Vol. 332: 281–289, 2007 | MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser | Published March 5

OPEN ACCESS

Marine proteomics

Brook L. Nunn[1,2], Aaron T. Timperman[3,*]

[1]Medicinal Chemistry Department, University of Washington, Box 355350, Seattle, Washington 98155, USA
[2]Marine Chemistry Department, University of Otago, Box 46, Dunedin, New Zealand
[3]C. Eugene Bennett Department of Chemistry, West Virginia University, Prospect St, Morgantown, West Virginia 26509-6045, USA

ABSTRACT: A wealth of information is recorded in a protein's primary sequence, which can be used to determine its biological function and origin, and provide clues to the mechanisms of degradation. In contrast to DNA, proteins and their amino acid constituents have demonstrated a wide-spread presence outside the cell, preserved in the environment. In marine samples, proteins are present as mixtures from numerous sources in a salty, complex matrix at low concentrations. As a result of these factors, studies of this nitrogen-based component in the oceans have previously been limited to bulk elemental and amino acid analyses; these analyses were incapable of providing details regarding protein sequence, function and source information. Advances in biological mass spectrometry now allow for the analysis and characterization of the protein component from the marine environment. Proteomic mass spectrometry is a high-throughput analysis of protein mixtures that does not require any prior knowledge of the original protein structures in the mixture, making it an ideal technique for marine studies. Potential marine applications of proteomics include: analyzing organisms cultured under different nutrient conditions to examine cellular expression and adaptation, profiling the marine dissolved and particulate organic matter pools to determine source information and understand long-term carbon preservation, and verifying genomic findings with proteomic analyses to determine which genes are translated and to what extent the protein is expressed. Although some major advances in marine studies and mass spectrometry have been made, there remains a significant amount of methods development and community education before the full potential of proteomics is reached.

KEY WORDS: Seawater · Protein · Mass spectrometry · Genomics · Dissolved organic matter · DOM · Particulate organic matter · POM

INTRODUCTION

Many researchers in the marine community focus their attention on better understanding the cycling and transfer of carbon in the world's oceans. The political and social awareness of global warming is currently a driving force toward an improved understanding of this pool of carbon and how it interacts with the atmosphere and influences the world's climate. As a result, there is an increasing need for more accurate models of the cycling and transfer of carbon throughout the oceans. Historically, oceanographers examined the carbon pool at the elemental level, analyzing bulk carbon concentrations, later followed by monomeric molecular level analysis including carbohydrates (Hecky 1973, Lyons et al. 1979, Cowie & Hedges 1984), hydrocarbons (Nissenbaum et al. 1971, Prahl 1985), lignin (Prahl 1985), and amino acids (King 1974, Lee & Cronin 1982, Henrichs & Farrington 1987, Lee 1988). Through more detailed, non-destructive analyses we can gain additional information on the origin and fate of these organic molecules. Recent advances and applications in molecular-level analyses, such as mass spectrometry and nuclear magnetic resonance (NMR), are now being applied to marine samples to gain a better understanding of size distribution and structures of the original organic polymeric molecules present (e.g. Minor et al. 2003, Kujawinskia et al. 2004, Li et al. 2004, Aluwihare et al. 2005).

One of the remaining untapped reservoirs of information is locked up in molecules that are common to all life and also persist in the environment as discrete

*Corresponding author. Email: atimperm@wvu.edu

units: proteins. Unlike the previous polymers analyzed (e.g. carbohydrates or hydrocarbons), proteins are the result of a precise arrangement of their monomeric constituents — amino acids — where the composition and sequence can be specific to source organisms and/or cellular function. Proteins comprise most of a cell's machinery and have many important functions including structural integrity, energy transfer, and cellular death. Protein expression is an important indicator of cellular state and can provide information on the activation of various cellular pathways, while the survival of particular proteins in the marine environment (e.g. in the dissolved or sedimentary pool) can provide insight into mechanisms that control the degradation of organic matter.

Proteins and their precursors, amino acids, are widespread in a variety of marine environments at significant enough concentrations to be considered an important contributor to the carbon and nitrogen pools (Hedges 1991, Benner et al. 1992, Keil et al. 1994, McCarthy et al. 1998, Horiuchi et al. 2004) (Fig. 1). Older protein-identification technologies allowed for the isolation and sequencing of proteins on a protein-by-protein basis. Because most marine investigators that are interested in this component of carbon are not looking for a specific protein, but instead the identification of any and all proteins in the system, the technology was not compatible with their requirements. Proteomics is a high-throughput analysis for the rapid identification of known or unknown protein mixtures in complex systems. The emergence of proteomics will allow investigators to sequence and enumerate as many proteins as possible from the system, and determine if these proteins change as a response to stimuli or environmental condition. With the improvements in technology and advancements in proteomics, marine investigations will now be able to gain greater information by examining these C and N components at a higher molecular level. Throughout the present study we discuss several themes and questions that have been previously approached by marine investigators; however, with the exception of a few studies (Tanoue 1996, Powell et al. 2005), all prior investigations on the protein component in the ocean have been limited to the analysis of amino acids rather than peptides and proteins. The goals of the present study are to: (1) introduce proteomic mass spectrometry and clarify some common misconceptions of data interpretation; (2) introduce potential applications of proteomics in the marine field; and (3) provide some ideas on how to advance the community at the pace of the technology. Although this technique is in its infancy in the marine field, it has the ability to provide many clues to the sources and transformation of carbon in the oceans.

