J. Stevenson. 1975. Amino acid analysis of soils and sediments: extraction and desalting. Soil Biol. Biochem. 7(2):143-151 [doi:10.1016/0038-0717(75)90012-7].

Collins, M. J., A. N. Bishop, and P. Farrimond. 1995. Sorption by mineral surfaces: Rebirth of the classical condensation pathway for kerogen formation? Geochim. Cosmochim. Acta 59(11):2387-2391 [doi:10.1016/0016-7037(95)00114-F].

Cowie, G. L., and J. I. Hedges. 1992a. Sources and reactivities of amino acids in a coastal marine environment. Limnol. Oceanogr. 37:703-724 [doi:10.4319/lo.1992.37.4.0703].

———, and ———. 1992b. Improved amino acid quantification in environmental-samples—charge-matched recovery standards and reduced analysis time. Mar. Chem. 37(3-4):223-238 [doi:10.1016/0304-4203(92)90079-P].

Craig, O. E., and M. J. Collins. 2000. An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. J. Immunol. Methods 236:89-97 [doi:10.1016/S0022-1759(99)00242-2].

Davis, J. A. 1982. Adsorption of natural dissolved organic matter at the oxide/water interface. Geochim. Cosmochim. Acta 46:2381-2393 [doi:10.1016/0016-7037(82)90209-5].

Davis, M. T., and others. 2001. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. II. Limitations of complex mixture analyses. Proteomics 1(1):108-117 [doi:10.1002/1615-9861(200101)1:1<108::AID-PROT108>3.0.CO;2-5].

Dong, H. P., D. Z. Wang, M. Dai, and H. S. Hong. 2010. Characterization of particulate organic matters in the water column of the South China Sea using a shotgun proteomic approach. Limnol. Oceanogr. 55(4):1565-1578 [doi:10.4319/lo.2010.55.4.1565].

Dufresne, A., and others. 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc. Nat. Acad. Sci. U.S.A. 100(17):10020-10025 [doi:10.1073/pnas.1733211100].

Dunning, J. D., B. J. Herren, R. W. Tipps, and R. S. Snyder. 1982. Fractionation of mineral species by electrophoresis. J. Geophys. Res. 87:781-788 [doi:10.1029/JB087iB13p10781].

Eng, J. K., 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. Amer. Soc. Mass Spectrom. 5:976-989 [doi:10.1016/1044-0305(94)80016-2].

———, B. Fischer, J. Grossmann, and M. J. MacCoss. 2008. A fast SEQUEST cross correlation algorithm. J. Proteome Res. 7(10):4598-4602 [doi:10.1021/pr800420s].

Fairbanks, G., C. Levinthal, and R. H. Reeder. 1965. Analysis of C14-labeled proteins by disc electrophoresis. Biochem. Biophys. Res. Commun. 20:393-399 [doi:10.1016/0006-291X(65)90589-9].

———, T. L. Steck, and D. F. H. Wallach. 1971. Electrophoretic analysis of the major proteins of the human erythrocyte membrane. Biochemistry 10:2606-2617 [doi:10.1021/bi00789a030].

Giovannoni, S. J., and others. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242-1245 [doi:10.1126/science.1114057].

Graves, P. R., and T. A. J. Haystead. 2002. Molecular biologist's guide to proteomics. Microbiol. Mol. Biol. Rev. 66(1):39-63 [doi:10.1128/MMBR.66.1.39-63.2002].

Greenland, D. J. 1971. Interactions between humic and fulvic acids and clays. Soil Sci. 111:34-41 [doi:10.1097/00010694-197101000-00004].

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

Henrichs, S. M., and S. F. Sugai. 1993. Adsorption of amino-acids and glucose by sediments of Resurrection Bay, Alaska, USA—functional-group effects. Geochim. Cosmochim. Acta 57(4):823-835 [doi:10.1016/0016-7037(93)90171-R].

Hirano, H., S. Komatsu, H. Takakura, F. Sakiyama, and S. Tsunasawa. 1992. Deblocking and subsequent microsequence analysis of N-alpha-blocked proteins electroblotted onto PVDF membrane. J. Biochem. 111(6):754-757.

Horsfall, I. M., and G. A. Wolff. 1997. Hydrolysable amino acids in sediments from the Porcupine Abyssal Plain, northeast Atlantic Ocean. Org. Geochem. 26(5-6):311-320 [doi:10.1016/S0146-6380(97)00015-6].

Kan, J., T. E. Hanson, J. M. Ginter, K. Wang, and F. Chen. 2005. Metaproteomic analysis of Chesapeake Bay bacterial communities. Saline Syst. 1(7):1-13.

