PROTEOMIC ANALYSIS OF THE COCCOLITHS OF THE COCCOLITHOPHORE SPECIES *EMILIANIA HUXLEYI*

By

SARAH ROSEMARIE NEWPORT

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Earth and Planetary Sciences

Written under the direction of

Paul Falkowski

And approved by

_____

_____

_____

New Brunswick, New Jersey

January 2022

ABSTRACT OF THE THESIS

PROTEOMIC ANALYSIS OF COCCOLITHS OF THE COCCOLITHOPHORE SPECIES *EMILIANIA HUXLEYI*

By

SARAH ROSEMARIE NEWPORT

Thesis Director:

Paul G. Falkowski

Biominerals are mineral-organic composite materials created by living organisms which often contain proteins that facilitate the mineralization process. Studying biominerals and biomineralizing proteins is important for understanding the molecular processes that make biomineralization possible and for investigating the origin and evolution of the process. Coccolithophores are abundant calcium carbonate producing biomineralizing organisms and are central to many biogeochemical cycles. However, while much work has been done to understand coccolith formation and composition, no biomineralizing protein has yet to be definitively identified in coccoliths. The purpose of this study was to show that proteins are occluded in the coccoliths and to begin identifying those that may be involved in coccolith mineralization. Coccolith samples were collected and cleaned from laboratory grown cultures. The liths were analyzed for protein content using nano LC-MS/MS, spectrophotometry, and gel electrophoresis. Mass spectrometry analysis found 100 discrete proteins in the sample. Of these, bioinformatic analysis identified 77 proteins with functional annotations, including six that may be potentially involved in biomineralization in some capacity.

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to Professor Paul Falkowski for being my advisor and mentor. 2020 and 2021 were unusually challenging years, with COVID-19 and social unrest upending the lives of so many, myself included. I was dispirited and close to withdrawing after my initial graduate school plans fell apart due to the pandemic, but Paul helped put me back on course. His support and encouragement mean so much to me. I am certain I would not be completing my degree had he not become my advisor.

I am also deeply grateful to Dr. Jeana Drake of UCLA for lending me her expertise in biomineralizing proteins. She devoted her own time to help me understand my results and answer my many, many questions. Her kindness and graciousness helped pull me through the more challenging moments of my work.

I want to thank my committee members Professor Kenneth Miller and Professor Yair Rosenthal for their contributions and encouragement. Their support and patience will not be forgotten.

I would like to offer my special thanks to Falkowski lab manager Dr. Kevin Wyman for showing me the ropes of the lab and helping me with every step of my laboratory work. I also am grateful to the members of both the Falkowski and Bidle research groups for offering their friendship and help when I needed it.

I wish to acknowledge all involved at Rutgers University and especially my department who kept cool heads during uncertain times and ensured that the work could go on safely. I am especially grateful to Professor Miller and Professor Jim Wright who offered their support and eased my academic concerns following the COVID-19 fallout. A sincere thank you to all my EPS friends for keeping me sane and supporting me.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

.

LIST OF TABLES

Table **Page**

INTRODUCTION

A biomineral is a mineral-organic composite material produced by living organisms (Weiner & Dove, 2003). They are truly representive of the convergence of geologic and biologic disciplines. Over 60 different biominerals have been identified, consisting of widely different mineral compositions, and are found across all kingdoms of life (Lowenstam, 1989). The genesis of biomineralization represents a significant evolutionary milestone in the history of life. For these, and many other reasons, the biomineralization process represents an important system for study.

For the first three billion years of life on Earth, all organisms were soft bodied, and left virtually no physical trace of their existence in the geologic record. Prokaryotes such as bacteria were the first lifeforms to evolve the ability to mineralize, often passively. However, nearly 550 million years ago there was a sudden and widespread onset of biomineralization among eukaryotic macrobiota. The capability of eukaryotes to generate their own mineralized skeletons was one of the hallmarks of the "Cambrian Explosion." The significance of biomineralization cannot be understated considering how much of our understanding of Earth history is derived from the study of the skeletal remains of biomineralizing organisms. While much has been learned about the molecular mechanisms that produce biominerals, it remains unclear exactly how it is done and when this ability originated. A better understanding of the molecular processes that facilitate biomineralization provides insights into how, why, when, and where certain phyla and species emerged on our planet (Lowenstam, 1989).

Proteins in the organic matrix of many biominerals are hypothesized to serve as a framework for crystal nucleation and other mineralization processes. Many proteins

involved in biomineralization have been described and studied extensively (Chen *et al.*, 1984; Luo *et al.*, 1997; Marin *et al.*, 1996; Mass *et al.*, 2013; Michenfelder *et al.*, 2003). Identifying and studying the function of such proteins is critical for understanding biomineralization at a mechanical, molecular level. Coccolithophores are perhaps the most prolific and abundant producers of biogenic calcite on the planet, making them prime research subjects for biomineralization studies.

Studying the protein composition of coccoliths is challenging for many reasons. A primary challenge is the microscopic size of coccoliths. Individual coccoliths are only a few microns in size. Of this, only a small fraction is the organic matrix, roughly 3% of the coccolith by weight (Westbroek *et al.*, 1973). And of that, only a fraction would be protein. Thus, the concentration of protein is likely vanishingly small and difficult to analyze.

There are also difficulties with analysis techniques due to the tendency for such proteins to be acidic. In gel electrophoresis, a routine analysis for protein work, most bands are diffuse and smeared, even in two-dimensional gels, where glycosylation can lead to poor separation (Dauphin, 2001). There can also be great differences in the results of gel electrophoresis depending on the protocols and the species used (Gotliv *et al.*, 2003; Puverel *et al.*, 2005).

Previous studies suggest proteins may be identified in coccoliths, but it challenging to demonstrate. A 2016 study showed that coccoliths dissolved in ethylenediaminetetraacetic acid (EDTA) contain soluble acidic polysaccharides (Lee *et al.*, 2016). The same study also tested for protein, but the researchers were not able to detect any (Lee *et al.*, 2016). Another study in provided evidence that there is a small

concentration of protein in the organic fraction of coccoliths from *Pleurochrysis carterae* (Gal *et al.*, 2016). The researchers detected a protein band in silver stained SDS-polyacrylamide gel from the coccolith SOM, but there was no additional mass spectrometry analysis. However, their sample preparation preserved coccolith-associated organic material in order to keep it close to its native state (Gal *et al.*, 2016) and the protein may be part of the organic base plate.

This work sought to provide further evidence that there are indeed proteins incorporated within the mineral phase of coccoliths. The ultimate goal of this work is to begin identifying proteins in the mineral phase of coccoliths that may play important roles in mineralization to advance a better understanding of the biomineralization system.

