ABSTRACT

Title of Document:	A PROTEOMICS APPROACH TO THE EXAMINATION OF PROTEINS IN MARINE SYSTEMS
	Jessica Felicia Faux, Doctor of Philosophy, 2014
Directed By:	Professor H. Rodger Harvey, Marine, Estuarine, and Environmental Sciences

The response of global carbon and nitrogen cycles to future climate change is uncertain. In order to understand the impacts that future changes to climate will have on these cycles, a more detailed understanding of them is essential. This dissertation utilizes a combined approach of molecular biomarkers and proteomic investigations to elucidate historic source material contributions and microbial protein production to contribute to a more thorough understanding of the marine carbon and nitrogen cycles.

The examination of molecular organic biomarkers throughout an Arctic sediment core showed the dominant input in the area was from marine sources with lower but steady contributions from terrestrial sources during the Holocene. Attempts to recover proteins from deeper sediments to correlate with lipid biomarkers were unsuccessful but led to the optimization of an extraction protocol for an added protein standard, bovine serum albumin, from sediments. An investigation into the expressed proteome of the heterotrophic marine bacterium, *Ruegeria pomeroyi*, under environmentally realistic carbon supply conditions during exponential and stationary growth phases identified over 2000 proteins. The most abundant proteins identified were responsible for porins, transport, binding, translation, and protein refolding and could represent potential

biomarkers of bacterial processes and/or activity. A parallel study of *R. pomeroyi*, in which ^{13}C-labeled leucine was added to the culture during exponential growth phase, showed labeled incorporation ranging from 16 to 21% of the total proteins produced depending on growth phase. The widespread distribution of the label among the growth phases indicates active recycling by the bacteria. This study demonstrates a method through which bacterial protein synthesis can be tracked. A study of the marine diatom *Thalassiosira pseudonana* acclimated to iron replete or iron-limited conditions showed iron-limited organisms increased proteins involved in pathways associated with intracellular protein recycling, the pentose phosphate pathway, lower photosynthetic energy production, enhancement of photorespiration, and increased polysaccharide production. This application of proteomics to the examination of proteins in marine sediments, a marine diatom, and a heterotrophic marine bacterium shows the potential for these techniques to help elucidate the fate of proteins in marine environments and could be used in conjunction with well-established molecular organic marker studies.

A PROTEOMICS APPROACH TO THE EXAMINATION OF PROTEINS IN MARINE SYSTEMS

By

Jessica Felicia Faux

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor H. Rodger Harvey, Chair
Professor Johan Schijf
Research Professor Lee Cooper
Professor Allen Place
Dr. Byron Crump
Professor Neil Blough, Dean's Representative

Acknowledgements

I want to give my utmost thanks to my research advisor, Dr. H. Rodger Harvey, for all his support and guidance throughout my graduate career at Chesapeake Biological Laboratory. I will be forever grateful for his belief in my ability to accomplish the goal of completing this dissertation. He always seemed to know what I could do even when I sincerely doubted myself. For that I am so very thankful. I also extend my greatest thanks to Dr. Brook Nunn without whose help I would not have been able to study proteomics. She too offered much support and guidance throughout my graduate career. I would like to sincerely thank the members of my doctoral committee, Drs. Johan Schijf, Lee Cooper, Byron Crump, Allen Place, and Neil Blough. I am very grateful for their participation and encouragement. I want to acknowledge Dr. Hafiz Ahmed for his contributions. I want to extend my thanks to Dr. Carys Mitchelmore and Hannah Pie for allowing me to use their lab space and for all their assistance with the cell culturing accomplished during this dissertation research. My thanks also go out to Dr. Thomas Miller for his encouragement and support. I am grateful to the Nutrient Analytical Services Laboratory for CHN analysis, to David Morris and Dr. Stephen Macko for isotopic analysis, to Dr. Leonid Polyak for Healy sediment samples, to Dr. Dennis Darby for advice on sedimentation rates, to Dr. Mary Ann Moran for her generous gift of a *Ruegeria pomeroyi* culture and to Christa Smith for all her help with culturing, and to National Science Foundation for funding this research.

I sincerely thank past and present members of the Harvey Lab. They have been extremely supportive and have helped in my training in so many ways. My warmest thanks go out to Laura Belicka who was instrumental in helping me start out on this

incredible journey and who has remained a dear friend. So many thanks go out to Karen Taylor, Rachel Pleuthner, Eli Moore, Tanya Muniak, and Molly Mikan. I am so lucky to call you all friends as well as colleagues.

I have been so lucky to have a support system of friends and family to help me along the way. I especially thank my parents, Gary and Patricia Gregory, for always encouraging my education and for helping me to be able to finish this dissertation. I thank my mother- and father-in-law, Darrell and Kathleen Faux, for helping when I needed to travel to finish my research. I thank my friends, the Robbins', the Muja's, and the Madden's for entertainment and distraction throughout the many years that this journey took. Lastly, I owe all of my success to the amazing encouragement and support of my husband, Timothy Faux, and my daughter, Aislinn Faux. Without Tim I wouldn't even have known about the program at CBL and without his constant love and his belief that I could do this I never would have been able to complete this journey. And to my daughter, I thank you for the smiles you give me every day that make this whole process worthwhile.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Research Overview

One of the most debated science policy topics today is climate change. Reviews of anthropogenic alterations of carbon dioxide (CO_2) levels and other greenhouse gases indicate that these changes have led to increasing global temperatures, which have the potential to lead to changes in climate (IPCC 2007). The global cycles of carbon (C) and nitrogen (N) are intrinsic to life on Earth; future changes to the global climate hold the potential to alter the functioning of these cycles. An in-depth understanding of them is necessary to predict future impacts of a changing climate on these cycles.

One of the most obvious regions where changes to biogeochemical cycles could manifest themselves is at polar latitudes. The polar region is predicted to see the most significant changes in temperature over the next 100 years (Macdonald et al., 2004) and Arctic environments have already experienced the impacts of warming temperatures. These increases in temperature are likely to cause complex alterations to many, if not all, components of the biogeochemical cycles in these areas, the impacts of which are not fully understood. For example, decreases in ice cover could lead to enhanced primary production, but loss of ice algal communities and replacement with pelagic communities along with alterations to upwelling, wind-driven vertical mixing, and freshwater inputs must also be considered when determining the extent of the alteration (Macdonald et al., 2004; Stabeno et al., 2012). Increased temperatures will likely lead to enhanced microbial recycling and increased supplies of organic material, potentially from permafrost, which have been sequestered from these cycles for thousands to millions of

years. How this potential increase in cycling and input of materials will affect C and N cycles or sequestration of organic matter is unknown.

While this region may demonstrate the earliest and most severe repercussions of alterations in the biogeochemical cycles, globally, all ecosystems are likely to be affected by climate change. Ecosystems may exhibit alterations similar to those in the polar regions, or they may experience other, more localized changes. Regardless of the ecosystem, a more comprehensive understanding of the functioning of biogeochemical cycles will facilitate the prediction of the impacts of future climate change scenarios.

This dissertation began with the broad objective of examining present and past C cycling as a means to detail what might be expected with future climate change in the Arctic. An investigation of the organic matter inputs to a sediment core in the Arctic Ocean was undertaken with the goal of elucidating how those inputs have changed over time and as a means of corroborating any alterations in organic matter deposition to past changes in climate. The expectation was that understanding what had occurred in the past would help clarify potential outcomes for future climate change in the Arctic.

The study of the Arctic sediment core utilized molecular biomarkers and elemental composition measurements as a means of determining the sources of organic matter inputs. A great amount of variability was not observed in the sediment record and most of the organic matter deposited was marine in origin. Lipid biomarkers and elemental composition measurements can provide information on the specificity of inputs of organic matter; however, they do not provide a great deal of information about the organisms producing the organic matter or the conditions under which it was produced. The ability to determine more conclusively what organisms were the sources of the

organic matter and what biological processes were occurring at the time of deposition would provide important information with which to make conclusions on past organic matter deposition and cycling.

The study of the proteins in this environmental system would provide greater detail on the sources of the contributed organic matter and the specific proteins produced may give insight into what was occurring at the time of deposition. The focus of this dissertation project transitioned to encompass this new line of investigation with the expectation that the identification of peptides and by extension, proteins in dated marine sediment sequences could result in the development of new protein biomarkers that could be correlated to known lipid biomarkers.

Under this second sub-study, I analyzed a sediment core to determine what fraction of proteins or peptides remained intact and identifiable in marine sediments over a sequence from surface to deeper, older sediment horizons. Based upon a previously published analysis of surface sediments from the Bering Sea (Moore, Nunn, Goodlett, et al., 2012), I used their extraction protocol to isolate any proteins present in the sediment core. However, the initial results from those extractions showed no identifiable proteins. While identification of preserved proteins was ultimately unsuccessful, these efforts led to an optimized, new extraction protocol that can be applied in future work.

The recovery of a bovine serum albumin (BSA) standard added to sediments indicated that the extraction protocols were functional, but the question remained as to why there were no proteins identifiable in the sediment. Amino acids present in the sediments that were used in the optimization experiments demonstrate the presence of fractions of proteinaceous material. The uncertainties are as follows: Do these particular

sediments lack identifiable peptides because they originated from organisms for which we don't have sequenced genomes and identified proteomes? Or is it possible that in sediments that are older and deeper than surficial horizons that the interactions between the proteinaceous material and any potential matrices are too strong to allow for the removal of intact peptides via non-hydrolyzing extraction methods? Or has the organic material been recycled or degraded to the point that any intact peptides are no longer distinctive enough to be conclusively assigned to parent proteins? These uncertainties dictated a change in strategy away from pursuing my original goal of investigating the presence of proteins in older, deeper sediments.

Since proteins are essential biochemical components to all organisms and contain both nitrogen and carbon, they are readily recycled in marine systems by consumers. Bacteria play a major role in that recycling and in the resynthesis of proteins (Cole et al., 1988). Elucidation of bacterial protein production and recycling can enable evaluation of the role of bacteria in diminishing capabilities to identify proteins in older sediments, using currently available genomic tools. In addition, more detailed knowledge of the materials bacteria are utilizing and recycling will help enhance our understanding of marine C and N cycles.

In order to try to understand the lack of identifiable peptides or proteins in the marine sediments studied, I chose to begin with an initial examination of bacterial protein production and recycling. The focus of my research shifted to the following goals: (i) to examine the expressed proteome of a heterotrophic marine bacterium, *Ruegeria pomeroyi*, with environmentally realistic carbon substrates, throughout the growth curve and to evaluate any observed patterns in protein expression; (ii) to determine whether the

incorporation of a tracer level addition of an isotopically labeled amino acid to *R. pomeroyi* can be tracked in the proteins it produces; (iii) to examine the expressed proteomes of two *Thalassiosira pseudonana* cultures when exposed to low iron (Fe) or Fe replete conditions and what metabolic pathways were affected by the two Fe conditions.

1.2. Scientific Background

1.2.1. Proteins in marine environments

Protein accounts for the majority of the nitrogen present in organisms and also contributes a substantial amount of C with estimates of contributions to C as high as 67% in plankton (Wakeham et al., 1997). Proteins serve multiple functions in organisms including facilitating enzymatic processes, providing structural support and storage, protection from disease, and responses to stimuli to name but a few. Because of their necessity to so many cellular functions and the limited presence of biologically available nitrogen in many marine systems, proteins are typically rapidly recycled upon the death of an organism and their amino acids either used as energy or resynthesized.

Although intact proteins themselves are thought to be rapidly recycled (e.g. Nunn et al., 2010), amino acids are found throughout marine systems, and are important components of C and N pools (Keil et al., 2000 and references therein). The traditional view is that due to their ease of depolymerization, proteins are more readily recycled and thus unlikely to remain intact in marine systems long enough to be an important contributor in refractory material such as sediments. Yet, the ubiquity of amino acid contributions to sediments and the presence of high molecular weight preserved fractions

(Nguyen and Harvey, 2003) indicate a larger role of protein or its products as contributors to the undefined fraction of C and N seen in the geological record.

The identification of the proteinaceous material present in the unidentified fraction of sequestered organic material and/or the elucidation of the extent of protein recycling, will contribute to an improved understanding of C and N cycles. The persistence of particular proteins in the environment can provide insight into the mechanisms that control organic matter degradation (Nunn and Timperman, 2007) while, conversely, identifying and quantifying bacterial uptake of peptides and amino acids could provide information on exactly how much protein within marine systems is recycled. Both of these lines of investigation would provide better constraints on our understanding of the C and N cycles.

1.2.2. Protein Analysis - Proteomics

Early methods of examining protein cycling relied on amino acid analysis (see review by Nunn and Timperman, 2007). Amino acids represent a substantial portion of organic matter in sediments and can provide information on depositional conditions and the extent of the decay of the proteinaceous material (Keil et al., 2000 and references therein). The traditional technique to measure amino acids utilizes acid hydrolysis to destroy the amide bonds that link individual amino acids in the primary protein structure, which negates any possible inferences about the sequence, structure or function of the source protein (Nunn and Timperman, 2007). Total hydrolysable amino acids (THAAs) do not provide clues as to the source of proteins because the bulk amino acid compositions of many organisms are similar and, as such, do not allow for the distinction

of source contribution (Keil et al., 2000 and references therein). The proteinaceous material in environmental samples, whether from live organic communities or from preserved inorganic material, needs to be identified and studied to better understand the extent of nitrogen cycling in marine systems. Any live organisms will contain proteins, but McCarthy et al. (1997) and Zang et al. (2000) also demonstrated the presence of amide bonds in environmental samples indicating that amino acids can remain linked during preservation. These findings indicate that intact proteins and peptides are preserved and illustrate the need for a method to identify them in order to begin elucidating the mechanisms of preservation. Without knowing which proteins are preserved, we can't begin to determine how they were preserved.

Analysis and identification of any of these proteins or peptides, living or preserved, requires that they retain their primary structure, the linear sequence of their amino acids. The primary sequence of the protein determines how it is folded and its three-dimensional conformation that, in turn, determines its biological role within an organism. Traditional amino acid hydrolysis techniques are not appropriate for the identification of peptides or proteins as they destroy the primary structure. Before the identification of individual proteins can be attempted, proteinaceous material must first be efficiently extracted from samples. A balance must be struck in which the extraction method is rigorous enough to extract proteins, yet hydrolysis is limited to retain primary sequence information. Sediments in particular, present a complicated matrix of both intact and degrading material from which to extract proteins.

Once the proteinaceous material has been extracted, identification of the peptides or proteins within that extract can proceed. A protein can be identified by determining its

primary sequence. Older protein identification techniques provided primary sequences on a protein-by-protein basis that is not compatible with the desired analysis of all proteins in a sample (Nunn and Timperman, 2007) as it would take a great amount of time and resources to determine each individual protein. Advances in tandem mass spectrometry (MS/MS) now enable its utilization to measure spectra allowing for the direct determination of the amino acid sequence of the peptide or protein (Armengaud, 2013; Hunt et al., 1986). MS/MS can be used to determine the amino acid sequence of a protein directly from the measured spectra (*de novo* sequencing) or proteins can be identified by matching measured spectra to known protein or genomic sequences through database correlation (Hettich et al., 2013; Powell et al., 2005). These advances have given rise to the field of proteomics, the study of proteins, particularly their structure and function, from mixtures in complex systems. Its application to environmental samples is allowing researchers to sequence and identify large numbers of proteins.

A standard protocol for MS/MS-based protein identifications involves the measurement of spectra of small peptides, obtained through the cleavage of proteins, and is known as the 'bottom-up approach' or shotgun proteomics (Fig. 1.1) (Hettich et al., 2013). Briefly, the process begins with the extraction of peptides and proteins from a sample, followed by their enzymatic digestion, typically through a trypsin digest. The resultant peptides are separated using high-performance liquid chromatography (HPLC) and subsequently passed through a mass spectrometer (MS) where parent ion mass-to-charge (*m/z*) ratios are determined. Individual parent ions are further fragmented by the MS, and the resultant *m/z* ratio values produce the tandem mass spectrum (referred to as spectra, measured spectra, daughter ion scans, or MS2 scans). The measured spectra are

then compared to theoretical spectra that are predicted from protein or genomic databases. If the spectra are complete enough, the peptide sequence can be mathematically deduced directly from measured spectra themselves (*de novo* sequencing). The presence of more than one peptide unique to the same protein is used to infer the presence of the parent protein in the sample (Armengaud, 2013; Nunn and Timperman, 2007). Inferring the presence of the intact protein from the detection of its component peptides may be problematic when dealing with protein identifications from refractory or preserved materials. It is possible, however, to determine the observed fraction of a protein's amino acid sequence (i.e., its sequence coverage), which can provide a measure of confidence in the conclusion that the intact protein was present in the sample.

While shotgun proteomics offers many benefits over previous techniques, there are several complexities that must be taken into consideration and accounted for within any analysis. The first issue deals with the relative abundance of proteins within a sample. The more abundant proteins within a sample will likely yield more peptides. Because peptides are isolated for fragmentation and sequence determination based on the intensities of their peaks, the more abundant peptides will likely be chosen and may mask those peptides of lower abundance (Armengaud, 2013). In addition, since the measured spectra are computationally linked to a protein, redundant, homologous, or isobaric (peptides with the same molecular weight) peptides may be assigned to multiple proteins causing complications within an analysis (Hettich et al., 2013). In order to account for these potential misidentifications of proteins, an evaluation of false-positives can be

calculated using a database in which the same peptide sequences are read in reverse (Armengaud, 2013).

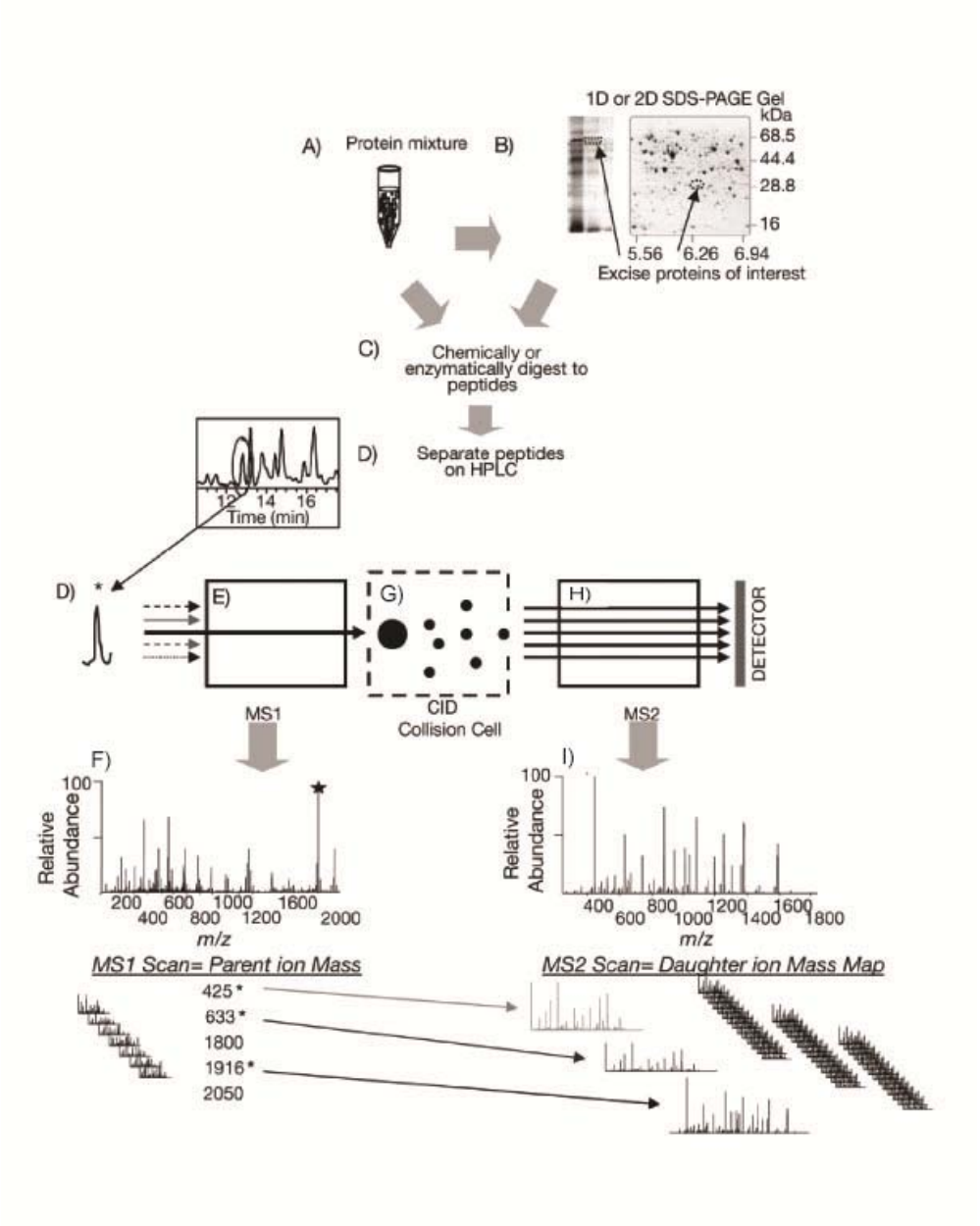


Figure 1.1. Diagram of a bottom-up approach of the identification of peptides using tandem mass spectrometry. Proteins in a mixture or isolated from gel electrophoresis (A and B) are enzymatically digested into peptides (C). Peptides are then separated by HPLC (D). Peptides elute from the HPLC and are ionized and analyzed by the first mass spectrometer (E). Mass to charge ratios are measured and the parent-ion scan (MS1 scan) is produced (F). In data dependent analysis single peptides are isolated based on the intensity of their peaks (i.e. the three most intense peaks (F)) from the parent-ion scan for fragmentation (G) and sequence determination. Each selected ion from the parent-ion scan is then individually fragmented and sent to the second mass spectrometer (H) to be analyzed which produces the daughter-ion scans (MS2 scans) (I). The daughter-ion scans are then correlated to theoretical spectra from a provided database. CID: collision induced dissociation. Adapted from Nunn and Timperman, 2007.

1.2.3. Application of Proteomics to Marine Samples

The application of proteomics to environmental samples has allowed the examination of individual eukaryotic and prokaryotic microbial marine organisms, communities of organisms, and proteinaceous material in the water column and sediments. Proteomics has enabled the documentation of individual organisms from diverse environments (Armengaud, 2013) and has include investigations into the expressed proteomes of organisms, the presence of specific proteins, imposing stress conditions and tracking alterations in the proteome, and examining proteins in the exoproteome, the proteins excreted from the cell (e.g. Christie-Oleza and Armengaud, 2010; Christie-Oleza, Fernandez, et al., 2012; Christie-Oleza, Pina-Villalonga, et al., 2012; Green et al., 2013; Nunn et al., 2013; Nunn et al., 2009). While the study of individual species provides valuable information, it would also be desirable to study the interactions and metabolic exchanges among organisms of a community.

Many marine organisms aren't amenable to culturing, however, and cannot be studied as individuals in a laboratory setting (Schweder et al., 2008; Stewart, 2012). The application of metaproteomics, which is the study of all the proteins obtained directly from environmental samples, is helping overcome these culturing limitations and providing insight into community structures (Hettich et al., 2013; Morris et al., 2010; Schweder et al., 2008; Williams et al., 2012). These approaches are rapidly advancing our knowledge of marine organisms and their communities. While proteomics and metaproteomics enable the identification of proteins in organisms and systems that could not be studied before, there remain limitations to these techniques. There is still the need for effective extraction of proteins from samples or cell fractions depending on the target

proteome, but the largest limitation in proteomics and more specifically, metaproteomics, is the lack of genomic sequence information on the majority of marine microbial communities (Chao and Hansmeier, 2012; Hettich et al., 2013; Schweder et al., 2008). This particular limitation may not be an obstacle for very much longer, however. The decreasing costs and more wide-spread availability of DNA sequencing is likely to allow the determination of increasing numbers of genome sequences for marine organisms. In addition, experimental proteomic data is now being used to revise and improve gene annotation of existing genomes (Armengaud, 2013; Christie-Oleza, Miotello, et al., 2012). This can be expected to lead to more, greatly improved, genome sequences with which to identify protein expression in complex marine microbial communities.

Studies have also used proteomics to evaluate degrading or preserved proteins in marine systems. Powell et al. (2005) determined the presence of bacterial membrane/envelope proteins as well as what they describe as the first detection of a dissolved protein that is not a bacterial membrane protein, specifically ribulose bisphosphate carboxylase. Nunn at al. (2010) followed diatom proteins through 23 days of degradation and observed that remaining proteins were those associated with membranes, had glycan modifications, or could aggregate into a supermolecule. Moore, Nunn, Goodlett, et al. (2012) identified progressively decreasing numbers of proteins in samples collected at the chlorophyll maximum, to sinking particles collected in sediment traps, to surface sediments, and to suspended particles.

One notable result of these studies compared to those of individual microbial organisms or a microbial community is the number of identified proteins. There are far fewer proteins found in these studies than those of individual organisms or communities

which can return hundreds to one or two thousand confidently identified proteins. As was discussed previously, THAAs can represent substantial amounts of the organic nitrogen fraction in sediments (Burdige and Martens, 1988; Keil et al., 2000 and references therein) but the number of identifiable proteins in preserved or degrading samples can vary greatly. Powell et al. (2005) found between 42 to 79 peptides in seawater samples depending on the method used. Moore, Nunn, Goodlett, et al. (2012) identified more than 200 proteins in surface waters with decreased numbers of proteins identified in the deeper water column and surface sediments. Nunn et al. (2010) identified 121 proteins at the start of a laboratory degradation experiment followed by a precipitous drop to 4 proteins at the end of the degradation.

The differences seen in these studies could be attributed to the degree of degradation and the ability to extract proteinaceous material from the sample. Moore, Nunn, Goodlett, et al. (2012) propose that the higher number of identifications seen in the sediment trap samples is explained by the presence of fresher material that has aggregated and is falling through the water column while suspended particles, having remained longer in the water column, have been exposed to degradation for a greater period of time, resulting in a smaller number of identified proteins. The substantially higher number identified at the chlorophyll maximum is to be expected as a consequence of increased productivity (and recent cell growth) at that depth. During degradation, microbial activity alters the structure of proteins through hydrolysis to smaller peptides and potentially adds modifications (e.g., glycosylation, methylation, etc.). Since the identification of a protein is the result of a comparison of measured spectra to hypothetical spectra generated from the genomic information, lack of knowledge of these

modifications makes protein identifications extremely difficult. In addition, if the protein has been truncated to peptides too small to be uniquely identified to individual proteins, no protein identification will be possible. These are additional obstacles that need to be overcome when identifying proteins in certain marine samples. More study of individual organisms and their communities, their roles in nutrient cycling and the nutrient cycling pathways will provide information on what state the resulting, preserved proteins are in.

1.3. Research Objectives and Approach

A complete understanding of the C and N cycles is necessary in order to anticipate alterations to these cycles brought about by climate changes that affect biogeochemical cycles. Proteins are an important component in all biological organisms and are part of the C and N cycles, but we do not have complete knowledge of exactly what proteins are synthesized, how proteins are recycled, and the ultimate fate of proteins in marine systems. This dissertation takes two approaches to address some of these deficiencies by investigating the presence of proteins in marine sediments and evaluating which proteins are synthesized by marine microbes.

Chapter 2 details the effort to identify and quantify organic matter inputs to a sediment core collected from the Northern Chukchi Sea over the last 8800 years. The goal of this investigation was to identify the sources of organic matter and follow the changes in deposition over the Holocene period to track environmental and climatic changes occurring in this area over time that were recorded in the sediments of the core. Identifying and understanding past historical changes may provide insights in

determining how certain regions will change or adapt to future climate changes. The following research question was the primary goal:

1. Can past climatic or environmental changes be determined by identifying changes to organic matter source inputs throughout a sediment core?

This question was investigated with a multi-proxy approach that included molecular organic markers (fatty acid, alkane, alcohol and sterol analysis), a novel proxy, the Branched and Isoprenoid Tetraethers (BIT) index (a Glycerol Dialkyl Glycerol Tetraether lipid analysis), and bulk C and stable isotope measurements. These measures were used to examine sediments from 18 sampling horizons within the core. I was responsible for conducting all sediment analyses and data analysis and compiled these into the resulting published work.

Chapter 3 details a methodology for the optimization of protein extractions from marine sediments and assesses the effectiveness of proteomic databases in the identification of proteins in marine samples. As was discussed previously, proteomics offers the enhanced ability to study proteins in environmental samples, but its efficacy is still limited to the quality and quantity of protein extractable from marine samples, as well as to the quality of the database against which measured spectra are searched. With these limitations in mind, the following research questions were investigated:

1. Which extraction method, when applied to marine sediments, results in the greatest recovery of protein?
2. Which database, when used to search protein extracts from marine sediments, results in the highest level of confidence in protein identifications?

This chapter details my efforts as well as those of another graduate student, which resulted in a group-produced publication. I was responsible for conducting, analyzing and documenting the evaluation of optimized extraction methods for proteins from marine sediments. The experimental work was conducted on sediments from the same horizon (20-22 cm depth) in a core taken from the Bering Sea. The analysis of surface sediments and the subsequent evaluation of the effectiveness of protein identifications through various databases were conducted by E.K. Moore.

Chapter 4 details studies on the expressed proteome of *Ruegeria pomeroyi*. Our objective was to track the total proteins produced and identified through various growth phases and to determine any observed patterns in protein expression. Specifically, the following research question was investigated:

1. What proteins are marine bacteria actively synthesizing and are there any changes in protein expression during different growth phases?

To address this question, *R. pomeroyi* was cultured under environmentally realistic C conditions and samples were collected at four time points throughout its growth phases representing early and late exponential phase and early and late stationary phase.

Chapter 5 details the investigation of the incorporation of a tracer level addition of labeled amino acid to *R. pomeroyi*. The ability to track the specific production of proteins provides an additional tool for understanding the extent to which bacterial recycling or remineralization affects marine organic matter. The goal was to determine whether the incorporation of a labeled amino acid, ^{13}C-leucine, to a heterotrophic marine

bacterium could be tracked in the proteins that the bacteria synthesized. Specifically, the following research question was investigated:

1. Can the production of proteins incorporating an isotopically labeled amino acid be tracked in the expressed proteome of *R. pomeroyi*?

To address this question *R. pomeroyi* was cultured under environmentally realistic C conditions with the addition of ^{13}C-leucine at two time points. Samples were collected at four time points throughout the bacterium's growth phases representing early and late exponential phase and early and late stationary phase.

Chapter 6 details the study of the expressed proteomes of *Thalassiosira pseudonana* when acclimated to low iron (Fe) or Fe replete conditions and what metabolic pathways were affected by the two Fe conditions. Understanding the physiological strategies employed by marine microbial organisms, and the proteins expressed because of these strategies, will further enhance our knowledge of the role of these organisms in global biogeochemical cycles. Specifically, the following question was investigated:

1. Can alterations to the expressed proteome be determined when a marine diatom is acclimated to low Fe or Fe replete conditions?

T. pseudonana was cultured in artificial seawater and was acclimated for 10 generations to low Fe or Fe replete conditions. It was then transferred to a new culture and sampled at a mid-exponential growth phase. This chapter details my efforts as well as that of our collaborator, Dr. Brook Nunn. I was responsible for the evaluation of the proteins that were significantly up- or down-regulated.

1.4. Research Significance

Potentially changing climate scenarios dictate that global biogeochemical cycles are very likely to be affected. Our understanding of the role of proteins within these cycles is still limited, which limits predictive capabilities. This dissertation research was designed to begin to elucidate details of the cycling of proteins through marine organic matter. There is a need to establish the extent of the bacterial contribution to the production of proteins as an initial means to determine their role in the recycling of marine organic matter. Specifically, this dissertation provides initial information on the protein production of a cosmopolitan, heterotrophic, marine bacterium. These finding will provide the initial steps needed for further investigations tracking bacterial production and recycling of proteins in marine systems.

Chapter 2: Organic Sources and Carbon Sequestration in Holocene Shelf Sediments from the Western Arctic Ocean

A major portion of this chapter has been published as "Organic sources and carbon sequestration in Holocene shelf sediments from the western Arctic Ocean". Faux, J.F, Belicka, L.L., Harvey, H.R. 2011. Continental Shelf Research, 31, 1169-1179.

2.1. Abstract

To quantify changes in organic carbon inputs and preservation, sediments from the western Arctic Ocean, specifically the Northern Chukchi Sea, spanning the last 9000 years of the Holocene period were collected during the Healy-Oden Trans-Arctic Expedition (HOTRAX) expedition and analyzed. The multi-proxy approach included molecular organic markers, bulk carbon and isotope measurements plus more recent approaches to terrestrial carbon estimation (the Branched and Isoprenoid Tetraethers (BIT) index). The upper 1100 cm of the core, corresponding to the last 7.4 ka, showed a relatively stable total organic carbon content of 1.13% to 1.38% which decreased below 1100 cm to 0.6%. C:N ratios ranged from 8.4 to 10.83 over the Holocene time period examined. The distribution of *n*-alcohols and *n*-alkanes revealed major contributions from long-chain *n*-alcohols and *n*-alkanes while minimal contributions were seen from short-chain *n*-alkanes. The majority of the total fatty acids was comprised of saturated and monounsaturated fatty acids with short-chain and long-chain saturated fatty acids present in similar concentrations throughout most of the core and monounsaturated fatty acids decreasing down-core. Total sterol concentrations showed considerable inputs from marine sterols, $C_{28}\Delta^{5,22}$, $C_{28}\Delta^{5,24(28)}$ and dinosterol, as well as $C_{29}\Delta^{5}$, typically considered a terrestrial marker. The BIT indices for core sediments ranged from 0.021 to 0.216 with minor changes seen in older sequences. Overall, organic biomarkers indicate marine sources as the more dominant input of organic matter with lower but continual

contributions from terrestrial sources at this location during the Holocene. The remarkable consistency among multiple molecular organic markers of both marine and terrestrial origin over the Holocene period encompassed by the core suggests that sinking material and surface sediments were heavily influenced by bottom currents and mixing processes prior to their deposition.

2.2. Introduction

Anthropogenic alterations of atmospheric carbon dioxide (CO_2) and other greenhouse gases have the potential to lead to changes in climate with projections that the high latitudes will experience the most significant alterations in temperature (IPCC, 2007). Some of these changes are already being observed including decreases in the amount of sea ice, changes to algal distributions (Li et al., 2009), increased coastal erosion, and the melting of permafrost (Macdonald et al., 2004). These changes have the potential to alter the balance of the marine and terrigenous components of the arctic organic carbon cycle and, in turn, feedbacks to the global carbon cycle through changes in carbon sequestration. Historical changes in carbon source delivery as recorded in ocean sediments can provide an understanding of the linkages between climate and carbon biogeochemical cycling and sequestration. Bulk organic matter properties such as total organic carbon (TOC), hydrogen index (HI), and carbon and nitrogen isotopic composition (Knies and Stein, 1998; Schubert et al., 2001; Yamamoto et al., 2008) have been used to reconstruct processes affecting organic carbon deposition in various parts of the Arctic. Molecular organic biomarkers, including *n*-alkanes, alkanoic acids and sterols, have been used successfully as proxies for the determination of organic matter

sources in the Arctic (Belicka et al., 2004; Birgel, et al., 2004; Fahl and Stein, 1997). Such organic markers can provide the specificity of inputs, whether marine or terrestrial, the impact of water column processes on their preservation (e.g. Conte et al., 2003), and over sediment sequences the depositional changes over long time scales (Stein and Macdonald, 2004).

This study examined sediments from the Northern Chukchi Sea to identify and quantify organic matter inputs over the last 8800 years. The multi-proxy approach included molecular organic markers as well as bulk carbon and stable isotope measurements to identify the sources of organic matter and follow the changes in deposition over the Holocene period along the Western Arctic Shelf.

2.3. Materials and Methods

2.3.1. Study area, sampling and core dating

During the HOTRAX expedition of 2005 (Darby, Polyak, et al., 2009) a jumbo piston core (JPC), designated HLY0501-JPC5, was collected from the northern Chukchi Sea (72.6942 N, 157.5177 W; Fig. 2.1). This core, hereafter referred to as JPC5, was approximately 16.7 m in length and was collected in 415 m of water. A parallel trigger core (TC) was used to reconstruct the uppermost sediment section lost during the piston coring. Subsamples were made available by Leonid Polyak at Byrd Polar Research Institute and were stored in a -70°C freezer prior to analysis. Samples discussed here represent subsections approximately every 100 cm with the deepest analyzed sample at a corrected depth of 1315 -1317 cm. In order to determine depositional age, an age model for the TC and JPC was utilized from Darby, Ortiz, et al. (2009), who conducted

radiocarbon dating on various bivalves from six intervals (unadjusted depths: 37 to 880.5 cm; adjusted depths: 112 to 955.5) in JPC5. Based on the dates obtained the following age model was constructed, which was used as the basis for age calculations in the present study:

Average age (ka (kilo-annum)) = 0.0067*Depth

In a related study by McKay et al. (2008) in the same location, slightly differing reservoir ages were proposed. This leads to an alternative age structure for the core and by extension, differing offsets for the TC and JPC. These constraints are discussed in the results.

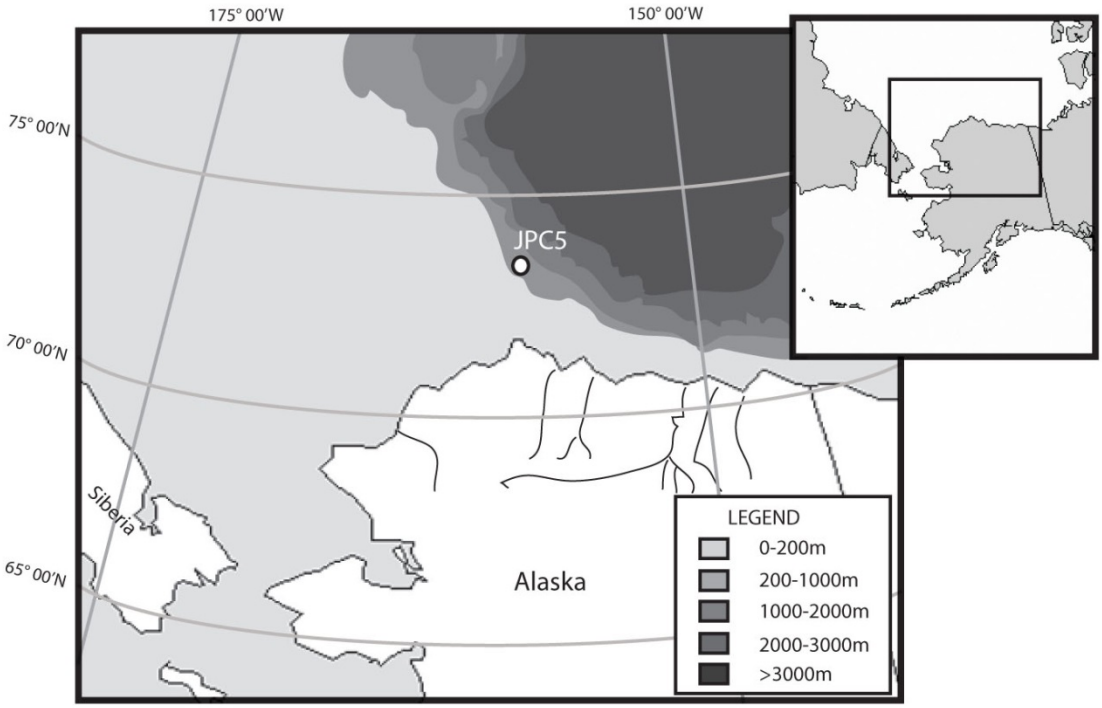


Figure 2.1. Map of the western Arctic Ocean showing the sampling station for sediment core HLY0501-JPC5 collected as part of the HOTRAX expedition in the Northern Chukchi Sea.

2.3.2. Sediment preparation and analysis

Each sample was transferred to a glass I-CHEM jar, covered with aluminum foil and lyophilized. Samples were then homogenized with a mortar and pestle prior to extraction.

Sediment organic carbon and nitrogen content was determined with an elemental analyzer (Exeter) and stable isotopic composition was measured with an Optima stable isotope ratio mass spectrometer using standard methods (Analysis conducted at University of Virginia). Internal laboratory reference gases for carbon and nitrogen were calibrated against the respective international standards, NBS-19 and atmospheric N_2. Isotopic results are reported in delta notation (δ) as per mil deviations (‰) from the corresponding international standards of Vienna Pee Dee Belemnite (V-PDB) and atmospheric N_2 (air). Analytical precision for carbon and nitrogen was within $\pm$ 0.2‰.

For lipid analysis, approximately 3 g of dry sediment was extracted with a mixture of CH_2Cl_2:MeOH (1:1 v/v) in a solvent-rinsed test tube as described in Belicka and Harvey (2009) with minor modification. Briefly, sediments were extracted three times with sonication followed by centrifugation to pellet sediment solids. To each sample and blank, 5α-cholestane and nonadecanoic acid (C19:0) were added as internal standards. For the third and final extraction, equal amounts of CH_2Cl_2:MeOH (1:1 v/v) and ultrapure water were added to the sample along with approximately 2 to 5 mL of CH_2Cl_2. The CH_2Cl_2 layer was removed and added to the other extracts and the solvent was removed by rotary evaporation.

Total extracts were subject to alkaline hydrolysis to release ester-bound lipids and isolate neutral and polar lipid fractions. Methanolic KOH (0.5N) and ultrapure water

were added (7:1 v/v) and the samples were heated at 70°C for 30 min, following which, an additional 1 mL of ultrapure water was added and gently mixed. Neutral lipids were subsequently extracted three times with a mixture of hexane/diethyl ether (9:1 v/v). The remaining fraction of the extract was acidified by the addition of concentrated HCl to a final pH < 2. The polar lipids were then extracted three times with the hexane/diethyl ether (9:1 v/v) mixture. The neutral fraction was derivatized with BSTFA (N,O-bis(trimethyl-silyl)trifluoroacetamide) amended with 25% pyridine at 50°C for 15 min. The polar fraction was methylated with boron trifluoride in methanol (BF_3:MeOH) at 70°C for 30 min and further purified by partitioning into hexane/diethyl ether (9:1 v/v) three times.

Individual compounds in both fractions were quantified by capillary gas chromatography (GC) with an Agilent 6890 GC System using a J&W DB-5MS column (60 m length x 320 μm i.d. x 0.25 μm film) and flame ionization detector. Instrument methods were identical to those described in Belicka et al. (2002) with the exception of an initial injector temperature of 250°C. Lipids were identified by mass spectrometry using an Agilent 6890 Series GC System integrated to a 5973 Network Mass Selective Detector (MSD) with helium as the carrier gas at a flow rate of 1.4 mL min^{-1}. The MSD was operated at 280°C in electron ionization mode with a mass range of 50-650 amu.

2.3.3. Glycerol dialkyl glycerol tetraether lipids (GDGTs) sample preparation and analysis

Total lipid extracts for the majority of GDGT samples were prepared as described above with a subset of samples (75-77, 115-117, 185-187, 275-277, and 395-397cm)

extracted using a Dionex ASE 300 Accelerated Solvent Extractor. The extraction was completed through three cycles with each cycle at 100°C and 1500 psi for 15 min and flushing with 16.5 mL CH_2Cl_2:MeOH (3:1 v/v).

GDGTs were analyzed using the method of Hopmans et al. (2004) with minor modifications. Briefly, total lipid extracts were partitioned into apolar and polar fractions with alumina column chromatography using hexane:CH_2Cl_2 (9:1 v/v) and CH_2Cl_2:MeOH (1:1 v/v), respectively. Polar lipids were then evaporated, redissolved in hexane:propanol (99:1 v/v) and filtered through a 0.2 μm pore size, 13 mm diameter Teflon filter in a stainless steel filter holder with a luer-lock syringe. GDGTs were analyzed using an Agilent 1100 Series High Performance Liquid Chromatography system coupled to an Agilent 1100 Series SL Ion Trap Mass Spectrometer with an atmospheric pressure chemical ionization interface. Separation was achieved on a Grace Prevail Cyano column (2.1 x 150 mm; 3μm particle size). The sample run time was 50 min with a constant flow rate of 0.2 mL min^{-1}. The run began with an isocratic hold at 99% hexane 1% propanol for 5 min followed by a linear gradient to 1.8% propanol over 45 min. A reverse gradient brought the propanol back to 1% over 5 min followed by a 10 min flush of 99:1 hexane:propanol.

BIT indices were calculated using areas of the $[M + H]^+$ peaks as described in Hopmans et al. (2004). Briefly, the relative abundances of the branched GDGTs (I + II + III) and crenarchaeol (IV) (structures are given in Hopmans et al., 2004) were compared as follows: BIT = (I + II + III) / (I + II + III + IV).

2.4. Results

2.4.1. Corrected core depths and core dating

An estimated offset of the trigger (TC) and piston core was calculated to be 75 cm by Darby, Ortiz, et al. (2009) based on radiocarbon dating (AMS) of various bivalves. This offset is used to align sediment horizons and all depths referenced refer to corrected depths unless otherwise stated. An alternative alignment was presented by McKay et al. (2008) based on differing estimations of corrected ages for the core. In both studies, radiocarbon ages were corrected for marine reservoir effects and stable carbon isotope fluctuations using CALIB5.0.2 (Darby, Ortiz, et al., 2009; McKay et al., 2008; Stuiver et al., 2005); however, Darby, Ortiz, et al. (2009) used a ΔR value, the regional difference compared with the average global reservoir correction, of 460 while McKay et al. (2008) used a ΔR value of 0. This produced somewhat differing results of corrected sediment dates. The use of the higher ΔR by Darby, Ortiz, et al. (2009) to account for older Pacific water that influences this system is supported by mollusk studies from coastal environments, but due to the fact that the core was taken from deeper waters at the edge of the slope they acknowledge that there is more uncertainty in samples farther offshore. Atlantic water exposure is proposed by McKay et al. (2008) as justification for use of a ΔR of 0. Yet work by Weingartner et al. (2005) and Pickart et al. (2005) note that Pacific waters entering through the Bering Strait make their way through various pathways across the Chukchi Sea and into the Arctic Ocean thereby contributing to primary production in the Chukchi and potentially altering the ^{14}C signature. Pickart et al. (2005) found Pacific waters traveling along the shelf break near the area where JPC5 was collected and state that a significant fraction of Pacific waters make their way to the Fram

Strait. In addition, they describe eddies that form in the region near where JPC5 was collected and suggest that these eddies transport water offshore. Because of the strong presence of Pacific water in the Western Chukchi and the potential for eddy formation and transport of water offshore, the use of the ΔR value of 460 from Darby, Ortiz, et al. (2009) was chosen for the age determinations.

Applying the age model of Darby, Ortiz, et al. (2009) to the sampled horizons of the TC and JPC5 returned an age for TC surface sediments of 0.0067 ka. The top of the piston core (75-77cm) was calculated to be 0.5092 ka and the deepest horizon analyzed (1315-1317cm) at 8.82 ka. Four TC horizons (0-2, 40-42, 80-82, 120-122) were combined with 14 JPC horizons to construct a consolidated core that, for ease of discussion, will be referred to as JPC5.

2.4.2. Bulk sedimentary organic matter

Total organic carbon over the depth analyzed ranged from 0.6% to 1.38% (Table 2.1; Fig. 2.2). Organic carbon remained fairly constant in the top 1100 cm of the core, varying from 1.13% to 1.38%, and decreased to 0.6% below 1100 cm. Sedimentary organic nitrogen varied little, ranging from 0.19% in surficial sediments to 0.07% at 1315 cm (Fig. 2.2). C:N ratios showed subtle changes over time, ranging from 8.4 to 10.8 (Fig. 2.2). The $\delta^{13}C$ values ranged from -26.2‰ to -21.6‰ with the most depleted values found in the deeper core sediments (Fig. 2.2). To enhance the measures here with a finer scale sampling, results from McKay et al. (2008) are included in an expanded table of sediment properties (Table A2.1).

Table 2.1. Bulk Parameters of sediment core HLY0501-JPC5 from the northern Chukchi Sea. These parameters are expanded to include data from McKay et al. (2008) in Table A2.1. TOC: total organic carbon

Calculated Age (ka)*	Corrected Sample Depth (cm)**	TOC (%)	Nitrogen (%)	C:N	$\delta^{13}C$ (‰)
0.0067	0-2	1.38	0.19	8.40	-21.73
0.2747	40-42	1.36	0.15	10.62	-22.15
0.5092	75-77	1.26	0.16	9.44	-23.07
0.5427	80-82	1.33	0.16	9.78	-21.56
0.7772	115-117	1.31	0.16	9.70	-21.82
0.8107	120-122	1.27	0.14	10.62	-22.32
1.2462	185-187	1.25	0.14	10.59	-23.48
1.8492	275-277	1.21	0.16	8.61	-21.72
2.6532	395-397	1.18	0.15	9.11	-22.61
3.4572	515-517	1.17	0.14	9.85	-22.26
4.2612	635-637	1.21	0.16	8.91	-22.88
5.0652	755-757	1.18	0.13	10.83	-24.01
5.6012	835-837	1.17	0.16	8.79	-24.10
5.8692	875-877	1.15	0.15	9.11	-22.55
6.6732	995-997	1.17	0.13	10.62	-23.38
7.4772	1115-1117	1.14	0.14	9.61	-24.03
8.0132	1195-1197	0.86	0.11	9.16	-25.24
8.8172	1315-1317	0.60	0.07	10.07	-26.15

* Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); ** Depth correction applied as per Darby, Ortiz, et al. (2009).

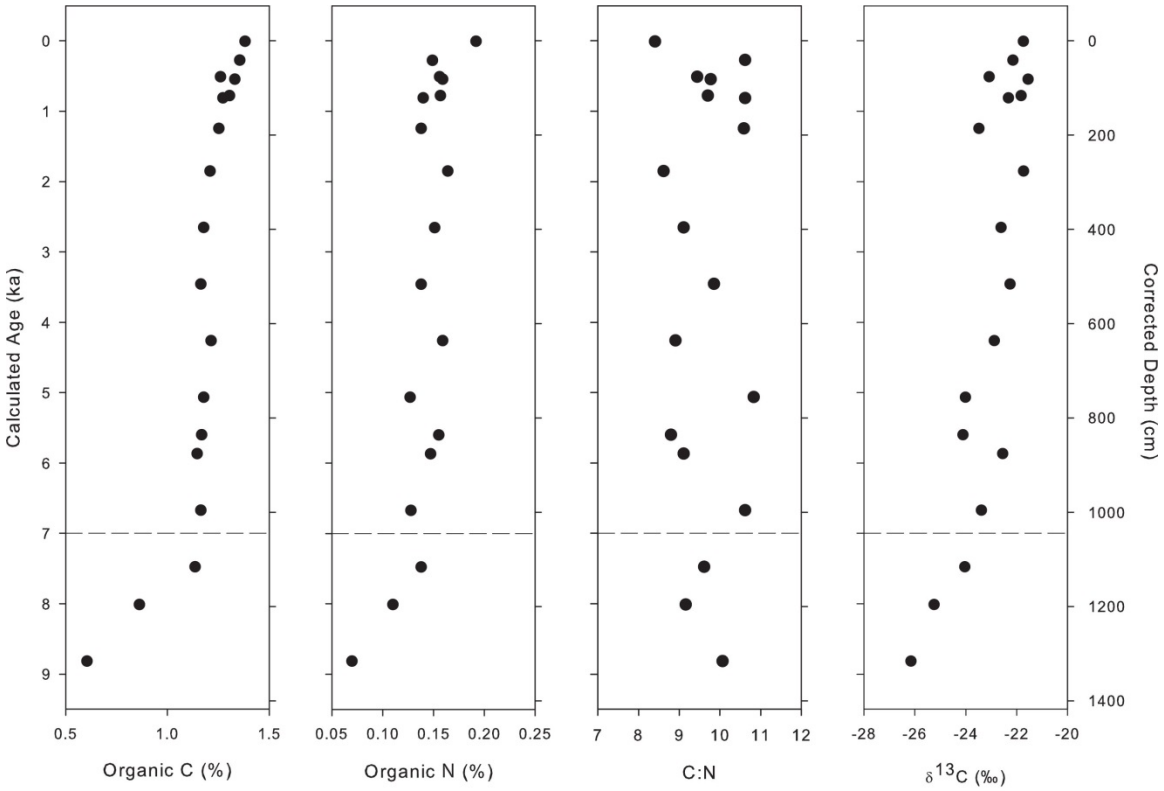


Figure 2.2. Bulk properties, including organic carbon (%), organic N (%), atomic C:N ratio, and $\delta^{13}C$ (‰), of sediment core HLY0501-JPC5 from the northern Chukchi Sea. Points are plotted at midpoint of sampled horizon. Calculated age is based on mid-point of corrected depths and determined based upon age model from Darby, Ortiz, et al. (2009). Dashed line indicates potential end of transitional period of sea level rise.

2.4.3. Molecular biomarker distributions

2.4.3.1. Fatty acids

Total fatty acids throughout the core ranged from 1331.3 to 3444.2 μg g^{-1} OC (Table 2.2). No distinct patterns were seen in distribution over the depths of the entire core, although more variation was observed in shallower horizons. Concentrations showed little variation throughout the upper 1100 cm of the core with the lowest concentration at the 115-117 cm interval and the highest concentration at the 755-757 cm interval. Total fatty acids decreased in the two deepest horizons sampled, 1196 cm and 1316 cm, of 1884 μg g^{-1} OC and 1331 μg g^{-1} OC, respectively (Fig. 2.3). Among fatty

acid classes, saturated fatty acids were the dominant contributor throughout the core (comprising 66.3 to 92.6% of the total fatty acids) of which short-chain fatty acids (C_{14} – C_{21}) accounted for 32.6 to 73.9% and long-chain fatty acids (C_{22}-C_{32}) contributed 26.1 to 67.4% (Table 2.3). There was a substantial input from C16:0n and C18:0n with C16:0n representing 13.3 to 20.6% and C18:0n representing 8.3 to 41.4% of total fatty acids. Short and long-chain saturated fatty acids were generally similar in concentration throughout much of the core with three near surface horizons having higher concentrations of long-chain fatty acids (Fig. 2.3). Long and short-chain fatty acids diverged below 1100 cm, with the long-chain fraction decreasing in concentration in the two deepest horizons and the short-chain fraction remaining steady. This decrease in long-chain fatty acids appears to affect the total fatty acid concentration creating a similar pattern in the long chain fatty acids and the total fatty acids in the deepest horizons (Fig. 2.3).

Monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), branched fatty acids and dicarboxylic acids (DCA) were all present in lower concentrations than saturated acids with PUFA, branched FA (data not shown), and DCA being found in very low concentrations (Fig. 2.3). The MUFAs comprised 4.6 to 24.0% of the total fatty acids and were dominated by the C_{16} and C_{18} MUFAs. The C_{16} MUFAs were major contributors with 0.69 to 13.9% while the sum of the C_{18} MUFAs amounted to 2.52 to 8.72% of total fatty acids. The overall trend for MUFAs was to decrease down-core with highest concentrations occurring at the 0-2 cm, 395-397 cm, and 755-757 cm horizons (Fig. 2.3). PUFAs and DCAs showed little variation with depth with the exception of a peak in DCAs in the 835-837 cm horizon (Fig. 2.3). Branched FAs

showed the highest concentration in the 0-2 cm horizon and then remained fairly constant throughout the remainder of the core, before decreasing in the two deepest horizons (data not shown).

Table 2.2. Concentrations of major lipid markers (as μg g^{-1} OC), and Branched and Isoprenoid Tetraether index (BIT index) of sediment core HLY0501-JPC5 from the northern Chukchi Sea. The BIT Indices for the 40-42, 80-82, and 120-122 cm intervals of the TC were not measured.

Calculated Age (ka)*	Corrected Sample Depth (cm)**	Total Fatty Acids	Total *n*-alkanes	Total *n*-alcohols	HPA Ratio	BIT Index
0.0067	0-2	2898.5	282.0	693.3	0.31	0.028
0.2747	40-42	2630.6	228.7	503.4	0.31	-
0.5092	75-77	2698.6	352.6	772.5	0.31	0.022
0.5427	80-82	2843.9	222.3	578.7	0.24	-
0.7772	115-117	2119.8	545.6	961.2	0.35	0.022
0.8107	120-122	2891.9	294.4	595.1	0.33	-
1.2462	185-187	2768.1	377.9	888.7	0.28	0.029
1.8492	275-277	2814.2	342.5	685.0	0.34	0.029
2.6532	395-397	3055.6	385.4	799.8	0.32	0.030
3.4572	515-517	2779.0	424.6	844.9	0.34	0.035
4.2612	635-637	3002.1	471.4	909.6	0.35	0.044
5.0652	755-757	3444.2	505.3	1071.4	0.32	0.048
5.6012	835-837	2561.4	522.2	437.6	0.53	0.049
5.8692	875-877	2614.6	505.7	862.3	0.37	0.056
6.6732	995-997	2539.3	405.1	938.4	0.28	0.065
7.4772	1115-1117	2579.1	623.9	1177.4	0.33	0.090
8.0132	1195-1197	1883.5	512.7	444.0	0.48	0.137
8.8172	1315-1317	1331.2	291.2	287.6	0.43	0.216

* Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); ** Depth correction applied as per Darby, Ortiz, et al. (2009).

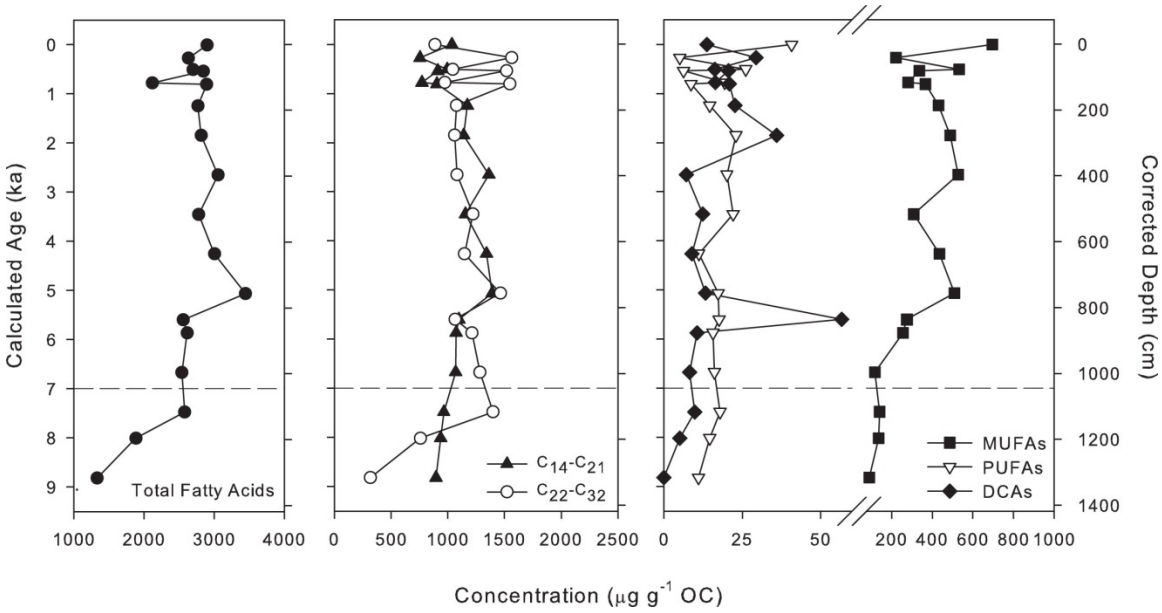


Figure 2.3. The fatty acid distribution seen over time in sediments from the northern Chukchi Sea shelf. The C_{14}-C_{21} and C_{22}-C_{32} designations refer to saturated fatty acids. Depth correction and calculated age were determined following Darby, Ortiz, et al. (2009). Dashed line indicates potential end of transitional period of sea level rise. OC: organic carbon, MUFA: monounsaturated fatty acid, PUFA: polyunsaturated fatty acid, DCA: dicarboxylic acid. Note scale changes between individual graphs.

Table 2.3. The distribution of total fatty acids and major groups (as μg g^{-1} OC) in sediments in piston core collected in HLY0501-JPC5 from the northern Chukchi Sea.

Calculated Age (ka)*	Corrected Sample Depth (cm)**	Total Fatty Acids	C_{14}-C_{21} Saturated Fatty Acids	C_{22}-C_{32} Saturated Fatty Acids	MUFA	PUFA	DCA
0.0067	0-2	2898.5	1035.3	887.8	695.6	40.8	13.7
0.2747	40-42	2630.6	754.7	1563.3	217.5	5.1	29.4
0.5092	75-77	2698.6	996.9	1042.8	531.2	26.1	16.3
0.5427	80-82	2843.9	914.0	1516.7	336.6	6.2	20.6
0.7772	115-117	2119.8	771.8	973.5	280.8	19.3	16.4
0.8107	120-122	2891.9	905.0	1546.7	365.5	8.7	20.8
1.2462	185-187	2768.1	1174.0	1075.9	430.7	14.6	22.7
1.8492	275-277	2814.2	1141.5	1061.3	488.1	23.0	36.0
2.6532	395-397	3055.6	1363.1	1080.5	527.5	20.1	7.1
3.4572	515-517	2779.0	1156.8	1220.6	308.0	22.0	12.4
4.2612	635-637	3002.1	1341.8	1146.4	434.8	11.1	9.0
5.0652	755-757	3444.2	1390.8	1463.6	508.8	17.2	13.3
5.6012	835-837	2561.4	1096.7	1064.2	275.1	17.6	56.7
5.8692	875-877	2614.6	1072.5	1215.1	256.2	15.7	10.5
6.6732	995-997	2539.3	1068.8	1282.3	118.1	16.0	8.1
7.4772	1115-1117	2579.1	965.2	1400.8	141.3	17.9	9.8
8.0132	1195-1197	1883.5	939.2	757.6	136.4	14.5	5.1
8.8172	1315-1317	1331.2	898.4	317.0	90.8	11.0	0.0

* Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); ** Depth correction applied as per Darby, Ortiz, et al. (2009).

2.4.3.2. Alkanes, Alcohols and Sterols

Concentrations of *n*-alcohols ranged from 287.6 to 1177.4 μg g^{-1} OC down-core. Long-chain *n*-alcohols (C_{22}-C_{26}) were greater in concentration than shorter chain alcohols (C_{14}-C_{21}) and varied from 203.5 to 838.7 μg g^{-1} OC. The lowest *n*-alcohol concentration was found in the deepest horizon with low concentrations also seen at 835-837 and 1195-1197cm (Fig. 2.4). Concentrations of *n*-alkanes ranged from 291.2 to 623.9 μg g^{-1} OC with long-chain *n*-alkanes (C_{23}-C_{33}) ranging from 167.3 to 418.7 μg g^{-1} OC with an odd-over-even predominance throughout the core (Fig. 2.4). There were minimal

contributions of short-chain *n*-alkanes (C_{17}, C_{19}). As with the fatty acids, there were slight variations among alcohols and alkanes near the top of the core.

The Higher Plant Alkanes, or HPA index, was calculated based upon concentrations of $C_{24\text{-}28}$ even-chain *n*-alcohols and $C_{27\text{-}31}$ odd-chain *n*-alkanes using the formula of Westerhausen et al., (1993):

HPA = [$C_{24\text{-}28}$ even *n*-alcohols] / [$C_{24\text{-}28}$ even *n*-alcohols + $C_{27\text{-}31}$ odd *n*-alkanes].

Values of the HPA index ranged from 0.24 to 0.53 (Table 2.2). The majority of the time span sampled ranged between 0.28 and 0.35 and showed only minor variation with depth.

Total sterol concentrations ranged from 96.6 to 435.1 μg g^{-1} OC. Of the fourteen predominant sterols, the six most abundant sterols included: 24-methlycholesta-5,22-dien-3β-ol ($C_{28}\Delta^{5,22}$), 24-methylcholesta-5,24(28)-dien-3β-ol ($C_{28}\Delta^{5,24(28)}$), 4α,24-dimethylcholestan-3β-ol, 4α,23,24-trimethylcholest-22-en-3β-ol (dinosterol), cholest-5-en-3β-ol (cholesterol), and 24-ethylcholest-5-en-3β-ol ($C_{29}\Delta^{5}$). Concentrations of these sterols varied over time, with the greatest changes occurring near surface. The one exception was a shift at the 835-837 cm horizon (5.62 ka) where all sterols showed a decrease to concentrations comparable to those seen in the oldest two horizons at 8.01 to 8.8 ka. The $C_{29}\Delta^{5}$ showed two peaks in concentration in the 755-757 cm and 1115-1117 cm horizons with 63.8 and 73.5 μg g^{-1} OC, respectively. $C_{28}\Delta^{5,22}$, $C_{28}\Delta^{5,24(28)}$, 4α,24-dimethylcholestan-3β-ol, dinosterol, and cholesterol all vary between 2.1 and 67.5 μg g^{-1} OC while $C_{29}\Delta^{5}$ shows slightly higher concentrations varying from 15.7 to 73.5 μg g^{-1} OC throughout the core (Fig. 2.5).

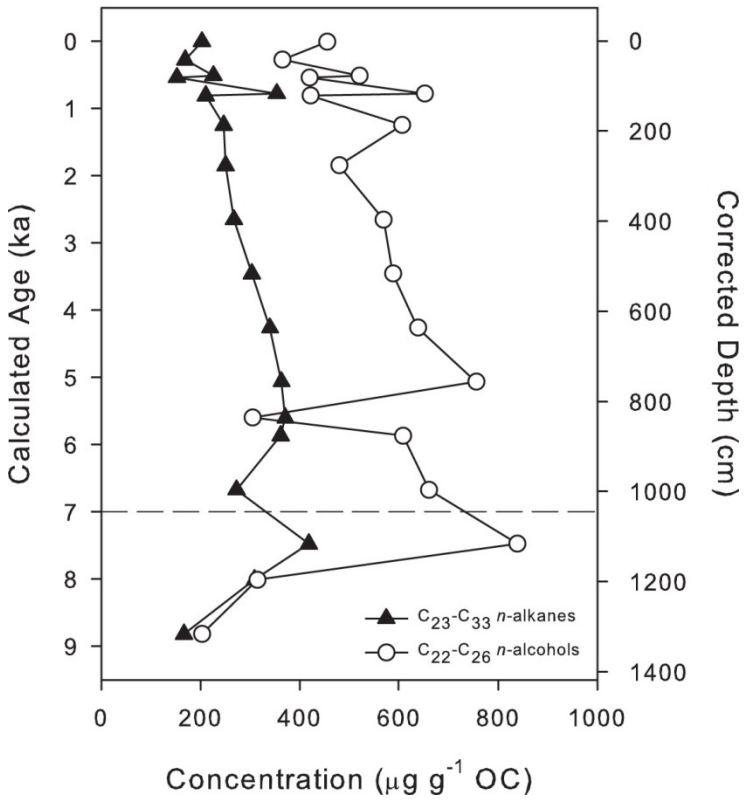


Figure 2.4. *n*-alcohol and *n*-alkane distributions in sediment core HLY0501-JPC5 from the northern Chukchi Sea. Points plotted at midpoint of sampled horizon. Dashed line indicates potential end of transitional period of sea level rise. OC: organic carbon.

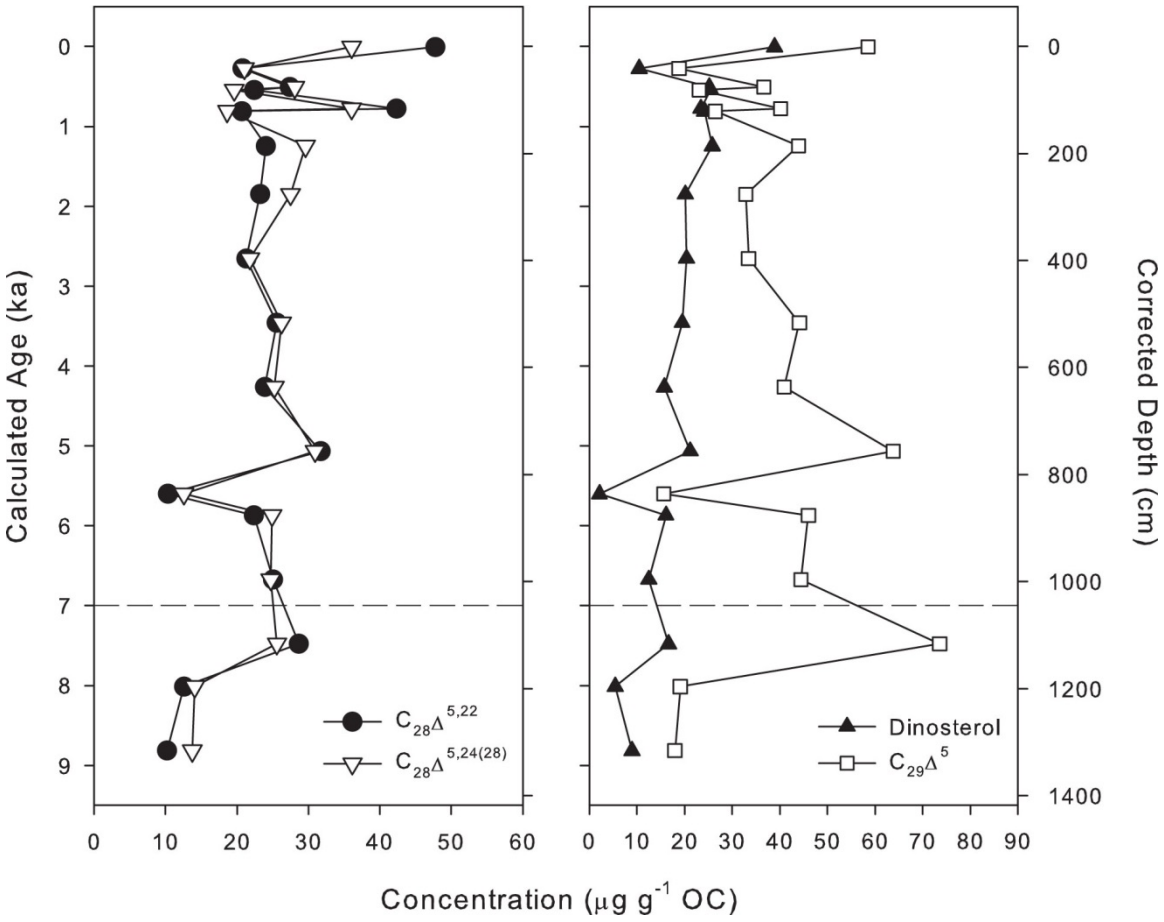


Figure 2.5. The abundance and distribution of major sterols seen in dated horizons of sediment core HLY0501-JPC5 from the northern Chukchi Sea. Sterols include the $C_{28}\Delta^{5,22}$: 24-methlycholesta-5,22-dien-3β-ol; $C_{28}\Delta^{5,24(28)}$: 24-methylcholesta-5,24(28)-dien-3β-ol; $C_{29}\Delta^{5}$: 24-ethylcholest-5-en-3β-ol; Dinosterol: 4a,23,24-trimethylcholest-22-en-3β-ol. Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009). Dashed line indicates potential end of transitional period of sea level rise. OC: organic carbon.

2.4.3.3. GDGT Lipids

Ratios of GDGTs were utilized to calculate the BIT Index. All samples contained branched GDGTs (I-III), crenarchaeol (IV) and the ubiquitous GDGT V. BIT indices showed slight increases down-core from 0.028 at the surface to 0.216 at depth (Fig. 2.6). BIT indices correlated significantly with $\delta^{13}C$ values down-core ($r = -0.857$; $p < 0.0001$; $n = 15$, using SAS 9.2) (Fig. 2.8).

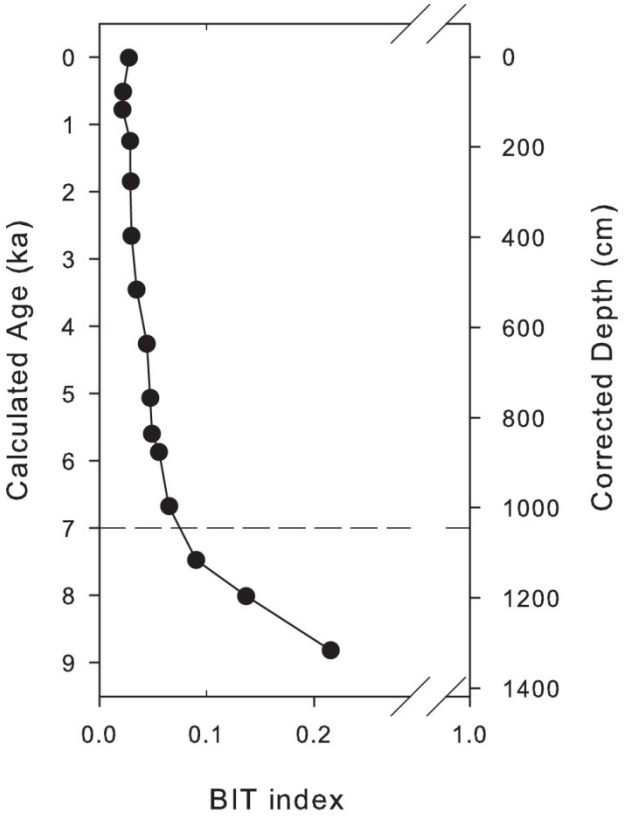


Figure 2.6. The Branched and Isoprenoid Tetraether index (BIT index) observed in sediment core HLY0501-JPC5 from the northern Chukchi Sea. Points plotted at midpoint of sampled horizon. Dashed line indicates potential end of transitional period of sea level rise. Data for the 40-42, 80-82 and 120-122 cm intervals of the TC were not measured.

2.5. Discussion

The multi-proxy approach including bulk parameters, molecular biomarker distributions and novel proxies, i.e. BIT index, show an Arctic Holocene sediment sequence with remarkable homogeneity. Higher density sampling of bulk properties in the same core was conducted by McKay et al. (2008) and show similar trends with depth for TOC and carbon isotopic signatures (Table A2.1). Bulk parameters for JPC5 suggest a predominantly marine source for the organic material comprising this core with

continual, but somewhat lower inputs of terrestrial derived organic materials. Organic carbon content itself remains fairly consistent throughout the core with concentrations only decreasing in the deepest horizons, suggesting either reduced amounts of organic carbon (allochthonous or autochthonous) being sequestered at that time (8 to 9 ka) or enhanced degradation of the organic fraction compared to more modern sediments in the upper segments of the core. Given the constant sedimentation for JPC5 (approx. 149 cm Kyr^{-1}; Darby, Ortiz, et al. 2009) together with the small variation observed in organic carbon content, it seems unlikely that substantial changes in organic carbon accumulation rate have occurred in this location.

The C:N ratios throughout JPC5 (8.4 to 10.8) fall within the typical reported ratios of 4 to 10 for algal organic matter production (Meyers, 1997). These values are consistent with those found by Belicka and Harvey (2009) for surface sediments from the continental slope at the nearby East Hanna Shoal region in which C:N ratios (for sites EHS6 and EHS9) were 8.1 and 7.2. In addition, C:N ratios of 9.22 in surface sediments of the adjacent Chukchi shelf and 10.38 in Chukchi Basin surface sediments have been reported (Belicka et al. 2002), two locations that bracket the sample setting for JPC5. Areas such as the Beaufort shelves which receive large amounts of terrestrial material from the Mackenzie River tend to have C:N ratios of 9 to 14 (Goñi et al., 2000). It should be noted that the calculation of C:N ratios in the present study utilized standard values of total nitrogen present in the samples. The JPC5 sediments were acidified before analysis eliminating most of the inorganic nitrogen and the resultant N values were accepted as organic N. Schubert and Calvert (2001) demonstrated that values obtained for $C:N_{tot}$ and $C:N_{org}$ can vary significantly in certain regions of the Arctic

which they attribute to the presence of bound ammonium to clay. Increased amounts of bound inorganic nitrogen would lower observed C:N ratios, thus suggesting a more marine character of the overall sediment material. They found that this scenario could be of consequence in regions with elevated clay content in sediments. The amount of bound inorganic nitrogen was not determined for JPC5 samples, and thus the C:N ratios described here may be maximal. The use of multiple other proxies supports the conclusion that the majority of contributed organic material to JPC5 is indeed marine.

Identification of organic matter source through the use of isotopic carbon measurements ($\delta^{13}C$) can be more difficult in Arctic sediments than in sediments in low latitudes (Mueller-Lupp et al., 2000; Schubert and Calvert, 2001). While the terrestrial end-member is consistently defined at -26 to -28‰ (Goñi et al., 2005; Mueller-Lupp et al., 2000; Naidu et al., 2000, Schubert and Calvert, 2001), the marine end-member is highly variable and far from constant. The variability in the $\delta^{13}C$ for marine carbon in the Arctic stems from several factors, including differing fractionation by phytoplankton as a result of increased CO_2 solubility, differences in cell sizes and growth rates (Laws et al., 1995) as well as from inputs of sea-ice algae which can be isotopically enriched (Gradinger, 2009). Observed values for "marine carbon" $\delta^{13}C$ values in the Arctic have ranged from -16 to -27‰ (Goericke and Fry, 1994) with several marine end members proposed including: -19 to -24‰ in the Beaufort Sea (Goñi et al., 2000), -21 and -24‰ for the Northern Bering/Chukchi Seas and Beaufort Sea, respectively (Naidu et al., 2000), and -19 to -21.3‰ based on modeled data from the Yermak Plateau (Schubert and Calvert, 2001). Given this range of possible values, any estimation of carbon source based on carbon isotopic values alone is subject to substantial uncertainty.

In the Holocene sediments examined in this study, values of $\delta^{13}C$ ranged from -26.2 to -21.6 ‰, with the deepest two segments having the most depleted values. These values are similar to those in the Alaskan Beaufort Sea of -24.6 to -18.6 ‰ (Belicka et al., 2009). In the northern Bering Sea and the Chukchi Sea additional values of -24 to -20 ‰ and -24 to -21 ‰, respectively, are reported (Naidu et al., 2000) and Mueller-Lupp et al. (2000) report values in the Laptev Sea of -26.6 ‰ near the Lena delta and -22.8 ‰ on the continental slope. This progression of $\delta^{13}C$ values seaward in the Bering, Chukchi and Laptev Seas from depleted to more enriched (lighter to heavier) appear to be due to the mixing of terrestrial and marine organic carbon with the most enriched values reflective of largely increased marine inputs (Mueller-Lupp et al., 2000; Naidu et al., 2000). Yet, the strict assignment of a $\delta^{13}C$ value at which those inputs become entirely, or mostly, marine is difficult. For example, Naidu et al. (2000) inferred that the more depleted values (-24 ‰) in the Bering and Chukchi Seas are due to greater influence from terrestrial sources; yet they rely on a similar value in the Beaufort Sea to define the marine end-member. Clearly the marine isotopic end-member is challenging to define and may be specific for each region of the Arctic. Given that, the majority of the $\delta^{13}C$ values determined for JPC5 fall within the large accepted range for a marine end-member in the Arctic, yet detailed biomarker information shows there is undoubtedly a contribution from terrestrial sources. Due to the uncertainty in defining a $\delta^{13}C$ marine end-member, the use of end-member mixing models to determine organic matter inputs at this location in the western Arctic is subject to large variability and limited utility (Belicka and Harvey, 2009). Instead, lipid biomarkers were used in an attempt to constrain the relative contributions of terrestrial and marine sources.

Multivariate statistical approaches have shown promise in previous studies to constrain relative contributions from marine and terrestrial sources to Arctic sediment (Belicka and Harvey, 2009; Yunker et al., 1995, 2005). Attempts to apply these approaches to JPC5 were limited by the absence (likely through degradation) of many of the labile biomarkers indicative of specific organic matter sources and necessary to generate the proportional contributions from each source with confidence. In addition, these statistical techniques are most effective when multiple sample locations are used. Given these caveats, however, we can make initial estimates of the source contributions to JPC5 using the proportions of biomarker contributions determined by Belicka and Harvey (2009) for nearby locales. Nineteen of the 25 sediment cores used by Belicka and Harvey (2009) in their determination of organic source contributions flanked the JPC5 site, allowing a reasonable extrapolation of their results as an exploratory determination of inputs. Using their calculated contributions on the parallel biomarkers seen in JPC5, we can estimate the distributions of marine and terrestrial sources in JPC5 over time (Fig. 2.7). These estimates support the argument that terrestrial and marine sources of organic matter are very similar in magnitude down-core, and are in agreement with only minor changes seen down-core for bulk parameters. As marine organic matter is commonly accepted as labile, its presence in similar proportions to terrestrial material suggests it was a major source of initially deposited material. This infers a predominantly marine source of organic matter to JPC5 with significant and continual contributions from terrestrial sources.

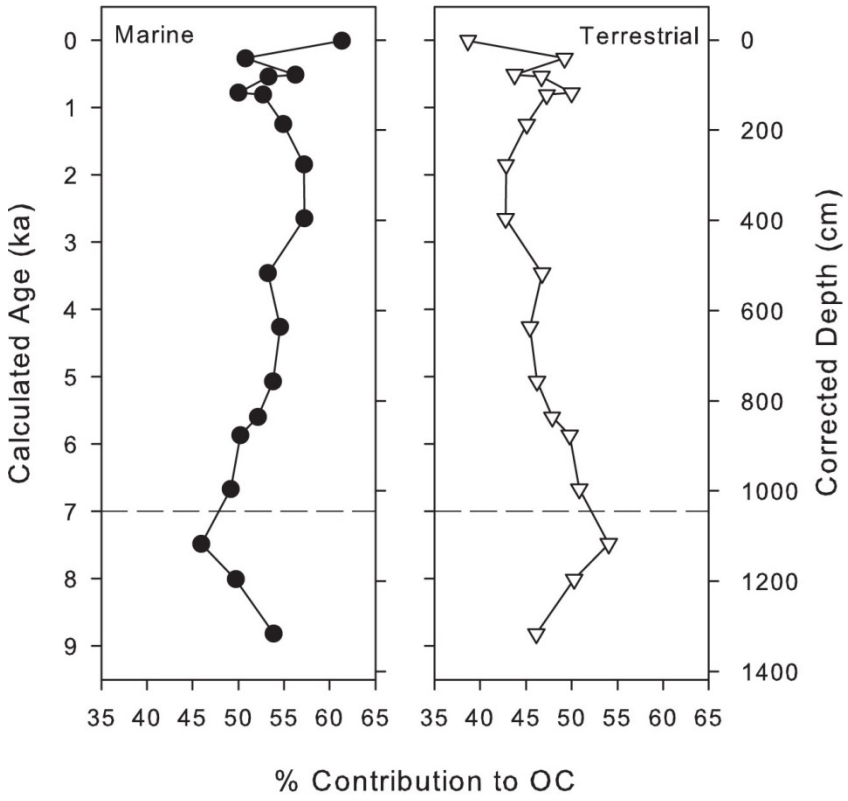


Figure 2.7. The fraction of marine and terrestrial organic matter over time seen in sediment core HLY0501-JPC5 from the northern Chukchi Sea. Note that estimates provide for two major sources accounting for all organic matter present over the time span examined. Depths are shown as the mid-point of the sediment section sampled. Calculated age is based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009). Dashed line indicates potential end of transitional period of sea level rise. OC: organic carbon.