Keil, R. G., D. B. Montlucon, F. G. Prahl, and J. I. Hedges. 1994. Sorptive preservation of labile organic matter in marine sediments. Nature 370(6490):549-552 [doi:10.1038/370549a0].

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(20):5383-5392 [doi:10.1021/ac025747h].

Knicker, H., and P.G. Hatcher. 1997. Survival of protein in an organic-rich sediment: possible protection by encapsulation in organic matter. Naturwissenschaften 84(6):231-234 [doi:10.1007/s001140050384].

Kuster, B., A. P. Hunter, S. F. Wheeler, R. A. Dwek, and D. J. Harvey. 1998. Structural determination of N-linked carbohydrates by matrix-assisted laser desorption/ionization mass spectrometry following enzymatic release within sodium dodecyl sulphate polyacrylamide electrophoresis gels: Application to species-specific glycosylation of alpha(1)-acid glycoprotein. Electrophoresis 19(11):1950-1959 [doi:10.1002/elps.1150191113].

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685 [doi:10.1038/227680a0].

Laver, W. G. 1964. Structural studies on the protein subunits from three strains of influenza virus. J. Mol. Biol. 9:109-124 [doi:10.1016/S0022-2836(64)80094-2].

Limmer, A. W., and A. T. Wilson. 1980. Amino-acids in buried paleosols. J. Soil Sci. 31(1):147-153 [doi:10.1111/j.1365-2389.1980.tb02072.x].

Lourenco, S. O., E. Barbarino, U. M. L. Marquez, and E. Aidar. 1998. Distribution of intracellular nitrogen in marine microalgae: basis for the calculation of specific nitrogen-to-protein conversion factors. J. Phycol. 34(5):798-811 [doi:10.1046/j.1529-8817.1998.340798.x].

Maizel, J. V. 2000. SDS polyacrylamide gel electrophoresis. Trends Biochem. Sci. 25(12):590-592 [doi:10.1016/S0968-0004(00)01693-5].

Mayer, L., L. Benninger, M. Bock, D. DeMaster, Q. Roberts, and C. Martens. 2002. Mineral associations and nutritional quality of organic matter in shelf and upper slope sediments off Cape Hatteras, USA: a case of unusually high loadings. Deep-Sea Res. II. 49(20):4587-4597 [doi:10.1016/S0967-0645(02)00130-3].

Mayer, L. M. 1994. Surface area control of organic carbon accumulation in continental shelf sediments. Geochim. Cosmochim. Acta 58(4):1271-1284 [doi:10.1016/0016-7037(94)90381-6].

McRoy, C. P. 1987. Global maximum of primary production in the North Bering Sea. E.O.S. Comm. 68:172.

Mintrop, L., and J. C. Duinker. 1994. Depth profiles of amino-acids in porewater of sediments from the Norwegian—Greenland Sea. Oceanol. Acta 17(6):621-631.

Morris, R. M., B. L. Nunn, C. Frazar, D. R. Goodlett, Y. S. Ting, and G. Rocap. 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J. 4(5):673-685 [doi:10.1038/ismej.2010.4].

Nesvizhskii, A., A. Keller, E. Kolker, and R. Aebersold. 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75(17):4646-4658 [doi:10.1021/ac0341261].

Nguyen, R. T., and H. R. Harvey. 2001. Protein preservation in marine systems: hydrophobic and other non-covalent associations as major stabilizing forces. Geochim. Cosmochim. Acta 65:1467-1480 [doi:10.1016/S0016-7037(00)00621-9].

———, and ———. 2003. Preservation via macromolecular associations during *Botryococcus braunii* decay: proteins in the Pula Kerogen. Org. Geochem. 34:1391-1403 [doi:10.1016/S0146-6380(03)00154-2].

Nunn, B. L., and R. G. Keil. 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].

———, and A. T. Timperman. 2007. Marine proteomics. Mar. Ecol. Progr. Ser. 332:281-289 [doi:10.3354/meps332281].

———, Y. S. Ting, L. Malmstrom, Y. S. Tsai, A. Squier, D. R. Goodlett, and H. R. Harvey. 2010. The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation. Limnol. Oceanogr. 55(4):1790-1804 [doi:10.4319/lo.2010.55.4.1790].

Ogunseitan, O. A. 1993. Direct extraction of proteins from environmental samples. J. Microbiol. Methods 17(4):273-281 [doi:10.1016/0167-7012(93)90056-N].