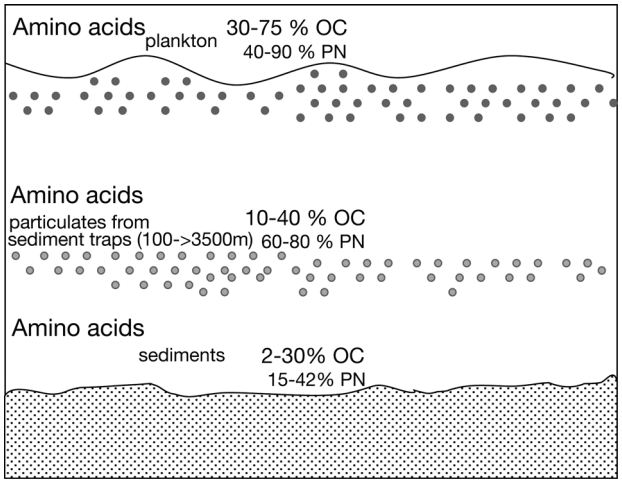

Fig. 1. Amino acids contribute a significant fraction of the percentage of organic carbon (OC) and particulate nitrogen (PN) in plankton (Siezen & Mague 1978, Lee & Cronin 1982, Lee & Olson 1984, Nguyen & Harvey 1994, 1997, Wakeham et al. 1997, Keil 1999 and references therein), particulates from sediment traps over a range of depths (Nguyen & Harvey 1994, Wakeham et al. 1997, Keil 1999 and references within), and in the coastal and deep ocean sediments (Wakeham et al. 1997, Keil 1999 and references therein, Nunn 2004). This, combined with other experimental evidence, strongly suggests that knowledge of the cycling and preservation of proteins in the marine environment is critical for understanding the global carbon cycle

METHODS OF PROTEIN ANALYSIS

Proteins are polymers consisting of a mixture of 20 genetically encoded amino acid monomers. The objective of protein analyses is to determine the order and number of amino acid residues that are covalently linked in a linear chain, referred to as the primary sequence. The primary sequence dictates how the protein is folded locally (secondary structure) and what form it takes 3-dimensionally (tertiary structure), which ultimately results in its biological role. The initial starting point for primary sequence analysis is to disrupt or denature its 3-dimensional structure, thereby unfolding the protein to make it more accessible for analysis. Previously, the majority of oceanic protein analyses involved complete hydrolysis of all peptide linkages, breaking proteins into the original amino acid monomers: complex mixtures of proteins and peptides were chemically hydrolyzed (150°C, 6 N HCl, 1 h) to amino acids for interpretation (e.g. Cowie & Hedges 1992, Keil & Kirchman 1993, McCarthy et al. 1997, Nunn & Keil 2005). As a result, any information that might have been gained pertaining to the sequence, structure, function or source of the protein was lost. Advances in biological mass spectrometry allow for mixtures of proteins to be analyzed from

more complex matrices and their primary sequences to be determined, thereby making the technique more informative to oceanographers.

Proteomic methods can be divided into techniques that analyze peptide fragments from the proteins, referred to as bottom-up protein analysis (Fig. 2), and those that analyze whole proteins in the mass spectrometer (MS), a top-down protein analysis (Reid & McLuckey 2002). In the bottom-up approach, proteins are cleaved into peptides to produce shorter segments