BACKGROUND

Biomineralization

In modern life, biomineralization is widespread and very diverse. Earlier investigations into biomineralization generally assumed that biological crystallization followed a classical nucleation pathway (Weiner & Addadi, 2011). Through this pathway, crystallization occurs when a saturated solution reaches a critical point where the dissolved chemical species begin to coalesce into a nucleus, seeding crystal growth. However, subsequent research has revealed that many biominerals are formed from disordered, amorphous precursor phases via nonclassical nucleation (Lowenstam & Weiner, 1985; Weiner & Addadi, 2011).

Though it remains an area of active investigation, at least three common biological pathways have been proposed to explain the formation of biogenic crystals among differing organisms: (1) mineralization within a syncytium (large vesicle), (2) mineralization within intracellular vesicles, and (3) mineralization on an extracellular organic matrix (Weiner & Addadi, 2011). Pathway 1 is demonstrated in echinoderm skeleton biomineralization while pathway 2 is exemplified by the production of coccoliths. The formation of mollusk shells, bone, enamel, and radial foraminifera tests demonstrate mineralization via pathway 3 (Weiner & Addadi, 2011). The macromolecules that form the extracellular matrix influence crystal mechanics and texture, while also acting as crystal growth inhibitors (Herman *et al.*, 1988).

Biominerals can be composed of a variety of metals and anions, but in general there are three main classes: calcium carbonates, calcium phosphates, and silicates. Calcium carbonates, frequently in the form of calcite or aragonite, are the most

commonly occurring form of biomineral. The discontinuous distribution of these three classes of biominerals across all of Eukarya suggests numerous instances of convergent evolution, horizontal gene transfer, or some combination of both (Knoll, 2003).

Calcium carbonates are the most commonly produced biominerals in modern oceans. The biogenic formation of carbonates has a profound impact on the global carbon cycle and represents a significant sink of seawater calcium carbonate (Berner, 2004; Lowenstam, 1989). Despite favorable conditions for calcium carbonate precipitation in modern seawater (>10 mM calcium and 2mM bicarbonate), the mineral generally does not spontaneously precipitate—its production requires biological assistance (Drake *et al.*, 2014). Enzymes such as carbonic anhydrase as well as acidic proteins allow living organisms to overcome kinetic barriers required to produce calcium carbonate crystals (Jackson *et al.*, 2007; Le Roy *et al.*, 2014; Lowenstam, 1989; Rodriguez-Navarro *et al.*, 2019).

Biomineralizing proteins

Proteins are fundamental macromolecules responsible for nearly every task of cellular life. They do the work of living by providing cell structure, maintaining pH balance, removing waste, and many other vital functions. Some play critical roles in the biomineralization process. These biomineralizing proteins can nucleate new crystal growth, control ion fluxes, and cease crystallization (Lowenstam, 1989). Biomineralizing proteins are even able to stabilize unusual crystal structures that would otherwise be unstable, allowing for mineral forms that would not precipitate abiotically (Lowenstam, 1989).

Calcium carbonate biominerals produced by eukaryotes are all organo-mineral composites, with the organic phase constituting only a small fraction of the total biomineral. By weight, this generally ranges from less than 0.1% to around 3% (Marin & Luquet, 2007). This organic matrix is composed of a mixture of different macromolecules, including polysaccharides, lipids, and proteins and can be separated into soluble (SOM) and insoluble (IOM) portions. The organic matrix performs essential functions in mineralization (Simkiss & Wilbur, 1989). It can operate like a template for carbonate deposition, influence the growth of calcium carbonate crystals in particular directions, and inhibit growth in others (Lowenstam, 1989).

In the proteinaceous fraction of the organic matrix, unusually acidic proteins are key components (Marin & Luquet, 2007). These proteins tend to be rich in glutamic or aspartic acid (Young, 1971) which provide negatively charged sites at biological pH. The negatively charged sites interact strongly with cations, such as calcium. Due to this, biomineralizing proteins frequently have the ability to bind Ca^{2+} reversibly (Verret *et al.*, 2010) making acidic proteins central to classical hypotheses on biomineralization (Lowenstam, 1989).

Much is known about the amino acid composition of the organic matrix, but there is less knowledge of the characterization of the corresponding proteins (Marin & Luquet, 2007). This is largely due to the technical difficulties in analysis which arise from the unusually highly charged and highly acidic nature of such proteins, which tend not to separate well and are intimately associated with the mineral phase (Gotliv *et al.*, 2003; Tambutté *et al.*, 2007).

Biomineralizing proteins have been identified in the mineral phase of eggshells (Chien *et al.*, 2008), bones, and enamel (Bansal *et al.*, 2012; Termine *et al.*, 1981) in vertebrates. Many have also been identified in invertebrate marine organisms. These include sea urchins (Karakostis *et al.*, 2016; Shashikant *et al.*, 2018), corals (Fukuda *et al.*, 2003; Mass *et al.*, 2013), and molluscs (Bouyoucef *et al.*, 2018; Mann *et al.*, 2012; Politi *et al.*, 2007; Weiner & Hood, 1975). However, of these, few show much sequence similarity to one another (Nagasawa, 2004), complicating research efforts.

For calcium carbonate biominerals, proteins featuring an EF-hand motif are of particular interest. EF-hands are calcium-binding helix–loop–helix structural motifs found in hundreds of proteins (Kretsinger & Nockolds, 1973). Of these, several are known to affect mineralization, such as osteonectin (Bolander *et al.*, 1988; Delany *et al.*, 2000) and the oyster pearl protein EFCBP (Huang *et al.*, 2007). The EF-hand is the most common calcium binding motif found in proteins. Thus, discovering a protein with the EF-hand domain is a promising find in the search for mineralizing proteins. However, calcium is an important metal for life, and is involved in vital processes other than calcium carbonate formation. Proteins with the EF-hand motif, where known, perform diverse functions (Ikura, 1996) and are therefore not necessarily binding calcium for mineralization.

In 1998, Corstjens *et al.* identified a gene which encodes an acid-rich protein in *E. huxleyi* called GPA. It contains 24% acidic residues and a sequence similar to the calcium binding EF-hand domain (Corstjens *et al.*, 1998). GPA is also closely associated with polysaccharides. Due to these factors, it has been suggested that it may be involved in coccolith mineralization. Currently, most of the literature regarding the theorized

coccolithophore proteome only focuses on GPA. However, a 2006 study investigating *E. huxleyi* biomineralizing proteins found no compelling evidence for it being directly involved in calcification, contradicting the notion that GPA is central to the process (Quinn *et al.*, 2006). However, their study does not rule out GPA from participating in some aspect of the process. Finding and characterizing biomineralizing proteins in coccoliths is key to unraveling open questions about the calcification process.