Pederson, T. 2008. Turning a PAGE: the overnight sensation of SDS-polycrylamide gel electrophoresis. FASEB J. 22(4):949-953.

Pruitt, K., T. Tatusova, and D. Maglott. 2002. The Reference Sequence (RefSeq) Project. *In* J. McEntyre and J.Ostell [eds.], The NCBI handbook. National Center for Biotechnology Information.

Quince, C., T. P. Curtis, and W. T. Sloan. 2008. The rational exploration of microbial diversity. ISME J. 2:1997-2006 [doi:10.1038/ismej.2008.69].

Rashid, M. A., D. E. Buckley, and K. R. Robertson. 1972. Interactions of a marine humic acid with clay minerals and a natural sediment. Geoderma 8:11-27 [doi:10.1016/0016-7061(72)90029-8].

Reisfeld, R. A., U. J. Lewis, and D. E. Williams. 1962. Disc electrophoresis of basic proteins and peptides on polyacrylamide gels. Nature 195:281-283 [doi:10.1038/195281a0].

Reuss, F. F. 1809. The first observation of electrokinetic interactions. Mem. Soc. Imperiale Naturalists de Moscow 2:327.

Roth, L. C., and H. R. Harvey. 2006. Intact protein modification and degradation in estuarine environments. Mar. Chem. 102:33-45 [doi:10.1016/j.marchem.2005.10.025].

Sambrotto, R. N., H. J. Niebauer, J. J. Goering, and R. L. Iverson. 1986. Relationships among vertical mixing, nitrate uptake, and phytoplankton growth during spring bloom in the S-E Bering Sea. Cont. Shelf Res. 5:161-198 [doi:10.1016/0278-4343(86)90014-2].

Scherl, A., S. A. Shaffer, G. K. Taylor, H. D. Kulasekara, S. I. Miller, and D. R Goodlett. 2008. Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. Anal. Chem. 80(4):1182-1191 [doi:10.1021/ac701680f].

Schulze, W. X., G. Gleixner, K. Kaiser, G. Guggenberger, M. Mann, and E. D. Schulze. 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia 142:335-343 [doi:10.1007/s00442-004-1698-9].

Shapiro, A. L., E. Vinuela, J. V. Maizel Jr. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochem. Biophys. Res. Commun. 28:815-820 [doi:10.1016/0006-291X(67)90391-9].

Shevchenko, A., M. Wilm, O. Vorm, and M. Mann. 1996. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Anal. Chem. 68(5):850-858 [doi:10.1021/ac950914h].

Spahr, C. S., and others. 2001. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. Proteomics 1(1):93-107 [doi:10.1002/1615-9861(200101)1:1<93::AID-PROT93>3.0.CO;2-3].

Tran, J. C., and A. A. Doucette. 2009. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. Anal. Chem. 81(15):6201-6209 [doi:10.1021/ac900729r].

Waldhier, M. C., K. Dettmer, M. A. Gruber, and P. J. Oefner. 2010. Comparison of derivatization and chromatographic methods for GC-MS analysis of amino acid enantiomers in physiological samples. J. Chromatogr. B 878(15-16):1103-1112 [doi:10.1016/j.jchromb.2010.03.021].

Walsh, J. J., and others. 1989. Carbon and nitrogen cycling within the Bering Chukchi Seas—Source regions for organic-matter effecting AOU demands of the Arctic-Ocean. Progr. Oceanogr. 22(4):277-359 [doi:10.1016/0079-6611(89)90006-2].

Wang, X. C., and C. Lee. 1993. Adsorption and desorption of aliphatic amines, amino acids and acetate by clay minerals and marine sediments. Mar. Chem. 44:1-23 [doi:10.1016/0304-4203(93)90002-6].

Washburn, M. P., D. Wolters, and J. R. Yates. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19:242-247 [doi:10.1038/85686].

Yi, E. C., M. Marelli, H. Lee, S. O. Purvine, R. Aebersold, J. D. Aitchison, and D. R. Goodlett. 2002. Approaching complete peroxisome characterization by gas-phase fractionation. Electrophoresis 23:3205-3216 [doi:10.1002/1522-2683(200209)23:18<3205::AID-ELPS3205>3.0.CO;2-Y].

Yooseph, S., and others. 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol. 5(3):432-466 [doi:10.1371/journal.pbio.0050016].

Zang, X., J. D. H. van Heemst, 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. Org. Geochem. 31(7-8):679-695 [doi:10.1016/S0146-6380(00)00040-1].

Submitted 19 July 2011
Revised 16 March 2012
Accepted 6 April 2012