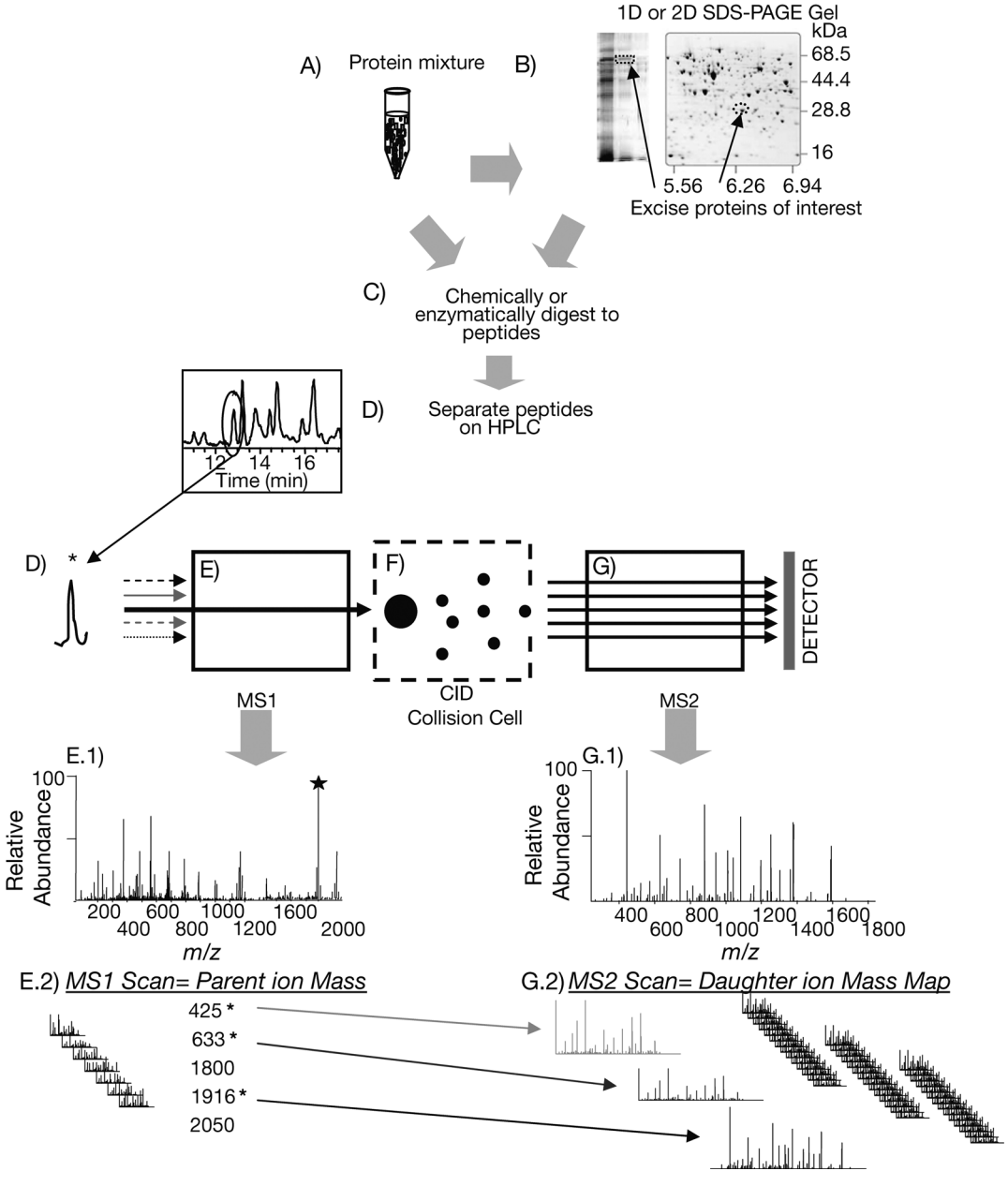

Fig. 2. Bottom-up proteomic project for protein identification from peptides using tandem mass spectrometry (MS). (A) Protein mixtures or (B) isolated proteins from gel electrophoresis can be sequenced using tandem MS approaches. The individual protein or mixture is first chemically or enzymatically digested into peptides (C). Complex mixture of peptides are then separated using inline HPLC (D) prior to injection and ionization in the mass spectrometer. As individual peptides (D) elute off the chromatography column they are ionized and analyzed in the first mass spectrometer (E: MS1). Mass to charge (m/z) ratios are measured (E.1), yielding the parent-ion scan. The analyst can then isolate single peptides (e.g. the 3 most intense peaks; E.2) from the parent-ion scan for fragmentation (F) and sequence determination. Each selected ion from the parent-ion scan is then individually fragmented and sent to the second MS (G: MS2), yielding daughter-ion scans (G.1). Sequence analysis is then performed using all parent-ion scans (E.2) and their respective daughter-ion scans (G.2). Interpretation of daughter-ion scans for the purpose of peptide sequencing is described in Fig. 3

that are more amenable to sequencing in the MS than whole proteins. Because peptides are unique to specific proteins, peptide tags or short peptide sequences that are determined experimentally can then be used to search the databases for the parent protein (e.g. Powell et al. 2005). Identification of more than 1 peptide unique to a protein is commonly used to infer the presence of the entire intact protein. As a result, bottom-up analyses excel at protein identification when combined with database searches. In contrast, the top-down approach analyzes whole proteins in the MS and can provide complete sequence coverage. The top-down method is therefore best suited for the analysis of protein modifications such as phosphorylations. Fragmentation of whole proteins (top-down) or peptides (bottom-up) can be achieved in the MS using one of a variety of dissociation technologies (e.g. electron capture dissociation). The present study focuses on the bottom-up approach of peptide sequencing and protein identification.

PEPTIDE SEQUENCING USING TANDEM MASS SPECTROMETRY

A basic knowledge of the fundamentals of peptide sequencing by tandem mass spectrometry is essential for understanding the potential applications for this technology; a more detailed description can be found in a number of recent publications (e.g. Fenn et al. 1989, Mann & Wilm 1994). Tandem mass spectrometry takes advantage of 3 properties of proteins: (1) the building blocks of proteins are known; (2) proteins can be cleaved into peptides; and (3) protonated peptides fragment in a predictable manner, producing product ion spectra that are reproducible and interpretable. The most commonly used proteomic method begins with the isolation of proteins using gel electrophoresis, followed by excision from the gel and proteolytic digestion of the protein using an enzyme, typically trypsin (Fig. 2A–C). The resulting peptides are then extracted and separated using reversed-phase high-performance liquid chromatography (HPLC, Fig. 2D), ionized, and the parent ion mass to charge (m/z) ratios are measured in the MS (Fig. 2E). An individual peptide parent ion can then be selected and isolated for fragmentation in the MS (Fig. 2F); the resulting m/z ratio values of the fragmented parent ions are measured, yielding a tandem mass spectrum (Fig. 2G). This ion isolation process is critical because it ensures that the fragment ions are from the selected parent ion, making this method extremely well suited to the analysis of complex mixtures.