Coccolithophores

Coccolithophores are unicellular phytoplankton that inhabit the photic zone of the entire global ocean (Lowenstam, 1989). They are photosynthetic and thus play a significant role in marine ecosystems as primary producers. They also influence ocean-atmosphere CO_2 exchange through respiration (Rost & Riebesell, 2004). Many species of coccolithophores produce calcium carbonate "scales" that encase the coccolithophore cell, forming the coccosphere. These were first observed and named coccoliths in 1858 by Thomas Huxley (Huxley, 1858). As coccolithophores are found in huge numbers around the globe and are both photosynthesizers as well as calcium carbonate biomineralizers, they are extremely important players in multiple marine biogeochemical cycles.

Despite their small size, research suggests that coccolithophores are responsible for half of all calcium carbonate precipitation in the oceans (Milliman, 1993) and produce about half of all exported marine calcium carbonate (Broecker & Clark, 2009; Schiebel, 2002). It has even been suggested that *Emiliania huxley*i, the most abundant coccolithophorid species, may actually be the most abundant calcium carbonate

producing species on earth (Westbroek *et al.*, 1984). Coccolithophores are able to grow in such numbers as to form blooms that are so large, they can be seen from space (Smyth *et al.*, 2004).

The liths produced by coccolithophores eventually settle to the ocean floor and can be preserved as sediment over time. Coccoliths are resistant to dissolution allowing these sediments to be preserved in huge deposits as chalk, which are responsible for striking geologic features, including the famous White Cliffs of Dover. Additionally, sinking coccoliths can serve as an aggregate material which transports particulate organic carbon from the surface to the deep ocean (de Vargas *et al.*, 2007; Klaas & Archer, 2002).

Calcifying coccolithophores also play an important role in ocean-atmosphere gas exchange. They are photosynthetic, so they capture carbon dioxide and release oxygen. However, the chemical reaction of calcification releases carbon dioxide into the surface waters and atmosphere (de Vargas *et al.*, 2007).

Mineralized coccolithophores first evolved in the early Mesozoic (Harper & Knoll, 1975). The earliest coccolith fossils date to the Triassic Period, roughly 220 million years old (Bown *et al.*, 2004). Coccolithophores began to propagate into the open global ocean during the Cretaceous. During this time, carbonate deposition switched from shallow seas to the ocean depths (Hay, 2004). These vast deposits of coccoliths and other calcareous nannoplankton from this time are responsible for the name Cretaceous, which is derived from the Latin word for chalk. This switch radically altered the marine carbon cycle.

In the time since coccolithophores first began mineralizing, there have been numerous changes to the morphological structures of coccoliths that correlate with major changes in geologic history (de Vargas *et al.*, 2007). The close association of coccolith morphology changes due to mass extinctions or other biological turnover events make coccolith fossils useful for paleoceanographers and chronostratigraphers. This also suggests that perhaps the mode of calcification or even the crystal structures of coccoliths may be associated with a coccolithophore's ability to survive changing environments.

Biologically controlled calcification comes with a high physiological cost. Such a heavy cost implies there must be an advantage to calcification and raises the question: Why do coccolithophores calcify? Montiero et al (2016) suggest the origin of calcification in the coccolithophores was likely to protect against predation by grazers, such as pteropods. However, calcification provides other protections which may have given calcified coccolithophores other evolutionary advantages, such as protections from ultraviolet photodamage and defense from bacterial and viral infection (Monteiro *et al.*, 2016).

Coccolithogenesis

A rather fascinating and unusual feature of coccolith formation is that calcification takes place intracellularly (Paasche, 1962) in alkaline Golgi-derived vesicles (Manton, 1967). The coccolith is then extruded to the outside of the cell (Brownlee & Taylor, 2004; Manton, 1967; Taylor *et al.*, 2007). This method of mineralization requires a great deal of biological control. The coccolithophore cell must maintain net fluxes of Ca^{2+} and inorganic carbon from outside the cell to the intracellular vesicle where the

calcification takes place (Brownlee & Taylor, 2004). Calcium concentration must also be tightly controlled to ensure the mineral does not precipitate as the polymorph apatite (Constantz, 1986).

The coccolithogenesis process is complex, highly controlled, and is not yet fully understood in molecular detail. The liths are first formed by secretion of calcium carbonate inside the specialized vesicle on an organic template base plate (Marsh *et al.*, 1992). The crystal spacing and orientation is precisely controlled forming intricate, interlocking crystals that make up the coccolith, which vary in morphology at the species and intraspecies level (Geisen *et al.*, 2004). The process continues in the vesicle until the coccolith is fully formed. At this stage, the coccolith-containing vesicle migrates toward the outside of the cell and merges with the cell membrane to release the coccolith (Taylor *et al.*, 2007). The complete, fully crystallized coccolith is then incorporated into the coccosphere (Drescher *et al.*, 2012).

The coccoliths remain loosely connected outside of the cell membrane, likely held together by sticky organic macromolecules or by the coccoliths interlocking (Takano *et al.*, 1994; Young & Henriksen, 2003). It has been shown coccoliths in many species are coated in acidic polysaccharides that can also be incorporated into the coccolith mineral phase (Borman *et al.*, 1982; Gal *et al.*, 2016; van der Wal *et al.*, 1983). The polysaccharides have been implicated in calcium binding during lith formation (Marsh & Dickinson, 1997) and protecting the coccosphere from dissolution (Hassenkam *et al.*, 2011).

Emiliania huxleyi

As previously discussed, *Emiliania huxleyi* is the most abundant coccolithophorid species on Earth and has been studied extensively. The genome of *E. huxleyi* strain CCMP1516 was fully sequenced in 2013 by Read et al., providing a new, crucial reference for novel biological studies. Their work predicted 30,569 protein-coding genes and of those, identified dozens of genes potentially involved in biomineralization processes (Read *et al.*, 2013). For these reasons, *E. huxleyi* was selected as the model organism for this study.

METHODS

Culturing coccolithophores

A great deal of coccoliths needed to be grown for the experimental work. The readily available *Emiliania huxleyi* strain CCMP374 was not producing coccoliths, so a culture of strain CCMP371 was obtained from Bigelow laboratories. After growing for a few days, it was verified that strain CCMP371 was producing coccoliths. In culture, the media was cloudy white, and it was confirmed to be calcifying under SEM (Figure 2). Strain CCMP371 proved to be an excellent organism with which to work as it produced plentiful coccoliths readily.