Evidence for the loss of marine organic matter is found in the dominance of the total fatty acids by saturated and monounsaturated forms, suggesting both substantial reworking of the sediments and removal of the more labile polyunsaturated fatty acids common to algal sources. The delivery and preservation of terrestrial organic matter is reflected by the presence of long-chain fatty acids (C_{22}-C_{32}) found in JPC5, which showed an even-over-odd predominance, the presence of which is widely accepted as an indicator of terrestrial contributions (Belicka et al., 2009; Fahl and Stein, 1999) though

there are reports of long-chain fatty acids being of microalgal or bacterial origin (Birgel et al. 2004, and references therein). Although PUFAs are present throughout JPC5, their concentrations are much lower than those reported for surface sediments in nearby areas (Belicka et al., 2002; Belicka and Harvey, 2009). The low amounts of PUFAs present in sediments are not surprising given the expected losses typically seen during degradation. Fahl and Stein (1997), found the majority of surface sediments in their study sites in the Laptev Sea dominated by saturated fatty acids and short-chain MUFAs, attributing this to degradation and autoxidation of the unsaturated bonds. A number of the monounsaturated C_{18} fatty acids were also prominent contributors. Kattner et al. (1983) found 18:1ω9 as one of the main fatty acids present in a spring phytoplankton bloom, hinting that in our samples it indicates marine inputs. Long-chain dicarboxylic acid (DCA) concentrations, likely representative of higher plant inputs, throughout the core were lower than concentrations found in surface sediments in nearby locations (Belicka and Harvey, 2009), though their presence again points to the presence of a terrestrial component in this core.

Within total *n*-alcohol concentrations, long-chain alcohols (C_{22}, C_{24}, C_{26}) were present in the highest abundance and demonstrated an even-over-odd predominance indicative of terrestrial inputs (Volkman, 2006). The long-chain alcohol concentrations in the 835-837 cm horizon deviated from the general trend down the core showing one of the lowest concentrations within the core. A similar decrease was seen in sterol concentrations for this horizon, yet long-chain alkane and fatty acid concentrations did not show this deviation. This pattern is difficult to interpret because both terrestrial (long-chain alcohols and $C_{29}\Delta^5$) and marine ($C_{28}\Delta^{5,22}$, $C_{28}\Delta^{5,24(28)}$, and dinosterol)

biomarkers show parallel decreases in concentrations. A related measure, the Higher Plant Index, has been used to assess the extent of degradation of terrestrial matter (Westerhausen et al., 1993), but remains fairly consistent throughout the core. Although the absolute values for this index are regionally specific, here the lack of variation demonstrates that deposited organic material has been largely homogenous over long time periods. It argues that the organic material that has accumulated at this site experienced mixing prior to deposition.

The sterol distribution, notably the high concentrations of cholesterol, $C_{28}\Delta^{5,22}$ and $C_{28}\Delta^{5,24(28)}$ provide further evidence for abundant marine organic matter and secondary production. In JPC5, $C_{28}\Delta^{5,22}$ and $C_{28}\Delta^{5,24(28)}$ comprise up to approximately 25% of the total sterols, and cholesterol up to approximately 22%. The former two compounds are thought to be diatom derived (Volkman, 1986), while cholesterol likely reflects broad eukaryotic inputs from both the water column and sedimentary residents (Volkman, 2003). Dinosterol is highly specific to dinoflagellate inputs (Volkman, 2006). McKay et al. (2008) recently used counts of dinocyst (dinoflagellate cysts) assemblages as a proxy for sea surface conditions recorded in JPC5. An attempt was made to compare numbers of dinocysts obtained by McKay et al. (2008) to actual dinosterol concentrations determined in JPC5. Although both suggest dinoflagellate inputs to sediments throughout the almost 9000 years of deposition, statistical evaluation revealed no significant correlation. Despite the high concentrations of these marine biomarkers, we cannot ignore the substantial contribution of $C_{29}\Delta^{5}$ to the total sterols (Fig. 2.5). Although occasionally derived from plankton, in polar regions, this sterol is considered a

terrestrial indicator and inputs correspond well to other terrestrial markers such as α-amyrin in nearby regions of the Arctic (Belicka and Harvey, 2009).

As an independent measure of terrestrial carbon inputs (Sinninghe Damsté et al., 2009), the BIT index provides an additional metric to constrain the origin of organic carbon present in sediments. The BIT index relies on the ratio of branched GDGTs to the sum of the branched GDGTs plus the isoprenoid crenarchaeol to determine the source of organic matter inputs. Branched GDGTs are thought to be produced by organisms living in terrestrial environments while crenarchaeol appears to be produced by marine planktonic archaea. Ratios closer to zero indicate primarily marine sources of organic matter while ratios closer to one correlate with terrestrial organic matter sources (Hopmans et al., 2004). The BIT indices for JPC5 ranged from 0.021 to 0.216, with lowest values in surface sediments and increasing only slightly to the deepest interval. The low values obtained in upper core horizons are comparable to those found in surface sediments of nearby areas by Belicka and Harvey (2009) (East Hanna Shoal stations 6 and 9; 0.03 and 0.04 respectively). We can compare the BIT indices and $\delta^{13}C$ values of JPC5 to examine their prediction of terrestrial inputs. A simple correlation of isotopic values versus the BIT index values finds they are significantly correlated (Fig. 2.8). Both proxies shift toward values in the deepest core segments that may be considered indicative of more terrestrial inputs, i.e., larger BIT indices and more depleted $\delta^{13}C$ values. Together with the stable isotopic data from McKay et al. (2008) it suggests that terrestrial organic inputs to this shelf environment may have been greater in the early Holocene. Yet, corresponding increases of other terrestrial markers such as long-chain n-alkanes, n-alcohols, and n-alkanoic acids, in early Holocene sediments were not apparent,

and even decreased in these deepest segments. Estimates of the contributions of terrestrial and marine sources (Fig. 2.7) suggest that the two sources are roughly equivalent in the second deepest horizon, with marine sources comprising a larger proportion of the organic matter remaining in the deepest horizon. As pointed out by Belicka and Harvey (2009), estimates of terrestrial organic matter from the BIT index are likely not directly comparable to estimates from other lipid biomarkers as these proxies may estimate separate pools of terrestrial organic matter, such as peat or soil organic matter versus higher plant wax inputs. While the BIT index suggests a greater proportion of terrestrial inputs from soils or peats during the early Holocene, biomarker signatures indicate that an increased fraction of marine carbon was deposited and preserved over time.

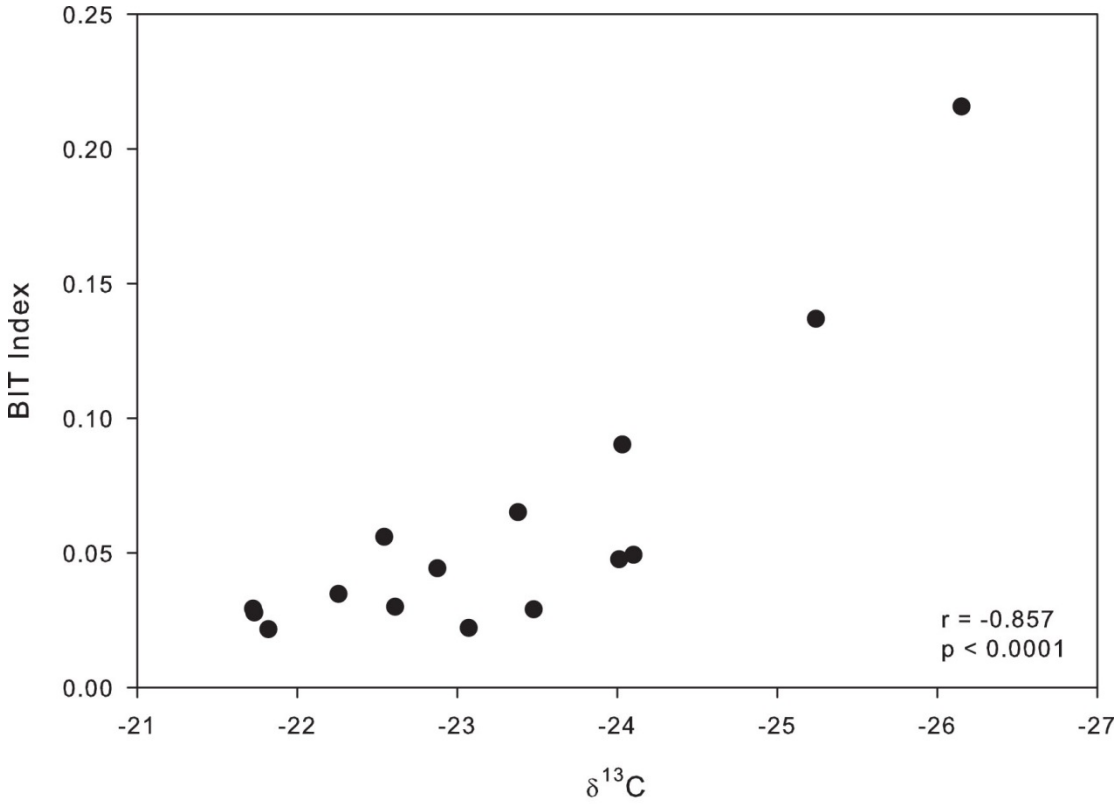


Figure 2.8. Correlation between Branched and Isoprenoid Tetraether indices (BIT indices) and $\delta^{13}C$ values in sediment core HLY0501-JPC5 from the northern Chukchi Sea. The correlation coefficient (r) was determined using SAS 9.2 with an n = 15.

An important caveat with all the proxies used in this study is their source specificity. Fahl and Stein (1999) have discussed the observation that a number of the marine indicators can also be attributed to freshwater organisms under some conditions. In continental margins such as the Laptev Sea and Arctic Ocean, where different transport mechanisms of terrestrial organic matter than those seen in lower latitudes are important, the use of biomarker proxies is more complex and they have suggested that the presence of short-chain fatty acids, dinosterol, and $C_{28}\Delta^{5,22}$ in shelf sediments might also be indicative of freshwater inputs as opposed to marine inputs (Fahl and Stein, 1999).

However, in slope sediments, they attribute higher concentrations of short-chain fatty acids, dinosterol, and $C_{28}\Delta^{5,22}$ to marine origins. Our study site on the Chukchi slope is unlikely to receive major riverine inputs as opposed to locations further east and west near abundant river outfalls and unlikely to receive significant amounts of short-chain fatty acids, dinosterol, and $C_{28}\Delta^{5,22}$ seen from freshwater organisms. In addition, the use of multiple proxies alleviates the uncertainties of relying on isolated single markers for identification of organic matter origins.

The JPC5 sediments show remarkably consistent distributions down-core, yet the region experienced several changes in sea level and coverage of sea ice (Macdonald et al., 2004). McKay et al. (2008) cite a decrease in C:N ratios and an increase in $\delta^{13}C$ in younger sediments of JPC5 as an indication that more terrestrial material entered the region in the early Holocene. Our own analysis did not find a marked change in C:N, yet we did observe a subtle increase in $\delta^{13}C$ (-26 to -24‰) and a small shift in the BIT index to more marine character in the younger sediments of JPC5. There was no indication of an increase in other terrestrially derived markers in the two deepest intervals of the core. Long-chain *n*-alkanes, *n*-alcohols and long-chain fatty acids all decreased in the oldest sediment horizons and showed no alterations that would indicate changes in inputs of terrestrial materials during the early Holocene. Instead, the decrease in total organic carbon and parallel decrease seen in terrestrial markers (with concentrations based on organic content) in the older Holocene sediments argues that these terrestrial sources were decreased relative to the marine markers. Further, the diatom markers $C_{28}\Delta^{5,22}$, $C_{28}\Delta^{5,24(28)}$ and dinosterol, which decline at the 1195-1197 cm horizon but then remain steady, and the short-chain fatty acids, which don't show a marked decrease in the oldest

horizons, suggest that terrestrial carbon inputs were not greater at the beginning of the Holocene. As pointed out previously, the BIT index may be measuring a different portion of terrestrial organic matter than the other lipid biomarkers, suggesting that the relative composition of the terrestrial matter may have been altered through the Holocene but that the total inputs may not have changed. In addition, proportions of terrestrial to marine organic matter do not indicate substantial changes down-core.

We had hypothesized that changes in organic matter source deposition throughout the Holocene, instigated by changes in sea level, would be apparent in the core. Sea level rise would cause the points of major deposition on the shelf to move landward as the coastline retreated inland which should be recorded as a progression from terrestrial to marine organic matter deposition in the core over time. Estimates indicate that sea level changes in the Arctic occurred between 9 and 11 ka in the Kara and Laptev Seas (Mann et al., 1995; Stein and Fahl, 2000; Stein et al., 2001) with one estimate of transitional periods lasting to 7 ka (Mueller-Lupp et al., 2000) and the stabilization of sea levels to modern conditions between 5 and 6 ka (Darby et al., 2006; Mueller-Lupp et al., 2000). The end to this transitional period at 7 ka is indicated in the included figures as a point of reference. The shifts seen in the organic carbon content and $\delta^{13}C$ values measured in this study fall between approximately 9 to 7 ka which corresponds to the estimates by Mueller-Lupp et al. (2000) of a transitional period lasting until 7 ka, although as discussed previously, other proxies do not support the conclusion of a major shift in inputs in this deeper portion of the core. This is supported by an examination of the location of JPC5 with respect to present and historic coastlines. The age model constructed for JPC5 dates the deepest analyzed horizon at approximately 8.8 ka that

would be near or after estimates of sea level rise in the Arctic. Estimates predict the level of the Arctic seas was approximately 60 to 70 m lower at the time sea levels began to rise (Mueller-Lupp et al., 2000; Stein and Fahl, 2000). Water depths at JPC5 are greater than 400 m suggesting that its location would not be affected by the advance and retreat of coastlines. Further, the depth profiles for the Chukchi shelf and slope show that Barrow Canyon crosses the likely transport path for terrestrial material being carried towards JPC5 that might also act as a preferential deposition zone for terrestrial materials. This combination of significant water depth and physical factors may help explain the lack of substantial changes in organic matter inputs to JPC5 during period of sea level change in the Arctic.

The observed homogeneity throughout the examined sediments of JPC5 is highly suggestive of mixing prior to deposition that results in time-dependent variation in marine versus terrestrial organic carbon inputs being obscured in the sedimentary record. Current flow throughout the Chukchi allows waters to mix as they travel from the Pacific through the Chukchi (Weingartner et al., 2005). As previously mentioned, Darby, Ortiz, et al. (2009) found that the region where JPC5 was recovered had constant, high sedimentation rates throughout the core of ~149 cm Kyr^{-1}. This high sedimentation rate and observed sediment particle size distributions indicate that this slope has some form of current deposition (Darby, Ortiz, et al., 2009). Darby, Ortiz, et al. (2009) speculated that the Beaufort Undercurrent might have been more constant through the Holocene, thereby accounting for the consistent sedimentation rates and that, as it passes the canyon margin, it may separate, thus causing deposition at JPC5. If the Beaufort Undercurrent splits in the area of JPC5, it would be slowed, and slower moving water would deposit particles

that are too large to remain suspended. They also suggested transport and deposition of sediment from sea ice. Sea ice leaves a record of change in sediments (Müller et al., 2009) and Eicken et al. (2005) found that sea ice potentially transports more sediment than previously thought in the Chukchi and western Beaufort Seas. While sea ice may have been an additional contributor to the sediments that accumulated on this section of the Western Arctic Shelf, evidence points to substantial mixing by bottom currents in Chukchi slope prior to accumulation (Darby, Ortiz, et al., 2009). Our interpretation is that these currents transport resuspended sediments from shallower regions that having already undergone degradation, mix the more recalcitrant terrestrial and residual marine input. This allows any significant changes in hydrological forcing and organic inputs experienced over the Holocene period examined to be blended prior to deposition, with little additional change thereafter other than expected diagenetic losses.

Polyak et al. (2009) examined another core (HLY0501-6JPC) collected during the HOTRAX expedition located in 673 m of water and determined that the upper sedimentary unit of the core, ca. 8.5 ka, represented Holocene sediments. They state that the sediments underlying this upper unit were much more heterogeneous than the Holocene sediments. Yamamoto and Polyak (2009) and Adler et al. (2009) examined a core (HLY0503-8JPC) from approximately 2800 m depth covering time spans of up to 130 ka. The studies of HLY0503-8JPC found changes in sedimentary proxies (i.e., sediment color, density, sand content, etc.) (Adler et al., 2009) and lipid biomarkers (Yamamoto and Polyak, 2009) within the core which were attributed to glacial-interglacial period variation. The presence of observed variations within this much deeper core seems to indicate that it is not only the changing coast lines that are affecting

the deposition of sediments and organic matter but other mechanisms as well such as sea ice transport or sub-surface current transport. The increased heterogeneity of the older sediments in the HLY0501-6JPC core and the changes seen in older sediments of the HLY0503-8JPC core may indicate that had deeper sediments of JPC5 been examined evidence of changing organic matter inputs may have been seen.

2.6. Conclusion

Evaluation of bulk parameters, molecular biomarker distributions and molecular proxies, i.e. BIT index, indicate that the organic material comprising an approximately 8800-year sediment record for the northern Chukchi Sea is primarily marine derived organic matter with moderately smaller contributions from terrestrial sources. The lack of obvious changes in the balance between marine and terrestrial organic carbon deposition throughout the core as determined by traditional lipid biomarkers indicates that the consistency of inputs to the core throughout the Holocene appears to be due to the mixing of material by bottom currents prior to deposition. Examination of older sediments from JPC5 may offer more information on glacial-interglacial transitions at this location.

2.7. Acknowledgments for this chapter

This work was supported by grant ARC-0612580 as part of the Arctic Natural Sciences Program of the National Science Foundation. We thank David Morris and Dr. Stephen Macko for isotopic analysis, Leonid Polyak for samples and many useful

discussions and Dennis Darby for advice on sedimentation rates. Jochen Knies and Ruediger Stein provided many useful comments which strengthened the final manuscript.

2.8. Addendum (not included in published paper)

It was determined through the research detailed in this chapter that lipid biomarkers did not provide evidence of great variations in inputs, and it appears that the core location was more influenced by mixing than changes to sea level or deposition. Lipid biomarkers used can provide some specificity of inputs, but they do not provide a great deal of information about the organisms producing the lipids or the conditions under which they were produced. After this initial investigation it was hypothesized that determination of the proteins in environmental systems may provide greater detail on the organic matter sources and what was occurring at the time of deposition. It was decided that another Arctic core would be used for proteomic analysis of down-core horizons and identified proteins would be corroborated to known lipid biomarkers. Samples for this subsequent analysis were obtained from a multi-core collected during the spring Bering Sea Ecosystem Study (BEST) cruise, 2010, in the Bering Sea, from a location designated as Mn-14 (59.90339 N, 175.80530 W). This core was approximately 35 cm in length and was collected in 130 m of water. The core was sub-sampled in 1 or 2 cm increments down the core. The outcome of the analysis of this second core is discussed in Chapter 3.

Chapter 3: Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments

A major portion of this chapter appears as "Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments". Moore, E.K., Nunn, B.L., Faux, J.F, Goodlett, D.R., Harvey, H.R. 2012. Limnology and Oceanography Methods, 10, 353-366.

3.1. 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 prior to 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, bovine serum albumin (BSA) from older (>1ky) sediments was optimized. Protein extraction from sediments via direct electrophoresis of a slurry mixture and the urea 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.

3.2. Introduction

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 30-40% of particulate nitrogen in marine sediments (Cowie and Hedges, 1992a; Grutters et al., 2001). While 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 proteins and/or their component peptides in marine sediment is required.

Extracting proteins from sediment for subsequent analysis has long been a challenge (Belluomini et al., 1986; Craig and Collins, 2000; Nunn and Keil, 2006; Ogunseitan, 1993). The potential interferences present in sediment or soils include protein binding to the mineral matrix (Collins et al., 1995; Keil et al., 1994; Mayer, 1994), 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 such as detergents to solubilize proteins can be effective, this 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 sedimentary minerals appears to be surface adsorption (Mayer, 1994; Mayer et al., 2002). Various mechanisms have been proposed for 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., Fairbanks et al., 1971; Laver, 1964; Maizel, 2000; Pederson, 2008; Reisfeld et al., 1962; Shapiro et al., 1967). This includes electrophoretic separation

to purify proteins from cell cultures prior to analysis with mass spectrometry (Botelho et al., 2010; Tran and Doucette, 2009). Here we present a modified electrophoretic approach as an extraction and preparative technique for complex environmental samples prior to high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) analysis for 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; Kuster, et al., 1998; Shevchenko et al., 1996). 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 SDS-PAGE to sediments 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 (Dong et al., 2010; Morris et al., 2010; Schulze et al., 2005). The extraction, isolation and analysis of proteins and peptides using MS/MS are only the first steps. In order 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 a high diversity of unknown organisms, incomplete protein databases, and diverse mixtures of proteins that yield limited concentrations of specific sequences for detection and identification (Bastida et al., 2009; Graves and Haystead, 2002; Quince et al., 2008).

The goals of this work were twofold. The first goal was to optimize the extraction of proteins from marine sediments with a broadly applicable methodology. The second 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 in which the buffer-solubilized material is separated from particles and loaded directly onto gels, and 2) a novel slurry extraction method in which buffer-solubilized material remains with the sediment particles and both are loaded on to the gel together. Two electrophoresis gels were compared including preparatory tube gels and standard 1-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. 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; 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.

3.3. Materials and Methods

3.3.1. 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 of extraction buffer in duplicate Falcon Tubes to yield a 5:1 buffer:sediment ratio (*v/v*). Tubes were sonicated on ice for 60 seconds using pulse sonication (Bronson microprobe, at 20 kHz). Sediment extracts were centrifuged to remove particles from the extraction liquid (5,000 *x g*, 10 minutes, 4^{o}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 that 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 seconds 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. 3.1). The top 5 cm of the gel was then excised for 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.

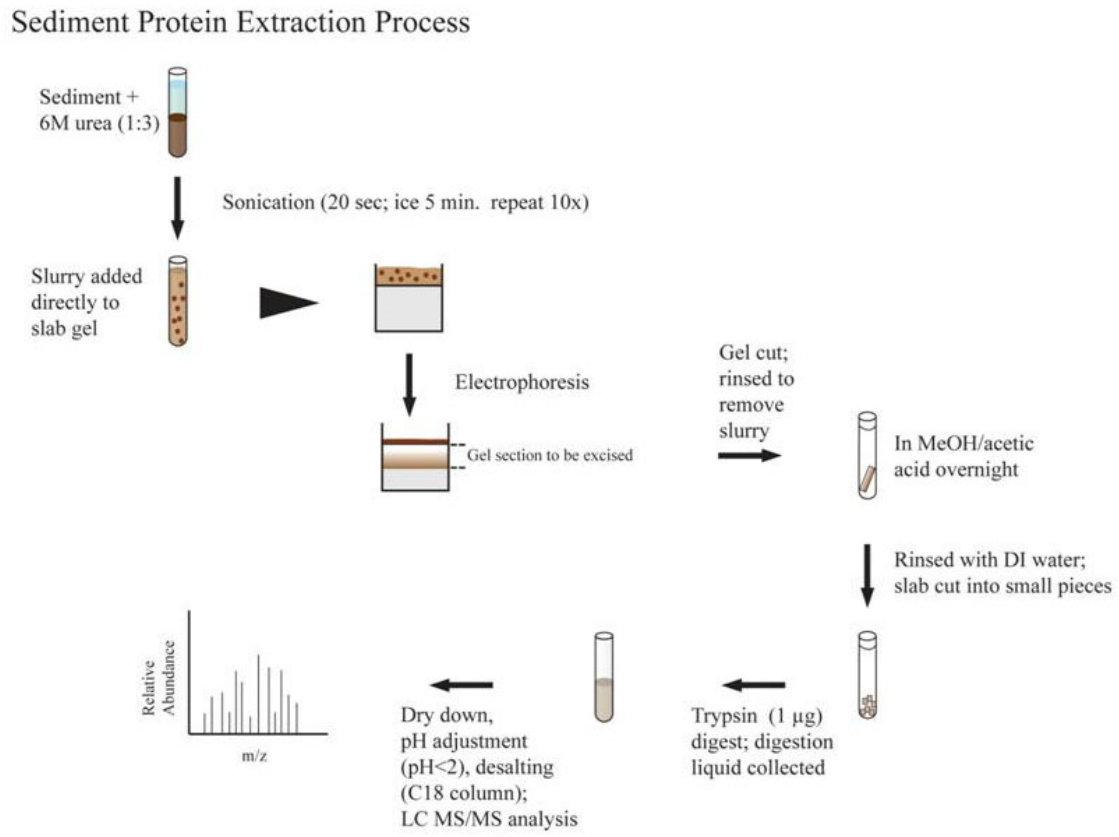


Figure 3.1. Schematic workflow for the slurry approach for extraction, purification, and digestion of sedimentary samples prior to LC/MS analysis.

3.3.2. 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 tested treatments in 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 pre-cast 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 methods of sample preparation 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 of extraction buffer was mixed with 250 µL of sediment, the exception being those extractions that employed only urea. To those samples 300 µL of urea was added. All extraction buffers contained the following (with the exception of the urea only extractions): 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). Samples extracted with 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 sec 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 loaded on 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 loaded 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 in order to rinse away excess SDS, and placed in a fixing solution of 50% methanol, 10% acetic acid overnight.

3.3.3. 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. Flat gels were rinsed with 8 mL of ammonium bicarbonate and acetonitrile and tube gels were rinsed with 18 mL. The rinse cycle was repeated twice more and gel pieces were dried via speedvac for 45 min. To the gel pieces, 1 μg of 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 samples was adjusted with 5% formic acid to a pH $\leq$ 2 and the sample was run through a

C18 desalting centrifuge column (Nest Group) following which samples were dried and volumes were adjusted in preparation for analysis.

3.3.4. Direct digestion of sediment

Prior to shotgun proteomic analysis, 100 mg of sediment were mixed with 300 µL of 6 M urea and 50 mM ammonium bicarbonate. The sediments were then sonicated using a Bronson sonicating microprobe, at 20 kHz for 60 sec. 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 hr (37^{o}C). Proteins were then alkylated by adding 60 µL of 200 mM iodoacetamide and incubated in the dark for 1 hr. After the addition and incubation of 60 µL of dithiothreitol (1 hr 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 x g, 20 minutes) and the supernatant containing the digest and buffer was removed. The sediments were then washed 3 times with 1 mL 25 mM ammonium bicarbonate, centrifuged, and supernatants combined. The volume was reduced via speedvac to ~10 µL, and 200 µL of 5% ACN, 0.1 % trifluoroacetic acid was added prior to desalting the peptides using a C18 desalting centrifuge column (NEST group). Sample pH was adjusted to <2 using small additions of 10% TFA and samples were desalted using the protocol provided by manufacturer.

3.3.5. Mass Spectrometry

Protein analyses for all samples were conducted using standard shotgun proteomic techniques employing nanocapillary HPLC-MS/MS as described previously (Aebersold and Goodlett, 2001; Washburn et al., 2001). Samples were introduced into a hybrid linear ion Orbitrap (LTQ-OT) mass spectrometer (Thermo Fisher, San Jose, CA) via a NanoAcquity high performance liquid chromatography (HPLC) system (Waters, Beverley, MA). Trapping and analytical capillary columns were packed in-house using a pressurized cylinder (Brechbuhler AG, Schlieren Switzerland). Magic C18 (5 μm diameter, 100 Å pore size) particles (Michrom Bioresource, Auburn, CA) were slurried with analytical grade MeOH and placed in the pressurized cylinder to pack columns using 6.895 x 10^6 Pa nitrogen. Trapping column capillaries were 20 mm x 100 μm i.d., while the analytical column dimensions were 150 mm x 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 (Manchester, U.K.). Based on parallel amino acid measurements of each sample, 1 μg of 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 x 10^7. The LTQ-OT XL (ThermoFisher, San Jose, CA) was operated in positive ion mode using data-dependent acquisition. 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 data dependent acquisition from the following *m/z* ranges: *350-444, 444-583, 583-825, 825-1600* (Davis et al., 2001; Scherl et al., 2008; Spahr et al., 2001; Yi, et al., 2002).

3.3.6. 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, as the high variability of post-translational modifications make confident identification difficult. Four protein sequence databases were generated. The FASTA for these were generated by concantenating databases. The four databases were 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 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 non-redundant 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 to 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 Dalton (Da) on cysteine (resulting from IAM alkylation) and 16 Da on methionine from oxidation. Predicted fragmentation vs. observed tandem mass spectra were statistically evaluated with PeptideProphet software in order 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 would 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.

3.3.7. 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 hydrolysed for 4 hours at 110^{o}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^{o}C and separated through a DB-5MS (0.25 mm ID, 30 m) GC column with hydrogen as the carrier gas. Amino acid identification utilized 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. BSA was analyzed in parallel to correct for responses among individual amino acids and calculation of molar ratios.

3.4. Results

3.4.1. Sediment properties and amino acids

The surface sediment examined in this study contained 0.48% organic carbon (OC) and 0.06% particulate nitrogen (PN) while the deeper sediments utilized in the optimization experiments contained 1.07% OC and 0.15% PN. Amino acids represented 1.39 $\pm$ 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 contained 2.59 $\pm$ 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 vs. THAA content of sediments as a proxy for total protein, and was calculated for surface sediment samples resulting in efficiencies of 12.5% $\pm$ 1.1 for the traditional buffer surface sediment extraction, and 100.6% $\pm$ 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. 3.2).

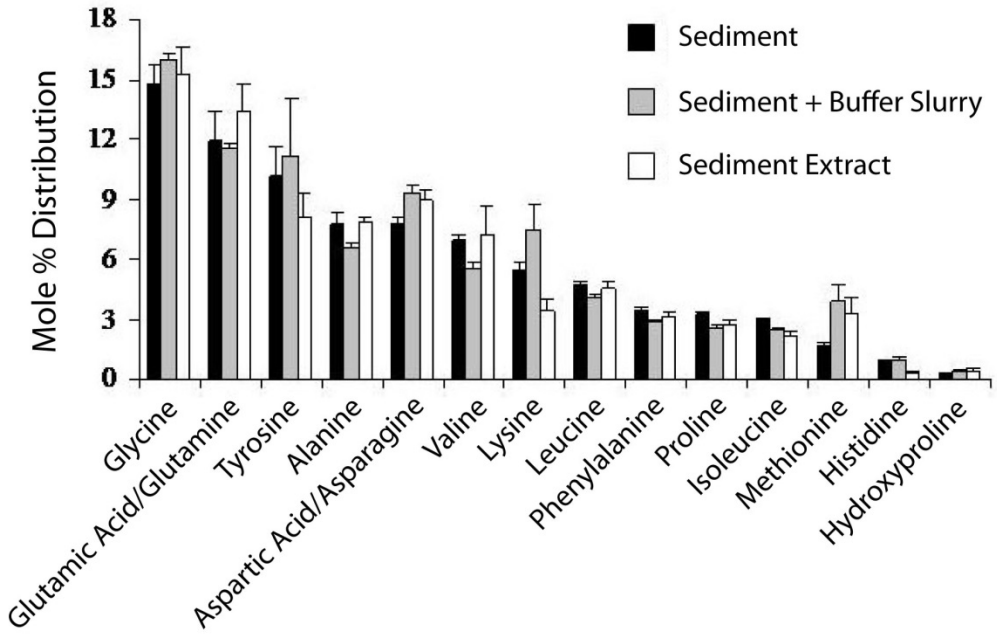


Figure 3.2. The comparative distribution of amino acids observed in Bering Sea shelf sediments using the two extraction approaches verses acid-hydrolysed intact sediments.

3.4.2. 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 this database containing proteomes from one diatom and two marine bacteria, 302 unique peptides were identified from the slurry tube gel method, which equated to 126 protein identifications (Table 3.1, Table A3.1). 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 in each of the four methods resulted in identifications from 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. 3.3A). 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 3.2). The majority of proteins identified were involved in metabolic processes. The biggest difference in the distribution of metabolic proteins from 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 3.1, Table A3.1). There were fewer peptides identified using each extraction method for the GOS search compared to the Thaps search, but the majority of proteins identified from the slurry tube gel method with the GOS search still resulted in 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. 3.3B), and seven proteins were identified with peptide sequences that were identical 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).

Tandem 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 3.1). Fewer peptides and proteins were identified using the NR database compared to 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 with the NR database (Table A3.1). The majority of direct digest identified proteins were bacterial in origin. The NCBI-Refined database search yielded more 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.

Table 3.1: Total proteins identified by extraction method and proteomic database.

Database	Extraction Method	Proteins	Peptides	T. pseudo	T. pseudo+
Thaps	Direct Digest	6	7	6	0
	Slurry Flat Gel	31	82	30	1
	Traditional Tube Gel	60	149	60	0
	Slurry Tube Gel	126	302	122	0
GOS	Direct Digest	4	4	0	0
	Traditional Tube Gel	63	130	37	10
	Slurry Tube Gel	114	257	87	7
NCBI-NR	Direct Digest	16	16	2	0
	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 3.2: 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

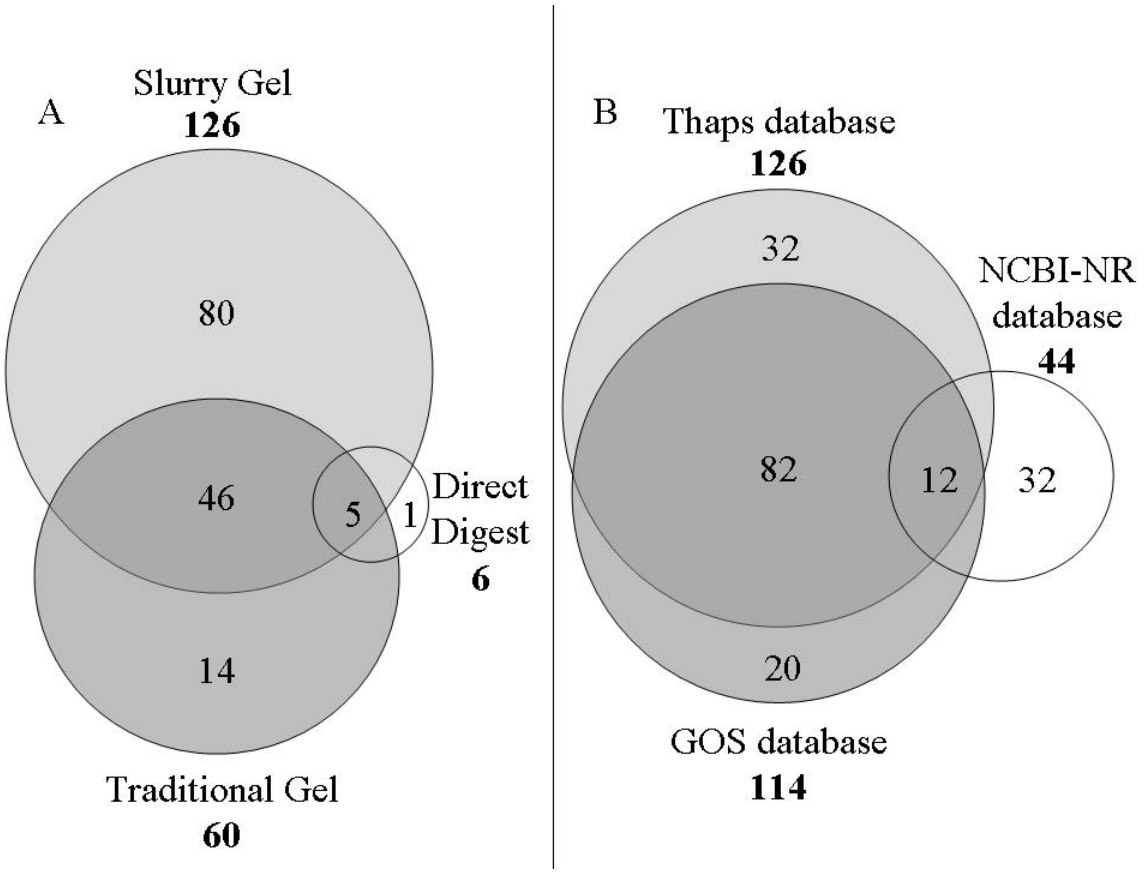


Figure 3.3. Venn diagrams of (A) the number of proteins identified in common between the slurry tube gel, traditional tube gel, and direct digest methods searched against the Thaps database; (B) number of proteins in common between the Thaps, GOS/Thaps, and NCBI-NR database searches of the slurry tube gel method.

3.4.3. Protein recovery optimization from sediments

There was recovery of BSA standard from sediment samples with most extraction methods tested (Table 3.3). The exceptions 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 3.3). 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 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.

Table 3.3. Results of protein search for BSA 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
Slab				
	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

3.5. 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.

Because the protein content of marine sediments may be limited by degradation or masked by interactions with various matrices, it is beneficial to optimize extraction techniques in order to maximize recovery. Chen et al. (2008) demonstrate the applicability of sequence coverage as a useful measure of protein expression. While sequence coverage is not equivalent to extraction efficiency, it is assumed that the greater the sequence coverage, the more peptides recovered from the parent protein, therefore the better extraction method. Though the utilization 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. Using this metric, comprehensive testing of the extraction of added BSA from sediments found that extraction with 6 M urea and the placement of the slurry directly on a flat gel was most effective (Fig. 3.1).

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 thousands of spectra) using the GOS database consumed >720 hours, and the NCBI-NR database >1080 hours, using an 800 CPU cluster. This is a substantial time investment compared to the Thaps database searches of <5 hours. 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 as 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. The reverse sequences are utilized to determine false discovery rates. 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. 3.4). 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 made with the large databases searched against Bering Sea surface sediment extract digests.

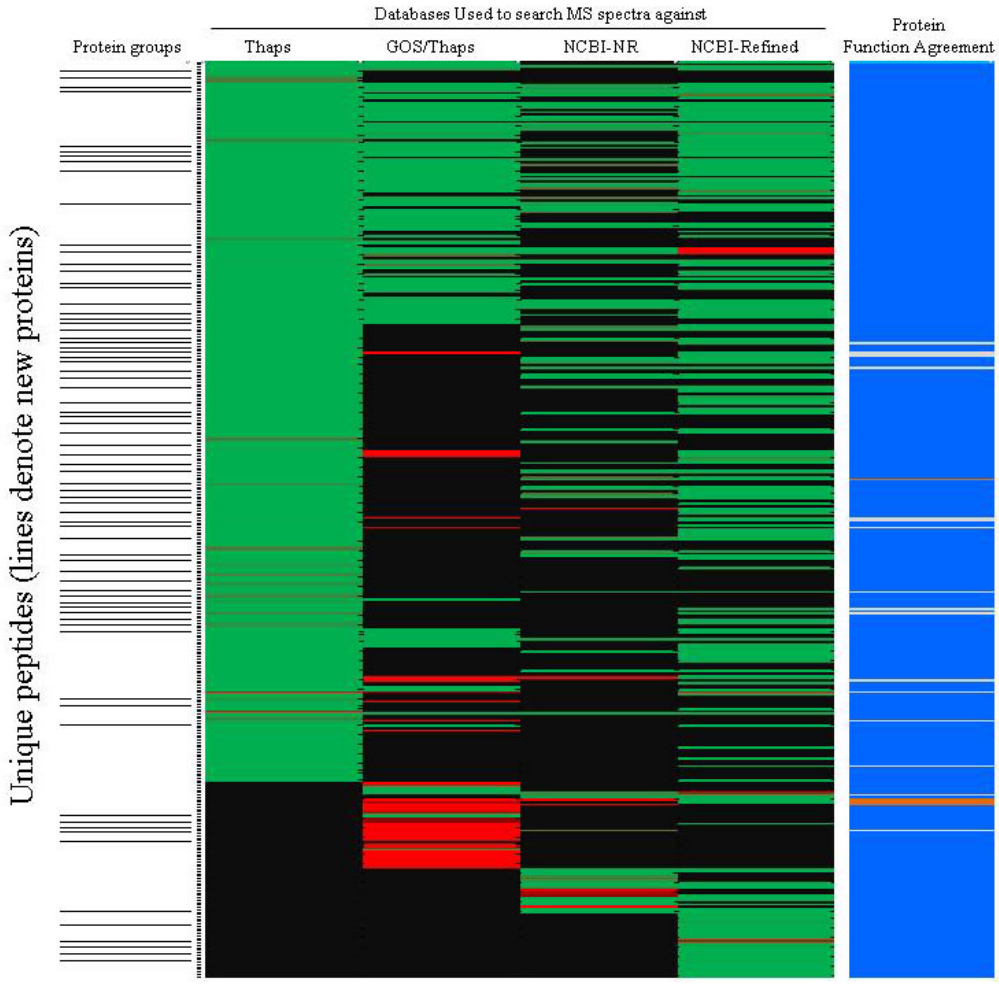


Figure 3.4. Species assignment and protein function assignment comparisons from SEQUEST search of surface sediment slurry extraction of Bering Sea sediment. Mass spectra were obtained from 4 different databases: Thaps (column 1); GOS/Thaps (column 2); NCBI-NR (column 3); NCBI-Refined (column 4). 385 unique peptides are represented along the vertical axis and represented as dots between the first column containing protein groups and the Thaps column. 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.

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 (ΔCorr). Typically, an assignment is made if the xcorr > 2.0 and ΔCorr > 0.1. As an example of a ΔCorr determination: when using a larger database there is more peptide sequence similarity (e.g. SEVSALLGR, SEVSAILGR – two peptides differing by only a leucine and isoleucine residue both of which have the same mass). As a result, PeptideProphet will assign a low ΔCorr to the second best peptide match because it can't determine that it is different from the first match. A low ΔCorr 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. 3.3B). 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, 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.

3.6. Conclusions

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 in order 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 prior to enzymatic digestion and MS analysis. It may be possible to achieve similar results through the use of a dialysis bag with an electric current. Dialysis bags, however, are usually used to separate small molecules (e.g., salts, reducing agents, etc.) from the desired proteins. It is thought that many of the substances interfering with protein extraction are larger molecules (e.g., humics or other organic matter). These larger molecules would remain in the dialysis bag with the proteins allowing for continued interaction after the application of the electrical current. The subsequent digestion may not be as successful as if these larger substances had been separated in the gel.

The determined extraction 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.

3.7. Acknowledgements for this chapter

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.

Chapter 4: Proteomic expression of a heterotrophic marine bacterium

4.1. Abstract

The distribution of total proteins expressed by the heterotrophic marine bacterium, *Ruegeria pomeroyi*, under environmentally realistic carbon conditions during exponential and stationary growth phases were identified and analyzed. Over 2000 proteins produced were identified, with little variation seen among replicates and subtle differences among the expressed proteins between time points. The most abundant proteins identified were responsible for porins, transport, binding, translation, and protein refolding. The total number of observed proteins decreased over time as did the number of proteins common between adjacent time points. Despite such decreases, at least 75% of proteins common among adjacent time points remained. The proteins that were not common between time points were responsible for similar biological processes at each time point. There were few proteins known to be significantly up- or down-regulated and those that were also shared biological functionalities. An examination of the functional annotation of expressed proteins found clusters most enriched within the samples contained proteins that contributed to RNA subunit structure and RNA binding; ribosome assembly and biogenesis; translation and amino acid, amine, organic acid, and oxoacid biosynthesis and metabolism. The biological functions identified among the most abundant proteins and the most prevalent clusters suggest an organism that maintained active growth and cell turnover well into stationary phase with no obvious patterns of changing biological function. The presence of abundant porins and transport proteins relative to the other proteins produced could represent potential proteins to be used as biomarkers of bacterial processes and/or activity.

4.2. Introduction

The global cycles of carbon and nitrogen are essential to life on Earth and a thorough understanding of them is necessary to understand the cycling of nutrients occurring today and to assess the potential impacts of future changes on these cycles. The carbon and nitrogen cycles include both organic and inorganic reservoirs with the organic fraction comprised of multiple biochemical compartments, most broadly encompassing proteins, lipids, and carbohydrates. In marine organisms, proteins can account for the largest fraction of constituent components (Brown, 1991) potentially representing more than 50% of the organic matter and 85% of the organic nitrogen (Tanoue, 1995). Traditionally, proteins have been perceived as labile in the environment and, due to their ease of depolymerization, readily recycled (Hollibaugh and Azam, 1983). On the other hand, the presence of amino acids and high molecular weight preserved fractions (Nguyen and Harvey, 2003) in sediments indicate a larger role of protein or its products, as an important contributor to the poorly characterized fraction of carbon and nitrogen seen in the geological record. In order to fully elucidate the mechanisms at work in these global cycles an improved understanding of the cycling, degradation, and preservation of proteins is needed, specifically an evaluation of the role of bacteria.

The task of examining the role of bacteria in marine environments is a difficult one. The majority of marine bacteria have not been cultured (Schweder et al., 2008), or are not culturable, under laboratory conditions (Stewart, 2012). Although metaproteomic analysis of all the proteins obtained from an environmental sample has great potential, it is limited by the availability of bacterial peptide sequence information (Schweder et al.,

2008). In addition, although the numbers of bacteria in marine samples are high (e.g. 10^4 to 10^6 cells mL^{-1}; Bianchi and Giuliano, 1996), the biomass, and by extension amount of protein, they contribute can be much less than that contributed by other organisms. This can result in small numbers of protein identifications due to the low abundance of bacterial peptides (Moore, Nunn, Goodlett, et al., 2012).

Because of these limitations, initial steps need to be taken to more fully understand those organisms that can be cultured and to improve the chances of protein identification through metaproteomics. With the ultimate goal to determine the extent to which bacteria participate in organic matter recycling and degradation, investigations to more fully understand what proteins marine bacteria produce and utilize under normal and altered environmental conditions should be undertaken. The examination of these expressed proteins will, potentially, lead to the discovery of sets of proteins that are known to be produced under certain conditions by select, most, or all bacteria within laboratory cultures, and these proteins can be selectively searched for in environmental samples.

The goal of this work was to determine the proteins expressed by a heterotrophic marine bacterium under environmentally realistic carbon conditions throughout its growth phases. *Ruegeria pomeroyi* is a wide-spread, generalist, marine bacterium which has been studied in several capacities from the investigation and documentation of its genome, to studies of specific proteins such as those involved in its ability to utilize dimethylsulphoniopropionate (DMSP), to the proteins produced under adverse laboratory conditions (e.g. Christie-Oleza, Pina-Villalonga, et al., 2012; Moran et al., 2004; Reisch

et al., 2013). Our objective was to determine overall proteome expression and document observed patterns in protein expression throughout the growth curve.

4.3. Materials and Methods

4.3.1. R. pomeroyi culture conditions

R. pomeroyi DSS-3 was generously provided by Dr. Mary Ann Moran, University of Georgia on agar plates. Upon arrival it was transferred to a liquid culture of nutrient-rich marine media comprised of 2.5 g tryptone, 4 g yeast extract, and 20 g sea salts L^{-1} of water. Cultures and blanks were established in 20 mL and 100 mL vessels and placed on a laboratory bench top at approximately 23°C. Cultures were transferred five times over a fifteen-day period. On the fifth transfer, the amount of tryptone and yeast in the media was decreased by half. Over the following two months culturing continued and the amount of tryptone and yeast continually decreased until it reached 12.5% of the original media concentration.

Cultures were then transferred to a marine basal media using 10 mM glucose as a carbon source, salinity 28. The marine basal medium was comprised of 250 mL of basal medium (150 mL 1 M Tris HCl pH 7.5, 87 mg K_2HPO_4, 1.5 g NH_4Cl, 375 mL water), 50 mL FeEDTA stock (50 mg FeEDTA, 100 mL water), 699 mL sea salt solution (20 g sea salts, 699 mL water), 1 mL vitamin supplement (2 mg biotin, 2 mg folic acid, 10 mg pyroxidoxine-HCl, 5 mg riboflavin, 5 mg thiamine, 5 mg nicotinic acid, 5 mg pantothenic acid, 0.1 mg cyanocobalamin, 5 mg *p*-aminobenzoic acid). Over a two-month period, transferring approximately every 48-72 hours, the cultures were continued and the

amount of glucose decreased to 0.1 mM and a growth curve at that concentration was determined.

4.3.2. Determination of growth curve

In order to determine a growth curve upon which to base the experimental culture sampling, a culture, grown in marine basal media (0.1 mM glucose), was established and sampled every eight hours for 48 hours. The total volume of each sub-sample was 5 mL and ranged from 1 to 100% culture with concentrations decreasing through time. The samples were killed with a 2% formalin addition and filtered onto black 0.2 μm polycarbonate filters for cell counts. Fluoroshield with DAPI histology mounting medium (Sigma-Aldrich) was used to stain cells for counting (Porter and Fieg, 1980). The cell counts were accomplished by picking random locations on the slide; 15 fields were counted on each slide using an Olympus BX50 microscope at 1250x magnification. Each time point was also run through a spectrophotometer (set at 600 nm) as a secondary measure to compare to the cell counts. The curves for the counts and the spectrophotometer were comparable with each other so the growth curve was established using the cell count data.

4.3.3. Experimental culture conditions

Experimental cultures were grown in marine basal media with 0.1 mM glucose. Three biological replicates were used. A blank was also established and maintained under the same conditions minus the addition of any organism. Cultures and blanks were established and grown in 2 L and 1 L flasks, respectively, on a laboratory bench top at approximately 23°C.

4.3.4. Sampling procedures

Sampling from the bacterial cultures was undertaken at 25 hours, 38 hours, and 44 hours, based on the initial growth curve described in section 4.3.2. These time points were selected because of the correspondence to the early exponential growth phase, late exponential growth phase, and early stationary growth phase, respectively. A fourth sampling point, at 66 hours, was also collected to represent late stationary phase based on the results from the initial growth curve. For time point 1 (T_1), time point 2 (T_2), and time point 3 (T_3), all three biological replicates were sampled and processed as subsequently described for proteomic analysis and protein quantification. Due to limitations imposed by increased bacterial numbers on filtration time, for time point 4 (T_4), only one replicate was sampled. The replicate was split and analyzed three times as described for proteomic analysis.

At T_1, 750 mL was sampled from each culture for proteomic analysis and total protein quantification via BCA assay (500 mL for proteomic analysis and 250 mL for total protein via BCA analysis). 12 mL of a 10% (w/v) solution of trichloroacetic acid

(TCA) was added to the 750 mL sample to precipitate the proteins. The sample was refrigerated at 4°C for one hour prior to filtering through a 0.2 μm polycarbonate filter. At T_2, 375 mL was sampled for proteomic analysis and protein quantification (250 mL for proteomic analysis and 125 mL for total protein via BCA analysis). 10 mL of a 10% (w/v) solution of TCA was added to the sample. At T_3, 150 mL of culture was sampled (100 mL for proteomic analysis and 50 mL for total protein via BCA analysis) and 2 mL of a 10% (w/v) solution of TCA was added. At T_4, 150 mL of culture from one replicate was sampled (100 mL for proteomic analysis and 50 mL for total protein via BCA analysis) and 2 mL of a 10% (w/v) solution of TCA was added. Refrigeration and filtering for T_2, T_3, and T_4 is as described for T_1. Blanks were sampled at T_2 and T_3. Blanks were not sampled at T_1 or T_4 but cultures showed no visible changes when compared to the other time points. Blanks were processed as described above for samples.

4.3.5. Protein extraction, BCA quantification, and protein digestion

A total of 300 μL of 2 M urea was added to filters for both the proteomic analysis and the bicinchoninic acid (BCA) assay, the filters were covered with deionized water and sonicated for one minute and placed on ice. This procedure was repeated two times and the three extracts were combined and dried in a speedvac until only a few μL remained. To bring the urea concentration down to 6 M, 300 μL of DI water was added to the extract.

Total protein concentrations were measured on each sample using a Pierce BCA assay kit. Manufacturer's instructions were followed for the preparation and analysis of

standards and samples. Briefly, dilution standards were created using an albumin (BSA) standard ranging from 0.5 µg/mL to 200 µg/mL. In a test tube, 1 mL of standard or sample was combined with 1 mL of working reagent and the test tube was incubated in a water bath for 60 minutes at 60°C. The sample was then read in a spectrophotometer at 562 nm. Averaged blank-corrected readings for the standards were used to create a standard curve against which the unknown sample concentrations were determined.

For protein digestion in preparation of proteomic analysis, 19.8 µL of 1.5M Tris buffer was added (pH 8.8) along with 7.5 µL of 200mM TCEP and the samples were incubated for one hour at 37°C. Following incubation, 60 µL of 200 mM iodoacetamide (IAM) was added to alkylate cysteine residues, samples were vortexed, and allowed to sit for one hour in the dark. To neutralize the excess IAM and to reduce the disulfide bonds, 60 µL of 200mM dithiothreitol (DTT) was added and the samples were allowed to sit for one hour. Following this reduction, 2.4 mL of 25 mM ammonium bicarbonate (NH_4HCO_3) was added to dilute the urea after which 600 µL of methanol was added to each sample. Each sample next received trypsin at an enzyme:protein ratio of 1:50, was vortexed, and was incubated overnight at 37°C. Samples were desalted using a macro-spin C18 column (NestGroup). The columns were saturated with a solvent mixture of 80% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) and equilibrated with a solvent mixture of 5% ACN, 0.1% TFA. The sample was added to the column and the column was washed with 5% ACN, 0.1% TFA. The sample was eluted with 80% ACN, 0.1% TFA. The sample was dried to near dryness in a speedvac and brought up to a volume yielding approximately 1 µg protein/2 µL with 5% ACN, 0.1% formic acid (FA) for analysis by mass spectrometry (MS).

4.3.6. Mass spectrometry

Digested peptides within the samples were separated prior to introduction into the mass spectrometer through reverse-phase chromatography which utilized an analytical column (20 cm long, 75 μm id fused silica capillary column packed with C18 particles (Magic C18AQ, 100 Å, 5 μm; Michrom, Bioresources) preceded by a pre-column (2 cm long, 100 μm (Magic C18AQ, 200 Å, 5 μm; Michrom). Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-acetonitrile gradient (5 to 35% acetonitrile in 90 min). Tandem mass spectrometry (LC-MS/MS) of the peptides was performed on a QExactive mass-spectrometer (Thermo Fisher) and on a LTQ-Orbitrap hybrid mass-spectrometer (Thermo Fisher) for the blanks.

4.3.7. Database Search

Mass spectral results were interpreted and searched with SEQUEST (v. 2012.01.0) (Eng et al., 2008) to match observed spectra to theoretical spectra generated from the predicted peptide sequences determined by the search database chosen. All CID spectra were searched against the most recent version of the *R. pomeroyi* proteome (ASM1196v2; ncbi.nlm.nih.gov/genome; 4,278 proteins) and 50 common protein contaminants in the lab (e.g. keratins, trypsin). The searches were conducted with enzyme specificity for trypsin and a +6 Da modification specific to leucine residues. The alteration to the leucine was searched because these samples were run on the MS at the same time as another study using a ^{13}C-leucine label and the potential for carry-over between samples needed to be taken into consideration. Match probability between the

observed and theoretical spectra was set at 90% ($p<0.1$) on ProteinProphet and PetptideProphet (Keller et al., 2002). Proteins with only one identified peptide were excluded.

4.3.8. DAVID

The database for annotation, visualization and integrated discovery (DAVID) bioinformatics resource (http://david.abcc.ncifcrf.gov/home.jsp) was used to enable an interpretation of the biological importance of the proteins identified in this study as well as an analysis of the enrichment of those proteins within the samples. This resource was used to evaluate the relative presence of types of proteins within a sample compared to what the expected presence of those proteins would be in the normal proteome (Huang et al., 2009). Specifically, proteins are determined to be enriched if they are present to a greater extent (higher ratio) within the sample than would be expected by random chance when compared to the ratio of that protein to the entire genome. Thus DAVID can provide information on the apparent enrichment of types of proteins relative to what would be expected for the organism given its genome background.

The lists of total observed proteins within each time point were submitted to DAVID and the results were analyzed using its Functional Annotation Clustering tool. This analysis provides results as clusters of similar protein annotations. These annotations provide a description of a protein's biological function. The data reported includes the total number of proteins in the sample associated with each annotation within the cluster (count) as well as a group enrichment score for the cluster. The group enrichment score is the geometric mean (in –log scale) of the p-values of the annotations

included within the cluster ("Functional Annotation Tool"). DAVID determines the p-value of an annotation by examining the frequency of the expression of the proteins involved in the annotation within the sample compared to those proteins' frequencies within the entire genome. Lower p-values for annotations are associated with higher likelihood that the biological function or process is enriched within the sample and not occurring as a matter of chance. Since the group enrichment term is calculated from the individual p-values within the cluster, those clusters with the smaller p-values will be more enriched within the sample, i.e. higher enrichment scores ("Functional Annotation Tool"). In general, $p < 0.01$ is used to determine significant enrichment. For the analysis of the total proteins seen, however, an even stricter threshold was used of $p < 0.001$.

4.3.9. Label-free protein quantification (Qspec)

For protein quantification, spectral counts of proteins in all replicates were used. Spectral counting tallies the number of spectra seen for peptides from a specific protein. These counts are used to estimate relative abundance of the protein within the sample with the accepted assumption that the more abundant a protein is within a sample the more peptides it would generate upon digestion and hence the more spectra that would be detected from those peptides (Liu et al., 2004). Qspec is an application created to determine differences in protein populations using spectral count data (Choi et al., 2008) and was used in this study to determine significantly up- or down-regulated proteins among the time points. In this study, a new version of Qspec (QPROT 1.2.2, qspec-param command) was used as opposed to the original version discussed in Chapter 6. Qspec still takes into account the size of a protein and normalizes total spectral counts if

needed. Consideration of the length of a protein is necessary because longer proteins could potentially produce more identifiable peptides thereby elevating their spectral counts and artificially increasing their relative abundance.

This new implementation uses Z-statistics, which is computed as the natural log fold change, divided by its standard error, and determination of the regulation of the protein is made by comparing the false discovery rate (FDRup or FDRdown) to your statistical threshold (i.e. an FDRup = 0.003 would indicate a significantly up-regulated protein because this value is less than p=0.05) (H. Choi, personal communication, October 11, 2013). Results for this study are reported as proteins that were significantly up- or down-regulated in the second time point of a stated comparison and these proteins were considered to be significantly altered if the FDRup or FDRdown was $p < 0.05$.

4.3.10. Interactive Pathways Explorer (iPath v2)

In order to visualize the biological employment of the identified proteins, iPath software (http://pathways.embl.de/; Yamada et al., 2011) was used to display the metabolic pathways functioning within *R. pomeroyi*. This software uses various identifiers to categorize the observed proteins by pathway. Enzyme commission numbers (EC numbers) and UniProt identifiers (http://www.uniprot.org/) for the observed proteins in this investigation were entered into the iPath software.

4.4. Results

4.4.1. R. pomeroyi growth curve and protein quantifications

An initial growth curve was established for *R. pomeroyi* by cell count and was confirmed by spectrophotometric analysis at an optical density of 600 nm (OD_{600}). Table 4.1 and Figure 4.1 show the cell counts and OD_{600} readings and the determined growth curve for *R. pomeroyi* under the growth conditions described previously.

Table 4.1. Initial cell counts and optical densities (at 600nm) for *R. pomeroyi* cultures.

Time (hr)	Count	OD_{600}
0	1.3×10^4	0.003
8	2.5×10^4	0.002
16	2.2×10^5	0.002
24	2×10^6	0.004
32	6×10^6	0.016
40	2×10^7	0.017
48	1.5×10^7	0.019

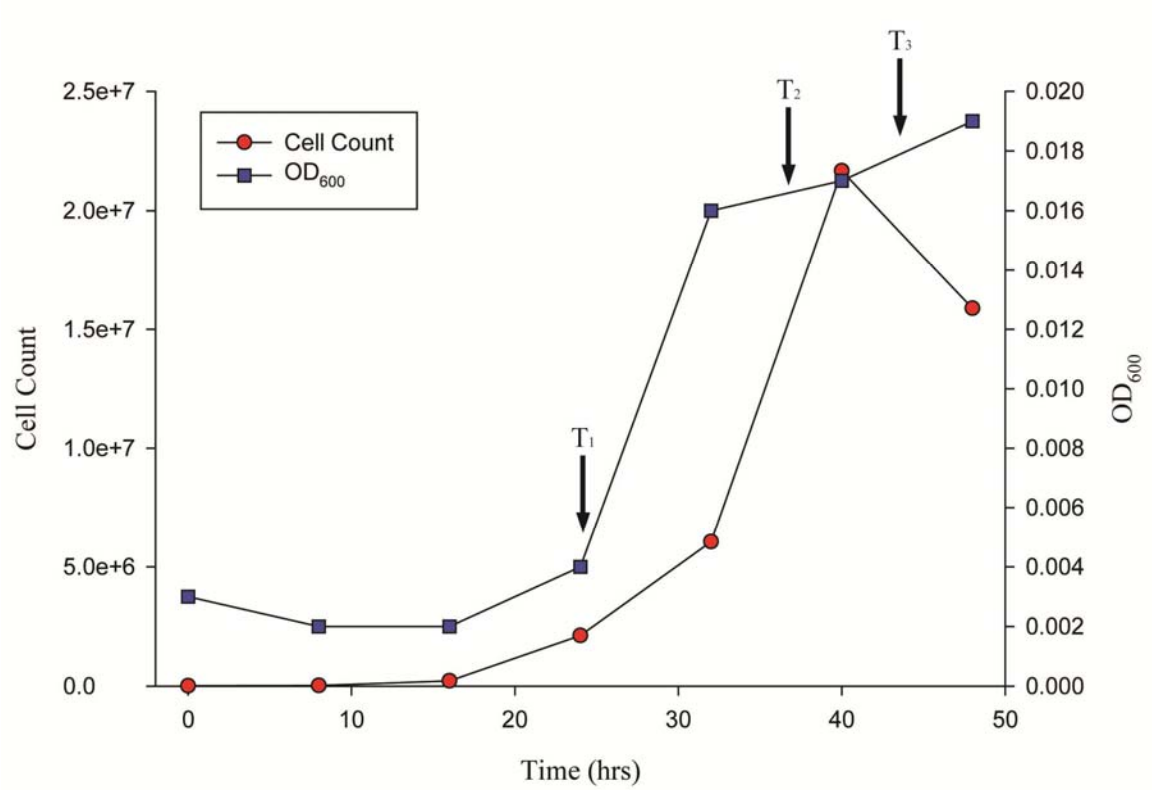


Figure 4.1. Initial growth curve determined for *R. pomeroyi* upon which the experimental design for the culturing was based. For the experiment T_4 was collected at 66 hours but cell counts were not determined.

4.4.2. Observed proteins

In total, 2037 total proteins were observed among all the replicates across all time points. This equated to the expression of approximately 48% of the theoretical proteome. Of those total proteins, 774 were seen in all replicates at all-time points.

A Cochran Q test was performed on the replicates at each time point to determine whether the identified proteins were significantly different. This analysis was based on the presence or absence of a protein in a replicate. This analysis showed that the replicates T_1, T_2, and T_3 were statistically different ($p<0.05$). The results of the three analyses of the T_4 sample were statistically indistinguishable. This was to be expected

because, for that time point, one sample was run three times through the MS as was explained previously. Though the replicates for the first three time points were statistically different, there was a similarity of 80% or greater among the proteins identified for replicates of each time point.

The number of observed proteins decreased through time (Table 4.2; Fig. 4.2) as the culture progressed from early exponential growth phase to late stationary phase. Of the total 2037 proteins identified across all time points, more than half, 1164, were identified at all four time points (Fig. 4.3), though not necessarily in every replicates of each time point. T_1 had the most unique proteins (200), but in contrast T_3 and T_4 had only 11 and 12 unique proteins, respectively (Fig. 4.3). The number of shared proteins between consecutive time points decreased as time progressed from 1581 (T_1 to T_2) to 1326 shared proteins (T_2 to T_3) to 1201 (T_3 to T_4) (Fig. 4.3). These shared proteins account for more than 75% of the proteins at a time point depending on the comparison.

Protein analysis was undertaken on the experimental blanks at T_2 and T_3. Only two *R. pomeroyi* proteins were identified at T_2 and only three were identified at T_3 (Table 4.2). This equates to 0.1% and 0.2% of the proteins identified in the samples at those time points.

Table 4.2. Sum of *R. pomeroyi* proteins observed in all replicates of a time point. * Proteins were defined as having match probability of 90% or greater by ProteinProphet AND PetptideProphet and having 2 or more independent spectra identified for the protein.

Time Point	Time (hrs)	Growth State	Proteins observed
T_1	25	Early exponential	1818
T_2	38	Late exponential	1774
T_2 (Blank)	38	Late exponential	2
T_3	44	Early stationary	1363
T_3 (Blank)	44	Early stationary	3
T_4	66	Late stationary	1352

* The lists of proteins found in the three replicates for each time point were combined and duplicate, or over-lapping, proteins were deleted to compile a list of all the proteins seen within a time point.

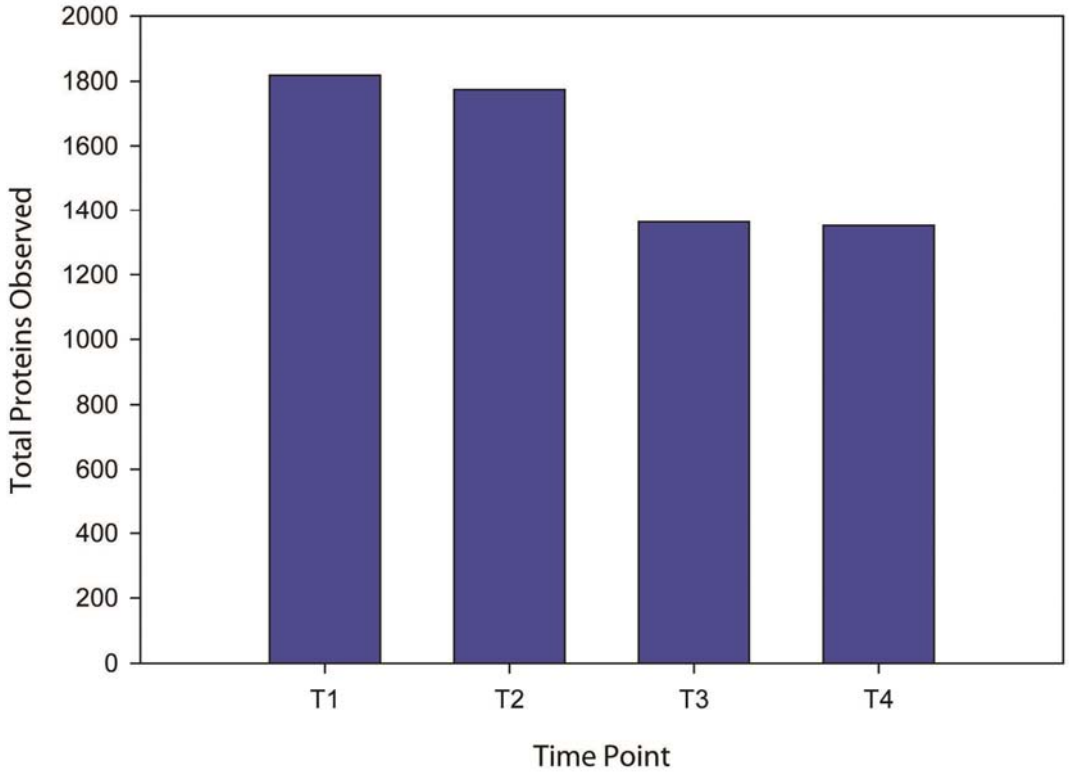


Figure 4.2. Sum of *R. pomeroyi* proteins observed in all replicates of a time point. The lists of proteins found in the three replicates for each time point were combined and duplicate, or over-lapping, proteins were deleted to compile a list of all the proteins seen at each time point. Proteins were defined as having match probability of 90% or greater by ProteinProphet and PetptideProphet and having 2 or more independent spectra identified for the protein.

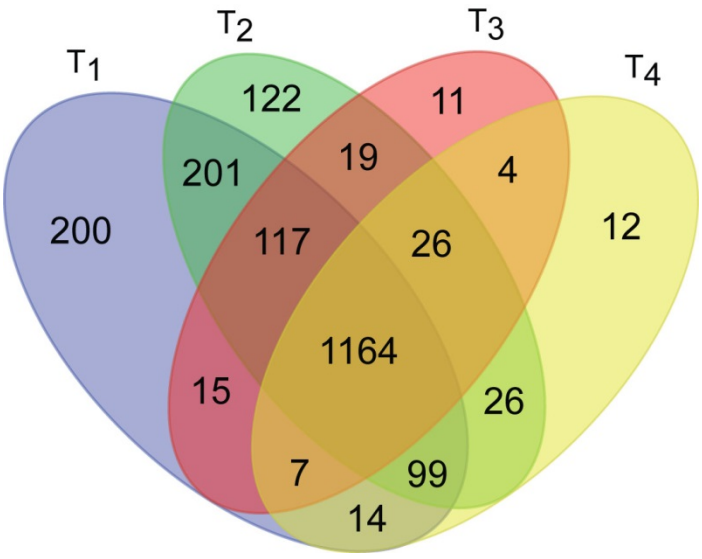


Figure 4.3. Comparison of the total proteins observed among the four time points. The Venn diagram was created using the software available on http://bioinformatics.psb.ugent.be/webtools/Venn/.

4.4.3. Protein abundance and definition

Protein abundance is defined here through the semi-quantitative use of spectral counts. Proteins were ranked according to the summed number of spectral counts in each replicate for a time point. The top 1% of proteins for each replicate were chosen. The lists of proteins from the three replicates of each time point were combined and duplicate proteins eliminated to give cumulative lists of 13 to 18 proteins. The lists from all time points were then combined to produce one final list of the 23 most abundant proteins throughout the experiment (Table 4.3). The total number of spectral counts for each protein was then summed across replicates for each time point to give a final total for each protein at each time point (Fig. 4.4). It should be noted that the size of the individual proteins was not taken into consideration for this analysis so the spectral count data displayed here has not been normalized to protein length. As can be seen in Figure

4.4, T_1 and T_2 have higher counts than T_3 and T_4 for all but two of the most abundant proteins. The five most abundant proteins have functions relating to porins, transport and binding activity, translation, and ATP binding and protein refolding (Table 4.3; Fig. 4.4).

Table 4.3. Most abundant (1%) proteins identified ranked by total number of spectral counts among all replicates in all time points.*

Protein (UniProt ID†)	T1			T2			T3			T4		
	Replicate A	Replicate B	Replicate C	Replicate A	Replicate B	Replicate C	Replicate A	Replicate B	Replicate C	Replicate A	Replicate B	Replicate C
Q5LMY2	856	908	1096	1012	846	904	639	850	574	643	648	580
Q5LV41	593	737	769	648	292	337	180	321	252	162	151	153
Q5LW24	398	737	773	799	402	323	161	269	147	180	192	183
Q5LMR5	449	524	498	497	297	332	222	328	382	272	285	296
Q5LV15	399	387	366	264	155	207	109	188	151	126	125	131
Q5LMQ5	241	226	256	208	154	166	140	173	145	148	145	141
Q5LS31	171	305	271	265	135	168	73	146	143	77	79	87
Q5LPN4	180	241	228	199	157	161	90	181	131	103	99	101
Q5LMQ6	187	204	220	204	127	135	118	142	127	127	127	123
Q5LNP1	150	234	212	178	146	135	93	129	107	114	119	108
Q5LRZ5	184	280	348	284	149	122	50	90	38	62	54	56
Q5LMR4	196	200	228	176	81	97	86	105	108	88	95	90
Q5LN23	154	160	169	159	113	109	76	107	91	105	102	101
Q5LMT9	124	218	212	175	85	111	67	118	96	77	82	73
Q5LQB9	144	173	182	193	72	83	64	89	56	70	69	69
Q5LU87	138	114	122	133	91	90	78	107	72	104	106	96
Q5LV02	126	112	130	128	97	102	81	103	80	96	94	99
Q5LSM4	129	100	104	123	99	115	66	104	73	119	103	105
Q5LX13	133	234	238	96	119	117	36	45	50	37	34	40
Q5LVA0	96	177	186	94	48	55	40	62	43	35	40	42
Q5LT95	54	142	73	75	85	52	49	124	48	55	55	56
Q5LRJ2	15	46	26	8	144	158	26	88	41	32	30	30
Q5LNH0	21	27	81	159	20	8	7	14	8	17	11	16

* Ranking was determined as follows: the proteins with the top 1% of independent spectral counts were obtained for each replicate. The lists of the replicates were combined and duplicates eliminated to produce a total list for the time point. This resulted in lists consisting of 13 to 18 proteins. The combined lists from the time points were compiled and duplicates removed to give the list above. This equates to 1.2 to 1.6% of the total proteins depending on the replicate; † Q5LMY2: porin activity, Q5LV41: D-xylose transport; monosaccharide binding, Q5LW24: transporter activity, Q5LMR5: GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity, Q5LV15: ATP binding; cytoplasm; protein refolding, Q5LMQ5: DNA binding; DNA-directed RNA polymerase activity; ribonucleoside binding; transcription, DNA-dependent, Q5LS31: amino acid binding, Q5LPN4:transporter activity, Q5LMQ6: DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-dependent, Q5LNP1: ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1), Q5LRZ5: cytoplasm; translation elongation factor activity, Q5LMR4: GTP binding; GTP catabolic process; GTPase

activity; cytoplasm; translation elongation factor activity, Q5LN23: 3'-5'-exoribonuclease activity; RNA binding; RNA processing; cytoplasm; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity, Q5LMT9: No definition, Q5LQB9: outer membrane-bounded periplasmic space; transport, Q5LU87: ATP binding; DNA binding; biotin carboxylase activity; gluconeogenesis; metal ion binding; pyruvate carboxylase activity, Q5LV02: ATP binding; adenylylsulfate kinase activity; sulfate adenylyltransferase (ATP) activity; sulfate assimilation, Q5LSM4: cellular amino acid metabolic process; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor, Q5LX13: calcium ion binding, Q5LVA0: RNA binding; ribosome; structural constituent of ribosome; translation, Q5LT95: carbon-monoxide dehydrogenase (acceptor) activity, Q5LRJ2: calcium ion binding, Q5LNH0: ATP hydrolysis coupled proton transport; ATP synthesis coupled proton transport; hydrogen ion transmembrane transporter activity; integral to membrane; lipid binding; plasma membrane; proton-transporting ATP synthase complex, coupling factor F(o).

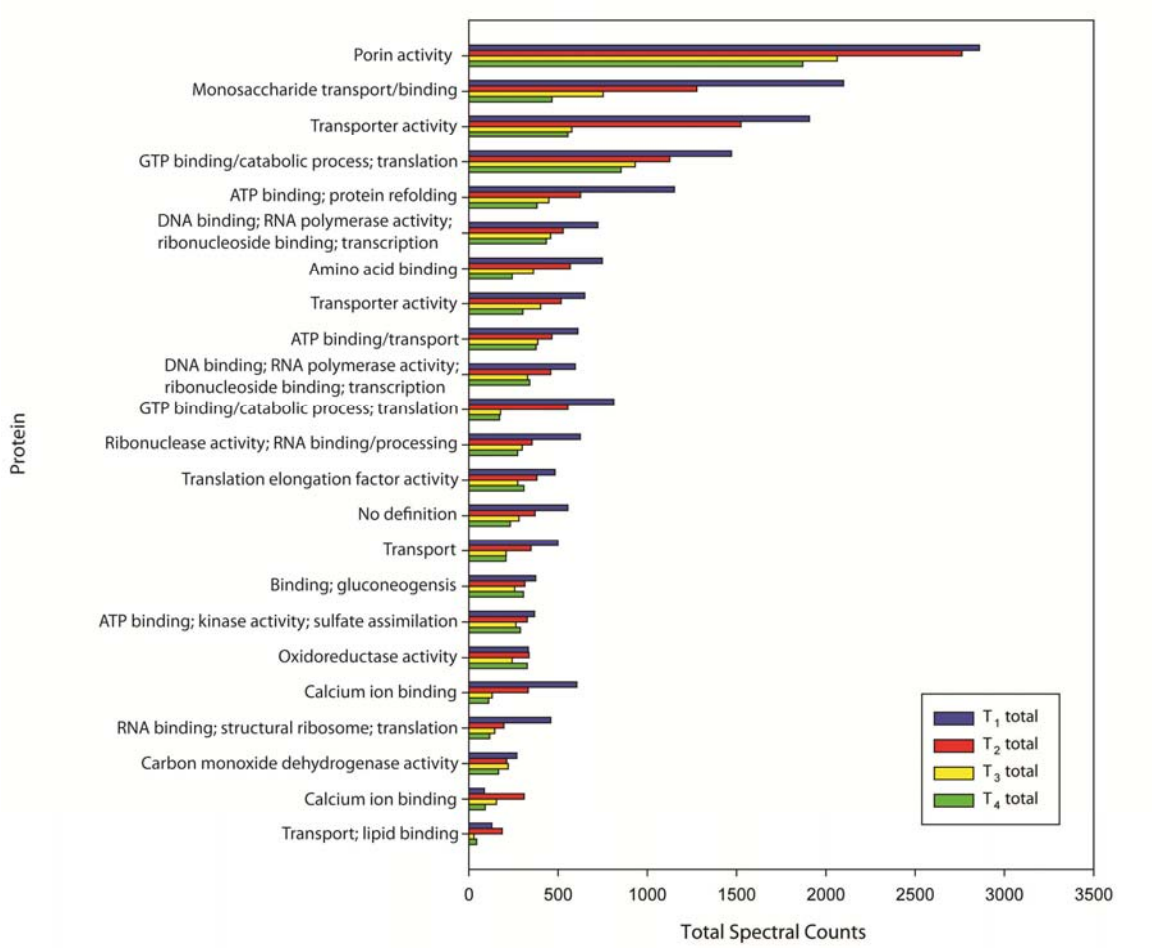


Figure 4.4. Most abundant proteins as defined by total spectral counts for each time point from all replicates.

4.4.4. Significantly up- and down-regulated proteins

The individual proteins present were compared by time point using Qspec to determine those proteins that might be significantly up- or down-regulated. Spectral counts, tallied from all three replicates, for each protein identified in T_1 were compared to those in T_2, T_3, and T_4. If a protein was significantly different in the second time point compared to T_1, either due to an increase or decrease in spectral counts, it is referred to as up- or down-regulated, respectively.

There were a total of 134 proteins significantly up- or down-regulated across the four time points (Table 4.4). There were 21 proteins up-regulated in T_2 that remained up-regulated in T_3 and T_4. There were 22 proteins down-regulated in T_2 that remained down-regulated in T_3 and T_4. There were 91 proteins whose regulation varied between time points, meaning that they were up- or down-regulated at certain time points and not others (Fig. 4.5). When comparing the biological functions of the proteins remaining up- and down-regulated, there were no functions altered that indicated a change in cellular physiology. Table A4.1 provides a complete list of up- and down-regulated proteins and their biological functions.

Table 4.4. Up- and down-regulated proteins. The latter time point is what is being referred to as up- or down-regulated. *

Time point comparison	Time (hours)	# of up-regulated proteins	# of down-regulated proteins
T_1 v T_2	25 v 38	36	40
T_1 v T_3	25 v 44	53	33
T_1 v T_4	25 v 66	40	49

* Proteins were considered to be up- or down-regulated if the corresponding false discovery rate (FDRup or FDRdown) was $p < 0.05$.

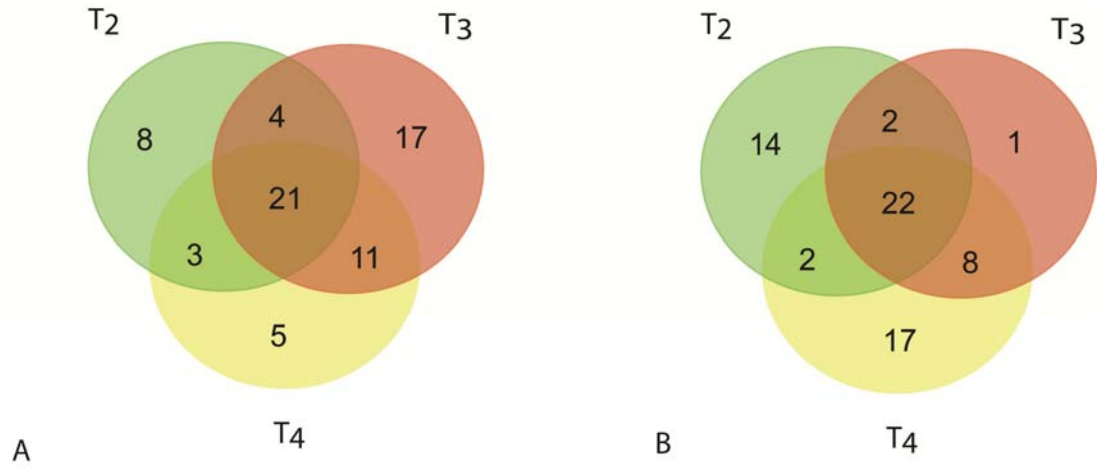


Figure 4.5. Distribution of up- and down-regulated proteins. (A) Proteins within each time point that were determined to be up-regulated compared to T1. (B) Proteins within each time point that were determined to be down-regulated compared to T1. Proteins were considered to be significantly up- or down-regulated if their FDRup or FDRdown was $p<0.05$. The Venn diagram was created using the software available on http://bioinformatics.psb.ugent.be/webtools/Venn/.

4.4.5. DAVID functional annotation results

The Functional Annotation Clustering tool within the DAVID bioinformatics resource was used to evaluate the biological functions and enrichment of proteins identified within this study. There were five prominent annotation clusters identified by DAVID (hereafter clusters 1-5). All four time points contributed proteins to these five clusters. There were an additional three clusters (clusters 6-8) that had proteins in at least 1 time point but not all (Table 4.5; Fig. 4.6 and 4.7). When evaluating the clusters based on protein count (Fig. 4.6), four of the five clusters containing proteins from all time points (clusters 1-3 and 5) were similar in the number of proteins in the cluster at each time point. One cluster, (cluster 4), however, showed a marked difference in the number of proteins contributing to that cluster from the four time points (Fig. 4.6). There were considerably more proteins contributed to this cluster at T_1 and T_2 than at T_3 and T_4. Although cluster 4 had the highest number of proteins (by count) in T_1 and T_2, when evaluated by enrichment score, its values were much more similar among the time points and it was not the cluster that was most enriched in the samples (Fig. 4.7). This seems to indicate that while there were many proteins involved in binding that were expressed, their relative presence in the samples compared to what would be expected from the entire proteome was not as great as the other clusters. When viewed by enrichment score, clusters 1-5 each have similar values among the time points and clusters 1 and 2 are more enriched in these samples (Fig. 4.7). Enrichment scores of clusters 6-8 were not displayed in the figure because not all time points had enriched proteins in those clusters.