In the positive ion mode, basic amino acid residues in the peptides are protonated. Frequently, tryptic peptides are doubly charged (+2) because both the amino-terminus (N-terminus) and the basic residue at the caboxy-terminus (C-terminus) are positively charged. The proton associated with the N-terminus in solution is mobile in the gas phase, allowing it to migrate along the peptide backbone and directing fragmentation to the adjacent amide bond. When fragmentation occurs at an amide bond, fragmentation ions that contain the N-terminal residue are called b-ions, whereas fragmentation ions that contain the C-terminal residue are referred to as y-ions. Fragmentation of a +2 parent ion typically results in a b- and y-ion that are each singly charged. Different members of the peptide ion population will typically break at different amide bonds, yielding b- and y-ion series (Fig. 3). The mass differences between singly charged ions that are contiguous in the series correspond to the amino acid residue masses; additionally, the residue order is encoded in the mass ladder (Fig. 3).

A single HPLC-MS run can produce thousands of spectra, making automated data filtering and interpretation a requirement (Hirosawa et al. 1993, Perkins et al. 1999). Automated analysis is typically achieved by comparing the experimentally obtained fragment ion spectra, with theoretical spectra mathematically predicted from the sequences in both genomic and protein databases. To perform correlative database sequence searching, the analyst typically provides the software with 3 pieces of information: the organism's full proteome (or genome for translation), the enzyme that was used for the digestion, and any chemical modifications or adducts that might be present (methylation, Na^+ adducts, etc.). Using scoring algorithms to rank the spectra, the software then returns a list of proteins, with their respective peptides identified, a final percent of protein sequenced, a correlation score, and HTML-links to the individual peptides' spectra for direct scrutiny. Since database correlation routines always return a match, proper filtering and manual verification are required to maintain reliability. For automated correlative database protein identifications, an important point of emphasis is that the interrogated protein or peptide sequence must be in the database in order for it to be properly identified. However, if the fragment-ion series is complete enough for a given peptide, the amino acid sequence can be mathematically interpreted directly from the tandem MS spectrum, either manually or using computerized algorithms (see Fig. 3). This ability to perform protein sequencing without depending on any prior knowledge of the amino acid sequence (de novo sequencing) is critical for environmental samples such as seawater, because only a small fraction of the contributing organisms' genomes have been sequenced (Powell et al. 2005). De novo sequencing programs are currently

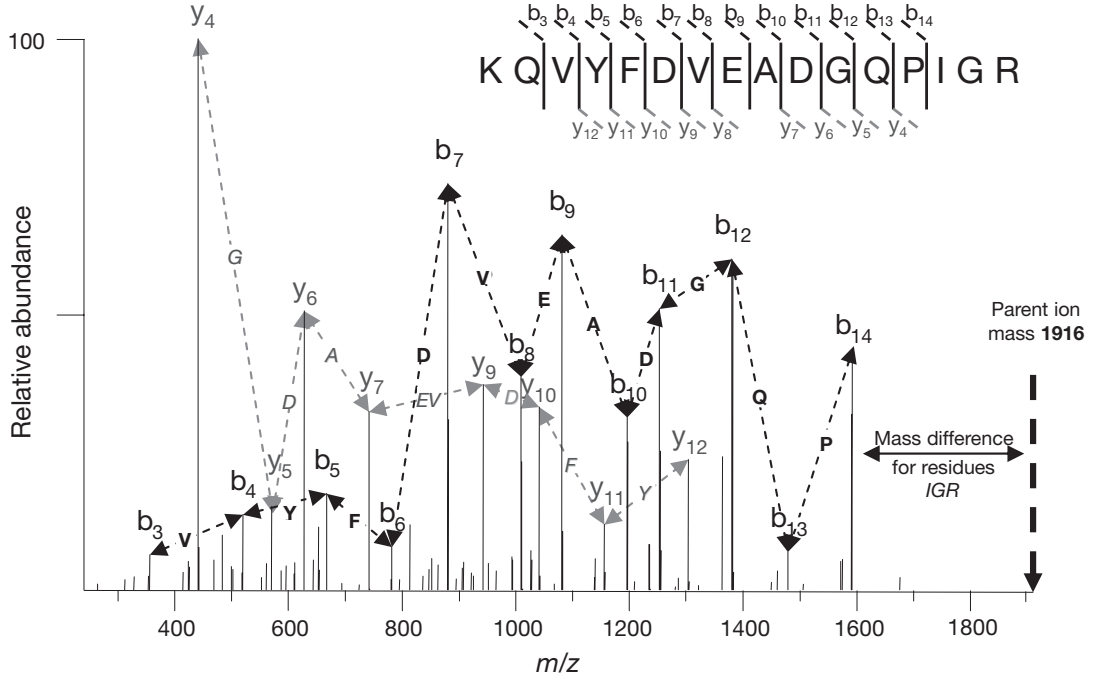

Fig. 3. How to de novo sequence a mass-to-charge daughter-ion spectra produced from the fragmentation of a tryptic peptide using tandem mass spectrometry (parent ion m/z = 1916 Da). Amino acid sequence of the original peptide denoted by amino acid single letter codes—amino-terminus is lysine (K), carboxy-terminus is arginine (R)—is indicated in the top-right corner. The b-ions (black) include all ions produced during fragmentation that retained the amino terminus (e.g. KQ, KQV, KQVY, etc.). The y-ion series (grey) consist of all ions produced during fragmentation that retained the carboxy-terminus (e.g. RGIP, RGIP, RGIPQ, etc.). The daughter-ion spectra is a plot of the m/z ratio of each ion produced during the fragmentation of the parent peptide and their relative abundance (y-axis). Mass differences between ions in the spectra are indicative of monoisotopic amino acid residue masses (e.g. b_{14} to b_{13} = 97 Da: Proline-**P**; b_{13} to b_{12} = 128 Da: Glutamine-**Q**). Dashed black lines indicate mass differences between b-ion series with the respective amino acid in the center of the line, and grey dashed lines with arrows are the mass differences between the y-ion series with their respective amino acids in the center of the line. Not all ions in both series need to be present to decipher the original amino acid sequence of the peptide

available and use scoring routines with similar caveats to database correlation programs.