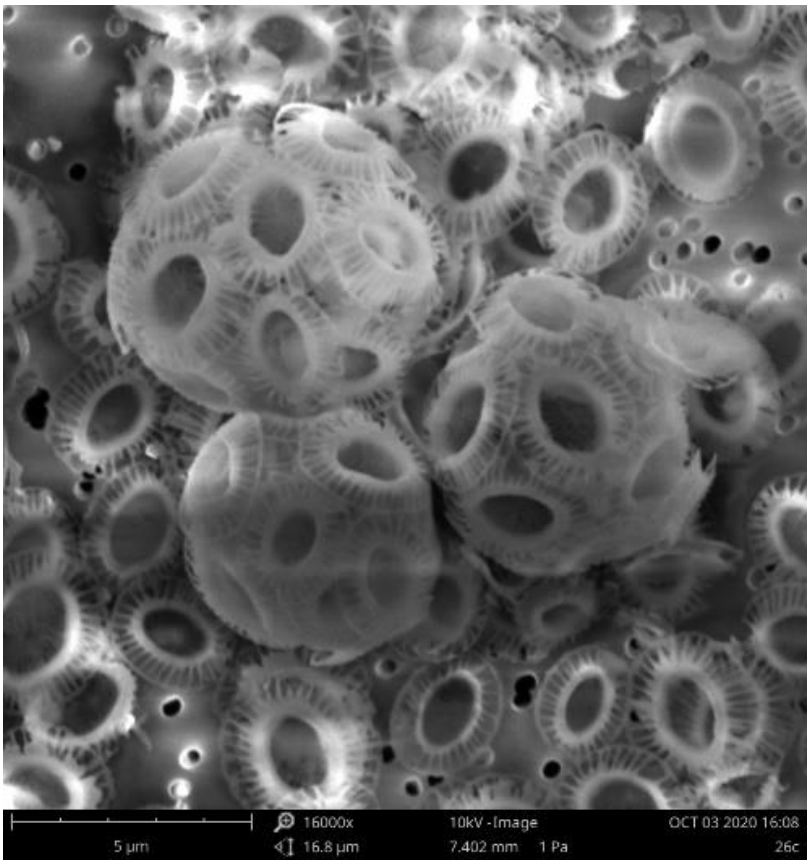

Figure 1. Scanning Electron Microscope image of *E. Huxleyi* strain CCMP371 confirming successful production of coccoliths in culture.

Emiliania huxleyi strain CCMP371 was grown in f/50 media (following Bigelow laboratory's amended Guillard and Ryther recipe) over the course of a year to collect the coccoliths (Guillard, 1975; Guillard & Ryther, 1962). A mass culture was started in four 20 L Nalgene carboys. The growing media was made using filtered natural seawater which was autoclaved to sterilize. The seawater was sourced from the New Jersey south shore at high tide and kept in large tanks at Rutgers Department of Marine and Coastal Sciences building in the dark for several months before being pumped through a sand filter to a holding tank for use. After cooling to room temperature, filter sterilized nutrients were added to make f/50 media.

Roughly 500 mL of previously prepared CCMP371 cultures were added to each carboy aseptically in a laminar flow hood. Each carboy was equipped with a filtered air intake and outlet to allow gas exchange, but no additional air was injected into the cultures. The cultures were placed on magnetic stirrers to prevent the buildup of organic material on the bottom and to encourage coccolithophore growth. They were left undisturbed in a climate-controlled incubator room for seven weeks. The temperature was kept at 26°C, light intensity was 31.6 μmol m^{-2} s^{-1}, and lamps were on a light/dark cycle of 14/10 hours.

Isolating coccoliths

Once the coccolithophores in culture had produced plentiful coccoliths, the liths needed to be collected and isolated. The cultures were spun down to remove the growth media and capture the coccoliths by slow, continuous flow centrifugation. The resulting

concentrated coccolithophore culture was a slimy, green-white mixture of shed coccoliths and coccolithophore cells which was stored in 50 mL Falcon tubes at 4°C.

The concentrated cells were then lysed via osmotic stress to separate any attached coccoliths. The cells were lysed in a 1 M KNO_3 solution following an adapted protocol written by Or Eliason and forwarded by Christopher Johns (personal communication, 10/15/2020). 25 mL of 1 M KNO_3 was added to the tubes containing the concentrated culture and left rocking at 4°C overnight. After the first treatment, the tubes were centrifuged, and it was evident that the liths were separating from the cells. There was an accumulation of white coccoliths at the bottom of the tubes and dark green cellular material in the supernatant. The supernatant was discarded and the KNO_3 treatment was repeated three times to ensure the entirety of the cells were lysed.

The remaining organic material was separated from the coccoliths using Percoll (pH 8.0). Approximately 20 mL of undiluted Percoll was added to the samples and vortexed to mix well. After a quick spin in the centrifuge (1 minute at 3300 rcf), the organics rose to the surface of the Percoll and the coccoliths accumulated at the bottom of the tube. The organic matter was carefully removed and discarded using a pipette. This process was repeated until there was no visible green material in the samples. The Percoll was cleaned off by washing the liths in filter sterilized Milli-Q water three times.

Cleaning coccoliths

The isolated coccoliths were then cleaned further by oxidizing residual organic material. They were first soaked overnight in filter sterilized 10% bleach. They were then

cleaned using a protocol provided by Dr. Jeana Drake (personal communication, 3/2/2021). A solution of 50% H_2O_2 (33%) and 50% NaOCl (6%) was applied to the coccoliths and vortexed. The liths in solution were left overnight then rinsed with Milli-Q water three times. This process was repeated three times to ensure all organic matter was fully oxidized. The clean coccoliths were then flash frozen in liquid nitrogen and lyophilized. The liths were stored dry at room temperature for experimentation.

A fifth 20 L carboy culture was prepared a few months later and processed in the same way as the previous cultures. The coccoliths produced here were used in the second sample sent to proteomics. These coccoliths were cleaned additionally by soaking in a highly concentrated alcoholic hydroxide solution overnight. They were rinsed clean with Milli-Q water ten times and stored in ethanol at 4°C for experimentation. Roughly 10 grams of coccoliths were collected after processing all the cultures.

Sample preparation for proteomic analysis

The two samples were sent to the Rutgers Center for Advanced Biotechnology and Medicine (CABM) for proteomic analysis. The first sample was dissolved in 10% EDTA and given to Dr. Haiyan Zheng for analysis without further concentrating or processing. The samples were analyzed by LC-MS.