In general, the 913 proteins identified by DAVID as being enriched within the samples showed a general biological function in replication and growth. Specifically,

they reflected proteins involved in binding, the biosynthesis and metabolism of amino acids, the functioning of RNA, ribosomes, and translation as well as the biosynthesis and metabolism of nucleotides and ribonucleotides in order of decreasing protein counts per cluster. In addition, they have functionalities that contribute to binding, ligase activity, and oxidoreductase activity (Table 4.5; Fig. 4.7). Table A4.2 gives a complete list of the annotation categories within the clusters and the protein counts associated with each at all four time points.

Table 4.5. Functional annotation descriptions for clusters identified by DAVID.

Cluster	Description
1	Proteins involved in RNA subunit structure and RNA binding; ribosome assembly and biogenesis; translation
2	Proteins involved in amino acid biosynthesis and metabolism; amine biosynthesis and metabolism; organic acids biosynthesis; and oxoacid metabolism
3	Proteins involved in purine, nucleotide and ribonucleotide biosynthesis and metabolism
4	Proteins involved in binding (ATP, nucleoside, ribonucleotide)
5	Proteins involved in tRNA aminoacylation; amino acid activation; RNA metabolism (all these involved in protein synthesis); ligase activity
6	Proteins involved in coenzyme and cofactor metabolism
7	GTP binding; guanyl nucleotide and ribonucleotide binding
8	Oxidoreductase activity using NAD(P)H

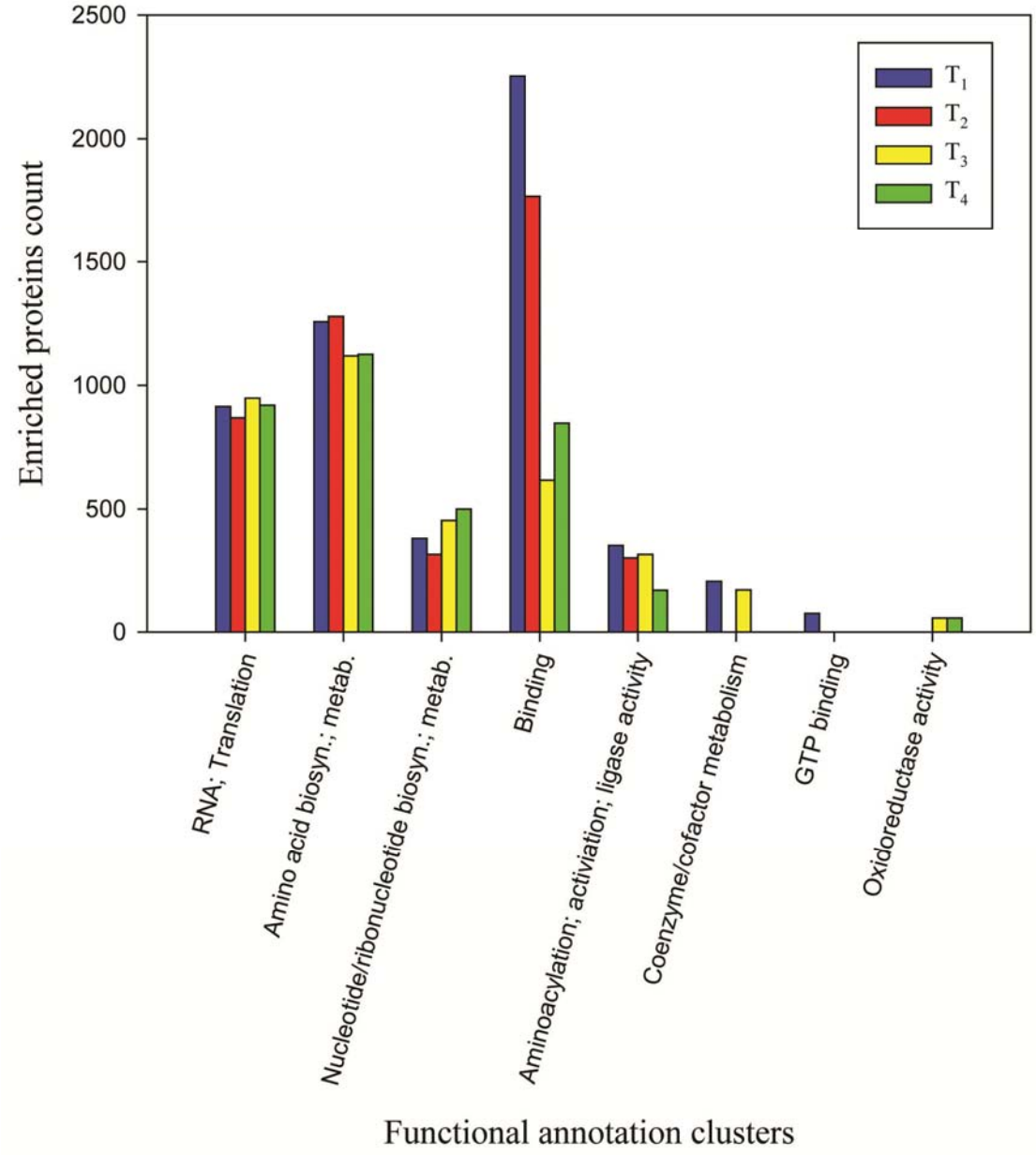


Figure 4.6. Sum of the proteins enriched within functional annotation clusters via DAVID analysis. Proteins are categorized as enriched if they are present to a greater extent than would be expected by chance when compared to the protein's ratio to the entire genome. DAVID groups annotations with common proteins together to create clusters of functional annotations.

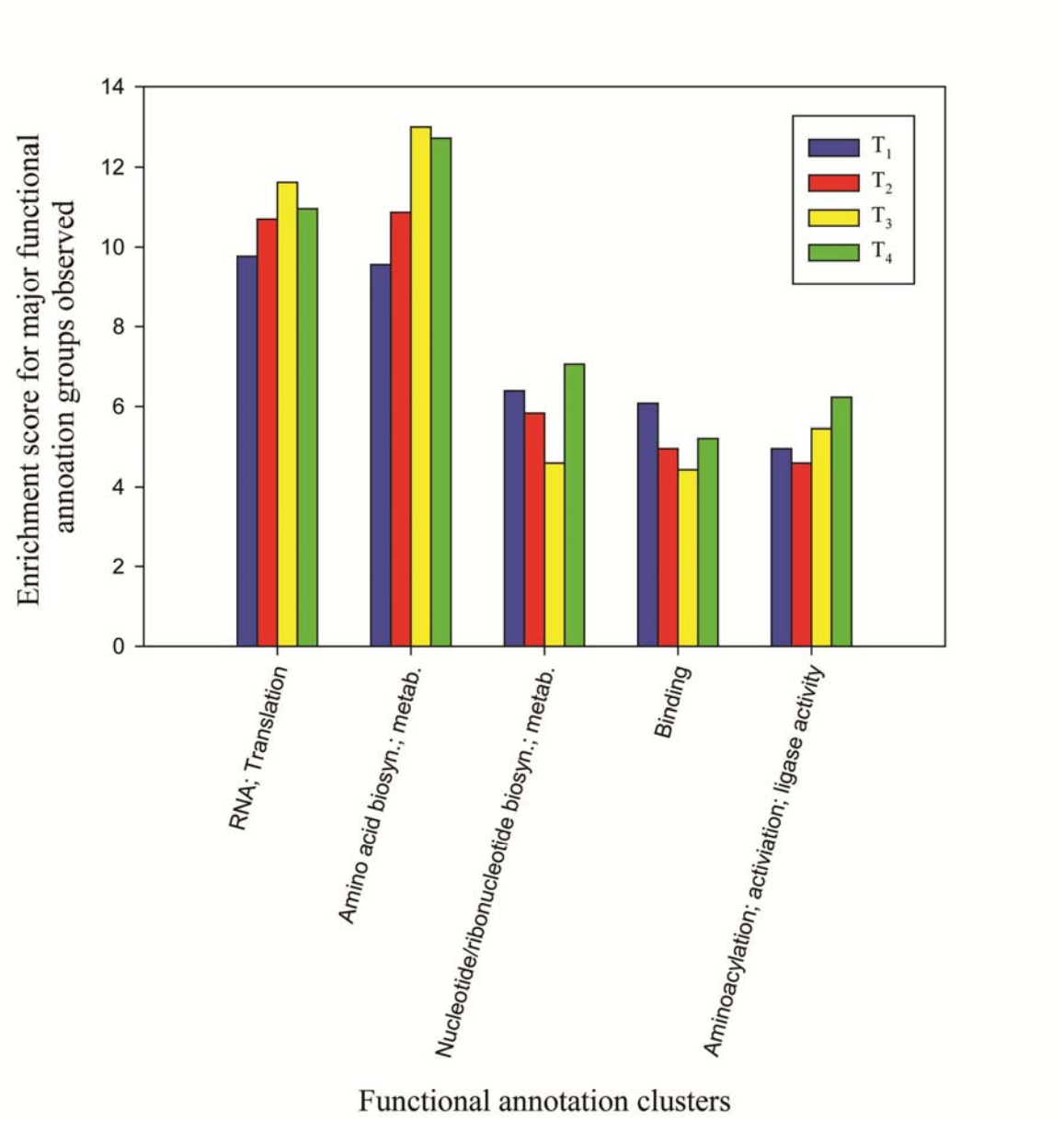


Figure 4.7. Enrichment score for the major functional annotation groups observed within the experiment. DAVID groups annotations with common proteins together to create clusters of functional annotations. Protein lists are determined to be enriched if the number of proteins representing a certain annotation are present in a greater ratio than would be expected by chance when compared to the expression of that annotation within the entire genome. A p-value is determined for each annotation. The enrichment score is the geometric mean, in a –log scale, of the cluster's members' p-values. The more significant the p-value, the less likely it is that the proteins observed were present to that extent by chance and the more likely that the annotation within which they fall is enriched within the biological sample. The lower the p-values of the group, the greater the enrichment score. Enrichment score provides a means to rank biological significance of multiple clusters.

4.4.6. iPath results

Figure 4.8 displays the metabolic pathway map of the shared proteins at all four time points. The pathways displayed are described by the coordinating colored boxes. The iPath software displays the chemical compounds, or metabolites, of a pathway as nodes, or corners, and the enzymes, or proteins, as the lines. This investigation did not study the metabolites. Of the 1164 total proteins entered, approximately 26%, or 308 proteins, were mapped to at least one pathway within iPath. It can be seen in Figure 4.8 that at least some proteins from all the pathways were identified in this investigation.

Figures 4.9, 4.10 and 4.11 display the pathways generated by the proteins unique to growth at 25, 38, and 66 hours, respectively. These proteins are found in only one time point and no others. The proteins for growth at 44 hours were also submitted to iPath but those proteins were not matched to any of the elements on the pathway map.

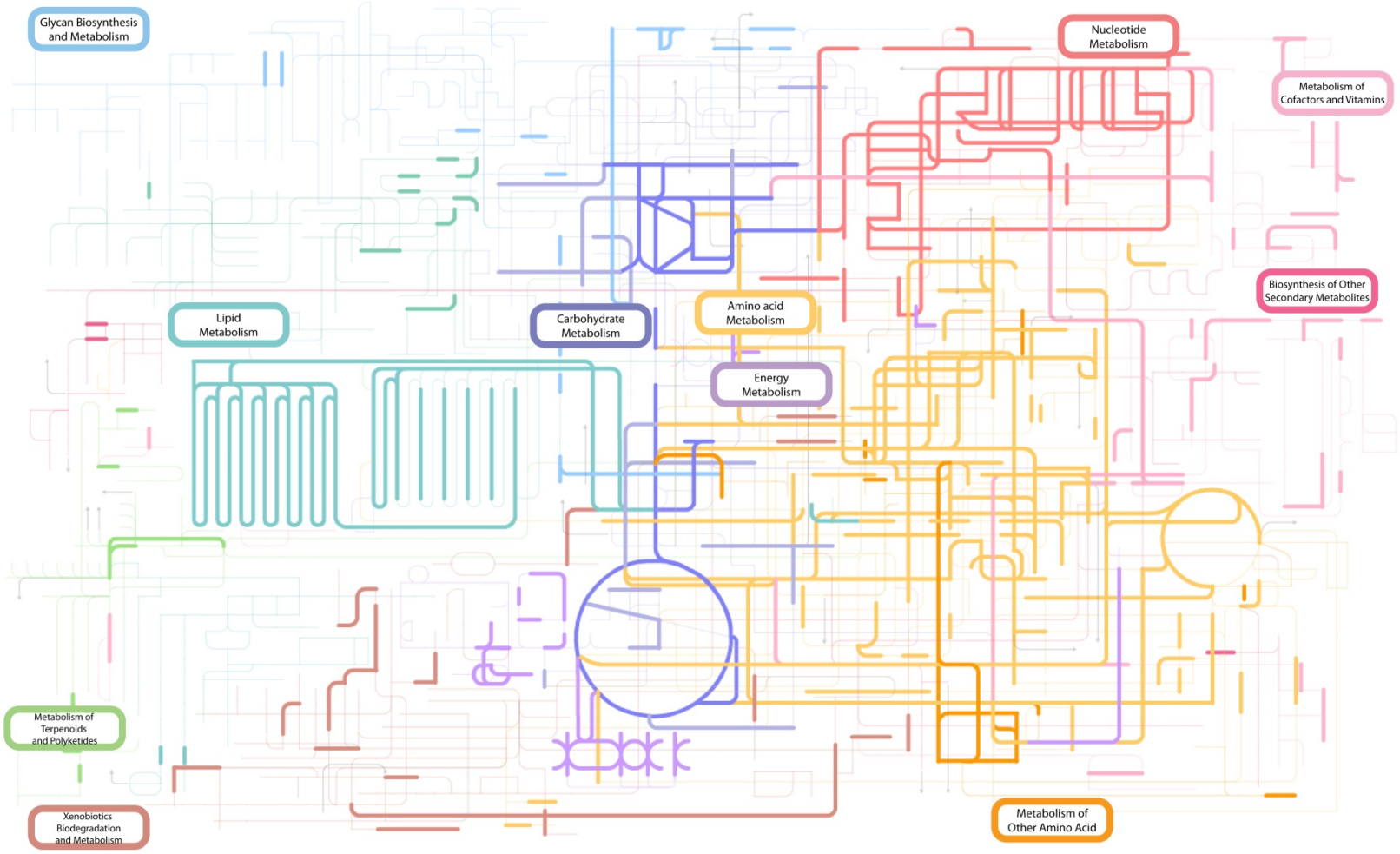


Figure 4.8. Metabolic biochemistry map of shared proteins identified in all time points for *R. pomeroyi*. The map represents both metabolites and proteins. The metabolites are displayed as nodes, or corners, and the proteins are represented by the lines within the diagram. This study did not measure metabolites. The pathways displayed in various colors are described in the corresponding colored boxes.

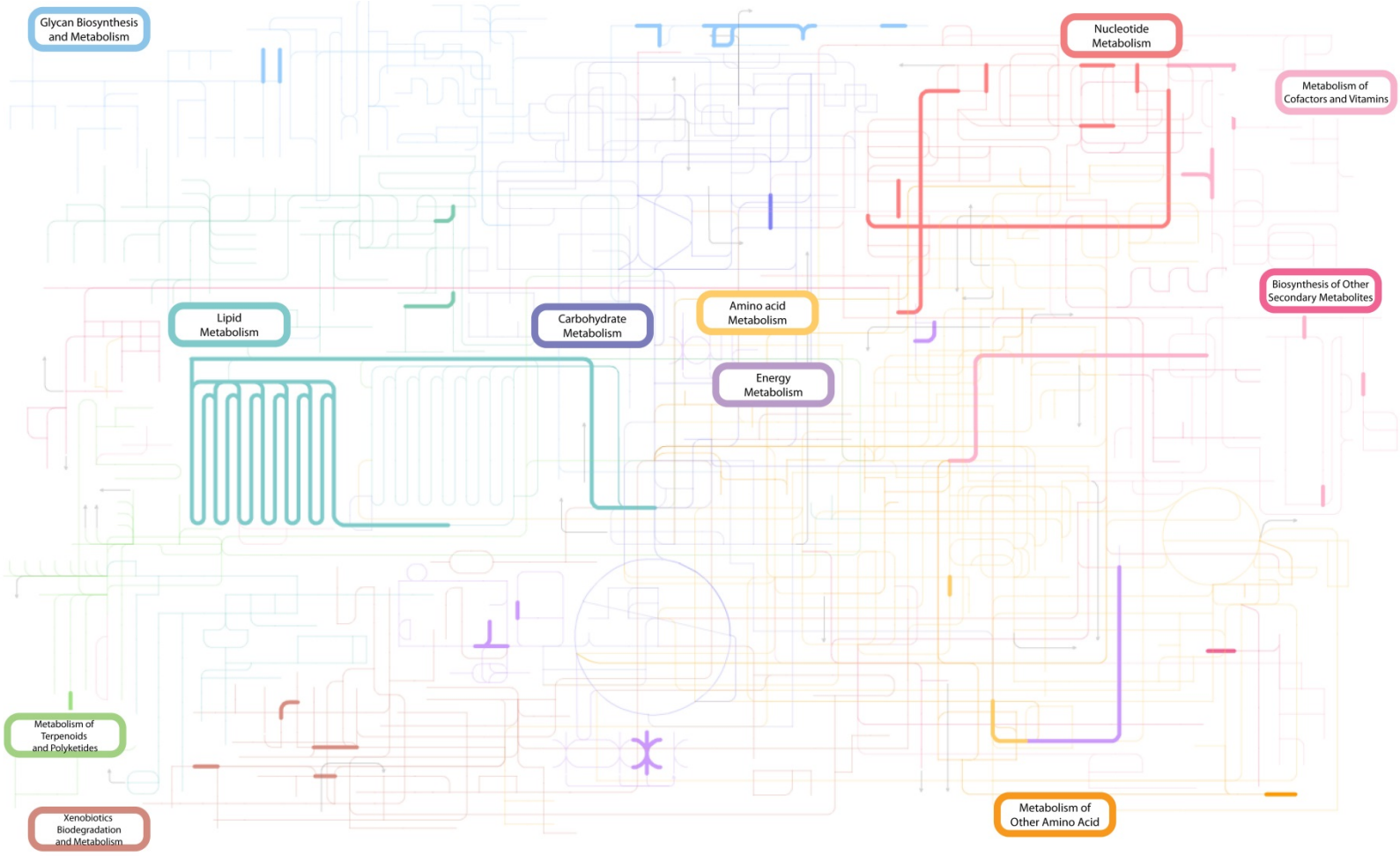


Figure 4.9. Metabolic biochemistry map of proteins identified only in T_1 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

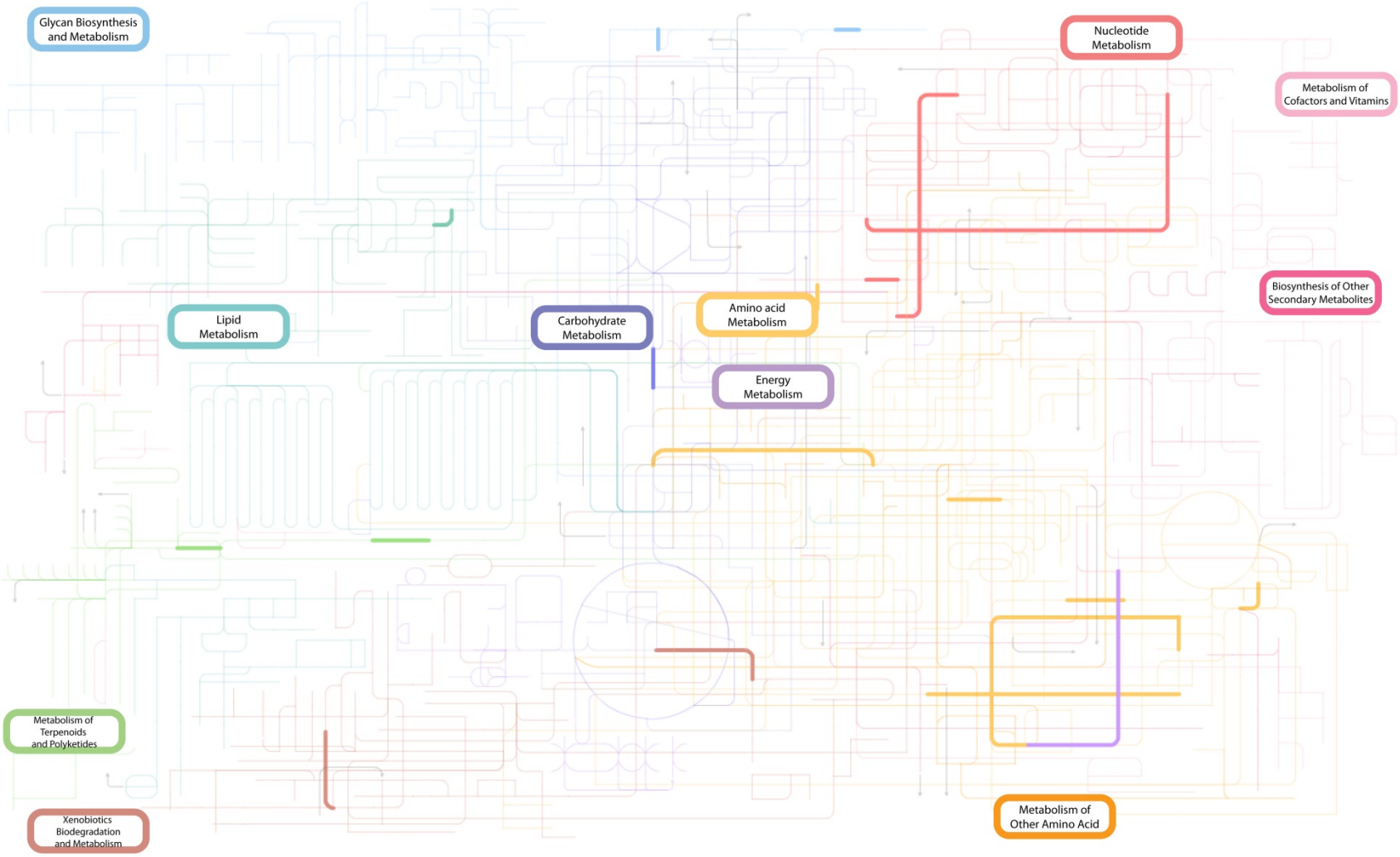


Figure 4.10. Metabolic biochemistry map of proteins identified in only T_2 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

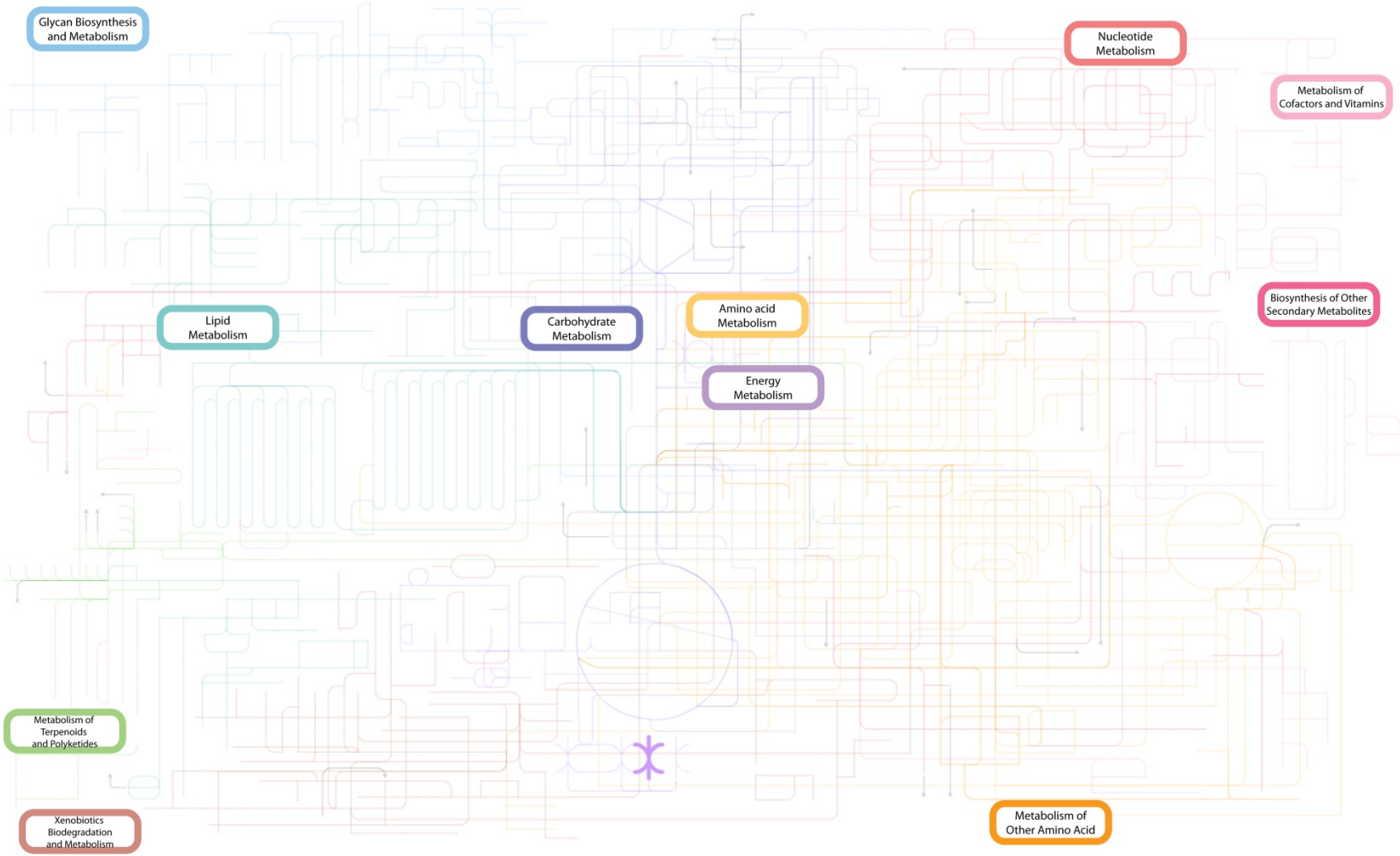


Figure 4.11. Metabolic biochemistry map of proteins identified in only T_4 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

4.5. Discussion

4.5.1. Protein expression during R. pomeroyi growth

Detailed protein analysis found that there was subtle variation among the three replicates at each time point, and minor variation over the growth phases of *R. pomeroyi*. At each time point the replicates were either statistically the same or had a similarity of 80% or greater among the expressed proteins. In a study of *R. pomeroyi* by Christie-Oleza, Fernandez, et al. (2012), technical and biological replicates also showed close similarity in a cluster analysis. This close similarity among replicates suggests *R. pomeroyi* is predictable in the proteins produced over its growth cycle. It also lends confidence that culture conditions in this investigation were maintained among the replicates throughout the investigation and variability was low.

The total number of proteins expressed at each time point decreased over time, which is not surprising since bacterial growth rate slows as a culture progresses from exponential phase through stationary phase (Barton, 2005). The number of shared proteins between adjacent time points also decreased over time although the shared proteins still accounted for at least 75% of the proteins identified at each time point. Although there was a decrease in the number of shared proteins, the large number of proteins in common suggests that the majority of the same biological functions occurred at all four time points.

Of the proteins unique to individual time points, T_1 had the highest number, 200, while T_3 and T_4 had far fewer with only 11 and 12, respectively. When examining the biological functions attributed to these unique proteins, various types of binding and transport are represented, as well as involvement in DNA metabolism, repair, binding and

unwinding. These biological functions are also found among the proteins shared between time points. Given these similarities, even though the individual proteins are unique to certain time points, the biological functions appear to be similar in all time points. The fact that there are 200 unique proteins in T_1 involved in these processes may indicate that at that time point cell replication was occurring to a greater extent than at later time points and a greater variety of proteins are required to facilitate this. This seems a logical assumption based on the fact that T_1 is occurring in early exponential phase when cell replication is likely to be high. Although there may have been a slight change in the quantity of proteins produced through the time points, there was not a great amount of change in the biological function of this bacterium as it progressed through consecutive growth phases. There were no changes that indicate a culture undergoing any obvious changes or stresses over time.

4.5.2. Proteins abundance and functionalities

Of the 23 most abundant proteins, the top five by means of spectral abundance had noticeably higher spectral counts with some time points exceeding 1000 counts. The remaining proteins were represented at relatively similar levels below 750 counts. This could indicate that the first five are needed by the bacterium in greater quantities with the remainder needed to a lesser extent. All proteins listed as abundant were identified in this investigation were found at all four time points. Prior studies have shown that in *R. pomeroyi*, the presence of natural microbial communities or the origin of the sea water did not cause a strong variation in the protein pattern (Christie-Oleza, Fernandez, et al., 2012). The similarity among the time points in this study and lack of change also

observed by Christie-Oleza, Fernandez, et al. (2012) may indicate that *R. pomeroyi* will consistently produce porin and transport proteins to a greater extent than other proteins.

The expressed proteins that were most prominent were porins or were involved in transport, binding, translation, and protein refolding. Moran et al. (2004) found that *R. pomeroyi* had the highest proportion of genes coding for signal transduction, transport and binding proteins so these finding are not unexpected. Christie-Oleza, Fernandez, et al. (2012) also found transporters and porins in *R. pomeroyi* cultures incubated in natural sea water. Porins have been detected in various ocean samples (Tanoue et al. 1995; Yamada and Tanoue, 2003; Yamada and Tanoue, 2009) and are thought to remain largely intact (Tanoue et al., 1995). As it is generally accepted that proteins are readily broken down and recycled or further degraded, the finding of porins in ocean samples points to their resistance to enzymatic degradation, potentially due to their occurrence in membranes. This resistance appears to allow them to remain intact to a degree where enough peptides are produced to enable identification. Dong et al. (2013) determined that in a survey of high molecular weight DOM (HMW-DOM), transporters were frequently present and Sowell et al. (2011) found that transport proteins were the most abundant category of proteins in a coastal ecosystem. In a metaproteomic study of the Southern Ocean, Williams et al. (2012) determined that proteins synthesized by *Alphaproteobacteria*, to which the *Roseobacter* clades, including *R. pomeroyi*, belong, were mostly ATP-binding cassette (ABC) transporters or other active transport systems.

Much of the genomic and proteomic information concerning marine bacteria has yet to be documented. In addition, there is very little known about the contribution of bacteria to the marine environment. However, this culture study using media and a

culture study using seawater (Christie-Oleza, Fernandez, et al., 2012) produced similar abundant proteins as have been seen in oceanic samples indicating that porins and transport proteins could be used as potential biomarkers of bacterial processes and/or activity. Moreover, because they have apparent refractory characteristics, the potential detection of porins in marine sediments may prove to be of value for tracking bacterial inputs. Of course, porins and transport proteins in oceanic samples are not necessarily limited to synthesis from bacteria. Nevertheless, more study of these proteins as potential markers seems advisable.

4.5.3. Significantly up- and down-regulated proteins

For the large number of total proteins identified relatively few were shown to be significantly up- or down-regulated. Because of the small number of proteins significantly altered it is difficult to draw many conclusions. While there were proteins that remained up- or down-regulated through the time points compared, an inspection of their biological function showed no apparent novel processes either being turned on (up-regulated) or off (down-regulated). The proteins remaining up- or down-regulated, had functionalities that could be described as basic housekeeping functions of a cell. When the biological functions of the proteins altered in only one comparison (i.e. only up- or down-regulated in T_2, T_3, or T_4) were reviewed, a handful of different proteins responsible for similar functions were found to be both up- and down-regulated. For example, in the T_1 to T_2 comparison, different proteins involved in metal ion binding and DNA binding were both up- and down-regulated. This indicates that while the presence or absence of an individual protein may be significant, their associate functionalities do

not seem to be altered across the time points and further supports the determination that there is little variation in biological function between the time points and various growth phases.

4.5.4. DAVID results

The most enriched proteins in all growth phases in this study were those responsible for various functions of replication: RNA subunit structure and RNA binding; ribosome assembly and biogenesis; translation and amino acid, amine, organic acid, and oxoacid biosynthesis and metabolism. This consistency indicates that the same biological processes appear to be occurring throughout the time points. As discussed previously, the decrease seen in the total number of proteins identified at T_3 and T_4 (stationary phase) and lower spectral counts of the most abundant proteins at those time points could potentially indicate a decrease in growth. The DAVID analysis, however, shows the cluster responsible for RNA, ribosomes and translation as being more enriched at T_3 and T_4 than T_1 and T_2 (exponential phase) with equal or slightly more total proteins. This would seem to contradict the idea that growth had slowed greatly in T_3 and T_4.

The cluster containing proteins responsible for amino acid biosynthesis and metabolism was also more enriched in time points 3 and 4 than 1 and 2. Moran et al. (2007) found evidence in a comparison of three *Roseobacter* genomes, of which *R. pomeroyi* was one, to support the idea that amino acid metabolism is particularly important and closely regulated among *Roseobacters*. The nucleotide and ribonucleotide biosynthesis cluster was also more enriched in T_4 than T_1 and T_2 and the aminoacylation; amino acid activation; RNA metabolism cluster was also more enriched in T_3 and T_4 than

T_1 and T_2. These enrichments may result from more proteins shifting to these purposes as nutrients declined. It is difficult to say for certain what purpose the enrichment of these proteins would serve in those later time points but it is important to note the difference for future comparisons.

4.5.5. Implications from iPath results

The iPath software provided a visualization of the biological functions/pathways that were being used by the cells within this study. When the shared proteins identified in all four time points are input, most of the pathways on the map are illuminated identifying proteins from all the major metabolic pathways (Fig. 4.8). Several pathways appear to have most of the pathway illuminated, including amino acid metabolism, nucleotide metabolism, and carbohydrate metabolism including the citric acid cycle (circular portion at the bottom of the iPath map). These findings corroborate the results from DAVID. Figures 4.9 through 4.11 identify the pathways illuminated by the unique proteins at time points 1, 2, and 4 and demonstrate visually that though the proteins are unique, the biological (i.e. metabolical) functions attributed to these proteins are not. The same pathways are identified in the unique proteins as in the shared proteins among the time points.

4.6. Conclusions

Identification and examination of the proteins produced by *R. pomeroyi* under environmentally realistic carbon conditions throughout its growth phases showed very little variation among replicate cultures and little difference among the proteins produced

at the four time points investigated. The small differences seen between the time points and analysis of the functionalities of the proteins produced at each time point indicate that the majority of the same biological functions were occurring throughout the study. The most abundant proteins identified were responsible for porins, transport, binding, translation, and protein refolding. The functional annotation clusters that were most enriched contained proteins that contributed to RNA subunit structure and RNA binding; ribosome assembly and biogenesis; translation and amino acid, amine, organic acid, and oxoacid biosynthesis and metabolism. The biological functions identified among the most abundant proteins and the enriched clusters suggest an organism that maintained active growth well into stationary phases with no obvious evidence of changing biological function. The presence of abundant porins and transport proteins in relative to other proteins synthesized could represent potential proteins to be used as biomarkers of bacterial processes and/or activity.

Chapter 5: Tracking proteomic expression of a tracer level addition of labeled amino acid in a heterotrophic marine bacterium

5.1. Abstract

A heterotrophic marine bacterium, *Ruegeria pomeroyi*, was cultured with a tracer level addition of ^{13}C-labeled leucine (average addition: 600-800nM). Bacterial protein biosynthesis which incorporated the label was tracked through exponential and stationary growth phases. Over 1900 total proteins were identified, with a total of 479 proteins incorporating the isotopic leucine label. Label incorporation ranged between 16 and 21% of the total proteins depending on growth phase. Almost one quarter (23%) of the proteins that incorporated the label were found in all four time points and adjacent time points had between 44 and 53% of their labeled proteins in common. The most abundant proteins identified in this investigation were involved in porins, transport, binding and translation and almost all contained the label. One functional annotation cluster was determined to be prominent and was enriched at all four time points. This cluster contained proteins involved in ribosomal biosynthesis, RNA binding, translation, and protein metabolism. The finding of these biological functions as most prevalent indicates that the cells were undergoing active growth through all growth phases. The widespread distribution of the label among the time points indicates that the isotopic label was recycled by the bacteria with little remineralization. This study demonstrates the successful tracking of the uptake and incorporation of an isotopically labeled amino acid and although no patterns to its uptake were distinguished this demonstrates a method by which bacterial protein synthesis can be tracked.

5.2. Introduction

Ecosystems are sustained by biogeochemical cycling through organic and inorganic reservoirs of major elements such as carbon, nitrogen and phosphorus, which are essential to the survival of organisms. There are many processes that occur to allow the cycling of these nutrients through the various biogeochemical cycles such as uptake and incorporation of inorganic nutrients, consumption by heterotrophic consumers, and the eventual degradation and remineralization of organic material. The ability to study these processes will facilitate our understanding of the mechanisms that are responsible for maintaining ecosystems as well as to help predict potential impacts of future changes.

Protein synthesis and consumption in the nitrogen cycle has yet to be fully explored because of the limited ability to examine proteins in environmental samples. The rapidly advancing field of proteomics, a high-throughput analysis for the identification of protein mixtures in complex systems, is providing insights into many processes that have not been previously examined (Armengaud, 2013; Hettich et al., 2013). An improved understanding of the cycling, degradation, and preservation of proteins is needed, specifically an examination of the role of bacteria.

Identification of bacterial proteins in marine environments can be very challenging. The majority of marine bacteria cannot be cultured in a laboratory (Schweder et al., 2008; Stewart, 2012) and while the application of metaproteomics, the study of all the proteins obtained from an environmental sample, can help alleviate this problem, protein identification through this type of analysis is still limited by the availability of bacterial peptide sequence information (Schweder et al., 2008). In

addition, bacterial proteins are very often masked by the more abundant inputs of eukaryotic marine organisms.

To enhance our understanding of the role of bacteria in microbial recycling in elemental cycles, investigation into the synthesis and recycling of proteins by marine bacteria is necessary. Bacterial production has a highly significant relation to primary production with estimates of bacterial production accounting for 20% of primary production in pelagic systems meaning that about 40% of primary production fluxes through bacteria (Cole et al., 1988). When considering the entire water column those estimates increase to approximately 30% of primary production with 60% of the primary production expected to flux through bacteria (Cole et al., 1988). Bacteria are also the main contributors to the diagenesis of organic matter in marine sediments (Deming and Barross, 1993).

Protein synthesis and bacterial production studies have traditionally involved uptake experiments that utilized isotopically labeled thymidine and leucine. These measures of biomass production used incorporation of thymidine into DNA (as a proxy from cell division) (Tibbles and Harris, 1996) and the incorporation of leucine into protein (as a proxy of protein synthesis) (Chin-Leo and Kirchman, 1988; Kirchman et al., 1985; Simon and Azam, 1989). While these measurements have been the standard to provide information on biomass production and have been essential in our ability to estimate nutrient cycling in marine systems (Chin-Leo and Kirchman, 1988) they don't allow us to identify individual protein expression by bacteria.

The goal of this study was to determine the incorporation of a tracer level addition of labeled amino acid by a heterotrophic marine bacterium. The ability to track the

production of proteins will enable a much more in-depth understanding of the extent to which bacteria synthesize, recycle, or remineralize organic material. For example, this will allow determination of whether the label is utilized in only certain proteins or if it is incorporated universally and, if used in selective proteins, whether there is an order to the types of proteins synthesized with the labeled substrate. In addition, the successful tracking of a labeled amino acid will potentially enable the tracking of other more complicated substrates and how they may be utilized or incorporated by bacteria thus expanding our knowledge of bacterial protein synthesis and recycling.

5.3. Materials and Methods

5.3.1. R. pomeroyi culture conditions and initial determination of growth kinetics

All culture conditions and the determination of a growth curve were the same as described in Chapter 4, sections 4.3.1 and 4.3.2. For cell counts, OD_{600} readings, and the determined growth curve for *R. pomeroyi* refer to Table 4.1 and Figure 4.1.

5.4.2. Experimental culture conditions

Cultures were grown in marine basal media with 0.1 mM glucose. A ^{13}C-labeled leucine substrate was added to the experimental groups. Three biological replicates were used. A blank was also established and maintained under the same conditions minus the addition of any organism. The labeled leucine was added at 23 and 36 hours after inoculation of the cultures and averaged 616 nM for the first addition and 795 nM for the second addition. The label was added in these concentrations because of the lack of an appropriate balance with which to measure a smaller amount of labeled leucine necessary

to make a less concentrated addition. The label was added twice in this initial attempt to track its incorporation as we were unsure if it would be rapidly remineralized and therefore not able to be seen in the proteins produced. Cultures and blank were grown in 2 L and 1 L flasks, respectively, at approximately 23°C on a laboratory bench top.

5.3.3. Sampling procedures, protein extraction, BCA quantification, protein digestion and mass spectrometry

All materials and methods for sampling procedures, protein extractions, quantification, digestion and mass spectrometery are the same as detailed in Chapter 4, sections 4.3.4, 4.3.5, and 4.3.6.

5.3.4. Database Search

Mass spectral results were interpreted and searched with SEQUEST (v. 2012.01.0) (Eng et al., 2008) to match observed spectra to theoretical spectra generated from the predicted peptide sequences determined by the search database chosen. All CID spectra were searched against the most recent version of the *R. pomeroyi* proteome (ASM1196v2; ncbi.nlm.nih.gov/genome; 4,278 proteins) and 50 common protein contaminants in the lab (e.g. keratins, trypsin). The searches were conducted with enzyme specificity for trypsin and a +6 Da modification to leucine residues in order to monitor and track the $^{13}C_6$-leucine additions. Match probability between the observed and theoretical spectra was set at 90% ($p<0.1$) on ProteinProphet and PeptideProphet (Keller et al., 2002). Proteins with only one identified peptide were excluded.

5.3.5. DAVID

The DAVID bioinformatics resource (http://david.abcc.ncifcrf.gov/home.jsp), as detailed in Chapter 4, section 4.3.8, was used to interpret the biological importance of the proteins identified in this study. Results are reported as what proteins were enriched with modified Fisher p-values (also called EASE scores) <0.01.

5.3.6. Label-free protein quantification and Interactive Pathways Explorer (iPath v2)

The determination of significantly up- or down-regulated proteins and the employment of iPath software was as described in Chapter 4, sections 4.3.9 and 4.3.10.

5.4. Results

5.4.1. Total proteins identified

In total, 1927 total proteins were observed among all the replicates across all time points. Of those observed proteins, 808 were seen in all replicates at all time points. Statistical analysis, a Cochran Q test, was performed on the replicates of each time point to determine whether the total proteins seen among the time points were statistically the same. This was conducted based on the presence or absence of a protein in a replicate. It was determined that the replicates of T_1 and the repeated analyses of the sample from T_4 were statistically the same. This was to be expected for the T_4 sample due to the fact that for that time point one sample was run three times through the MS as biological replicates were not available. Though the replicates for the second and third time points were not statistically the same, there was a similarity of 82% or greater among the

proteins identified for the replicates of a time point for all time points in the study. The number of observed proteins decreased from T_1 to T_2 but remained relatively constant from T_2 through T_4 as the culture progressed from late exponential phase through stationary phase (Table 5.1; Fig. 5.1).

Of the 1927 identified proteins, 479 incorporated the ^{13}C-leucine label during synthesis. The number of labeled proteins seen at the four time points ranged from 238 to 299 (Table 5.1). Though the number of total proteins identified was higher at T_1 than the other three time points by more than 200 proteins, the percentage of labeled proteins identified within each time point showed much less variation among the four time points (Table 5.1; Fig. 5.1) ranging from 16 to 21% with an average of 18% of the identified proteins carrying a label.

Protein analysis was also done on the experimental blanks at T_2 and T_3. Only eight *R. pomeroyi* proteins were identified at T_2 and only nine were identified at T_3 (Table 5.1). This equates to 0.5% and 0.6% of the total proteins identified at those time points. None of the proteins identified in the blanks contained a label.

Table 5.1. Sum of *R. pomeroyi* proteins observed in all replicates by time point.* Proteins were defined as having match probability of 90% or greater by ProteinProphet AND PeptideProphet and having 2 or more independent spectra identified for the protein.

Time Point	Time (hrs)	Growth State	Proteins observed	Labeled proteins observed	% of observed proteins labeled
T_1	25	Early exponential	1719	286	17
T_2	38	Late exponential	1445	299	21
T_2 (Blank)	38	Late exponential	8	0	0
T_3	44	Early stationary	1489	256	17
T_3 (Blank)	44	Early stationary	9	0	0
T_4	66	Late stationary	1452	238	16

* The lists of proteins found in the three replicates for each time point were combined and duplicate, or over-lapping, proteins were deleted to compile a list of all the proteins seen within a time point.

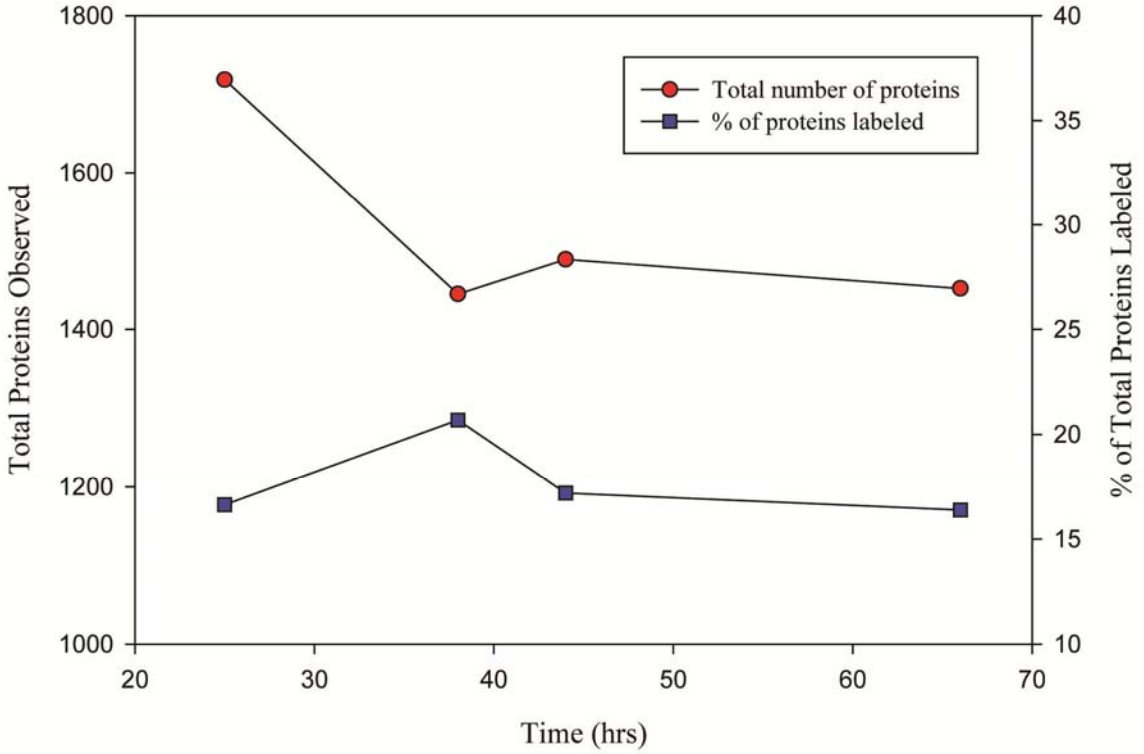


Figure 5.1. Sum of *R. pomeroyi* proteins observed and the percentage of that total that was labeled with ^{13}C-leucine in all replicates at the four time point plotted by hours. The lists of proteins found in the three replicates for each time point were combined and duplicate, or over-lapping, proteins were deleted to compile a list of all the proteins seen within a time point. Proteins were defined as having match probability of 90% or greater by ProteinProphet and PeptideProphet and having 2 or more independent spectra identified for the protein.

5.4.2. Labeled proteins by time point

There were 108 labeled proteins identified among all four time points (Fig. 5.2). These proteins account for 23% of the 479 labeled proteins seen among all four time points. When examining individual time points, the labeled proteins unique to those time points vary from as few as 28 up to 72 accounting for 11 to 25% of the labeled proteins observed (Fig. 5.2; Table 5.2). T_1 has the most labeled proteins unique to a single time

point with 72, or 25%. Additional comparisons were made between the time points as follows: T_1 to T_2, T_2 to T_3, T_3 to T_4, and T1 to T_4 (Table 5.3; Fig. 5.3). The comparisons of adjacent time points show that the percentage of labeled proteins in common between the time points doesn't vary greatly, from 44 to 53% (Table 5.3). When comparing the earliest and latest time points (T_1 and T_4), however, the number of labeled proteins in common decreases slightly to 37% (Table 5.3).

A heat map (Fig. 5.4) was generated to visualize the labeled proteins identified by gene ontology category giving a visual representation of the labeling of proteins over time. Each gene ontology category is represented by a black line in the far left column. Each smaller black line in the second column represents a protein which falls under that ontology category. This visualization shows that a few of the broader gene ontology categories show more proteins labeled in the earlier time points with labeled proteins decreasing over time while a few others show increases in labeling over time. Two categories of proteins involved in transport activity show a slight increase in labeling over time while two categories of proteins involved with RNA binding, ribosomal structure and translation show slight decreases in labeling over time. Overall, however, this visual representation shows labeling equally dispersed throughout the four time points. Table A5.1 provides a complete list of all labeled proteins by time point.

Table 5.2. Labeled proteins by time point.

Time Point	Labeled proteins observed	Labeled proteins unique* to the time point	% of labeled proteins unique to each time point	# of unique proteins that were up- or down-regulated†
T_1	286	72	25	40
T_2	299	56	19	27
T_3	256	28	11	16
T_4	238	30	13	19

* Unique proteins are those proteins found only in the referenced time point; † Proteins were considered to be up- or down-regulated if the corresponding false discovery rate (FDRup or FDRdown) was $p < 0.05$.

Table 5.3. Comparison between the time points of the labeled proteins identified.

Time points compared	Total proteins in both time points	# proteins in common	% of proteins in common	# proteins unique* to 1st time point	# proteins unique to 2nd time point
T_1 - T_2	405	180	44	106	119
T_1 - T_3	378	164	43	122	92
T_1 - T_4	382	142	37	144	96
T_2 - T_3	368	187	51	112	69
T_2 - T_4	366	171	47	128	67
T_3 - T_4	323	171	53	85	67

* Unique proteins are those proteins found only in the referenced time point.

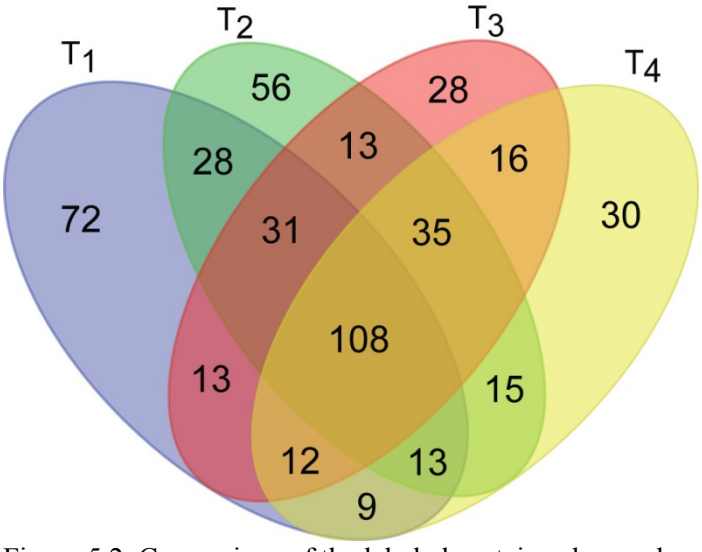


Figure 5.2. Comparison of the labeled proteins observed among the four time points. The Venn diagram was created using the software available on http://bioinformatics.psb.ugent.be/webtools/Venn/.

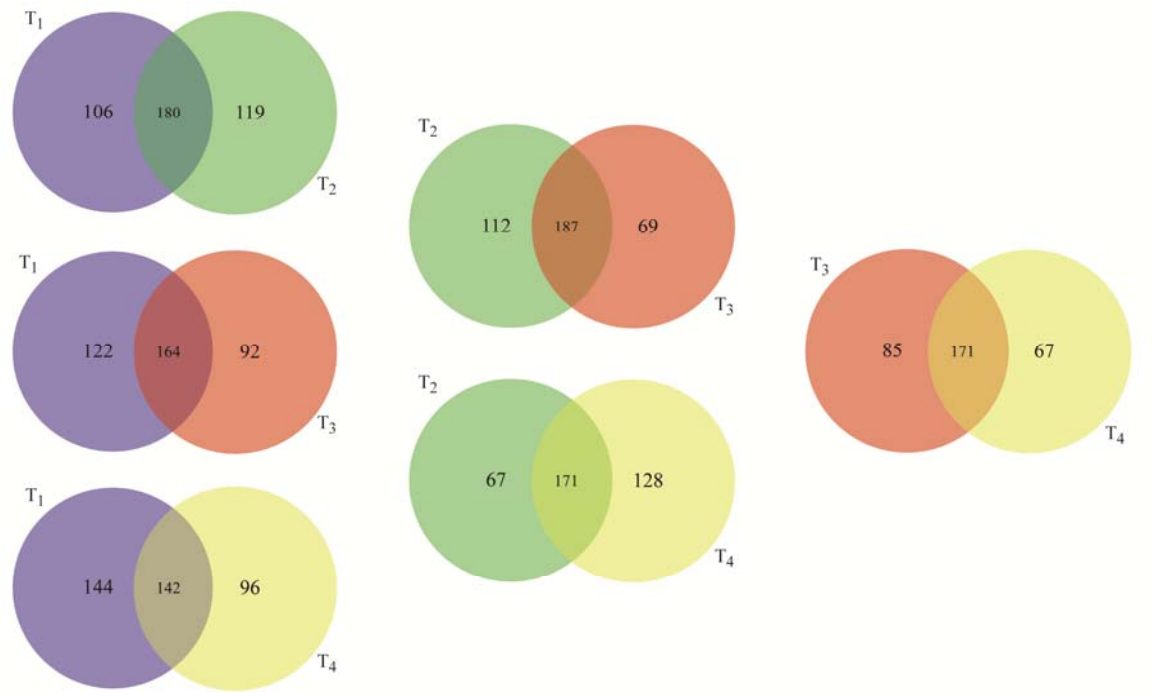


Figure 5.3. Time point comparisons of labeled proteins. The Venn diagrams were calculated using the software available on http://bioinformatics.psb.ugent.be/webtools/Venn/.

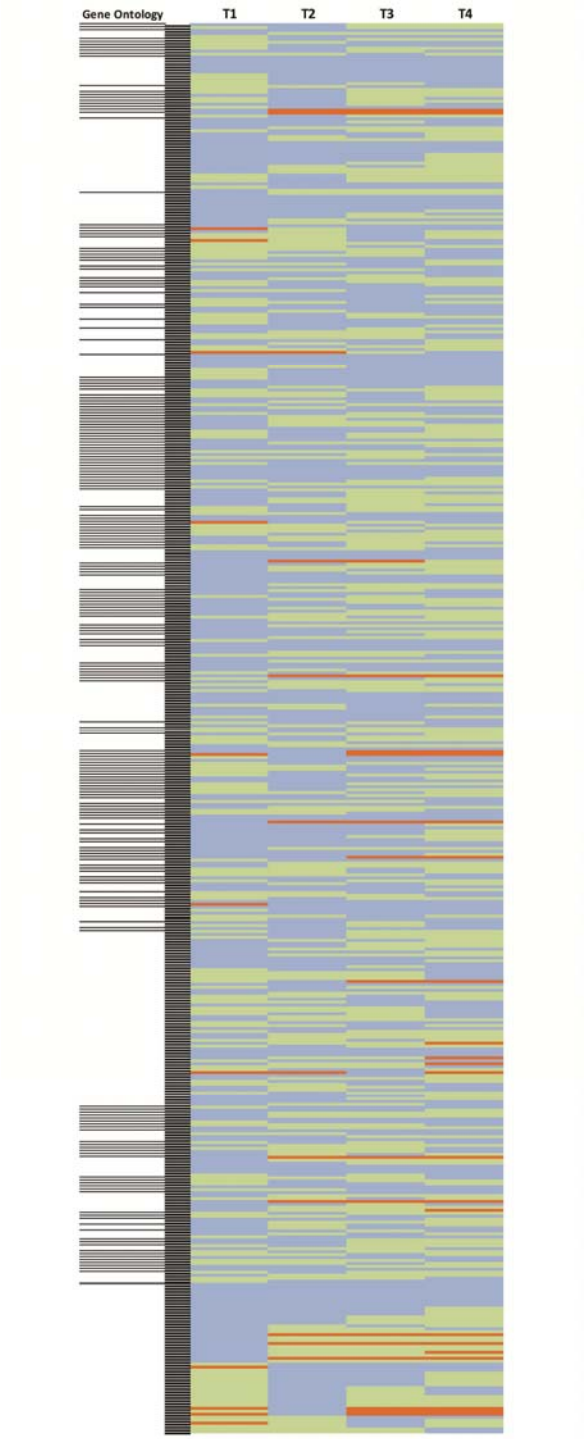


Figure 5.4. Heat map of the observed labeled proteins grouped by gene ontology. Blue indicates the presence of a labeled protein in that time point, green the presence of the protein but the absence of the label in that time point, and orange the absence of the protein in that time point. Black lines on the left under the gene ontology heading indicate a gene ontology category; the smaller lines in the second column indicate the multiple proteins present that fall under that category.

5.4.3. Most abundant proteins

Protein abundance was determined through the semi-quantitative measure of spectral counts. To determine the abundance of proteins within each sample, proteins were ranked by spectral counts. The top 1% of proteins for each replicate from each time point was chosen. The lists of proteins from the three replicates of each time point were combined and duplicate proteins eliminated to give cumulative lists of 14 to 16 proteins. The lists from all time point were then combined to produce one final list of the 17 most abundant proteins throughout the experiment (Table 5.4). The total number of spectral counts from the three replicates was then summed for each time point to give a final total for each protein (Fig. 5.5). It should be noted that the size of the individual proteins was not taken into consideration for this analysis so the spectral count data used here has not been normalized to protein length.

The four most abundant proteins display a great deal more spectral counts than the remaining 12 proteins all of which remain at or just below 1000 counts (Fig. 5.5). T_1 and T_2 have higher counts than T_3 and T_4 for all 17 proteins (Fig. 5.5). The four most abundant proteins deal with porins, transport and binding activity, and translation (Table 5.4; Fig. 5.5). The vast majority of the most abundant proteins were also labeled proteins (Table 5.4).

Table 5.4. Most abundant (1%) proteins identified ranked by total number of spectral counts among all replicates in all time points.

Protein (UniProt ID†)	T1			T2			T3			T4		
	Replicate A	Replicate B	Replicate C	Replicate A	Replicate B	Replicate C	Replicate A	Replicate B	Replicate C	Replicate A	Replicate B	Replicate C
Q5LMY2	**857**	**1048**	**1073**	**1676**	**1220**	**816**	**937**	**849**	**965**	**897**	**948**	**974**
Q5LV41	**728**	**671**	**785**	**647**	**674**	**379**	**380**	**314**	**548**	**351**	**387**	**345**
Q5LW24	**735**	**538**	**768**	**744**	**521**	**324**	**443**	**245**	**375**	**371**	**382**	**388**
Q5LMR5	**472**	**467**	**493**	**808**	**869**	**369**	**369**	**311**	**477**	**378**	**355**	**376**
Q5LV15	**369**	**363**	**373**	**356**	**336**	**207**	192	**139**	**182**	**162**	**166**	**166**
Q5LMQ5	**253**	**214**	**232**	**387**	**372**	**172**	**184**	**170**	**191**	**163**	152	**171**
Q5LS31	**240**	**262**	**240**	**284**	**331**	**179**	**159**	**152**	**228**	**146**	**168**	**155**
Q5LPN4	**201**	**198**	**206**	**377**	**382**	**182**	**171**	**137**	**188**	**167**	**158**	**154**
Q5LNP1	**210**	**180**	**206**	**398**	**332**	**144**	**187**	**129**	**162**	**142**	**154**	**147**
Q5LMQ6	**233**	**194**	**219**	**329**	**287**	**148**	**175**	**146**	**157**	**165**	**164**	**156**
Q5LMR4	**216**	**180**	**227**	**288**	**250**	**116**	**135**	**101**	**109**	**116**	**121**	**109**
Q5LN23	**162**	**156**	**166**	299	**283**	122	**136**	**102**	**148**	**128**	**129**	117
Q5LRZ5	**191**	**262**	**311**	**208**	**191**	**110**	**108**	**76**	**121**	86	**94**	**92**
Q5LMT9	139	**168**	**192**	**254**	**237**	114	**105**	**109**	**136**	**117**	**120**	**116**
Q5LQB9	**148**	**163**	**169**	**274**	**162**	**101**	**119**	**89**	**142**	**125**	**119**	**118**
Q5LNN9	**135**	**137**	**126**	**280**	**237**	**117**	**116**	**84**	**114**	**96**	**111**	**103**
Q5LX13	**201**	**131**	**197**	**262**	**235**	**117**	**43**	66	**86**	63	60	57

***** Ranking was determined as follows: the proteins with the top 1% of independent spectral counts were obtained for each replicate. The lists of the replicates were combined and duplicates eliminated to produce a total list for the time point. This resulted in lists consisting of 14 to 16 proteins. The combined lists from the time points were compiled and duplicates removed to give the list above. This equates to 1.2 to 1.6% of the total proteins depending on the replicate; † Q5LMY2: porin activity, Q5LW24: transporter activity, Q5LV41: D-xylose transport; monosaccharide binding, Q5LMR5: GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity, Q5LV15:ATP binding; cytoplasm; protein refolding, Q5LMQ5: DNA binding; DNA-directed RNA polymerase activity; ribonucleoside binding; transcription, DNA-dependent, Q5LS31: amino acid binding, Q5LMQ6: DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-dependent; Q5LMR4: GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity, Q5LNP1: ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1), Q5LPN4:transporter activity, Q5LX13: calcium ion binding, Q5LRZ5: cytoplasm; translation elongation factor activity, Q5LN23: 3'-5'-exoribonuclease activity; RNA binding; RNA processing; cytoplasm; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity, Q5LQB9: outer membrane-bounded periplasmic space; transport, Q5LMT9: No report, Q5LNN9: ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1); proton-transporting ATPase activity, rotational mechanism; Bold = labeled proteins.

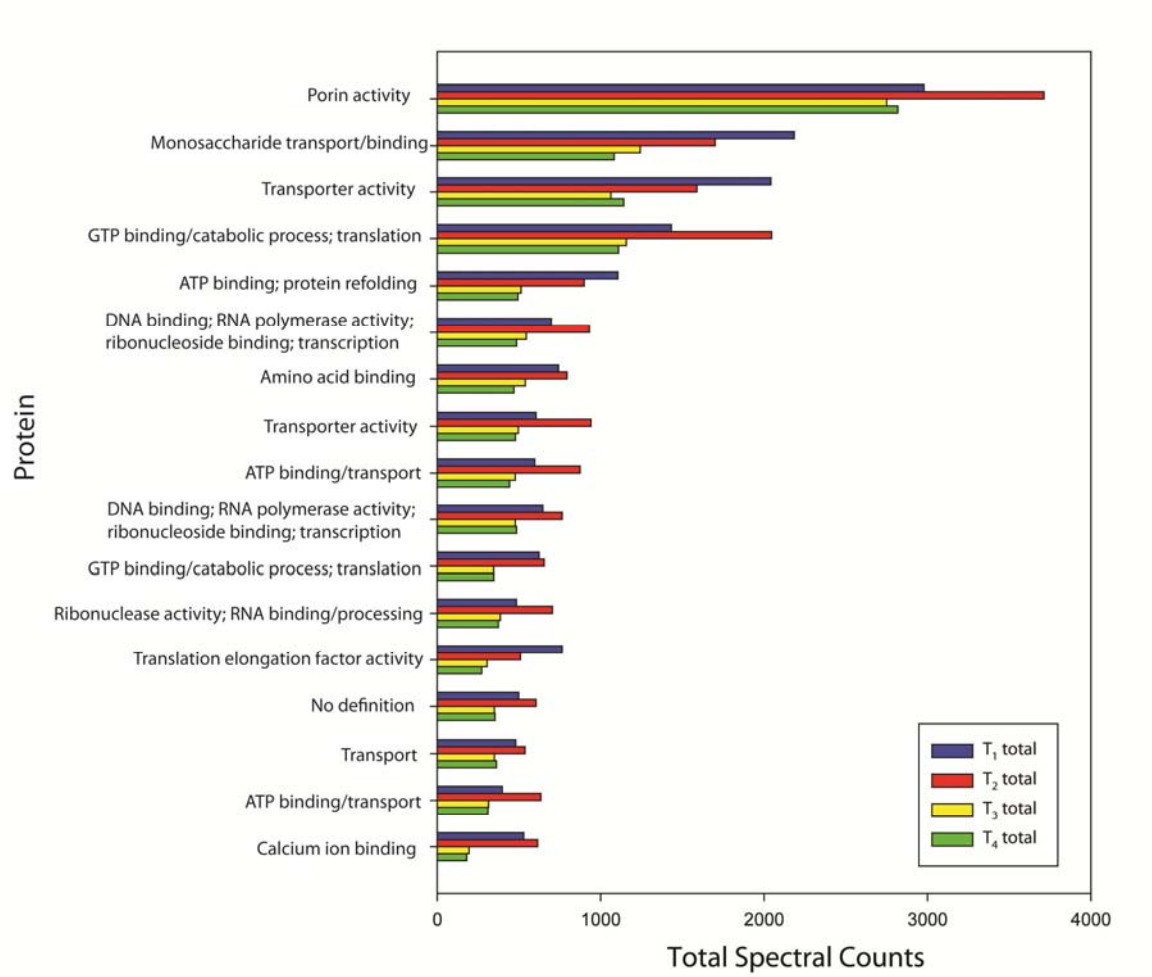


Figure 5.5. Most abundant labeled proteins defined by total spectral counts for each time point from all replicates.

5.4.4. Significantly up- and down-regulated proteins

Comparisons were made between T_1 and T_2, T_3, and T_4 using spectral counts and the label-free protein quantification program, Qspec. There were a total of 339 proteins that were up- or down-regulated in at least one comparison (Table 5.5; Table A5.2). There were 45 proteins up-regulated in T_2 that remained up-regulated in T_3 and T_4. There

were 29 proteins down-regulated in T_2 that remained down-regulated in T_3 and T_4. There were 265 proteins whose regulation varied between time points, meaning that they were up- or down-regulated at certain time points and not others (Fig. 5.6). Table A5.2 provides a complete list of up- and down-regulated proteins and their biological functions.

Table 5.5. Up- and down-regulated proteins as compared between time points. * The latter time point is what is being referred to as up- or down-regulated.

Time point comparison	Time (hours)	# of up-regulated proteins	# of down-regulated proteins
T_1 v T_2	25 v 38	114	115
T_1 v T_3	25 v 44	123	62
T_1 v T_4	25 v 66	128	58

* Proteins were considered to be up- or down-regulated if the corresponding false discovery rate (FDRup or FDRdown) was $p < 0.05$.

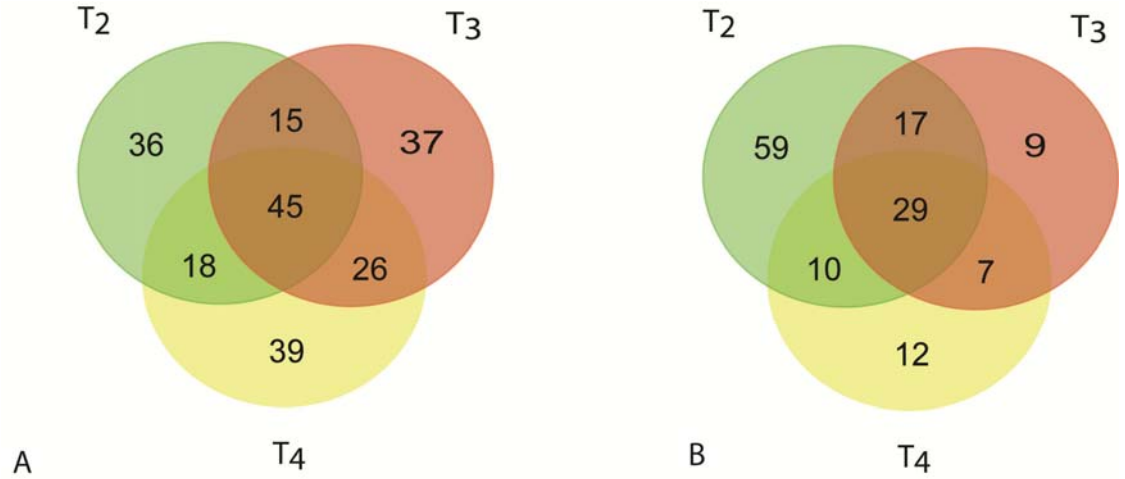


Figure 5.6. Distribution of up- and down-regulated proteins. (A) Proteins within each time point that were determined to be up-regulated compared to T_1. (B) Proteins within each time point that were determined to be down-regulated compared to T_1. Proteins were considered to be significantly up- or down-regulated if their FDRup or FDRdown was $p<0.05$. The Venn diagram was created using the software available on http://bioinformatics.psb.ugent.be/webtools/Venn/.

5.4.5. DAVID results

The lists of all labeled proteins identified at each time point were analyzed via the Functional Annotation Clustering tool in DAVID, resulting in identification of clusters of similar annotations. I report here the number of proteins (count) associated with each annotation within the cluster. While DAVID also has the ability to assign an enrichment score to the identified clusters, many of the clusters in this study were not well represented at every time point, or were absent, so to do a comparison of the enrichment scores would provide little value to the analysis.

For the purposes of our analysis, some clusters were manually combined from DAVID outputs. This was done because some clusters in the DAVID output were comprised of the same individual annotations within the cluster across all time points, but

other clusters had some, but not all, of the same individual annotations that may be present in the same cluster at a different time point. To make the comparisons uniform, the individual annotations within a cluster were combined across the four time points.

Although the majority of the clusters were not present in all time points, there were three clusters, 1 through 3, which were present in all time points (Table 5.6; Fig. 5.7). Cluster 1 was comprised of proteins involved in RNA subunit structure, RNA binding, ribosome assembly and biogenesis and translation. Cluster 2 incorporated proteins involved in various monosaccharide and carbohydrate metabolic and catabolic processes, glycolysis and gluconeogenesis. Cluster 3 included proteins involved in acetyl-CoA biosynthesis, metabolism and catabolism, the citric acid cycle, and respiration. The remaining seven clusters are present in some but not all time points and incorporate proteins responsible for various cell functions such as amino acid biosynthesis and metabolism; nucleotide and nucleoside biosynthesis and metabolism and ATP synthesis and transport; translation elongation; aminoacylation, RNA metabolism and ligase activity; translation and protein synthesis; binding; and protein folding (Table 5.6; Fig. 5.7).

The proteins that were significantly up- or down-regulated were also analyzed using DAVID employing a p-value <0.01 to determine enriched annotations. There was only one cluster identified at this statistical threshold in all the comparisons. To get an initial estimate of what biological functions were up- or down-regulated we raised the threshold to $p<0.1$. There were clusters identified for almost all comparisons at this new range. Because there is less statistical confidence with the higher p-value, a manual comparison of the gene ontology categories associated with the altered proteins was also

done. Comparisons, done with DAVID and manually, showed that there were few to no functions altered in the time point comparisons that would give an indication of a change in the behavior of the cells.

Table 5.6. DAVID results for all labeled proteins. Some clusters were manually compiled as not all annotation categories were enriched in all time points though the categories should clearly be grouped based on biological function.

		Counts			
Cluster	Gene Ontology Category or Pathway	T_1	T_2	T_3	T_4
1	GO:0003735~structural constituent of ribosome	42	33	25	21
	GO:0006412~translation	55	46	39	36
	GO:0005198~structural molecule activity	42	34	27	22
	GO:0019843~rRNA binding	27	22	16	11
	sil03010:Ribosome	42	33	25	21
	GO:0044267~cellular protein metabolic process	63	53	46	43
	GO:0022618~ribonucleoprotein complex assembly	24	21	16	13
	GO:0042255~ribosome assembly	24	21	16	13
	GO:0003723~RNA binding	32	28	21	16
	GO:0019538~protein metabolic process	67	59	53	49
	GO:0022613~ribonucleoprotein complex biogenesis	25	23	17	14
	GO:0042254~ribosome biogenesis	25	23	17	14
	GO:0034622~cellular macromolecular complex assembly	25	22	16	13
	GO:0034621~cellular macromolecular complex subunit organization	26	26	18	16
	GO:0065003~macromolecular complex assembly	25	23	16	13
	GO:0043933~macromolecular complex subunit organization	26	27	18	16
	GO:0034645~cellular macromolecule biosynthetic process	67	-	-	-
		637	494	386	331
2	GO:0006090~pyruvate metabolic process	8	-	-	7
	GO:0005996~monosaccharide metabolic process	14	21	14	14
	GO:0006006~glucose metabolic process	11	18	13	12
	GO:0044262~cellular carbohydrate metabolic process	20	26	17	17
	GO:0019318~hexose metabolic process	11	19	13	13
	GO:0046365~monosaccharide catabolic process	-	15	12	10
	GO:0019320~hexose catabolic process	-	15	12	10
	GO:0006007~glucose catabolic process	-	15	12	10
	GO:0006096~glycolysis	-	10	7	7
	GO:0044275~cellular carbohydrate catabolic process	-	15	12	10
	GO:0046164~alcohol catabolic process	-	15	13	11
	GO:0016052~carbohydrate catabolic process	-	15	12	10
	sil00010:Glycolysis / Gluconeogenesis	-	16	-	12
		64	200	137	143
3	GO:0006091~generation of precursor metabolites and energy	-	36	20	-
	GO:0006084~acetyl-CoA metabolic process	10	14	10	11
	GO:0046356~acetyl-CoA catabolic process	8	11	8	7
	GO:0006099~tricarboxylic acid cycle	8	11	8	7
	GO:0009060~aerobic respiration	9	13	9	8
	GO:0051187~cofactor catabolic process	8	11	8	7
	GO:0009109~coenzyme catabolic process	8	11	8	7
	sil00020: Citrate cycle	12	14	11	12
	GO:0015980~energy derivation by oxidation of organic compounds	-	21	-	-
	GO:0045333~cellular respiration	-	15	-	-
	GO:0006732~coenzyme metabolic process	-	23	-	-
	GO:0006086~acetyl-CoA biosynthetic process from pyruvate	-	-	-	4
	GO:0006085~acetyl-CoA biosynthetic process	-	-	-	4
		63	180	82	67
4	GO:0043436~oxoacid metabolic process	64	58	-	-
	GO:0046394~carboxylic acid biosynthetic process	30	-	-	-
	GO:0016053~organic acid biosynthetic process	30	-	-	-
	GO:0006520~cellular amino acid metabolic process	39	36	-	-

	GO:0008652~cellular amino acid biosynthetic process	22	23	-	-
	GO:0044106~cellular amine metabolic process	39	-	-	-
	GO:0009309~amine biosynthetic process	22	23	-	-
		246	140	-	-
5	GO:0006164~purine nucleotide biosynthetic process	11	10	-	9
	GO:0006163~purine nucleotide metabolic process	11	-	-	-
	GO:0015985~energy coupled proton transport, electrochem. gradient	6	6	-	-
	GO:0015986~ATP synthesis coupled proton transport	6	6	-	-
	GO:0034220~ion transmembrane transport	6	6	-	-
	GO:0015992~proton transport	7	-	-	6
	GO:0006818~hydrogen transport	7	-	-	-
	GO:0015672~monovalent inorganic cation transport	10	-	-	-
	GO:0046961~proton-trans. ATPase activity, rotational mechanism	4	4	-	-
	GO:0042777~plasma membrane ATP synthesis proton transport	4	4	-	-
	GO:0009145~purine nucleoside triphosphate biosynthetic process	-	7	-	6
	GO:0009206~purine ribonucleoside triphosphate biosynthetic process	-	7	-	6
	GO:0009201~ribonucleoside triphosphate biosynthetic process	-	7	-	6
	GO:0009152~purine ribonucleotide biosynthetic process	-	10	-	9
	GO:0006119~oxidative phosphorylation	-	8	-	7
	GO:0009142~nucleoside triphosphate biosynthetic process	-	7	-	6
	GO:0009144~purine nucleoside triphosphate metabolic process	-	7	-	6
	GO:0009205~purine ribonucleoside triphosphate metabolic process	-	7	-	6
	GO:0009199~ribonucleoside triphosphate metabolic process	-	7	-	6
	GO:0009150~purine ribonucleotide metabolic process	-	10	-	9
	GO:0009141~nucleoside triphosphate metabolic process	-	7	-	-
	GO:0055085~transmembrane transport	-	8	-	-
	GO:0009260~ribonucleotide biosynthetic process	-	11	-	-
		72	139	-	82
6	GO:0008135~translation factor activity, nucleic acid binding	-	-	6	6
	GO:0003746~translation elongation factor activity	-	-	4	4
	IPR004161:Translation elongation factor EFTu/EF1A, domain 2	-	-	4	4
	GO:0006414~translational elongation	-	-	4	4
		-	-	18	18
7	GO:0006418~tRNA aminoacylation for protein translation	-	-	8	9
	GO:0043039~tRNA aminoacylation	-	-	8	9
	GO:0043038~amino acid activation	-	-	8	9
	GO:0016876~ligase activity, aminoacyl-tRNA & related compounds	-	-	-	8
	GO:0004812~aminoacyl-tRNA ligase activity	-	-	-	8
	GO:0016875~ligase activity, forming carbon-oxygen bonds	-	-	-	8
	sil00970:Aminoacyl-tRNA biosynthesis	-	-	-	9
		-	-	24	60
8	IPR004161:Translation elongation factor EFTu/EF1A, domain 2	5	-	-	-
	IPR000795:Protein synthesis factor, GTP-binding	5	-	-	-
	GO:0003924~GTPase activity	6	-	-	-
	GO:0019001~guanyl nucleotide binding	8	-	-	-
		24	-	-	-
9	GO:0043176~amine binding	-	-	12	-
	GO:0031406~carboxylic acid binding	-	-	12	-
	GO:0016597~amino acid binding	-	-	9	-
		-	-	33	-
10	GO:0003755~peptidyl-prolyl cis-trans isomerase activity	-	-	5	-
	GO:0016859~cis-trans isomerase activity	-	-	5	-
	GO:0006457~protein folding	-	-	7	-
		-	-	17	-

-, indicates that the annotation category was not determined to be enriched at that time point.

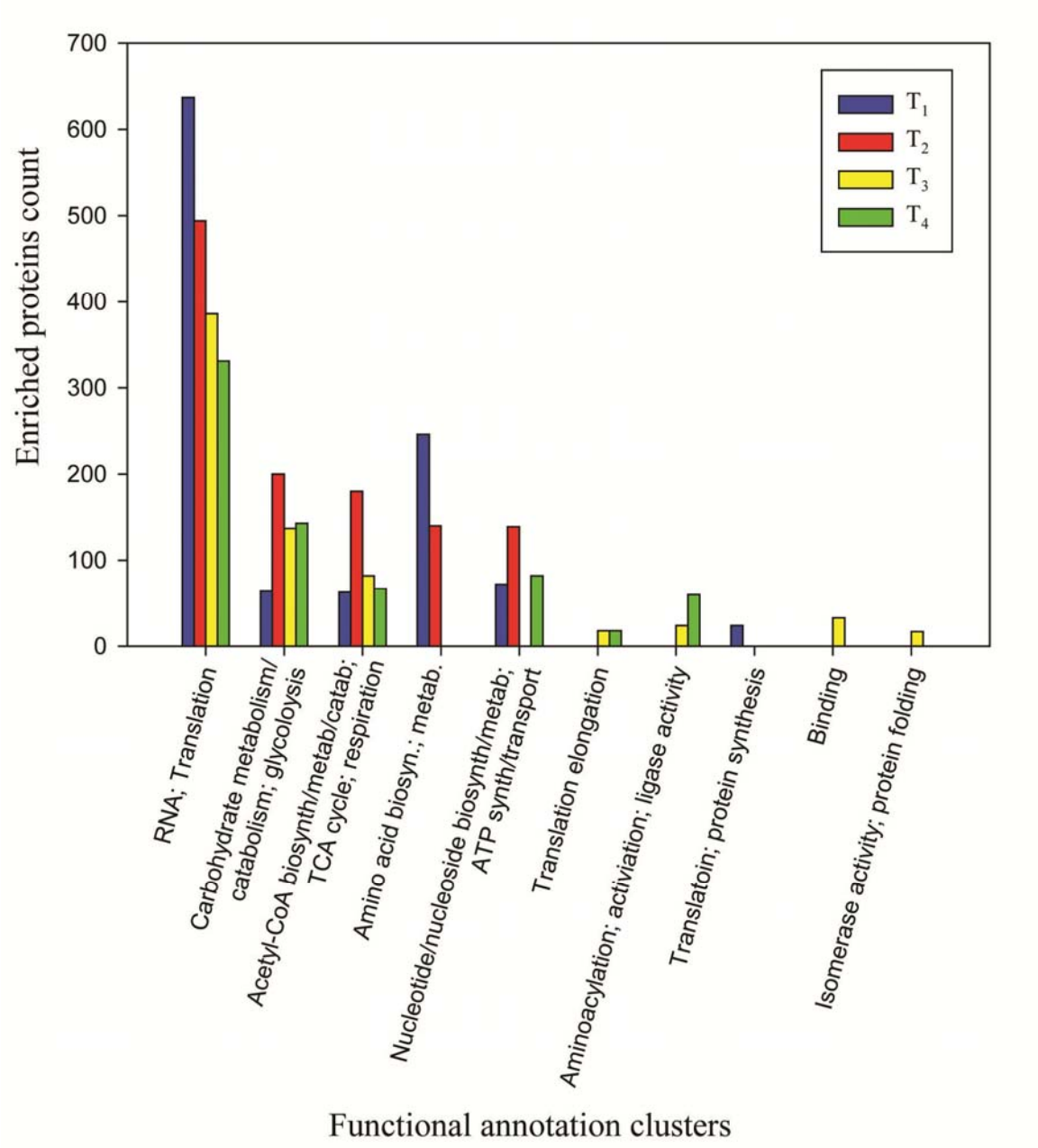


Figure 5.7. Sum of the proteins shown to be enriched within functional annotation clusters by DAVID. Proteins are determined to be enriched if they are present to a greater extent than would be expected by random chance when compared to that protein's ratio to the entire genome. DAVID groups annotations with common proteins together to create clusters of functional annotations.

5.4.6. iPath results

The Interactive Pathways Explorer software (iPath v2) was used to evaluate the expressed proteins identified within the investigation (http://pathways.embl.de/). Figure 5.8 displays the pathways generated when the labeled proteins shared among all four time points of this investigation were entered into iPath. Of the 108 labeled proteins identified among the four time points, 29, or approximately 27%, were mapped to at least one pathway. Figures 5.9, 5.10, 5.11 and 5.12 display the pathways generated by the proteins unique to time points 1, 2, 3, and 4, respectively. .