For either approach it is important to note that the quality of the data is a major factor in the reliability and confidence of the sequences obtained. All aspects of sample preparation, separation and MS analysis can affect the data quality. Mass spectrometers that provide higher mass accuracy, resolving power, and signal-to-noise ratios produce higher quality data that will provide greater data reliability.

APPLICATIONS FOR MARINE PROTEOMICS

Current work in marine proteomics can be divided into 2 broad areas: the recovery and analysis of proteins from the marine environment, and the analysis of proteins from cultured organisms. Characterization of proteins directly from seawater can be used to determine the structures of proteins that are resistant to degradation and accumulate to detectable levels. The isolation and characterization of proteins from marine samples, such as seawater, porewater, particulates, and sediments, will greatly improve our understanding of the sources and mechanisms that control the cycling and long-term preservation of organic matter. For years the marine community has been limited to the examination of amino acids; full characterization of proteins from marine samples will provide a description of dissolved organic matter (DOM) components at the molecular level. Through the sequencing of these proteins and peptides, we can potentially gain information about the presence or past existence of an organism in a sample and the original function of the protein. We can also identify specific protein families, domains or themes preferentially preserved or any chemical modifications or adductions that might have enhanced the proteins' preservation. Through the combination of all these analyses we can greatly exceed previous elemental-level investigations by providing clues to what environmental conditions might encourage or discourage long-term preservation of carbon and nitrogen within the ocean. In a more directed strategy, specific proteins have been injected

into sediments and their degradation followed as a function of time (Nunn et al. 2003). These studies have shown that the model proteins used degrade rapidly, in the order of weeks.

Cultures of marine plankton can be used to determine what proteins from their genome are expressed and the relative levels at which these proteins are excreted, or released, into the surrounding environment. Since an organism's expressed proteome is dynamic, cellular protein expression changes as a function of environmental conditions. As a result, proteomics allows investigators to determine how organisms are able to biochemically cope and respond to varying environmental stresses. For example, proteins excreted from an organism into the surrounding medium could have one of numerous functions, including organism-to-organism communication or signaling (e.g. Wisniewska et al. 2003), or as an aid in the digestion or acquisition of nutrients, or as a microbial deterrent (e.g. Thomas et al. 2004). Isolating and sequencing these excreted proteins can inform us as to how the organisms biochemically manage and respond to their surroundings. In many parts of the world's oceans, different nutrients are in high demand as a result of being present at very dilute concentrations. A wide variety of organisms have adapted to these nutrient-deplete conditions and grow opportunistically when conditions are favorable. Thus, a long-standing question in the marine community concerns how these organisms are able to sequester the required nutrients from such dilute conditions. In many cases these questions can be answered using differential quantitative proteomics on organisms grown in culture with and without specific nutrients.

Controlled studies of cultured marine organisms can also improve our understanding of which peptide-linked molecules are most likely to contribute to the dissolved and particulate organic matter pools. Both relative protein expression levels and resulting protein products after extensive degradation can be analyzed and potentially quantified. Studies such as these may also provide information on relative resistance of different proteins to degradation, allowing for their selective enrichment and providing clues on long-term preservation (e.g. Nunn et al. 2003, Squier & Harvey 2006). Further insight into which degradation processes are most important may also be gained by controlled exposure of proteins to different enzymes, bacteria, light, or abiotic reactants. Using protein mass spectrometry, sequences and relative quantities of resulting peptide end-products can be obtained (Nunn et al. 2003, Peers & Price 2006). These types of experiments can provide the foundation for understanding which protein components comprise the recalcitrant dissolved and particulate organic matter pools in the ocean.

USING PROTEOMICS TO COMPLEMENT GENOMIC FINDINGS IN MARINE ECOSYSTEMS

In the past decade, ocean-based genomics has begun to explore the diversity, cellular evolution and adaptive abilities of marine organisms. Although this has provided the community with the beginnings of a database of microbial and eukaryotic blueprints, it does not necessarily translate into biochemical expression or phenotype. Genomics demonstrates which genes are shared, but proteomics can show clearer relationships by illustrating functional similarities and phenotypic variances. Through the use of pure genome sequences, open reading frames (ORFs) can be predicted, but they cannot be used to determine if or when transcription takes place or to what degree a protein is expressed. Proteomics can provide the researcher with more than the hypothetical cellular scenario. With a well-designed experiment, investigators can examine the conditions under which a protein is expressed (Nilsson & Davidson 2000, Kislinger & Emili 2003), its cellular location (Dunkley et al. 2004), the relative quantities (Yao et al. 2001, Molloy et al. 2005), and what protein–protein interactions take place (Giot et al. 2003, Schweitzer et al. 2003).