The second sample was also dissolved in 10% EDTA. After dissolution, the sample was dialyzed in phosphate buffered saline in a 10K MWCO Thermo Scientific™ Slide-A-Lyzer™ cassette following manufacturer's instructions. The dialyzed sample was concentrated by centrifugation in a 15 mL 3 kDa MilliporeSigma™ Amicon™ filter

following manufacturer's instructions. This second sample was analyzed by the same method as the first sample, but in reverse mode to target any highly acidic proteins.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

The LC-MS/MS work and subsequent database search was performed by Dr. Haiyan Zheng of Rutgers Center for Advanced Biotechnology and Medicine (CABM). Samples were analyzed by LC-MS using Nano LC-MS/MS (Dionex Ultimate 3000 RLSCnano System, Thermofisher) interfaced with Eclipse (ThermoFisher). The following was written by Dr. Zheng with only minor edits by me (personal communication, 11/9/21):

> Three microliters of 12.5 µL in-gel digested using trypsin. Sample P was loaded on to a fused silica trap column (Acclaim PepMap 100, 75UMx2CM, ThermoFisher). After washing for 5 min at 5 µl/min with 0.1% TFA, the trap column was brought in-line with an analytical column (Nanoease MZ peptide BEH C18, 130A, 1.7 µm, 75umx250mm, Waters) for LC-MS/MS. Peptides were fractionated at 300 nL/min using a segmented linear gradient 4-15% B in 30 min (where A: 0.2% formic acid, and B: 0.16% formic acid, 80% acetonitrile), 15-25% B in 40min, 25-50% B in 44min, and 50-90% B in 11min. Solution B then returns at 4% for 5 minutes for the next run.
>
> The scan sequence began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, scan range from M/Z 375–1500, automatic gain control (AGC) target 1E6, maximum injection time 100 ms). The top S (3 sec) duty cycle scheme were used to determine the number of MS/MS performed for each cycle. Parent ions of charge 2-7 were selected for MS/MS and dynamic exclusion of 60 seconds was used to avoid repeat sampling. Parent masses were isolated in the quadrupole with an isolation window of 1.2 m/z, automatic gain control (AGC) target 1E5, and fragmented with higher-energy collisional dissociation with a normalized collision energy of 30%. The fragments were scanned in Orbitrap with resolution of 15,000. The MS/MS scan ranges were determined by the charge state of the parent ion, but lower limit was set at 110 amu.

MS Database Search

The following was written by Dr. Zheng with only minor edits by me (personal communication, 11/9/21):

> The peak list of the LC-MS/MS were generated by Thermo Proteome Discoverer (v. 2.1) into MASCOT Generic Format (MGF) and searched against GCF_000372725_protein and common lab contaminants (CRAP) using an in house version of X!Tandem (GPM Fury) (Craig & Beavis, 2004).
>
> Search parameters were the following: fragment mass error, 20 ppm; parent mass error, +/- 7 ppm; fixed modification, no fixed modification; variable modifications: methionine monooxidation for the primary search, asparagine deamination, tryptophan oxidation and dioxidation, methionine dioxidation, and glutamine to pyro-glutamine were considered at the refinement stage.
>
> Protease specificity: trypsin (C-terminal of R/K unless followed by P with 1 missed cleavage during the preliminary search and 5 missed cleavages during refinement. Minimum acceptable peptide and protein expectation scores were set at 10^{-2} and 10^{-4}, respectively. The overall peptide false positive rate (Gupta *et al.*, 2011) was 0.07%.

Data analysis from mass spectrometry

The protein sequences retrieved from Dr. Zheng's mass spectrometry analysis were analyzed using Blast2GO version 6.0.3 (Conesa & Götz, 2008; Conesa *et al.*, 2005; Götz *et al.*, 2008). Blast2GO is a bioinformatics software tool for the functional annotation of novel sequence data. It uses the Basic Local Alignment Search Tool (BLAST) algorithm to search databases for homologous sequences and transfers existing functional annotation from characterized sequences to novel ones. Each protein's FASTA (Pearson & Lipman, 1988) file containing their amino acid sequence was downloaded from the NCBI website. These files were given as input to Blast2GO, and all functions were run using the program's default parameters, except for changing the E-value cut off from 10^{-6} to 10^{-10} in the annotation step.

The sequences were queried using Blast2Go functions in the following order, BLAST (program Blastp), Gene Ontology (GO) mapping, InterProScan, and Annotation. All query sequences with BLAST hits were assigned with a description that is the most probable function of the detected hits. Annotations were assigned for sequences that qualify the threshold E value of $\leq 10^{-10}$. Available annotations and GO terms from the BLAST search were statistically analyzed by Blast2GO and the most probable annotations were assigned to the sequences. These annotations are presented in Table 1 in the Appendix section. Protein sequences without any Blast hits are left unannotated.

For each protein, a simple calculation was performed to quantify the percent acidity from the sequences contained in their FASTA files. The total of aspartic acid (D) and glutamic acid (E) was divided by the total number of amino acids in each sequence to determine if any were particularly acid rich.

Additional tests for protein detection

Other methods were employed to test for the presence of protein in the coccoliths in addition to mass spectrometry. The Pierce™ BCA Protein Assay Kit was used to measure the concentration of protein in a sample of dissolved, dialyzed, and concentrated coccoliths following the microplate procedure included in the kit. The assay was spectrophotometrically analyzed on a microplate reader at 562 nm. The carbon/nitrogen ratio of samples of cleaned coccoliths, whole coccolithophore cells, decalcified cells, and decalcified coccoliths were measured using a Carlo Erba Instruments NA 1500 Series 2 nitrogen/carbon/sulfur analyzer.

A DeNovix NanoDrop spectrophotometer was used as a first pass indicator for the presence of protein throughout this project as it only requires a single microliter drop of sample to analyze. Samples analyzed with the NanoDrop were prepared in different ways and had varying levels of concentrations. Each sample was measured for absorbance at 280nm wavelength using the Protein A280 setting.

Gel electrophoresis was performed on two occasions. On the first attempt, a sample of liths dissolved in 10% EDTA and concentrated in an Amicon filter without dialysis was tested (sample 1). The second run tested samples which had been prepared using three different methods—by dissolving liths in EDTA and concentrating (sample 2); dissolving liths in EDTA, dialyzing, then concentrating (sample 3); and dissolving the liths in acetic acid and concentrating (sample 4). Each attempt was run using Bio-Rad 50 µL 10 well Mini-PROTEAN TGX precast stain-free polyacrylamide gels following manufacturer's instructions. The gels were imaged using Bio-Rad ChemiDoc MP Imaging System using the "best sensitivity" protocol, optimized for low protein concentrations. The second attempt of gel electrophoresis was also silver stained using the Pierce™ Silver Stain for Mass Spectrometry kit.