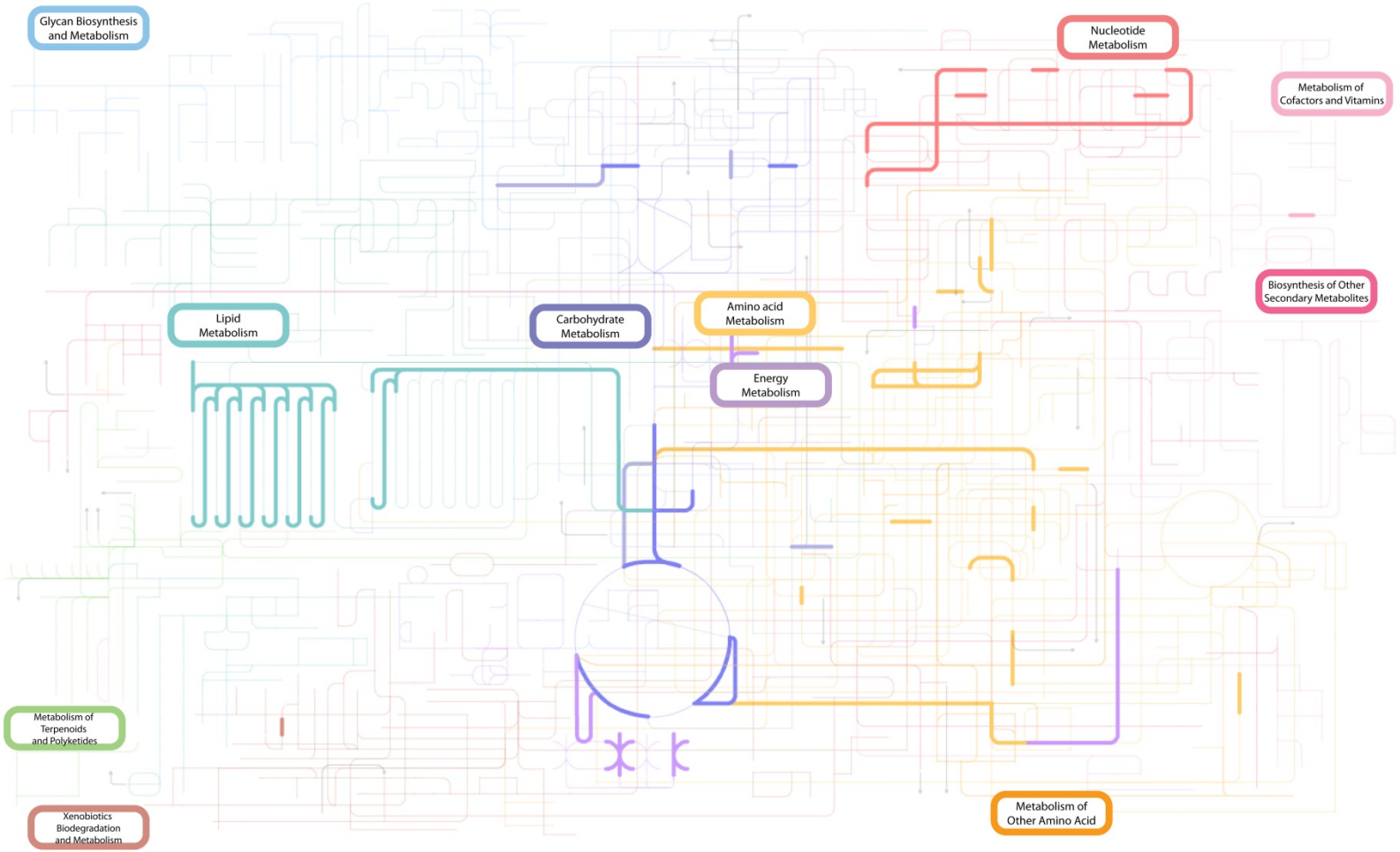


Figure 5.8. Metabolic biochemistry map of the labeled proteins identified in all time points for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

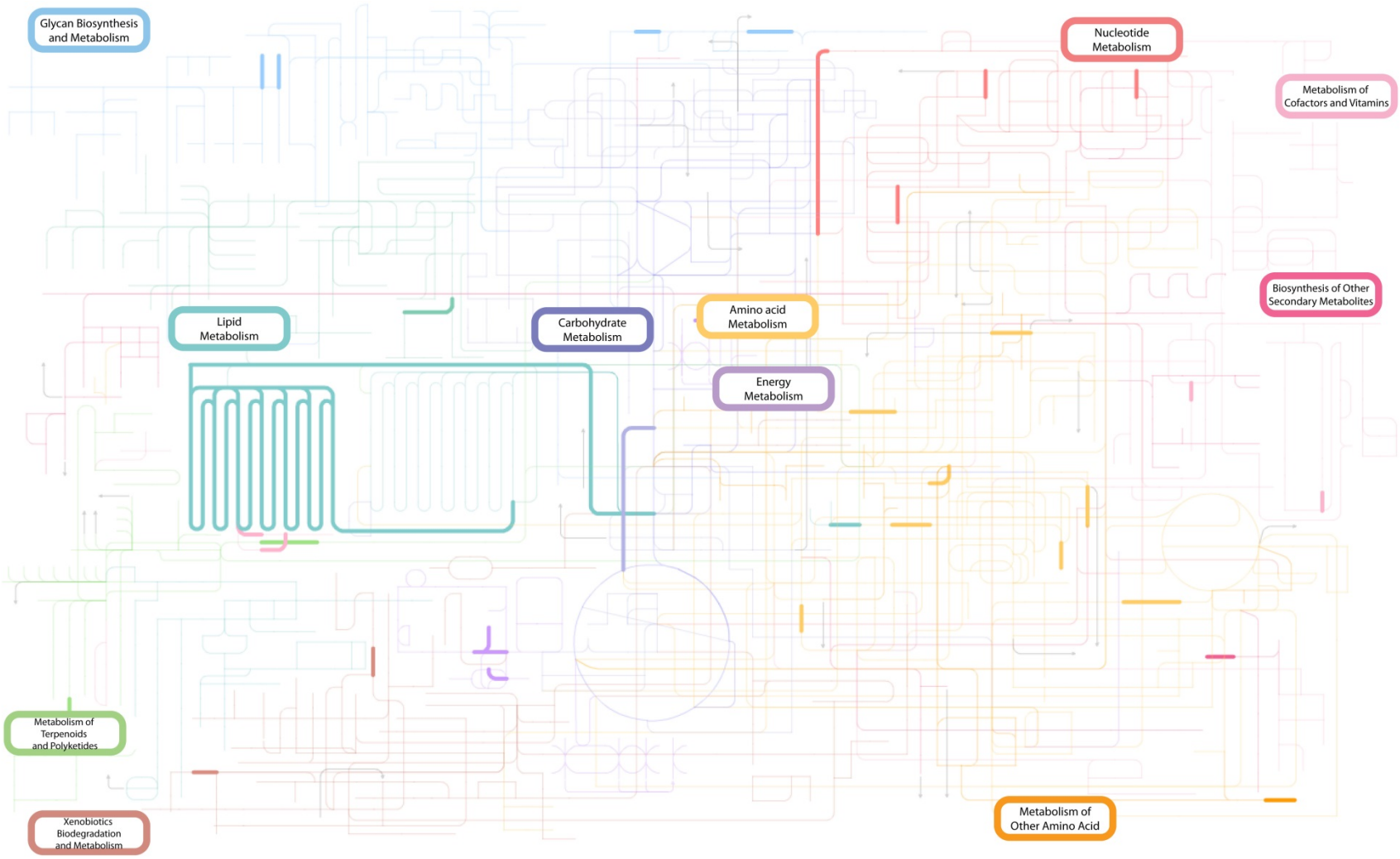


Figure 5.9. Metabolic biochemistry map of labeled proteins identified only in T_1 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

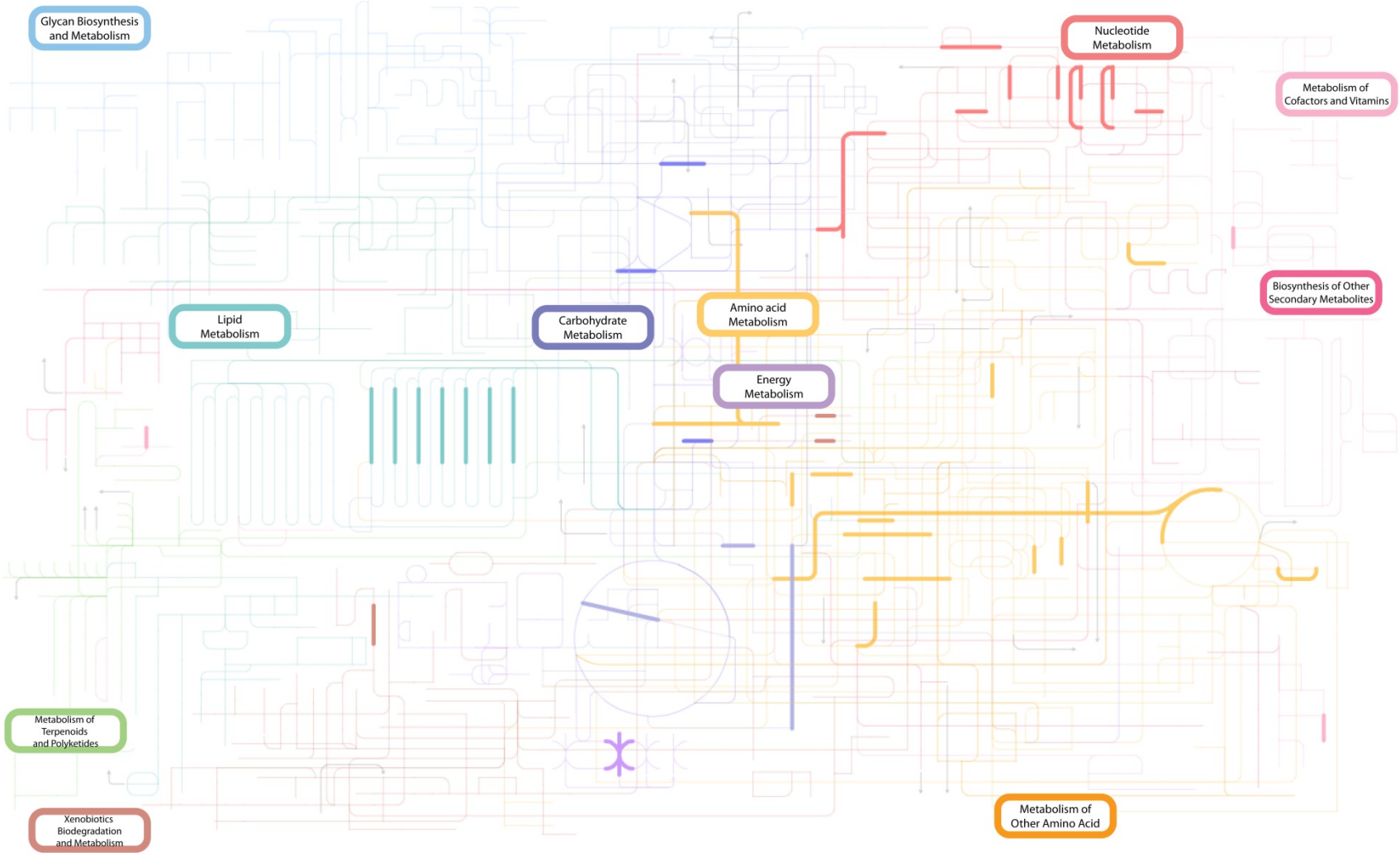


Figure 5.10. Metabolic biochemistry map of labeled proteins identified only in T_2 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

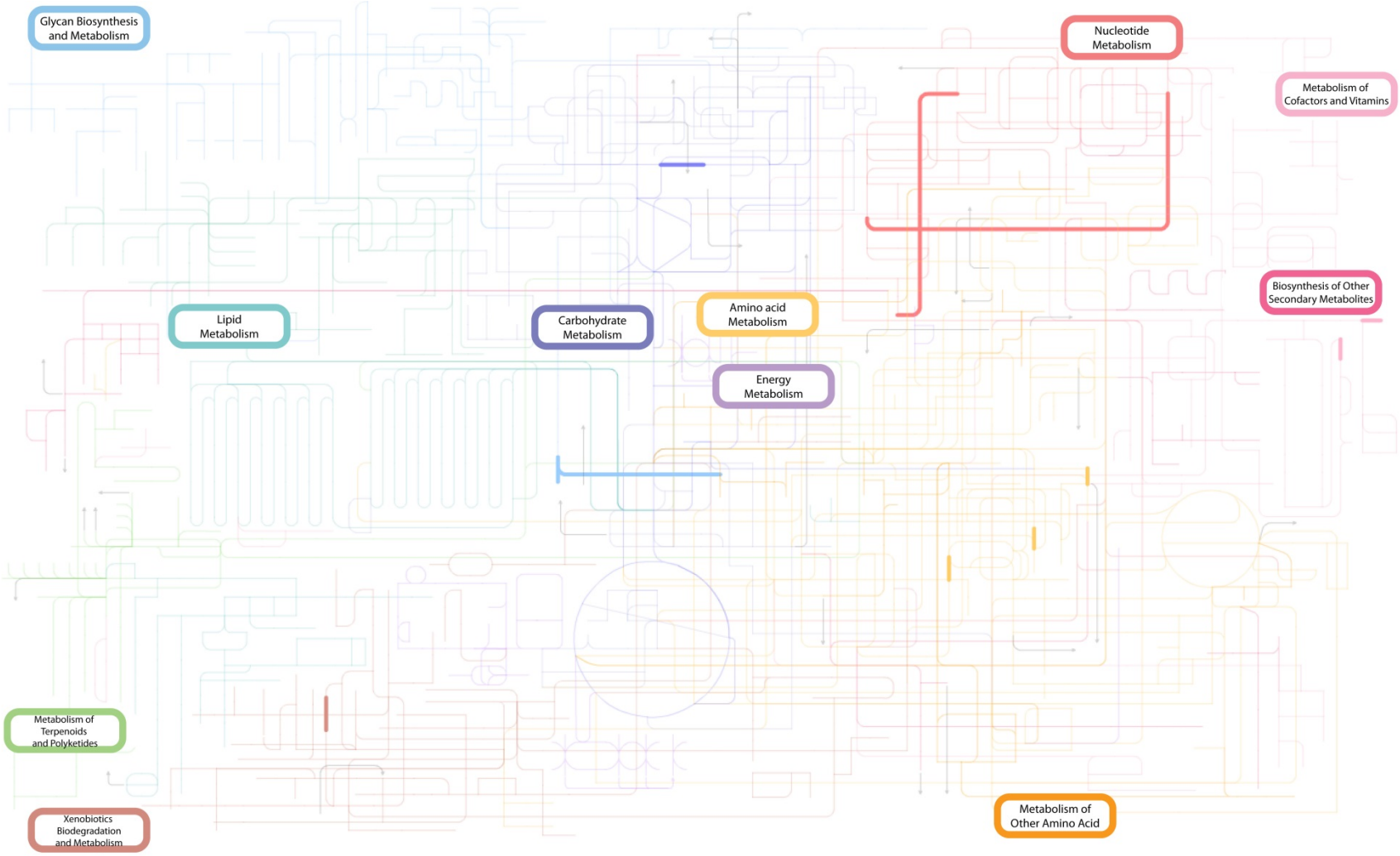


Figure 5.11. Metabolic biochemistry map of labeled proteins identified only in T_3 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

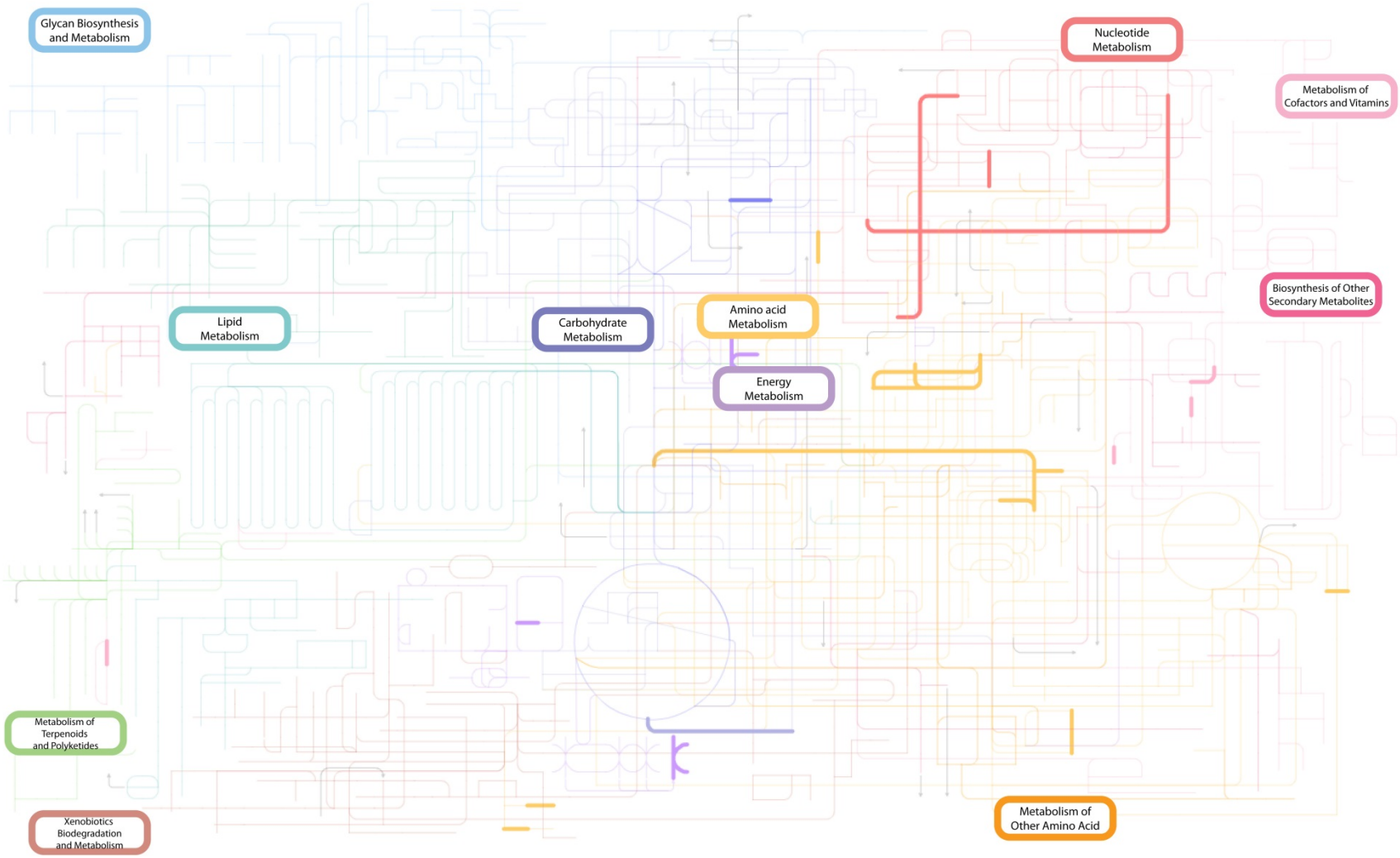


Figure 5.12. Metabolic biochemistry map of labeled proteins identified only in T_4 for *R. pomeroyi*. The pathways displayed in various colors are described in the corresponding colored boxes.

5.5. Discussion

5.5.1. Protein synthesis

There was not a large amount of variation in the protein expression among the three replicates at each time point, nor was there much variation among the time points. At each time point the replicates were either statistically indistinguishable or had a similarity of 82% or greater among the expressed proteins. As discussed in Chapter 4, where there was a similarity of 80% or greater between time points, a cluster analysis of Christie-Oleza, Fernandez, et al. (2012) indicated that the grouping of biological and technical replicates together was appropriate. The similarity between the replicates may indicate that when subjected to similar culture conditions, and potentially environmental conditions, *R. pomeroyi* is predictable in the proteins it produces. This also indicates that culture conditions were maintained among replicates with an outcome of low variability.

In total, over 1900 proteins were identified throughout the time points with 42% (808 proteins) being seen at every time point within all replicates. When examining proteins that were seen at each time point, though not in all replicates of that time point, the similarity among the four time points increases to 63%, or 1216 proteins.

The total number of proteins decreased by 274 proteins from T_1 to T_2 but then remained relatively steady through T_3 and T_4. Of the 1927 identified proteins, 479 proteins among the four time points were shown to contain the ^{13}C-leucine label. When examined by time point, 16 to 21% of the proteins per time point were labeled. Though the total proteins decreased from T_1 to T_2 the percentage of proteins labeled at the T_2 time point increased to 21% relative to 17% of the proteins labeled at T_1. This increase relative to the total proteins could be due to the second addition of label at 36 hours after

the start of the culture, two hours prior to the sampling of T_2 at 38 hours. The percentage of labeled proteins decreases again to 17% in T_3 and 16% in T_4. The subsequent decreases could be due to multiple reasons including the remineralization of some of the label to CO_2 as the culture progressed.

5.5.2. Labeled proteins by growth phase

Of the total proteins identified that carried a label, roughly one quarter, 23%, were identified at all four growth phases. Labeled proteins unique to one time point, i.e., proteins not seen at any other time points, accounted for 11 to 25% of the labeled proteins decreasing from 25% at T_1 to 11 and 13% at T_3 and T_4, respectively. Gene ontology of the unique proteins in each growth phase, show that the biological functions are similar among the phases even though the proteins are unique. The unique proteins during growth have functions that include RNA biosynthesis, translation, protein metabolism, amino acid activation and protein transport. The decrease in the number of unique proteins in the later time points (stationary phase) may be due to a decrease in the rate of cell turn over and protein synthesis so fewer numbers of proteins are needed and the variety of proteins made is decreased. The fact that the function of the unique proteins is similar indicates that protein expression and thus function are rather consistent over the experimental time period.

When comparing adjacent time points and the labeled proteins in each, the similarity in labeled proteins doesn't vary extensively. In the T_1 to T_2 (exponential phase), T_2 to T_3 (exponential to stationary phase), and T_3 to T_4 comparisons (stationary phase), the percentage of proteins in common between the compared time points varies

from 44 to 53%. The proteins similar in the T_1 to T_4 comparison (early exponential to late stationary phase) does decrease to only 37% of the proteins in common. An examination of the biological function of the proteins unique to each time point within the comparison still shows similarity of function all dealing with what would be considered standard housekeeping functions of a cell, replication, protein production and carbohydrate metabolism.

The hope of this experiment was to determine if the incorporation of an isotopic label could be tracked in the proteins produced. There was successful incorporation of the label in the identified proteins within the investigation. The findings show that there was relatively uniform incorporation of the label across all time points. Figure 5.4 is a heat map which helps with a visualization of this. As can be seen in the figure the majority of the proteins are present in all four time points (blue and green) with very few not seen (orange) in some time points. The last gene ontology category at the bottom which contains the most missing proteins has no defined gene ontology and those proteins are defined as unreviewed proteins so no inference can be made at this time as to their biological function or to any implications of their absence. This figure illustrates that there is no obvious pattern in labeling among the proteins. It would seem that no major protein groups were turned on or off by the cell during progression from exponential to stationary phase. Although no pattern could be distinguished, this might not be the case under altered conditions (e.g., stress or using a culture of specialist bacteria). Since the label was found distributed across all time points, but only added to the culture in exponential phase, the label appears to be predominantly recycled and reused.

5.5.3. Most abundant proteins

An examination of the most abundant proteins identified within the investigation showed that the four most abundant proteins had higher total spectral counts than the remaining 13 abundant proteins. The remaining proteins all had counts of similar magnitude. All of the abundant proteins had fewer spectral counts in T_3 and T_4 than in T_1 and T_2. This could be due to the fact that they represent stationary phase where cell growth and replication could have slowed causing the need for the number of copies of these proteins to decrease. The vast majority of the proteins identified as being abundant in this investigation were also labeled in at least one replicate of a time point with only one protein not carrying a label in any replicate at T_4. This indicates that the label is actively being taken up and is being used for those proteins that we can infer are important in the reproduction and propagation of these cells.

The four most abundant proteins were involved in porin activity, transport or transporter activity, binding and translation. Other abundant proteins had biological functions to participate in transcription and translation as well as protein refolding, amino acid binding, ATP binding and transport, all activities inherent in the replication of cells and production of proteins. These are similar functions to those seen in our investigation of the total proteins produced by *R. pomeroyi* during its life cycle though this is not unexpected as the culture conditions were the same between the two studies. Porin, transport and ribosomal proteins were also identified in culture by Christie-Oleza, Fernandez, et al. (2012) and transport proteins were seen in high molecular weight dissolved organic matter (HMW-DOM) (Dong et al., 2013). In addition, porins were

found in various oceanic samples (Tanoue et al., 1995; Yamada and Tanoue, 2003; Yamada and Tanoue, 2009).

Christie-Oleza and Armengaud (2010) and Christie-Oleza, Pina-Villalonga, et al. (2012) found transporter proteins to be one of the most abundant proteins identified in a study of *Roseobacter* exoproteomes. Christie-Oleza, Pina-Villalonga, et al. (2012) also determined that porins were high in some of the strains studied. Based on proteomic data, Christie-Oleza, Pina-Villalonga, et al. (2012) determined that these bacteria use one of four strategies to uptake nutrients and compete with the community. They were the production of transporters for nutrient uptake, synthesis of flagella for motility, secretion of adhesion proteins or the export of toxins. In this study only one toxin was identified though it was not labeled. This toxin was found in all four time points but it had very low spectral counts. The toxin found in this study was not one identified by Christie-Oleza, Pina-Villalonga, et al. (2012). One flagellum protein was identified as labeled in T_2, T_3, and T_4 though it was not one of the most abundant proteins. This protein was seen by Christie-Oleza, Pina-Villalonga, et al. (2012) in each of their seawater conditions, with and without the community and in all replicates. While the exoproteome was not studied in this investigation, the presence of multiple transporter proteins and their abundance relative to the other proteins may indicate that the cultures studied here were using that strategy to obtain nutrients. There is always the possibility that only one toxin and one flagellar protein were identified because they were immediately excreted and one of the other strategies may be utilized but it seems unlikely.

If the ultimate goal is to determine functional biomarkers of bacterial contribution, based on the findings of this investigation and by the studies discussed

above, porins or transport proteins might be viable options by which to try to identify the presence of bacteria. It should be noted that the findings of porins and transport proteins in oceanic samples and HMW-DOM are not necessarily limited to contributions from bacteria. Though this is the case, their presence in marine samples, as well as in bacterial cultures, support the potential use of these proteins as identifiable bacterial proteins. Investigation into their potential to hold up to deposition and degradation would need to be done but the findings of porins in oceanic samples, of which dissolved organic matter (DOM) is refractory points to their potential longevity in the environment. Transport proteins found in HMW-DOM may be an indicator for newer less degraded bacterial contributions. In Moore, Nunn, Faux et al. (2012) efforts to extract proteins from deeper, older sediments were not successful. This was due in part to the lack of genomic information on the organisms present with which to create a searchable database. Once proteins can be identified more reliably in sediment samples it can be tested to see if porins are found there as well. If this is possible they may be the first step in constructing reliable protein biomarkers.

5.5.4. Significantly up- and down-regulated proteins

Of the labeled proteins identified within this investigation, 71% were significantly up- or down-regulated at one of the time point comparisons made. It would be expected that this up- or down-regulation would give an indication of something changing within the population from early exponential to late stationary growth phases but that was not the case here. Based on the gene ontologies of the proteins up- or down-regulated, nothing can be inferred about the cell states. This was true for those proteins that were

up- or down-regulated and remained so across the growth phases as well as for those proteins which were up- or down-regulated and were unique to one time point comparison. Analysis showed that biological functions identified among the up-regulated proteins were often duplicated in the down-regulated proteins thus providing no evidence that any cell processes were being altered, enhanced, or diminished in any way. This leads to the conclusion that while the presence or absence of individual proteins may differ significantly throughout time the functions associate with those proteins don't appear to be altered across the growth phases.

5.5.5. Functional annotation of proteins

When the labeled proteins identified in this investigation were analyzed through the functional annotation software DAVID, one cluster emerged as the most prominent (highest protein counts) among all four time points. This cluster included proteins involved ribosome biosynthesis, RNA binding, translation, and protein metabolism. This is an indication that at all time points active protein synthesis is occurring and that even in late stationary phase the cells don't appear to be under significant stress or entering cell death phase as no enrichment of proteins associated with those conditions was seen.

There were two other clusters identified in all time points among the labeled proteins, though both had much lower total protein counts within all four time points than the first cluster. These two clusters contained proteins dealing with carbohydrate metabolism and glycolysis and with the citric acid cycle and acetyl-CoA metabolism. As these two clusters contain proteins that function within glycolysis and the citric acid cycle

and thereby supply the bacterium with energy it is not surprising to see them represented within a population that seems to be maintaining cell turnover.

The remaining seven clusters were not identified at all time points though they include proteins involved with housekeeping tasks within the cell such as amino acid biosynthesis, transport, translation, protein synthesis and binding. The lack of representation of proteins within these clusters at all time points serves to support what was discussed earlier in that the label was taken up and incorporated indiscriminately by *R. pomeroyi* and not only by specific proteins functioning within a narrower biological focus.

5.5.6. Protein involvement in metabolic pathways

Figure 5.8, the map illustrating the shared proteins present among all time points, shows that the main metabolic pathways on the map, nucleotide metabolism, carbohydrate metabolism, amino acid metabolism, lipid metabolism, and energy metabolism are illuminated to some extent. This provides further support of the DAVID results and the list of the most abundant proteins that indicated the *R. pomeroyi* cultures were involved in replication and growth. Figures 5.9 to 5.12 show the labeled proteins unique to time points 1 through 4, respectively. There are fewer pathways illuminated in these figures than in the map displaying the shared labeled proteins. Several of the pathways identified are similar among the time points. This demonstrates that though while there were unique proteins labeled within the time points, they are responsible for similar biological functions within the four time points with no anticipated functionalities being turned “on or off”.

5.6. Conclusion

Investigation into the proteins produced by *R. pomeroyi* grown in culture with the tracer level addition of labeled ^{13}C-leucine identified more than 1900 total proteins of which 479 were labeled. Little variation was seen among the replicates of each time point and no great differences were determined in total proteins seen among the time points. The most abundant proteins in this investigation were responsible for porins, transport, binding, and translation and almost all contained the label. The one prominent functional annotation cluster identified among the labeled proteins in all time points contained proteins that contributed to ribosomal biosynthesis, RNA binding, translation, and protein metabolism. These biological functions indicate that the cells in all time points were undergoing active turnover and the label was taken up during cell replication. There was no discernible pattern in the types of proteins which incorporated the label or to when it was incorporated, with the label being seen in very similar proportions (16 to 21% of total proteins) in all four time points. As the label was seen in all four time points but only added in exponential phase it would appear that the label is recycled by the bacteria with little remineralization. There appeared no obvious pattern to which labeled proteins were up- or down-regulated. The various different proteins that were up- or down-regulated were involved in some of the same biological processes giving no indication that any process was being altered, enhanced, or diminished in any way. This study shows the successful tracking of the uptake of an isotopically labeled amino acid and indicates that, in the future, the utilization of other, more complex, labeled substrates may provide even more information about bacterial proteins synthesis and the recycling of organic matter.

The study described here was a follow-up to the initial investigation of the expressed proteome of *R. pomeroyi* under environmentally realistic carbon conditions. In total, 2037 proteins were identified in the proteome study discussed in Chapter 4. In this study, 1927 total proteins were identified. A comparison of the two shows that 1826 proteins are the same within the two studies accounting for 90% of the identified proteins in the proteome investigation discussed in Chapter 4 and 95% of the proteins identified in this labeling investigation. While only labeled proteins were evaluated and discussed here, there were many similarities between the labeled proteins identified and the total expressed proteome. Those proteins determined to be present in highest abundance in the total expressed proteome were also the most abundant labeled proteins identified here. No patterns were distinguishable in the proteins shown to be significantly up- or down-regulated either in the total proteome or the labeled proteins. Evaluation of the biological functions associated with the identified proteins in the previous study and this one show a bacterium maintaining cell growth throughout its growth phases. The previous study of the expressed proteome has served as almost a control with which to compare this study and the similarities detailed above indicate that the addition of the label was not to an extent to alter the functioning of this bacterium.

Chapter 6: Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation

A major portion of this chapter has been published in "Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation". Nunn, B.L., Faux, J.F, Hippmann, A.A., Moldonado, M.T., Harvey, H.R., Goodlett, D.R., Boyd, P.W., Strzepek, R.F. 2013. PLoS ONE, 8(10): e75653. Doi: 10.1371/journal.pone.0075653.

6.1. Abstract

Phytoplankton growth rates are limited by the supply of iron (Fe) in approximately one third of the open ocean, with major implications for carbon dioxide sequestration and carbon (C) biogeochemistry. To date, understanding how alteration of Fe supply changes phytoplankton physiology has focused on traditional metrics such as growth rate, elemental composition, and biophysical measurements such as photosynthetic competence (F_v/F_m). Researchers have subsequently employed transcriptomics to probe relationships between changes in Fe supply and phytoplankton physiology. Recently, studies have investigated longer-term (i.e. following acclimation) responses of phytoplankton to various Fe conditions. In the present study, the coastal diatom, *Thalassiosira pseudonana,* was acclimated (10 generations) to either low or high Fe conditions, i.e. Fe-limiting and Fe-replete. Quantitative proteomics and a newly developed proteomic profiling technique that identifies low abundance proteins were employed to examine the full complement of expressed proteins and consequently the metabolic pathways utilized by the diatom under the two Fe conditions. A total of 1850 proteins were confidently identified, nearly tripling previous identifications made from differential expression in diatoms. Given sufficient time to acclimate to Fe limitation, *T. pseudonana* up-regulates proteins involved in pathways associated with intracellular protein recycling, thereby decreasing dependence on extracellular nitrogen (N), C and Fe.

The relative increase in the abundance of photorespiration and pentose phosphate pathway proteins reveal novel metabolic shifts, which create substrates that could support other well-established physiological responses, such as heavily silicified frustules observed for Fe-limited diatoms. Here, we discovered that proteins and hence pathways observed to be down-regulated in short-term Fe starvation studies are constitutively expressed when *T. pseudonana* is acclimated (i.e., nitrate and nitrite transporters, Photosystem II and Photosystem I complexes). Acclimation of the diatom to the desired Fe conditions and the comprehensive proteomic approach provides a more robust interpretation of this dynamic proteome than previous studies.

6.2. Introduction

Field studies have demonstrated that phytoplankton stocks across the world's oceans are frequently limited by Fe (e.g., Arrigo et al., 2000; Boyd et al., 2000; Coale et al., 2003; Gervais et al., 2002). The adverse effects of low Fe concentrations on primary production are well established in ~ 30% of the world's oceans, the High Nitrate Low Chlorophyll (HNLC) regions (Behrenfeld and Kolber, 1999; Boyd et al., 1996; Martin et al., 1994). The widespread Fe limitation of phytoplankton in HNLC waters has major implications for the ocean C cycle and has led to modelling efforts to link the cycling and bioavailability of Fe to atmospheric draw-down of CO_2 into the ocean (Lefevre and Watson, 1999; Sarmiento et al., 2010). More fundamental research into the biochemical basis of long-term physiological acclimation used by diatoms to survive in low Fe environments can provide insights that will facilitate better modeling of global ocean biogeochemistry.

Over the last two decades, experimental studies to better understand the role of Fe in phytoplankton physiology have used a wide range of approaches from elemental (Marchetti and Cassar, 2009) and biophysical analyses (Greene et al., 1991; Strzepek et al., 2012) to "omics" (Marchetti et al., 2012; Scheele and Holstein, 2002). A number of cellular strategies have been identified for diatoms residing in Fe-sufficient waters. For example, LaRoche et al. (1996) reported that diatoms had significantly higher ratios of the Fe-S protein ferredoxin relative to the non-ferrous flavodoxin. Recently, Whitney et al. (2011) demonstrated that the expression of these proteins is controlled by a diel periodicity. In Fe-sufficient waters, Marchetti and Cassar (2009) revealed that open ocean pennate diatoms possess the ability to capitalize on such high Fe conditions by storing excess Fe within the protein ferritin, but centric diatoms in such offshore waters do not appear to have this protein. In contrast, open ocean centric diatoms survive with extremely low cellular Fe requirements by parsimonious modifications to their photosynthetic architecture (Strzepek and Harrison, 2004).

In the last decade, several investigators have utilized two diatom genomes (*T. pseudonana* and *Phaeodactylum tricornutum*) to examine gene expression under Fe-limiting conditions for a range of physiological processes including cell-death (Scheele and Holstein, 2002), Fe acquisition (Kustka et al, 2007), N assimilation (Allen et al., 2011) and silica bioprocessing (Durkin et al., 2011) in Fe-limited cultures. Transcriptomic studies can provide a better understanding of how diatoms respond to rapid (i.e. quasi-instantaneous) Fe additions or gradual (i.e. days to weeks) decreases in dissolved Fe concentrations due to drawdown by phytoplankton (see Table 6.1). The ease of real-time PCR gene expression, a useful discovery tool for field-based molecular

indicators, is also providing information on individual gene expression across a range of Fe culture conditions (Allen et al., 2011; Kustka et al., 2007; Whitney et al., 2011).

Recent trends in "omic" examinations of physiological responses of diatoms have focused on longer-term acclimation responses to a given environmental condition. For example, recent studies have demonstrated that, given sufficient time (conventionally 10 generations) (Wood et al., 2005) diatoms and other organisms acclimate to conditions of decreasing nutrient supply that are distinct from short-term stress responses (Lacerda et al., 2007; Whitney et al., 2011). As more "omic" approaches are applied to diatoms, correlations and causations among transcriptomics, proteomics, and metabolomics can be better scrutinized. Since proteomics identifies those proteins that were made and expressed in the cell at the time of harvest, profiling this signature can provide identities of enzymes utilized in various metabolic pathways following acclimation. Although many diatom studies thus far have quantitatively examined gene expression, none have taken advantage of recent developments in proteomic profiling that reveal low abundance proteins.

This study takes advantage of these technologies thereby providing the full complement of proteins expressed and new viewpoints of the effects of acclimation on the proteome. Specifically, the aim of our study was to identify, using the proteome, physiological strategies (during mid-exponential growth) of diatoms exposed to long-term acclimation to high or low Fe conditions. We examined the proteome of cultures acclimated to specific Fe conditions for 10 generations. Proteomics can be used to decipher biochemical pathways (Aebersold and Mann, 2003; Nilsson et al., 2010) and these pathways are likely to be altered depending on whether diatoms encounter quasi-

instantaneous or gradual changes in Fe supply. In our study we used a data-independent acquisition strategy of shotgun proteomic profiling (McDonald and Yates, 2002; Nunn and Tipperman, 2007; Panchaud et al., 2009) to survey the dynamic diatom proteome and map the biochemical pathways utilized by the coastal diatom *T. pseudonana* to survive under low Fe conditions.

Table 6.1. Methodological details from some recent studies using "omics" to study Fe limitation on diatoms.

Study	Protocols	Locale or culture studied	Manipulation	Growth stage	Number of proteins identified	Number of transcripts identified
This Study	Proteomic profiling	*T. pseudonana* CCMP1335	Acclimated to Fe-replete and steady-state Fe limitation	Harvested at mid-exponential growth phase	1850	N/A
Lommer et al. 2012	Transcriptomics ; qRT-PCR; 2D SDS PAGE LC/MS/MS	*T. oceanica* CCMP1005	Fe-replete and Fe stressed cultures	Harvested at late exponential growth phase	767	Chloroplast reads: 2026 (-Fe), 14931(+Fe); Mitochondrial reads: 31261(-Fe), 18136(+Fe).
Durkin et al. 2012	Transcriptomics	Pseudo-nitzschia multiseries CLN 17	Field acclimated; Before and after Fe enrichment	Mid-exponential and stationary (nutrient limited) growth phases	N/A	Using 454-sequencing: Fe-limited surface @ Sta. P: 26; Fe (+) surface @ Sta. P: 37; Puget Sound Surface: 37. SOLiD sequencing: 0 to 375 silicic acid transporter sequence reads detected for each 454-derived SIT sequence.
Marchetti et al. 2012	Meta-transcriptomics	Field samples from low Fe waters of the northeast subarctic Pacific Ocean	Field acclimated; Before and after Fe enrichment	Line P, northeast subarctic Pacific Ocean; took samples when Fe-limited and after 98 hr Fe addition incubation	N/A	Transcripts for 2845 genes were differentially expressed in +Fe vs. ambient; 3888 in +Fe vs. control.
Whitney et al. 2011	qRT-PCR; gene expression	*T. pseudonana* CCMP1335, *T. weissflogii* CCMP1010	Acclimated to Fe-deplete and Fe-replete, some with copper limitation	Fe-limiting acclimation or rapid Fe stress (cultures from Fe-replete media were transferred to media with no added Fe)	N/A	No value provided for numbers of transcripts.
Allen et al. 2011	PCR and Western blots	*Phaeodactylum tricornutum* CCMP2561	Acclimated to Fe-replete and Fe-deplete with varied N sources		212	228 contigs were identified as differentially up-regulated in Fe-limited treatment – these were assembled to 212 predicted proteins.

Mock et al. 2008	Transcriptomics and proteomics	*T. pseudonana* CCMP1335	Nutrient replete, Si and Fe stressed cultures	Early-stationary growth phase	349	Transcripts for >8000 predicted genes; 75- 84 genes induced by various conditions
Kustka et al. 2007	qRT-PCR	*T. pseudonana* CCMP1335 *Phaeodactylum tricornutum* CCMP630	Acclimated to Fe limitation; Fe-resupply	Steady-state mid-exponential growth phase	N/A	The abundance of specific gene transcripts relative to a housekeeping gene was reported; no values provided.

qRT-PCR = quantitative reverse transcription polymerase chain reaction; 2D SDS PAGE LC/MS/MS= 2 dimensional SDS PAGE gel electrophoresis followed by tandem mass spectrometry protein identifications on individual gel spots

6.3. Materials and Methods

6.3.1. Acclimation of T. pseudonana culture

The diatom *T. pseudonana* was selected as the study organism; it has a well characterized physiology, genome, and proteome (Arnbrust et al., 2004; Carvalho and Lettieri, 2011; Dyhrman et al., 2012; Nunn et al., 2009). *T. pseudonana* clone 3H (CCMP1335, mean diameter 4 μm) was obtained from the Provasoli-Guillard Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME, USA). Cells were grown axenically using semi-continuous batch culturing and the chemically well-defined artificial seawater medium AQUIL (Price et al., 1988; Sunda et al., 2005). The cultures were grown in 28 mL polycarbonate tubes and were acclimated to either Fe-limiting (pFe 20.5, where pFe = -log $[Fe^{3+}]$ and $[Fe]_{total}$ = 42 nmol L^{-1}) or Fe sufficient conditions (pFe 19, $[Fe]_{total}$ = 1.37 μmol L^{-1}), with speciation calculated using MINEQL (Westall et al., 1976). The AQUIL medium used in this study included 100 μmol L^{-1} EDTA to buffer the trace metals, and was prepared with a chemical composition identical to that described by Maldonado et al. (2006). All cultures were maintained at 19 ± 1°C and a continuous light intensity of 150 μmol quanta m^{-2} s^{-1}, provided with cool-white fluorescent lights. The growth rates (d^{-1}) of the cultures were monitored daily using in vivo chlorophyll a (Chl) fluorescence measurements with a Turner Designs AU-10 Fluorometer (Sunnyvale, CA, USA). Cell density (cells mL^{-1}) and cell volume (fL $cell^{-1}$; fL = femtoliter = 10^{-15} L) were determined on freshly harvested cells using a Coulter Z2 Particle Count and Size Analyzer. Sterile trace metal-clean techniques were used during all experiments and manipulations.

A Phyto-PAM fluorometer (Walz) equipped with a Phyto-ED emitter-detector unit was used to measure the maximum photochemical efficiency of photosystem II (F_v/F_m). All fluorescence measurements were made once algal samples had been dark-acclimated for 30 – 45 minutes at 19 ± 1°C. The fluorometer was programmed to acquire multiple turnover saturations of PSII using a 200 – 300 ms saturation flash applied at 30 s intervals. The inter-pulse interval was determined from the minimum time required for fluorescence to relax to pre-saturation levels. The primary signals (F_o and F_m) measured by the Phyto-PAM were obtained by excitation of the sample with measuring light beams at 4 different wavelengths (470, 520, 645 and 665 nm). The mean value of these 4 signals was calculated for each saturation flash, and a minimum of 10 saturation flashes was averaged for each sample. A Water-S stirring device (Walz) was used to keep samples suspended, and was shut off 10 seconds before applying each saturation flash to minimize signal noise. Differences in growth rates, F_v/F_m, and cell volumes between Fe treatments were assessed using a two-tailed T-test ($p < 0.05$) following Levene's Test for homoscedasticity.

Once the cells were acclimated to the specific Fe levels (growth rates in 10 successive transfers varied by less than 15%, Brand et al. 1981), they were used to inoculate 1 L cultures in polycarbonate bottles. These large volume cultures were monitored daily by measuring cell density and volume, as well as Chl fluorescence. At mid-exponential phase, cells were harvested by filtration onto 47 mm diameter polycarbonate filters, resuspended in a small volume of media and pelleted by centrifugation (4˚C, 7,500 x g, 5min) into four separate Eppendorf tubes. The pellets were frozen in liquid nitrogen and stored at -80 ˚C.

6.3.2. Proteomic sample preparation

Details of cellular preparations for proteomic analyses can be found in Nunn et al. (2009). Briefly, 6 M urea was added to the cell pellets and they were lysed using a titanium micro-probe sonicator. Between each sonication event of 10-15s, samples were immersed in liquid nitrogen to rapidly freeze them. After 10 sonication events, disulfide bonds were reduced with dithiothreitol and alkylated with iodoacetamide. Ammonium bicarbonate was added to dilute the urea prior to the addition of MeOH. The combination of urea, a strong denaturing agent, and methanol helped to solubilize membrane proteins. No cellular or chemical fractionations were conducted prior to Trypsin digestions in order to reduce protein loss. Each sample received Trypsin at an enzyme: protein ratio of 1:50, vortexed, and incubated on a Thermomixer (800 rpm) overnight at 37^{o}C. Samples were desalted using a macro-spin C18 column (NestGroup) following the manufacturers guidelines prior to analysis by mass spectrometry (MS). Peptide concentrations were measured on each sample using the Thermo Scientific NanoDrop 2000/2000c Spectrophotometer. The peptide bond absorbance was monitored at 205 nm UV wavelength and samples were diluted to yield a final concentration of 100 μg protein mL^{-1}.

6.3.3. Mass Spectrometry

Samples were separated and introduced into the mass spectrometer (MS) by reverse-phase chromatography using an 15 cm long, 75 μm i.d. fused silica capillary column packed with C18 particles (Magic C18AQ, 100 Å, 5μm; Michrom, Bioresources,

Inc., CA) fitted with a 2 cm long, 100 µm i.d. precolumn (Magic C18AQ, 200Å, 5µ; Michrom). Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-acetonitrile gradient (5-35% acetonitrile in 60 min). Mass spectrometry was performed on two Thermo Fisher (San Jose, CA) hybrid tandem mass spectrometers: the linear ion trap Velos (LTQ-VELOS) and the linear ion trap –Orbitrap (LTQ-OT). Based on peptide concentrations, a total of 1 µg of peptide digest in 10 µl of 5% ACN, 0.1% formic acid was sampled per LC-MS analysis. For quantitative analyses, the 4 biological splits from Fe-limited (Tp_1^-, Tp_2^-, Tp_3^-, Tp_4^-) and Fe-replete(Tp_1^+, Tp_2^+, Tp_3^+, Tp_4^+) diatoms were analysed on the LTQ-OT with four gas phase fractionations using data-dependent acquisition , as is outlined in Nunn et al. (2009), culminating in a total of 16 analyses per condition. As previously mentioned, no chemical or cellular fractions were completed on the whole cell lysates. Gas phase fractionations in the mass spectrometer provide the user with an accurate means for isolating peptides based on the *m/z* ratio with no sample loss. Gas phase fractions (GPF) selected were optimized using the *Thalassiosira pseudonana* genome to predict the best *m/z* windows (350-444, 444-583, 583-825, 825-1600 *m/z*) (Scherl et al., 2008). This set of quadruplicate GPF analyses on the LTQ-OT provided a dataset from which statistical confidence could be applied to determine significantly up- or down-regulated proteins with respect to the alternate cell state using *QSpec* (see below).

In addition, the Fe-replete (Tp_1^+) and Fe-limited (Tp_1^-) cell lysates were analyzed on the LTQ-VELOS using a data-independent analysis known as the PAcIFIC method (Panchaud et al., 2009). This method is ideal for examinations of low abundance proteins and yields a rapid automated profile of all proteins expressed at the time of harvest (down

to an estimated copy number of 100 proteins per cell) (Panchaud et al., 2009). This dataset provided the means to interrogate the presence or absence of full metabolic pathways. Rather than requiring the mass spectrometer to select ions for fragmentation based on its observation in the MS1 scan, it is programmed to fragment all ions in a narrow range of *m/z* units. Slight modifications to the published PAcIFIC method were made to adapt the protocol to the LTQ-VELOS' faster duty cycle. Each method file included a 60 min linear HPLC gradient of 5-35% ACN and the MS covered a 31.5 *m/z* mass range, rather than 15 *m/z* units. This resulted in a total of 32 method files per PAcIFIC analytical cycle, covering a full *m/z* range of 400-1400 *m/z*. Triplicate PAcIFIC analytical cycles were completed on both the Fe-replete (Tp_1^+) and Fe-limited (Tp_1^-) samples, yielding a total of 96 MS experiments.

Confirmation of the up- or down-regulation of a pathway was confirmed using the quadruplicate, statistically rigorous, LTQ-OT dataset. The PAcIFIC dataset provides information on the presence or absence of a larger number of proteins and can provide spectra from low abundant proteins, but the quadruplicate sets of 4 GPF analyses provided data for spectral counting with statistical confidence.

6.3.4. Protein Database Searching and Mass Spectrometry Data Interpretation

All tandem mass spectrometry results were searched and interpreted with an in-house copy of SEQUEST (PVM v.27 20070905) following the parameters outlined in Nunn et al. (2009). The protein database used for correlating spectra with protein identifications was generated by combining the latest release version 3.0 of the nuclear *T. pseudonana* predicted protein database (www.jgi.doe.gov), the unmapped sequences

(Thaps3_bd; www.jgi.gov), 302 *T. pseudonana* proteins from the PubMed Entrenz Protein database, and 50 common contaminants. SEQUEST parameters included: reverse concatenated sequence database search, no enzyme specificity, cysteine modification of 57 Da (resulting from the iodoacetamide), and modifications on methionine of 15.999 Da (oxidation). Minimum protein and peptide thresholds were set at p> 0.95 on ProteinProphet and PeptideProphet (Keller et al., 2002). The SEQUEST criteria for a doubly charged peptide used a correlation factor (Xcorr) >2.5, a cross-correlation factor ΔCorr >0.1 and for a triply charged peptides the Xcorr minimum was 3.5. Protein identifications from the whole-cell lysates were accepted by ProteinProphet if the above mentioned thresholds were passed, two or more peptides were identified (PeptideProphet), and at least one termini was tryptic. Using concatenated target-decoy database searches, false-discovery rates (FDR) were calculated according to Elias and Gygi (2007) and were all <1%.

6.3.5. Label-free protein Quantification (QSpec)

For protein quantification, 4 biological splits were collected from both the Fe-limited and Fe-replete cultures. Each biological split received four 90-minute gas phase fraction analyses on the LTQ-OT, resulting in 16 tandem MS experiments per culture condition. This dataset was used in order to determine relative quantities of protein expression from the Fe-replete and Fe-deplete cultures. The common method of spectral counting was employed to determine relative quantities (Chen et al., 2008; Choi et al., 2008; Old et al., 2005). Spectral counting sums up the number of identified peptide tandem mass spectra resulting from a specific protein in order to estimate abundance of

that protein relative to other proteins in the sample (Liu et al., 2004). Final spectral counts of each protein result from the accumulation of all spectral counts derived from all 4 GPF analyses per analytical set. Spectral counting data were filtered at $p > 0.95$ protein and peptide probability using PeptideProphet; proteins with only one peptide identified were excluded. Significance analyses were completed using *QSpec* in order to determine if proteins were significantly up- or down-regulated based on the quadruplicate sets of 4 GPF analyses. *QSpec* was designed specifically for interpreting differences in protein populations determined from tandem mass spectrometry spectral counts (Choi et al., 2008). *QSpec* takes into account the size of the protein and normalizes total spectral counts achieved per analytical set if needed. *QSpec* is reported using a fold change difference in abundance with a log base 2 scale. This provides an easy way to examine the data because a reported positive fold change indicates up-regulation in Fe-limited cells (e.g. +1 fold means twice, or 2^1, as many spectra were observed in the Fe-limited relative to the Fe-replete cells), and a negative fold change indicates down-regulation of a protein in Fe-limited cells. A reported fold change of zero indicates no significant difference was measured between the quadruplicate data sets from Fe-limited and Fe-replete diatoms. Proteins were considered to be up- or down- regulated if the reported Bayes Factor was > 10, the corresponding FDR < 1%, and the fold difference observed was >0.5.

6.3.6. iPath: metabolic pathway mapping

To better illustrate the potential physiological implications of the changes to the *T. pseudonana* proteome, we mapped expressed proteins in the context of overall

metabolic pathways using iPath software (http://pathways.embl.de) (Letunic et al., 2008; Yamada et al., 2011). Each protein's observed presence was determined from the LTQ-VELOS-generated PAcIFIC data. Approximately 50 % of the proteins in the *T. pseudonana* proteome have iPath identifiers, such as enzyme commission numbers (EC numbers), or associated KEGG pathways (Kyoto Encyclopaedia of Genes and Genomes) and can be mapped using iPath software.

6.4. Results

Iron-limited cells grew significantly slower (0.68 ± 0.26 d^{-1}; mean ± standard error) than Fe-replete cells (1.71 ± 0.25 d^{-1}) ($p < 0.05$, two-tailed t-test; Table 6.2). The photochemical efficiency of PSII, F_v/F_m, was lower in Fe-limited cells (0.54 ± 0.01) compared to Fe-replete cells (0.67 ± 0.01; Table 6.2). Fe-limited cells of *T. pseudonana* had significantly lower specific growth rates and F_v/F_m, and were significantly smaller, than Fe-replete cells ($p < 0.05$) (Table 6.2).

A total of 1850 proteins (containing 2 or more peptides) were identified from the combined Fe-limited and Fe-replete datasets, providing the largest expressed protein dataset of diatoms to date (see Table S1: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653). Of the 1257 identifications from Fe- replete cells, 85% overlapped with the Fe-limited diatoms, revealing 188 and 593 proteins uniquely expressed under Fe-replete and Fe-limiting conditions, respectively (Fig. 6.1).

Our GPF dataset revealed 131 proteins that were significantly up- or down-regulated (*QSpec*; threshold of >10 Bayes factor). *QSpec* identified 77 proteins to be up-

regulated in Fe-limited cells, and 54 to be up-regulated in Fe-replete cells, by more than a 0.5 fold change in $\log_2$ scale (i.e. 0 fold change indicates no difference in abundance was measured; Table A6.2). Unlike transcript regulation, quadruplicate datasets of protein expression with >0.5 fold difference is considered a significant difference in protein abundance (Choi et al., 2008; Lommer et al., 2012). Gene ontology categories represented by the up-regulated proteins from each cell culture condition in Table A6.2 were analyzed using DAVID software to identify which metabolic pathways had a significant number of proteins up-regulated and were associated with a specific physiological process and/or biochemical pathway (Table 6.3 and 6.4). MS analyses on Fe-replete cells indicate that 20 of the proteins up-regulated (*QSpec*-analysis) were involved in translation and 19 were involved in photosynthesis and light harvesting (Table 6.3). Other well-represented biochemical categories in Fe-replete cells included ion transporters, macromolecular biosynthetic processes, and proteins involved in aspects of genetic information and processing other than translation. Although Fe-limited cells up-regulated more proteins, the metabolic pathways associated with these proteins were more diverse, resulting in fewer, well-represented pathways. When acclimated to Fe-limited conditions, *T. pseudonana* up-regulated proteins involved in the metabolic processing of sugars, amino acid metabolism, and protein transport and localization (Table 6.4).

Characterization of metabolic pathways was improved by identifying low abundance proteins using PAcIFIC mass spectrometry (Fig. 6.2 and 6.3). Twice as many protein identifications from the lipopolysaccharide biosynthesis pathway were identified in the Fe-limited cultures (14 proteins) compared to the Fe- replete cultures (7 proteins;

see Table S3: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653). A few of the proteins in lipopolysaccharide biosynthesis pathway were also significantly up-regulated in Fe-limited cultures (UDP-glucose 6-dehydrogenase, +1.8 fold and UDP-glucose pyrophosphorylase, +1.23 fold; Table A6.2). Other critical enzymes in the pathway were uniquely identified in Fe-limited cells, such as UDP-sugar pyrophosphorylase UDP-glucose 4 epimerase. The pentose phosphate pathway was also better represented (twice as many proteins) in the Fe-limited compared to Fe-replete cultures (see Table S4: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653, Supporting Fig. 6.1A and 6.1B). Transketolase, an enzyme involved in pentose phosphate pathway and C fixation, was up-regulated ~2-fold in Fe-limited cells, supporting the pathway analysis (Table A6.2, see Table S4: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653 and Fig. 6.3). The rate-controlling enzyme for the pentose phosphate pathway, glucose-6-phosphate dehydrogenase, was only confidently identified in Fe-limited cells. Many of the metabolites resulting from the pentose phosphate pathway enter glycolysis, and although all the glycolysis enzymes are constitutively expressed in Fe-replete and Fe-limited cultures, several of the key enzymes were up-regulated in the Fe-limited cultures (Table A6.2).

The eight glycolysis-specific proteins that are up-regulated include fructose-1,6-bisphosphate aldolase precursor (+1.78 fold), glucose-6-phosphate isomerase (+1.75 fold), fructose bisphosphate aldolase (+1.65 fold), phosphofructokinase (PFK-1;+1.50

fold), enolase (+1.37 fold), fructose-1,6-bisphosphate aldolase precursor (+0.65 fold), glyceraldehydes-3-phosphate dehydrogenase precursor (GADPH; +0.6 fold), and phosphoglycerate kinase precursor (+0.5 fold) (Table A6.2 and see Table S5: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653).

In the Fe-limited cultures, 35 more proteins from the amino acid biosynthesis and degradation pathways were identified (total of 112 proteins; Fig. 6.3 and see Table S6: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653). In particular, 15 proteins involved in the biosynthesis of phenylalanine, tyrosine, and tryptophan were identified in Fe-limited cultures, whereas only 5 enzymes involved this pathway were identified in the Fe-replete cultures. PAcIFIC analyses identified two aminotransferases unique to the Fe-limited cultures, and two additional aminotransferases were also statistically determined to be up-regulated in the GPF dataset (Table A6.2). Many proteins involved in amino acid metabolism are also associated with N metabolism (Table 6.5). Nitrate and nitrite transporters were identified in both cell states, in addition to urea transporters and the urease enzyme (Table 6.5). The Ni-ABC transporter was, however, the most down-regulated protein in Fe-limited cells in this study (-2.87 fold; Table A6.2). Ni-ABC transporters utilize the ATP-binding cassette to help energetically transport Ni across membranes. The nickel transported into the cell is an essential co-factor for the urease enzyme. NADPH nitrite reductase was confidently identified in both cell states (20 unique peptides identified in each cell state), whereas the ferredoxin nitrite reductase was only identified in the Fe-limited cultures (6 unique peptides in Fe-limited diatoms only). In the case of all of other N metabolic enzymes, a greater number of unique peptides were identified in all analyses of the Fe-replete cultures.

All proteins involved in light-harvesting antennae, photosynthetic C fixation and ATP synthase were constitutively expressed (see Table S7: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653). However, many of those proteins were identified to be down-regulated in the Fe-limited cultures. A total of 9 PSII complex proteins (Psb A, B, C, D, E, F, H, V, Y) and five photosystem I proteins (Psa B, D, E, F, L) were identified in both Fe-replete and Fe-limited conditions. Photosystem proteins that were identified in Fe-limited cultures yielded lower peptide spectral counts in all subunits except Cytochrome b_{559} (PsbF) and PSII reaction center D2 (PsbD) (Fig. 6.4). Quantitative analyses revealed that 3 subunits of the chloroplast-localized ATP synthase complex (AtpA, AtpE, and AtpG), the Cytochrome b_6f complex (PetA, PetB), photosystem I ferredoxin binding complex (PsaD), and PSII reaction center D1 (PsbA) were all down-regulated (>0.5 fold) in the Fe-limited cultures (Table A6.2). Many of these proteins have direct Fe requirements (e.g. heme Fe and Fe-S proteins of the Cyt b_6f complex (Pet A, Pet B, Pet C), or are associated with Fe within a complex (e.g. photosystem I reaction center protein PsaD binding ferredoxin, and PSII reaction center protein D1 associated with a non-heme Fe). Nine fucoxanthin chlorophyll *a/c* binding proteins (FCP) were identified in both cultures, one was significantly up-regulated in Fe-limited cultures, and two were significantly down-regulated (Table A6.2). The FCP that was up-regulated in Fe-limited cultures has been associated with light mediated oxidative stress (Lommer et al., 2012).

All proteins for the uptake and fixation of CO_2 were present in the diatoms under both Fe conditions (see Table S8: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653). The

large subunit of RuBisCO, 1,5-bisphosphate carboxylase/oxygenase (RbcL), responsible for catalyzing the first step of C fixation, was observed to be up-regulated in Fe-limited diatom cultures (Table A6.2). Because RuBisCO can use either CO_2 or O_2 as a substrate, many photosynthetic organisms, including diatoms, employ a C concentrating mechanism to increase the catalytic efficiency of RuBisCO toward CO_2. Phosphoenolpyruvate carboxylase (PEPC2), which utilizes bicarbonate in order to carboxylate phosphoenolpyruvate into oxaloacetate, was down-regulated in Fe-limited cultures (-1.24 fold; Table A6.2).

Table 6.2. Specific growth rates (d^{-1}), photochemical efficiency of photosystem II (F_v/F_m), and cell volumes of Fe-replete and Fe-limited cultures of the diatom *T. pseudonana* CCMP1335 used for proteomic analyses.

Treatment	Specific growth rate (d^{-1})	F_v/F_m	Cell volume (fL $cell^{-1}$)
Fe-replete	1.71 ± 0.25	0.67 ± 0.01	5.47 ± 0.14
Fe-limited	0.68 ± 0.26	0.54 ± 0.01	3.93 ± 0.08

Values are mean ± standard error; fL = femtoliter = 10^{-15} liter

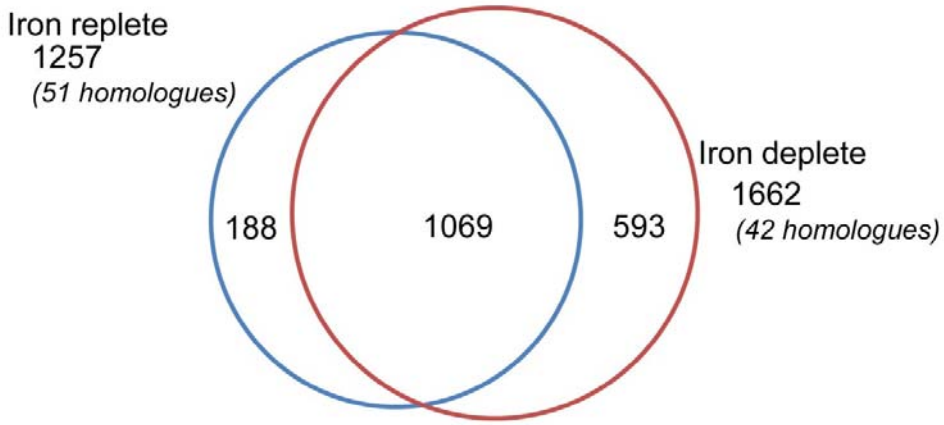


Figure 6.1. Venn diagram of number of proteins identified in Fe-replete and Fe-limited *T. pseudonana*. Proteins results presented were confidently identified from triplicate PAcIFIC analyses on the LTQ-VELOS. Fe-replete (blue) and Fe-limited (red) conditions were harvested at mid-exponential growth phase after acclimation. Numbers parenthetically annotated indicate homologous protein identifications.

Table 6.3. Biological processes up-regulated in Fe-replete *T. pseudonana* as reported by DAVID Biological Process Term level 4 analysis of gene ontology categories.

Biological Process Term	Count	%	P-Value
translation	20	30.8	1.2E-09
generation of precursor metabolites and energy	15	23.1	1.2E-07
photosynthesis	11	16.9	2.1E-06
photosynthesis, light reaction	8	12.3	1.8E-05
proton transport	7	10.8	3.2E-05
hydrogen transport	7	10.8	3.2E-05
ATP biosynthetic process	7	10.8	1.7E-04
ATP metabolic process	7	10.8	1.7E-04
energy coupled proton transport, down electrochem. gradient	6	9.2	2.3E-04
ATP synthesis coupled proton transport	6	9.2	2.3E-04
ion transmembrane transport	6	9.2	2.6E-04
purine nucleoside triphosphate biosynthetic process	7	10.8	2.7E-04
purine ribonucleoside triphosphate metabolic process	7	10.8	2.7E-04
ribonucleoside triphosphate biosynthetic process	7	10.8	2.7E-04
purine nucleoside triphosphate metabolic process	7	10.8	2.7E-04
nucleoside triphosphate biosynthetic process	7	10.8	2.7E-04
ribonucleoside triphosphate metabolic process	7	10.8	2.7E-04
purine ribonucleoside triphosphate biosynthetic process	7	10.8	2.7E-04
nucleoside triphosphate metabolic process	7	10.8	2.9E-04
plasma membrane ATP synthesis coupled proton transport	3	4.6	6.8E-04
oxidative phosphorylation	6	9.2	7.0E-04

Count: total proteins up-regulated in Fe-replete cultures that correlated with the Biological Process; %: Percent of total proteins associated with that term; P-value: probability that the number of proteins identified to be up-regulated from that biological process is significant with respect to the total number of proteins from the *T. pseudonana* proteome associated with that process (reported p-value threshold <0.05)

Table 6.4. Biological processes up-regulated in Fe-limited *T. pseudonana* as reported by DAVID Biological Process Term level 4 analysis of gene ontology categories.

Biological Process Term	Count	%	P-Value
hexose metabolic process	6	7.8	1.1E-03
monosaccharide metabolic process	6	7.8	1.2E-03
glucose metabolic process	5	6.5	4.7E-03
glycolysis	4	5.2	6.8E-03
aspartate family amino acid biosynthetic process	3	3.9	1.2E-02
hexose catabolic process	4	5.2	1.4E-02
glucose catabolic process	4	5.2	1.4E-02
monosaccharide catabolic process	4	5.2	1.4E-02
aspartate family amino acid metabolic process	3	3.9	1.7E-02
cellular carbohydrate catabolic process	4	5.2	2.0E-02
alcohol catabolic process	4	5.2	2.0E-02
cellular metabolic compound salvage	2	2.6	3.6E-02
asparagine metabolic process	2	2.6	3.6E-02
asparagine biosynthetic process	2	2.6	3.6E-02
establishment of protein localization	5	6.5	3.9E-02
protein transport	5	6.5	3.9E-02
protein localization	5	6.5	4.6E-02
intracellular protein transport	4	5.2	4.9E-02
generation of precursor metabolites and energy	6	7.8	5.2E-02
cellular protein localization	4	5.2	5.6E-02
cellular macromolecule localization	4	5.2	5.6E-02
carbohydrate catabolic process	4	5.2	6.2E-02
intracellular transport	4	5.2	7.6E-02

Count: total proteins up-regulated in Fe-limited cultures that correlated with the Biological Process; %: Percent of total proteins associated with that term; P-value: probability that the number of proteins identified to be up-regulated from that biological process is significant with respect to the total number of proteins from the *T. pseudonana* proteome associated with that process (reported p-value threshold <0.05)

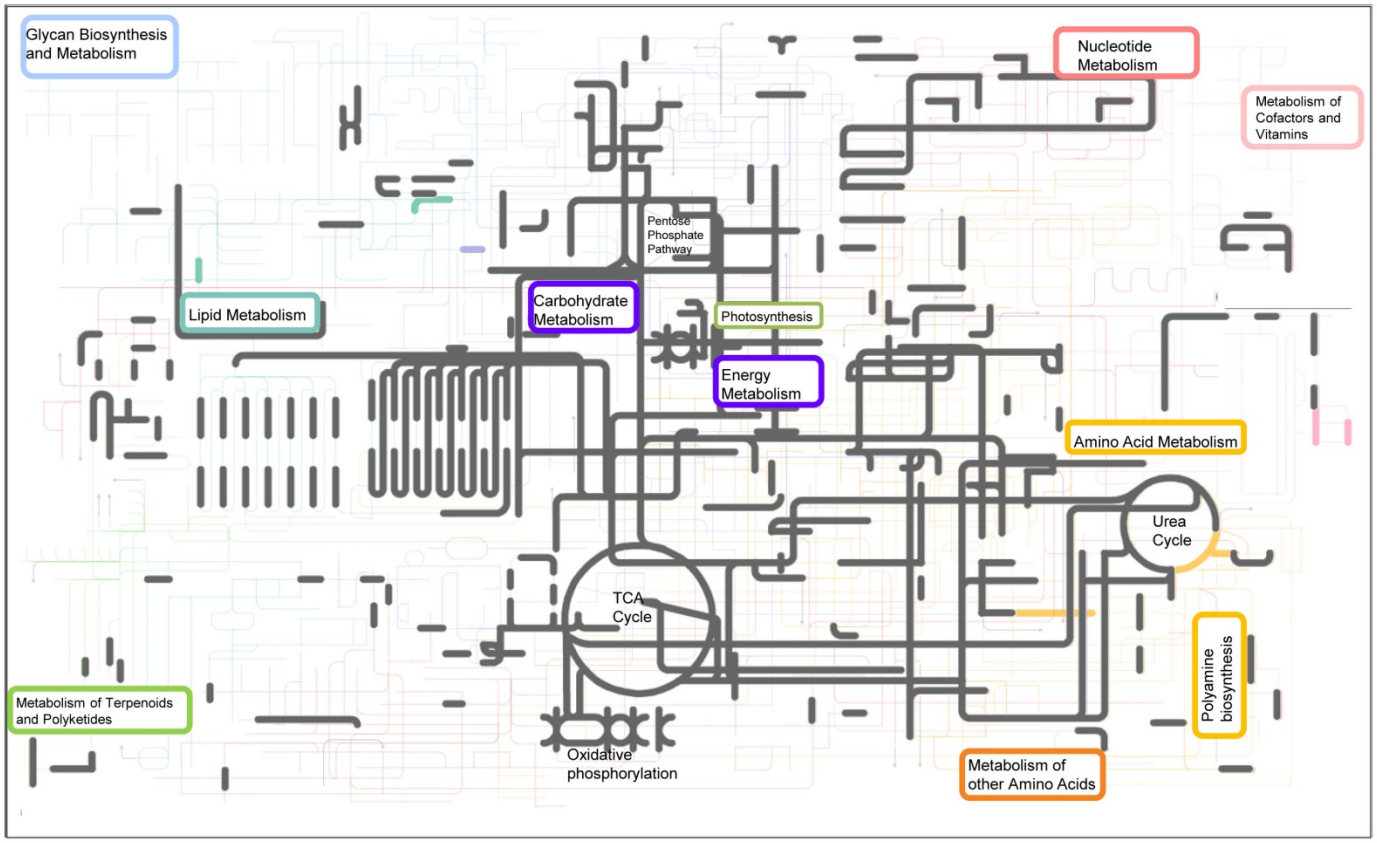


Figure 6.2. Metabolic biochemistry map of proteins expressed and identified in Fe-replete *T. pseudonana*. Map includes data from triplicate PAcIFIC analyses on a tandem mass spectrometer from *Thalassiosira pseudonana* acclimated to Fe-replete conditions. Each node (or corner) represents a metabolite and the lines connecting the nodes represent an enzyme (i.e. protein). Metabolites were not measured in this study. Proteins that were identified in both Fe-replete and Fe-limited cultures are highlighted in grey. Proteins that were identified to be unique to the Fe-replete cultures are indicated in color. From top left – light blue: sugar and glycan biosynthesis, light purple: starch and sucrose metabolism (including photosynthesis, oxidative phosphorylation, carbon fixation), dark purple: glycolysis-gluconeogenesis (including the citric acid cycle), red: nucleotide metabolism, teal: lipid metabolism, orange: amino acid metabolism (including urea cycle).

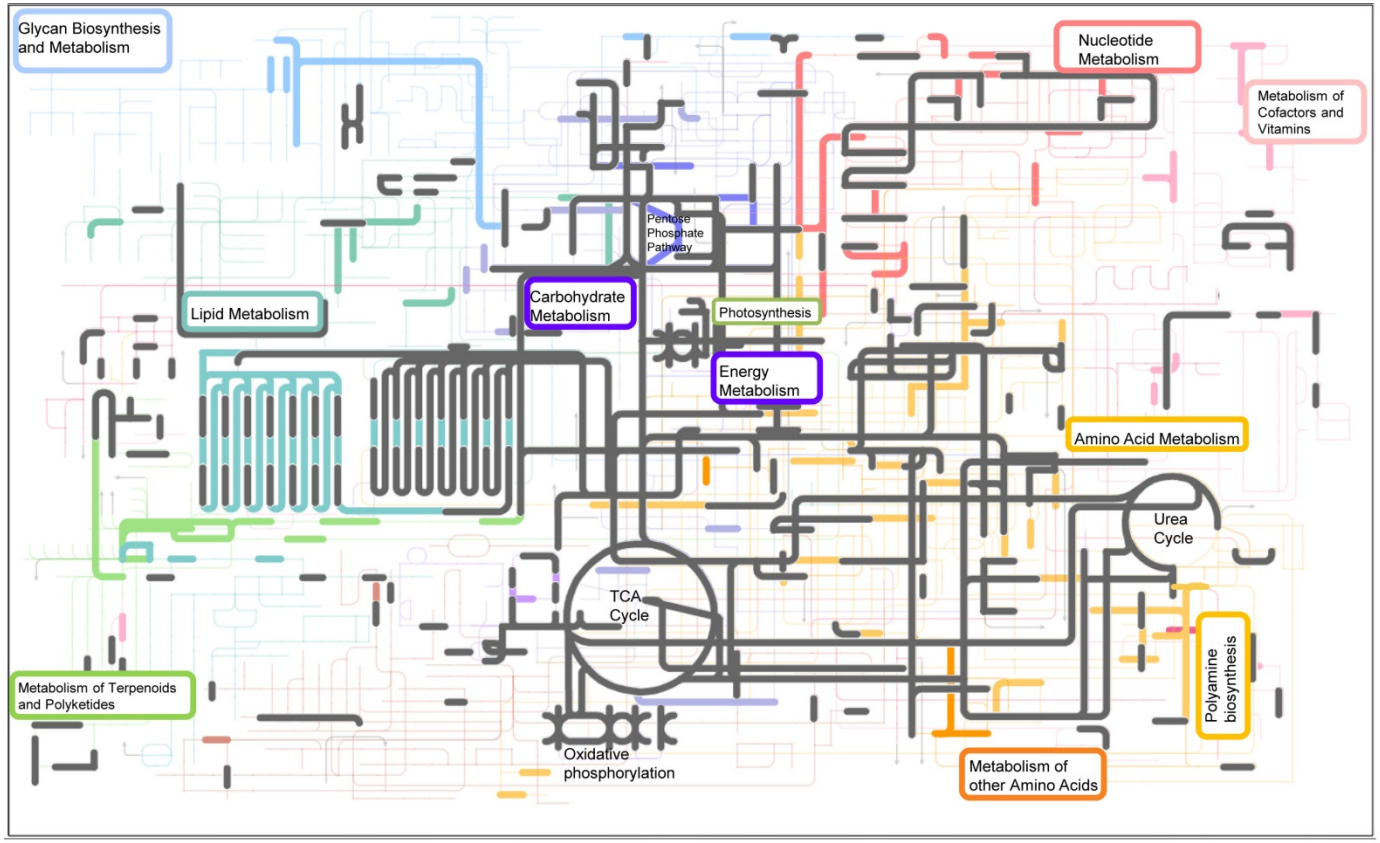


Figure 6.3. Metabolic biochemistry map of proteins expressed and identified in Fe-limited *T. pseudonana*. Map includes data from triplicate PAcIFIC analyses on a tandem mass spectrometer from *Thalassiosira pseudonana* acclimated to Fe-limitation. Each node (or corner) represents a metabolite and the lines connecting the nodes represent an enzyme (i.e. protein). Metabolites were not measured in this study. Proteins that were identified in both Fe-replete and Fe-limited cultures are highlighted in grey. Proteins that were identified to be unique to Fe-limited cultures are indicated in color. From top left – light blue: sugar and glycan biosynthesis, light purple: starch and sucrose metabolism (including photosynthesis, oxidative phosphorylation, carbon fixation), dark purple: glycolysis-gluconeogenesis (including the citric acid cycle), red: nucleotide metabolism, teal: lipid metabolism, orange: amino acid metabolism (including urea cycle).

Table 6.5. Spectral count data and QSpec statistical analyses for key nitrogen metabolism, urea cycle, and spermine synthesis proteins identified in all experiments.

		Fe replete					Fe limited					QSpec Statistics		
		4GPFS/replicate				Triplicates on VELOS	4GPFS/replicate				Triplicates on VELOS			
Global ID Number	Protein Function	Tp_1^+	Tp_2^+	Tp_3^+	Tp_4^+	PAcIFIC Tp_1^+	Tp_1^-	Tp_2^-	Tp_3^-	Tp_4^-	PAcIFIC Tp_1^-	Bayes factor	Log_2 fold change	Significantly up/down regulated
224013245	glutamine synthase	78	75	76	80	69	44	53	44	52	57	112.00	0.46	
223999327	carbamoyl-phosphate synthetase III; CPSase III	19	20	27	28	36	23	21	21	23	37	0.52	0.07	
224009263	glutamate synthase	5	7	6	8	19	4	4	3	5	26	1.75	0.42	
224011010	ferredoxin-dependent glutamate synthase	4	2	2	4	14	6	4	5	4	21	1.20	0.41	
209583549	argininosuccinate synthase	6	5	6	6	10	11	13	7	8	20	3.37	0.48	
223995983	nitrite reductase (NAD(P)H) large subunit	13	16	10	12	21	15	17	10	13	20	0.56	0.06	
224009908	aspartate-ammonia ligase	1	2	0	0	12	5	8	8	6	19	40.64	1.63	YES
224013082	N-acetylornithine aminotransferase	2	1	4	2	7	6	8	6	7	16	28.15	0.91	YES
223993805	OTC: ornithine transcarbamylase	4	4	3	3	10	3	2	2	3	14	1.03	0.25	
224000259	oat-prov protein-ornithine aminotransferase	5	8	8	8	10	11	8	6	7	14	0.62	0.09	
223996511	spermine synthase	5	11	6	8	5	19	16	11	15	14	43.36	0.66	YES
223999447	asparagine synthetase	0	0	0	0	4	2	3	4	6	12	10.25	1.63	YES
224010906	nitrate reductase	2	2	1	1	14	0	0	0	1	11	1.09	0.84	
224014054	hydrolase/ nickel ion binding / urease	2	2	2	2	7	1	1	1	1	11	1.15	0.44	
224001546	spermidine/putrescine ABC transporter	2	3	2	4	7	2	3	2	5	11	0.59	0.06	
223999105	glutamate dehydrogenase	1	1	3	3	5	2	1	1	2	10	0.71	0.20	
224005475	argininosuccinate lyase	0	0	0	0	3	1	1	0	1	8	1.46	0.77	
223993093	ornithine cyclodeaminase	2	0	0	0	9	2	1	2	0	8	0.71	0.48	
224012585	glutamine synthetase	0	0	0	0	3	2	3	2	2	7	3.03	1.34	
224003823	NADH glutamate synthase small chain	2	0	1	1	8	3	3	1	1	7	0.93	0.42	

224007933	solute:sodium symporter/ urea transporter	7	5	6	5	11	4	4	5	5	7	0.77	0.22
223997254	nitrite transporter NAR1	2	3	4	3	8	1	2	1	1	7	2.53	0.66
223999185	ferredoxin nitrite reductase	0	0	0	0	0	4	4	1	0	6	4.56	1.34
223995861	carbamate kinase	0	0	0	0	0	0	1	1	0	6	1.40	0.58
224006530	mitochondrial glycine decarboxylase T-protein	2	1	1	1	0	2	3	3	3	5	1.71	0.57
223998258	putative nitrate transporter	2	0	1	1	11	0	0	1	0	5	1.08	0.62
209585936	nitrate transporter	2	2	4	5	4	2	4	3	3	4	0.72	0.05
224008166	aminomethyltransferase	0	0	0	0	2	0	0	0	0	2	1.00	0.00
224008464	nitrate reductase apoenzyme	0	0	0	0	0	0	0	1	0	2	0.99	0.34
209586149	arginase	0	1	2	1	2	0	0	0	0	0	1.00	0.00
224014840	putative leucine dehydrogenase	0	0	0	0	0	0	0	0	0	2	Identified only using PAcIFIC	
223998388	glycine cleavage protein (aminomethyltransferase)	0	0	0	0	0	0	0	0	0	2	Identified only using PAcIFIC	

Each replicate analysis (i.e. 1, 2, 3, 4) is indicative of the summation of spectral counts for that particular protein identified in 4 gas phase fractions (GPFs). Spectral counts resulting from triplicate analytical cycles of PAcIFIC were added together to provide a final count for the PAcIFIC analysis. In order to be considered significantly up or down regulated two criteria were met by QSpec: Bayes factor >10, and $\log_2(-Fe / +Fe)>0.5$. Proteins that do not have QSpec information were only identified using the data-independent PAcIFIC method and could not be statistically evaluated. This list does not include all proteins involved in nitrogen metabolism (e.g. amino acids biosynthesis and degradation).