Because the ocean is one of the most dynamic environments in which organisms live, the success of a species depends on its ability to rapidly adapt to varying light, temperature and nutrient sources. Close examination of the genomes of oceanic microbes has already demonstrated that many of these organisms have the blueprints for diverse suites of organic and inorganic nitrogen and carbon transporters (Palenik et al. 2003, Armbrust et al. 2004). Proteomics can clarify if and to what extent various pathways are utilized, which environmental triggers act on the system, and relative protein-level response times. Additional information on protein expression levels in combination with gene expression will help investigators to clarify phylogenetic roots and possibly endosymbiotic events by highlighting dormant pseudo-genes, protein-level amino acid migrations, and mutations (Coin & Durbin 2004, Jaffe et al. 2004, Wirth et al. 2005). Through the use of proteomics, we may be able to simplify ocean-wide genomic investigations that are attempting to decipher evolutionary changes from ancestral cells. For example, instead of a broad-based survey of oceanic genomes, we can narrow the focus to a few directed analyses of proteins involved in specific biochemical pathways (Bibby et al. 2001, Strzepek & Harrison 2004, Peers & Price 2006).

THE FUTURE OF MARINE PROTEOMICS

If proteomic technology is beyond its tenth year (Wasinger et al. 1995), why is it that the marine field is

only recently beginning to use it as a tool to answer some of the community's questions? For many environmental investigators, molecular-level analyses have been impractical. The 3 primary reasons why the marine science field has taken so long to adopt the new technology are instrument availability, financial resources, and availability of trained personnel.

Excluding the proteomic investigation of cultured marine organisms, environmental protein analysts must contend with mixtures of proteins present at very low concentrations combined with complex matrices and relatively unknown sources. Prior to the recent investigation where large volumes (~100 l) of water were ultrafiltered to permit mass spectrometric characterization of the dissolved proteins (Powell et al. 2005), all previous investigations of this C and N pool involved amino acid hydrolysis and derivitization (Wakeham & Lee 1989, Cowie & Hedges 1992, McCarthy et al. 1997) as a means to circumvent low analyte concentrations present in high levels of contaminants. As proteomic technology is quickly being adopted by a number of different laboratories to investigate a wide variety of biological questions, rapid innovations and advances are being made to improve detection limits, sensitivity, and contaminant tolerance.

General improvements in the proteomics field are taking place, but because marine applications are in their infancy and it is such a specialized niche, there must first be substantial advances in the development of new methods. To analyze dissolved, exuded, preserved or particulate protein fractions from the ocean, samples must be collected (e.g. using sediment traps, large volumes of water, cultured organisms), extracted (e.g. chemically), de-contaminated (e.g. via chromatography, other chemical separation), isolated, or concentrated (e.g. via ultrafiltration, chemical precipitation, dialysis). MS techniques and instrumentation must then be optimized and a rigorous method for data analysis must be developed (e.g. de novo analysis, sequence homology searches) and validated (i.e. molecular weight or isoelectric point verification, immunoassays, or MS identification of synthetic peptides). To date, one of the primary limits for large-scale proteomic analyses is the lack of a marine genomic or proteomic database to search. In short, to finalize organism-level proteomic projects, there is a need for complete marine genomes. Several authors have addressed the complexity of this task because of the difficulty in isolating and culturing marine microbes (Beja 2004, Falkowski & de Vargas 2004, Hess 2004, Venter et al. 2004). Another obstacle that must be overcome before the completion of marine proteomics projects is the lack of facilities dedicated to large environmental protein discovery projects (not medical use). Typically only small projects are tackled as 'pet projects' by proteomic facilities and investigators, and often there is neither sufficient time nor instrumental resources to adequately develop techniques and identify marine proteins. This situation strongly implies the need to encourage funding for larger collaborative groups that include investigators not typically involved in the marine or oceanographic community.

In order to investigate some of the larger marine proteomics questions or to complement marine genomes with proteomes, funding for environmental research will need to increase. An efficient proteomics facility typically requires several qualified, full-time technical staff to work together as a team to complete full annotations. The technical support includes people trained in wet-laboratory chemical preparations, protein chromatography, methods development, and instrumental optimization and maintenance, in addition to IT staff. Unlike the genomics field, there are to date no large proteomics facilities partially dedicated to helping answer environmental questions (e.g. Joint Genome Institute, California, USA). Focusing the funding on a few environmental proteomic centers may alleviate this problem and allow marine investigators to continue to explore and collect ancillary data from all over the world's oceans. The available funding also plays a role in the skilled personnel available for completing marine proteomics-based projects. Many students, doctorates, or staff trained in proteomic mass spectrometry can be easily enticed to migrate to the life sciences divisions where funding is higher, jobs are more prevalent, and resources are seemingly unlimited relative to environmental research. In order for marine proteomics to flourish, trained personnel will need to be recruited, and an awareness of the importance of solving global environmental questions must become a priority for both government and community.

CONCLUSIONS

Moving beyond the analysis of elemental concentrations and amino acids is the next step toward advancing the science of marine organic chemistry. Because proteins are an intricate arrangement of 20 amino acids, each one can be specific to both a function and a source. Recent advancements in the field of biological mass spectrometry now provide an avenue through which to analyze the proteomics of different marine systems. A recent study by Powell et al. (2005) demonstrated how this high-throughput analysis allowed them to investigate the DOM pool without the need for specialized techniques that only identify expected targets (e.g. enzyme assays, antibody assays, fluorescent tags). As a discovery driven science, proteomics allows users to identify complete unknowns without missing

unanticipated interactions. This dramatically improves the range of applications within the marine field for which this technique can be employed. However, because the marine field consists of such diverse environments and matrices in which these proteins reside (e.g. phytoplankton, sediment, hydrothermal vents), a great amount of methods development remains to be completed.