RESULTS

Coccolith processing

The clean, lyophilized coccoliths were examined under SEM to confirm the samples were free of any remnant cellular material (**Figure 2**). There is evidence of clumping, which may be due to lingering organics, or electrostatic charge.

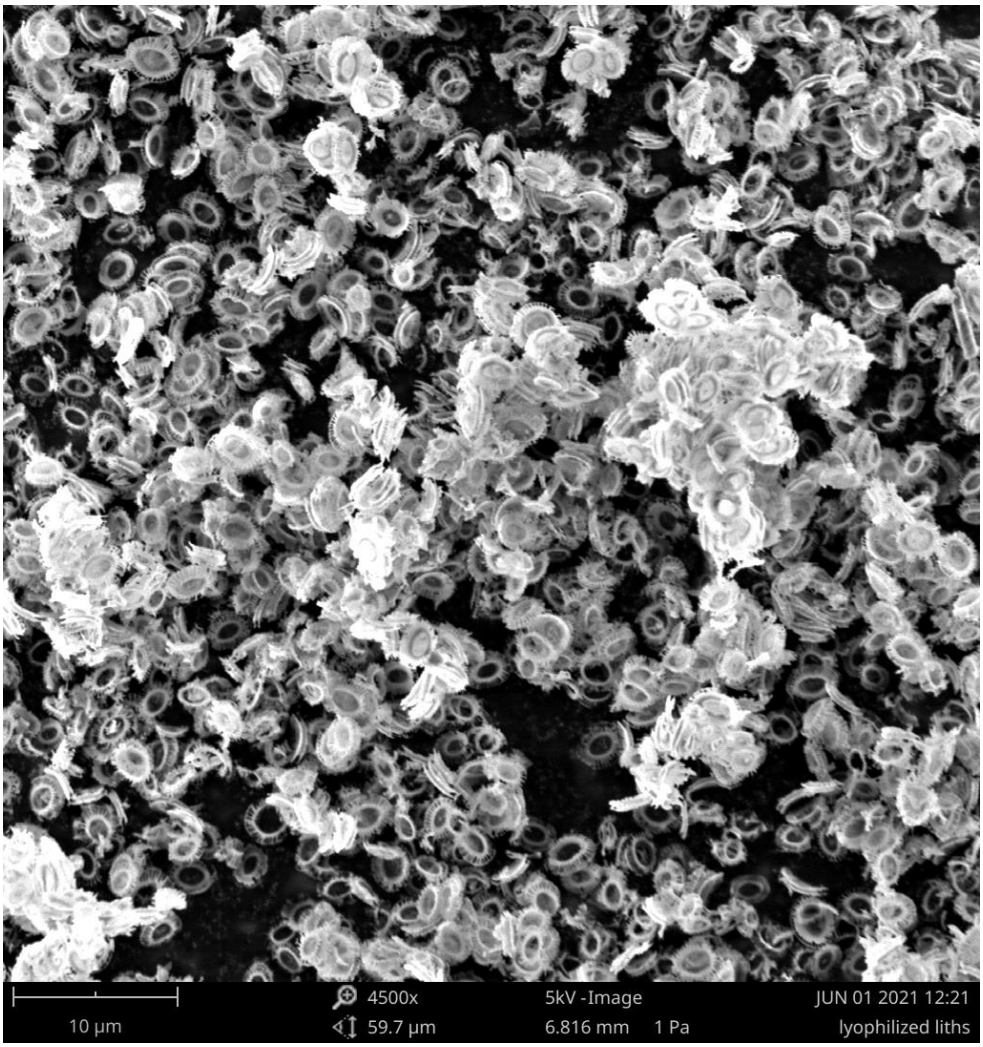

Figure 2. SEM image of cleaned, lyophilized liths.

Mass Spectrometry

The first run through mass spectrometry returned 100 discrete proteins. Of these, 23 did not have any functional annotations, suggesting these proteins may be unique to coccolithophores. This result is not too surprising as there are relatively few phytoplankton species which have been fully sequenced as well as the lack of knowledge about proteins involved in biomineralization processes across species. The 77 distinctive functional annotations that were assigned to the rest were mostly localized to the cell membrane. There were no extracellular GO terms, suggesting minimal contamination in the sample. Two proteins were somewhat acid rich at 20-23% (P46 and P55, Table 1). All 100 proteins averaged out at 11% acidic. However, the sequences in the FASTA files retrieved from NCBI may not be completely predicted as they are all lacking the "*" at the end, which signifies the stop codon. Most begin with M for methionine, which is expected, but the C-terminal end is not explicitly indicated. The calculations were conducted assuming the sequences are correct and complete, but there is inherent uncertainty due to the missing stop symbol.

The second sample which was run in reverse mode only returned five discrete proteins. All five of these were also detected in the first run—no additional proteins were acquired. These five all had functional annotations; three are histones and two are putative actins. Histones and actins do not appear to be involved in biomineralizing processes and may be human contaminants. These proteins are not highly acidic either, ranging from only 7-13%.

The mass spectrometry, bioinformatic, and acidity calculation results are included in the appendix (Table 2 and Table 3). Table 2 lists the 77 proteins with functional

annotations and Table 3 lists the 23 proteins with no annotations. The most notable results are presented here and described in the following discussion:

Table 1. notable results from mass spectrometry and bioinformatic analysis

Protein	Accession number	Description	% acidic
P10	XP_005762630	putative death-specific protein with Ca binding EF hand domain	17.5
P15	XP_005757874	protein disulfide isomerase	15.2
P31	XP_005785519	protein disulfide-isomerase	12.8
P36	XP_005783192	heat shock cognate 70 kDa protein 2-like	15.5
P46	XP_005764796	heat shock protein 90	20.3
P55	XP_005766276	pentapeptide repeat-containing protein	23.0
P71	XP_005785898	putative lipoprotein	7.9
P72	XP_005773365	putative membrane protein	14.6
P76	XP_005763039	transmembrane protein 87A-like	8.3
P77	XP_005778320	V-type proton ATPase subunit B, brain isoform	13.2

Returned sequences with e-values $\leq 10^{-10}$ are presented in order of decreasing e-value. "Description" is the associated annotation in Blast2GO.

Additional tests for protein

The sample tested using the BCA assay turned visibly purple, indicative of a protein signal. From spectrophotometry analysis, the sample tested had a protein concentration of 19.1 μg/mL.