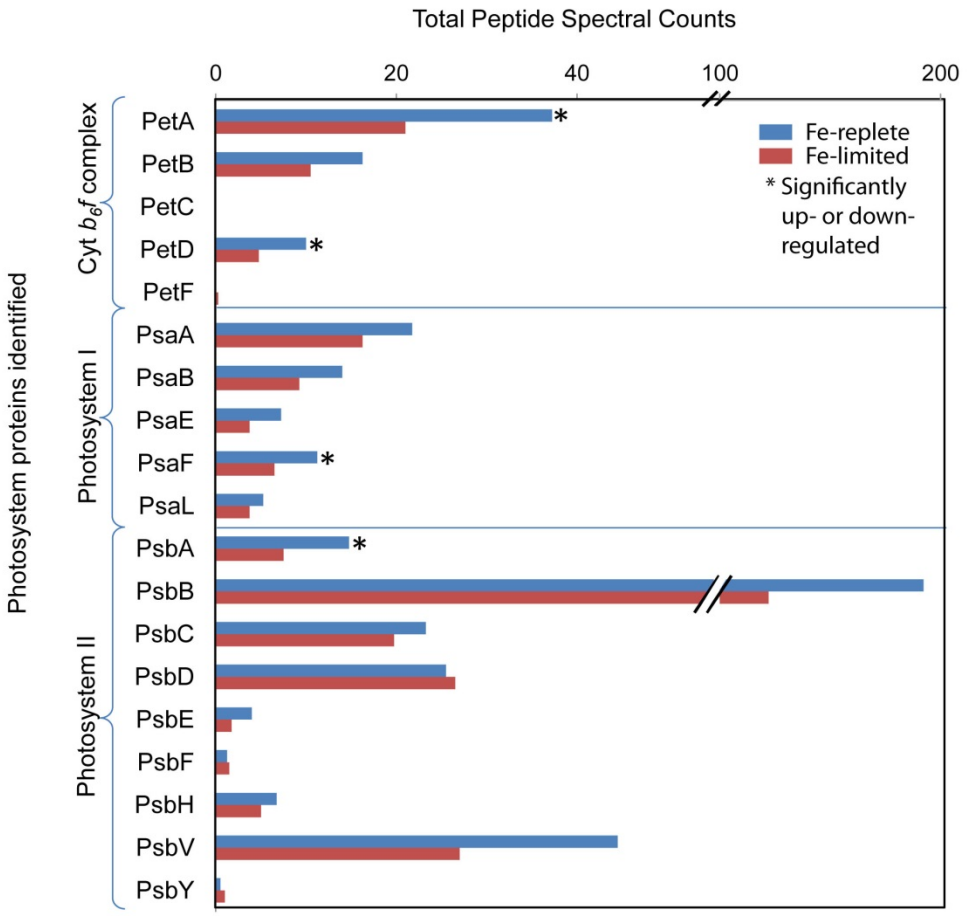


Figure 6.4. Total peptide spectral counts from photosystem complex subunits. Spectral counts result from quadruplicate analyses on Fe-replete (blue) and Fe-limited (red) cultures. Photosystem II requires 2-3 atoms of Fe per complex, Cytochrome (Cyt) b_6f complex requires 6 Fe atoms per complex, and photosystem I requires 12 Fe atoms per complex. "*" indicates that the protein was determined to be significantly up- or down-regulated by QSpec (i.e. Bayes Factor > 10 and $\log_2$ fold change >0.5).

6.5. Discussion

The goal of this study was to better understand the physiological changes in whole-cell metabolism of the cosmopolitan diatom *T. pseudonana* when acclimated to Fe limitation by examining the expressed proteome at mid-exponential steady-state growth. Several investigators have reported gene expression and enzyme assay responses to rapid Fe-limitation (Milligan and Harrison, 2000; Whitney et al., 2011) and Fe-enrichments or resupply (for examples see Table 6.1) (Kustka et al., 2007). Whole-cell diatom responses examined thus far have also revealed physiological responses of field samples before and after quasi-instantaneous Fe enrichments using transcriptomics (Durkin et al., 2012; Marchetti et al., 2012). The relationship between genomics, transcriptomics, proteomics, and metabolic strategies utilized by an organism is currently being scrutinized (Fernie and Stitt, 2012; Foss et al., 2011; Konopka and Wilkins, 2012). Recent studies of functional RNA in humans have shown that although much of the genome is transcribed, not all of the transcripts are translated into functional proteins (Birney et al., 2007; Dyhrman et al., 2012; Foss et al., 2011). This suggests that pure genomic analysis or transcriptomic tiling arrays may be limited in their ability to estimate cellular processes and that proteomics may provide an additional metric for determining metabolic strategies. Lommer et al. (2012) demonstrated that *T. oceanica* had a very high dynamic range of transcript expression and a relatively narrow range of protein expression in response to varying Fe conditions, suggesting that diatoms, like other organisms, may rapidly increase transcript abundances, but not translate them to proteins.

There is also evidence that proteins can be synthesized within seconds (Lacerda et al., 2007; Lehninger, 1965), therefore sufficient acclimation to the desired environmental

perturbation is essential in order to avoid capturing short-term stress responses. For example, a time series study demonstrated the dynamic proteomic response of a bacterial culture to cadmium exposure in soils (Lacerda et al., 2007); within 15 minutes of exposure to cadmium, the culture manipulated the proteome rapidly enough to be detected using MS-based proteomics, revealing shifts in physiology that were not detected using phylogenetic tools. In order to better understand the adaptive strategies of diatoms to Fe limitation, cultures were acclimated to Fe limitation over 10 generations prior to harvesting them for proteomic analyses. The combination of data independent mass spectrometry and such acclimation in our study revealed that many proteins previously determined to be absent in Fe-limited cells actually exhibit a low-level constitutive expression. Although most metabolic pathways involved proteins represented in both the Fe-replete and Fe-limited cells, the Fe-limited cells up-regulated enzymes involved in the pentose phosphate pathway and intracellular protein transport and recycling (Fig. A6.1). In contrast, Fe-replete diatoms up-regulated proteins dedicated to cell division, photosynthesis and the production of macromolecules for energy (Fig. A6.2). The procedure to construct an interactive version of iPath Figures A6.1 and A6.2 is provided with the raw data in Text A6.1. While our primary goal was to understand how diatoms acclimate to steady-state Fe limitation, the metabolic pathways identified to be enhanced in Fe-replete diatoms illustrate that Fe-limitation studies inherently include growth limiting metabolic shifts. In order to tease apart generic growth limitation and Fe-limitation, one would need to do a large-scale study examining the effects of multiple limiting nutrients (e.g. light, temperature, and other nutrients), including individual nutrient limitations and a matrix of co-limitations. This was beyond the scope of this

study. In the remainder of the Discussion section, we outline seven important cellular ramifications resulting from the acclimation of our model diatom to differing Fe conditions.

6.5.1. Intracellular conservation of nitrogen in Fe-limited cells

Although nitrate is the most abundant N source in the ocean, its use can be constrained by Fe availability as the enzymes involved in nitrate assimilation, nitrate and nitrite reductase, require Fe as a cofactor. Several studies have examined whether Fe-limited diatoms have a preference for, and the ability to utilize, oxidized or reduced forms of N (Maldonado and Price, 1996; Milligan and Harrison, 2000; Raven et al., 1999). Regardless of the N source provided to Fe-limited diatoms, the C:N ratio does not change significantly, suggesting N-uptake enzymes are functioning regardless of the Fe-cofactors required (Maldonado and Price, 1996). That said, Milligan and Harrison (2000) demonstrated that Fe limited diatoms build up an internal pool of NO_2^- due to the lack of cellular reducing power present in the cell to adequately use nitrite reductase. Furthermore, nitrite reductase requires 5 Fe atoms per active enzyme, whereas nitrate reductase only requires two. As a result, Fe limited cells have inactive nitrite reductases, resulting in the accumulation of intracellular nitrite and the retardation of nitrate assimilation. Proteomic profiling of the Fe-limited diatom at mid-exponential growth provided evidence of low-level expression of nitrate and nitrite reductases (Table 6.5). As no ammonium (reduced N source) was added to the culture medium, these findings suggest that Fe-limited diatoms continue to uptake oxidized nitrogen despite a weakened nitrogen assimilation pathway due to the lack of reducing power (as was demonstrated by

Milligan and Harrison (2000)). This suggests that oxidized N assimilation is diminished in Fe-limited cells in concert with diminished Fe-limited growth rates. Our proteomic profiles suggest that Fe-limited *T. pseudonana* are primed to compensate for this diminished reducing power and a compromised nitrate assimilation pathway by up-regulating intracellular N recycling pathways. In Fe-limited cells we identified more enzymes involved in intracellular protein trafficking, protein and peptide bond breaking, and the general re-distribution of N and C from protein backbones (Table 6.4). Evidence for this enzyme-based machinery in Fe-limited cells includes: the unique identification of 14 proteins with proteasome subunits, up-regulation of 2 aminotransferases, the complete pathway for acyl-tRNA biosynthesis, up-regulation of multiple peptidase enzymes, and the identification of 112 proteins involved in amino acid metabolism. Recycling intracellular N-rich proteins would enable the diatom to reduce resources allocated to nitrogen uptake and assimilation (nitrate, nitrite, urea, ammonia), while permitting the cell to regenerate new proteins for growth (e.g. see Fig. 6.5) (Allen et al., 2011). Lommer et al. (2012) revealed transcriptomic evidence of a similar "biomass recycling" response when *T. oceanica* was subjected to Fe stress and harvested at late-exponential growth phase. This efficient intracellular N recycling strategy may also contribute to the higher diversity of proteins observed in the Fe-limited diatom at mid-exponential growth (Fig. 6.1 and 6.3).

Fe-limited *T. pseudonana* expressed 35 more enzymes for amino acid metabolism compared to the Fe-replete cells (Fig. 6.3 and Table S6: http://www-ncbi-nlm-nih-gov.proxy-um.researchport.umd.edu/pmc/articles/PMC3797725/#s6title). These included multiple aminotransferases, which can yield many fates for amino acids

including re-organization into new amino acids or complete intracellular recycling to alpha keto acids, ammonia, or pyruvate (Berg et al., 2002). Argininosuccinate lyase (ArgH), an enzyme that degrades argininosuccinate to arginine and fumerate, was discovered to have a higher protein abundance in the Fe-limited diatoms (Table 6.5). This suggests the activation of the argininosuccinate shunt during iron limitation and its discovery supports our hypothesis of enhanced intracellular recycling of N-containing compounds. Since arginine can be easily mobilized, degraded, or incorporated into proteins, amino acids, and polyamines (e.g. see Fig. 6.5), it may play a critical role in diatom N recycling and storage (Llacer et al., 2008). Iron-limited *T. pseudonana* expressed an incomplete urea cycle and up-regulated proteins involved in polyamine synthesis (Table 6.5 and Fig. 6.2, 6.3). The alteration of these biochemical pathways was also noted by Mock et al. (2008). The one enzyme from the urea cycle that we did not detect in Fe-limited cells (Fig. 6.3), arginase, is responsible for the breakdown of arginine into urea. This process appears to primarily occur in Fe-replete cells (Nunn et al., 2009) and in both cell states when they reach stationary growth (data not shown).

Urea can also provide N for diatoms (Price and Harrison, 1988). No significant difference in expression of the urease enzyme was detected between the Fe-replete and Fe-limited cultures. Although previous reports have observed the urease enzyme to be present under Fe and N limitation (Allen et al., 2011; Milligan and Harrison, 2000), urease requires nickel (Ni) as a cofactor when processing urea (Egleston and Morel, 2008; Price and Morel, 1991). The Ni-ABC transporter (NikA) was identified to be the most down-regulated protein in Fe-limited cells (-2.87 fold; Table A6.2). Egleston and Morel (2008) demonstrated that the presence of Ni is essential for urease when diatoms

are using urea as their primary N source. The down-regulation of the Ni-transporting enzymes suggests that urease was not actively utilized and urea was not a primary N source during Fe limitation. Marchetti et al. (2012) conducted Fe-resupply experiments on natural diatom communities that induced the generation of urea cycle transcripts, indicating the pathway's close ties to Fe availability. This tight interlinking of the urea cycle to Fe supply suggests that urea might be a N storage molecule under Fe-replete conditions.

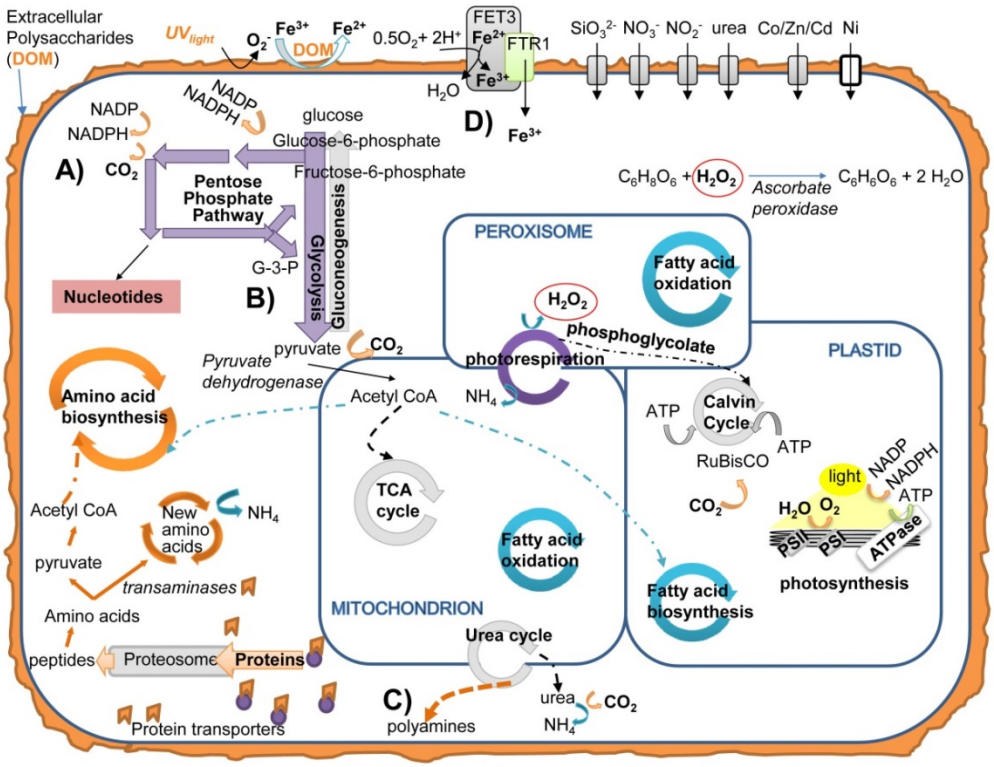


Figure 6.5. Stylized representation of diatom cell biochemistry when acclimated to Fe-limitation. Not all metabolic pathways are shown. Black and grey pathways and proteins indicate presence during Fe-limitation, white proteins indicate significantly down-regulated proteins, colored pathways were significantly up-regulated in Fe-limited cells compared to Fe-replete cells. A) Pentose Phosphate Pathway, B) The rejoining of the pentose phosphate pathway with glycolysis and generation of pyruvate. C) Polyamine synthesis using spermine synthase. D) Proposed reduction of Fe^{3+} and eventual transport of Fe into cells. Purple: photosynthesis and glycolysis-gluconeogenesis, red: nucleotide metabolism, teal: lipid metabolism, orange: amino acid metabolism.

6.5.2. Glycolysis and the pentose phosphate shunt in Fe-limited cells

Glucose is a primary product resulting from photosynthesis and it can be used in many ways to generate energy in the cell. The up-regulation of the pentose phosphate pathway may allow Fe-limited cells to bypass the first step of glycolysis, which uses ATP, and provide them with excess CO_2 and reducing equivalents of NADPH (Fig. 6.5A). These reducing substrates may be particularly important in Fe-limited cells by lessening oxidative stress and maintaining cellular integrity. This re-direction of glucose also provides a biochemical route for the cells to create nucleotides from the 5-C molecule ribulose-5-phosphate. The pentose phosphate pathway eventually re-connects with the glycolysis pathway and produces pyruvate, a useful cellular currency (see Fig. 6.5B). In the mitochondria, pyruvate may be converted to acetyl CoA by pyruvate dehydrogenase, which was up-regulated in Fe-limited cells (+1.2 fold). Acetyl-CoA can then either enter the citric acid cycle, amino acid biogenesis, or fatty acid biosynthesis – pathways that had the full complement of necessary enzymes identified in Fe-limited diatoms (Fig. 6.3 and 6.5).

The enzymes involved in glycolysis and the pentose phosphate pathway identified in this study have been previously shown to play a primary role in the conversion of glucose to energy in times of low photosynthetic competence or activity, such as under Fe limitation (Espen et al., 2000; Nisi and Zocchi, 2000; Reinhardt et al., 2006; Wang et al., 2001). The over-expression of enzymes in the glycolytic pathway by photosynthetic organisms stressed by Fe has been previously noted in green alga (Reinhardt et al., 2006), tomatoes (Wang et al., 2001), and in cucumber roots (Espen et al., 2000; Nisi and Zocchi, 2000). To our knowledge, this is the first time that the pentose phosphate pathway has

been identified to play a key role in the acclimation of diatoms to Fe limitation by providing cells additional reducing equivalents via the pentose phosphate shunt.

6.5.3. Photosynthetic energy production in Fe-deplete cells

In most photosynthetic organisms, light-driven adenosine triphosphate (ATP) synthesis via photosynthesis is the primary pathway for generating chemical energy that can be used to drive other metabolic pathways. Many of the photosynthetic enzymes were previously identified to be the most abundant in nutrient replete *T. pseudonana* cells (Nunn et al., 2009). We would expect the down-regulation of photosynthetic proteins when Fe limits diatoms because many require Fe as a cofactor in order to transfer electrons to the ATP synthase complex (Raven et al., 1999). Photosystem I proteins and Cytochrome b_6f require Fe and are down-regulated in Fe-limited cultures (see Table S7: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075653) (Greene et al., 1991; Lommer et al., 2012; Strzepek and Harrison, 2004). This study revealed that *T. pseudonana* acclimated to low Fe conditions retains the core complexes of both PSI and PSII, but most subunits of these complexes are down-regulated in Fe-limited cells (Fig. 6.4). One key function of these photosynthetic complexes is to pump protons across the thylakoid membrane of the chloroplast, which drives the production of ATP by ATP synthase. Under Fe limitation, three chloroplast-localized ATP synthase subunits are down-regulated (Table A6.2).

The presence of both photosystem complexes suggests that they are working collaboratively, but they are generating fewer protons to drive ATP synthase, resulting in fewer ATP synthase complexes being present. This would suggest that the limited Fe in

the cell is retained in the PSII and PSI complexes and to control growth, ATP synthase is actively synthesized or recycled, depending on growth status. This can result in a reduction of net energy gained and slower growth rates in Fe-limited diatoms (Table 6.2).

6.5.4. Enhancement of photorespiration in Fe-limited cells

Two enzymes responsible for supplying substrates for the C-concentrating mechanism (CCM) are carbonic anhydrase and phosphoenolpyruvate carboxylase (PEPC2). The higher abundance of PEPC2 in Fe-replete cells compared to Fe-limited cells suggests that when diatoms are provided excess nutrients they are capable of actively avoiding photorespiration by using CCMs, whereas the low-abundance of PEPC2 in Fe-limited cultures may force them to photorespire (Marchetti et al., 2012). An enhanced role for photorespiration under Fe limitation is supported by the unique identification of two photorespiration enzymes using the LTQ-VELOS PAcIFIC profile (phosphoglycolate phosphatase and glycine decarboxylase). Photorespiration can result in a net loss of C and the increased presence of proteins in this pathway provides a biochemical explanation for the reduced C:N ratio found in *T. pseudonana* acclimated to Fe limitation (Maldonado and Price, 1996).

The consequences of Fe-limited diatoms photorespiring, rather than employing a CCM, include net cellular C and N loss, slower growth rates, and the production of ammonia. Ammonia has many potentially useful fates in diatoms including (but not limited to) amino acid generation and polyamine synthesis (Fig. 6.5C). Spermine synthase, a key enzyme in polyamine synthesis, which helps precipitate silica on cell wall frustules, was also identified to be up-regulated in Fe-limited cells (Table 6.5, Table

A6.2). We propose that photorespiration, through its link with spermine synthase and its substrate ammonia, may underlie the increased Si:N ratio (Timmermans et al., 2004) and the thickened silica frustule observed in Fe-limited *T. pseudonana* (Marchetti and Cassar, 2009; Wilken et al., 2011).

6.5.5. Iron acquisition and endocytic recycling in Fe-limited cells

We predicted that *T. pseudonana* could increase production of proteins involved in Fe acquisition when limited by Fe. Unlike oceanic pennate diatoms, such as *Pseudo-nitzschia* sp. and *Fragilariopsis* sp., *T. pseudonana* does not have the gene encoding ferritin, a protein used for intracellular Fe storage (Marchetti and Cassar, 2009). The lack of ferritin may contribute to the higher extracellular Fe levels (or demand) for *T. pseudonana* to thrive compared to oceanic diatoms.

Iron permease, FTR1, was up-regulated in the Fe-limited cells (Table A6.2; +0.77 fold). FTR1 requires the copper-requiring ferroxidase (FET3) to function properly (Kwok et al., 2006; Strochlic et al., 2008). The FET3 enzyme was not identified in our study, possibly because it is membrane-bound and difficult to extract and ionize (Santoni et al., 2000). Strochlic et al. (2008) observed that yeast cells maintain and recycle the FTR1 complex under Fe limitation, whereas during Fe-replete conditions the cells targeted this complex for endocytic recycling, regenerating the protein when needed. This may explain the observed up-regulation of FTR1 enzyme in Fe-limited cells. Notably, we did not detect ferric reductase, an enzyme noted to reduce Fe^{3+} for uptake by the FET3-FTR transport system (see section 6.5.6).

6.5.6. Polysaccharide biosynthesis and its potential role in Fe uptake

Multiple enzymes involved in polysaccharide biosynthesis were up-regulated in Fe-limited cultures. High concentrations of extracellular polysaccharides are produced by Fe-limited diatoms and have a role in increasing Fe bioavailability (Hassler et al., 2009; Hassler et al., 2011). Detailed examinations of controlled substrates in seawater demonstrated that the photoreduction of ferric iron (Fe^{3+}) occurs in the presence of acidic sugars under UV light (Kuma et al., 1995; Ozturk, et al., 2004; Rijkenberg et al., 2010). These findings led to the examination of photoreduction of Fe^{3+} in the presence of diatom polysaccharide exudates (Steigenberger et al., 2010). Steigenberger et al. (2010) demonstrated that the combination of UV light, diatom exudates and reactive oxygen species reduced Fe^{3+} to ferrous iron (Fe^{2+}) and the presence of the exudates further stabilized the Fe^{2+} for eventual biological uptake.

Typically, the redox state of iron in seawater is Fe^{3+}, thus requiring reduction prior to biological import into the cell. This is frequently carried out by ferric reductase, which is part of the membrane-bound FET3-FTR transport system. Ferric reductase was not identified in our proteomic analysis because it is either in low abundance in the cells or it was not efficiently extracted from the cell because it is membrane-bound. The up-regulation of enzymes used to produce polysaccharides in the Fe-limited culture might have provided a stabilizing effect for Fe^{2+} in the surrounding seawater. Having more available ferrous iron for uptake could supplement a diminished or absent ferric reductase uptake pathway.

6.5.7. Rapid sinking via aggregation

Although this study did not include the analysis of transparent exopolymer production during Fe-limited growth, several studies have reported a thickening or the extracellular matrix and mucus (Mari et al., 2005; Mock et al., 2008). Extracellular polysaccharides have also been observed to enhance the diatoms' ability to aggregate, or raft together (Boyd et al., 2005; Mari et al., 2005; Mari et al., 2012; Mock et al., 2008). This rafting of cells is also observed in the open ocean when cells have reached stationary phase and is thought to be a contributor to increased organic C export to the ocean floor (Jackson et al., 2005; Passow et al., 2001). Thus, the elevated production of polysaccharides by Fe-limited *T. pseudonana* could enhance the overall export rate of the cells out of the photic zone. In addition, this study, and others, have observed that when diatoms are Fe-limited they increase the thickness of the silica frustule and decrease the size of the cells (Table 6.2) (Leynaert et al., 2004; Marchetti and Cassar, 2009; Mock et al., 2008; Muggli et al., 1996). As a result, this increases the rate at which these diatom cells sink to the ocean floor. A rapid sinking rate from the upper ocean to the deep ocean might reduce bacterial colonization and the ultimate degradation rate of diatom organic matter. Consequently, those regions where diatoms are Fe-limited there should be enhanced algal C export from the upper ocean to deeper waters and sediments below. The identification of diatom photosystem proteins in sediment traps intercepting particles from the water column, and from ocean floor sediments, validates that their rapid vertical transport from the upper ocean does enhance C preservation (Moore, Nunn, Goodlett, et al., 2012; Nunn et al., 2010).

6.6. Conclusion

Using comprehensive whole-cell proteomic approaches we evaluated metabolic acclimation of the diatom *T. pseudonana* to steady-state Fe limitation. When this diatom is acclimated to Fe limitation, the proteins expressed suggest that intracellular N and Fe recycling are used to conserve essential resources during mid-exponential growth. Up-regulation of transaminases and proteolytic enzymes allows the diatom cell to harvest N from amino acids without transporting equivalent quantities of N into the cells using N-specific transporters. Intracellular recycling would decrease the amount of C and N needed by the cell. The lack of Ni-ABC transporters, an enzyme found to be necessary for the function of the urease enzyme, suggests that the bioavailable urea is not being utilized when the diatom is Fe-limited. Here, we propose a strategy as to how Fe-limited diatoms might continue to grow with limited assimilation of new N.

When acclimated to low levels of Fe, this diatom is able to continue photosynthesis to generate ATP despite the high Fe requirements of photosynthetic electron transport. A decrease in PSII photosynthetic efficiency, as observed in this study (Table 6.2), is a signature characteristic of Fe-limitation (Behrenfeld and Milligan, 2013). Likewise, intracellular Fe concentrations are reduced when diatoms are Fe-limited (Maldonado and Price, 1996; Strzepek and Harrison, 2004; Sunda and Huntsman, 1997). This suggests that to maintain steady-state growth when Fe-limited, Fe must be conserved within the cell and unique mechanisms to acquire Fe may exist. We propose, based on previous experiments (Steigenberger et al., 2010) and the proteins and pathways presented in this study, that increased extracellular polysaccharides combined with UV light and H_2O_2 production from heightened photorespiration might make ferrous Fe more

bioavailable. That said, we also note that whole-cell proteomic profiling is not optimized for extracting membrane-bound proteins, thus possibly not providing some membrane-bound Fe-specific transporters.

Further adaptations for subsistence under Fe limitation include the enhanced use of the pentose phosphate pathway. This bypass of the glycolysis pathway yields substrates that are important to nucleotide, amino acid, and fatty acid synthesis. The intracellular storage and movement of N-containing compounds in these Fe-limited diatoms is one of the more unique and unanticipated results from this study. In addition, rather than the ammonia by-product resulting from photorespiration being a costly substrate to discard, diatoms might shuttle ammonia to the cytoplasm-located urea cycle enzymes. Although during Fe-limitation the enzymes necessary to complete the full urea cycle are not present, multiple enzymes suggest that the N-rich substrates are shunted from the cycle and packaged into arginine and low molecular weight polyamines. The up-regulation of spermine synthase suggests that polyamines, a key molecule required for silica precipitation (Kröger et al., 2000), are in abundance in these Fe-limited diatoms. This would suggest that the thick silica frustule and enhanced sinking rate of Fe-limited diatoms might result from photorespiration by-products. These polyamines are also a reserve for N for subsequent utilization.

Here we have outlined seven essential cellular processes expressed in the proteome that are either enhanced or modified by diatoms acclimated to Fe-limited growth conditions. Using a detailed proteomic profiling method, it was possible to follow metabolic pathways that are modified and determine if nutrients or metabolic substrates were re-directed to maintain cellular homeostasis. Due to the rapid and

dynamic nature of protein and transcript expression, acclimation to a defined nutrient condition is essential to avoid harvesting cells that have not fully acclimated physiologically to an environmental change. Given the proliferation of environmental manipulation experiments – for example in climate change research on oceanic biota (Boyd and Hutchins, 2012) - it is also essential to report all details of growth conditions utilized.

6.7. Acknowledgements for this chapter

We would like to thank Jimmy Eng for his computational support and the MacCoss lab for helpful remarks and comments on this manuscript.

Chapter 7: Concluding Perspectives

7.1. Summary of the Key Findings from this Dissertation

7.1.1. Lipid biomarker approach to the determination of historical carbon inputs to marine sediments

The first part of my dissertation used a multi-proxy approach including molecular organic markers, bulk carbon and isotope measurements, as well as more recent approaches to terrestrial carbon estimation (the BIT index of Branched and Isoprenoid Tetraethers) to determine organic matter source inputs to a sediment core in the Arctic Ocean (Chapter 2). We were successful in determining and analyzing source inputs and concluded that organic biomarkers indicate marine sources as the more dominant input of organic matter with lower but continual contributions from terrestrial sources at this location during the time period encompassed within the core. The consistency seen among multiple molecular organic markers of both marine and terrestrial origin throughout the core suggests that sinking material or surface sediments were heavily influenced by bottom currents or other mixing processes prior to their deposition. No alterations in organic markers indicate sea level changes over the last 10ka at this particular core location. This could be because the core was not sampled to a deep enough depth to correspond with sea level changes or because its depth of more than 400m masks any changes in sea level height and its proximity to Barrow Canyon acts as a preferential deposition site for terrestrial material before it reaches the sampling location. While these approaches provide needed information on organic matter source inputs, the knowledge gained about the specific organisms responsible for the production of the observed organic matter is limited. The homogeneity seen in this studied core prompted

the incorporation of proteomic analyses with these traditional biogeochemical approaches as a means to more fully develop a picture of marine or terrestrial conditions at the time of deposition.

7.1.2. The extraction and identification of proteins from preserved material in marine sediments

The second part of my dissertation research (Chapter 3) attempted to identify protein contributions to marine sediments in the hope of correlating those proteins identified to molecular organic markers discussed in Chapter 2. The unsuccessful identification of proteins in the sediment core with which I was working prompted further investigation into the optimization of an extraction method for proteins from marine sediments. It was determined that extraction with urea buffer and the subsequent placement of the slurry extract onto precast 12% Bis-Tris 1-dimension, flat gels optimized the recovery of BSA added to the sediments. While this method showed successful recovery of the added protein, it was less successful at recovering identifiable preserved proteins from the sediment. A second component of this research, conducted by fellow graduate student, E. K. Moore, evaluated the ability of multiple databases to confidently identify proteins from surface marine sediments. Findings showed that searching measured spectra against larger databases containing more diverse proteomes did not result in a greater number of protein identifications and that, regardless of the protein database used, identified peptides correlated to proteins with the same function across taxa. The database comparisons indicate that determining taxonomic-level information remains a challenge in environmental samples comprised of unknown

populations of organisms, but show that it is possible to confidently assign functions of the identified proteins. The ability to identify proteins from the surface sediments and not from deeper, older sediments raises the question as to what is preventing these identifications.

As was shown with the database comparisons, taxonomic identifications are challenging even when searching large, comprehensive databases. The main limitation of environmental proteomics, at this time, is the lack of sequence information on most marine organisms. It is entirely possible that the lack of identifications from the deeper sediments of the core is due to the lack of a complete database that incorporates the genomic data of the organisms which contributed the proteins to the sediments. The other limitation when examining older sediments is the potential that, over the course of time, the interactions between the proteinaceous material and any potential matrices are too strong to allow for the removal of intact peptides via non-hydrolyzing extraction methods. Another alternative is that the lack of identified proteins could be the result of the organic material having undergone recycling or degradation to the point that any intact peptides are no longer distinctive enough to be conclusively assigned to parent proteins. If this is the case, we may not have seen the recoveries of BSA that we did had the incubation been allowed to proceed for several weeks or months. Clearly, more research is needed before we can fully understand the fate of proteins in marine systems.

7.1.3. Utilization of a model organism to elucidate expressed proteins and to track the incorporation of an isotopically labeled amino acid in the expressed proteome

The third part of my research (Chapters 4 and 5) shifted focus from attempting to identify organic matter inputs, both lipid and protein, in preserved organic material, to utilizing a model organism to address some of the limitations documented in the previous chapter (Chapter 3). It became apparent that to ultimately elucidate the role of proteins in the larger biogeochemical cycles, investigation would have to start at the beginning by first identifying what proteins a marine organism is producing and then determining whether that production can be tracked. These determinations could eventually lead to the discovery of sets of proteins that are known to be produced under certain conditions by select, most, or all organisms and that could be selectively searched for in environmental samples. Strategically, a large benefit would ensue from investigations into the synthesis and recycling of proteins through marine bacteria. The ability to track the specific production of amino acids, peptides, or proteins will enable a much more in-depth understanding of the extent to which bacteria recycle or remineralize organic material.

Chapters 4 and 5 detailed studies into the expressed proteomes of a model, heterotrophic, marine bacterium, *R. pomeroyi*. This bacterium was grown under environmentally realistic C conditions and in Chapter 4 the entire expressed proteome was evaluated. Over 2000 proteins were identified with little difference among the expressed proteins observed during growth. The most abundant proteins identified were responsible for porins, transport, binding, translation, and protein refolding. An examination of functional annotation clusters showed that those clusters most enriched

within the samples contained proteins contributing to RNA subunit structure and RNA binding; ribosome assembly and biogenesis; translation and amino acid, amine, organic acid, and oxoacid biosynthesis and metabolism. These biological functions suggest an organism that maintained active growth and cell turnover well into stationary phase with no obvious patterns of changing biological function evident at different phases of growth.

In Chapter 5, *R. pomeroyi*, was provided a tracer level addition of a ^{13}C-labeled leucine at two different time points and the proteins produced which incorporated the label were detailed at four time points throughout exponential and stationary growth phases. Over 1900 total proteins were identified, with a total of 479 proteins incorporating the leucine label and accounting for between 16 and 21% of the total proteins identified depending on the time point. Almost all of the most abundant proteins identified in this investigation contained ^{13}C-leucine and also were involved in porins, transport, binding and translation. One functional annotation cluster of proteins was determined to be more prominent and was enriched at all four time points. This cluster contained proteins involved in ribosomal biosynthesis, RNA binding, translation, and protein metabolism. The finding of these biological functions, as in the results from Chapter 4, indicates cells were actively undergoing turnover through all growth phases. The label was actively taken up at all four time points with no distinguishable pattern found to its incorporation. The widespread distribution seems to indicate that the label was recycled by the bacteria with little remineralization.

These studies demonstrate that the proteome expression of model organisms can be successfully established and show that the tracking of the incorporation of an isotopically labeled substrate is possible. The presence of porins and transport proteins in

abundance relative to the other proteins in both experiments suggests that these groups could represent potential biomarkers of bacterial processes and/or activity. The successful tracking of the uptake and incorporation of an isotopically labeled amino acid demonstrates a method by which bacterial protein synthesis can be tracked and monitored under varying conditions.

7.1.4. Utilization of a model organism to track expressed proteins when acclimated to low iron or iron replete conditions and what metabolic pathways were affected by the two iron conditions.

The fourth and final portion of this dissertation research (Chapter 6) investigates a marine diatom, *Thalassiosira pseudonana*, and the effect acclimation to low Fe or Fe replete culture conditions has on its expressed proteome.

A total of 1850 proteins were identified in this investigation. When acclimated to Fe limitation, *T. pseudonana* up-regulates proteins involved with intracellular protein recycling. The benefit of this up-regulation would be in the decreased dependence on extracellular N, C, and Fe. In addition, proteins, and related synthesis pathways, that were observed to be down-regulated in short-term Fe starvation studies were expressed when *T. pseudonana* was acclimated to low Fe conditions. These expressed proteins include nitrate and nitrite transporters and Photosystem I and II complexes. Relative increases in the abundance of photorespiration and pentose phosphate pathway proteins were also documented indicating metabolic changes that could support other physiological responses, such as the heavily silicified frustules observed in Fe-limited diatoms (Marchetti and Cassar, 2009; Wilken et al., 2011). As this study shows, the expressed proteins of phytoplankton communities in marine environments will

potentially allow the tracking of eukaryotic proteins as well as bacterial contributions. In addition, changes identified here encourage the potential use of proteomics in determining past inorganic marine conditions based on the expressed proteomes that are identified in preserved samples.

7.2. Concluding Remarks

The studies detailed in this dissertation demonstrate the potential of proteomics to help elucidate the presence of and possibly even the fate of proteins in marine environments. Proteomics allows for the identification of proteins produced by individual organisms, communities of organisms, and potentially preserved proteins in sediments. While there is still a great deal of work to be done before proteomics can be uniformly applied to marine samples, great strides are being made in the sequencing of marine genomes, in expansion of the capabilities of the tandem mass spectrometers which analyze these samples, and in software designed to correlate and analyze the resulting data. In addition, studies such as these will eventually combine to create a much more cohesive view of the role of proteins and by extension C and N in global biogeochemical cycles.

These studies create a foundation from which additional experimentation can potentially examine the expressed proteomes of *R. pomeroyi* or *T. pseudonana* under altered (i.e. stressed) culturing conditions. Uncertainties include whether the abundant proteins observed here are also produced under stress or if their abundance changes with variable conditions. These results also open up the possibility of using these model organisms for more complex labeling studies to determine whether the recycling of the

label can be tracked into the recipient organisms. On a community-level, it is also not known which organisms alter their expressed proteome when faced with limiting resources or competition; labeled substrates could be tracked through the community and not just one organism in culture. The list of possibilities is great and once advancements in the field overcome the current limitations imposed by the lack of sequenced genomes. Advances in genomic technologies may quickly reduce this challenge.

Appendices

Appendix A.2.

Table A2.1. Expanded data set of bulk parameters of sediment core HLY0501-JPC5 to include data from McKay et al. (2008). Shaded: Data from McKay et al. (2008)*. TOC: total organic carbon

Calculated Age (ka)**	Corrected Depth (cm)***	TOC (%)	Nitrogen (%)	C:N	$\delta^{13}C$ (‰)
0.0067	0-2	1.38	0.19	8.40	-21.73
0.0536	7-9	1.68	0.23	7.30	-22.52
0.0804	11-13	1.63	0.24	6.90	-
0.1340	19-21	1.50	0.21	7.20	-
0.1876	27-29	1.51	0.20	7.50	-
0.2412	35-37	1.47	0.21	6.90	-22.35
0.2747	40-42	1.36	0.15	10.62	-22.15
0.2948	43-45	1.50	0.21	7.00	-
0.3484	51-53	1.53	0.21	7.40	-
0.4020	59-61	1.42	0.21	6.90	-
0.4556	67-69	1.42	0.20	7.10	-22.46
0.5092	75-77	1.26	0.16	9.44	-23.07
0.5092	75-77	1.47	0.20	7.30	-
0.5427	80-82	1.33	0.16	9.78	-21.56
0.5628	83-85	1.45	0.20	7.10	-
0.5896	87-89	1.51	0.20	7.70	-22.74
0.6164	91-93	1.48	0.20	7.40	-
0.6432	95-97	1.53	0.19	8.20	-
0.6700	99-101	1.54	0.21	7.50	-22.73
0.7370	109-111	1.49	0.20	7.60	-
0.7772	115-117	1.31	0.16	9.70	-21.82
0.8040	119-121	1.51	0.20	7.50	-22.84
0.8107	120-122	1.27	0.14	10.62	-22.32
0.8576	127-129	1.51	0.20	7.50	-
0.9648	143-145	1.46	0.20	7.40	-
1.0184	151-153	1.50	0.20	7.60	-22.96
1.1256	167-169	1.48	0.19	7.80	-
1.2429	184-186	1.49	0.19	7.80	-23.07
1.2462	185-187	1.25	0.14	10.59	-23.48
1.3501	200-202	1.47	0.19	7.80	-
1.4573	216-218	1.48	0.18	8.20	-23.18
1.6181	240.5-242.5	1.51	0.19	8.00	-
1.8492	275-277	1.21	0.16	8.61	-21.72
1.8861	280.5-282.5	1.46	0.18	8.10	-22.97
2.3182	345-347	1.48	0.18	8.10	-23.12
2.5326	377-379	1.43	0.18	8.00	-23.08
2.6532	395-397	1.18	0.15	9.11	-22.61
2.7470	409-411	1.39	0.17	8.10	-23.24
2.9614	441-443	1.44	0.18	8.00	-23.31
3.3701	502-504	1.46	0.18	8.30	-23.42
3.4572	515-517	1.17	0.14	9.85	-22.26
3.5845	534-536	1.44	0.17	8.30	-23.360
4.0133	598-600	1.50	0.18	8.40	-23.44
4.2612	635-637	1.21	0.16	8.91	-22.88
4.4254	659-661	1.46	0.18	8.20	-23.50
4.6398	691-693	1.40	0.17	8.10	-23.56
5.0652	755-757	1.18	0.13	10.83	-24.01

5.0686	755.5-757.5	1.51	0.17	9.00	-23.66
5.2729	786-788	1.43	0.16	9.00	-23.70
5.6012	835-837	1.17	0.16	8.79	-24.10
5.7017	850-852	1.45	0.16	8.80	-23.92
5.8692	875-877	1.15	0.15	9.11	-22.55
6.1305	914-916	1.37	0.16	8.50	-24.11
6.3483	946-948	1.32	0.15	8.80	-24.07
6.6732	995-997	1.17	0.13	10.62	-23.38
6.7771	1010.5-1012.5	1.22	0.15	7.90	-24.15
7.2059	1074.5-1076.5	1.34	0.16	8.40	-24.39
7.4772	1115-1117	1.14	0.14	9.61	-24.03
7.8524	1171-1173	1.17	0.13	8.80	-24.74
8.0132	1195-1197	0.86	0.11	9.16	-25.24
8.2812	1235-1237	0.94	0.11	8.60	-24.95
8.4856	1265.5-1267.5	0.92	0.11	8.50	-24.94
8.8072	1213.5-1315.5	0.84	0.10	8.70	-25.09
8.8172	1315-1317	0.60	0.07	10.07	-26.15

- not detected; *There is an average offset of 0.23% between the values of TOC determined by this study and McKay et al. (2008). This may be due to different methodologies utilized in the determination of TOC; ** Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); *** Depth correction applied as per Darby, Ortiz, et al. (2009). McKay et al. (2008) data adjusted using same depth correction.

Table A2.2. The distribution of saturated fatty acids (μg g^{-1} OC) of sediment core HLY0501-JPC5 from the western Arctic Ocean.

		Saturated Fatty Acids																	
Calculated Age (ka)*	Corrected Sample Depth (cm)**	C14:0n	C15:0n	C16:0n	C17:0n	C18:0n	C20:0n	C21:0n	C22:0n	C23:0n	C24:0n	C25:0n	C26:0n	C27:0n	C28:0n	C29:0n	C30:0n	C31:0n	C32:0n	Total
0.0067	0-2	99.1	31.1	525.1	26.7	240.4	90.5	22.4	142.0	51.2	233.3	60.7	196.4	42.1	105.1	16.8	27.2	4.3	8.7	1923.1
0.2747	40-42	19.9	6.8	350.5	17.1	251.1	87.4	21.9	166.8	71.2	388.2	111.9	398.1	83.4	230.2	31.0	54.9	8.7	18.9	2318.0
0.5092	75-77	67.4	22.8	486.7	22.3	277.5	97.8	22.5	156.5	48.3	248.1	67.0	240.9	51.8	146.3	22.6	41.7	6.3	13.2	2039.7
0.5427	80-82	25.2	8.6	400.3	19.2	344.4	95.0	21.5	165.7	65.1	378.9	106.5	385.4	81.3	215.6	27.5	53.8	11.6	25.1	2430.7
0.7772	115-117	59.2	18.9	369.0	20.4	191.4	91.8	21.1	136.3	45.4	232.3	65.3	228.7	50.4	138.3	18.6	38.8	6.5	12.9	1745.3
0.8107	120-122	30.4	14.4	400.4	26.6	315.0	94.5	23.7	171.7	72.9	395.0	107.8	383.4	76.5	229.3	28.8	53.2	8.1	20.0	2451.7
1.2462	185-187	48.8	19.4	527.6	27.3	417.2	109.5	24.1	157.1	53.6	258.5	72.7	251.0	55.4	150.3	22.4	36.9	4.6	13.5	2249.9
1.8492	275-277	62.5	22.3	532.3	25.9	374.7	100.5	23.2	139.1	49.9	248.3	73.1	257.3	54.8	153.3	23.0	41.9	5.3	15.4	2202.8
2.6532	395-397	75.1	29.6	622.8	28.1	461.4	122.6	23.4	166.6	50.2	259.1	69.7	252.8	52.9	150.4	19.7	38.3	7.0	13.8	2443.6
3.4572	515-517	63.7	24.5	522.3	26.0	370.0	124.8	25.4	177.7	57.3	289.2	83.2	284.7	62.5	170.9	24.9	46.0	6.3	17.9	2377.4
4.2612	635-637	68.4	23.4	605.1	30.7	460.2	128.4	25.5	188.8	56.5	281.1	77.9	262.8	55.5	151.4	19.4	34.6	5.9	12.7	2488.2
5.0652	755-757	61.8	25.8	635.3	30.1	454.6	152.5	30.8	215.7	71.0	355.3	100.0	338.7	73.7	205.0	28.3	53.4	7.8	14.6	2854.5
5.6012	835-837	63.5	22.5	494.4	28.7	302.7	155.8	29.1	216.5	61.4	293.3	76.9	230.8	45.3	100.4	12.9	19.6	2.5	4.6	2161.0
5.8692	875-877	52.0	23.2	457.0	28.1	337.1	145.5	29.7	206.7	69.1	319.1	88.2	262.6	56.7	143.0	19.4	34.2	4.9	11.2	2287.6
6.6732	995-997	52.6	20.2	431.8	25.4	351.7	157.4	29.6	230.0	69.7	346.2	89.8	269.4	55.7	151.1	20.0	34.6	4.7	10.9	2351.2
7.4772	1115-1117	51.3	19.6	411.9	26.3	250.7	171.3	34.1	295.5	82.6	387.6	96.7	272.5	55.6	145.8	16.6	32.4	5.4	9.9	2366.0
8.0132	1195-1197	32.6	13.4	387.5	21.1	375.8	91.3	17.6	172.0	39.1	203.7	49.1	160.3	28.8	73.1	8.1	16.2	2.2	4.9	1696.8
8.8172	1315-1317	9.8	3.7	257.7	20.4	551.5	46.0	9.2	63.7	18.0	96.4	23.0	75.5	12.4	28.0	-	-	-	-	1215.4

- not detected; * Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); ** Depth correction applied as per Darby, Ortiz, et al. (2009); Abbreviations: OC: organic carbon

Table A2.3. Distribution of monounsaturated fatty acids (μg g^{-1} OC) of sediment core HLY0501-JPC5 from the western Arctic Ocean.

		Monounsaturated Fatty Acids (MUFA)											
Calculated Age (ka)**	Corrected Sample Depth (cm)***	C14:1	C16:1*	C16:1ω7	C16:1ω5	C16:1*	C17:1*	C17:1ω8	C17:1ω6	C18:1ω9	C18:1ω7	C18:1ω5	Total
0.0067	0-2	3.4	47.3	291.1	15.8	48.5	-	16.7	20.2	101.1	146.1	5.5	695.6
0.2747	40-42	-	4.2	3.8	84.0	21.7	4.2	3.7	5.2	23.0	63.3	8.6	221.7
0.5092	75-77	11.3	12.1	234.1	21.5	67.5	5.2	5.0	10.8	81.3	76.6	6.0	531.2
0.5427	80-82	3.8	4.0	0.0	146.8	25.8	4.8	-	4.0	57.9	78.3	11.2	336.6
0.7772	115-117	3.5	2.3	110.7	6.6	54.0	1.4	4.2	5.9	50.6	40.4	1.2	280.8
0.8107	120-122	2.5	5.7	0.0	149.3	27.4	3.1	1.2	4.4	67.8	100.6	3.7	365.5
1.2462	185-187	9.7	2.7	235.1	18.7	27.6	2.0	4.1	2.5	59.7	63.0	5.7	430.7
1.8492	275-277	14.6	4.9	318.8	14.1	21.2	2.2	4.8	9.9	56.1	39.6	1.9	488.1
2.6532	395-397	32.4	4.7	373.1	14.1	8.1	0.6	2.9	4.8	56.4	28.3	2.1	527.5
3.4572	515-517	15.5	0.9	176.2	9.0	7.7	0.9	2.2	7.3	54.4	33.0	1.0	308.0
4.2612	635-637	39.6	3.8	194.1	10.6	9.6	1.2	4.6	6.2	69.1	96.0	-	434.8
5.0652	755-757	18.0	3.9	343.2	11.3	5.7	-	3.8	6.9	61.1	54.9	-	508.8
5.6012	835-837	-	4.3	130.1	15.9	8.3	-	1.7	3.4	75.3	32.4	3.7	275.1
5.8692	875-877	16.2	1.5	76.0	5.6	5.0	-	4.6	6.1	88.9	52.3	-	256.2
6.6732	995-997	-	1.8	36.4	5.4	4.6	-	2.3	3.6	49.5	14.5	-	118.1
7.4772	1115-1117	-	2.7	41.6	5.3	4.0	-	2.2	2.3	59.1	24.0	-	141.3
8.0132	1195-1197	-	1.8	41.3	2.4	5.1	-	-	1.9	69.2	14.7	-	136.4
8.8172	1315-1317	-	-	9.2	-	-	-	-	-	59.5	17.7	4.4	90.8

- not detected; * position of unsaturation not determined; ** Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); *** Depth correction applied as per Darby, Ortiz, et al. (2009); Abbreviations: OC: organic carbon

Table A2.4. Distribution of n-alcohols (μg g^{-1} OC) of sediment core HLY0501-JPC5 from the western Arctic Ocean.

		n-alcohols																		
Calculated Age (ka)**	Corrected Sample Depth (cm)***	C14	C15*	C15*	C15	C16	C17*	C17	C18	C19	C20	C21	C22	C23*	C23	C24	C25	C26	C27	Total
0.0067	0-2	5.7	2.0	1.9	4.5	36.6	4.2	4.0	30.1	6.2	63.8	17.1	178.3	3.7	25.4	133.0	16.7	144.7	15.2	693.3
0.2747	40-42	1.3	0.6	0.6	1.3	14.7	1.0	2.0	13.3	2.6	35.2	11.8	129.3	3.2	21.4	111.0	16.5	125.4	12.2	503.4
0.5092	75-77	7.2	2.1	2.0	5.1	34.6	2.1	6.3	29.3	6.5	65.6	19.6	200.6	9.0	30.2	155.4	14.7	164.8	17.4	772.5
0.5427	80-82	2.1	1.1	0.6	1.8	17.3	1.4	2.9	14.9	3.5	40.4	13.4	151.0	3.7	24.1	126.9	19.0	142.3	12.5	578.7
0.7772	115-117	8.4	2.2	2.7	5.6	44.3	2.5	7.9	34.0	6.8	76.4	23.5	239.1	10.8	39.5	198.9	24.0	214.7	20.0	961.2
0.8107	120-122	2.8	1.5	0.7	1.9	23.0	1.9	3.1	15.6	3.4	40.3	14.3	153.4	4.9	26.9	130.2	19.0	138.0	14.2	595.1
1.2462	185-187	5.8	2.3	2.6	5.0	39.0	2.9	7.0	32.3	6.3	74.8	22.2	224.8	7.9	36.8	186.8	20.4	195.8	15.9	888.7
1.8492	275-277	2.3	1.5	1.8	3.4	27.8	1.6	5.3	24.3	5.0	52.1	15.8	162.5	-	28.2	145.1	17.2	172.3	18.9	685.0
2.6532	395-397	3.4	1.7	2.6	3.2	26.6	3.2	4.8	25.6	5.6	60.7	18.7	196.1	6.1	33.0	176.0	14.4	197.7	20.3	799.8
3.4572	515-517	4.6	2.1	2.0	4.1	32.4	2.9	7.3	30.8	5.8	64.6	20.0	202.6	7.5	34.5	181.4	18.1	204.3	19.8	844.9
4.2612	635-637	4.1	2.0	1.9	4.3	33.6	2.6	6.9	29.0	6.3	69.7	21.6	218.5	9.3	37.6	196.8	18.2	223.3	23.9	909.6
5.0652	755-757	4.6	2.6	2.3	4.4	36.1	3.1	8.5	29.9	6.5	77.2	24.8	245.3	8.8	46.0	231.3	28.4	279.4	32.2	1071.4
5.6012	835-837	1.1	1.0	0.7	2.5	11.1	4.6	3.9	10.9	3.2	30.5	11.8	96.1	7.5	18.6	90.5	9.5	118.6	15.5	437.6
5.8692	875-877	3.2	1.9	1.6	3.5	26.7	3.7	7.1	24.8	5.9	66.0	21.7	206.3	5.5	37.0	185.7	21.3	216.9	23.4	862.3
6.6732	995-997	3.9	2.4	1.7	4.5	31.8	3.1	7.8	27.7	6.3	73.5	23.9	223.9	9.6	40.0	199.1	18.1	237.5	23.7	938.4
7.4772	1115-1117	4.1	2.4	1.4	4.5	33.5	2.4	10.0	32.1	8.2	96.6	33.4	295.1	8.3	55.4	254.5	23.0	289.1	23.5	1177.4
8.0132	1195-1197	2.4	0.9	1.0	2.3	13.4	1.4	7.1	12.9	-	35.4	11.2	105.2	5.9	17.9	92.9	5.9	117.0	11.0	444.0
8.8172	1315-1317	-	0.5	0.5	0.8	7.6	-	2.1	10.7	2.3	20.6	7.3	67.6	-	11.0	59.3	15.3	76.6	5.4	287.6

- not detected; * denotes branched alcohols; ** Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); *** Depth correction applied as per Darby, Ortiz, et al. (2009); Abbreviations: OC: organic carbon

Table A2.5. Distribution of n-alkanes (μg g^{-1} OC) of sediment core HLY0501-JPC5 from the western Arctic Ocean.

		n-alkanes																
Calculated Age (ka)*	Corrected Sample Depth (cm)**	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	Total
0.0067	0-2	1.7	2.6	5.3	7.5	12.9	13.1	28.8	14.4	33.9	13.6	48.8	7.9	36.7	-	41.9	-	12.9	282.0
0.2747	40-42	0.8	0.7	1.9	3.1	7.1	7.8	19.1	13.6	29.3	15.1	45.2	9.7	33.2	3.2	29.2	-	12.9	228.7
0.5092	75-77	6.6	8.5	14.4	11.6	18.1	16.8	35.7	19.4	41.2	16.2	59.3	10.4	43.3	4.4	39.2	-	7.4	352.6
0.5427	80-82	1.7	1.2	3.7	4.8	8.3	9.2	20.7	14.8	31.4	15.1	49.0	10.3	0.0	3.7	35.7	-	16.2	222.3
0.7772	115-117	6.1	8.9	16.6	16.1	25.4	25.2	53.0	37.3	65.7	34.6	87.9	21.9	70.3	-	65.3	-	11.4	545.6
0.8107	120-122	2.5	2.3	4.3	5.5	10.3	10.5	24.1	17.1	37.2	18.9	55.3	12.4	40.8	4.6	35.4	-	17.7	294.4
1.2462	185-187	-	3.6	10.3	12.1	22.0	18.3	40.3	27.3	48.2	24.0	65.8	13.3	46.0	-	38.3	-	8.4	377.9
1.8492	275-277	-	-	4.0	7.1	15.0	15.9	35.2	20.3	43.8	18.2	66.0	11.4	50.4	-	48.0	-	7.2	342.5
2.6532	395-397	-	1.9	8.2	9.8	18.1	16.1	37.7	24.5	49.8	22.3	71.6	11.1	52.1	5.4	48.9	-	7.8	385.4
3.4572	515-517	-	3.4	9.2	10.6	18.7	17.7	42.1	25.8	52.5	22.1	78.2	13.9	63.3	-	59.7	-	7.4	424.6
4.2612	635-637	-	3.5	9.6	10.9	19.1	19.2	44.2	27.0	60.0	24.8	89.8	16.7	70.3	-	70.6	-	5.6	471.4
5.0652	755-757	-	4.8	10.7	12.0	21.4	20.5	49.1	29.9	62.1	25.4	93.8	16.6	75.3	-	75.9	-	7.8	505.3
5.6012	835-837	1.9	3.9	8.7	10.0	20.2	21.4	50.8	30.6	68.8	27.6	100.9	18.7	77.8	7.5	61.6	-	11.7	522.2
5.8692	875-877	0.5	3.4	9.7	11.9	21.6	21.8	51.4	30.6	65.0	26.5	95.6	16.0	72.4	-	71.2	1.3	6.5	505.7
6.6732	995-997	-	2.6	9.9	11.6	21.6	21.7	47.1	27.4	55.7	20.8	73.1	12.3	48.5	-	48.2	-	4.7	405.1
7.4772	1115-1117	1.8	6.0	15.1	16.2	29.3	29.2	67.2	40.8	81.2	33.0	111.4	19.4	79.3	-	79.5	1.9	12.7	623.9
8.0132	1195-1197	2.7	8.9	20.1	21.4	29.9	27.8	52.5	30.3	60.0	25.0	78.9	18.0	63.5	-	54.5	-	19.2	512.7
8.8172	1315-1317	-	-	5.6	11.8	17.4	19.0	30.5	18.6	33.1	17.9	40.3	13.7	34.0	7.8	29.4	4.9	7.3	291.2

* Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); ** Depth correction applied as per Darby, Ortiz, et al. (2009); - not detected; Abbreviations: OC: organic carbon

Table A2.6. Distribution of sterols (μg g^{-1} OC) of sediment core HLY0501-JPC5 from the western Arctic Ocean.

		Sterols														
Calculated Age (ka)*	Corrected Sample Depth (cm)**	C26Δ5,22	C26Δ22	27norC26Δ25,22	C27Δ5,22	C27Δ5	C28Δ5,22	C28Δ22	C28Δ5,24(28)	dMeΔ5	C29Δ5	dMeΔ0	Dinosterol	tMeSΔ0	tMeRΔ0	Total
0.0067	0-2	17.21	10.19	8.35	23.36	67.52	47.76	19.84	36.07	9.44	58.47	30.43	38.93	21.39	18.05	435.1
0.2747	40-42	12.18	-	2.07	11.90	32.00	20.80	6.35	21.04	2.64	18.83	7.75	10.50	8.27	7.44	154.0
0.5092	75-77	13.73	9.88	2.05	39.33	43.28	27.40	11.77	28.13	4.20	36.65	26.32	25.16	16.35	12.43	314.6
0.5427	80-82	8.01	-	1.92	11.38	48.22	22.42	6.67	19.61	4.04	23.09	25.03	25.49	10.81	9.80	191.5
0.7772	115-117	12.71	12.29	9.59	25.11	44.43	42.33	14.21	36.08	4.45	40.11	26.42	23.48	16.88	12.34	338.8
0.8107	120-122	12.27	-	1.69	9.75	45.78	20.71	5.20	18.65	4.35	26.45	17.02	24.17	11.87	9.38	190.3
1.2462	185-187	10.63	11.49	1.88	13.55	36.72	24.05	11.71	29.57	4.85	43.89	23.67	25.78	20.57	14.05	288.6
1.8492	275-277	9.70	9.25	9.48	12.05	32.32	23.24	11.33	27.47	4.87	32.87	28.20	20.17	16.57	12.20	266.6
2.6532	395-397	9.14	10.33	2.57	9.60	29.92	21.31	10.04	21.82	3.91	33.43	22.01	20.42	15.34	10.40	234.6
3.4572	515-517	10.20	10.24	8.93	10.03	32.01	25.54	14.64	26.22	4.20	44.13	24.19	19.50	15.66	10.39	275.0
4.2612	635-637	10.68	11.65	8.55	9.45	29.77	23.94	14.82	25.25	3.08	40.85	23.42	15.75	12.96	10.60	258.7
5.0652	755-757	14.41	16.45	8.72	12.72	39.80	31.73	19.32	30.97	3.78	63.77	26.34	21.17	19.17	15.49	348.5
5.6012	835-837	5.05	4.55	0.86	3.81	13.18	10.33	5.09	12.56	1.20	15.65	11.16	2.13	2.14	2.46	96.6
5.8692	875-877	8.89	10.32	5.14	8.74	28.87	22.36	13.57	24.94	3.55	46.00	19.68	16.11	12.87	9.26	246.3
6.6732	995-997	9.79	11.18	5.04	9.45	27.98	25.04	15.52	24.70	1.73	44.45	23.96	12.48	11.03	8.80	247.6
7.4772	1115-1117	8.69	11.99	8.03	10.67	29.79	28.67	20.88	25.61	4.28	73.53	22.82	16.68	14.98	15.40	317.3
8.0132	1195-1197	3.27	4.94	-	3.57	15.56	12.60	8.97	14.08	1.24	19.09	14.23	5.50	5.86	4.34	122.8
8.8172	1315-1317	1.81	4.62	-	5.82	23.60	10.20	7.43	13.78	0.70	17.96	9.87	8.92	6.15	4.86	122.7

* Calculated age based on mid-point of corrected depths and determined with age model from Darby, Ortiz, et al. (2009); ** Depth correction applied as per Darby, Ortiz, et al. (2009); - not detected

Abbreviations: C26Δ5,22: 24-norcholesta-5,22-dien-3b-ol; C26Δ22: 24-nor-5a-cholest-22-en-3b-ol; 27norC26Δ25,22: 27-nor-24-cholesta-5,22-dien-3b-ol; C27Δ5,22: cholesta-5,22-dien-3b-ol; C27Δ5: cholest-5-en-3b-ol; C28Δ5,22: 24-methlycholesta-5,22-dien-3b-ol; C28Δ22: 24-methlycholest-22-en-3b-ol; C28Δ5,24(28): 24-methylcholesta-5,24(28)-dien-3b-ol; dMeΔ5: 23,24-dimethylcholest-5-en-3b-ol; C29Δ5: 24-ethylcholest-5-en-3b-ol; dMeΔ0: 4a,24-dimethylcholestan-3b-ol; Dinosterol: 4a,23,24-trimethylcholest-22-en-3b-ol; tMeSΔ0: 4a,23s,24r-trimethylcholestan-3b-ol; tMeRΔ0: 4a,23r,24r-trimethylcholestan-3b-ol; OC: organic carbon

Appendix A.3.

Table A3.1. Summary list of all proteins identified in each sample database search. Includes identified species, biological function, cellular compartment (Comp): C = Chloroplast; S = Secretory; M = Mitochondria; N = Nucleus; U = Uncharacterized Compartment, molecular weight (MW), isoelectric point (pI), percent sequence coverage (SC).

A. Slurry Tube Gel – Thaps Database

Protein	Annotation	Species	Function	Comp	MW	pI	SC
30S ribosomal protein S1	jgi 15259	*T. pseudonana*	Binding, RNA	R	31770	4.6	3.8
30S ribosomal protein S2	gi 118411126	*T. pseudonana*	Translation	C	25635	9.3	11.9
30S ribosomal protein S5	gi 118411207	*T. pseudonana*	Translation	C	19297	10.3	11.7
30S ribosomal protein S6	gi 118411222	*T. pseudonana*	Translation	C	12043	9.8	10.7
30S ribosomal protein S7	gi 118411217	*T. pseudonana*	Translation	C	17730	10.5	21.2
3-deoxy-7-phosphoheptulonate synthase	jgi 2790	*T. pseudonana*	Biosynthesis, amino acid	C	53939	6.0	2.5
40S ribosomal protein S11	jgi 22535	*T. pseudonana*	Translation	R	19118	10.3	8.5
40S ribosomal protein S23	jgi 28209	*T. pseudonana*	Translation	R	15734	10.5	12.6
40S ribosomal protein SA	jgi 21871	*T. pseudonana*	Translation	R	27261	5.9	10.3
50S ribosomal protein L11	gi 118411123	*T. pseudonana*	Translation	C	14880	9.7	9.2
50S ribosomal protein L14	gi 118411201	*T. pseudonana*	Translation	C	13433	10.3	13.2
50S ribosomal protein L19	gi 118411099	*T. pseudonana*	Translation	C	13812	10.6	25.0
50S ribosomal protein L2	gi 118411193	*T. pseudonana*	Translation	C	30675	10.9	5.5
50S ribosomal protein L22	gi 118411196	*T. pseudonana*	Translation	C	12986	10.3	11.3
50S ribosomal protein L23	gi 118411192	*T. pseudonana*	Transferase	C	11152	9.9	10.2
50S ribosomal protein L24	gi 118411202	*T. pseudonana*	Biosynthesis, amino acid	C	8682	10.1	20.8
50S ribosomal protein L3	gi 118411190	*T. pseudonana*	Translation	C	22012	10.1	26.1
50S ribosomal protein L4	gi 118411191	*T. pseudonana*	Translation	C	24203	10.2	7.9
50S ribosomal protein L5	gi 118411203	*T. pseudonana*	Translation	C	27571	9.7	3.8
60 kDa chaperonin	gi 118411188	*T. pseudonana*	Metabolic	C	57361	5.2	3.2
6-phosphogluconate dehydrogenase	jgi 33343	*T. pseudonana*	Pentose Phosphate Shunt	S	53348	5.6	9.8
Abnormal wing discs	jgi 6290	*T. pseudonana*	Diphosphate kinase	U	17236	5.5	13.8
Acetyl-CoA carboxylase	jgi 6770	*T. pseudonana*	Ligase	C	228295	5.0	3.2
Acidic ribosomal phosphoprotein P0	jgi 25812	*T. pseudonana*	Biogenesis	R	34116	4.8	3.7

Actin A	jgi 25772	*T. pseudonana*	Cytoskeleton	S	41791	5.0	4.8
Adenosinetriphosphatase	jgi 40156	*T. pseudonana*	Transport, proton	S	39935	7.6	10.6
Adenylate kinase	jgi 31809	*T. pseudonana*	Metabolic, nucleic acid	S	24598	6.3	4.4
ATP synthase CF0 B chain subunit I	gi 118411110	*T. pseudonana*	Photosynthesis	C	20029	9.8	8.4
ATP synthase CF0 B' chain subunit II	gi 118411109	*T. pseudonana*	Photosynthesis	C	17373	4.6	9.6
ATP synthase CF0 C chain subunit III	gi 118411108	*T. pseudonana*	Photosynthesis	C	8166	5.0	50.0
ATP synthase CF1 alpha chain	gi 118411112	*T. pseudonana*	Photosynthesis	C	53989	5.0	8.5
ATP synthase CF1 beta chain	gi 118411134	*T. pseudonana*	Photosynthesis	C	51143	4.7	30.2
ATP synthase CF1 delta chain	gi 118411111	*T. pseudonana*	Photosynthesis	C	21077	9.2	16.0
ATP/ADP translocator	jgi 39143	*T. pseudonana*	Transport	S	32254	9.4	9.9
ATP-dependent clp protease ATP-binding subunit	gi 118411220	*T. pseudonana*	Catalytic activity	C	102150	6.5	3.0
BiP	jgi 27656	*T. pseudonana*	Morphogenesis, Cell	C	70451	4.7	3.6
CbbX protein homolog	jgi 40193	*T. pseudonana*	Binding, ATP	C	35036	5.3	18.3
CDC48/ATPase	jgi 267952	*T. pseudonana*	Binding, ATP	S	89464	4.8	3.1
Cell division protein FtsH2	jgi 31930	*T. pseudonana*	Binding, Zn	S	61956	5.1	1.9
Cell division protein FtsH-like protein	gi 118411141	*T. pseudonana*	Photosynthesis	C	70206	5.1	6.9
CG11154-PA, isoform A	jgi 41256	*T. pseudonana*	Transport, proton	S	53388	5.1	9.4
Chloroplast light harvesting protein isoform 12, 18 kDa	jgi 33606	*T. pseudonana*	Photosynthesis	C	18463	4.6	7.6
Chloroplast light harvesting protein isoform 12, 26 kDa	jgi 270092	*T. pseudonana*	Photosynthesis	C	26078	5.5	6.2
Chloroplast light harvesting protein isoform 15	jgi 2845	*T. pseudonana*	Photosynthesis	C	21873	5.1	5.4
Chloroplast light harvesting protein isoform 5	jgi 32723	*T. pseudonana*	Photosynthesis	C	19175	5.2	10.6
Citrate synthase, mitochondrial precursor	jgi 11411	*T. pseudonana*	Transferase	M	52269	6.2	2.3
Coatomer protein complex, subunit gamma 2	jgi 269663	*T. pseudonana*	Transport, protein	S	100270	5.1	1.1
Cyclophilin	jgi 29244	*T. pseudonana*	Folding, Protein	S	20899	6.9	9.2
Cytochrome b559 alpha chain	gi 118411160	*T. pseudonana*	Photosynthesis	C	9514	5.6	21.4
Cytochrome b6	gi 118411154	*T. pseudonana*	Photosynthesis	C	23906	9.2	6.0
Cytochrome f	gi 118411137	*T. pseudonana*	Photosynthesis	C	33988	8.2	4.1
Elongation factor 2	jgi 269148	*T. pseudonana*	GTPase	S	91887	6.0	8.7
Elongation factor alpha-like protein	jgi 41829	*T. pseudonana*	GTPase	S	49969	8.7	16.8
Enoyl-acyl carrier reductase	jgi 32860	*T. pseudonana*	Oxidation reduction	S	32813	5.1	3.5
Eukaryotic translation initiation factor 4A2 isoform 2	jgi 9716	*T. pseudonana*	Binding, DNA	S	42405	5.6	11.9
Ferredoxin component	jgi 29842	*T. pseudonana*	Oxidoreductase	C	18511	8.9	11.7

Fructose-1,6-bisphosphate aldolase precursor	jgi 428	*T. pseudonana*	Glycolysis	S	39810	4.8	8.1
Fucoxanthin chlorophyl a/c protein, 21 kDa	jgi 38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyll a /c protein, 20 kDa	jgi 38494	*T. pseudonana*	Photosynthesis	C	20354	4.5	11.6
Fucoxanthin chlorophyll a /c protein, 21 kDa	jgi 42962	*T. pseudonana*	Photosynthesis	C	21515	5.1	11.0
Fucoxanthin chlorophyll a/c binding protein, 22 kDa	jgi 264921	*T. pseudonana*	Photosynthesis	C	22205	4.6	8.6
Fucoxanthin chlorophyll a/c binding protein, 22.6 kDa	jgi 268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	17.1
GDP-mannose 3,5-epimerase	jgi 41548	*T. pseudonana*	Coenzyme	S	40707	5.1	3.3
GDP-mannose dehydratase	jgi 40586	*T. pseudonana*	Coenzyme	S	40412	5.9	4.2
Geranyl-geranyl reductase	jgi 10234	*T. pseudonana*	Metabolic, Aromatic Compound	C	47230	5.9	10.3
Glucose-6-phosphate dehydrogenase	jgi 34514	*T. pseudonana*	Metabolic, glucose	S	57149	7.6	2.2
Glyceraldehyde-3-phosphate dehydrogenase	jgi 28334	*T. pseudonana*	Glycolysis	M	36574	6.1	10.8
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi 31383	*T. pseudonana*	Glycolysis	S	39587	5.3	14.1
Glycolaldehydetransferase	jgi 21175	*T. pseudonana*	Transport	M	71708	5.0	2.3
Heat shock protein 70	jgi 269120	*T. pseudonana*	Heat shock	S	71187	4.8	8.1
Heat shock protein Hsp90	jgi 6285	*T. pseudonana*	Heat shock	S	80242	4.7	2.0
Histone H2A.1	jgi 19793	*T. pseudonana*	Binding, DNA	N	13053	10.4	7.3
Histone H4	jgi 3184	*T. pseudonana*	Binding, DNA	N	11384	11.5	35.9
HLA-B associated transcript 1	jgi 269556	*T. pseudonana*	Hydrolase	S	49087	5.2	2.8
Hsp70-type chaperone	gi 118411189	*T. pseudonana*	Transcription	C	65339	4.8	2.3
HSP90-like protein	jgi 22766	*T. pseudonana*	Heat shock	S	80966	4.7	2.0
Hypothetical protein CBG08717	jgi 269322	*T. pseudonana*	Transport, Proton	S	58009	5.8	13.8
Hypothetical protein CdifQ_02003487	jgi 36462	*T. pseudonana*	Metabolic	U	21340	4.8	6.7
Hypothetical protein FG01081.1	jgi 25949	*T. pseudonana*	Translation	R	20124	9.8	15.3
Hypothetical Protein No BLAST result	jgi 23918	*T. pseudonana*	n.a.	U	15261	10.0	11.6
Hypothetical Protein No BLAST result	jgi 11169	*T. pseudonana*	n.a.	S	31192	4.8	14.0
Hypothetical protein SPBC29A3.04	jgi 269961	*T. pseudonana*	n.a.	C	28946	10.5	3.8
Isocitrate dehydrogenase NADP-dependent, monomeric type	jgi 1456	*T. pseudonana*	Citric Acid Cycle	C	73809	5.7	2.6
Ketol-acid reductoisomerase	jgi 23228	*T. pseudonana*	Biosynthesis, Amino Acid	S	58240	5.1	3.2
L4/L1	jgi 22610	*T. pseudonana*	Translation	R	40991	10.3	7.7
Nucleoside diphosphate kinase	jgi 31091	*T. pseudonana*	Diphosphate kinase	S	16917	8.3	11.3

Oxygen-evolving enhancer protein 1 precursor	jgi 34830	*T. pseudonana*	Photosynthesis	S	29136	5.2	20.4
Phosphoglycerate kinase precursor	jgi 35712	*T. pseudonana*	Glycolysis	S	42256	5.0	10.5
Photosystem I ferredoxin-binding protein	gi 118411153	*T. pseudonana*	Photosynthesis	C	15518	9.6	46.8
Photosystem I p700 chlorophyll A apoprotein A	gi 118411096	*T. pseudonana*	Photosynthesis	C	83642	7.3	3.2
Photosystem I p700 chlorophyll A apoprotein B	gi 118411097	*T. pseudonana*	Photosynthesis	C	82090	7.6	5.2
Photosystem I protein F	gi 118411168	*T. pseudonana*	Photosynthesis	C	20362	8.9	11.9
Photosystem I protein L	gi 118411163	*T. pseudonana*	Photosynthesis	C	15704	9.3	18.2
Photosystem II 10 kDa phosphoprotein	gi 118411116	*T. pseudonana*	Photosynthesis	C	7388	6.0	21.2
Photosystem II 11 kD protein	jgi 3258	*T. pseudonana*	Photosynthesis	C	19602	9.6	5.7
Photosystem II chlorophyll A core antenna apoprotein	gi 118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	22.6
Photosystem II chlorophyll A core antenna apoprotein CP43	gi 118411149	*T. pseudonana*	Photosynthesis	C	51845	7.7	10.2
Photosystem II protein Y	gi 118411171	*T. pseudonana*	Photosynthesis	C	4006	12.5	22.2
Photosystem II reaction center protein D1	gi 118411180	*T. pseudonana*	Photosynthesis	C	39699	5.3	18.3
Photosystem II reaction center protein D2	gi 118411148	*T. pseudonana*	Photosynthesis	C	39064	5.6	17.7
PsbV	gi 118411100	*T. pseudonana*	Photosynthesis	C	17841	7.7	11.0
Putative ribosomal protein S18	jgi 26893	*T. pseudonana*	Translation	R	17159	10.8	21.9
Ribosomal protein L27	jgi 39735	*T. pseudonana*	Translation	R	16090	10.7	9.5
Ribosomal protein L5	jgi 802	*T. pseudonana*	Translation	C	35283	8.6	3.9
Ribosomal protein S12	jgi 37628	*T. pseudonana*	Translation	R	12611	6.2	9.6
Ribosomal protein S13	jgi 26221	*T. pseudonana*	Translation	R	17054	10.4	24.5
Ribosomal protein S3	jgi 28049	*T. pseudonana*	Translation	M	29161	9.3	12.2
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi 118411104	*T. pseudonana*	Photosynthesis	C	54325	6.2	33.9
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	gi 118411103	*T. pseudonana*	Photosynthesis	C	15843	5.1	19.4
Rubisco expression protein	gi 118411164	*T. pseudonana*	Photosynthesis	C	32381	5.9	10.5
S-adenosyl methionine synthetase	jgi 21815	*T. pseudonana*	Metabolism, one carbon	S	50359	5.2	3.2
S-adenosyl-L-homocysteinas protein	jgi 28496	*T. pseudonana*	Metabolism, one carbon	S	52309	5.1	11.4
SPAC22H10.12c	jgi 26136	*T. pseudonana*	GTPase	S	49514	5.4	4.9
Structural constituent of ribosome, 13 kDa	jgi 262056	*T. pseudonana*	Translation	R	13713	9.9	9.7
Structural constituent of ribosome, 14 kDa	jgi 31084	*T. pseudonana*	Translation	R	14755	10.4	9.4
Transaldolase	jgi 27187	*T. pseudonana*	Metabolism,	S	34855	4.8	9.1

Translation elongation factor Tu	gi 118411218	*T. pseudonana*	Carbohydrate Biosynthesis, Protein	M	44458	4.9	11.7
Tubulin beta chain	jgi 31569	*T. pseudonana*	Microtubule based movement	S	49670	4.9	2.2
Ubiquitin	jgi 40669	*T. pseudonana*	Modification, Protein	S	17567	9.9	16.3
UDP-glucose pyrophosphorylase	jgi 262059	*T. pseudonana*	Nucleotidyl-transferase activity	C	47135	5.4	1.9
Unknown	jgi 39299	*T. pseudonana*	Binding, GTP	U	20905	6.8	6.0
Unknown protein	jgi 4820	*T. pseudonana*	n.a.	S	45854	6.6	2.9
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi 2233	*T. pseudonana*	Transport, Proton	S	16720	5.6	10.8
Vacuolar ATP synthase subunit A	jgi 37123	*T. pseudonana*	Transport, Proton	S	68343	5.0	6.0
DNA-directed RNA polymerase subunit gamma	gi 33862040	*Prochlorococcus marinus*	Transcription	S	71995	6.6	1.7
Photosystem II PsbD protein D2	gi 33861713	*Prochlorococcus marinus*	Photosynthesis	S	39325	5.6	6.2

B. Slurry Tube Gel – GOS/Thaps Database

Protein	Annotation	Species	Function	Comp	MW	pI	SC
30S ribosomal protein S5	gi 118411207	*T. pseudonana*	Translation	C	19297	10.3	11.7
30S ribosomal protein S6	gi 118411222	*T. pseudonana*	Translation	C	12043	9.8	10.7
30S ribosomal protein S7	gi 118411217	*T. pseudonana*	Translation	C	17730	10.5	21.2
3-deoxy-7-phosphoheptulonate synthase	jgi 2790	*T. pseudonana*	Biosynthesis, amino acid	C	53939	6.0	2.5
40S ribosomal protein S11	jgi 22535	*T. pseudonana*	Translation	R	19118	10.3	8.5
40S ribosomal protein S23	jgi 28209	*T. pseudonana*	Translation	R	15734	10.5	7.7
40S ribosomal protein SA p40	jgi 21871	*T. pseudonana*	Translation	R	27261	5.9	10.3
50S ribosomal protein L11	gi 118411123	*T. pseudonana*	Translation	C	14880	9.7	9.2
50S ribosomal protein L14	gi 118411201	*T. pseudonana*	Translation	C	13433	10.3	13.2
50S ribosomal protein L23	gi 118411192	*T. pseudonana*	Transferase	C	11152	9.9	10.2
50S ribosomal protein L3	gi 118411190	*T. pseudonana*	Translation	C	22012	10.1	26.1
50S ribosomal protein L4	gi 118411191	*T. pseudonana*	Translation	C	24203	10.2	7.9
60 kDa chaperonin	gi 118411188	*T. pseudonana*	Metabolism	C	57361	5.2	3.2
6-phosphogluconate dehydrogenase	jgi 33343	*T. pseudonana*	Pentose phosphate shunt	S	53348	5.6	9.8
Acetyl-CoA carboxylase	jgi 6770	*T. pseudonana*	Ligase	C	228295	5.0	1.5

Acidic ribosomal phosphoprotein P0	jgi 25812	*T. pseudonana*	Biosynthesis	R	34116	4.8	3.7
Adenosinetriphosphatase	jgi 40156	*T. pseudonana*	Transport, Proton	S	39935	7.6	7.4
Adenylate kinase	jgi 31809	*T. pseudonana*	Metabolism, nucleic acid	S	24598	6.3	4.4
ATP synthase CF0 B chain subunit I	gi 118411110	*T. pseudonana*	Photosynthesis	C	20029	9.8	8.4
ATP synthase CF0 B' chain subunit II	gi 118411109	*T. pseudonana*	Photosynthesis	C	17373	4.6	9.6
ATP synthase CF0 C chain subunit III	gi 118411108	*T. pseudonana*	Photosynthesis	C	8166	5.0	50
ATP synthase CF1 alpha chain	gi 118411112	*T. pseudonana*	Photosynthesis	C	53989	5.0	6
ATP synthase CF1 beta chain	gi 118411134	*T. pseudonana*	Photosynthesis	C	51143	4.7	25.1
ATP/ADP translocator	jgi 39143	*T. pseudonana*	Transport	S	32254	9.4	9.9
ATP-dependent clp protease ATP-binding subunit	gi 118411220	*T. pseudonana*	Catalytic activity	C	102150	6.5	1.5
CbbX protein homolog	jgi 40193	*T. pseudonana*	Binding, ATP	C	35036	5.3	15.1
Cell division protein FtsH-like protein	gi 118411141	*T. pseudonana*	Photosynthesis	C	70206	5.1	3.9
CG11154-PA, isoform A	jgi 41256	*T. pseudonana*	Transport, Proton	S	53388	5.1	6.6
Chloroplast light harvesting protein isoform 12	jgi 270092	*T. pseudonana*	Photosynthesis	C	26078	5.5	6.2
Chloroplast light harvesting protein isoform 15	jgi 2845	*T. pseudonana*	Photosynthesis	C	21873	5.1	5.4
Chloroplast light harvesting protein isoform 5	jgi 32723	*T. pseudonana*	Photosynthesis	C	19175	5.2	10.6
Coatomer protein complex, subunit gamma 2	jgi 269663	*T. pseudonana*	Transport, Protein	S	100270	5.1	1.1
Cytochrome b6	gi 118411154	*T. pseudonana*	Photosynthesis	C	23906	9.2	6
Cytochrome f	gi 118411137	*T. pseudonana*	Photosynthesis	C	33988	8.2	4.1
Elongation factor 2	jgi 269148	*T. pseudonana*	GTPase	S	91887	6.0	8.7
Elongation factor alpha-like protein	jgi 41829	*T. pseudonana*	GTPase	S	49969	8.7	7.1
Enoyl-acyl carrier reductase	jgi 32860	*T. pseudonana*	Oxidation reduction	S	32813	5.1	3.5
Eukaryotic translation initiation factor 4A2 isoform 2	jgi 9716	*T. pseudonana*	Binding, DNA	S	42405	5.6	3.5
Fructose-1,6-bisphosphate aldolase precursor	jgi 428	*T. pseudonana*	Glycolysis	S	39810	4.8	8.1
Fucoxanthin chlorophyl a/c protein, 21.8 kDa	jgi 38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyll a /c protein, 20.3 kDa	jgi 38494	*T. pseudonana*	Photosynthesis	C	20354	4.5	11.6

Fucoxanthin chlorophyll a/c binding protein, 22.2 kDa	jgi 264921	*T. pseudonana*	Photosynthesis	C	22205	4.6	8.6
Fucoxanthin chlorophyll a/c binding protein, 22.6 kDa	jgi 268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	17.1
GDP-mannose 3,5-epimerase	jgi 41548	*T. pseudonana*	Coenzyme	S	40707	5.1	3.3
Geranyl-geranyl reductase	jgi 10234	*T. pseudonana*	Metabolic, Aromatic Compound	C	47230	5.9	10.3
Glyceraldehyde-3-phosphate dehydrogenase	jgi 28334	*T. pseudonana*	Glycolysis	M	36574	6.1	4.4
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi 31383	*T. pseudonana*	Glycolysis	S	39587	5.3	5.3
Glycolaldehydetransferase	jgi 21175	*T. pseudonana*	Transport	M	71708	5.0	2.3
Heat shock protein 70	jgi 269120	*T. pseudonana*	Morphogenesis, Cell	S	71187	4.8	8.1
Hsp70-type chaperone	gi 118411189	*T. pseudonana*	Transcription	C	65339	4.8	2.3
HSP90-like protein	jgi 22766	*T. pseudonana*	Binding, ATP	S	80966	4.7	2
Hypothetical protein CBG08717	jgi 269322	*T. pseudonana*	Transport, Proton	S	58009	5.8	7.7
Hypothetical protein CdifQ_02003487	jgi 36462	*T. pseudonana*	Metabolic	U	21340	4.8	6.7
Hypothetical protein FG01081.1	jgi 25949	*T. pseudonana*	Translation	R	20124	9.8	7.9
Iron-sulfur cluster formation ABC transporter	gi 118411105	*T. pseudonana*	Photosynthesis	C	54338	6.9	2.5
Ketol-acid reductoisomerase	jgi 23228	*T. pseudonana*	Biosynthesis, amino acid	S	58240	5.1	3.2
L4/L1	jgi 22610	*T. pseudonana*	Translation	R	40991	10.3	7.7
Oxygen-evolving enhancer protein 1 precursor	jgi 34830	*T. pseudonana*	Photosynthesis	S	29136	5.2	23.6
Phosphoglycerate kinase precursor	jgi 35712	*T. pseudonana*	Glycolysis	S	42256	5.0	3.2
Phosphoribulokinase	jgi 4376	*T. pseudonana*	Biosynthesis	C	42389	4.9	3.3
Photosystem I ferredoxin-binding protein	gi 118411153	*T. pseudonana*	Photosynthesis	C	15518	9.6	20.1
Photosystem I p700 chlorophyll A apoprotein A	gi 118411096	*T. pseudonana*	Photosynthesis	C	83642	7.3	3.2
Photosystem I p700 chlorophyll A apoprotein B	gi 118411097	*T. pseudonana*	Photosynthesis	C	82090	7.6	4.1
Photosystem I protein F	gi 118411168	*T. pseudonana*	Photosynthesis	C	20362	8.9	11.9
Photosystem I protein L	gi 118411163	*T. pseudonana*	Photosynthesis	C	15704	9.3	18.2
Photosystem II 10 kDa phosphoprotein	gi 118411116	*T. pseudonana*	Photosynthesis	C	7388	6.0	21.2
Photosystem II 11 kD protein	jgi 3258	*T. pseudonana*	Photosynthesis	C	19602	9.6	5.7

Photosystem II chlorophyll A core antenna apoprotein	gi 118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	22.6
Photosystem II reaction center protein D1	gi 118411180	*T. pseudonana*	Photosynthesis	C	39699	5.3	13.1
Photosystem II reaction center protein D2	gi 118411148	*T. pseudonana*	Photosynthesis	C	39064	5.6	17.9
PsbV	gi 118411100	*T. pseudonana*	Photosynthesis	C	17841	7.7	11
Ribosomal protein L27	jgi 39735	*T. pseudonana*	Translation	R	16090	10.7	9.5
Ribosomal protein L5	jgi 802	*T. pseudonana*	Translation	C	35283	8.6	3.9
Ribosomal protein S12	jgi 37628	*T. pseudonana*	Translation	R	12611	6.2	9.6
Ribosomal protein S18	jgi 26893	*T. pseudonana*	Translation	R	17159	10.8	21.9
Ribosomal protein S3	jgi 28049	*T. pseudonana*	Translation	M	29161	9.3	7.8
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi 118411104	*T. pseudonana*	Photosynthesis	C	54325	6.2	27.3
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	gi 118411103	*T. pseudonana*	Photosynthesis	C	15843	5.1	19.4
Rubisco expression protein	gi 118411164	*T. pseudonana*	Photosynthesis	C	32381	5.9	10.5
S-adenosyl methionine synthetase	jgi 21815	*T. pseudonana*	Metabolic, One carbon	S	50359	5.2	3.2
S-adenosyl-L-homocysteinas protein	jgi 28496	*T. pseudonana*	Metabolic, One carbon	S	52309	5.1	7.9
SPAC22H10.12c	jgi 26136	*T. pseudonana*	GTPase	S	49514	5.4	4.9
Structural constituent of ribosome	jgi 31084	*T. pseudonana*	Translation	R	14755	10.4	9.4
Transaldolase	jgi 27187	*T. pseudonana*	Metabolism, carbohydrate	S	34855	4.8	9.1
Translation elongation factor Tu	gi 118411218	*T. pseudonana*	Biosynthesis, Protein	M	44458	4.9	8.1
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi 2233	*T. pseudonana*	Transport, Proton	S	16720	5.6	10.8
Vacuolar ATP synthase subunit A	jgi 37123	*T. pseudonana*	Transport, Proton	S	68343	5.0	4.8
50S ribosomal protein L22	gi 118411196	*T. pseudonana*, JCVI_PEP-1096694753977	Translation	C	12986	10.3	11.3
Putative CDC48/ATPase	jgi 267952	*T. pseudonana*, JCVI_PEP-1096692712161	Binding, ATP	S	89464	4.8	7.3
Glucose-6-phosphate dehydrogenase	jgi 34514	*T. pseudonana*, JCVI_PEP-1096675866109	Metabolism, Glucose	S	57149	7.6	2.8
Citrate synthase	jgi 11411	*T. pseudonana*, JCVI_PEP-	Transferase	M	52269	6.2	6

		1096675019495					
Core histone	jgi 3184	*T. pseudonana*, JCVI_PEP-1096673929279	Nucleosome assembly	N	15268	11.5	25.2
Photosystem II chlorophyll A core antenna apoprotein CP43	gi 118411149	*T. pseudonana*, JCVI_PEP-1096672603429	Photosynthesis	C	51845	7.7	16.1
GDP-mannose 4,6-dehydratase	jgi 40586	*T. pseudonana*, JCVI_PEP-1096671414083	Metabolic, GDP-mannose	S	40656	5.9	6.2
Unknown	JCVI 1096683416369	Alpha proteobacterium	n.a.	S	24195	6.9	14.3
Isocitrate dehydrogenase, NADP-dependent	JCVI 1096689319217	*Azotobacter vinelandii*	Metabolism, Isocitrate process	S	26694	8.6	7.7
2-alkenal reductase	JCVI 1096666945533	*Bursaphelenchus xylophilus*	Modification, RNA	S	50833	n.a.	12.1
DNA-directed RNA polymerase, beta' subunit	JCVI 1096665662829	*Candidatus Pelagibacter ubique*	Transcription	S	154508	n.a.	2
FMN-dependent dehydrogenase	JCVI 1096674553715	*Candidatus Pelagibacter ubique*	Oxidation reduction	S	24460	8.8	9.6
Penicillin-binding protein, 1A family	JCVI 1096687481801	*Candidatus Pelagibacter ubique*	Biosynthesis, peptidoglycan	S	31421	9.3	6.1
Adenylylsulfate reductase, alpha subunit	JCVI 1096672730853	*Desulfovibrio vulgaris*	Oxidation reduction	S	15446	6.2	5.5
ATP synthase F1, alpha subunit	JCVI 1096675315161	*Emiliania huxleyi*	Synthesis, ATP	C	53427	n.a.	4.2
Unknown	JCVI 1096665982711	*Flavobacteriales bacterium ALC-1*	n.a.	S	n.a.	n.a.	8.8
Beta-tubulin	JCVI 1096668084639	*Gastrostyla steinii*	Microtubule based movement	S	25577	4.7	12.5
Unknown	JCVI 1096673080787	*Hydrogenobaculum sp.*	n.a.	S	30796	9.1	4.9
Extracellular solute-binding protein, family 1	JCVI 1096666167981	*Labrenzia aggregata*	Transport	S	70965	n.a.	4.8

Glyceraldehyde-3-phosphate dehydrogenase, type I	JCVI 1096682420143	*Microscilla marina*	Glycolysis	S	27653	4.8	6.2
Argininosuccinate lyase	JCVI 1096681073423	*Natronomonas pharaonis*	Biosynthesis, arginine	S	22803	10.0	11.2
Unknown	JCVI 1096694739943	*Porphyra yezoensis*	n.a.	U	44881	6.4	15.8
Actin	JCVI 1096673373927	*Prymnesium parvum*	Binding, Protein	S	28829	n.a.	5
Fructose-bisphosphate aldolase class-I	JCVI 1096678563861	*Pseudoalteromonas atlantica*	Glycolysis	S	29476	4.4	4.9
Peptidase M24	JCVI 1096682068811	*Pseudoalteromonas atlantica*	Proteolysis	S	49553	6.3	2.1
Unknown	JCVI 1096691696875	*Synechococcus sp.*	n.a.	S	32778	5.5	7.8
Actin	JCVI 1096668934071	*Vannella ebro*	Binding, Protein	S	25450	5.4	17.5

Appendix A.4.