After sufficient methods development and general cataloguing of marine proteomes has occurred, biogeochemists will better be able to model the evolution and cycling of carbon pools within the ocean. We can begin to survey how different marine organisms' proteomes adapt to dynamic nutrient conditions, and which proteins are expressed in the cell, released into the environment, and passed between trophic levels. This information will provide great insight into which proteins are preserved in the environment and whether chemical modifications play a role in their ultimate preservation. The culmination of numerous marine proteomic studies has the potential to allow global-carbon investigators to model how marine organisms will respond to future anthropogenic perturbations and release proteins into the environment for long-term preservation. Integrating these techniques into the marine field is the next logical step to advancing oceanic environmental research.

Acknowledgements. We thank P. Boyd, C. Nunn, B. Gallis, and reviewers for their comments and suggestions for improving this manuscript. B.L.N. thanks the Office of Polar Programs for her NSF postdoctoral fellowship. This work was supported by National Science Foundation grant numbers 0444148 and OCE-0453737.

LITERATURE CITED

Aluwihare LI, Repeta DJ, Pantoja S, Johnson CG (2005) Two chemically distinct pools of organic nitrogen accumulate in the ocean. Science 308:1007–1010

Armbrust EV, Berges JA, Bowler C, Green BR and 41 others (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306: 79–86

Beja O (2004) To BAC or not to BAC: marine ecogenomics. Curr Opin Biotechnol 15:187–190

Benner R, Pakulski JD, McCarthy M, Hedges JI, Hatcher PG (1992) Bulk chemical characteristics of dissolved organic matter in the ocean. Science 255:1561–1564

Bibby TS, Nield J, Barber J (2001) Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. Nature 412:743–745

Coin L, Durbin R (2004) Improved techniques for the identification of pseudogenes. Bioinformatics 20(Suppl 1): I94–I100

Cowie GL, Hedges JI (1984) Determination of neutral sugars in plankton, sediments, and wood by capillary gas chromatography of equilibrated isomeric mixtures. Anal Chem 56:497–504

Cowie GL, Hedges JI (1992) Sources and reactivities of amino acids in a coastal marine environment. Limnol Oceanogr 37:703–724

Dunkley TPJ, Dupree P, Watson RB, Lilley KS (2004) The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in *Arabidopsis thaliana*. Biochem Soc Trans 32: 520–523

Falkowski PG, de Vargas C (2004) Genomics and evolution. Shotgun sequencing in the sea: a blast from the past? Science 304:58–60

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Analytical instrumentation: electrospray ionization for mass spectrometry of large biomolecules. Science 246: 64–71

Giot L, Bader JS, Brouwer C, Chaudhuri A and 45 others (2003) A protein interaction map of *Drosophila melanogaster*. Science 302:1727–1736

Hecky RE, Mopper K, Kilham P, Degens ET (1973) The amino acid and sugar composition of diatom cell-walls. Mar Biol 19:323–331

Hedges J (1991) Lignin, cutin, amino acids and carbohydrate analyses of marine particulate organic matter. In: Hurd DC, Spencer DW (eds) Marine particles: analysis and characterization. American Geophysical Union, Washington, DC, p 129–137

Henrichs SM, Farrington J (1987) Early diagenesis of amino acids and organic matter in two coastal marine sediments. Geochim Cosmochim Acta 51:1–15

Hess WR (2004) Genome analysis of marine photosynthetic microbes and their global role. Curr Opin Biotechnol 15: 191–198

Hirosawa M, Hoshida M, Ishikawa M, Toya T (1993) Mascot–multiple alignment system for protein sequences based on 3-way dynamic-programming. Comp Appl Biosci 9: 161–167

Horiuchi T, Takano Y, Ishibashi J, Marumo K, Kobayashi K (2004) Amino acids in water samples from deep sea hydrothermal vents at Suiyo Seamount, Izu-Bonin Arc, Pacific Ocean. Org Geochem 35:1121–1128

Jaffe JD, Berg HC, Church GM (2004) Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics 4:59–77

Keil RG (1999) Early diagenesis of amino acids in high organic content marine sediments. Geochem Earth Surf 5:259–262

Keil RG, Kirchman DL (1993) Dissolved combined amino acids: chemical form and utilization by marine bacteria. Limnol Oceanogr 38:1256–1270

Keil RG, Montluçon DB, Prahl FG, Hedges JI (1994) Sorptive preservation of labile organic matter in marine sediments. Nature 370:549–551

King KJ (1974) Preserved amino acids from silicified protein in fossil Radiolaria. Nature 252:690–692

Kislinger T, Emili A (2003) Going global: protein expression profiling using shotgun mass spectrometry. Curr Opin Mol Theor 5:285–293

Kujawinskia EB, Vecchio RD, Blough NV, Klein GC, Marshall AG (2004) Probing molecular-level transformations of dissolved organic matter: insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Mar Chem 92:23–37

Lee C (1988) Amino acid chemistry and amine biogeochemistry in particulate material and sediments. In: Blackburn TH, Sorensen J (eds) Nitrogen cycling in coastal marine environments. SCOPE Series 33. John Wiley & Sons, p 125–141