Samples analyzed on the NanoDrop ranged from 0.032 – 6.501 mg/mL of protein. The wide range of values is due to testing different preparation methods and varying degrees of concentration. The highest measured concentration of 6.501 mg/mL was from the second sample sent to CABM for mass spectrometry analysis in reverse mode.

In both attempts of running gel electrophoresis, no discernable bands appeared. The first attempt shows a smeared band near the terminal end when the contrast was high (Figure 2). The second attempt that tested three preparation methods also did not show any nice, clear bands but smeared and bulging lanes, also at high contrast image (Figure

3). A silver stain was applied to the second gel as an attempt to bring out bands of very low protein concentration but was also unsuccessful due to stopping the development too early.

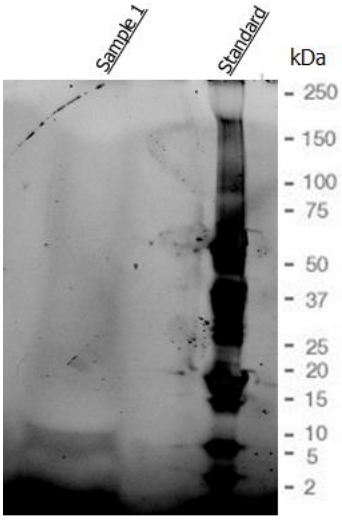

Figure 3. First gel electrophoresis run result, sample 1. No clear bands. Shown at high contrast, a wide, smeared, somewhat visible band appears at the terminal end.

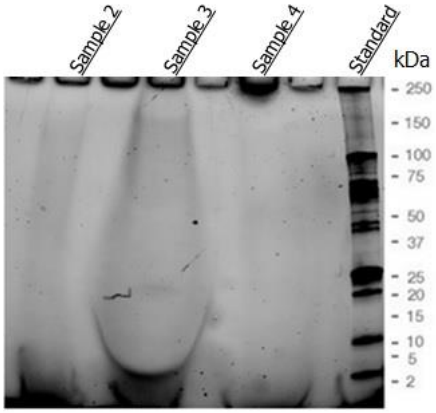

Figure 4. Second gel electrophoresis run result shown in high contrast. Samples 2 - 4 are shown in lanes to the left with the standard ladder on the right. The bulging in the middle lane is peculiar but not unheard of in gel electrophoresis. No discernable bands.

DISCUSSION

This work strongly suggests that proteins are indeed in the coccolith organic matrix, albeit in a very small quantity. While polysaccharides are certainly facilitating crystallization, proteins are quite likely also involved. As discussed earlier, the liths are formed within a vesicle, move through the cell fully formed, and are then excocytosed. This intracellular formation process provides ample opportunity for the mineral structure to incorporate proteins along the way. This work is but a first step in the study of coccolith biomineralizing proteins, but it lays the groundwork for further investigation.

From my analyses and research on the function of proteins annotated by Blast2GO, I have found six promising proteins that should be more thoroughly investigated. These are a putative calcium binding protein, membrane proteins, heat shock proteins, lipoproteins, disulfide isomerase, and a V-type H+-ATPase. We must also consider the 23 proteins with no annotations at all which could possibly be something worth investigating. Additionally, we can rule out many proteins which are not likely involved in mineralization, such as light harvesting proteins, to narrow the focus of future studies.

Certainly, a protein with a calcium binding EF hand domain is a candidate for being involved in the calcium carbonate mineralization process. This protein is described in annotation as a “putative death-specific protein with Ca binding EF hand domain” and its GO function is “calcium ion binding.” The protein is also slightly acidic at 17.5%, which is higher than average but not exactly acid rich. However, its description as a death-specific protein is less suggestive of an association with mineralization. Searching

the literature did not produce any evidence that death-specific proteins are related to biomineralization.

Membrane and transmembrane proteins may also be of interest. Transmembrane proteins permit the transport of substances across the cell membrane. As coccoliths are made intracellularly inside an organelle, perhaps these proteins play a role in facilitating the coccolith extrusion from the inside of the cell outward to the exterior. Obif and Bril are transmembrane proteins with published connections to biomineralization (Mizuhashi *et al.*, 2015; Moffatt *et al.*, 2008). The membrane proteins are unlikely to be directly involved in mineral precipitation but could possibly be engaged in some aspect of coccolith mineralization.

The other proteins that stand out as potentially having something to do with mineralization are also more likely to have only peripheral roles in the process, if they have any association at all. Searching the literature resulted in some interesting connections to mineralization, although in distantly related species. Heat shock proteins have been found to be involved in bone metabolism and bone-related diseases (Hang *et al.*, 2018). Lipoproteins have been implicated in pathological calcification in humans (Bouchareb *et al.*, 2015; Hara *et al.*, 2021). A study on protein disulfide-isomerase showed a connection to mineralization in osteoblast-like cells (Chen *et al.*, 2010) and subsequent work demonstrated that in mice, reducing levels of a disulfide-isomerase resulted in impaired bone formation (Wang *et al.*, 2014). Finally, V-type H+-ATPase has been found to affect calcification in many organisms, including a terrestrial isopod (Ziegler *et al.*, 2004), sea urchin embryos (Hu *et al.*, 2020), and the giant clam, *Tridacna squamosa* (Ip *et al.*, 2018).

There are some limitations of this study and room for improvement in the laboratory work. For one, coccoliths are relatively difficult to collect in the quantities needed for proteomic analyses. It has been a limiting factor and likely a reason why such a study has not been published to date. The supply of coccoliths can limit how much experimentation one can do; their use must be conservative unless there is a huge quantity with which to work. Also, while the processed liths appeared to be clean in SEM image, I cannot be certain they were entirely free of cellular material. Additionally, only the soluble matrix was analyzed, there could be important proteins in the insoluble portion of the coccoliths which were discarded. The cost of mass spectrometry analysis can also be prohibitive, limiting replication and access to interested researchers who may not have sufficient funding.

Despite these limitations, there should be additional studies on the coccolith proteome. It would be ideal to repeat this experiment to verify the results. Other coccolithophore species must also be analyzed and compared with this work, which is currently being undertaken by Dr. Jeana Drake at UCLA. The full genomes of eight strains of coccolithophores belonging to the genus *Gephyrocapsa* were recently sequenced (Bendif *et al.*, 2019), providing an opportunity to begin a cross-species comparative study of *E. huxleyi* and *Gephyrocapsa*. Dr. Drake is working on analyzing the coccolith proteome of *Gephyrocapsa* for a future collaborative publication.