Table A4.1. Proteins that were up- or down-regulated within the investigation. The latter time point is what is being referred to as up- or down-regulated. *

Protein	T_1 v T_2	T_1 v T_3	T_1 v T_4	Gene Ontology
Q5LTX5	Up	Up	Up	No definition
Q5LWL1	Up	Up	Up	cytoplasm; ribosome biogenesis
Q5LPU6	Up	Up	Up	cyclic nucleotide biosynthetic process; phosphorelay response regulator activity; phosphorus-oxygen lyase activity; regulation of transcription, DNA-dependent
Q5LV09	Up	Up	Up	'de novo' IMP biosynthetic process; 5-(carboxyamino)imidazole ribonucleotide mutase activity; lyase activity
Q5LPU9	Up	Up	Up	methanogenesis; methyltransferase activity
Q5LND9	Up	Up	Up	glutamate biosynthetic process; glutamate synthase activity
Q5LPU8	Up	Up	Up	hydrolase activity
Q5LP95	Up	Up	Up	No definition
Q5LQ46	Up	Up	Up	methyltransferase activity
Q5LQZ7	Up	Up	Up	ATP binding; ATP catabolic process; ATPase activity, coupled to transmembrane movement of substances; integral to membrane
Q5LW86	Up	Up	Up	No definition
Q5LWE4	Up	Up	Up	No definition
Q5LWT6	Up	Up	Up	cytoplasm; dimethylallyl diphosphate biosynthetic process; hydrolase activity; isopentenyl-diphosphate delta-isomerase activity; isoprenoid biosynthetic process; metal ion binding
Q5LTW7	Up	Up	Up	catalytic activity; regulation of biosynthetic process
Q5LS80	Up	Up	Up	ligase activity, forming carbon-nitrogen bonds
Q5LLS0	Up	Up	Up	No definition
Q5LTG0	Up	Up	Up	outer membrane-bounded periplasmic space; transport
Q5LS64	Up	Up	Up	fumarate hydratase activity; fumarate metabolic process; tricarboxylic acid cycle; tricarboxylic acid cycle enzyme complex
Q5LVP4	Up	Up	Up	3'-5' exonuclease activity; DNA binding; DNA replication; DNA-directed DNA polymerase activity
Q5LTX0	Up	Up	Up	No definition
Q5LUI7	Up	Up	Up	No definition
Q5LWC8	Up	Up	-	metalloendopeptidase activity; proteolysis
Q5LV36	Up	Up	-	gamma-glutamyltransferase activity; glutathione metabolic process

Q5LN82	Up	Up	-	RNA binding; RNA metabolic process; cytoplasm; exoribonuclease II activity; nucleic acid phosphodiester bond hydrolysis
Q5LR15	Up	Up	-	ATP binding; ATP catabolic process; cytochrome complex assembly; heme-transporting ATPase activity; outer membrane-bounded periplasmic space; plasma membrane
Q5LQ93	Up	-	Up	carbohydrate binding; carbohydrate metabolic process; sequence-specific DNA binding transcription factor activity
Q5LMZ5	Up	-	Up	biosynthetic process; pyridoxal phosphate binding; transaminase activity
Q5LWA5	Up	-	Up	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LPW4	-	Up	Up	ATP binding; DNA binding; DNA modification; Type I site-specific deoxyribonuclease activity
Q5LNY1	-	Up	Up	membrane
Q5LP88	-	Up	Up	oxidoreductase activity; zinc ion binding
Q5LNA9	-	Up	Up	hydrolase activity
Q5LS03	-	Up	Up	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
Q5LQ06	-	Up	Up	arsenite methyltransferase activity; cytosol
Q5LSD8	-	Up	Up	No definition
Q5LVP6	-	Up	Up	2 iron, 2 sulfur cluster binding; UDP-N-acetylmuramate dehydrogenase activity; electron carrier activity; flavin adenine dinucleotide binding; metal ion binding; xanthine dehydrogenase activity; xanthine oxidase activity
Q3V7J3	-	Up	Up	cytoplasm; pyridoxine 5'-phosphate synthase activity; pyridoxine biosynthetic process
Q5LX28	-	Up	Up	ubiquinone biosynthetic process
Q5LW38	-	Up	Up	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LLL6	Up	-	-	adenyl nucleotide binding
Q5LNA3	Up	-	-	oxidoreductase activity
Q5LWD9	Up	-	-	No definition
Q5LT31	Up	-	-	sequence-specific DNA binding
Q5LQB9	Up	-	-	outer membrane-bounded periplasmic space; transport
Q5LN66	Up	Down	-	4 iron, 4 sulfur cluster binding; cytoplasm; metal ion binding; rRNA (adenine-C2-)-methyltransferase activity; rRNA base methylation; rRNA binding; tRNA (adenine-C2-)-methyltransferase activity; tRNA binding
Q5LQ03	Up	-	-	intein-mediated protein splicing
Q5LMS4	Up	-	-	No definition
Q5LQK1	-	Up	-	ATP binding; ATP catabolic process; ATPase activity
Q5LWB7	-	Up	-	No definition
Q5LTW9	-	Up	-	No definition
Q5LPB8	-	Up	-	cytidine deaminase activity; zinc ion binding

Q5LRX8	-	Up	-	No definition
Q5LMQ3	-	Up	-	LSU rRNA binding; ribosome; ribosome biogenesis; structural constituent of ribosome; translation
Q5LXE1	-	Up	-	L-malate dehydrogenase activity; cellular carbohydrate metabolic process; malate metabolic process; tricarboxylic acid cycle
Q5LX75	-	Up	-	'de novo' UMP biosynthetic process; 'de novo' pyrimidine nucleobase biosynthetic process; amino acid binding; aspartate carbamoyltransferase activity; cellular amino acid metabolic process
Q5LMN1	-	Up	-	coproporphyrinogen oxidase activity; cytoplasm; iron-sulfur cluster binding; porphyrin-containing compound biosynthetic process
Q5LWA4	-	Up	-	transferase activity
Q5LTX1	-	Up	-	hydrolase activity
Q5LQ87	-	Up	-	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LNW6	-	Up	-	protein secretion
Q5LQ48	-	Up	-	No definition
Q5LS75	-	Up	-	RNA binding; RNA processing; ribonuclease activity
Q5LR82	-	Up	-	cysteine biosynthetic process from serine; cytoplasm; serine O-acetyltransferase activity
Q5LQK2	-	Up	-	3-hydroxyacyl-CoA dehydrogenase activity
Q5LQW0	-	-	Up	No definition
Q5LRU8	-	-	Up	No definition
Q5LP46	-	-	Up	cell wall macromolecule catabolic process
Q5LR07	-	-	Up	ATP binding; cytoplasm; selenocysteine biosynthetic process; selenocysteinyl-tRNA(Sec) biosynthetic process; serine-tRNA ligase activity; seryl-tRNA aminoacylation
Q5LQR1	-	-	Up	oxidation-reduction process; oxidoreductase activity
Q5LM79	Down	Down	Down	No definition
Q5LNE7	Down	Down	Down	carbohydrate metabolic process; cell wall macromolecule catabolic process; lysozyme activity; peptidoglycan catabolic process
Q5LUZ3	Down	Down	Down	No definition
Q5LNF3	Down	Down	Down	chorismate metabolic process; chorismate mutase activity
Q5LPN7	Down	Down	Down	hydrolase activity
Q5LP94	Down	Down	Down	5-formyltetrahydrofolate cyclo-ligase activity; ATP binding; folic acid-containing compound biosynthetic process
Q5LSV2	Down	Down	Down	cytoplasm; translational termination
Q5LQJ1	Down	Down	Down	No definition
Q5LUZ8	Down	Down	Down	intracellular; sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity; transcription, DNA-dependent

Q5LQJ3	Down	Down	Down	DNA binding; nucleotide binding; regulation of transcription, DNA-dependent
Q5LRC8	Down	Down	Down	adenyl nucleotide binding
Q5LQY8	Down	Down	Down	No definition
Q5LNF5	Down	Down	Down	N-acetyltransferase activity
Q5LRF6	Down	Down	Down	ATP binding; integral to membrane; phosphorelay sensor kinase activity; signal transduction by phosphorylation
Q5LRB1	Down	Down	Down	ATP binding; ATP catabolic process; ATP-binding cassette (ABC) transporter complex; ATPase activity; DNA binding; SOS response; cytoplasm; excinuclease ABC activity; excinuclease repair complex; nucleic acid phosphodiester bond hydrolysis; nucleotide-excision repair; plasma membrane; transport; zinc ion binding
Q5LPA3	Down	Down	Down	No definition
Q5LQW3	Down	Down	Down	alanine:sodium symporter activity; membrane
Q5LQQ3	Down	Down	Down	hydrolase activity
Q5LU58	Down	Down	Down	UDP-N-acetylmuramate dehydrogenase activity; cell cycle; cell division; cytoplasm; flavin adenine dinucleotide binding; peptidoglycan biosynthetic process; regulation of cell shape
Q5LNG3	Down	Down	Down	aminoacyl-tRNA editing activity; aminoacyl-tRNA ligase activity; regulation of translational fidelity
Q5LRM3	Down	Down	Down	4 iron, 4 sulfur cluster binding; cytoplasm; lipoate synthase activity; metal ion binding; protein lipoylation
Q5LS82	Down	Down	Down	No definition
Q5LNC3	Down	Down	-	phenylacetate-CoA ligase activity
Q5LMH8	Down	Down	-	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides; metal ion binding
Q5LQY7	Down	-	Down	aminomethyltransferase activity; cytoplasm; glycine catabolic process; sarcosine oxidase activity; tetrahydrofolate metabolic process
Q5LPA8	Down	-	Down	cytoplasm; glycine biosynthetic process from serine; glycine hydroxymethyltransferase activity; pyridoxal phosphate binding; tetrahydrofolate interconversion
Q5LRP8	-	Down	Down	No definition
Q5LPP5	-	Down	Down	ATP binding; ATP catabolic process; ATPase activity
Q5LRU9	-	Down	Down	No definition
Q5LMA4	-	Down	Down	carbohydrate metabolic process; catalytic activity
Q5LU74	-	Down	Down	UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelyl-D-alanyl-D-alanine:undecaprenyl-phosphate transferase activity; cell cycle; cell division; integral to membrane; peptidoglycan biosynthetic process; phospho-N-acetylmuramoyl-pentapeptide-transferase activity; plasma membrane; regulation of cell shape
Q5LP33	-	Down	Down	oxidoreductase activity

Q5LTW6	-	Down	Down	No definition
Q5LRA3	-	Down	Down	cytoplasm; isomerase activity; queuosine biosynthetic process; transferase activity
Q5LP43	Down	-	-	intracellular; sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity; transcription, DNA-dependent
Q5LQY3	Down	-	-	No definition
Q5LPB5	Down	-	-	carbon-sulfur lyase activity
Q5LR36	Down	-	-	No definition
Q5LMQ0	Down	-	-	large ribosomal subunit; rRNA binding; regulation of translation; structural constituent of ribosome; tRNA binding; translation
Q5LNC5	Down	-	-	No definition
Q5LPQ3	Down	-	-	metabolic process; pyridoxal phosphate binding; transaminase activity
Q5LN54	Down	-	-	hydrolase activity
Q5LVE9	Down	-	-	2 iron, 2 sulfur cluster binding; electron carrier activity; metal ion binding; oxidoreductase activity; phenylacetate catabolic process; transport
Q5LMR4	Down	-	-	GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity
Q5LR94	Down	-	-	ATP binding; cytoplasm; glycolysis; phosphoglycerate kinase activity
Q5LTZ5	Down	-	-	sequence-specific DNA binding
Q5LNT3	Down	-	-	protein methyltransferase activity; ribosome
Q5LNF6	Down	-	-	N-acetyltransferase activity
Q5LTX6	-	-	Down	DNA binding; response to stress; sequence-specific DNA binding transcription factor activity; transcription, DNA-dependent
Q5LSH2	-	-	Down	4 iron, 4 sulfur cluster binding; anaerobic respiration; electron carrier activity; formate dehydrogenase (NAD+) activity; formate dehydrogenase complex; molybdenum ion binding
Q5LQM1	-	-	Down	agmatinase activity; metal ion binding
Q5LQ51	-	-	Down	proteolysis; serine-type endopeptidase activity
Q5LTZ3	-	-	Down	intracellular; sequence-specific DNA binding transcription factor activity
Q5LPK0	-	-	Down	No definition
Q5LP97	-	-	Down	No definition
Q5LST0	-	-	Down	5,1No definition-methylenetetrahydrofolate-dependent tRNA (m5U54) methyltransferase activity; cytoplasm; flavin adenine dinucleotide binding; methylenetetrahydrofolate-tRNA-(uracil-5-)-methyltransferase (FADH2-oxidizing) activity
Q5LLV4	-	-	Down	ATP binding; acetyl-CoA carboxylase activity; acetyl-CoA carboxylase complex; fatty acid biosynthetic process; malonyl-CoA biosynthetic process; transferase activity; zinc ion binding
Q5LLN6	-	-	Down	No definition
Q5LMW1	-	-	Down	DNA binding; DNA methylation; N-methyltransferase activity

Q5LLU9	-	-	Down	hydrolase activity
Q5LQV3	-	-	Down	outer membrane-bounded periplasmic space
Q5LNG2	-	-	Down	4 iron, 4 sulfur cluster binding; NAD biosynthetic process; cytoplasm; metal ion binding; quinolinate biosynthetic process; quinolinate synthetase A activity; transferase activity, transferring alkyl or aryl (other than methyl) groups
Q5LQY6	-	-	Down	sarcosine oxidase activity; tetrahydrofolate metabolic process
Q5LS76	-	-	Down	ATP binding; intracellular signal transduction; nucleoside-triphosphatase activity; phosphorelay response regulator activity; regulation of transcription, DNA-dependent; sequence-specific DNA binding; transcription, DNA-dependent
Q5LQI5	-	-	Down	GTP metabolic process; dGTPase activity; magnesium ion binding; phosphoric diester hydrolase activity

* Proteins were considered to be up- or down-regulated if the corresponding false discovery rate (FDRup or FDRdown) was $p < 0.05$.

Table A4.2. DAVID results for all non-overlapping proteins. The number of proteins associated with each annotation category is given as counts.

		Counts			
Cluster	Gene Ontology Category or Pathway	T_1	T_2	T_3	T_4
1	GO:0006412~translation	95	97	92	92
	GO:0019538~protein metabolic process	207	205	180	168
	GO:0003735~structural constituent of ribosome	53	53	52	52
	GO:0044267~cellular protein metabolic process	152	149	136	129
	GO:0003723~RNA binding	66	63	61	60
	GO:0019843~rRNA binding	35	35	35	35
	GO:0042255~ribosome assembly	33	33	32	32
	GO:0022618~ribonucleoprotein complex assembly	33	33	32	32
	GO:0042254~ribosome biogenesis	46	45	43	41
	GO:0022613~ribonucleoprotein complex biogenesis	46	45	43	41
	GO:0034621~cellular macromolecular complex subunit organization	50	49	46	45
	GO:0005198~structural molecule activity	56	61	56	55
	GO:0034622~cellular macromolecular complex assembly	42	-	39	38
	sil03010:Ribosome	-	-	52	52
	GO:0043933~macromolecular complex subunit organization	-	-	49	48
		914	868	948	920
2	GO:0044271~nitrogen compound biosynthetic process	210	202	184	180
	GO:0008652~cellular amino acid biosynthetic process	93	97	90	87
	GO:0006520~cellular amino acid metabolic process	167	173	148	152
	GO:0009309~amine biosynthetic process	99	102	94	92
	GO:0043436~oxoacid metabolic process	273	276	222	235
	GO:0016053~organic acid biosynthetic process	119	123	113	110
	GO:0046394~carboxylic acid biosynthetic process	119	123	113	110
	GO:0044106~cellular amine metabolic process	177	182	155	159
		1257	1278	1119	1125
3	GO:0055086~nucleobase, nucleoside and nucleotide metabolic process	104	100	88	93
	GO:0034654~nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process	72	68	64	67
	GO:0034404~nucleobase, nucleoside and nucleotide biosynthetic process	72	68	64	67
	GO:0009259~ribonucleotide metabolic process	39	39	38	38
	GO:0009260~ribonucleotide biosynthetic process	38	38	37	37
	GO:0009165~nucleotide biosynthetic process	56	-	51	53
	GO:0009150~purine ribonucleotide metabolic process	-	-	27	28
	GO:0006163~purine nucleotide metabolic process	-	-	30	31
	GO:0009152~purine ribonucleotide biosynthetic process	-	-	26	27

	GO:0006164~purine nucleotide biosynthetic process	-	-	28	29
	GO:0009156~ribonucleoside monophosphate biosynthetic process	-	-	-	15
	GO:0009161~ribonucleoside monophosphate metabolic process	-	-	-	15
		381	313	453	500
4	GO:0000166~nucleotide binding	324	313	250	257
	GO:0032553~ribonucleotide binding	234	222	183	186
	GO:0032555~purine ribonucleotide binding	234	222	183	186
	GO:0017076~purine nucleotide binding	278	268	-	218
	GO:0001882~nucleoside binding	256	248	-	-
	GO:0005524~ATP binding	206	-	-	-
	GO:0001883~purine nucleoside binding	255	247	-	-
	GO:0030554~adenyl nucleotide binding	255	247	-	-
	GO:0032559~adenyl ribonucleotide binding	211	-	-	-
		2253	1767	616	847
5	GO:0034660~ncRNA metabolic process	55	50	45	-
	GO:0006399~tRNA metabolic process	44	-	36	-
	GO:0043038~amino acid activation	27	27	26	26
	GO:0006418~tRNA aminoacylation for protein translation	26	26	25	25
	GO:0043039~tRNA aminoacylation	26	26	25	25
	GO:0016070~RNA metabolic process	78	76	66	-
	GO:0004812~aminoacyl-tRNA ligase activity	23	23	22	23
	GO:0016876~ligase activity, form aminoacyl-tRNA and related cmpds	23	23	22	23
	GO:0016875~ligase activity, forming carbon-oxygen bonds	23	23	22	23
	sil00970:Aminoacyl-tRNA biosynthesis	25	25	24	24
		350	299	313	169
6	GO:0051186~cofactor metabolic process	121	-	98	-
	GO:0006732~coenzyme metabolic process	84	-	72	-
		205	-	170	-
7	GO:0005525~GTP binding	25	-	-	-
	GO:0032561~guanyl ribonucleotide binding	25	-	-	-
	GO:0019001~guanyl nucleotide binding	25	-	-	-
		75	-	-	-
8	GO:0016655~ oxidoreductase activity, acting on NADH or NADPH	-	-	22	22
	GO:0016651~oxidoreductase activity, acting on NADH or NADPH	-	-	34	34
		-	-	56	56

- : indicates that the annotation category was not determined to be enriched at that time point.

Appendix A.5.

Text A5.1. Description of ship-board labeling experiment.

Chapter 5 investigated whether the incorporation of tracer level addition of labeled amino acid to a model, heterotrophic marine bacterium could be tracked in the proteins it produces. Prior to attempting that investigation, a field-based, shipboard incubation experiment was completed with water collected at approximately 30 m depth from the Chukchi Sea during the summer of 2010. An isotopically labeled amino acid (^{13}C-leucine, concentration of 1 μM) was added to filtered (3 μm) natural seawater.

Carboys were kept in the dark at 4°C for the duration of the incubation and were shaken daily. Sampling was conducted at 49.5, 97, and 170 hours after the labeled leucine was added. The samples were filtered onto 47 mm polycarbonate filters for proteomic analysis or onto 25 mm GF/F filters for bulk isotope analysis. At the first two time points, 800 mL was filtered for proteomic analysis and 100 mL was filtered for isotope analysis. At the final time point, 500 mL was filtered for proteomic analysis and 200 mL was filtered for isotope analysis. All filters were rinsed with 3 mL of 5% TCA and then 3 mL of 80% ethanol. Two samples for bulk isotope analysis were collected at each time point. One sample from the first time point was analyzed twice for bulk isotope analysis, both samples from the second and third time points were analyzed twice for bulk isotope analysis. Proteomic analysis was also conducted on each time point.

Figure 5.1 shows the incorporation of the label into bacterial protein as measured by standard combustion methods for stable isotope analysis with a reference point at time zero. There was a substantial increase in the bulk isotopic in the second and third time points of the incubation. In addition, we were able to confidently identify one labeled

protein, ATP synthase. This provided evidence that the incorporation of label into bacterial proteins could be tracked and served as the basis of the experiment detailed in this chapter.

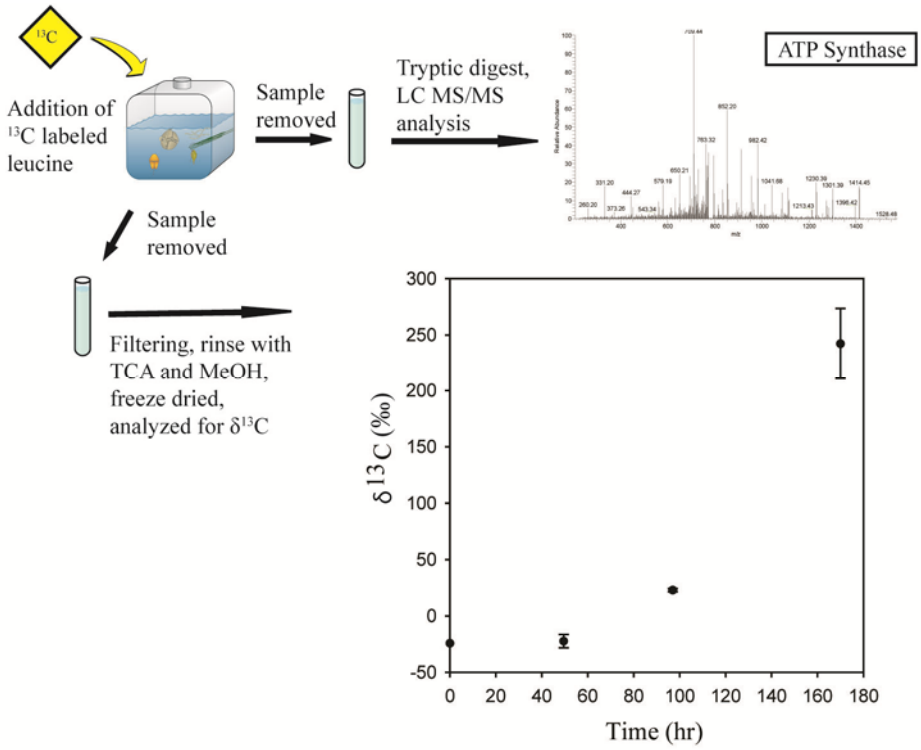


Figure A5.1. Representation of methods used and the results of the uptake of isotopically labeled leucine (^{13}C-leucine) in shipboard incubations of waters collected from the Chukchi Sea. Samples were collected for bulk carbon isotope analysis displayed here as δ^{13}C (‰) (referenced to V-PDB) with standard error bars. The point at time zero represents the expected *in situ* δ^{13}C value determined as a mean of observed particulate organic matter (POM) δ^{13}C values in polar regions (Guo et al., 2004; Hobson et al., 2002; Iken et al., 2005). ATP synthase was identified through proteomic analysis of the ^{13}C labeled proteins.

Table A5.1. Total labeled proteins identified within the investigation. Presence in a time point is denoted by X. X* represents the presence of a labeled leucine within the protein. Proteins absent from a time point are denoted by -.

	Time Points				
Protein (UniProt ID)	T_1	T_2	T_3	T_4	Gene Ontology
Q5LV47	X*	X*	X	X	xylulokinase activity; xylulose metabolic process
Q5LRI5	X	X*	X	X	XMP salvage; magnesium ion binding; plasma membrane; purine ribonucleoside salvage; xanthine phosphoribosyltransferase activity
Q5LT94	X*	X*	X*	X*	UDP-N-acetylmuramate dehydrogenase activity; flavin adenine dinucleotide binding
Q5LKG9	X*	X	X	X	UDP-N-acetylmuramate dehydrogenase activity; flavin adenine dinucleotide binding
Q5LL21	X	X*	X*	X*	UDP-N-acetylmuramate dehydrogenase activity; flavin adenine dinucleotide binding
Q5LU58	X	X*	X	X	UDP-N-acetylmuramate dehydrogenase activity; cell cycle; cell division; cytoplasm; flavin adenine dinucleotide binding; peptidoglycan biosynthetic process; regulation of cell shape
Q5LL90	X	X	X*	X*	UDP-N-acetylgalactosamine biosynthetic process; UDP-N-acetylglucosamine 1-carboxyvinyltransferase activity; cell cycle; cell division; cytoplasm; peptidoglycan biosynthetic process; regulation of cell shape
Q5LU63	X	X*	X*	X*	UDP-N-acetyl-D-glucosamine:N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelyl-D-alanyl-D-alanine-diphosphoundecaprenol 4-beta-N-acetylglucosaminlytransferase activity; UDP-N-acetylgalactosamine biosynthetic process; carbohydrate binding; cell cycle; cell division; lipid glycosylation; peptidoglycan biosynthetic process; plasma membrane; regulation of cell shape; undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase activity
Q5LX28	X	X*	X	X	ubiquinone biosynthetic process
Q5LV87	X*	X*	X	X*	tryptophan synthase activity
Q5LWF6	X	X	X*	X	tRNA binding; tRNA nucleotidyltransferase activity; tRNA processing; tRNA-specific ribonuclease activity
Q5LP55	X*	X*	X*	X*	transporter activity
Q5LPN4	X*	X*	X*	X*	transporter activity
Q5LU48	X*	X*	X*	X*	transporter activity
Q5LVJ5	X*	X*	X*	X*	transporter activity
Q5LVT7	X*	X*	X*	X*	transporter activity
Q5LW24	X*	X*	X*	X*	transporter activity
Q5LSC9	X	X*	X*	X*	transporter activity
Q5LT71	X	X*	X*	X*	transporter activity
Q5LWP6	X	X*	X*	X*	transporter activity
Q5LUC3	X	X*	X	X*	transporter activity
Q5LVC9	X	X	X*	X*	transferase activity, transferring acyl groups other than amino-acyl groups
Q5LMX7	X	X*	X	X	transferase activity

Q5LWA4	X	X*	X	X	transferase activity
Q5LXH5	X*	X*	X	X	transaminase activity
Q5LRR9	X	X*	X	X*	thiamine biosynthetic process; thiamine-phosphate diphosphorylase activity
Q5LTJ9	X	X*	X	X	sterol binding
Q5LRZ4	X*	X*	X	X*	small ribosomal subunit; structural constituent of ribosome; translation
Q5LLU2	X	X*	X*	X*	S-formylglutathione hydrolase activity; carboxylesterase activity; formaldehyde catabolic process
Q5LX50	X*	-	-	-	sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity; transcription, DNA-dependent
Q5LLS1	X*	-	-	-	S-(hydroxymethyl)glutathione dehydrogenase activity; ethanol oxidation; zinc ion binding
Q5LKV7	X	X*	X	X	S-(hydroxymethyl)glutathione dehydrogenase activity; ethanol oxidation; zinc ion binding
Q5LMR2	X*	X*	X*	X*	rRNA binding; small ribosomal subunit; structural constituent of ribosome; tRNA binding; translation
Q5LMR3	X*	X*	X	X*	rRNA binding; small ribosomal subunit; structural constituent of ribosome; tRNA binding; translation
Q5LNM2	X*	X*	X*	X*	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LW41	X*	X	X	X	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LW58	X	X*	X	X	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LW34	X*	X*	X*	X	rRNA binding; ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LW46	X*	X	X*	X	rRNA binding; ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LW55	X*	X	X	X	rRNA binding; ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LLU6	X*	X*	X*	X*	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LLU7	X*	X*	X*	X*	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW43	X*	X*	X*	X*	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW44	X*	X*	X*	X*	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LLT1	X*	X*	X*	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LMP9	X*	X*	X*	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LR48	X*	X*	X*	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW33	X*	X*	X	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LR50	X*	X	X*	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LR49	X*	X	X	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LWV5	X*	X	X	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LMG5	X	X*	X	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW42	X	X*	X	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW45	X	X*	X	X	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LVA0	X*	X*	X*	X*	RNA binding; ribosome; structural constituent of ribosome; translation
Q5LWL2	X	X*	X*	X*	RNA binding; nucleotide binding; regulation of DNA-dependent transcription, termination; sequence-specific DNA binding transcription factor activity
Q5LPT9	X*	X	X	X	RNA binding; hydrolase activity, acting on ester bonds; metal ion binding

Q5LLU5	X*	X	X	X	ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LMQ4	X*	X*	X*	X*	ribosome; structural constituent of ribosome; translation
Q5LNF2	X*	X*	X*	X*	ribosome; structural constituent of ribosome; translation
Q5LSS4	X*	X*	X*	X*	ribosome; structural constituent of ribosome; translation
Q5LSS5	X*	X*	X*	X*	ribosome; structural constituent of ribosome; translation
Q5LUS9	X*	X*	X*	X*	ribosome; structural constituent of ribosome; translation
Q5LW50	X*	X*	X*	X	ribosome; structural constituent of ribosome; translation
Q5LW31	X*	X*	X	X*	ribosome; structural constituent of ribosome; translation
Q5LRY1	X*	X*	X	X	ribosome; structural constituent of ribosome; translation
Q5LNE8	X*	X	X*	X*	ribosome; structural constituent of ribosome; translation
Q5LP83	X*	X	X	X	ribosome; structural constituent of ribosome; translation
Q5LQW2	X*	X*	X*	X*	response to stress
Q5LN03	-	X	X*	X*	regulation of transcription, DNA-dependent
Q5LSA5	X*	X	X*	X	pteridine-containing compound metabolic process
Q5LR69	X	X	X*	X	provirus
Q5LLV9	X	X	X*	X	proteolysis; serine-type peptidase activity
Q5LQ51	-	X	X*	X*	proteolysis; serine-type endopeptidase activity
Q5LL24	X	X	X	X*	proteolysis; serine-type endopeptidase activity
Q5LPW0	X	X	X*	X	proteolysis; serine-type D-Ala-D-Ala carboxypeptidase activity
Q5LU20	X	X	X	X*	protein-L-isoaspartate (D-aspartate) O-methyltransferase activity
Q5LNW6	X	X*	X	X*	protein secretion
Q5LNT3	X	X	X	X*	protein methyltransferase activity; ribosome
Q5LWN3	X	X*	X*	X*	primary metabolic process
Q5LR11	X*	X*	X*	X	P-P-bond-hydrolysis-driven protein transmembrane transporter activity; integral to membrane; intracellular protein transmembrane transport; plasma membrane; protein targeting; protein transport by the Sec complex
Q5LR12	X	X	X*	X	P-P-bond-hydrolysis-driven protein transmembrane transporter activity; integral to membrane; intracellular protein transmembrane transport; plasma membrane; protein targeting; protein transport by the Sec complex
Q5LMY2	X*	X*	X*	X*	porin activity
Q5LNN8	X	X	X	X*	plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1)
Q5LXG2	X*	X*	X*	X*	plasma membrane
Q5LVQ2	X*	X*	X	X*	plasma membrane
Q5LNC3	X*	X	X	X	phenylacetate-CoA ligase activity
Q5LLK1	X	X*	X	X	periplasmic space; thiamine biosynthetic process

Q5LNU4	X	X*	X*	X	periplasmic space; protein import
Q5LMU3	X*	X*	X*	X*	periplasmic space; polyamine binding; polyamine transport
Q5LT08	X	X	X*	X*	periplasmic space; polyamine binding; polyamine transport
Q5LR96	X*	X*	X*	X*	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LR95	X*	X*	X*	X	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LKE3	X	X*	X*	X*	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LWL7	X	X	X*	X	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LXC7	X	X*	X*	X*	oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding; tricarboxylic acid cycle
Q5LX26	X*	X*	X*	X*	oxidoreductase activity; zinc ion binding
Q5LSB6	X*	X	X*	X*	oxidoreductase activity; zinc ion binding
Q5LX27	X	X*	X	X*	oxidoreductase activity; zinc ion binding
Q5LP39	X	X*	X	X	oxidoreductase activity; zinc ion binding
Q5LN30	X	X*	X*	X*	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LVG7	X	X*	X*	X*	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LLJ1	X*	X*	X*	X	oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor; sodium ion transport
Q5LKG7	X*	X*	X	X*	oxidoreductase activity
Q5LNA3	X*	X	X	X	oxidoreductase activity
Q5LP33	X	X	X	X*	oxidoreductase activity
Q5LVQ7	X	X	X	X*	oxidoreductase activity
Q5LV37	X*	X*	X*	X	oxidation-reduction process; oxidoreductase activity
Q5LM28	X*	X	X*	X	oxidation-reduction process; oxidoreductase activity
Q5LMA9	X	X*	X	X	oxidation-reduction process; oxidoreductase activity
Q5LMX2	X	X	X*	X	oxidation-reduction process; oxidoreductase activity
Q5LTK9	-	-	X	X*	oxidation-reduction process; oxidoreductase activity
Q5LM71	X*	X*	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LM98	X*	X*	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LQB9	X*	X*	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LVG4	X*	X*	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LQ68	X*	X	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LKZ1	X	X*	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LQ89	X	X*	X*	X*	outer membrane-bounded periplasmic space; transport
Q5LWY4	X	X*	X*	X*	outer membrane-bounded periplasmic space
Q5LU19	X	X*	X*	X*	outer membrane; protein transport; transporter activity
Q5LSV7	X*	X*	X*	X*	oligopeptide transporter activity

Q5LTI3	X*	X*	X*	X*	O-acetylhomoserine aminocarboxypropyltransferase activity; cysteine metabolic process; pyridoxal phosphate binding
Q5LLU8	X*	X	X*	X	nucleotide binding; rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LRS7	X	X*	X	X	NADP binding; glucose-6-phosphate dehydrogenase activity; pentose-phosphate shunt
Q5LP19	X	X	X*	X	NADP binding; glucose-6-phosphate dehydrogenase activity; pentose-phosphate shunt
Q5LSN3	X	X	X*	X	NADP binding; amino acid binding; cellular amino acid metabolic process; homoserine dehydrogenase activity
Q5LPR9	X*	X*	X*	X	NADH dehydrogenase (quinone) activity
Q5LPM8	X*	X	X*	X*	NAD(P)+ transhydrogenase (AB-specific) activity; nucleotide binding; proton transport
Q5LPM9	X	X*	X	X	NAD(P)+ transhydrogenase (AB-specific) activity; NADP binding; integral to membrane; plasma membrane
Q5LNG0	X*	X	X*	X*	NAD biosynthetic process; nicotinate-nucleotide diphosphorylase (carboxylating) activity
Q5LVK0	X*	X*	X*	X*	NAD binding; NADP binding; glucose metabolic process; oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LPR7	X	X*	X	X*	NAD binding; NADH dehydrogenase (quinone) activity; photosynthesis, light reaction; plasma membrane; quinone binding; transport
Q5LL27	X*	X	X	X	NAD binding; histidine biosynthetic process; histidinol dehydrogenase activity; zinc ion binding
Q5LWG9	X*	X	X*	X*	NAD binding; cytosol; malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity; malate dehydrogenase (oxaloacetate-decarboxylating) activity; malate metabolic process; manganese ion binding; transferase activity, transferring acyl groups
Q5LPB6	X*	X	X	X*	NAD binding; cytosol; malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity; malate dehydrogenase (oxaloacetate-decarboxylating) activity; malate metabolic process; manganese ion binding; transferase activity, transferring acyl groups
Q5LUU0	X*	X	X	X	N2-acetyl-L-ornithine:2-oxoglutarate 5-aminotransferase activity; arginine biosynthetic process; cytoplasm; pyridoxal phosphate binding
Q5LW56	X*	X*	X*	X	mRNA binding; rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LM72	X	X*	X*	X	monooxygenase activity; oxidation-reduction process
Q5LVD5	X	X*	X	X	methyltransferase activity
Q5LUX0	X	X*	X	X	methylmalonyl-CoA epimerase activity; pyruvate fermentation to propionate
Q5LPQ8	X*	X*	X*	X*	methylcrotonoyl-CoA carboxylase activity
Q5LTV6	X*	X	X	X	metal ion binding; transporter activity
Q5LSA3	X*	X*	X*	X	metal ion binding; transketolase activity
Q5LQZ2	X*	X*	X*	X*	metal ion binding; superoxide dismutase activity; superoxide metabolic process
Q5LVC3	X	X	X	X*	metal ion binding; pentose-phosphate shunt; ribulose-phosphate 3-epimerase activity
Q5LU89	X*	X*	X	X	metabolic process; pyridoxal phosphate binding; transaminase activity
Q5LWU0	X	X	X	X*	menaquinone biosynthetic process; methyltransferase activity; ubiquinone biosynthetic process

Q5LSL1	X*	X*	X*	X*	membrane; polysaccharide transmembrane transporter activity
Q5LNA5	X	X*	X	X	lysine biosynthetic process; saccharopine dehydrogenase (NAD+, L-lysine-forming) activity
Q5LMQ3	X*	X*	X*	X*	LSU rRNA binding; ribosome; ribosome biogenesis; structural constituent of ribosome; translation
Q5LN58	X*	X*	X*	X*	L-serine biosynthetic process; O-phospho-L-serine:2-oxoglutarate aminotransferase activity; cytoplasm; pyridoxal phosphate binding
Q5LN57	X*	X*	X*	X*	L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
Q5LM44	X*	X*	X*	X*	L-phenylalanine:2-oxoglutarate aminotransferase activity; aromatic-amino-acid:2-oxoglutarate aminotransferase activity; biosynthetic process; cellular amino acid metabolic process; pyridoxal phosphate binding
Q5LXE1	X*	X*	X*	X	L-malate dehydrogenase activity; cellular carbohydrate metabolic process; malate metabolic process; tricarboxylic acid cycle
Q5LX36	X	X*	X	X	L-isoleucine transaminase activity; L-leucine transaminase activity; L-valine transaminase activity; branched-chain amino acid metabolic process
Q5LS40	X*	X	X	X	lipid A biosynthetic process; transferase activity, transferring acyl groups other than amino-acyl groups
Q5LRT0	X	X*	X*	X*	ligase activity
Q5LW40	X*	X	X	X	large ribosomal subunit; structural constituent of ribosome; translation
Q5LW38	X*	X*	X	X*	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LW57	X*	X*	X	X	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LW48	X*	X	X	X	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LW59	X*	X	X	X	large ribosomal subunit; rRNA binding; structural constituent of ribosome; transferase activity; translation
Q5LMQ0	X*	X*	X	X*	large ribosomal subunit; rRNA binding; regulation of translation; structural constituent of ribosome; tRNA binding; translation
Q5LMI1	X	X*	X	X	kinase activity
Q5LPR2	X	X*	X	X	isomerase activity
Q5LRI3	X	X	X*	X*	isomerase activity
Q5LKR6	X*	X*	X*	X*	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
Q5LS95	X*	X*	X*	X*	iron-sulfur cluster binding; oxidoreductase activity, acting on the aldehyde or oxo group of donors; thiamine pyrophosphate binding
Q5LUX1	-	X*	X	X*	intracellular signal transduction; phosphorelay response regulator activity; regulation of transcription, DNA-dependent
Q5LSF8	X	X	X*	X	integral to membrane; transmembrane transport; transporter activity
Q5LV40	X	X	X	X*	integral to membrane; plasma membrane; transporter activity
Q5LW11	X	X	X*	X	integral to membrane; plasma membrane; protein insertion into membrane; protein transport
Q5LSU8	X	X*	X	X	integral to membrane; metalloendopeptidase activity; plasma membrane; proteolysis
Q5LP40	X*	X	X	X	inositol monophosphate 1-phosphatase activity; inositol phosphate dephosphorylation;

					phosphatidylinositol phosphorylation
Q5LPR1	X*	X	X	X	hydroxymethylglutaryl-CoA lyase activity; metabolic process
Q5LP16	X*	X*	X	X*	hydrolase activity; sterol binding
Q5LPE4	X	X*	X	X	hydrolase activity, acting on ester bonds; metal ion binding; nucleotide binding; nucleotide catabolic process
Q5LPP2	X	X	X	X*	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
Q5LM01	X*	X*	X*	X*	hydrolase activity
Q5LQ96	X*	X*	X*	X*	hydrolase activity
Q5LQQ3	X*	X*	X*	X*	hydrolase activity
Q5LKE5	X*	-	-	X	hydrolase activity
Q5LS84	X*	X*	X	X	homocysteine S-methyltransferase activity
Q5LU44	X*	X	X	X	hippurate hydrolase activity
Q5LPC2	X*	X	X*	X	GTP binding; UMP salvage; magnesium ion binding; uracil phosphoribosyltransferase activity; uracil salvage
Q5LU54	X*	X*	X	X	GTP binding; GTPase activity; barrier septum assembly; cell division site; cytokinesis by binary fission; cytoplasm; protein complex; protein polymerization
Q5LWL4	X*	X*	X*	X*	GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation initiation factor activity
Q5LMR4	X*	X*	X*	X*	GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity
Q5LMR5	X*	X*	X*	X*	GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity
Q5LRU3	X*	X	X	X	GTP binding; GTP catabolic process; GTPase activity
Q5LV85	X*	X*	X	X	GTP binding
Q5LSU6	X*	X	X	X*	Gram-negative-bacterium-type cell outer membrane assembly; cell outer membrane; integral to membrane; plasma membrane; protein insertion into membrane
Q5LMG3	X*	X*	X*	X*	glycolysis; magnesium ion binding; potassium ion binding; pyruvate kinase activity
Q5LR89	X	X*	X	X*	glycolysis; intracellular membrane-bounded organelle; pyruvate dehydrogenase (acetyl-transferring) activity
Q5LLG8	X*	X	X	X	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
Q5LLZ6	X*	X*	X	X	glutamate biosynthetic process; glutamate synthase (NADPH) activity
Q5LVR7	X*	X*	X*	X	gamma-glutamyltransferase activity; glutathione metabolic process
Q5LN84	X*	X	X*	X*	gamma-aminobutyric acid catabolic process; succinate-semialdehyde dehydrogenase [NAD(P)+] activity
Q5LSN4	X*	X*	X	X	fructose 1,6-bisphosphate 1-phosphatase activity; gluconeogenesis; glycerol metabolic process; metal ion binding
Q5LNV2	X*	X	X	X	folic acid-containing compound biosynthetic process; histidine biosynthetic process; methenyltetrahydrofolate cyclohydrolase activity; methionine biosynthetic process; methylenetetrahydrofolate dehydrogenase (NADP+) activity; purine nucleotide biosynthetic process;

					tetrahydrofolate interconversion
Q5LUQ2	X	X	X*	X	flavin adenine dinucleotide binding; oxidoreductase activity; sulfur oxidation
Q5LTP2	X*	X	X	X	flavin adenine dinucleotide binding; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen; ubiquinone biosynthetic process
Q5LPQ5	X*	X*	X*	X*	flavin adenine dinucleotide binding; isovaleryl-CoA dehydrogenase activity
Q5LQ59	X*	X	X	X	ferredoxin-NADP+ reductase activity
Q5LP20	X*	X*	X*	X*	Entner-Doudoroff pathway; phosphogluconate dehydratase activity
Q5LWT8	X*	X*	X	X	enoyl-CoA hydratase activity
Q5LKE2	X*	X*	X*	X	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process
Q5LRI6	X*	X	X	X	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process
Q5LWU3	X	X*	X*	X*	energy reserve metabolic process; transferase activity, transferring acyl groups other than amino-acyl groups
Q5LXE8	X*	X*	X*	X*	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
Q5LX54	X*	X*	X*	X*	electron carrier activity; heme binding; iron ion binding; ubiquinol-cytochrome-c reductase activity
Q5LWT7	X*	X*	X*	X*	electron carrier activity; heme binding; iron ion binding
Q5LMM5	X*	X	X	X	electron carrier activity; heme binding; iron ion binding
Q5LLJ8	X	X	X	X*	electron carrier activity; heme binding
Q5LVI2	X*	X*	X*	X*	electron carrier activity; flavin adenine dinucleotide binding
Q5LVI1	X*	X	X	X	electron carrier activity
Q5LV41	X*	X*	X*	X*	D-xylose transport; monosaccharide binding
Q5LV46	X*	X*	X*	X*	D-xylose metabolic process; cytoplasm; magnesium ion binding; pentose-phosphate shunt; xylose isomerase activity
Q5LTY7	X*	X	X*	X*	dTMP biosynthetic process; flavin adenine dinucleotide binding; thymidylate synthase (FAD) activity
Q5LMP7	X	X*	X	X	DNA-dependent transcription, termination; positive regulation of transcription elongation from RNA polymerase II promoter; transcription antitermination
Q5LRJ1	X*	-	-	-	DNA replication; cobalamin binding; cobalt ion binding; deoxyribonucleotide biosynthetic process; nucleotide binding; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
Q5LQI1	X	X*	X*	X	DNA repair; exodeoxyribonuclease III activity; nucleic acid phosphodiester bond hydrolysis
Q5LS96	X*	X	X	X	DNA binding; sequence-specific DNA binding transcription factor activity; transcription, DNA-dependent
Q5LQJ3	X	X	X	X*	DNA binding; nucleotide binding; regulation of transcription, DNA-dependent
Q5LXF9	X*	X	X	X	DNA binding; intracellular; regulation of transcription, DNA-dependent
Q5LST7	X	X*	X	X	DNA binding; intracellular signal transduction; phosphorelay response regulator activity; regulation of transcription, DNA-dependent

Q5LMQ6	X*	X*	X*	X*	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-dependent
Q5LW32	X*	X*	X*	X*	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-dependent
Q5LMQ5	X*	X*	X*	X*	DNA binding; DNA-directed RNA polymerase activity; ribonucleoside binding; transcription, DNA-dependent
Q5LRY0	X*	X*	X*	X*	DNA binding; DNA repair; rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LQJ4	X	X	X*	X	DNA binding; DNA recombination; regulation of transcription, DNA-dependent; regulation of translation; transcription, DNA-dependent
Q5LV98	X*	X	X*	X*	DNA binding; DNA recombination; chromosome; regulation of transcription, DNA-dependent; regulation of translation; transcription, DNA-dependent
Q5LTY5	X*	X*	X*	X*	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LWA5	X*	X*	X*	X*	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LMD9	X	X*	X*	X	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LM20	X*	X*	X*	X*	DNA binding
Q5LXC8	X	X*	X	X*	dihydrolipoyllysine-residue succinyltransferase activity; oxoglutarate dehydrogenase complex; tricarboxylic acid cycle
Q5LR87	X*	X	X*	X*	dihydrolipoyllysine-residue acetyltransferase activity; pyruvate dehydrogenase complex; pyruvate metabolic process
Q5LQ42	X	X*	X	X	'de novo' UMP biosynthetic process; magnesium ion binding; orotate phosphoribosyltransferase activity
Q5LM52	X	X*	X	X	'de novo' L-methionine biosynthetic process; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; aspartate-semialdehyde dehydrogenase activity; cytoplasm; diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process
Q5LN38	X*	X	X*	X	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
Q5LS81	X	X*	X	X	'de novo' IMP biosynthetic process; ATP binding; phosphoribosylaminoimidazolesuccinocarboxamide synthase activity
Q5LV10	X	X*	X	X	'de novo' IMP biosynthetic process; ATP binding; metal ion binding; phosphoribosylaminoimidazole carboxylase activity
Q5LS98	X	X	X*	X*	'de novo' IMP biosynthetic process; ATP binding; cytoplasm; phosphoribosylformylglycinamidine synthase activity
Q5LQ19	X*	X	X	X	'de novo' IMP biosynthetic process; amidophosphoribosyltransferase activity; iron-sulfur cluster binding; metal ion binding; nucleoside metabolic process; purine nucleobase biosynthetic process
Q5LTU4	X*	X*	X*	X*	'de novo' AMP biosynthetic process; GTP binding; adenylosuccinate synthase activity; cytoplasm; magnesium ion binding
Q5LP36	X*	X*	-	-	cytosol; methionine biosynthetic process; methylenetetrahydrofolate reductase (NADPH) activity; tetrahydrofolate interconversion

Q5LSQ3	-	X*	-	-	cytoplasm; urea catabolic process; urease activity
Q5LSV2	X	X*	X*	X*	cytoplasm; translational termination
Q5LRZ5	X*	X*	X*	X*	cytoplasm; translation elongation factor activity
Q5LLN4	X	X*	X	X	cytoplasm; protein folding; protein tetramerization; protein transport
Q5LNS5	X	X	X*	X*	cytoplasm; pentose-phosphate shunt; sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glyceronetransferase activity
Q5LXH0	X	X	X*	X	cytoplasm; iron ion binding; iron-sulfur cluster assembly; iron-sulfur cluster binding
Q5LNQ6	X	X*	X*	X*	cytoplasm; inorganic diphosphatase activity; manganese ion binding
Q5LPK9	X	X*	X	X	cytoplasm; glyoxylate cycle; malate synthase activity; tricarboxylic acid cycle
Q5LPA8	X*	X*	X*	X	cytoplasm; glycine biosynthetic process from serine; glycine hydroxymethyltransferase activity; pyridoxal phosphate binding; tetrahydrofolate interconversion
Q5LR37	X*	X*	X*	X*	cytoplasm; glutamate-ammonia ligase activity; glutamine biosynthetic process; nitrogen fixation
Q5LQ75	X	X*	X	X	cytoplasm; gluconeogenesis; glycolysis; pentose-phosphate shunt; triose-phosphate isomerase activity
Q5LRS9	X	X*	X*	X*	cytoplasm; gluconeogenesis; glucose-6-phosphate isomerase activity; glycolysis
Q5LUZ9	X	X*	X	X	cytoplasm; flavin adenine dinucleotide binding; removal of superoxide radicals; thioredoxin-disulfide reductase activity
Q5LWM0	X	X	X*	X	cytoplasm; ferrochelatase activity; heme biosynthetic process; metal ion binding
Q5LRV9	X*	X*	X	X	cysteine desulfurase activity; cysteine metabolic process; pyridoxal phosphate binding
Q5LR83	X*	X*	X*	X*	cysteine biosynthetic process from serine; cysteine synthase activity; transferase activity
Q5LS49	X	X	X	X*	cysteine biosynthetic process from serine; cysteine synthase activity; transferase activity
Q5LTP7	X*	X*	X*	X*	coenzyme binding; isoleucine biosynthetic process; ketol-acid reductoisomerase activity; valine biosynthetic process
Q5LPJ3	X*	X	X	X	cobalamin biosynthetic process
Q5LUE9	X	X	X	X*	cobalamin binding; metal ion binding; methylmalonyl-CoA mutase activity
Q5LPJ1	X	X	X*	X*	cellular process; hydrolase activity
Q5LRH1	X*	X*	X*	X*	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
Q5LSM4	X*	X*	X*	X*	cellular amino acid metabolic process; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
Q5LXH3	X*	-	-	-	cellular amino acid metabolic process; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
Q5LUJ2	X*	X*	X*	X	cell wall macromolecule catabolic process; transporter activity
Q5LQ10	X*	X*	X*	X*	cell wall macromolecule catabolic process
Q5LQL4	X*	X*	X*	X	cell surface; extracellular region; glycolysis; magnesium ion binding; phosphopyruvate hydratase activity; phosphopyruvate hydratase complex
Q5LMY9	X*	X*	X*	X	cell redox homeostasis; electron carrier activity; glycerol ether metabolic process; protein disulfide

					oxidoreductase activity
Q5LLP8	X*	X*	X	X	cell redox homeostasis; electron carrier activity; glycerol ether metabolic process; protein disulfide oxidoreductase activity
Q5LXD1	X*	X*	X*	X	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
Q5LMI7	X*	X*	X*	X*	cell outer membrane; integral to membrane; plasma membrane
Q5LRJ4	X*	X*	X*	X*	cell outer membrane; integral to membrane
Q5LWC2	X*	X	X	X	cell morphogenesis
Q5LR43	X*	X*	X*	X*	cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization; protein transport
Q5LP64	X*	X*	X*	X	catalytic activity; response to tellurium ion
Q5LTW7	X*	X*	-	-	catalytic activity; regulation of biosynthetic process
Q5LTQ9	X	X*	X*	X*	catalytic activity; metabolic process
Q5LNB6	X*	X	X	X	catalytic activity
Q5LLG6	X*	X	X	X	catalase activity; heme binding; hydrogen peroxide catabolic process; metal ion binding
Q5LMF5	X	X	X*	X	carboxylic acid metabolic process; cellular aromatic compound metabolic process; citrate (pro-3S)-lyase activity; citrate lyase complex; malyl-CoA lyase activity; metal ion binding
Q5LPB0	X	X	X*	X	carbon-nitrogen ligase activity, with glutamine as amido-N-donor
Q5LMZ4	X	X	X	X*	carbon-nitrogen ligase activity, with glutamine as amido-N-donor
Q5LT95	X*	X*	X*	X*	carbon-monoxide dehydrogenase (acceptor) activity
Q5LRI0	X	X*	X	X	carbohydrate metabolic process; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
Q5LTA6	X*	X	X*	X	calcium ion binding; membrane; outer membrane-bounded periplasmic space; oxidoreductase activity, acting on CH-OH group of donors
Q5LRJ2	X*	X*	X*	X*	calcium ion binding
Q5LX13	X*	X*	X*	X	calcium ion binding
Q5LSJ5	X	X	X*	X*	C4-dicarboxylate transport; outer membrane-bounded periplasmic space
Q5LKP6	X	X	X	X*	C4-dicarboxylate transmembrane transporter activity; outer membrane-bounded periplasmic space
Q5LSC2	X	X*	X	X	branched-chain amino acid transmembrane transporter activity
Q5LR60	X*	X*	X*	X*	beta-ketoacyl-acyl-carrier-protein synthase II activity; fatty acid biosynthetic process
Q5LMV3	-	X*	X*	X*	bacterial-type flagellum; ciliary or bacterial-type flagellar motility; structural molecule activity
Q5LSI0	X*	X*	X*	X*	ATPase activity, coupled to transmembrane movement of substances; outer membrane-bounded periplasmic space
Q5LSH7	X	X*	X*	X*	ATPase activity, coupled to transmembrane movement of substances; anion transport; outer membrane-bounded periplasmic space
Q5LNC1	X*	X*	X*	X*	ATPase activity, coupled to transmembrane movement of substances; amino acid binding; branched-chain amino acid transmembrane transporter activity
Q5LM59	X	X	X*	X	ATPase activity, coupled to transmembrane movement of substances; amino acid binding; branched-

					chain amino acid transmembrane transporter activity
Q5LLF8	X	X*	X*	X*	ATPase activity, coupled to transmembrane movement of substances; amino acid binding; amino acid transport
Q5LPT4	X*	X*	X	X*	ATP synthesis coupled electron transport; NADH dehydrogenase (ubiquinone) activity; integral to membrane
Q5LPS5	X	X*	X	X*	ATP synthesis coupled electron transport; NADH dehydrogenase (ubiquinone) activity; electron carrier activity; iron-sulfur cluster binding; membrane
Q5LNH0	X*	X*	X*	X*	ATP hydrolysis coupled proton transport; ATP synthesis coupled proton transport; hydrogen ion transmembrane transporter activity; integral to membrane; lipid binding; plasma membrane; proton-transporting ATP synthase complex, coupling factor F(o)
Q5LLV3	X	X*	X	X	ATP binding; tetrahydrofolylpolyglutamate biosynthetic process; tetrahydrofolylpolyglutamate synthase activity
Q5LLL2	X*	X*	X	X	ATP binding; S-adenosylmethionine biosynthetic process; cytoplasm; magnesium ion binding; methionine adenosyltransferase activity; one-carbon metabolic process
Q5LQK5	X	X*	X	X	ATP binding; RNA binding; cytoplasm; tyrosine-tRNA ligase activity; tyrosyl-tRNA aminoacylation
Q5LWJ6	X*	X	X*	X	ATP binding; protein folding; response to stress
Q5LNP0	X*	X*	X	X	ATP binding; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1); proton-transporting ATPase activity, rotational mechanism
Q5LNP2	X*	X*	X	X	ATP binding; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1); proton-transporting ATPase activity, rotational mechanism
Q5LUP9	X*	X	X	X*	ATP binding; nucleoside-triphosphatase activity; protein folding; zinc ion binding
Q5LXC4	X*	X*	X*	X	ATP binding; magnesium ion binding; manganese ion binding; succinate-CoA ligase (ADP-forming) activity; tricarboxylic acid cycle
Q5LS76	X*	X	X	X	ATP binding; intracellular signal transduction; nucleoside-triphosphatase activity; phosphorelay response regulator activity; regulation of transcription, DNA-dependent; sequence-specific DNA binding; transcription, DNA-dependent
Q5LTJ7	X*	X*	X*	X*	ATP binding; helicase activity; nucleic acid binding
Q5LP80	X*	X*	X*	X	ATP binding; glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity; translation
Q5LV38	X*	X*	X*	X*	ATP binding; glucokinase activity; glycolysis
Q5LLM8	X*	X*	X	X*	ATP binding; DNA-dependent transcription, termination; RNA binding; RNA-dependent ATPase activity; helicase activity; regulation of transcription, DNA-dependent
Q5LRU0	X	X*	X*	X*	ATP binding; DNA recombination; DNA repair; DNA-dependent ATPase activity; SOS response; cytoplasm; damaged DNA binding; single-stranded DNA binding
Q5LWU8	X	X	X	X*	ATP binding; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological

					change; DNA-dependent DNA replication; chromosome; cytoplasm; magnesium ion binding
Q5LNX4	X	X	X	X*	ATP binding; DNA binding; DNA topoisomerase type I activity; DNA topological change; chromosome; metal ion binding
Q5LML3	X	X	X	X*	ATP binding; DNA binding; DNA polymerase III complex; DNA replication; DNA-directed DNA polymerase activity; nucleoside-triphosphatase activity
Q5LNT7	X	X*	-	-	ATP binding; DNA binding; DNA duplex unwinding; DNA recombination; DNA repair; Holliday junction helicase complex; SOS response; four-way junction helicase activity
Q5LU87	X*	X*	X*	X*	ATP binding; DNA binding; biotin carboxylase activity; gluconeogenesis; metal ion binding; pyruvate carboxylase activity
Q5LUQ0	X	X*	X	X	ATP binding; cytoplasm; proteolysis; serine-type endopeptidase activity
Q5LV15	X*	X*	X*	X*	ATP binding; cytoplasm; protein refolding
Q5LV16	X*	X	X	X	ATP binding; cytoplasm; protein folding
Q5LUD3	X	X	X*	X	ATP binding; cytoplasm; proline-tRNA ligase activity; prolyl-tRNA aminoacylation
Q5LTX8	X	X*	X*	X	ATP binding; cytoplasm; metal ion binding; threonine-tRNA ligase activity; threonyl-tRNA aminoacylation
Q5LTL1	X	X	X*	X*	ATP binding; cytoplasm; metal ion binding; methionine-tRNA ligase activity; methionyl-tRNA aminoacylation
Q5LMS3	X*	X*	X*	X*	ATP binding; cytoplasm; magnesium ion binding; phenylalanine-tRNA ligase activity; phenylalanyl-tRNA aminoacylation; tRNA binding; tRNA processing
Q5LW87	X	X	X	X*	ATP binding; cytoplasm; lysine-tRNA ligase activity; lysyl-tRNA aminoacylation; tRNA binding
Q5LVN3	X	X	X	X*	ATP binding; cytoplasm; histidine-tRNA ligase activity; histidyl-tRNA aminoacylation
Q5LR94	X	X*	X	X*	ATP binding; cytoplasm; glycolysis; phosphoglycerate kinase activity
Q5LRH2	X*	X	X	X*	ATP binding; cytoplasm; glutamate-tRNA ligase activity; glutamyl-tRNA aminoacylation; tRNA binding
Q5LVJ2	X*	X	X	X	ATP binding; cytoplasm; gluconeogenesis; phosphoenolpyruvate carboxykinase (ATP) activity
Q5LQP4	X	X*	X*	X*	ATP binding; CTP biosynthetic process; GTP biosynthetic process; UTP biosynthetic process; cytoplasm; metal ion binding; nucleoside diphosphate kinase activity
Q5LX46	X	X	X*	X	ATP binding; cob(I)yrinic acid a,c-diamide adenosyltransferase activity; cobalamin biosynthetic process
Q5LPK1	X*	X*	X*	X*	ATP binding; carbon-nitrogen ligase activity, with glutamine as amido-N-donor; translation
Q5LUF3	X	X*	X	X	ATP binding; biotin carboxylase activity; metal ion binding; propionyl-CoA carboxylase activity
Q5LPR0	X*	X	X	X	ATP binding; biotin carboxylase activity; metal ion binding; methylcrotonoyl-CoA carboxylase activity
Q5LLT8	X*	X	X	X	ATP binding; ATP-dependent helicase activity; nucleic acid binding
Q5LTH1	X	X	X*	X	ATP binding; ATP-dependent helicase activity; nucleic acid binding
Q5LVT6	X*	X	X	-	ATP binding; ATP-binding cassette (ABC) transporter complex; ATPase activity, coupled to transmembrane movement of substances; carbohydrate transport
Q5LNU8	X	X*	X	X	ATP binding; ATPase activity; cytokinesis; integral to membrane; metalloendopeptidase activity; plasma

					membrane; protein catabolic process; proteolysis; zinc ion binding
Q5LNN9	X*	X*	X*	X*	ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1); proton-transporting ATPase activity, rotational mechanism
Q5LNP1	X*	X*	X*	X*	ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1)
Q5LXC5	X*	X*	X*	X*	ATP binding; ATP citrate synthase activity; cofactor binding; succinate-CoA ligase (ADP-forming) activity
Q5LQ83	X*	X	X	-	ATP binding; ATP catabolic process; ATP-dependent peptidase activity; cellular response to stress; cytoplasm; misfolded or incompletely synthesized protein catabolic process; sequence-specific DNA binding; serine-type endopeptidase activity
Q5LMU0	X	X*	X*	X*	ATP binding; ATP catabolic process; ATP-binding cassette (ABC) transporter complex; plasma membrane; polyamine-transporting ATPase activity
Q5LRB1	X*	X	X	-	ATP binding; ATP catabolic process; ATP-binding cassette (ABC) transporter complex; ATPase activity; DNA binding; SOS response; cytoplasm; excinuclease ABC activity; excinuclease repair complex; nucleic acid phosphodiester bond hydrolysis; nucleotide-excision repair; plasma membrane; transport; zinc ion binding
Q5LP06	X	X	X	X*	ATP binding; ATP catabolic process; ATPase activity; membrane; peptide transporter activity
Q5LWU6	X	X*	X*	X	ATP binding; ATP catabolic process; ATPase activity; catalytic activity
Q5LLV6	-	-	X*	-	ATP binding; ATP catabolic process; ATPase activity, coupled to transmembrane movement of substances; integral to membrane
Q5LV39	X*	X*	X*	X	ATP binding; ATP catabolic process; ATPase activity
Q5LS33	X*	X	X	X	ATP binding; ATP catabolic process; ATPase activity
Q5LW21	X	X*	X*	X	ATP binding; ATP catabolic process; ATPase activity
Q5LW96	X	X	X*	X*	ATP binding; ATP catabolic process; ATPase activity
Q5LUX6	X*	X	X	X*	ATP binding; aspartate-tRNA ligase activity; aspartyl-tRNA aminoacylation; cytoplasm; nucleic acid binding
Q5LTQ5	X*	X	X*	X*	ATP binding; arginine-tRNA ligase activity; arginyl-tRNA aminoacylation; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
Q5LWG3	X	X*	X	X	ATP binding; arginine biosynthetic process; argininosuccinate synthase activity; cytoplasm
Q5LP30	X*	X*	X*	X	ATP binding; aminoacyl-tRNA editing activity; cytoplasm; regulation of translational fidelity; valine-tRNA ligase activity; valyl-tRNA aminoacylation
Q5LNR7	X*	X*	X	X	ATP binding; aminoacyl-tRNA editing activity; cytoplasm; isoleucine-tRNA ligase activity; isoleucyl-tRNA aminoacylation; regulation of translational fidelity; zinc ion binding

Q5LV02	X*	X*	X*	X*	ATP binding; adenylylsulfate kinase activity; sulfate adenylyltransferase (ATP) activity; sulfate assimilation
Q5LME7	X*	X	X	X*	ATP binding; acetyl-CoA carboxylase activity; acetyl-CoA carboxylase complex; fatty acid biosynthetic process; malonyl-CoA biosynthetic process
Q5LXC0	X	X*	X*	X*	arginine biosynthetic process via ornithine; argininosuccinate lyase activity; cytoplasm
Q5LN29	X*	X*	X*	X	antioxidant activity; peroxiredoxin activity; response to oxidative stress
Q5LRI2	X*	X	X	X	anthranilate synthase activity; tryptophan biosynthetic process
Q5LM19	X*	X	X	X	AMP nucleosidase activity; AMP salvage
Q5LSF5	X*	X*	X*	X*	AMP binding; ATP binding; acetate-CoA ligase activity; acetyl-CoA biosynthetic process from acetate; metal ion binding
Q5LQN5	X	X*	X*	X*	aminopeptidase activity; cytoplasm; manganese ion binding; metalloexopeptidase activity; proteolysis
Q5LVN2	X	X	X	X*	aminoacyl-tRNA ligase activity; cytoplasm; tRNA aminoacylation for protein translation
Q5LV77	X*	X	X	X	amino acid transport; outer membrane-bounded periplasmic space
Q5LUN2	X*	X*	X*	X*	amino acid binding; branched-chain amino acid transmembrane transporter activity; outer membrane-bounded periplasmic space
Q5LP17	X	X	X*	X*	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate
Q5LUT9	X*	X*	X	X	amino acid binding; arginine biosynthetic process; cytoplasm; ornithine carbamoyltransferase activity
Q5LS31	X*	X*	X*	X*	amino acid binding
Q5LV45	X	X*	X	X*	aldose 1-epimerase activity; carbohydrate binding; hexose metabolic process
Q5LQW3	X*	X	X	X	alanine:sodium symporter activity; membrane
Q5LP93	X	X	X	X*	adenyl nucleotide binding; magnesium ion transmembrane transporter activity; membrane
Q5LLR0	X*	X	X	X*	adenosylhomocysteinase activity; cytoplasm; one-carbon metabolic process
Q5LM18	X*	X	X*	X	adenine catabolic process; adenine deaminase activity
Q5LS02	X*	-	-	-	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
Q5LS03	X*	X	X	X	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
Q5LSU3	X*	X*	X	X	acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase activity; cytoplasm; lipid A biosynthetic process
Q5LWL6	X*	X*	X*	X*	acetyl-CoA:L-glutamate N-acetyltransferase activity; arginine biosynthetic process; cytoplasm; glutamate N-acetyltransferase activity
Q5LXA8	X*	X*	X*	X*	acetyl-CoA C-acetyltransferase activity
Q5LR88	X*	X*	X*	X*	acetyl-CoA biosynthetic process from pyruvate; pyruvate dehydrogenase (acetyl-transferring) activity
Q5LRA7	X	X*	X	X*	acetyl-CoA biosynthetic process from pyruvate; cell redox homeostasis; cytosolic pyruvate dehydrogenase complex; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
Q3V7J4	X	X	X	X*	acetolactate synthase activity; branched-chain amino acid biosynthetic process; flavin adenine dinucleotide binding; magnesium ion binding; thiamine pyrophosphate binding
Q5LXA9	X	X*	X	X	acetoacetyl-CoA reductase activity; cytoplasm; poly-hydroxybutyrate biosynthetic process

Q5LRS8	X	X*	X*	X	6-phosphogluconolactonase activity; pentose-phosphate shunt
Q5LV90	X*	X*	X*	X*	5S rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LNP4	X	X	X*	X*	5-phosphoribose 1-diphosphate biosynthetic process; ATP binding; cytoplasm; kinase activity; magnesium ion binding; nucleotide biosynthetic process; ribonucleoside monophosphate biosynthetic process; ribose phosphate diphosphokinase activity
Q5LPS2	X*	X*	X*	X*	4 iron, 4 sulfur cluster binding; FMN binding; NAD binding; NADH dehydrogenase (ubiquinone) activity; metal ion binding
Q5LTJ5	X*	X	X	X	4 iron, 4 sulfur cluster binding; electron carrier activity; metal ion binding
Q5LN98	X	X	X*	X	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
Q5LRM3	X*	-	-	-	4 iron, 4 sulfur cluster binding; cytoplasm; lipoate synthase activity; metal ion binding; protein lipoylation
Q5LR20	X*	X*	X	X*	4 iron, 4 sulfur cluster binding; citrate hydro-lyase (cis-aconitate-forming) activity; isocitrate hydro-lyase (cis-aconitate-forming) activity; metal ion binding
Q5LSH2	X*	X	X	X	4 iron, 4 sulfur cluster binding; anaerobic respiration; electron carrier activity; formate dehydrogenase (NAD+) activity; formate dehydrogenase complex; molybdenum ion binding
Q5LQ99	X*	X	X	-	4 iron, 4 sulfur cluster binding; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity; iron ion binding; isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway; terpenoid biosynthetic process
Q5LLK7	X	X*	X	X	3-phosphoshikimate 1-carboxyvinyltransferase activity; aromatic amino acid family biosynthetic process; chorismate biosynthetic process; cytoplasm
Q5LKE7	X	X	X*	X*	3-oxoadipate enol-lactonase activity; beta-ketoadipate pathway; cellular aromatic compound metabolic process
Q5LNX1	X*	X*	X*	X	3-oxoadipate CoA-transferase activity; cellular aromatic compound metabolic process
Q5LNX0	X*	X	X	X	3-oxoadipate CoA-transferase activity; cellular aromatic compound metabolic process
Q5LKE1	X*	X	X*	X	3-oxoacyl-[acyl-carrier-protein] synthase activity
Q5LR56	X*	X	X*	X*	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; NAD binding; fatty acid biosynthetic process
Q5LWZ5	X*	X*	X	X*	3-isopropylmalate dehydrogenase activity; NAD binding; cytoplasm; leucine biosynthetic process; magnesium ion binding
Q5LX06	X*	X	X*	X	3-isopropylmalate dehydratase activity; 4 iron, 4 sulfur cluster binding; leucine biosynthetic process; metal ion binding
Q5LX07	X	X*	X*	X*	3-isopropylmalate dehydratase activity; 3-isopropylmalate dehydratase complex; leucine biosynthetic process
Q5LVB0	X*	X*	X	X	3-hydroxyisobutyrate dehydrogenase activity; coenzyme binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity; valine metabolic process

Q5LRR0	X*	X*	X	X	3-hydroxybutyrate dehydrogenase activity
Q5LPC8	X	X*	X	X	3-hydroxyacyl-CoA dehydrogenase activity; coenzyme binding; fatty acid metabolic process
Q5LQK2	X*	X	X	X*	3-hydroxyacyl-CoA dehydrogenase activity
Q5LS28	X	X*	X	X	3-deoxy-7-phosphoheptulonate synthase activity; aromatic amino acid family biosynthetic process
Q5LKE9	X	X	X*	X	3-carboxy-cis,cis-muconate cycloisomerase activity; metabolic process
Q5LN23	X*	X*	X*	X*	3'-5'-exoribonuclease activity; RNA binding; RNA processing; cytoplasm; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity
Q5LWV3	X	X	X	X*	3'-5' exonuclease activity; DNA binding; DNA polymerase III complex; DNA replication; DNA-directed DNA polymerase activity; cytoplasm
Q5LLS7	X	X	X*	X	3'-5' exonuclease activity; 5'-3' exonuclease activity; DNA binding; DNA replication; DNA-directed DNA polymerase activity
Q5LNJ7	X*	X*	X	X	3 iron, 4 sulfur cluster binding; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase activity; dimethylallyl diphosphate biosynthetic process; isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway; metal ion binding; terpenoid biosynthetic process
Q5LP21	X*	X*	X*	X*	2-dehydro-3-deoxy-phosphogluconate aldolase activity; 4-hydroxy-2-oxoglutarate aldolase activity
Q5LX56	X	X*	X	X	2 iron, 2 sulfur cluster binding; membrane; metal ion binding; ubiquinol-cytochrome-c reductase activity
Q5LLL7	X*	X	X	X	2 iron, 2 sulfur cluster binding; Gram-negative-bacterium-type cell wall; electron carrier activity; integral to membrane; metal ion binding; oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor; plasma membrane; sodium ion transport
Q5LKG8	X	X	X	X*	2 iron, 2 sulfur cluster binding; electron carrier activity; metal ion binding; oxidoreductase activity
Q5LT96	X	X*	X*	X*	2 iron, 2 sulfur cluster binding; carbon-monoxide dehydrogenase (acceptor) activity; electron carrier activity; metal ion binding
Q5LLR2	X*	X*	X*	X*	No definititon
Q5LLR9	X*	X*	X*	X*	No definititon
Q5LMT9	X*	X*	X*	X*	No definititon
Q5LPK0	X*	X*	X*	X*	No definititon
Q5LPV3	X*	X*	X*	X*	No definititon
Q5LRE3	X*	X*	X*	X*	No definititon
Q5LRP8	X*	X*	X*	X*	No definititon
Q5LXE6	X*	X*	X*	X*	No definititon
Q5LM34	X*	X*	X*	X	No definititon
Q5LN64	X*	X*	X*	X	No definititon
Q5LWE4	X*	X*	X*	X	No definititon
Q5LMJ7	X*	X*	X	X	No definititon
Q5LPW1	X*	X*	X	X	No definititon
Q5LS37	X*	X*	X	X	No definititon