Lee C, Cronin C (1982) The vertical flux of particulate nitrogen in the sea: decomposition of amino acids in the Peru upwelling area and the equitorial Pacific. J Mar Res 40: 227–251

Lee C, Olson BL (1984) Dissolved, exchangeable and bound aliphatic amines in marine sediments: initial results. Org Geochem 6:259–263

Li L, Zhao Z, Huang W, Peng P, Sheng G, Fua J (2004) Characterization of humic acids fractionated by ultrafiltration. Org Geochem 35:1025–1037

Lyons WB, Gaudette HE, Hewitt AD (1979) Dissolved organic matter in pore water of carbonate sediments from Bermuda. Geochim Cosmochim Acta 43:433–437

Mann M, Wilm M (1994) Error tolerant dentification of peptides in sequence databases by Peptide Sequence Tags. Anal Chem 66:4390–4399

McCarthy M, Pratum T, Hedges J, Benner R (1997) Chemical composition of dissolved organic nitrogen in the ocean. Nature 390:150–154

McCarthy MD, Hedges JI, Benner R (1998) Major bacterial contribution to marine dissolved organic nitrogen. Science 281:231–234

Minor EC, Wakeham SG, Lee C (2003) Changes in the molecular-level characteristics of sinking marine particles with water column depth. Geochim Cosmochim Acta 67: 4277–4288

Molloy MP, Donohoe S, Brzezinski EE, Kilby GW, Stevenson TI, Baker JD, Goodlett DR, Gage DA (2005) Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling. Proteomics 5: 1204–1208

Nguyen RT, Harvey HR (1994) A rapid micro-scale method for the extraction and analysis of protein in marine samples. Mar Chem 45:1–14

Nguyen RT, Harvey HR (1997) Protein and amino acid cycling during phytoplankton decomposition in oxic and anoxic waters. Org Geochem 27:115–128

Nilsson CL, Davidson P (2000) New separation tools for comprehensive studies of protein expression by mass spectrometry. Mass Spectr Rev 19:390–397

Nissenbaum A, Baedecker MJ, Kaplan IR (1971) Studies on dissolved organic matter from interstitial water of a reducing marine fjord. In: Gaentner HRV, Wehner H (eds) Advances in organic geochemistry. Pergamon Press, New York, p 427–440

Nunn BL (2004) Moving beyond amino acids: examinations of the protein component in marine sediments. PhD dissertation, University of Washington, Seattle

Nunn BL, Keil RG (2005) Size distribution and chemistry of proteins in Washington coast sediments. Biogeochemistry 75:177–200

Nunn BL, Norbeck A, Keil RG (2003) Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. Mar Chem 83:59–73

Palenik B, Brahamsha B, Larimer FW, Land M and 11 others (2003) The genome of a motile marine *Synechococcus*. Nature 424:1037–1042

Peers G, Price NM (2006) Copper-containing plastocyanin used for electron transport by an oceanic diatom. Nature 441:341–344

Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567

Powell MJ, Sutton JN, Del Castillo CE, Timperman AI (2005) Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. Mar Chem 95:183–198

Prahl FG (1985) Chemical evidence of differential particle dispersal in the southern Washington coastal environment. Geochim Cosmochim Acta 49:2533–2539

Reid GE, McLuckey SA (2002) 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectr 37: 663–675

Schweitzer B, Predki P, Snyder M (2003) Microarrays to characterize protein interactions on a whole-proteome scale. Proteomics 3:2190–2199

Siezen RJ, Mague TH (1978) Amino acids in suspended particulate matter from oceanic and coastal waters of the Pacific. Mar Chem 6:215–231

Squier AH, Harvey HR (2006) Applying proteomics to geochemistry: tracking individual proteins during marine diatom degradation using liquid chromatography-tandem mass spectrometry. EOS Trans: AGU Ocean Sci Meet Suppl 87, Abstr OS25L-01

Strzepek RF, Harrison PJ (2004) Photosynthetic architecture differs in coastal and oceanic diatoms. Nature 431: 689–692

Tanoue E (1996) Characterization of the particulate protein in Pacific surface waters. J Mar Res 54:967–990

Thomas X, Destoumieux-Garzon D, Peduzzi J, Afonso C and 7 others (2004) Siderophore peptide, a new type of post-translationally modified antibacterial peptide with potent activity. J Biol Chem 279:28233–28242

Venter JC, Remington K, Heidelberg JF, Halpern AL and 19 others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74

Wakeham SG, Lee C (1989) Organic geochemistry of particulate matter in the ocean: the role of particulates in oceanic sedimentary cycling. Org Geochem 14:83–96

Wakeham SG, Lee C, Hedges J, Hernes PJ, Peterson ML (1997) Molecular indicators of diagenetic status in marine organic matter. Geochim Cosmochim Acta 61:5363–5369

Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX and 6 others (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. Electrophoresis 16: 1090–1094

Wirth B, Louis VL, Potier S, Souciet JL, Despons L (2005) Paleogenomics or the search for remnant duplicated copies of the yeast DUP240 gene family in intergenic areas. Mol Biol Evol 22:1764–1771

Wisniewska J, Trejgell A, Tretyn A (2003) Plant signaling peptides. Acta Physiol Plant 25:105–122

Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C (2001) Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal Chem 73:2836–2842

Editorial responsibility: Howard Browman (Associate Editor-in-Chief), Storebø, Norway

Submitted: July 19, 2006; Accepted: November 8, 2006
Proofs received from author(s): February 20, 2007