Q5LR03	X*	X	X*	X	No definititon
Q5LX99	X*	X	X	X*	No definititon
Q5LKS5	X*	X	X	X	No definititon
Q5LLQ0	X*	-	-	-	No definititon
Q5LMY1	X*	X	X	X	No definititon
Q5LPY0	X*	X	X	X	No definititon
Q5LQ60	X*	-	-	-	No definititon
Q5LR02	X*	X	X	X	No definititon
Q5LSB3	X*	X	X	X	No definititon
Q5LSH0	X*	X	X	-	No definititon
Q5LTW6	X*	X	X	X	No definititon
Q5LTW9	X*	-	-	-	No definititon
Q5LUU9	X*	X	X	X	No definititon
Q5LRB9	X	X*	X*	X*	No definititon
Q5LTH6	-	X*	X*	X*	No definititon
Q5LX94	X	X*	X*	X*	No definititon
Q5LLZ2	X	X*	X*	X	No definititon
Q5LMX1	X	X*	X*	X	No definititon
Q5LP81	X	X*	X*	X	No definititon
Q5LS99	X	X*	X*	X	No definititon
Q5LXB2	X	X*	X*	X	No definititon
Q5LME6	X	X*	X	X*	No definititon
Q5LQM9	X	X*	X	X*	No definititon
Q5LRU2	X	X*	X	X*	No definititon
Q5LLS0	X	X*	X	X	No definititon
Q5LRG2	X	X*	X	X	No definititon
Q5LRU8	X	X*	X	X	No definititon
Q5LTY8	X	X*	X	X	No definititon
Q5LUS1	-	X*	-	-	No definititon
Q5LUS7	X	X*	-	-	No definititon
Q5LW54	-	X*	-	-	No definititon
Q5LME3	X	X	X*	X*	No definititon
Q5LPX9	X	X	X*	X	No definititon
Q5LS60	-	X	X*	X	No definititon
Q5LVQ5	X	X	X*	X	No definititon

Q5LP97	X	X	X	X*	No definititon
Q5LS82	X	X	X	X*	No definititon

Table A5.2. Proteins that were up- or down-regulated within the investigation. The latter time point is what is being referred to as up- or down-regulated.*

Protein (UniProt ID)	T_1vT_2	T_1vT_3	T_1vT_4	Gene Ontology
Q5LQP4	Up	Up	Up	ATP binding; CTP biosynthetic process; GTP biosynthetic process; UTP biosynthetic process; cytoplasm; metal ion binding; nucleoside diphosphate kinase activity
Q5LSF5	Up	Up	Up	AMP binding; ATP binding; acetate-CoA ligase activity; acetyl-CoA biosynthetic process from acetate; metal ion binding
Q5LXC0	Up	Up	Up	arginine biosynthetic process via ornithine; argininosuccinate lyase activity; cytoplasm
Q5LNP1	Up	Up	Up	ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1)
Q5LNN9	Up	Up	Up	ATP binding; ATP hydrolysis coupled proton transport; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1); proton-transporting ATPase activity, rotational mechanism
Q5LV02	Up	Up	Up	ATP binding; adenylylsulfate kinase activity; sulfate adenylyltransferase (ATP) activity; sulfate assimilation
Q5LMY2	Up	Up	Up	porin activity
Q5LXA8	Up	Up	Up	acetyl-CoA C-acetyltransferase activity
Q5LWY4	Up	Up	Up	outer membrane-bounded periplasmic space
Q5LN30	Up	Up	Up	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LWL2	Up	Up	Up	RNA binding; nucleotide binding; regulation of DNA-dependent transcription, termination; sequence-specific DNA binding transcription factor activity
Q5LU48	Up	Up	Up	transporter activity
Q5LSV2	Up	Up	Up	cytoplasm; translational termination
Q5LRB9	Up	Up	Up	No definition
Q5LUS9	Up	Up	Up	ribosome; structural constituent of ribosome; translation
Q5LSI0	Up	Up	Up	ATPase activity, coupled to transmembrane movement of substances; outer membrane-bounded periplasmic space
Q5LT71	Up	Up	Up	transporter activity
Q5LRU0	Up	Up	Up	ATP binding; DNA recombination; DNA repair; DNA-dependent ATPase activity; SOS response; cytoplasm; damaged DNA binding; single-stranded DNA binding

Q5LMR5	Up	Up	Up	GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation elongation factor activity
Q5LU19	Up	Up	Up	outer membrane; protein transport; transporter activity
Q5LNQ6	Up	Up	Up	cytoplasm; inorganic diphosphatase activity; manganese ion binding
Q5LRP8	Up	Up	Up	No definition
Q5LQN5	Up	Up	Up	aminopeptidase activity; cytoplasm; manganese ion binding; metalloexopeptidase activity; proteolysis
Q5LRS9	Up	Up	Up	cytoplasm; gluconeogenesis; glucose-6-phosphate isomerase activity; glycolysis
Q5LSC9	Up	Up	Up	transporter activity
Q5LX07	Up	Up	Up	3-isopropylmalate dehydratase activity; 3-isopropylmalate dehydratase complex; leucine biosynthetic process
Q5LMU0	Up	Up	Up	ATP binding; ATP catabolic process; ATP-binding cassette (ABC) transporter complex; plasma membrane; polyamine-transporting ATPase activity
Q5LWP6	Up	Up	Up	transporter activity
Q5LKZ1	Up	Up	Up	outer membrane-bounded periplasmic space; transport
Q5LWU3	Up	Up	Up	energy reserve metabolic process; transferase activity, transferring acyl groups other than amino-acyl groups
Q5LL21	Up	Up	Up	UDP-N-acetylmuramate dehydrogenase activity; flavin adenine dinucleotide binding
Q5LRT0	Up	Up	Up	ligase activity
Q5LR83	Up	Up	Up	cysteine biosynthetic process from serine; cysteine synthase activity; transferase activity
Q5LXC7	Up	Up	Up	oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding; tricarboxylic acid cycle
Q5LMV3	Up	Up	Up	bacterial-type flagellum; ciliary or bacterial-type flagellar motility; structural molecule activity
Q5LV46	Up	Up	Up	D-xylose metabolic process; cytoplasm; magnesium ion binding; pentose-phosphate shunt; xylose isomerase activity
Q5LQQ3	Up	Up	Up	hydrolase activity
Q5LPN4	Up	Up	Up	transporter activity
Q5LWN3	Up	Up	Up	primary metabolic process
Q5LKE3	Up	Up	Up	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LSS5	Up	Up	Up	ribosome; structural constituent of ribosome; translation
Q5LSH7	Up	Up	Up	ATPase activity, coupled to transmembrane movement of substances; anion transport; outer membrane-bounded periplasmic space

Q5LT95	Up	Up	Up	carbon-monoxide dehydrogenase (acceptor) activity
Q5LUN2	Up	Up	Up	amino acid binding; branched-chain amino acid transmembrane transporter activity; outer membrane-bounded periplasmic space
Q5LRE3	Up	Up	Up	No definition
Q5LVI2	Up	Up	-	electron carrier activity; flavin adenine dinucleotide binding
Q5LMQ5	Up	Up	-	DNA binding; DNA-directed RNA polymerase activity; ribonucleoside binding; transcription, DNA-dependent
Q5LMD9	Up	Up	-	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LXE1	Up	Up	Down	L-malate dehydrogenase activity; cellular carbohydrate metabolic process; malate metabolic process; tricarboxylic acid cycle
Q5LRJ2	Up	Up	Down	calcium ion binding
Q5LWU6	Up	Up	-	ATP binding; ATP catabolic process; ATPase activity; catalytic activity
Q5LWL4	Up	Up	-	GTP binding; GTP catabolic process; GTPase activity; cytoplasm; translation initiation factor activity
Q5LW21	Up	Up	-	ATP binding; ATP catabolic process; ATPase activity
Q5LS99	Up	Up	-	No definition
Q5LXB2	Up	Up	-	No definition
Q5LNU4	Up	Up	-	periplasmic space; protein import
Q5LWA5	Up	Up	-	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LLZ2	Up	Up	-	No definition
Q5LR43	Up	Up	-	cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization; protein transport
Q5LTX8	Up	Up	-	ATP binding; cytoplasm; metal ion binding; threonine-tRNA ligase activity; threonyl-tRNA aminoacylation
Q5LXC8	Up	-	Up	dihydrolipoyllysine-residue succinyltransferase activity; oxoglutarate dehydrogenase complex; tricarboxylic acid cycle
Q5LU87	Up	-	Up	ATP binding; DNA binding; biotin carboxylase activity; gluconeogenesis; metal ion binding; pyruvate carboxylase activity
Q5LUC3	Up	-	Up	transporter activity
Q5LR20	Up	Down	Up	4 iron, 4 sulfur cluster binding; citrate hydro-lyase (cis-aconitate-forming) activity; isocitrate hydro-lyase (cis-aconitate-forming) activity; metal ion binding
Q5LNW6	Up	-	Up	protein secretion

Q5LR89	Up	-	Up	glycolysis; intracellular membrane-bounded organelle; pyruvate dehydrogenase (acetyl-transferring) activity
Q5LNC1	Up	-	Up	ATPase activity, coupled to transmembrane movement of substances; amino acid binding; branched-chain amino acid transmembrane transporter activity
Q5LP55	Up	Down	Up	transporter activity
Q5LXG2	Up	-	Up	plasma membrane
Q5LPS5	Up	-	Up	ATP synthesis coupled electron transport; NADH dehydrogenase (ubiquinone) activity; electron carrier activity; iron-sulfur cluster binding; membrane
Q5LVT7	Up	-	Up	transporter activity
Q5LW32	Up	-	Up	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-dependent
Q5LR94	Up	-	Up	ATP binding; cytoplasm; glycolysis; phosphoglycerate kinase activity
Q5LME6	Up	-	Up	No definition
Q5LX26	Up	-	Up	oxidoreductase activity; zinc ion binding
Q5LRA7	Up	-	Up	acetyl-CoA biosynthetic process from pyruvate; cell redox homeostasis; cytosolic pyruvate dehydrogenase complex; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
Q5LLF8	Up	-	Up	ATPase activity, coupled to transmembrane movement of substances; amino acid binding; amino acid transport
Q5LP16	Up	Down	Up	hydrolase activity; sterol binding
Q5LW96	-	Up	Up	ATP binding; ATP catabolic process; ATPase activity
Q5LPK0	-	Up	Up	No definition
Q5LN57	Down	Up	Up	L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
Q5LS98	-	Up	Up	'de novo' IMP biosynthetic process; ATP binding; cytoplasm; phosphoribosylformylglycinamidine synthase activity
Q5LMT9	-	Up	Up	No definition
Q5LRI3	-	Up	Up	isomerase activity
Q5LNP4	-	Up	Up	5-phosphoribose 1-diphosphate biosynthetic process; ATP binding; cytoplasm; kinase activity; magnesium ion binding; nucleotide biosynthetic process; ribonucleoside monophosphate biosynthetic process; ribose phosphate diphosphokinase activity
Q5LNS5	-	Up	Up	cytoplasm; pentose-phosphate shunt; sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glyceronetransferase activity
Q5LT08	-	Up	Up	periplasmic space; polyamine binding; polyamine transport
Q5LTQ9	-	Up	Up	catalytic activity; metabolic process

Q5LQ89	-	Up	Up	outer membrane-bounded periplasmic space; transport
Q5LVJ5	-	Up	Up	transporter activity
Q5LVG7	-	Up	Up	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LMQ6	-	Up	Up	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-dependent
Q5LME3	-	Up	Up	No definition
Q5LTL1	-	Up	Up	ATP binding; cytoplasm; metal ion binding; methionine-tRNA ligase activity; methionyl-tRNA aminoacylation
Q5LV38	-	Up	Up	ATP binding; glucokinase activity; glycolysis
Q5LQ51	-	Up	Up	proteolysis; serine-type endopeptidase activity
Q5LKE7	-	Up	Up	3-oxoadipate enol-lactonase activity; beta-ketoadipate pathway; cellular aromatic compound metabolic process
Q5LX94	-	Up	Up	No definition
Q5LL90	-	Up	Up	UDP-N-acetylgalactosamine biosynthetic process; UDP-N-acetylglucosamine 1-carboxyvinyltransferase activity; cell cycle; cell division; cytoplasm; peptidoglycan biosynthetic process; regulation of cell shape
Q5LPJ1	-	Up	Up	cellular process; hydrolase activity
Q5LM44	-	Up	Up	L-phenylalanine:2-oxoglutarate aminotransferase activity; aromatic-amino-acid:2-oxoglutarate aminotransferase activity; biosynthetic process; cellular amino acid metabolic process; pyridoxal phosphate binding
Q5LPM8	Down	Up	Up	NAD(P)+ transhydrogenase (AB-specific) activity; nucleotide binding; proton transport
Q5LLU2	-	Up	Up	S-formylglutathione hydrolase activity; carboxylesterase activity; formaldehyde catabolic process
Q5LSJ5	-	Up	Up	C4-dicarboxylate transport; outer membrane-bounded periplasmic space
Q5LTY8	Up	-	-	No definition
Q5LMP7	Up	-	-	DNA-dependent transcription, termination; positive regulation of transcription elongation from RNA polymerase II promoter; transcription antitermination
Q5LW42	Up	-	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LUZ9	Up	-	-	cytoplasm; flavin adenine dinucleotide binding; removal of superoxide radicals; thioredoxin-disulfide reductase activity
Q5LTJ9	Up	-	-	sterol binding
Q5LNU8	Up	-	-	ATP binding; ATPase activity; cytokinesis; integral to membrane; metalloendopeptidase activity; plasma membrane; protein catabolic process; proteolysis; zinc ion binding
Q5LKV7	Up	-	-	S-(hydroxymethyl)glutathione dehydrogenase activity; ethanol oxidation; zinc ion binding

Q5LR96	Up	-	-	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LPM9	Up	-	-	NAD(P)+ transhydrogenase (AB-specific) activity; NADP binding; integral to membrane; plasma membrane
Q5LSS4	Up	-	-	ribosome; structural constituent of ribosome; translation
Q5LS81	Up	-	-	'de novo' IMP biosynthetic process; ATP binding; phosphoribosylaminoimidazolesuccinocarboxamide synthase activity
Q5LSC2	Up	-	-	branched-chain amino acid transmembrane transporter activity
Q5LXA9	Up	-	-	acetoacetyl-CoA reductase activity; cytoplasm; poly-hydroxybutyrate biosynthetic process
Q5LVG4	Up	-	-	outer membrane-bounded periplasmic space; transport
Q5LMP9	Up	-	Down	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LX56	Up	-	-	2 iron, 2 sulfur cluster binding; membrane; metal ion binding; ubiquinol-cytochrome-c reductase activity
Q5LP81	Up	-	-	No definition
Q5LWZ5	Up	Down	-	3-isopropylmalate dehydrogenase activity; NAD binding; cytoplasm; leucine biosynthetic process; magnesium ion binding
Q5LW45	Up	-	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LUQ0	Up	-	-	ATP binding; cytoplasm; proteolysis; serine-type endopeptidase activity
Q5LWG3	Up	-	-	ATP binding; arginine biosynthetic process; argininosuccinate synthase activity; cytoplasm
Q5LQ42	Up	-	-	'de novo' UMP biosynthetic process; magnesium ion binding; orotate phosphoribosyltransferase activity
Q5LLK7	Up	-	-	3-phosphoshikimate 1-carboxyvinyltransferase activity; aromatic amino acid family biosynthetic process; chorismate biosynthetic process; cytoplasm
Q5LNM2	Up	-	-	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LQK5	Up	-	-	ATP binding; RNA binding; cytoplasm; tyrosine-tRNA ligase activity; tyrosyl-tRNA aminoacylation
Q5LWE4	Up	-	Down	No definition
Q5LW58	Up	-	-	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LS28	Up	-	-	3-deoxy-7-phosphoheptulonate synthase activity; aromatic amino acid family biosynthetic process
Q5LMG5	Up	-	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LLN4	Up	-	-	cytoplasm; protein folding; protein tetramerization; protein transport
Q5LMI1	Up	-	-	kinase activity

Q5LRS7	Up	-	-	NADP binding; glucose-6-phosphate dehydrogenase activity; pentose-phosphate shunt
Q5LMJ7	Up	-	-	No definition
Q5LPC8	Up	-	-	3-hydroxyacyl-CoA dehydrogenase activity; coenzyme binding; fatty acid metabolic process
Q5LM52	Up	-	-	'de novo' L-methionine biosynthetic process; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; aspartate-semialdehyde dehydrogenase activity; cytoplasm; diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process
Q5LPE4	Up	-	-	hydrolase activity, acting on ester bonds; metal ion binding; nucleotide binding; nucleotide catabolic process
Q5LN98	-	Up	-	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
Q5LRS8	-	Up	-	6-phosphogluconolactonase activity; pentose-phosphate shunt
Q5LKE2	-	Up	-	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process
Q5LMX1	-	Up	-	No definition
Q5LXH0	-	Up	-	cytoplasm; iron ion binding; iron-sulfur cluster assembly; iron-sulfur cluster binding
Q5LLT1	-	Up	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LR88	-	Up	-	acetyl-CoA biosynthetic process from pyruvate; pyruvate dehydrogenase (acetyl-transferring) activity
Q5LPQ5	-	Up	-	flavin adenine dinucleotide binding; isovaleryl-CoA dehydrogenase activity
Q5LQ10	-	Up	-	cell wall macromolecule catabolic process
Q5LT96	-	Up	-	2 iron, 2 sulfur cluster binding; carbon-monoxide dehydrogenase (acceptor) activity; electron carrier activity; metal ion binding
Q5LLV9	-	Up	-	proteolysis; serine-type peptidase activity
Q5LKE9	-	Up	-	3-carboxy-cis,cis-muconate cycloisomerase activity; metabolic process
Q5LSN3	-	Up	-	NADP binding; amino acid binding; cellular amino acid metabolic process; homoserine dehydrogenase activity
Q5LTJ7	-	Up	-	ATP binding; helicase activity; nucleic acid binding
Q5LUD3	-	Up	-	ATP binding; cytoplasm; proline-tRNA ligase activity; prolyl-tRNA aminoacylation
Q5LSF8	-	Up	-	integral to membrane; transmembrane transport; transporter activity
Q5LTY5	-	Up	-	DNA binding; cytoplasm; regulation of transcription, DNA-dependent
Q5LN29	-	Up	-	antioxidant activity; peroxiredoxin activity; response to oxidative stress

Q5LX46	-	Up	-	ATP binding; cob(I)yrinic acid a,c-diamide adenosyltransferase activity; cobalamin biosynthetic process
Q5LT94	-	Up	-	UDP-N-acetylmuramate dehydrogenase activity; flavin adenine dinucleotide binding
Q5LM71	-	Up	-	outer membrane-bounded periplasmic space; transport
Q5LLS7	-	Up	-	3'-5' exonuclease activity; 5'-3' exonuclease activity; DNA binding; DNA replication; DNA-directed DNA polymerase activity
Q5LR12	-	Up	-	P-P-bond-hydrolysis-driven protein transmembrane transporter activity; integral to membrane; intracellular protein transmembrane transport; plasma membrane; protein targeting; protein transport by the Sec complex
Q5LVQ5	-	Up	-	No definition
Q5LW11	-	Up	-	integral to membrane; plasma membrane; protein insertion into membrane; protein transport
Q5LWF6	-	Up	-	tRNA binding; tRNA nucleotidyltransferase activity; tRNA processing; tRNA-specific ribonuclease activity
Q5LTH1	-	Up	-	ATP binding; ATP-dependent helicase activity; nucleic acid binding
Q5LWL7	-	Up	-	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LRH1	-	Up	-	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
Q5LN23	Down	Up	-	3'-5'-exoribonuclease activity; RNA binding; RNA processing; cytoplasm; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity
Q5LLR2	-	Up	-	No definition
Q5LP19	-	Up	-	NADP binding; glucose-6-phosphate dehydrogenase activity; pentose-phosphate shunt
Q5LXE8	-	Up	-	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
Q5LPW0	-	Up	-	proteolysis; serine-type D-Ala-D-Ala carboxypeptidase activity
Q5LQJ4	-	Up	-	DNA binding; DNA recombination; regulation of transcription, DNA-dependent; regulation of translation; transcription, DNA-dependent
Q5LLR9	-	Up	-	No definition
Q5LP80	-	Up	-	ATP binding; glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity; translation
Q5LVC9	-	-	Up	transferase activity, transferring acyl groups other than amino-acyl groups
Q5LN84	Down	-	Up	gamma-aminobutyric acid catabolic process; succinate-semialdehyde dehydrogenase [NAD(P)+] activity
Q5LSU6	Down	Down	Up	Gram-negative-bacterium-type cell outer membrane assembly; cell outer membrane; integral to membrane; plasma membrane; protein insertion into membrane

Q5LWV3	-	-	Up	3'-5' exonuclease activity; DNA binding; DNA polymerase III complex; DNA replication; DNA-directed DNA polymerase activity; cytoplasm
Q5LMZ4	-	-	Up	carbon-nitrogen ligase activity, with glutamine as amido-N-donor
Q5LQB9	-	-	Up	outer membrane-bounded periplasmic space; transport
Q5LSL1	-	-	Up	membrane; polysaccharide transmembrane transporter activity
Q5LTQ5	Down	-	Up	ATP binding; arginine-tRNA ligase activity; arginyl-tRNA aminoacylation; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
Q5LSM4	-	Down	Up	cellular amino acid metabolic process; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
Q5LNT3	-	-	Up	protein methyltransferase activity; ribosome
Q5LMI7	Down	Down	Up	cell outer membrane; integral to membrane; plasma membrane
Q5LTY7	Down	-	Up	dTMP biosynthetic process; flavin adenine dinucleotide binding; thymidylate synthase (FAD) activity
Q5LS49	-	-	Up	cysteine biosynthetic process from serine; cysteine synthase activity; transferase activity
Q5LP17	-	-	Up	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate
Q5LRR9	-	-	Up	thiamine biosynthetic process; thiamine-phosphate diphosphorylase activity
Q5LU63	-	-	Up	UDP-N-acetyl-D-glucosamine:N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelyl-D-alanyl-D-alanine-diphosphoundecaprenol 4-beta-N-acetylglucosaminlytransferase activity; UDP-N-acetylgalactosamine biosynthetic process; carbohydrate binding; cell cycle; cell division; lipid glycosylation; peptidoglycan biosynthetic process; plasma membrane; regulation of cell shape; undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase activity
Q5LNX4	-	-	Up	ATP binding; DNA binding; DNA topoisomerase type I activity; DNA topological change; chromosome; metal ion binding
Q5LPR7	-	-	Up	NAD binding; NADH dehydrogenase (quinone) activity; photosynthesis, light reaction; plasma membrane; quinone binding; transport
Q5LX27	-	-	Up	oxidoreductase activity; zinc ion binding
Q5LW87	-	-	Up	ATP binding; cytoplasm; lysine-tRNA ligase activity; lysyl-tRNA aminoacylation; tRNA binding
Q5LV40	-	-	Up	integral to membrane; plasma membrane; transporter activity
Q5LLJ8	-	-	Up	electron carrier activity; heme binding
Q5LU20	-	-	Up	protein-L-isoaspartate (D-aspartate) O-methyltransferase activity
Q5LTK9	-	-	Up	oxidation-reduction process; oxidoreductase activity
Q5LML3	-	-	Up	ATP binding; DNA binding; DNA polymerase III complex; DNA replication; DNA-directed DNA polymerase activity; nucleoside-triphosphatase activity

Q5LN03	-	-	Up	regulation of transcription, DNA-dependent
Q3V7J4	-	-	Up	acetolactate synthase activity; branched-chain amino acid biosynthetic process; flavin adenine dinucleotide binding; magnesium ion binding; thiamine pyrophosphate binding
Q5LSV7	Down	Down	Up	oligopeptide transporter activity
Q5LL24	-	-	Up	proteolysis; serine-type endopeptidase activity
Q5LME7	Down	-	Up	ATP binding; acetyl-CoA carboxylase activity; acetyl-CoA carboxylase complex; fatty acid biosynthetic process; malonyl-CoA biosynthetic process
Q5LP21	Down	-	Up	2-dehydro-3-deoxy-phosphogluconate aldolase activity; 4-hydroxy-2-oxoglutarate aldolase activity
Q5LWU8	-	-	Up	ATP binding; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; DNA-dependent DNA replication; chromosome; cytoplasm; magnesium ion binding
Q5LNN8	-	-	Up	plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1)
Q5LTI3	-	-	Up	O-acetylhomoserine aminocarboxypropyltransferase activity; cysteine metabolic process; pyridoxal phosphate binding
Q5LMG3	-	-	Up	glycolysis; magnesium ion binding; potassium ion binding; pyruvate kinase activity
Q5LVN3	-	-	Up	ATP binding; cytoplasm; histidine-tRNA ligase activity; histidyl-tRNA aminoacylation
Q5LUE9	-	-	Up	cobalamin binding; metal ion binding; methylmalonyl-CoA mutase activity
Q5LVC3	-	-	Up	metal ion binding; pentose-phosphate shunt; ribulose-phosphate 3-epimerase activity
Q5LS82	-	-	Up	No definition
Q5LVK0	Down	Down	Down	NAD binding; NADP binding; glucose metabolic process; oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
Q5LW55	Down	Down	Down	rRNA binding; ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LW59	Down	Down	Down	large ribosomal subunit; rRNA binding; structural constituent of ribosome; transferase activity; translation
Q5LRY1	Down	Down	Down	ribosome; structural constituent of ribosome; translation
Q5LVJ2	Down	Down	Down	ATP binding; cytoplasm; gluconeogenesis; phosphoenolpyruvate carboxykinase (ATP) activity
Q5LR48	Down	Down	Down	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LVA0	Down	Down	Down	RNA binding; ribosome; structural constituent of ribosome; translation
Q5LQ83	Down	Down	Down	ATP binding; ATP catabolic process; ATP-dependent peptidase activity; cellular response to stress; cytoplasm; misfolded or incompletely synthesized protein catabolic process; sequence-specific DNA binding; serine-type endopeptidase activity

Q5LV47	Down	Down	Down	xylulokinase activity; xylulose metabolic process
Q5LVB0	Down	Down	Down	3-hydroxyisobutyrate dehydrogenase activity; coenzyme binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity; valine metabolic process
Q5LW38	Down	Down	Down	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LVI1	Down	Down	Down	electron carrier activity
Q5LUP9	Down	Down	Down	ATP binding; nucleoside-triphosphatase activity; protein folding; zinc ion binding
Q5LLU5	Down	Down	Down	ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LMQ0	Down	Down	Down	large ribosomal subunit; rRNA binding; regulation of translation; structural constituent of ribosome; tRNA binding; translation
Q5LNP0	Down	Down	Down	ATP binding; plasma membrane; plasma membrane ATP synthesis coupled proton transport; proton-transporting ATP synthase activity, rotational mechanism; proton-transporting ATP synthase complex, catalytic core F(1); proton-transporting ATPase activity, rotational mechanism
Q5LW48	Down	Down	Down	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LLZ6	Down	Down	Down	glutamate biosynthetic process; glutamate synthase (NADPH) activity
Q5LR95	Down	Down	Down	peptidyl-prolyl cis-trans isomerase activity; protein folding; protein peptidyl-prolyl isomerization
Q5LRZ5	Down	Down	Down	cytoplasm; translation elongation factor activity
Q5LPA8	Down	Down	Down	cytoplasm; glycine biosynthetic process from serine; glycine hydroxymethyltransferase activity; pyridoxal phosphate binding; tetrahydrofolate interconversion
Q5LLG6	Down	Down	Down	catalase activity; heme binding; hydrogen peroxide catabolic process; metal ion binding
Q5LW41	Down	Down	Down	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LTW7	Down	Down	Down	catalytic activity; regulation of biosynthetic process
Q5LS76	Down	Down	Down	ATP binding; intracellular signal transduction; nucleoside-triphosphatase activity; phosphorelay response regulator activity; regulation of transcription, DNA-dependent; sequence-specific DNA binding; transcription, DNA-dependent
Q5LU54	Down	Down	Down	GTP binding; GTPase activity; barrier septum assembly; cell division site; cytokinesis by binary fission; cytoplasm; protein complex; protein polymerization
Q5LW33	Down	Down	Down	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LV15	Down	Down	Down	ATP binding; cytoplasm; protein refolding
Q5LXC4	Down	Down	Down	ATP binding; magnesium ion binding; manganese ion binding; succinate-CoA ligase (ADP-forming) activity; tricarboxylic acid cycle
Q5LRJ1	Down	Down	-	DNA replication; cobalamin binding; cobalt ion binding; deoxyribonucleotide biosynthetic process; nucleotide binding; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor

Q5LKG7	Down	Down	-	oxidoreductase activity
Q5LV16	Down	Down	-	ATP binding; cytoplasm; protein folding
Q5LNH0	Down	Down	-	ATP hydrolysis coupled proton transport; ATP synthesis coupled proton transport; hydrogen ion transmembrane transporter activity; integral to membrane; lipid binding; plasma membrane; proton-transporting ATP synthase complex, coupling factor F(o)
Q5LLS1	Down	Down	-	S-(hydroxymethyl)glutathione dehydrogenase activity; ethanol oxidation; zinc ion binding
Q5LPB6	Down	Down	-	NAD binding; cytosol; malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity; malate dehydrogenase (oxaloacetate-decarboxylating) activity; malate metabolic process; manganese ion binding; transferase activity, transferring acyl groups
Q5LKR6	Down	Down	-	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
Q5LPS2	Down	Down	-	4 iron, 4 sulfur cluster binding; FMN binding; NAD binding; NADH dehydrogenase (ubiquinone) activity; metal ion binding
Q5LR49	Down	Down	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW24	Down	Down	-	transporter activity
Q5LW43	Down	Down	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LLR0	Down	Down	-	adenosylhomocysteinase activity; cytoplasm; one-carbon metabolic process
Q5LPT9	Down	Down	-	RNA binding; hydrolase activity, acting on ester bonds; metal ion binding
Q5LRM3	Down	Down	-	4 iron, 4 sulfur cluster binding; cytoplasm; lipoate synthase activity; metal ion binding; protein lipoylation
Q5LTA6	Down	-	Down	calcium ion binding; membrane; outer membrane-bounded periplasmic space; oxidoreductase activity, acting on CH-OH group of donors
Q5LW46	Down	-	Down	rRNA binding; ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LM34	Down	-	Down	No definition
Q5LLU7	Down	-	Down	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LXD1	Down	-	Down	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
Q5LW50	Down	-	Down	ribosome; structural constituent of ribosome; translation
Q5LWJ6	Down	-	Down	ATP binding; protein folding; response to stress
Q5LNE8	Down	-	Down	ribosome; structural constituent of ribosome; translation
Q5LP64	Down	-	Down	catalytic activity; response to tellurium ion
Q5LX06	Down	-	Down	3-isopropylmalate dehydratase activity; 4 iron, 4 sulfur cluster binding; leucine biosynthetic process; metal ion binding

Q5LRZ4	-	Down	Down	small ribosomal subunit; structural constituent of ribosome; translation
Q5LW57	-	Down	Down	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
Q5LMR3	-	Down	Down	rRNA binding; small ribosomal subunit; structural constituent of ribosome; tRNA binding; translation
Q5LRY0	-	Down	Down	DNA binding; DNA repair; rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LX13	-	Down	Down	calcium ion binding
Q5LNR7	-	Down	Down	ATP binding; aminoacyl-tRNA editing activity; cytoplasm; isoleucine-tRNA ligase activity; isoleucyl-tRNA aminoacylation; regulation of translational fidelity; zinc ion binding
Q5LW31	-	Down	Down	ribosome; structural constituent of ribosome; translation
Q5LRR0	Down	-	-	3-hydroxybutyrate dehydrogenase activity
Q5LXC5	Down	-	-	ATP binding; ATP citrate synthase activity; cofactor binding; succinate-CoA ligase (ADP-forming) activity
Q5LNB6	Down	-	-	catalytic activity
Q5LN38	Down	-	-	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
Q5LM19	Down	-	-	AMP nucleosidase activity; AMP salvage
Q5LR03	Down	-	-	No definition
Q5LQ59	Down	-	-	ferredoxin-NADP+ reductase activity
Q5LV77	Down	-	-	amino acid transport; outer membrane-bounded periplasmic space
Q5LSA5	Down	-	-	pteridine-containing compound metabolic process
Q5LR56	Down	-	-	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; NAD binding; fatty acid biosynthetic process
Q5LQ60	Down	-	-	No definition
Q5LR50	Down	-	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LS84	Down	-	-	homocysteine S-methyltransferase activity
Q5LP83	Down	-	-	ribosome; structural constituent of ribosome; translation
Q5LTV6	Down	-	-	metal ion binding; transporter activity
Q5LWG9	Down	-	-	NAD binding; cytosol; malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity; malate dehydrogenase (oxaloacetate-decarboxylating) activity; malate metabolic process; manganese ion binding; transferase activity, transferring acyl groups
Q5LPC2	Down	-	-	GTP binding; UMP salvage; magnesium ion binding; uracil phosphoribosyltransferase activity; uracil salvage

Q5LL27	Down	-	-	NAD binding; histidine biosynthetic process; histidinol dehydrogenase activity; zinc ion binding
Q5LUU0	Down	-	-	N2-acetyl-L-ornithine:2-oxoglutarate 5-aminotransferase activity; arginine biosynthetic process; cytoplasm; pyridoxal phosphate binding
Q5LPR1	Down	-	-	hydroxymethylglutaryl-CoA lyase activity; metabolic process
Q5LM98	Down	-	-	outer membrane-bounded periplasmic space; transport
Q5LLL7	Down	-	-	2 iron, 2 sulfur cluster binding; Gram-negative-bacterium-type cell wall; electron carrier activity; integral to membrane; metal ion binding; oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor; plasma membrane; sodium ion transport
Q5LNA3	Down	-	-	oxidoreductase activity
Q5LRH2	Down	-	-	ATP binding; cytoplasm; glutamate-tRNA ligase activity; glutamyl-tRNA aminoacylation; tRNA binding
Q5LSH2	Down	-	-	4 iron, 4 sulfur cluster binding; anaerobic respiration; electron carrier activity; formate dehydrogenase (NAD+) activity; formate dehydrogenase complex; molybdenum ion binding
Q5LKE1	Down	-	-	3-oxoacyl-[acyl-carrier-protein] synthase activity
Q5LN58	Down	-	-	L-serine biosynthetic process; O-phospho-L-serine:2-oxoglutarate aminotransferase activity; cytoplasm; pyridoxal phosphate binding
Q5LRU3	Down	-	-	GTP binding; GTP catabolic process; GTPase activity
Q5LQ68	Down	-	-	outer membrane-bounded periplasmic space; transport
Q5LM28	Down	-	-	oxidation-reduction process; oxidoreductase activity
Q5LRI6	Down	-	-	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process
Q5LSB3	Down	-	-	No definition
Q5LR02	Down	-	-	No definition
Q5LUX6	Down	-	-	ATP binding; aspartate-tRNA ligase activity; aspartyl-tRNA aminoacylation; cytoplasm; nucleic acid binding
Q5LV41	Down	-	-	D-xylose transport; monosaccharide binding
Q5LTW9	Down	-	-	No definition
Q5LMY1	Down	-	-	No definition
Q5LV98	Down	-	-	DNA binding; DNA recombination; chromosome; regulation of transcription, DNA-dependent; regulation of translation; transcription, DNA-dependent
Q5LLU8	Down	-	-	nucleotide binding; rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LM18	Down	-	-	adenine catabolic process; adenine deaminase activity
Q5LS33	Down	-	-	ATP binding; ATP catabolic process; ATPase activity

Q5LX99	Down	-	-	No definition
Q5LXF9	Down	-	-	DNA binding; intracellular; regulation of transcription, DNA-dependent
Q5LR87	Down	-	-	dihydrolipoyllysine-residue acetyltransferase activity; pyruvate dehydrogenase complex; pyruvate metabolic process
Q5LS95	Down	-	-	iron-sulfur cluster binding; oxidoreductase activity, acting on the aldehyde or oxo group of donors; thiamine pyrophosphate binding
Q5LKG9	Down	-	-	UDP-N-acetylmuramate dehydrogenase activity; flavin adenine dinucleotide binding
Q5LLU6	Down	-	-	rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LM20	Down	-	-	DNA binding
Q5LLG8	Down	-	-	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
Q5LNX0	Down	-	-	3-oxoadipate CoA-transferase activity; cellular aromatic compound metabolic process
Q5LQK2	Down	-	-	3-hydroxyacyl-CoA dehydrogenase activity
Q5LV87	-	Down	-	tryptophan synthase activity
Q5LMQ4	-	Down	-	ribosome; structural constituent of ribosome; translation
Q5LLM8	-	Down	-	ATP binding; DNA-dependent transcription, termination; RNA binding; RNA-dependent ATPase activity; helicase activity; regulation of transcription, DNA-dependent
Q5LWT8	-	Down	-	enoyl-CoA hydratase activity
Q5LV39	-	-	Down	ATP binding; ATP catabolic process; ATPase activity
Q5LP30	-	-	Down	ATP binding; aminoacyl-tRNA editing activity; cytoplasm; regulation of translational fidelity; valine-tRNA ligase activity; valyl-tRNA aminoacylation
Q5LV37	-	-	Down	oxidation-reduction process; oxidoreductase activity
Q5LV90	-	-	Down	5S rRNA binding; ribosome; structural constituent of ribosome; translation
Q5LW34	-	-	Down	rRNA binding; ribosome; structural constituent of ribosome; tRNA binding; translation
Q5LW56	-	-	Down	mRNA binding; rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
Q5LR37	-	-	Down	cytoplasm; glutamate-ammonia ligase activity; glutamine biosynthetic process; nitrogen fixation
Q5LSA3	-	-	Down	metal ion binding; transketolase activity

* Proteins were considered to be up- or down-regulated if the FDRup or FDRdown was $p < 0.05$.

Appendix A.6.

Table A6.2. A list of proteins identified to be significantly up- or down- regulated in Fe-limited cells by Qspec from analyses of 4 biological splits from each culture.

	Global ID number	Protein Function	QSpec Protein abundance Log_2 fold change	Cellular Processes
Down-regulated proteins in Fe-limited cells	224015655	nickel ABC transporter periplasmic protein	-2.87	TRANSPORTER,AAmet
	118411185	30S ribosomal protein S4	-1.63	TRANSLATION
	223994385	chloroplast 50S ribosomal protein L15	-1.61	TRANSLATION
	223993577	Predicted Protein	-1.38	
	209583663	Predicted protein B0432.4	-1.26	
	223998678	Phosphoenolpyruvate carboxylase (PEPC2)	-1.24	PHOTO, ENERGY
	118411124	50S ribosomal protein L1	-1.21	TRANSLATION
	224007537	ATP binding	-1.13	DNA, REG, ENERGY
	224002675	ribosomal protein L22	-1.12	TRANSLATION
	224010633	COG1233: Phytoene dehydrogenase and related proteins	-1.09	OXI-RED
	224009938	RsuA 16S rRNA U516 pseudouridine synthase	-1.08	TRANSLATION
	118411203	50S ribosomal protein L5	-1.08	TRANSLATION
	223994299	Predicted proteins -PFAM: HEC/Ndc80p family	-0.98	DNA
	118411141	cell division protein FtsH-like protein	-0.96	CD, DNA PROT_deg
	223993621	proteophosphoglycan ppg4	-0.95	CMI, CARBmet
	223995685	Predicted cell division protein FtsH2	-0.88	CD, DNA PROT_deg
	118411191	50S ribosomal protein L4	-0.85	TRANSLATION
	118411193	50S ribosomal protein L2	-0.84	TRANSLATION
	209583588	SHM1 glycine hydroxymethyltransferase	-0.82	AAmet
	224007497	Predicted: nicotinamide nucleotide transhydrogenase	-0.82	OXI-RED
	118411222	30S ribosomal protein S6	-0.82	TRANSLATION
	209586173	Ribosomal protein L3	-0.81	TRANSLATION
	223993421	Muc19: mucin 19	-0.81	CMI
	223996073	phosphoribosylpyrophosphate synthetase	-0.8	PPP
	118411201	50S ribosomal protein L14	-0.8	TRANSLATION
	223998400	heat shock protein 83	-0.79	REG
	118411123	50S ribosomal protein L11	-0.78	TRANSLATION
	224008965	60S ribosomal protein L6 CgRPL6	-0.78	TRANSLATION
	224006742	30S ribosomal protein S1	-0.77	TRANSLATION
	118411126	30S ribosomal protein S2	-0.77	TRANSLATION
	224014104	Predicted protein	-0.76	

	224015517	Predicted protein- DNA binding domain	-0.73	DNA
	224013552	pyruvate dehydrogenase (lipoamide) beta	-0.72	Glyc, AAmet
	223999739	chloroplast light harvesting protein isoform 12	-0.7	PHOTO
	224009568	Predicted protein	-0.67	
	224009197	Predicted protein similar to CG9888-PA	-0.67	TRANSLATION
	118411109	ATP synthase CF0 B' chain subunit II	-0.64	ENERGY
	224015308	Photosystem I light harvesting proteins	-0.64	PHOTO
	223995405	Predicted protein	-0.64	
	118411155	cytochrome b6-f complex subunit IV	-0.63	OXI-RED, PHOTO
	223998931	adenosinetriphosphatase	-0.63	ENERGY
	118411180	photosystem II reaction center protein D1	-0.61	PHOTO
	224001616	ATPase, E1-E2 type	-0.58	ENERGY
	118411153	photosystem I ferredoxin-binding protein	-0.58	PHOTO
	118411220	ATP-dependent clp protease ATP-binding subunit	-0.58	ENERGY
	223996813	ribosomal protein S4, Y-linked	-0.57	TRANSLATION
	223992923	Predicted Porin protein	-0.57	TRANSPORTER
	118411137	cytochrome f	-0.56	PHOTO
	118411190	50S ribosomal protein L3	-0.56	TRANSLATION
	224003409	Predicted Protein No BLAST result	-0.54	
	118411135	ATP synthase CF1 epsilon chain	-0.54	ENERGY
	224010635	geranyl-geranyl reductase	-0.53	OXI-RED
	224003107	oxygen-evolving enhancer protein 1 precursor	-0.51	PHOTO
	118411112	ATP synthase CF1 alpha chain	-0.5	ENERGY
Up-regulated proteins in Fe-limited cells	224005154	phosphoglycerate kinase precursor	0.5	Glyc
	224011888	predicted translation elongation factor G	0.5	TRANSLATION, CD
	223994191	Predicted protein	0.59	
	224002408	aconitate hydratase 2 (citrate hydro-lyase 2) (aconitase 2)	0.59	CARBmet, ENERGY
	223993043	glyceraldehyde-3-phosphate dehydrogenase precursor	0.6	Glyc
	224009658	phosphoadenosine-phosphosulphate reductase	0.6	OXI-RED, ENERGY
	223993867	putative CDC48/ATPase	0.63	ENERGY
	224000661	fructose-1,6-bisphosphate aldolase precursor	0.65	Glyc
	223996511	Spermine synthase	0.66	CMI
	209583455	copper-induced girdle band-associated cell surface protein	0.67	CMI
	223998614	Predicted protein	0.68	
	223999927	Phosphoglucomutase, cytoplasmic (PGM)	0.69	Glyc
	223997268	manganese superoxide dismutase	0.69	OXI-RED
	223999217	phosphoribulokinase	0.71	ENERGY

224012331	ascorbate peroxidase	0.73	CARBmet, AAmet
223993279	2-isopropylmalate synthase A	0.73	AAmet
118411218	translation elongation factor Tu	0.76	TRANSLATION, CD
223999031	CbbX protein homolog	0.76	ENERGY
224007002	Adenosine kinase	0.77	TRANSLATION
224007705	elongation factor alpha-like protein	0.77	TRANSLATION, CD
223993031	Fe transporter Ftr1	0.78	TRANSPORTER, IT
224010100	histone H2A.1	0.79	REG
224001660	GDP-mannose dehydratase	0.82	CARBmet
224013212	fucoxanthin chlorophyll a/c binding protein	0.87	PHOTO
224012529	HSP70-like protein	0.87	CMI
224000862	putative signal-transduction protein with CBS domains	0.87	REG
223992617	putative aminotransferase AGD2	0.87	AAmet
223995719	membrane-associated 30 kD protein-like	0.89	CMI, REG
223999607	clathrin binding	0.91	IT
224013082	N-acetylornithine aminotransferase	0.91	AAmet
224013261	Predicted 3-oxoacyl-[acyl-carrier-protein] synthase	0.92	FA
118411104	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	0.94	PHOTO
224001930	COG0473: Isocitrate/isopropylmalate dehydrogenase	1.02	AAmet
224002995	elongation factor 2	1.03	TRANSLATION
224001430	ATPREP1/ATZNMP; metalloendopeptidase	1.07	PROT_deg
224006554	Tubulin alpha-2 chain	1.1	CMI, IT
209583661	GTPAse: T06D8.1b	1.1	REG, CD, IT
224010070	serine hydroxymethyltransferase	1.14	AAmet
223993519	ADP-ribosylation factor	1.14	IT, REG
224013196	stress-inducible protein STI1 homolog	1.17	REG
223999013	protein kinase, putative	1.18	CARBmet, AAmet
224001286	2nd Hit Predicted protein CBG01077	1.19	DNA
118411164	Rubisco expression protein	1.19	PHOTO
223992817	predicted protein	1.21	
224003531	Pyruvate dehydrogenase	1.21	AAmet, Glyc, OXI-RED
223999929	UDP-glucose pyrophosphorylase	1.23	CARBmet, PPP
209583468	copper-induced girdle band-associated cell surface protein	1.24	CMI, CD
223998024	14-3-3 regulatory protein	1.25	REG
223994125	cysD: ATP-sulfurylase	1.3	ENERGY
223993693	Predicted: Bromodomain; RING3	1.35	CD, TRANSLATION

	223997516	Aspartate-semialdehyde dehydrogenase, USG-1 related	1.36	AAmet
	224001278	enolase	1.37	Glyc
	223994119	Predicted HSP70 heat shock 70kD protein 4	1.39	REG
	224001636	predicted protein cgd4_200	1.39	
	223996962	proliferating cell nuclear antigen	1.49	DNA
	223999219	Phosphofructokinase	1.5	Glyc
	224008957	Predicted protein: ProSite- Protamine P1 signature	1.53	CD, AAmet
	224008733	similar to coatomer protein complex, subunit gamma 2	1.54	TRANS
	223993357	Tubulin beta chain (Beta tubulin)	1.56	CMI
	223995669	Predicted RNA Helicase	1.58	TRANSLATION
	224013856	Hmg protein 1.2, isoform c	1.6	DNA
	223999447	asparagine synthetase	1.63	AAmet
	224009908	aspartate-ammonia ligase	1.63	AAmet
	223997294	fructose-bisphosphate aldolase	1.65	Glyc
	209586218	Predicted protein related to NonF protein	1.68	REG, PROT_deg
	224012695	ClpB protein	1.7	PROT_deg
	224014576	COG0166: Glucose-6-phosphate isomerase	1.75	Glyc
	223997294	fructose-1,6-biphosphate aldolase precursor	1.78	Glyc
	224015421	phosphoesterase	1.79	AAmet
	224009534	UDP-glucose 6-dehydrogenase	1.8	OXI-RED, PPP
	224010838	metalloendopeptidase	1.81	PROT_deg
	223996381	chloroplast cysteine synthase 1 precursor	1.87	AAmet
	224013436	Predicted protein MG01320.4	1.89	CARBmet
	223993045	transketolase	1.91	PPP
	224013544	Predicted protein	1.94	
	224012106	Predicted protein UM04584.1	2.14	AAmet
	223997748	Predicted protein	2.3	OXI-RED

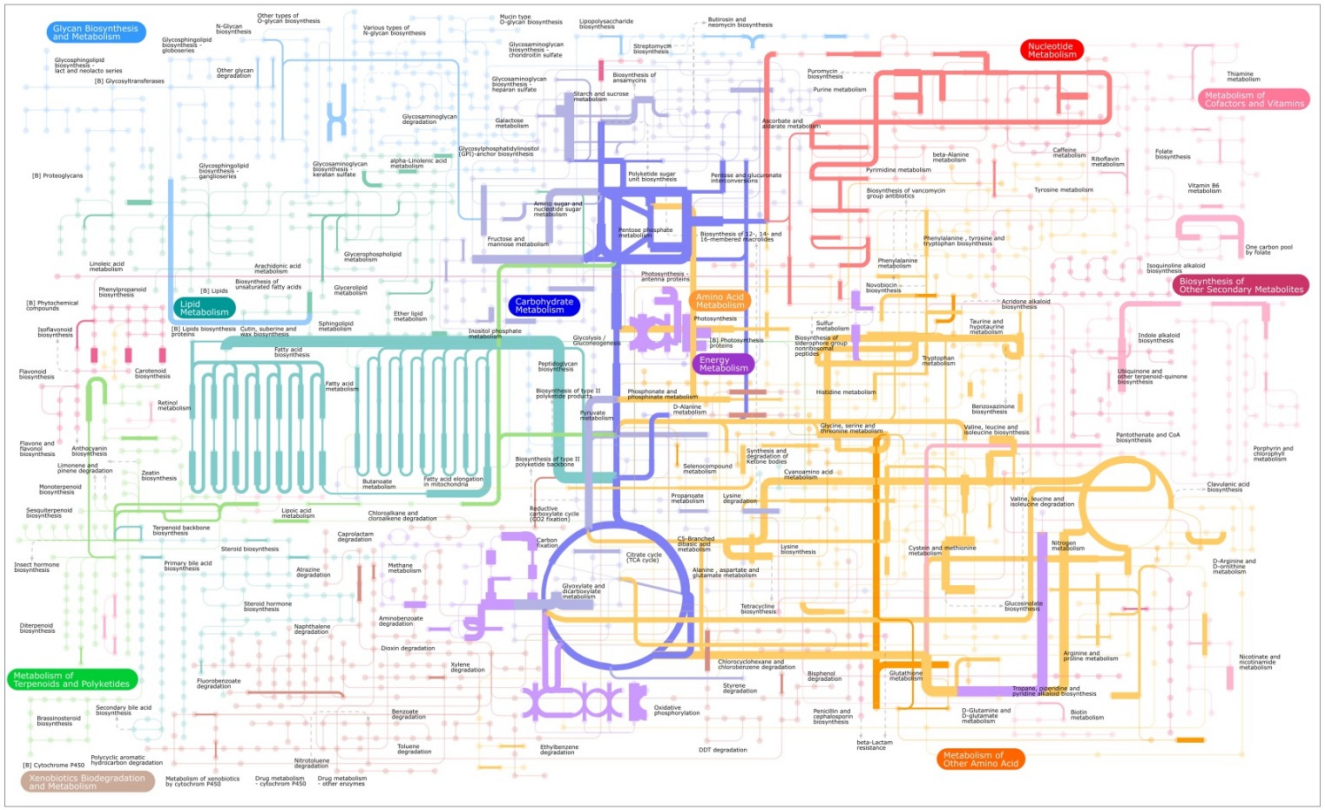


Figure A6.1. Metabolic biochemistry map and relative expression of proteins expressed and identified in Fe-limited *T. pseudonana*. Maps include relative expression data from triplicate PAcIFIC analyses on a tandem mass spectrometer from *Thalassiosira pseudonana* acclimated to Fe-limitation. Each node (or corner) represents a metabolite and the lines connecting the nodes represent an enzyme. A colored line represents proteins that were identified in the particular cell state. The thickness of the line is a function of the number of unique peptides identified from that particular protein [line thickness = 5* $\log_2$(number of unique peptides identified)]. This function was applied to visually express the larger range of protein expression while maintaining a line width between 5-20 pixels. Metabolites were not measured in this study. Colors from top left – light blue: sugar and glycan biosynthesis, light purple: starch and sucrose metabolism (including photosynthesis, oxidative phosphorylation, carbon fixation), dark purple: glycolysis-gluconeogenesis (including the citric acid cycle), red: nucleotide metabolism, teal: lipid metabolism, orange: amino acid metabolism (including urea cycle).

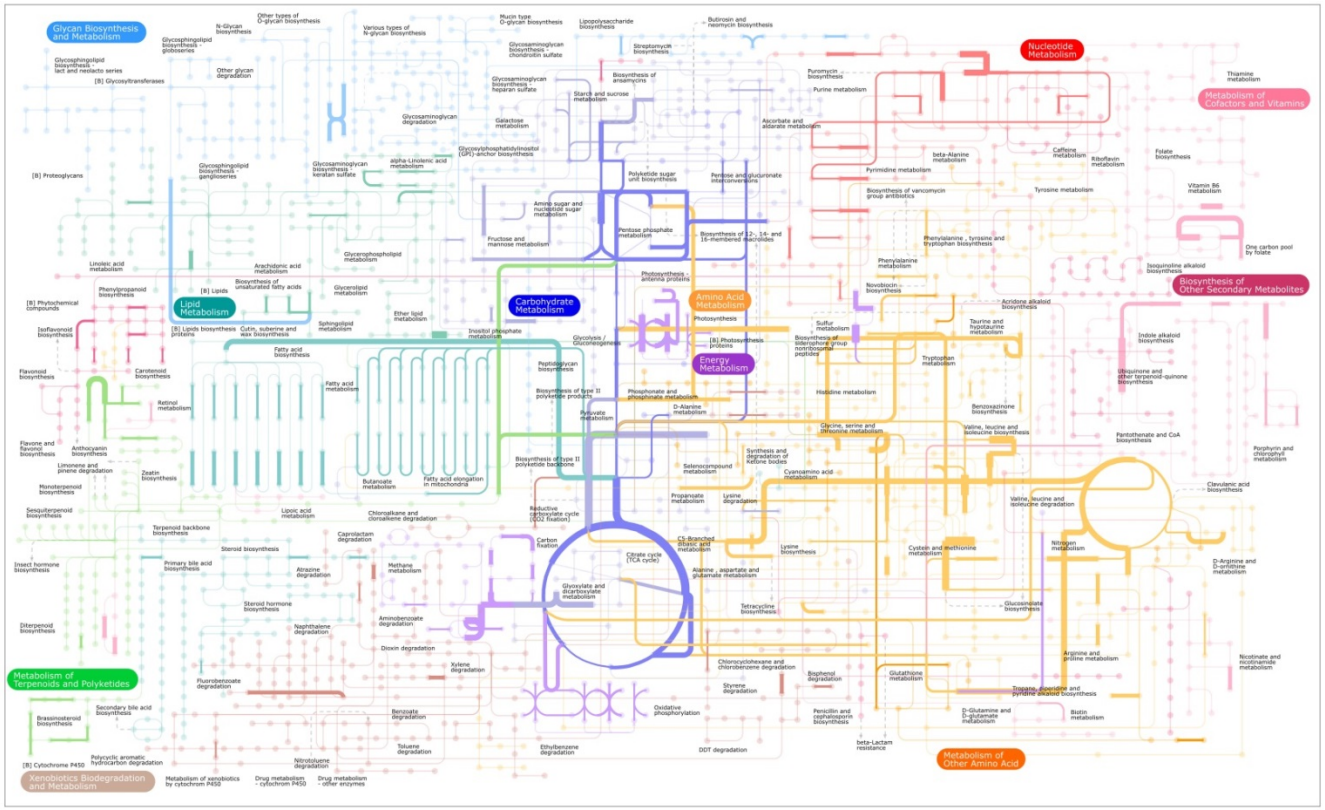
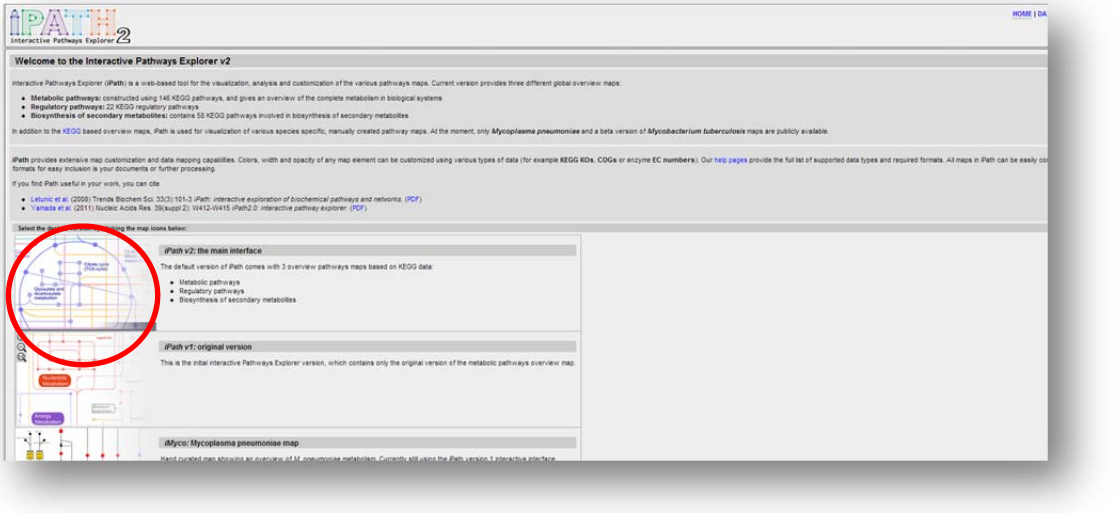


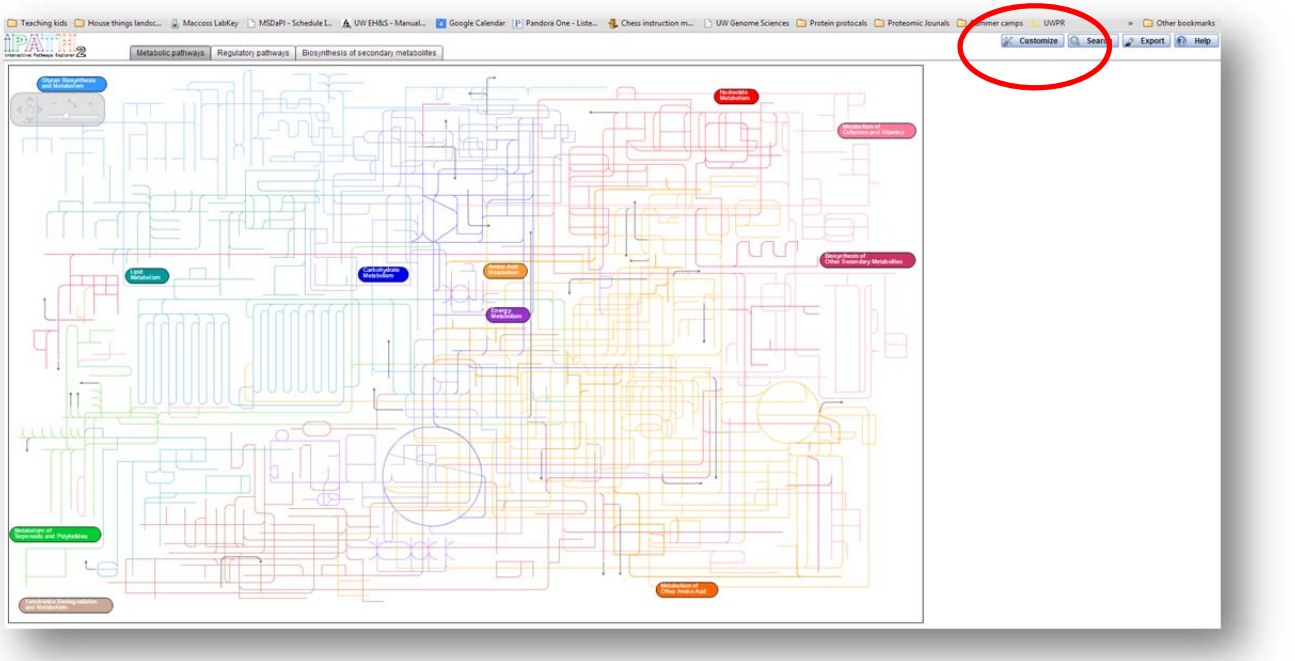
Figure A6.2. Metabolic biochemistry map and relative expression of proteins expressed and identified in Fe-replete *T. pseudonana*. Maps include relative expression data from triplicate PAcIFIC analyses on a tandem mass spectrometer from *Thalassiosira pseudonana* acclimated to Fe-replete conditions. Each node (or corner) represents a metabolite and the lines connecting the nodes represent an enzyme. A colored line represents proteins that were identified in the particular cell state. The thickness of the line is a function of the number of unique peptides identified from that particular protein [line thickness = 5* $\log_2$(number of unique peptides identified)]. This function was applied to visually express the larger range of protein expression while maintaining a line width between 5-20 pixels. Metabolites were not measured in this study. Colors from top left – light blue: sugar and glycan biosynthesis, light purple: starch and sucrose metabolism (including photosynthesis, oxidative phosphorylation, carbon fixation), dark purple: glycolysis-gluconeogenesis (including the citric acid cycle), red: nucleotide metabolism, teal: lipid metabolism, orange: amino acid metabolism (including urea cycle).

Text A6.1. Directions for making iPath Figures using proteomic data from Chapter 6.
This is a brief manual for submitting the raw data to generate more detailed and interactive maps identical to Figure A6.1 and A6.2 using iPath2.0.
Raw data to input is at the end of this document.

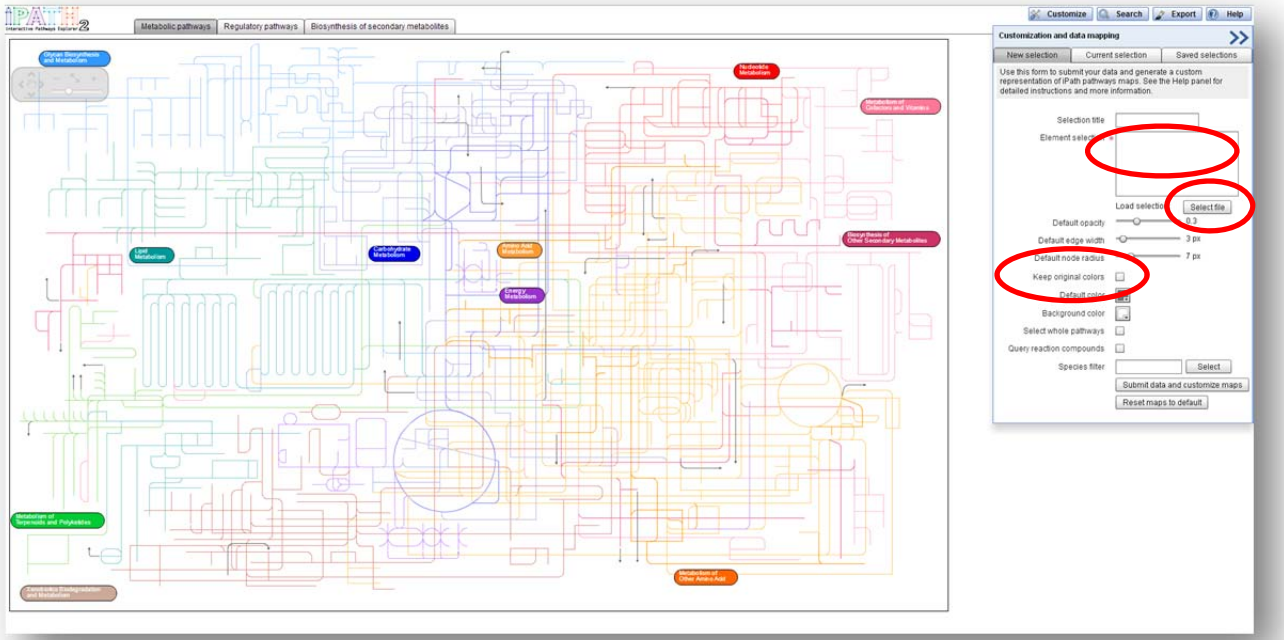
Go to http://pathways.embl.de/
Click on the image to select **iPath v 2: the main interface.**



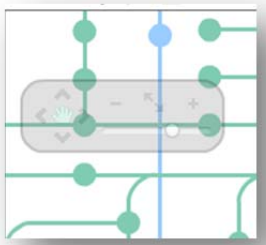
On the upper right corner select "Customize"



This is where you can upload data to generate an iPath map
You can provide a title to the plot by typing in the “Selection title” space

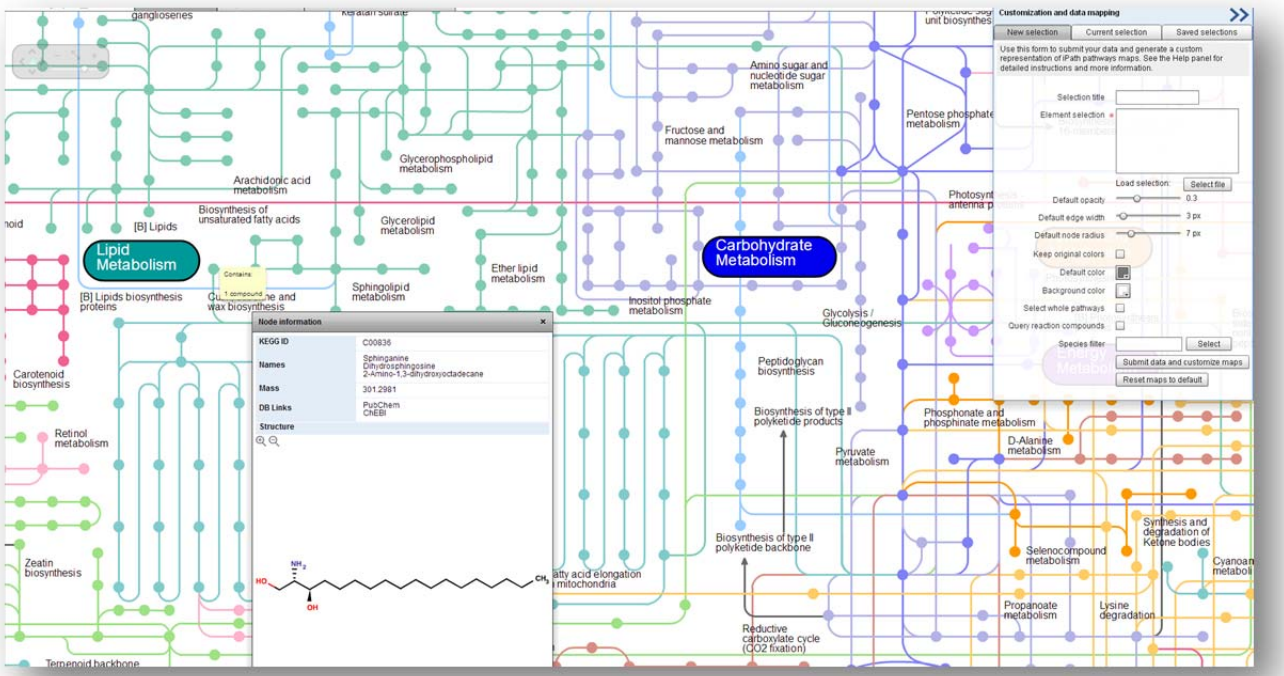


Copy the data that is listed below (at the end of this document) under either Fe-limited or Fe-replete and paste it into the window "Element Selection:. Do not change formatting. You can also save the lists at the bottom of the page as .txt files and upload them directly from your hard-drive using the "Select File" button.



Check the box labeled "keep original colors"
Click button for "Submit data and customize map"
Wait about 20 seconds for the map to be generated.

Now you can zoom in on the proteins identified that are present in the metabolic biochemistry map. Each line represents an enzyme and the nodes represent compounds. The upper left-hand corner has a picture of a "+" and "-" sign to zoom in and out or you can select the hand to push the sheet of paper around to look at different areas while you are zoomed in.



If you take your mouse pointer and hover over a line, you will see that a list of enzymes will be highlighted. You can then select a specific identified enzyme to get more information on that protein. You can also select nodes (dots on the plot). These dots represent different compounds that are synthesized by the enzyme that is represented by the line.

Fe-limited data- copy list below (end at the *) do not change formatting
K02637 1 W5
E1 1 W5
K01419 1 W5
E5.3.4.1 1 W5
K01103 1 W5
K07376 1 W5
K01829 1 W5
K00794 1 W5
E3.2.1.18 1 W5
K03667 1 W5
E3.2.1.18 1 W5
E1 1 W5
K04083 1 W5
E6.3.2.4 1 W5
E1 1 W5
K01426 1 W5
K02838 1 W5
E6.3.2.- 1 W5
E3.2.1.18 1 W5
K07200 1 W5
E5.3.4.1 1 W5
K04487 1 W5
E3 1 W5
E2.7.11.18 1 W5
E1 1 W5
K06980 1 W5
E3.2.1.18 1 W5
K08056 1 W5
K00927 1 W5
K00799 1 W5
K03076 1 W5
K04518 1 W5
K01893 1 W5
K01897 1 W5
K01549 1 W5
K00226 1 W5
E1.1.1.- 1 W5
K00391 1 W5
K03189 1 W5
K06972 1 W5
K02938 1 W5
K01086 1 W5
E3.6.3.14 1 W5
K08493 1 W5
K01530 1 W5
K03106 1 W5
K07517 1 W5
K00801 1 W5
E1.8.1.4 1 W5
E3.6.5.3 1 W5

E2.4.2.11 1 W5
K09838 1 W5
K01875 1 W5
K01463 1 W5
K06269 1 W5
K00698 1 W5
K08869 1 W5
K03694 1 W5
K08517 1 W5
E1 1 W5
E1.10.2.2 1 W5
E1.10.2.2 1 W5
E1.8.1.9 1 W5
K03977 1 W5
K01149 1 W5
E2.1.1 1 W5
K02738 1 W5
K00326 1 W5
K01267 1 W5
K09554 1 W5
K00863 1 W5
K03111 1 W5
K00762 1 W5
K09560 1 W5
K01770 1 W5
K01299 1 W5
K00301 1 W5
E1 1 W5
E5.2.1.8 1 W5
E1.11.1.5 1 W5
K06185 1 W5
K00384 1 W5
K00789 1 W5
K03263 1 W5
K09503 1 W5
K00765 1 W5
K05665 1 W5
K01802 1 W5
K00383 1 W5
E2.3.1.- 1 W5
K05765 1 W5
E1.3.1.33 1 W5
K03754 1 W5
K01876 1 W5
K00540 1 W5
E3.2.1.18 1 W5
K01358 1 W5
E3.2.1.18 1 W5
K01758 1 W5
K01423 1 W5
K03241 1 W5

K02959 1 W5
K09578 1 W5
K02921 1 W5
K04796 1 W5
K02924 1 W5
E1.1.1 1 W5
K07976 1 W5
K03177 1 W5
K03439 1 W5
E3.6.5.3 1 W5
E5.2.1.8 1 W5
K03868 1 W5
E5.3.99.2 1 W5
K08486 1 W5
K00616 1 W5
K03671 1 W5
K01265 1 W5
K03150 1 W5
K01809 1 W5
K01187 1 W5
E1.14.11.2 1 W5
E3.4.24 1 W5
K03097 1 W5
K00263 1 W5
E1.3.1.13 1 W5
K01784 1 W5
E2.7.1.- 1 W5
K00721 1 W5
E5.2.1.8 1 W5
E2.5.1.3 1 W5
K00899 1 W5
K00357 1 W5
E1.3.1.28 1 W5
K00212 1 W5
E1.1.1.65 1 W5
E5.2.1.8 1 W5
E6.3.2.4 1 W5
K07748 1 W5
K03131 1 W5
E1.4.99.1 1 W5
E3.4.19.1 1 W5
K03259 1 W5
K00850 1 W5
E3.6.1.3 1 W5
E3.6.1.3 1 W5
E2.1.1.127 1 W5
E5.2.1.8 1 W5
K01920 1 W5
E3.1.1.47 1 W5
K06119 1 W5
K01874 1 W5

K00837 1 W5
K02291 1 W5
K02013 1 W5
E2.7.1.37 1 W5
K00244 1 W5
E2 1 W5
K06067 1 W5
E5.3.1.9 1 W5
E2.7.1.37 1 W5
E3.2.1.18 1 W5
E3.2.1.18 1 W5
K00359 1 W5
K01028 1 W5
K01265 1 W5
E3.4.23.- 1 W5
K01284 1 W5
K00599 1 W5
K01303 1 W5
K01533 1 W5
K00697 1 W5
E2.3.1.74 1 W5
K01511 1 W5
E2.1.2.10 1 W5
K01850 1 W5
E2.3.1.48 1 W5
K03798 1 W5
E3.2.1.18 1 W5
K01872 1 W5
E3.1.1.3 1 W5
E3.2.1.3 1 W5
K03876 1 W5
K03107 1 W5
E2.7.1.37 1 W5
E5.2.1.8 1 W5
E2.3.1.48 1 W5
E6.3.2.- 1 W5
E3.6.1.3 1 W5
K01516 1 W5
E2.7.1.37 1 W5
K03456 1 W5
E3.2.1.3 1 W5
K00936 1 W5
K01878 1 W5
K02999 1 W5
E3.2.1.18 1 W5
K03240 1 W5
E3.2.1.3 1 W5
E6.3.2.19 1 W5
K03046 1 W7.92
K02694 1 W7.92
K02694 1 W7.92

K02723 1 W7.92
K02723 1 W7.92
K02891 1 W7.92
K02891 1 W7.92
K02997 1 W7.92
K02256 1 W7.92
K02261 1 W7.92
K03237 1 W7.92
E2.7.11.1 1 W7.92
K00812 1 W7.92
K00940 1 W7.92
K00697 1 W7.92
K06889 1 W7.92
K03006 1 W7.92
E3.4.21.92 1 W7.92
K02356 1 W7.92
K03839 1 W7.92
K03671 1 W7.92
K02976 1 W7.92
K00329 1 W7.92
K02971 1 W7.92
E6.3.2.19 1 W7.92
K03937 1 W7.92
E3.6.5.3 1 W7.92
K01165 1 W7.92
K01072 1 W7.92
K03940 1 W7.92
K01109 1 W7.92
E3.1.1.47 1 W7.92
K08727 1 W7.92
E2.3.1.48 1 W7.92
E3.1.3.41 1 W7.92
E1.3.1.33 1 W7.92
K01931 1 W7.92
K03031 1 W7.92
K00845 1 W7.92
K03715 1 W7.92
K07466 1 W7.92
K00891 1 W7.92
E2.5.1.16 1 W7.92
K01870 1 W7.92
K02885 1 W7.92
K04563 1 W7.92
E1.3.1.- 1 W7.92
K00795 1 W7.92
K03250 1 W7.92
K03243 1 W7.92
K05679 1 W7.92
E2.7.11.22 1 W7.92
E3.6.5.3 1 W7.92
K03076 1 W7.92

K03264 1 W7.92
K02978 1 W7.92
K01803 1 W7.92
K07575 1 W7.92
K03564 1 W7.92
K02876 1 W7.92
K02503 1 W7.92
K02958 1 W7.92
E2.7.1.- 1 W7.92
K03841 1 W7.92
E1.8.1.9 1 W7.92
K03100 1 W7.92
K03955 1 W7.92
K00036 1 W7.92
K03257 1 W7.92
K01637 1 W7.92
K01784 1 W7.92
K00559 1 W7.92
K02426 1 W7.92
K00658 1 W7.92
K03029 1 W7.92
E2.1.1.95 1 W7.92
K03033 1 W7.92
K02735 1 W7.92
K03094 1 W7.92
K02979 1 W7.92
K01736 1 W7.92
K09391 1 W7.92
K07127 1 W7.92
K01529 1 W7.92
K01090 1 W7.92
K00826 1 W7.92
K01529 1 W7.92
K03596 1 W7.92
E6.1.1.3 1 W7.92
K04382 1 W7.92
K03038 1 W7.92
E3.2.1.3 1 W7.92
K01754 1 W7.92
K01885 1 W7.92
E6.3.2.- 1 W7.92
K01735 1 W7.92
E3.6.1.- 1 W7.92
E3.6.1.3 1 W7.92
K00592 1 W7.92
E2.7.4.6 1 W7.92
E3.6.5.3 1 W7.92
K03941 1 W7.92
K01362 1 W7.92
K05805 1 W7.92
K01325 1 W7.92

K09550 1 W7.92
K08903 1 W7.92
K01520 1 W7.92
E2.3.1.12 1 W7.92
K08506 1 W7.92
E4.1.1.48 1 W7.92
K07566 1 W7.92
K06944 1 W7.92
E3.1.3.7 1 W7.92
E1 1 W7.92
E2.7.7.50 1 W7.92
K02357 1 W7.92
K03239 1 W7.92
K05768 1 W7.92
K01496 1 W7.92
K01825 1 W7.92
E1 1 W7.92
E2.7.7.6 1 W7.92
K03627 1 W7.92
K06158 1 W7.92
K01881 1 W7.92
K01887 1 W7.92
K08869 1 W7.92
E2.3.1.48 1 W7.92
K00624 1 W7.92
E3.6.5.1 1 W7.92
K10356 1 W7.92
K06948 1 W7.92
E1 1 W7.92
E3.4.24 1 W7.92
K07508 1 W7.92
E2.1.1.- 1 W7.92
E3.6.1.- 1 W7.92
K01099 1 W7.92
K01354 1 W7.92
K06957 1 W7.92
K01848 1 W7.92
E3.2.1.18 1 W7.92
K02709 1 W10
K02709 1 W10
K03043 1 W10
K03043 1 W10
K02949 1 W10
K02040 1 W10
K02125 1 W10
K03936 1 W10
K03935 1 W10
E2.7.1.- 1 W10
E1.5.1.2 1 W10
K02730 1 W10
K06027 1 W10

K01262 1 W10
E3.6.1.3 1 W10
E4.1.1.29 1 W10
K01883 1 W10
K01358 1 W10
K03039 1 W10
K02903 1 W10
K02956 1 W10
K02943 1 W10
E3.2.1.18 1 W10
E2.3.1.12 1 W10
K01829 1 W10
E3.2.1.3 1 W10
K03265 1 W10
E1 1 W10
E1 1 W10
K10251 1 W10
K03255 1 W10
E2.7.11.22 1 W10
E6.3.2.19 1 W10
K09496 1 W10
K00820 1 W10
E3.1.-.- 1 W10
K09498 1 W10
E5.4.2.8 1 W10
K09838 1 W10
E1.10.2.2 1 W10
K09272 1 W10
K00928 1 W10
K07515 1 W10
K05850 1 W10
K03100 1 W10
K06063 1 W10
K09015 1 W10
K02997 1 W10
K08678 1 W10
K03531 1 W10
K00164 1 W10
E1.14.11 1 W10
K01834 1 W10
K00872 1 W10
K00831 1 W10
K02500 1 W10
K02983 1 W10
K00813 1 W10
E5.1.3.3 1 W10
K00939 1 W10
K02838 1 W10
K00133 1 W10
K01267 1 W10
K01783 1 W10

K01783 1 W10
K01783 1 W10
K00344 1 W10
K00800 1 W10
K03778 1 W10
K09494 1 W10
K03030 1 W10
K01802 1 W10
K00324 1 W10
K03037 1 W10
E3.4.11.18 1 W10
K03266 1 W10
E5.2.1.8 1 W10
K08675 1 W10
K01802 1 W10
E1.11.1.5 1 W10
K00134 1 W10
K08906 1 W10
K03238 1 W10
K02951 1 W10
K03455 1 W10
K03704 1 W10
K03010 1 W10
K06067 1 W10
K01899 1 W10
K00814 1 W10
K03671 1 W10
K09458 1 W10
K02725 1 W10
E2.1.1 1 W10
K03236 1 W10
K04488 1 W10
K05993 1 W10
E3.6.5.3 1 W10
K00869 1 W10
E3.4.24.- 1 W10
K05610 1 W10
K03797 1 W10
K01465 1 W10
E1.14.11 1 W10
E3.1 1 W10
E3.6.4.1 1 W10
E3.6.1.3 1 W10
K06972 1 W10
K04523 1 W10
K02109 1 W11.61
K02109 1 W11.61
K02112 1 W11.61
K02112 1 W11.61
K02868 1 W11.61
K02868 1 W11.61

K02989 1 W11.61
E4.3.2.2 1 W11.61
K01441 1 W11.61
E2.3.1.48 1 W11.61
K00099 1 W11.61
K00022 1 W11.61
E2.4.1.34 1 W11.61
K01874 1 W11.61
K09499 1 W11.61
K01886 1 W11.61
E2.1.1.127 1 W11.61
K03531 1 W11.61
K07877 1 W11.61
K01840 1 W11.61
K02155 1 W11.61
E3.2.1.18 1 W11.61
E2.7.11.22 1 W11.61
K01148 1 W11.61
K01889 1 W11.61
K01072 1 W11.61
K02736 1 W11.61
E3.1.13.4 1 W11.61
E3.2.1.18 1 W11.61
K02868 1 W11.61
K00025 1 W11.61
K03943 1 W11.61
K03428 1 W11.61
K04802 1 W11.61
K03797 1 W11.61
K03061 1 W11.61
E3.6.5.3 1 W11.61
K01657 1 W11.61
K00260 1 W11.61
K00014 1 W11.61
K02964 1 W11.61
K00813 1 W11.61
K00955 1 W11.61
K02910 1 W11.61
K02973 1 W11.61
K03527 1 W11.61
E1 1 W11.61
K02917 1 W11.61
K01952 1 W11.61
K05808 1 W11.61
K00940 1 W11.61
K08902 1 W11.61
K02896 1 W11.61
K01938 1 W11.61
K07304 1 W11.61
K03070 1 W11.61
E3.1.2.16 1 W11.61

K08910 1 W11.61
K00942 1 W11.61
E1 1 W11.61
K00248 1 W11.61
K03531 1 W11.61
K00605 1 W11.61
K01358 1 W11.61
K09493 1 W11.61
K05915 1 W11.61
K02151 1 W11.61
K01625 1 W11.61
K01885 1 W11.61
E3.6.5.3 1 W11.61
K01951 1 W11.61
E1.2.4.1 1 W11.61
K01887 1 W11.61
K02975 1 W11.61
K00927 1 W11.61
E5.3.4.1 1 W11.61
K06942 1 W11.61
K01662 1 W11.61
K02891 1 W11.61
K00940 1 W11.61
K01529 1 W11.61
K06911 1 W11.61
K01599 1 W11.61
K03242 1 W11.61
K00945 1 W11.61
K02728 1 W11.61
K00232 1 W11.61
K00525 1 W11.61
K04508 1 W11.61
E1.14.11 1 W11.61
K02981 1 W12.92
K02981 1 W12.92
K02707 1 W12.92
K02707 1 W12.92
K02699 1 W12.92
K02699 1 W12.92
K02886 1 W12.92
K02985 1 W12.92
K00646 1 W12.92
E2.7.7 1 W12.92
K01676 1 W12.92
K02873 1 W12.92
K01175 1 W12.92
E1.9.3.1 1 W12.92
K01679 1 W12.92
K02912 1 W12.92
E2.1.1.45 1 W12.92
K01599 1 W12.92

K07976 1 W12.92
K03254 1 W12.92
K00366 1 W12.92
K02040 1 W12.92
K00088 1 W12.92
K00231 1 W12.92
K03686 1 W12.92
E3.4.24 1 W12.92
K04799 1 W12.92
K03064 1 W12.92
K06180 1 W12.92
K00799 1 W12.92
K03676 1 W12.92
K08776 1 W12.92
K00600 1 W12.92
K00295 1 W12.92
K02864 1 W12.92
K03036 1 W12.92
K09500 1 W12.92
K02889 1 W12.92
K01276 1 W12.92
K00926 1 W12.92
K00228 1 W12.92
E2.7.1.40 1 W12.92
K01890 1 W12.92
K02087 1 W12.92
K01529 1 W12.92
K00428 1 W12.92
K09503 1 W12.92
K00514 1 W12.92
E5.3.4.1 1 W12.92
K03062 1 W12.92
K07977 1 W12.92
K06942 1 W12.92
K02883 1 W12.92
K01939 1 W12.92
E1 1 W12.92
K01610 1 W12.92
K00162 1 W12.92
K06944 1 W12.92
K02693 1 W14.04
K03046 1 W14.04
K03046 1 W14.04
K02925 1 W14.04
K02993 1 W14.04
K01441 1 W14.04
K01992 1 W14.04
E3.2.1.18 1 W14.04
E2.5.1.47 1 W14.04
E1.2.4.1 1 W14.04
K00873 1 W14.04

E1 1 W14.04
K02949 1 W14.04
E3.6.3.14 1 W14.04
E3.1.3.16 1 W14.04
E6.3.2.- 1 W14.04
E1.11.1.7 1 W14.04
K08908 1 W14.04
K01915 1 W14.04
K00948 1 W14.04
K01256 1 W14.04
K02898 1 W14.04
E1.11.1.15 1 W14.04
K02731 1 W14.04
K00627 1 W14.04
E3.1.4.43 1 W14.04
K00161 1 W14.04
K03178 1 W14.04
K02974 1 W14.04
K00266 1 W14.04
K06118 1 W14.04
E3.6.1.3 1 W14.04
K01733 1 W14.04
K02991 1 W14.04
K00088 1 W14.04
E1.3.1.26 1 W14.04
E3.6.5.3 1 W14.04
K00003 1 W14.04
K02519 1 W14.04
K02726 1 W14.04
E3.6.1.3 1 W14.04
K03767 1 W14.04
K07976 1 W14.04
K03687 1 W14.04
K06212 1 W14.04
K01835 1 W14.04
K01803 1 W14.04
K00972 1 W14.04
E1.2.1.3 1 W14.04
K09488 1 W14.04
K02877 1 W14.04
K02870 1 W14.04
K07936 1 W14.04
K01710 1 W14.04
K07151 1 W14.04
E3 1 W14.04
K01855 1 W14.04
E2.1.1 1 W14.04
E3.4.21 1 W14.04
K01749 1 W14.04
K00939 1 W14.04
E2.1.1 1 W14.04

K02932 1 W15
E3.6.1.3 1 W15
K00930 1 W15
K01529 1 W15
K00145 1 W15
K01509 1 W15
E1 1 W15
E1 1 W15
E2.7.11.1 1 W15
K07976 1 W15
E2.3.1.48 1 W15
K03032 1 W15
K02894 1 W15
K03063 1 W15
K04499 1 W15
K01750 1 W15
K00698 1 W15
K02957 1 W15
K01256 1 W15
E1.1.1.18 1 W15
K02727 1 W15
K06445 1 W15
K09540 1 W15
K01254 1 W15
K00620 1 W15
K01755 1 W15
K02989 1 W15
K05864 1 W15
K03066 1 W15
K01358 1 W15
K06207 1 W15
K01778 1 W15
E1.2.1 1 W15
K00764 1 W15
K02882 1 W15
K02977 1 W15
K08738 1 W15
K01698 1 W15
K00627 1 W15
K07755 1 W15
E1.1.1.100 1 W15
K10413 1 W15
K02114 1 W15.85
K02114 1 W15.85
K02991 1 W15.85
E1.3.99.23 1 W15.85
K01961 1 W15.85
K02134 1 W15.85
E2.1.1 1 W15.85
K01868 1 W15.85
K07976 1 W15.85

K00908 1 W15.85
E1.8.1.4 1 W15.85
K02940 1 W15.85
K02969 1 W15.85
K01201 1 W15.85
K01933 1 W15.85
E1.1.1.31 1 W15.85
K02732 1 W15.85
K02150 1 W15.85
K01802 1 W15.85
E1 1 W15.85
K03942 1 W15.85
K02955 1 W15.85
K00101 1 W15.85
K01895 1 W15.85
K02149 1 W15.85
K01880 1 W15.85
K02900 1 W15.85
K01689 1 W15.85
K01803 1 W15.85
K02893 1 W15.85
K02690 1 W16.61
K02690 1 W16.61
K02870 1 W16.61
K02870 1 W16.61
K02635 1 W16.61
K02635 1 W16.61
K02930 1 W16.61
E3.2.1.3 1 W16.61
K01537 1 W16.61
K01624 1 W16.61
E2.3.1.48 1 W16.61
E2.7.1.150 1 W16.61
E1.4.1.4 1 W16.61
K02953 1 W16.61
K00297 1 W16.61
K02985 1 W16.61
K02729 1 W16.61
K00059 1 W16.61
K03405 1 W16.61
K00033 1 W16.61
K02641 1 W16.61
K01858 1 W16.61
K03253 1 W16.61
K02875 1 W16.61
K01869 1 W16.61
K04564 1 W16.61
E2.7.7 1 W16.61
K03065 1 W16.61
K02692 1 W17.3
K02692 1 W17.3

K03070 1 W17.3
K02479 1 W17.3
K02945 1 W17.3
K01586 1 W17.3
E3.6.-.- 1 W17.3
K01469 1 W17.3
E6.5.1.1 1 W17.3
E3.6.4.9 1 W17.3
E3.6.4.1 1 W17.3
K02981 1 W17.3
K00360 1 W17.3
K00963 1 W17.3
K00058 1 W17.3
K02987 1 W17.3
K01626 1 W17.3
K01687 1 W17.3
K00602 1 W17.3
K02866 1 W17.3
K02357 1 W17.3
K04078 1 W17.3
K01428 1 W17.3
K01069 1 W17.3
K01738 1 W17.3
K00382 1 W17.3
E3.4.24 1 W17.3
K01810 1 W17.3
K02993 1 W17.3
K03626 1 W17.3
K00820 1 W17.3
K01529 1 W17.3
K01711 1 W17.3
K01529 1 W17.3
K10046 1 W17.3
K00234 1 W17.3
E6.1.1.9 1 W17.3
K00873 1 W17.3
K02293 1 W17.3
K02115 1 W17.92
K02115 1 W17.92
K02634 1 W17.92
K02634 1 W17.92
K02703 1 W17.92
E3.4.24.64 1 W17.92
E3.2.1.3 1 W17.92
K02636 1 W17.92
K03028 1 W17.92
E2.5.1.47 1 W17.92
K01595 1 W17.92
K01953 1 W17.92
K01507 1 W17.92
K01873 1 W17.92

K00616 1 W17.92
K01834 1 W17.92
K04567 1 W17.92
K07374 1 W17.92
K10206 1 W17.92
K04564 1 W17.92
K02146 1 W17.92
K09565 1 W17.92
K02154 1 W17.92
K02872 1 W17.92
K02136 1 W17.92
K01897 1 W18.5
K01649 1 W18.5
K06959 1 W18.5
K06148 1 W18.5
K00600 1 W18.5
K02984 1 W18.5
E1 1 W18.5
K02937 1 W18.5
K08057 1 W18.5
K02960 1 W18.5
K06215 1 W18.5
K00855 1 W18.5
K00856 1 W18.5
K02932 1 W18.5
K02110 1 W19.04
K02110 1 W19.04
K00031 1 W19.04
K00430 1 W19.04
E2.5.1.22 1 W19.04
K00031 1 W19.04
K01955 1 W19.04
K00819 1 W19.04
K03252 1 W19.04
K02148 1 W19.04
K02908 1 W19.04
K01255 1 W19.04
K00208 1 W19.04
K03545 1 W19.04
K02962 1 W19.04
K00003 1 W19.04
K00325 1 W19.04
K00611 1 W19.04
K02720 1 W19.53
K01602 1 W19.53
K01602 1 W19.53
K00026 1 W19.53
K08769 1 W19.53
K02865 1 W19.53
K01835 1 W19.53
K02966 1 W19.53

E1.17.4.3 1 W19.53
K02901 1 W19.53
K02880 1 W19.53
K01870 1 W19.53
K00283 1 W19.53
K02705 1 W20
K02705 1 W20
K00958 1 W20
K03695 1 W20
K01696 1 W20
K01414 1 W20
K06028 1 W20
K02358 1 W20
K00818 1 W20
K02995 1 W20
K01903 1 W20
K03798 1 W20
K03969 1 W20
K01507 1 W20
K00052 1 W20
K02998 1 W20.44
K00058 1 W20.44
K02941 1 W20.44
K01624 1 W20.44
K09486 1 W20.44
K02031 1 W20.85
K00850 1 W20.85
E3.6.3.14 1 W20.85
E3.1.1.47 1 W20.85
K00134 1 W20.85
K02934 1 W20.85
E1.11.1.11 1 W20.85
K00927 1 W20.85
K00873 1 W20.85
K00012 1 W20.85
K02641 1 W20.85
K02706 1 W21.24
K02706 1 W21.24
K00820 1 W21.24
K01647 1 W21.24
K06972 1 W21.24
K01914 1 W21.24
K09291 1 W21.24
K00134 1 W21.24
K02137 1 W21.24
K07375 1 W21.24
E1 1 W21.61
K03695 1 W21.61
K00362 1 W21.61
K00134 1 W21.61
K02716 1 W21.61

K05236 1 W21.61
K01872 1 W21.61
K01940 1 W21.61
K02689 1 W21.96
K03798 1 W21.96
K04077 1 W21.96
K05862 1 W21.96
K00284 1 W21.96
K02936 1 W21.96
K01595 1 W21.96
K02355 1 W22.3
K05692 1 W22.3
K08912 1 W22.3
K07977 1 W22.3
K01689 1 W22.3
K04077 1 W22.62
K01897 1 W22.62
K02641 1 W22.62
K00615 1 W22.62
K06413 1 W22.62
K03696 1 W22.92
K03696 1 W22.92
K03696 1 W22.92
E1.11.1.11 1 W22.92
K01959 1 W22.92
K09485 1 W22.92
K02115 1 W22.92
K01845 1 W22.92
K03257 1 W22.92
K02930 1 W23.22
K01829 1 W23.22
K02925 1 W23.22
K03405 1 W23.5
K03405 1 W23.5
K03403 1 W23.5
K00265 1 W23.5
K01703 1 W23.5
E3.6.1.3 1 W23.77
K02147 1 W23.77
K00548 1 W23.77
K06630 1 W24.29
K05863 1 W24.29
K04043 1 W24.53
K00053 1 W24.77
K00789 1 W25
K01682 1 W25
K01552 1 W25
K04077 1 W25
K00392 1 W25.22
K05907 1 W25.22
K02145 1 W25.22

K04079 1 W25.22
K00956 1 W25.65
K01624 1 W25.85
K01955 1 W26.05
K02704 1 W26.43
K02704 1 W26.43
K02132 1 W26.61
K03235 1 W27.13
K04043 1 W27.13
K00134 1 W27.46
K03234 1 W27.77
K09490 1 W27.92
E3.2.1.3 1 W28.36
K00615 1 W28.64
K01251 1 W28.64
K01915 1 W29.16
K02358 1 W29.29
E3.6.3.14 1 W29.65
E3.6.3.14 1 W29.65
K02133 1 W29.65
K00927 1 W29.77
K04646 1 W30.11
K01601 1 W30.44
K01601 1 W30.44
K02112 1 W30.95
K03283 1 W32.05
K01961 1 W32.05
*

Fe replete data- copy list below (end at the *) do not change formatting
K02112 1 W31.96
E3.6.3.14 1 W31.05
E3.6.3.14 1 W31.05
K03283 1 W30.85
K01915 1 W30.54
K01601 1 W30.11
K01601 1 W30.11
K01961 1 W30.11
K02133 1 W29.16
K02704 1 W28.64
K02704 1 W28.64
K03235 1 W28.64
K04646 1 W27.77
E3.2.1.3 1 W27.62
K02132 1 W27.62
K02358 1 W27.3
K00615 1 W27.13
K09490 1 W26.61
K01251 1 W26.24
K04043 1 W26.05
K00927 1 W26.05
K01552 1 W25.85
K01955 1 W25.85
K02689 1 W25
K04077 1 W25
K03798 1 W24.53
K00956 1 W24.53
K03234 1 W24.29
K02115 1 W24.29
K05862 1 W24.04
K00134 1 W24.04
K05863 1 W24.04
K02930 1 W23.5
K01682 1 W23.5
K02995 1 W23.5
K02716 1 W23.5
E1 1 W23.22
K05907 1 W23.22
K04079 1 W23.22
K02870 1 W22.92
K02870 1 W22.92
K03696 1 W22.92
K03696 1 W22.92
K03696 1 W22.92
K00053 1 W22.92
K02641 1 W22.92
K03403 1 W22.92
K02934 1 W22.92
K02147 1 W22.92
K02706 1 W22.62

K02706 1 W22.62
K04043 1 W22.62
K02925 1 W22.62
K02962 1 W22.62
K02705 1 W22.3
K02705 1 W22.3
K00392 1 W22.3
K01959 1 W22.3
K02115 1 W21.96
K02115 1 W21.96
K02634 1 W21.96
K02634 1 W21.96
K01647 1 W21.96
K00362 1 W21.96
K02145 1 W21.96
K01829 1 W21.61
E3.6.3.14 1 W21.24
K01897 1 W21.24
K02941 1 W21.24
K00265 1 W21.24
K01624 1 W21.24
K02720 1 W20.85
K02703 1 W20.85
E3.6.1.3 1 W20.85
K02936 1 W20.85
K00134 1 W20.85
K02137 1 W20.85
K03798 1 W20.85
K01689 1 W20.85
K02641 1 W20.85
K02690 1 W20.44
K02690 1 W20.44
K04077 1 W20.44
K02981 1 W20.44
K02355 1 W20.44
E3.1.1.47 1 W20.44
K00134 1 W20.44
K02901 1 W20.44
K01870 1 W20.44
K01703 1 W20.44
K01845 1 W20.44
K00548 1 W20.44
K00789 1 W20
K02998 1 W20
K02031 1 W20
K00850 1 W20
K04077 1 W20
K02636 1 W20
K01595 1 W20
K02358 1 W20
K01595 1 W20

K02993 1 W20
K01529 1 W20
K00820 1 W19.53
K02945 1 W19.53
K08912 1 W19.53
K03252 1 W19.53
K03405 1 W19.04
K03405 1 W19.04
K02932 1 W19.04
E1.3.99.23 1 W19.04
E3.2.1.3 1 W19.04
K00360 1 W19.04
K05692 1 W19.04
K00284 1 W19.04
K02987 1 W19.04
K02966 1 W19.04
K02880 1 W19.04
K00101 1 W19.04
K01507 1 W19.04
K00325 1 W19.04
K02110 1 W18.5
K02110 1 W18.5
K02930 1 W18.5
K02886 1 W18.5
K01529 1 W18.5
K02865 1 W18.5
K00058 1 W18.5
K02937 1 W18.5
K00602 1 W18.5
K00283 1 W18.5
K02136 1 W18.5
K02932 1 W18.5
K03257 1 W18.5
K01602 1 W17.92
K01602 1 W17.92
K02692 1 W17.92
K02692 1 W17.92
K02989 1 W17.92
K02134 1 W17.92
K08769 1 W17.92
K00430 1 W17.92
E3.6.-.- 1 W17.92
K01914 1 W17.92
K02984 1 W17.92
K02969 1 W17.92
E1.11.1.11 1 W17.92
K06180 1 W17.92
K02866 1 W17.92
E1 1 W17.92
K02875 1 W17.92
K00820 1 W17.92

K02154 1 W17.92
K02872 1 W17.92
K03046 1 W17.3
K03046 1 W17.3
K02991 1 W17.3
K01897 1 W17.3
K00026 1 W17.3
E3.6.4.9 1 W17.3
K03255 1 W17.3
K01955 1 W17.3
K02908 1 W17.3
K01903 1 W17.3
K09485 1 W17.3
K06413 1 W17.3
K02882 1 W17.3
K01872 1 W17.3
K01939 1 W17.3
K06215 1 W17.3
K02109 1 W16.61
K02109 1 W16.61
K02981 1 W16.61
K02981 1 W16.61
K02925 1 W16.61
K02993 1 W16.61
K00031 1 W16.61
K01537 1 W16.61
K06959 1 W16.61
E6.5.1.1 1 W16.61
K06148 1 W16.61
K02949 1 W16.61
K06972 1 W16.61
E3.4.24.64 1 W16.61
K02894 1 W16.61
K06630 1 W16.61
K03695 1 W16.61
K06028 1 W16.61
K02964 1 W16.61
K02974 1 W16.61
K02991 1 W16.61
K01834 1 W16.61
K00819 1 W16.61
K02985 1 W16.61
K04078 1 W16.61
K00831 1 W16.61
K03253 1 W16.61
K00514 1 W16.61
K02146 1 W16.61
K07977 1 W16.61
K02960 1 W16.61
K01940 1 W16.61
K00611 1 W16.61

K02693 1 W15.85
K02868 1 W15.85
K02868 1 W15.85
K03043 1 W15.85
K03043 1 W15.85
K02635 1 W15.85
K02635 1 W15.85
K02479 1 W15.85
K00031 1 W15.85
K01750 1 W15.85
K02940 1 W15.85
K03254 1 W15.85
K02953 1 W15.85
K02040 1 W15.85
K02885 1 W15.85
K01733 1 W15.85
K02519 1 W15.85
K02150 1 W15.85
E1.17.4.3 1 W15.85
K02989 1 W15.85
K03405 1 W15.85
K01624 1 W15.85
K00927 1 W15.85
K00003 1 W15.85
K09486 1 W15.85
K00234 1 W15.85
K00052 1 W15.85
K02891 1 W15.85
K02985 1 W15
K01441 1 W15
K01961 1 W15
K02956 1 W15
E2.7.1.150 1 W15
K09291 1 W15
K03797 1 W15
K00266 1 W15
K09540 1 W15
K00955 1 W15
K00600 1 W15
K02910 1 W15
K02357 1 W15
K01516 1 W15
K06212 1 W15
K08057 1 W15
E1.11.1.5 1 W15
K00382 1 W15
K09503 1 W15
E5.3.4.1 1 W15
K02900 1 W15
K00873 1 W15
K02883 1 W15

E3.6.1.- 1 W15
K02893 1 W15
K00856 1 W15
E1.1.1.100 1 W15
K02964 1 W14.04
K02964 1 W14.04
K02114 1 W14.04
K02114 1 W14.04
K03070 1 W14.04
E3.6.1.3 1 W14.04
K00958 1 W14.04
K01649 1 W14.04
E1 1 W14.04
K02873 1 W14.04
E3.2.1.3 1 W14.04
E3.6.4.1 1 W14.04
K00908 1 W14.04
E1.8.1.4 1 W14.04
K02912 1 W14.04
K03428 1 W14.04
K01835 1 W14.04
K00058 1 W14.04
K06118 1 W14.04
K00818 1 W14.04
K01687 1 W14.04
K02973 1 W14.04
E1.3.1.26 1 W14.04
K01255 1 W14.04
K01428 1 W14.04
E3.6.1.3 1 W14.04
K02955 1 W14.04
K02876 1 W14.04
K07375 1 W14.04
K08902 1 W14.04
K06207 1 W14.04
E1.2.1 1 W14.04
K02877 1 W14.04
K06185 1 W14.04
K08738 1 W14.04
K00627 1 W14.04
K01529 1 W14.04
K02112 1 W12.92
K02112 1 W12.92
K03046 1 W12.92
K02707 1 W12.92
K02707 1 W12.92
K02694 1 W12.92
K02694 1 W12.92
K02949 1 W12.92
K03237 1 W12.92
E3.2.1.18 1 W12.92

K10251 1 W12.92
E2.3.1.48 1 W12.92
E1.9.3.1 1 W12.92
K01696 1 W12.92
K02868 1 W12.92
K09838 1 W12.92
E1.10.2.2 1 W12.92
K00963 1 W12.92
K03028 1 W12.92
K02957 1 W12.92
K02938 1 W12.92
K01414 1 W12.92
K00615 1 W12.92
K00627 1 W12.92
K00161 1 W12.92
E1 1 W12.92
K00295 1 W12.92
K05679 1 W12.92
K00134 1 W12.92
E3.6.5.3 1 W12.92
K02889 1 W12.92
K01802 1 W12.92
K03942 1 W12.92
K02729 1 W12.92
K05808 1 W12.92
K10206 1 W12.92
K01738 1 W12.92
K04564 1 W12.92
K00344 1 W12.92
K03545 1 W12.92
E2.1.1.95 1 W12.92
E3.4.24 1 W12.92
E1.11.1.11 1 W12.92
K02870 1 W12.92
K01529 1 W12.92
K01711 1 W12.92
K02977 1 W12.92
K07151 1 W12.92
E6.1.1.9 1 W12.92
E5.3.4.1 1 W12.92
E2.1.1 1 W12.92
E3.4.21 1 W12.92
K01749 1 W12.92
E1.14.19.3 1 W12.92
K05665 1 W12.92
K02125 1 W11.61
K03936 1 W11.61
E1.5.1.2 1 W11.61
E1.1.1.18 1 W11.61
E2.7.7 1 W11.61
E2.1.1 1 W11.61

E2.5.1.22 1 W11.61
E2.7.11.1 1 W11.61
K02155 1 W11.61
E2.7.11.22 1 W11.61
K07976 1 W11.61
K03076 1 W11.61
E1.3.99.23 1 W11.61
K08908 1 W11.61
K00948 1 W11.61
E1.4.1.4 1 W11.61
K02898 1 W11.61
K09272 1 W11.61
K01201 1 W11.61
E3.6.5.3 1 W11.61
K03686 1 W11.61
E3.1.4.43 1 W11.61
E1.3.1.- 1 W11.61
K01507 1 W11.61
K02864 1 W11.61
K01626 1 W11.61
E2.7.11.22 1 W11.61
K00872 1 W11.61
E1 1 W11.61
K02917 1 W11.61
K02983 1 W11.61
K00939 1 W11.61
K03066 1 W11.61
K00235 1 W11.61
K07976 1 W11.61
K00324 1 W11.61
K08910 1 W11.61
K05236 1 W11.61
K08906 1 W11.61
K03062 1 W11.61
K03626 1 W11.61
K04564 1 W11.61
K01803 1 W11.61
K00012 1 W11.61
K02975 1 W11.61
K00873 1 W11.61
K03065 1 W11.61
K01610 1 W11.61
K01698 1 W11.61
K01662 1 W11.61
K02293 1 W11.61
E3.6.5.3 1 W11.61
K01599 1 W11.61
K03242 1 W11.61
K00162 1 W11.61
E3.1 1 W11.61
E3.2.1.18 1 W11.61

K02709 1 W10
K02709 1 W10
K02699 1 W10
K02699 1 W10
K02997 1 W10
K00930 1 W10
K02903 1 W10
E3.2.1.18 1 W10
K01586 1 W10
K01624 1 W10
E3.2.1.3 1 W10
E6.3.2.19 1 W10
K03695 1 W10
K01679 1 W10
K02736 1 W10
K04499 1 W10
E3.1.-.- 1 W10
K01836 1 W10
K01599 1 W10
K03100 1 W10
K02731 1 W10
K02732 1 W10
K02727 1 W10
K01953 1 W10
E3.6.1.3 1 W10
K00164 1 W10
K00297 1 W10
K00088 1 W10
K01834 1 W10
K02148 1 W10
K01276 1 W10
K07374 1 W10
K01952 1 W10
K02726 1 W10
K00940 1 W10
E2.7.1.40 1 W10
K03969 1 W10
K02896 1 W10
K03687 1 W10
K00326 1 W10
K02641 1 W10
K01637 1 W10
K01529 1 W10
K01358 1 W10
K03238 1 W10
K01803 1 W10
K02879 1 W10
K03455 1 W10
K02149 1 W10
K07936 1 W10
K01689 1 W10

K03038 1 W10
K08956 1 W10
K10046 1 W10
K02918 1 W10
K00855 1 W10
K05805 1 W10
K01358 1 W10
K01758 1 W10
K04488 1 W10
K00945 1 W10
K02013 1 W10
E3.1.3.2 1 W10
E1.14.11 1 W10
K02723 1 W7.92
K02723 1 W7.92
K02891 1 W7.92
K02891 1 W7.92
K02256 1 W7.92
K02261 1 W7.92
E2.7.1.- 1 W7.92
K01441 1 W7.92
E2.3.1.48 1 W7.92
K02730 1 W7.92
E2.4.1.34 1 W7.92
E5.3.4.1 1 W7.92
K03531 1 W7.92
K01829 1 W7.92
K02947 1 W7.92
E2.3.1.12 1 W7.92
K00329 1 W7.92
E1 1 W7.92
K02971 1 W7.92
E1 1 W7.92
E2 1 W7.92
K01868 1 W7.92
K07976 1 W7.92
K01175 1 W7.92
K03937 1 W7.92
E3.6.5.3 1 W7.92
E2.7.11.22 1 W7.92
K09496 1 W7.92
K03940 1 W7.92
E3.1.3.16 1 W7.92
K00820 1 W7.92
E1.11.1.7 1 W7.92
E3.1.13.4 1 W7.92
E3.4.21.4 1 W7.92
E1.3.1.33 1 W7.92
K01915 1 W7.92
K03031 1 W7.92
K07976 1 W7.92

K04040 1 W7.92
K00226 1 W7.92
K00698 1 W7.92
E1.14.-.- 1 W7.92
E1.11.1.15 1 W7.92
E1.1.1.31 1 W7.92
K07517 1 W7.92
K03696 1 W7.92
K02997 1 W7.92
K09838 1 W7.92
K00813 1 W7.92
K03064 1 W7.92
K01254 1 W7.92
K00799 1 W7.92
K01873 1 W7.92
K00698 1 W7.92
K08869 1 W7.92
K01755 1 W7.92
K03264 1 W7.92
E1 1 W7.92
K00059 1 W7.92
K03767 1 W7.92
K02958 1 W7.92
K03100 1 W7.92
K00942 1 W7.92
K01802 1 W7.92
K00134 1 W7.92
K00559 1 W7.92
K01858 1 W7.92
K01358 1 W7.92
K09493 1 W7.92
K09560 1 W7.92
K02951 1 W7.92
K02151 1 W7.92
K01810 1 W7.92
K01625 1 W7.92
K02979 1 W7.92
K01880 1 W7.92
K07977 1 W7.92
E3.6.5.3 1 W7.92
K01090 1 W7.92
K03263 1 W7.92
K01951 1 W7.92
K05665 1 W7.92
K01887 1 W7.92
K01885 1 W7.92
E2.7.7 1 W7.92
E3.6.1.- 1 W7.92
E3.6.1.3 1 W7.92
K06942 1 W7.92
K00592 1 W7.92

K03754 1 W7.92
K07755 1 W7.92
K02734 1 W7.92
K01054 1 W7.92
K01325 1 W7.92
K00939 1 W7.92
E2.1.1 1 W7.92
K03241 1 W7.92
K02921 1 W7.92
K02728 1 W7.92
E3.4.24.- 1 W7.92
E5.2.1.8 1 W7.92
K00232 1 W7.92
K02357 1 W7.92
K03627 1 W7.92
K06158 1 W7.92
E2.1.1.127 1 W7.92
E2.3.1.74 1 W7.92
K10413 1 W7.92
K02999 1 W7.92
K00111 1 W7.92
K02469 1 W7.92
K02541 1 W7.92
K02040 1 W5
K03935 1 W5
K02837 1 W5
K00022 1 W5
K00940 1 W5
E2.5.1.47 1 W5
K01262 1 W5
K09499 1 W5
E4.1.1.29 1 W5
E2.1.1.127 1 W5
K04774 1 W5
K06889 1 W5
E3.6.3.16 1 W5
K03839 1 W5
K02976 1 W5
K02943 1 W5
E6.3.2.- 1 W5
K01469 1 W5
K01423 1 W5
E1 1 W5
K01840 1 W5
K00873 1 W5
E1 1 W5
E3.2.1.18 1 W5
E2.3.1.48 1 W5
E3.6.3.14 1 W5
K01165 1 W5
K03032 1 W5

E6.3.2.- 1 W5
K03063 1 W5
K06980 1 W5
E2.3.1.48 1 W5
K00927 1 W5
K09498 1 W5
E3.2.1.18 1 W5
E3.6.5.3 1 W5
E5.4.2.8 1 W5
K01931 1 W5
K03943 1 W5
K04518 1 W5
K04802 1 W5
K01897 1 W5
E1.11.1.7 1 W5
E2.7.1.137 1 W5
E1.1.1.- 1 W5
K00387 1 W5
K03061 1 W5
K03953 1 W5
K01256 1 W5
K01086 1 W5
E3.6.3.14 1 W5
K01933 1 W5
K03246 1 W5
E1.3.1.33 1 W5
K06063 1 W5
K06445 1 W5
K03531 1 W5
E3.6.5.3 1 W5
K03676 1 W5
E3.6.1.3 1 W5
K00620 1 W5
E1.14.11 1 W5
K03076 1 W5
K00698 1 W5
K00003 1 W5
K02978 1 W5
K01069 1 W5
E1.10.2.2 1 W5
E1.10.2.2 1 W5
K03977 1 W5
K00133 1 W5
E3.6.1.3 1 W5
K01890 1 W5
K07304 1 W5
K00033 1 W5
K03955 1 W5
K03030 1 W5
K01802 1 W5
K03037 1 W5

E3.4.11.18 1 W5
K01895 1 W5
E3.1.2.16 1 W5
E5.2.1.8 1 W5
K01652 1 W5
K03257 1 W5
K01476 1 W5
K00632 1 W5
K03627 1 W5
K03033 1 W5
K07942 1 W5
E1.2.1.3 1 W5
K03249 1 W5
K00301 1 W5
E1 1 W5
K01885 1 W5
K09488 1 W5
K03262 1 W5
K00789 1 W5
K06067 1 W5
K01710 1 W5
K01899 1 W5
K00814 1 W5
K03671 1 W5
E3.2.1.3 1 W5
K00927 1 W5
K00413 1 W5
E1.3.1.33 1 W5
K01855 1 W5
E3.6.5.3 1 W5
K01362 1 W5
K08023 1 W5
K07976 1 W5
E3.2.1.18 1 W5
K02959 1 W5
K02924 1 W5
K03177 1 W5
K03439 1 W5
E1.14.11 1 W5
K01187 1 W5
K03097 1 W5
E5.2.1.8 1 W5
E3.6.1.3 1 W5
K08869 1 W5
K00624 1 W5
K10356 1 W5
K06119 1 W5
K06948 1 W5
K01803 1 W5
K06067 1 W5
K03798 1 W5

K01872 1 W5
E5.2.1.8 1 W5
E6.3.2.- 1 W5
E2.7.1.37 1 W5
E4 1 W5
K03879 1 W5
K01611 1 W5
E3.1.13.- 1 W5
K01875 1 W5
K06924 1 W5
E5.2.1.8 1 W5
K10226 1 W5
K08869 1 W5
K03797 1 W5
K10256 1 W5
E2.7.7.6 1 W5
*

Bibliography

Adler, R.E., Polyak, L., Ortiz, J.D., Kaufman, D.S., Channell, J.E.T., Xuan, C., Grottoli, A.G., Sellén, E., Crawford, K.A. 2009. Sediment record from the western Arctic Ocean with an improved Late Quaternary age resolution: HOTRAX core HLY0503-8JPC, Mendeleev Ridge. Global and Planetary Change, 68 (1-2), 18-29.

Aebersold R. and Goodlett, D.R. 2001. Mass spectrometry in proteomics. Chemical Reviews, 101(2), 269-295.

Aebersold, R. and Mann, M. 2003. Mass spectrometry-based proteomics. Nature, 422: 198-207.

Allen, A.E., Dupont, C.L., Obornik, M., Horak, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H.H., Fernie, A.R. Bowler, C. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature, 473: 203-207.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam ,N.H., Zhou, S.G., Allen, A.E., Apt, K.E., Bechner, M., Brzezinski, M.A., Chaal, B.K., Chiovitti, A., Davis, A.K., Demarest, M.S., Detter, J.C., Glavina, T., Goodstein, D., Hadi, M.Z., Hellsten, U., Hildebrand, M., Jenkins, B.D., Jurka, J., Kapitonov, V.V., Kroger, N., Lau, W.W.Y., Lane, T.W., Larimer, F.W., Lippmeier, J.C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M.S., Palenik, B., Pazour, G.J., Richardson, P.M., Rynearson, T.A., Saito, M.A., Schwartz, D.C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F.P., Rokhsar, D.S. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science, 306(5693), 79-86.

Armengaud, J. 2013. Microbiology and proteomics, getting the best of both worlds! Environmental Microbiology, 15(1), 12-23.

Arrigo, K.R., DiTullio, G.R., Dunbar, R.B., Robinson, D.H., VanWoert, M., Worthen, D.L., Lizotte, M.P. 2000. Phytoplankton taxonomic variability and nutrient utilization and primary production in the Ross Sea. Journal of Geophysical Research, 105: 8827-8846.

Barton, L. 2005. Structural and Functional Relationships in Prokaryotes. New York, NY: Springer Science and Business Media, Inc.

Bastida, F., Moreno, J.L., Nicolas, C., Hernandez, T., Carcia, C. 2009. Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. European Journal of Soil Science, 60, 845-859.

Behrenfeld, M.J. and Kolber, Z.S. 1999. Widespread iron limitation of phytoplankton in the south Pacific Ocean. Science, 283: 840-843.

Behrenfeld, M.J. and Milligan, A.J. 2013. Photophysiological expressions of iron stress in phytoplankton. Annual Review of Marine Science, 5: 217-246.

Belicka, L.L. and Harvey, H.R. 2009. The sequestration of terrestrial organic carbon in Arctic Ocean sediments: A comparison of methods and implications for regional carbon budgets. Geochimica et Cosmochimica Acta, 73, 6231-6248.

Belicka, L.L., Macdonald, R.W., Harvey, H.R. 2002. Sources and transport of organic carbon to shelf, slope, and basin surface sediments of the Arctic Ocean. Deep-Sea Research I, 49, 1463-1483.

Belicka, L.L., Macdonald, R.W., Harvey, H.R. 2009. Trace element and molecular markers of organic carbon dynamics along a shelf-basin continuum in sediments of the western Arctic Ocean. Marine Chemistry, 115, 72-85.

Belicka, L.L., Macdonald, R.W., Yunker, M.B., Harvey, H.R. 2004. The role of depositional regime on carbon transport and preservation in Arctic Ocean sediments. Marine Chemistry, 86, 65-88.

Belluomini G., Branca, M., Calderoni, G., Schnitzer, M. 1986. Distribution and geochemical significance of amino-acids and amino-sugars in a clay suite of the Pliocene Pleistocene age from central Italy. Organic Geochemistry, 9(3), 127-133.

Berg, J.M., Tymoczko, J.L., Stryer, L. 2002. Biochemistry. New York: W.H. Freeman.

Bianchi, A. and Giuliano, L. 1996. Enumeration of viable bacteria in the marine pelagic environment. Applied and Environmental Microbiology, 62(1), 174-177.

Birgel, D., Stein, R., Hefter, J. 2004. Aliphatic lipids in recent sediments of the Fram Strait/Yermak Plateau (Arctic Ocean): composition, sources and transport processes. Marine Chemistry, 88, 127-160.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z.P., Snyder, M., Dermitzakis, E.T., et. al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 447: 799-816.

Botelho, D., Wall, M. J., Vieira, D. B., Fitzsimmons, S., Liu, F., Doucette, A. 2010. Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. Journal of Proteome Research, 9(6), 2863-2870.

Boyd, P.W., Muggli, D.L., Varela, D.E., Goldblatt, R.H., Chretien, R., Orians, K.J., Harrison, P.J. 1996. In vitro enrichment experiments in the NE subarctic Pacific. Marine Ecology Press Series, 136: 179-188.

Boyd, P.W., Watson, A.J., Law, C.S., Abraham, E.R., Trull, T., Murdoch, R., Bakker, D.C.E., Bowie, A.R., Buesseler, K.O., Chang, H., Charette, M., Croot, P., Downing, K., Frew, R., Gall, M., Hadfield, M., Hall, J., Harvey, M., Jameson, G., LaRoche, J., Liddicoat, M., Ling, R., Maldonado, M.T., McKay, R.M., Nodder, S., Pickmere, S., Pridmore, R., Rintoul, S., Safi, K., Sutton, P., Strzepek, R., Tanneberger, K., Turner, S., Waite, A., Zeldis, J. 2000. A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. Nature, 407: 695-702.

Boyd, P.W. and Hutchins, D.A. 2012. Understanding the responses of ocean biota to a complex matrix of cumulative anthropogenic change. Marine Ecology Progress Series, 470: 125-135.

Boyd, P.W., Strzepek, R., Takeda, S., Jackson, G., Wong, C.S., McKay, R.M., Law, C., Kiyosawa, H., Saito, H., Sherry, N., Johnson, K., Gower, J., Ramaiah, N. 2005. The evolution and termination of an iron-induced mesoscale bloom in the northeast subarctic Pacific. Limnology and Oceanography, 50: 1872-1886.

Burdige, D.J. and Martens, C.S. 1988. Biogeochemical cycling in an organic-rich coastal marine basin: 10. The role of amino acids in sedimentary carbon and nitrogen cycling. Geochimica et Cosmochimica Acta, 52(6), 1571-1584.

Brown, M.R. 1991. The amino acids and sugar composition of 16 species of microalgae used in mariculture. Journal of Experimental Marine Biology and Ecology, 145, 79-99.

Carvalho, R.N. and Lettieri, T. 2011 Proteomic analysis of the marine diatom *Thalassiosira pseudonana* upon exposure to benzo(a)pyrene. BMC Genomics, 12: 159.

Chao, T.C. and Hansmeier, N. 2012. The current state of microbial proteomics: Where we are and where we want to go. Proteomics, 12, 638-650.

Chen, J., Ryu, S., Gharib, S.A., Goodlett, D.R., Schnapp, L.M. 2008. Exploration of the normal human bronchioalveolar lavage fluid proteome. Proteomics Clinical Applications, 2(4), 585-595.

Chen, M., Huang, Y.P., Cai, P.G., Guo, L.D. 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., Shufeldt, R.C., Stevenson, F.J. 1975. Amino acid analysis of soils and sediments: extraction and desalting. Soil Biology and Biochemistry, 7(2), 143-151.

Chin-Leo, G. and Kirchman, D.L. 1988. Estimating bacterial production in marine waters from simultaneous incorporation of thymidine and leucine. Applied and Environmental Microbiology, 54(8), 1934-1939.

Choi, H., Fermin, D., Nesvizhskii, A.I. 2008. Significance analysis of spectral count data in label-free shotgun proteomics. Molecular Cellular Proteomics, 7: 2373–2385.

Christie-Oleza, J.A. and Armengaud, J. 2010. In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. Marine Drugs, 8, 2223-2239.

Christie-Oleza, J.A., Fernandez, B., Nogales, B., Bosch, R., Armengaud, J. 2012. Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. The ISME Journal, 6(1), 124-135.

Christie-Oleza, J.A., Miotello, G., Armengaud, J. 2012. High-throughput proteogenomics of *Ruegeria pomeroyi*: seeding a better genomic annotation for the whole marine *Roseobacter* clade.

Christie-Oleza, J.A., Pina-Villalonga, J.M., Bosch, R., Nogales, B., Armengaud, J. 2012. Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria. 2012. Molecular & Cellular Proteomics, 11(2), M111.013110.

Coale, K.H., Wang, X., Tanner, S.J., Johnson, K.S. 2003. Phytoplankton growth and biological response to iron and zinc addition in the Ross Sea and Antarctic Circumpolar Current along 170[deg]W. Deep Sea Research Part II, 50: 635-653.

Cole, J.J., Findlay, S. Pace, M.L. 1988. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Marine Ecology Progress Series, 43, 1-10.

Collins, M.J., Bishop, A.N., Farrimond, P. 1995. Sorption by mineral surfaces: Rebirth of the classical condensation pathway for kerogen formation? Geochimica et Cosmochimica Acta, 59(11), 2387-2391.

Conte, M.H., Dickey, T.D., Weber, J.C., Johnson, J.C., Knap, A.H. 2003. Transient physical forcing of pulsed export of bioreactive material to the deep Sargasso Sea. Deep-Sea Research I, 50, 1157-1187.

Cowie, G.L. and Hedges J.I. 1992a. Sources and Reactivities of Amino Acids in a Coastal Marine Environment. Limnology and Oceanography, 37, 703-724.

Cowie, G.L. and Hedges, J.I. 1992b. Improved amino acid quantification in environmental-samples - charge-matched recovery standards and reduced analysis time. Marine Chemistry, 37(3-4), 223-238.

Craig, O.E. and Collins, M.J. 2000. An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. Journal of Immunological Methods, 236, 89-97.

Darby, D.A., Ortiz, J., Polyak, L., Lund, S., Jakobsson, M., Woodgate, R.A. 2009. The role of currents and sea ice in both slowly deposited central Arctic and rapidly deposited Chukchi-Alaskan margin sediments. Global and Planetary Change, 68, 58-72.

Darby, D.A., Polyak, L., Jakobsson, M. 2009. The 2005 HOTRAX expedition to the Arctic Ocean. Global and Planetary Change, 68 (1-2), 1-4.

Darby, D.A., Polyak, L., Bauch, H.A. 2006. Past glacial and interglacial conditions in the Arctic Ocean and marginal seas – a review. Progress in Oceanography, 71, 129-144.

Davis, J. A. 1982. Adsorption of natural dissolved organic matter at the oxide/water interface. Geochimica et Cosmochimica Acta, 46, 2381-2393.

Davis, M.T., Spahr, C.S., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., Mort, J., Yu, W., Luethy, R., Patterson, S.D. 2001. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry - II. Limitations of complex mixture analyses. Proteomics, 1(1), 108-117.

Deming, J.W. and Barross, J.A. 1993. The early diagenesis of organic matter: bacterial activity. In: M.H. Engel and S.A. Macko (Editors), Organic geochemistry: principles and applications. Plenum Press, New York, NY, 119-144.

Dong, H.P., Wang, D.Z., Dai, M., Hong, H.S. 2010. Characterization of particulate organic matters in the water column of the South China Sea using a shotgun proteomic approach. Limnology and Oceanography, 55(4), 1565-1578.

Dong, H.P., Wang, D.Z., Xie, Z.X., Dai, M.H., Hong, H.S. 2013. Metaproteomic characterization of high molecular weight dissolved organic matter in surface waters in the South China Sea. Geochimica et Cosmochimica Acta, 109, 51-61.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E.V., Le Gall, F., Makarova, K.S., Ostrowski, M., Oztas, S., Robert, C., Rogozin, I.B., Scanlan, D.J., Tandeau de Marsac, N., Weissenbach, J., Wincker, P., Wolf, Y.I., Hess, W.R. 2003. Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly

minimal oxyphototrophic genome. Proceedings of the National Academy for Science USA, 100(17), 10020-10025.

Dunning, J.D., Herren, B.J., Tipps, R.W., Snyder, R.S. 1982. Fractionation of mineral species by electrophoresis. Journal of Geophysical Research, 87, 781-788.

Durkin, C.A., Marchetti, A., Bender, S.J., Truong, T., Morales, R., Armbrust, E.V. 2011. Diversity in iron-limited diatoms has varied effects on silicon cycling. Journal of Phycology, 47: S28-S28.

Durkin, C., Marchetti, A., Bender, S., Truong, T., Morales, R., Mock, T., Armbrust, E.V. 2012. Frustule-related gene transcription and the influence of diatom community composition on the silica precipitation in an iron-limited environment. Limnology and Oceanography, 57: 1619-1633.

Dyhrman, S.T., Jenkins, B.D., Rynearson, T.A., Saito, M.A., Mercier, M.L., Alexander, H., Whitney, L.P., Drzewianowski, A., Bulygin, V.V., Bertrand, E.M. 2012. The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. PLoS ONE, 7: e33768.

Eicken, H., Gradinger, R., Gaylord, A., Mahoney, A., Rigor, I., Melling, H. 2005. Sediment transport by sea ice in the Chukchi and Beaufort Seas: Increasing importance due to changing ice conditions? Deep-Sea Research II, 52, 3281-3302.

Egleston, E.S., Morel, F.M.M. 2008. Nickel limitation and zinc toxicity in a urea-grown diatom. Limnology and Oceanography, 53: 2462-2471.

Elias, J.E. and Gygi, S.P. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods, 4: 207-214.

Eng, J.K., McCormack, A.L., Yates, J.R. 1994. An approach to correlate tandem mass spectral data of peptides with amino-acid-sequences in a protein database. Journal of the American Society for Mass Spectrometry, 5, 976-989.

Eng, J.K., Fischer, B., Grossman, J., MacCoss, M.J. 2008. A fast SEQUEST cross correlation algorithm. Journal of Proteome Reasearch, 7(10), 4598-4602.

Espen, L., Dell'Orto, M., De Nisi, P., Zocchi, G. 2000. Metabolic responses in cucumber (Cucumis sativus L.) roots under Fe-deficiency: a ^{31}P-nuclear magnetic resonance in-vivo study. Planta, 210: 985-992.

Fahl, K. and Stein, R. 1997. Modern organic carbon deposition in the Laptev Sea and the adjacent continental slope: surface water productivity vs. terrigenous input. Organic Geochemistry, 26, No. 5/6, 379-390.

Fahl, K. and Stein, R. 1999. Biomarkers as organic-carbon-source and environmental indicators in the Late Quaternary Arctic Ocean: problems and perspectives. Marine Chemistry, 63, 293-309.

Fairbanks, G., Levinthal, C., Reeder, R.H. 1965. Analysis of C14-labeled proteins by disc electrophoresis. Biochemical and Biophysical Research Communications, 20, 393-399.

Fairbanks, G., Steck, T.L., Wallach, D.F.H. 1971. Electrophoretic analysis of the major proteins of the human erythrocyte membrane. Biochemistry, 10, 2606-2617.

Fernie, A.R. and Stitt, M. 2012. On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions scientific correspondence. Plant Physiology, 158: 1139-1145.

Foss, E.J., Radulovic, D., Shaffer, S.A., Goodlett, D.R., Kruglyak, L., Bedalov, A. 2011. Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. PLoS Biology, 9: e1001144.

"Functional Annotation Tool." DAVID Bioinformatics Resources 6.7. National Institute of Allergy and Infectious Diseases (NIAID). NIH, http://david.abcc.ncifcrf.gov/content.jsp?file=functional_annotation.html. 23 Feb. 2014.

Gervais, F., Riebesell, U., Gorbunov, M.Y. 2002. Changes in primary productivity and chlorophyll a in response to iron fertilization in the Southern Polar Frontal Zone. Limnology and Oceanography, 47: 1324–1335.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M., Rappe, M.S., Short, J.M., Carrington, J.C., and Mathur, E.J. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science, 309, 1242-1245.

Goericke, R. and Fry, B. 1994. Variations of marine plankton $\delta^{13}C$ with latitude, temperature, and dissolved CO_2 in the world ocean. Global Biogeochemical Cycles, 8, 85-90.

Goñi, M.A., Yunker, M.B., Macdonald, R.W., Eglinton, T.I. 2000. Distribution and sources of organic biomarkers in arctic sediments from the Mackenzie River and Beaufort Shelf. Marine Chemistry, 71, 23-51.

Goñi, M.A., Yunker, M.B., Macdonald, R.W., Eglinton, T.I. 2005. The supply and preservation of ancient and modern components of organic carbon in the Canadian Beaufort Shelf of the Arctic Ocean. Marine Chemistry, 93, 53-73.

Gradinger, R. 2009. Sea-ice algae: Major contributors to primary production and algal biomass in the Chukchi and Beaufort Seas during May/June 2002. Deep-Sea Research II, 56, 1201-1212.

Graves, P.R., Haystead, T.A.J. 2002. Molecular biologist's guide to proteomics. Microbiology & Molecular Biology Reviews, 66(1), 39-63.

Green, R.T., Todd, J.D., Johnston, A.W.B. 2013. Manganese uptake in marine bacteria; the novel MntX transporter is widespread in *Roseobacters*, *Vibrios*, *Alteromonadates* and the SAR11 and SAR116 clades. The ISME Journal, 7, 581-591.

Greene, R.M., Geider, R.J., Falkowski, P.G. 1991. Effect of Iron Limitation on Photosynthesis in a Marine Diatom. Limnology and Oceanography, 36: 1772-1782.

Greenland, D. J. 1971 Interactions between humic and fulvic acids and clays. Soil Science, 111, 34-41.

Grutters, M., van Raaphorst, W., Helder, W. 2001. Total hydrolysable amino acid mineralisation in sediments across the northeastern Atlantic continental slope (Goban Spur). Deep-Sea Research Part I, 48(3), 811-832.

Guo, L.D., Tanaka, T., Wang, D., Tanaka, N., Murata, A. 2004. Distributions, speciation and stable isotope composition of organic matter in the southeastern Bering Sea. Marine Chemistry, 91(1-4), 211-226.

Hassler, C.S., Havens, S.M., Bullerjahn, G.S., McKay, R.M.L., Twiss, M.R. 2009. An evaluation of iron bioavailability and speciation in western Lake Superior with the use of combined physical, chemical, and biological assessment. Limnology and Oceanography, 54: 987-1001.

Hassler, C.S., Schoemann, V., Nichols, C.M., Butler, E.C.V., Boyd, P.W. 2011. Saccharides enhance iron bioavailability to Southern Ocean phytoplankton. Proceedings of the National Academy of Sciences of the United States of America, 108: 1076-1081.

Henrichs, S.M. and Sugai, S.F. 1993. Adsorption of amino-acids and glucose by sediments of Resurrection Bay, Alaska, USA - functional-group effects. Geochimica et Cosmochimica Acta, 57(4), 823-835.

Hettich, R.L., Pan, C., Chourey, K., Giannone, R.J. 2013. Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. Analytical Chemistry, 85, 4203-4214.

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

Hobson, K.A., Fisk, A., Karnovsky, N., Holst, M., Gagnon, J.M., Fortier, M. 2002. A stable isotope ($\delta^{13}C$, $\delta^{15}N$) model for the North Water food web: implications for evaluating trophodynamics and the flow of energy and contaminants. Deep Sea Research Part II, 49(22-23), 5131-5150.

Hollibaugh, J.T. and Azam, F. 1983. Microbial degradation of dissolved proteins in seawater. Limnology and Oceanography, 28(6), 1104-1116.

Horsfall, I.M. and Wolff, G.A. 1997. Hydrolysable amino acids in sediments from the Porcupine Abyssal Plain, northeast Atlantic Ocean. Organic Geochemistry, 26(5-6), 311-320.

Hopmans, E.C., Weijers, J.W.H., Schefuβ, E., Herfort, L., Sinninghe Damsté, J.S., Schouten, S. 2004. A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. Earth and Planetary Science Letters, 224, 107-116.

Huang, D.W., Sherman, B.T., Lempicki, R.A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44-57.

Hunt, D.F., Yates III, J.R., Shabanowitz, J., Winston, S., Hauer, C.R. 1986. Protein sequencing by tandem mass spectrometry. Proc. Nat. Acad. Sci., 83, 6233-6237.

Iken, K., Bluhm, B.A., Gradinger, R. 2005. Food web structure in the high Arctic Canada Basin: evidence from delta C-13 and delta N-15 analysis. Polar Biology, 28(3), 238-249.

IPCC, 2007: Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jackson, G.A., Waite, A.M., Boyd, P.W. 2005. Role of algal aggregation in vertical carbon export during SOIREE and in other low biomass environments. Geophysical Research Letters, 32.

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

Kattner, G., Gercken, G., Eberlein, K. 1983. Development of lipids during a spring plankton bloom in the northern North Sea. Marine Chemistry, 14, 149-162.

Keil, R.G., Montlucon, D.B., Prahl, F.G., Hedges, J. I. 1994. Sorptive preservation of labile organic matter in marine sediments. Nature, 370(6490), 549-552.

Keil, R.G., Tsamakis, E., Hedges, J.I. 2000. Early diagenesis of particulate amino acids in marine systems. In: Goodfriend, G.A., Collins, M.J., Fogel, M.L., Macko, S.A., Wehmiller, J.F. (Eds.), Perspectives in Amino Acid and Protein Geochemistry. Oxford University Press, New York, pp. 69– 82.

Keller, A., Nesvizhskii, A.I., Kolker, E., Aebersold, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical Chemistry, 74(20), 5383-5392.

Kirchman, D., K'Nees, E., Hodson, R. 1985. Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. Applied and Environmental Microbiology, 49(3), 599-607.

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

Knies, J. and Stein, R. 1998. New aspects of organic carbon deposition and its paleoceanographic implications along the northern Barents Sea margin during the last 30,000 years. Paleoceanography, 13, No. 4, 384-394.

Konopka, A. and Wilkins, M.J. 2012. Application of meta-transcriptomics and proteomics to analysis of in situ physiological state. Frontiers in Microbiology, 3: 184.

Kröger, N., Deutzmann, R., Bergsdorf, C., Sumper, M. 2000. Species-specific polyamines from diatoms control silica morphology. Proceedings of the National Academy for Science, 97: 14133-14138.

Kuma, K., Nakabayashi, S., Matsunaga, K. 1995. Photoreduction of Fe(III) by Hydroxycarboxylic Acids in Seawater. Water Research, 29: 1559-1569.

Kuster, B., Hunter, A. P., Wheeler, S. F., Dwek, R. A., Harvey, D. J. 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.

Kustka, A.B., Allen, A.E., Morel, F.M.M. 2007. Sequence analysis and transcriptional regulation of iron acquisition genes in two marine diatoms. Journal of Phycology, 43: 715-729.

Kwok, E.Y., Severance, S., Kosman, D.J. 2006. Evidence for iron channeling in the Fet3p-Ftr1p high-affinity iron uptake complex in the yeast plasma membrane. Biochemistry, 45: 6317-6327.

Lacerda, C.M.R., Choe, L.H., Reardon, K.F. 2007. Metaproteomic analysis of a bacterial community response to cadmium exposure. Journal of Proteome Research, 6: 1145-1152.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680-685.

LaRoche, J., Boyd, P.W., McKay, R.M.L., Geider, R.J. 1996. Flavodoxin as an in situ marker for iron stress in phytoplankton. Nature, 382: 802-805.

Laver, W.G. 1964. Structural studies on the protein subunits from three strains of influenza virus. Journal of Molecular Biology, 9, 109-124.

Laws, E.A., Popp, B.N., Bidigare, R.R., Kennicutt, M.C., Macko, S.A. 1995. Dependence of phytoplankton carbon isotopic composition on growth rate and $[CO_2]_{aq}$: Theoretical considerations and experimental results. Geochimica et Cosmochimica Acta, 59 (6), 1131-1138.

Lefevre, N. and Watson, A.J. 1999. Modelling the geochemical cycle of iron in the oceans and its impact on atmospheric CO_2 concentrations. Global Biogeochemical Cycles, 13: 727-736.

Lehninger, A.L. 1965. The Molecular Basis of Biological Energy Transformations. In: Benjamin WA, editor. Bioenergetics. New York. pp. 258.

Letunic, I., Yamada, T., Kanehisa, M., Bork, P. 2008. iPath: interactive exploration of biochemical pathways and networks. Trends in Biochemical Sciences, 33: 101-103.

Leynaert, A, Bucciarelli, E., Claquin, P., Dugdale, R.C., Martin-Jézéquel, V., Pondaven, P., Ragueneau, O. 2004. Effect of iron deficiency on diatom cell size and silicic acid uptake kinetics. Limnology and Oceanography, 49: 1134-1143.

Li, W.K.W., McLaughlin, F.A., Lovejoy, C., Carmack, E.C. 2009. Smallest algae thrive as the Arctic Ocean freshens. Science, 326, 539.

Limmer, A.W. and Wilson, A.T. 1980. Amino-acids in buried paleosols. Journal of Soil Science, 31(1), 147-153.

Liu, H., Sadygov, R.G., Yates, J.R. 3rd. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical Chemistry 76: 4193–4201.

Llacer, J.L., Fita, I., Rubio, V. 2008. Arginine and nitrogen storage. Current Opinion in Structural Biology, 18: 673-681.

Lommer, M., Specht, M., Roy, A.S., Kraemer, L., Andreson, R., Gutowska, M.A., Wolf, J., Bergner, S.V., Schilhabel, M.B., Klostermeier, U.C., Beiko, R.G., Rosenstiel, P., Hippler, M., LaRoche, J. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biology, 13: R66.

Lourenco, S.O., Barbarino, E., Marquez, U.M.L., Aidar, E. 1998. Distribution of intracellular nitrogen in marine microalgae: basis for the calculation of specific nitrogen-to-protein conversion factors. Journal of Phycology, 34(5), 798-811.

Macdonald, R.W., Sakshaug, E., Stein, R. 2004. The Arctic Ocean: Modern status and recent climate change. In: R. Stein and R.W. Macdonald (editors), The Organic Carbon Cycle in the Arctic Ocean. Springer-Verlag, Berlin, pp.6-21.

Maizel, J.V. 2000. SDS polyacrylamide gel electrophoresis. Trends in Biochemical Sciences 25(12), 590-592.

Maldonado, M.T., Allen, A.E., Chong, J.S., Lin, K., Leus, D., Karpenko, N., Harris, S.L. 2006. Copper-dependent iron transport in coastal and oceanic diatoms. Limnology and Oceanography, 51: 1729-1743.

Maldonado, M.T. and Price, N.M. 1996. Influence of N substrate on Fe requirements of marine centric diatoms. Marine Ecology-Progress Series, 141: 161-172.

Mann, D.H. and Hamilton, T.D. 1995. Late Pleistocene and Holocene Paleoenvironments of the North Pacific Coast. Quaternary Science Review, 14, 449-471.

Marchetti, A. and Cassar, N. 2009. Diatom elemental and morphological changes in response to iron limitation: a brief review with potential paleoceanographic applications. Geobiology, 7: 419-431.

Marchetti, A., Schruth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, A.E., Armbrust, E.V. 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proceedings of the National Academy of Sciences of the United States of America, 109: E317-E325.

Mari, X., Rassoulzadegan, F., Brussaard, C.P.D., Wassmann, P. 2005. Dynamics of transparent exopolymeric particles (TEP) production by Phaeocystis globosa

under N- or P-limitation: a controlling factor of the retention/export balance. Harmful Algae, 4: 895-914.

Mari, X., Torreton, J.P., Claire, B.T.T., Bouvier, T., Thuoc, C.V., Lefebvre, J.P., Ouillon, S. 2012. Aggregation dynamics along a salinity gradient in the Bach Dang estuary, North Vietnam. Estuarine Coastal and Shelf Science, 96: 151-158.

Martin, J.H., Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S.J., Hunter, C.N., Elrod, V.A., Nowicki, J.L., Coley, T.L., Barber, R.T., Lindley, S., Watson, A.J., Vanscoy, K., Law, C.S., Liddicoat, M.I., Ling, R., Stanton,T., Stockel, J., Collins, C., Anderson, A., Bidigare, R., Ondrusek, M., Latasa, M., Millero, F.J., Lee, K., Yao, W., Zhang, J.Z., Friederich, G., Sakamoto, C., Chavez, F., Buck, K., Kolber, Z., Greene, R., Falkowski, P., Chisholm, S.W., Hoge, F., Swift, R., Yungel, J., Turner, S., Nightingale, P., Hatton, A., Liss, P., Tindale, N.W. 1994. Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. Nature, 371: 123-129.

Mayer, L.M. 1994. Surface area control of organic carbon accumulation in continental shelf sediments. Geochimica et Cosmochimica Acta, 58(4), 1271-1284.

Mayer, L., Benninger, L., Bock, M., DeMaster, D., Roberts, Q., Martens, C. 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 Research Part II, 49(20), 4587-4597.

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

McDonald, W.H. and Yates, J.R., 3rd (2002) Shotgun proteomics and biomarker discovery. Disease Markers, 18: 99-105.

McKay, J.L., de Vernal, A., Hillaire-Marcel, C., Not, C., Polyak, L., Darby, D. 2008. Holocene fluctuations in Arctic sea-ice cover: dinocyst-based reconstructions for the eastern Chukchi sea. Can. J. Earth Sci., 45, 1377-1397.

Meyers, P.A. 1997. Organic geochemical proxies of paleoceanographic, paleolimnologic, and paleoclimatic processes. Organic Geochemistry, 27, 213-250.

Milligan, A.J. and Harrison, P.J. 2000. Effects of non-steady-state iron limitation on nitrogen assimilatory enzymes in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae). Journal of Phycology, 36: 78-86.

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

Mock, T., Samanta, M.P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., Durkin, C., BonDurant, S.S., Richmond, K., Rodesch, M., Kallas, T., Huttlin, E.L., Cerrina, F., Sussmann, M.R., Armbrust, E.V. 2008. Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. Proceedings of the National Academy for Science USA, 105: 1579-1584.

Moore, E.K., Nunn, B.L., Faux, J.F., Goodlett, D.R. Harvey, H.R. 2012. Evaluation of electrophoretic protein extraction and database-drive protein identification from marine sediments. Limnology and Oceanography: Methods, 10, 353-366.

Moore, E.K., Nunn, B.L., Goodlett, D.R., Harvey, H.R. 2012. Identifying and tracking proteins through the marine water column: Insights into the inputs and preservation mechanisms of protein in sediments. Geochimica et Cosmochimica Acta, 83, 324-359.

Moran, M.A., Belas, R., Schell, M.A., Gonzalez, J.M., Sun, F., Binder, B.J., Edmonds, J., Ye, W., Orcutt, B., Howard, E.C., Meile, C., Palefsky, W., Goesmann, A., Ren, Q., Paulsen, I., Ulrich, L.E., Saunders, E., Buchan, A. 2007. Ecological genomics of marine *Roseobacters*. Applied and Environmental Microbiology, 73(14), 4559-4569.

Moran, M.A., Buchan, A., Gonzalez, J.M., Heidelberg, J.F., Whitman, W.B., Kiene, R.P., Henriksen, J.R., King, G.M., Belas, R., Fuqua, C., Brinkac, L., Lewis, M., Johnri, S., Weaver, B., Pai, G., Eisen, J.A., Rahe, E., Sheldon, W.M., Ye, W., Miller, T.R., Carlton, J., Rasko, D.A., Paulsen, I.T., Ren, Q., Daugherty, S.C., Deboy, R.T., Dodson, R.J., Durkin, A.S., Madupus, R., Nelson, W.C., Sullivan, S.A., Rosovitz, M.J., Haft, D.H., Selengut, J., Ward, N. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature, 432(7019), 910-913.

Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., Rocap, G. 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. The ISME Journal, 4, 673-685.

Mueller-Lupp, T., Bauch, H.A., Erlenkeuser, H., Hefter, J., Kassens, H., Thiede, J. 2000. Changes in the deposition of terrestrial organic matter on the Laptev Sea shelf during the Holocene: evidence from stable carbon isotopes. International Journal of Earth Sciences, 89, 563-568.

Muggli, D.L., Lecourt, M., Harrison, P.J. 1996. Effects of iron and nitrogen source on the sinking rate, physiology and metal composition of an oceanic diatom from the subarctic Pacific. Marine Ecology-Progress Series, 132: 215-227.

Müller, J., Massé, G., Stein, R., Belt, S. 2009. Variability of sea-ice conditions in the Fram Strait over the past 30,000 years. Nature Geoscience, 2, 772-776.

Naidu, A.S., Cooper, L.W., Finney, B.P., Macdonald, R.W., Alexander, C., Semiletove, I.P. 2000. Organic carbon isotope rations ($\delta^{13}C$) of Arctic Amerasian Continental shelf sediments. International Journal of Earth Sciences, 89, 522-532.

Nesvizhskii, A., Keller, A., Kolker, E., Aebersold, R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. Analytical Chemistry, 75(17), 4646-4658.

Nguyen, R.T. and Harvey, H.R. 2001. Preservation of protein in marine systems: hydrophobic and other noncovalent associations as major stabilizing forces. Geochimica et Cosmochimica Acta, 65 (9), 1467-1480.

Nguyen, R.T. and Harvey, H.R. 2003. Preservation via macromolecular associations during *Botryococcus braunii* decay: proteins in the Pula Kerogen. Organic Geochemistry, 34, 1391-1403.

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., 3rd, Bairoch, A., Bergeron, J.J.M. 2010. Mass spectrometry in high-throughput proteomics: ready for the big time. Nature Methods, 7: 681-685.

Nisi, P.D. and Zocchi, G. 2000. Phosphoenolpyruvate carboxylase in cucumber (Cucumis sativus L.) roots under iron deficiency: activity and kinetic characterization. Journal of Experimental Botany, 51: 1903-1909.

Nunn, B L., Aker, J. R., Shaffer, S. A., Tsai, Y. H., Strzepek, R. F., Boyd, P. W., Freeman, T. L., Brittnacher, M., Malmstrom, L., and Goodlett, D. R. 2009. Deciphering diatom biochemical pathways via whole cell proteomics. Aquatic Microbial Ecology, 55: 241-253.

Nunn, B.L., Faux, J.F., Hippmann, A.A., Maldonado, M.T., Harvey, H.R., Goodlett, D.R., Boyd, P.W., Strzepek, R.F. 2013. Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. PLoS ONE, 8(10): e75653. doi:10.1371/journal.pone.0075653

Nunn B. L. and Keil, R. G. 2006. A comparison of non-hydrolytic methods for extracting amino acids and proteins from coastal marine sediments. Marine Chemistry, 98, 31-42.

Nunn B. L. and Timperman, A. T. 2007. Marine Proteomics. Marine Ecology Progress Series, 332, 281-289.

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

Ogunseitan, O. A. 1993. Direct extraction of proteins from environmental samples. Journal of Microbiological Methods, 17(4), 273-281.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., Ahn, N.G.. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular and Cellular Proteomics, 4: 1487-1502.

Ozturk, M., Croot, P.L., Bertilsson, S., Abrahamsson, K., Karlson, B., David, R., Fransson, A., Sakshaug, E. 2004. Iron enrichment and photoreduction of iron under UV and PAR in the presence of hydroxycarboxylic acid: implications for phytoplankton growth in the Southern Ocean. Deep-Sea Research Part II, 51: 2841-2856.

Panchaud, A., Scherl, A., Shaffer, S.A., von Haller, P.D., Kulasekara, H.D., Miller, S.I., Goodlett, D.R. 2009. Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. Analytical Chemistry, 81: 6481-6488.

Passow, U., Shipe, R.F., Murray, A., Pak, D.K., Brzezinski, M.A., Alldredge, A.L. 2001. The origin of transparent exopolymer particles (TEP) and their role in the sedimentation of particulate matter. Continental Shelf Research, 21: 327-346.

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

Pickart, R.S., Weingartner, T.J., Pratt, L.J., Zimmermann, S., Torres, D.J. 2005. Flow of winter-transformed Pacific water into the Western Arctic. Deep-Sea Research II, 52, 3175-3198.

Polyak, L., Bischof, J., Ortiz, J.D., Darby, D.A., Channell, J.E.T., Xuan, C., Kaufman, D.S., Løvlie, R., Schneider, D.A., Eberl, D.D., Adler, R.E., Council, E.A. 2009. Late Quaternary stratigraphy and sedimentation patterns in the western Arctic Ocean. Global and Planetary Change, 68 (1-2), 5-17.

Porter, K.G. and Fieg, Y.S. 1980. The use of DAPI for identifying and counting aquatic microflora. Limnology and Oceanography, 25(5), 943-948.

Powell, M.J., Sutton, J.N., Del Castillo, C.E., Timperman, A.I. 2005. Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. Marine Chemistry, 95 (3-4), 183-198.

Price, N.M. and Harrison, P.J. 1988. Uptake of Urea-C and Urea-N by the Coastal Marine Diatom *Thalassiosira-Pseudonana*. Limnology and Oceanography, 33: 528-537.

Price, N.M., Harrison, G.I., Hering, J.G., Hudson, R.J., Nirel, P.M.V., Palenik, B., Morel, F.M.M. 1988. Preparation and chemistry of the artificial algal culture medium Aquil. Biological Oceanography, 6: 443-461

Price, N,M, and Morel, F.M.M. 1991. Colimitation of Phytoplankton Growth by Nickel and Nitrogen. Limnology and Oceanography, 36: 1071-1077.

Pruitt, K., Tatusova, T., Maglott, D. 2002. The Reference Sequence (RefSeq) Project. In The NCBI Handbook (ed. J. McEntyre, Ostell, J.). National Center for Biotechnology Information.

Quince, C., Curtis, T. P., Sloan, W. T. 2008. The rational exploration of microbial diversity. The ISME Journal, 2, 1997-2006.

Rashid, M. A., Buckley, D. E., Robertson, K. R. 1972. Interactions of a marine humic acid with clay minerals and a natural sediment. Geoderma, 8, 11-27.

Raven, J.A., Evans, M.C.W., Korb, R.E. 1999. The role of trace metals in photosynthetic electron transport in O-2-evolving organisms. Photosynthesis Research, 60: 111-149.

Reinhardt, I., Haebel, S., Herbik, A., Buckhout, T.J. 2006. Proteomic studies under iron stress: Iron deficiency-induced regulation of protein synthesis in the green alga *Chlamydomonas reingardtii*. In: Barton LL, Abadia J, editors. Iron Nutrition in Plants and Rhizospheric Microorganisms. Netherlands: Springer. pp. 371-393.

Reisch, C.R., Crabb, W.M., Gifford, S.M., Teng, Q., Stoudemayer, M.J., Moran, M.A., Whitman, W.B. 2013. Metabolism of dimethylsulphoniopropionate by *Ruegeria pomeroyi* DSS-3. Molecular Microbiology, 89(4), 774-791.

Reisfeld, R. A., Lewis, U. J., Williams, D. E. 1962. Disc electrophoresis of basic proteins and peptides on polyacrylamide gels. Nature, 195, 281-283.

Reuss, F. F. 1809. Mem. Soc. Imperiale Naturalists de Moscow 2, 327.

Rijkenberg, M.J.A., Fischer, A.C., Kroon, J.J., Gerringa, L.J.A., Timmermans, K.R., Wolterbeek, H.T., de Baar, H.J.W. 2005. The influence of UV irradiation on the photoreduction of iron in the Southern Ocean. Marine Chemistry, 93: 119-129.

Roth, L. C. and Harvey, H. R. 2006. Intact protein modification and degradation in estuarine environments. Marine Chemistry, 102, 33-45.

Sambrotto, R. N., Niebauer, H. J., Goering, J. J., Iverson, R. L. 1986. Relationships among vertical mixing, nitrate uptake, and phytoplankton growth during spring bloom in the S-E Bering Sea. Continental Shelf Research, 5, 161-198.

Santoni, V., Molloy, M., Rabilloud, T. 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis, 21: 1054-1070.

Sarmiento, J.L., Slater, R.D., Dunne, J., Gnanadesikan, A., Hiscock, M.R. 2010. Efficiency of small scale carbon mitigation by patch iron fertilization. Biogeosciences, 7: 3593-3624.

Scherl, A., Shaffer, S. A., Taylor, G. K., Kulasekara, H. D., Miller, S. I., Goodlett, D. R. 2008. Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. Analytical Chemistry, 80(4), 1182-1191.

Schubert, C.J. and Calvert, S.E. 2001. Nitrogen and carbon isotope composition of marine and terrestrial organic matter in Arctic Ocean sediments: implications for nutrient utilization and organic matter composition. Deep-Sea Research I, 48, 789-810.

Schubert, C.J., Stein, R., Calvert, S.E. 2001. Tracking nutrient and productivity variations over the last deglaciation in the Arctic Ocean. Paleoceanography, 16(2), 199-211.

Schulze, W.X., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., Schulze, E.D. 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia, 142, 335-343.

Schweder, T., Markert, S., Hecker, M. 2008. Proteomics of marine bacteria. Electrophoresis, 29, 2603-2616.

Shapiro, A.L., Vinuela, E., Maizel, J.V., Jr. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochemical and Biophysical Research Communications, 28, 815-820.

Scheele, U. and Holstein, S.E. 2002. Functional evidence for the identification of an Arabidopsis clathrin light chain polypeptide. FEBS Letters, 514: 355-360.

Shevchenko, A., Wilm, M., Vorm, O., Mann, M. 1996. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Analytical Chemistry, 68(5), 850-858.

Simon, M. and Azam, F. 1989. Protein content and protein synthesis rates of planktonic marine bacteria. Marine Ecology Progress Series, 51, 201-213.

Sinninghe Damsté, J.S.S., Ossebaar, J., Abbas, B., Schouten, S., Verschuren, D. 2009. Fluxes and distribution of tetraether lipids in an equatorial African lake: Constraints on the application of the TEX_{86} palaeothermometer and BIT index in lacustrine settings. Geochimica et Cosmochimica Acta, 73, 4232–4249

Sowell, S.M., Abraham, P.E., Shah, M., Verberkmoes, N.C., Smith, D.P., Barofsky, D.F., Giovannoni, S.J. 2011. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. ISME Journal, 5: 856-865.

Spahr, C.S., Davis, M.T., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., Mort, J., Courchesne, P.L., Chen, K., Wahl, R.C., Yu, W., Luethy, R., Patterson, S.D. 2001. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I. Profiling an unfractionated tryptic digest. Proteomics, 1(1), 93-107.

Stabeno, P.J., Farley, Jr., E.V., Kachel, N.B., Moore, S., Mordy, C.W., Napp, J.M., Overland, J.E., Pinchuk, A.I., Sigler, M.F. 2012. A comparison of the physics of the northern and southern shelves of the eastern Bering Sea and some implications for the ecosystem. Deep Sea Research II, 65-70, 14-30.

Steigenberger, S., Statham, P.J., Volker, C., Passow, U. 2010. The role of polysaccharides and diatom exudates in the redox cycling of Fe and the photoproduction of hydrogen peroxide in coastal seawaters. Biogeosciences, 7: 109-119.

Stein, R. and Fahl, K. 2000. Holocene accumulation of organic carbon at Laptev Sea continental margin (Arctic Ocean): sources, pathways, and sinks. Geo-Marine Letters, 20, 27-36.

Stein, R. and Macdonald, R.W. 2004. Geochemical proxies used for organic carbon source identification in Arctic Ocean sediments. In: R. Stein and R.W. Macdonald (editors), The Organic Carbon Cycle in the Arctic Ocean. Springer-Verlag, Berlin, pp.6-21.

Stein, R., Boucsein, B., Fahl, K., Garcia de Oteyza, T., Knies, J., Niessen, F. 2001. Accumulation of particulate organic carbon at the Eurasian continental margin during late Quaternary times: controlling mechanisms and paleoenvironmental significance. Global Planetary Change, 31, 97-104.

Stewart, E.J. 2012. Growing unculturable bacteria. Journal of Bacteriology, 194(16), 4151-4160.

Strochlic, T.I., Schmiedekamp, B.C., Lee, J., Katzmann, D.J., Burd, C.G. 2008. Opposing activities of the Snx3-retromer complex and ESCRT proteins mediate regulated cargo sorting at a common endosome. Molecular Biology of the Cell, 19: 4694-4706.

Strzepek, R.F. and Harrison, P.J. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. Nature, 431: 689-692.

Strzepek, R.F., Hunter, K.A., Frew, R.D., Harrison, P.J., Boyd, P.W. 2012. Iron-light interactions differ in Southern Ocean phytoplankton. Limnology and Oceanography, 57: 1182-1200.

Stuiver, M., Reimer, P.J., Reimer, R.W. 2005. Online radiocarbon calibration program CALIB 5.0.2. http://calib.qub.ac.uk/calib/.

Sunda, W.G. and Huntsman, S.A. 1997. Interrelated influence of iron, light and cell size on marine phytoplankton growth. Nature, 390: 389-392.

Sunda, W.G., Price, N.M., Morel, F.M.M. 2005. Trace metal ion buffers and their use in culture studies. In: Andersen RA, editor. Algal Culturing Techniques. Amsterdam: Acad. Press/Elsevier. pp. 35-63.

Tanoue, E. 1995. Detection of dissolved protein molecules in oceanic waters. Marine Chemistry, 51, 239-252.

Tanoue, E., Nishiyama, S., Kamo, M., Tsugita, A. 1995. Bacterial membranes: Possible source of a major dissolved protein in seawater. Geochimica et Cosmochimica Acta, 59 (12), 2643-2648.

Tibbles, B.J. and Harris, J.M. 1996. Use of radiolabelled thymidine and leucine to estimate bacterial production in soils from continental Antarctica. Applied and Environmental Microbiology, 62(2), 694-701.

Timmermans, K.R., van der Wagt, B., de Baar, H.J.W. 2004. Growth rates, half-saturation constants, and silicate, nitrate, and phosphate depletion in relation to iron availability of four large, open-ocean diatoms from the Southern Ocean. Limnology and Oceanography, 49: 2141-2151.

Tran, J.C. and Doucette, A.A. 2009. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. Analytical Chemistry, 81(15), 6201-6209.

Volkman, J.K. 1986. A review of sterol markers for marine and terrigenous organic matter. Organic Geochemistry, 9, 83-99.

Volkman, J.K. 2003. Sterols in microorganisms. Applied Microbiology Biotechnology, 60, 495-506.

Volkman, J.K. 2006. Lipid markers for marine organic matter. In: J.K. Volkman (editor), Handbook of Environmental Chemistry. Springer-Verlag, Berlin Heidelberg.

Waldhier, M.C., Dettmer, K., Gruber, M.A., Oefner, P.J. 2010. Comparison of derivatization and chromatographic methods for GC-MS analysis of amino acid enantiomers in physiological samples. Journal of Chromatography B, 878(15-16), 1103-1112.

Wakeham, S.G., Lee, C., Hedges, J.I., Hernes, P.J., Peterson, M.L. 1997. Molecular indicators of diagenetic status in marine organic matter, Geochimica et Cosmochimica Acta, 61, 5363-5369.

Walsh, J.J., McRoy, C.P., Coachman, L.K., Goering, J.J., Nihoul, J.J., Whitledge, T.E., Blackburn, T.H., Parker, P.L., Wirick, C.D., Shuert, P.G., Grebmeier, J.M., Springer, A.M., Tripp, R.D., Hansell, D.A., Djenidi, S., Deleersnijder, E., Henriksen, K., Lund, B.A., Andersen, P., Mullerkarger, F.E., Dean, K. 1989. Carbon and nitrogen cycling within the Bering Chukchi Seas - Source regions for organic-matter effecting AOU demands of the Arctic-Ocean. Progress in Oceanography, 22(4), 277-359.

Wang, X.C. and Lee, C. 1993. Adsorption and desorption of aliphatic amines, amino acids and acetate by clay minerals and marine sediments. Marine Chemistry, 44, 1-23.

Wang, Y.H., Garvin, D.F., Kochian, L.V. 2001. Nitrate-induced genes in tomato roots. Array analysis reveals novel genes that may play a role in nitrogen nutrition. Plant Physiology, 127: 345-359.

Washburn, M.P., Wolters, D., Yates, J.R. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology, 19, 242-247.

Weingartner, T., Aagaard, K., Woodgate, R., Danielson, S., Sasaki, Y., Cavalieri, D. 2005. Circulation on the north central Chukchi Sea shelf. Deep-Sea Research II, 52, 3150-3174.

Westall, J.C., Zachary, J.L., Morel, F.M.M. 1976. Mineql - General Algorithm for Computation of Chemical-Equilibrium in Aqueous Systems. Abstracts of Papers of the American Chemical Society, 172: 8-8.

Westerhausen, L., Poynter, J., Eglinton, G., Erlenkeuser, H., Sarnthein, M. 1993. Marine and terrigeneous origin of organic matter in modern sediments of the equatorial east Atlantic: the $\delta^{13}C$ and molecular record. Deep-Sea Research I, 40, 1087-1121.

Whitney, L.P., Lins, J.J., Hughes, M.P., Wells, M.L., Chappell, P.D., Jenkins, B.D. 2011. Characterization of putative iron responsive genes as species-specific indicators of iron stress in thalassiosiroid diatoms. Frontiers in Microbiology, 2: 234.

Wilken, S., Hoffmann, B., Hersch, N., Kirchgessner, N., Dieluweit, S., Rubner, W., Hoffmann, L.J., Merkel, R., Peeken, I. (2011) Diatom frustules show increased mechanical strength and altered valve morphology under iron limitation. Limnology and Oceanography, 56: 1399-1410.

Williams, T.J., Wilkins, D., Long, E., Evans, F., DeMaere, M.Z., Raftery, M.J., Cavicchioli, R. 2012. The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. Environmental Microbiology, 15(5), 1302-1317.

Wood, A.M., Everroad, R.C., Wingard, L.M. 2005. Measuring growth rates in microalgal cultures. In: Andersen RA, editor. Algal Culturing Techniques. Phycological Society of America: Elsevier, Amsterdam. pp. 269-285.

Yamada, N. and Tanoue, E. 2003. Detection and partial characterization of dissolved glycoproteins in oceanic waters. Limnology and Oceanography, 48(2), 1037-1048.

Yamada, N. and Tanoue, E. 2009. Similarity of electrophoretic dissolved protein spectra from coastal to pelagic seawaters. Journal of Oceanography, 65, 223-233.

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., Bork, P. 2011. iPath2.0: interactive pathway explorer. Nucleic Acids Research, 1-4, doi:10.1093/nar/gkr313.

Yamamoto, M. and Polyak, L. 2009. Changes in terrestrial organic matter input to the Mendeleev Ridge, western Arctic Ocean, during the Late Quaternary. Global and Planetary Change, 68 (1-2), 30-37.

Yamamoto, M., Tatsufumi, O., Sugisaki, S., Sakamoto, T. 2008. Late Pleistocene changes in terrestrial biomarkers in sediments from the central Arctic Ocean. Organic Geochemistry, 39, 754-763.

Yi, E.C., Marelli, M., Lee, H., Purvine, S.O., Aebersold, R., Aitchison, J.D., Goodlett, D.R. 2002. Approaching complete peroxisome characterization by gas-phase fractionation. Electrophoresis, 23, 3205-3216.

Yooseph, S., Sutton G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C.S., Li, H., Mashiyama, S.T., Joachimiak, M.P., van Belle, C., Chandonia, J.M., Soergel, D.A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B.J., Bafna, V., Friedman, R., Brenner, S.E., Godzik, A., Eisenberg, D., Dixon, J.E., Taylor, S.S., Strausberg, R.L., Frazier, M., Venter, J.C. 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biology, 5(3), 432-466.

Yunker, M.B., Belicka, L.L., Harvey, H.R., Macdonald, R.W. 2005. Tracing the inputs and fate of marine and terrigenous organic matter in Arctic Ocean sediments: A multivariate analysis of lipid biomarkers. Deep-Sea Research II, 52(24-26), 3478-3508.

Yunker, M.B., Macdonald, R.W., Veltkamp, D.J., Cretney, W.J. 1995. Terrestrial and marine biomarkers in a seasonally ice-covered Arctic estuary – integration of multivariate and biomarker approaches. Marine Chemistry, 49, 1-50.

Zang, X., van Heemst, J.D.H., Dria, K.J., Hatcher, P.G. 2000. Encapsulation of protein in humic acid from a histosol as an explanation for the occurrence of organic nitrogen in soil and sediment. Organic Geochemistry, 31(7-8), 679-695.