Additional work can isolate the proteins of particular interest for cloning and further investigation to verify and better understand their function with respect to biomineralization. Phyloproteomic analyses of coccolithophore and other marine species may reveal more about the origin of biomineralization and ultimately help fully

understand the process. While this study could not definitively identify a biomineralizing protein in the coccolith, it can serve to help progress the ongoing study of biomineralization.

REFERENCES CITED

Bansal, A. K., Shetty, D. C., Bindal, R., & Pathak, A. (2012). Amelogenin: A novel protein with diverse applications in genetic and molecular profiling. *Journal of oral and maxillofacial pathology : JOMFP, 16*(3), 395-399. doi:10.4103/0973-029X.102495

Bendif, E. M., Nevado, B., Wong, E. L. Y., Hagino, K., Probert, I., Young, J. R., Rickaby, R. E. M., & Filatov, D. A. (2019). Repeated species radiations in the recent evolution of the key marine phytoplankton lineage Gephyrocapsa. *Nature communications, 10*(1), 4234-4239. doi:10.1038/s41467-019-12169-7

Berner, R. A. (2004). *Phanerozoic Carbon Cycle: CO2 and O2*. Cary: Oxford University Press.

Bolander, M. E., Young, M. F., Fisher, L. W., Yamada, Y., & Termine, J. D. (1988). Osteonectin cDNA Sequence Reveals Potential Binding Regions for Calcium and Hydroxyapatite and Shows Homologies with Both a Basement Membrane Protein (SPARC) and a Serine Proteinase Inhibitor (Ovomucoid). *Proceedings of the National Academy of Sciences - PNAS, 85*(9), 2919-2923. doi:10.1073/pnas.85.9.2919

Borman, A. H., Jong, E. W., Huizinga, M., Kok, D. J., Westbroek, P., & Bosch, L. (1982). The Role in CaCO3 Crystallization of an Acid Ca2+-Binding Polysaccharide Associated with Coccoliths of Emiliania huxleyi. *European journal of biochemistry, 129*(1), 179-183. doi:10.1111/j.1432-1033.1982.tb07037.x

Bouchareb, R., Mahmut, A., Nsaibia, M. J., Boulanger, M.-C., Dahou, A., Lépine, J.-L., Laflamme, M.-H., Hadji, F., Couture, C., Trahan, S., Pagé, S., Bossé, Y., Pibarot, P., Scipione, C. A., Romagnuolo, R., Koschinsky, M. L., Arsenault, B. J., Marette, A., & Mathieu, P. (2015). Autotaxin Derived From Lipoprotein(a) and Valve Interstitial Cells Promotes Inflammation and Mineralization of the Aortic Valve. *Circulation (New York, N.Y.), 132*(8), 677-690. doi:10.1161/CIRCULATIONAHA.115.016757

Bouyoucef, M., Rakic, R., Gómez-Leduc, T., Latire, T., Marin, F., Leclercq, S., Carreiras, F., Serpentini, A., Lebel, J.-M., Galéra, P., & Legendre, F. (2018). Regulation of Extracellular Matrix Synthesis by Shell Extracts from the Marine Bivalve Pecten maximus in Human Articular Chondrocytes— Application for Cartilage Engineering. *Marine biotechnology (New York, N.Y.), 20*(4), 436-450. doi:10.1007/s10126-018-9807-7

Bown, P. R., Lees, J. A., & Young, J. R. (2004). Calcareous nannoplankton evolution and diversity through time. In *Coccolithophores* (pp. 481-508): Springer.

Broecker, W., & Clark, E. (2009). Ratio of coccolith CaCO3 to foraminifera CaCO3 in late Holocene deep sea sediments. *Paleoceanography, 24*(3), PA3205-n/a. doi:10.1029/2009PA001731

Brownlee, C., & Taylor, A. (2004). Calcification in coccolithophores: a cellular perspective. In *Coccolithophores* (pp. 31-49): Springer.

Chen, C.-C., Boskey, A. L., & Rosenberg, L. C. (1984). The inhibitory effect of cartilage proteoglycans on hydroxyapatite growth. *Calcified Tissue International, 36*(1), 285-290. doi:10.1007/BF02405332

Chen, J., Olivares-Navarrete, R., Wang, Y., Herman, T. R., Boyan, B. D., & Schwartz, Z. (2010). Protein-disulfide Isomerase-associated 3 (Pdia3) Mediates the Membrane Response to 1,25-Dihydroxyvitamin D3 in Osteoblasts. *The Journal of biological chemistry, 285*(47), 37041-37050. doi:10.1074/jbc.M110.157115

Chien, Y. C., Hincke, M. T., Vali, H., & McKee, M. D. (2008). Ultrastructural matrix–mineral relationships in avian eggshell, and effects of osteopontin on calcite growth in vitro. *Journal of structural biology, 163*(1), 84-99. doi:10.1016/j.jsb.2008.04.008

Conesa, A., & Götz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International journal of plant genomics, 2008*, 619832-619812. doi:10.1155/2008/619832

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674-3676. doi:10.1093/bioinformatics/bti610

Constantz, B. R. (1986). Coral skeleton construction; a physiochemically dominated process. *Palaios, 1*(2), 152-157. doi:10.2307/3514508

Corstjens, P. L. A. M., Van Der Kooij, A., Linschooten, C., Brouwers, G.-J., Westbroek, P., & Jong, E. W. d. V.-d. (1998). GPA, A CALCIUM-BINDING PROTEIN IN THE COCCOLITHOPHORID EMILIANIA HUXLEYI (PRYMNESIOPHYCEAE). *Journal of phycology, 34*(4), 622-630. doi:10.1046/j.1529-8817.1998.340622.x

Craig, R., & Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics, 20*(9), 1466-1467. doi:10.1093/bioinformatics/bth092

Dauphin, Y. (2001). Comparative studies of skeletal soluble matrices from some Scleractinian corals and Molluscs. *International journal of biological macromolecules, 28*(4), 293-304. doi:10.1016/S0141-8130(01)00124-6

de Vargas, C., Aubry, M.-P., Probert, I., & Young, J. (2007). Chapter 12 - Origin and Evolution of Coccolithophores: From Coastal Hunters to Oceanic Farmers. In (pp. 251,VIII-285,XI): Elsevier Inc.

Delany, A. M., Amling, M., Priemel, M., Howe, C., Baron, R., & Canalis, E. (2000). Osteopenia and decreased bone formation in osteonectin-deficient mice. *The Journal of clinical investigation, 105*(9), 1325-1325. doi:10.1172/JCI7039C1

Drake, J. L., Mass, T., & Falkowski, P. G. (2014). The evolution and future of carbonate precipitation in marine invertebrates: Witnessing extinction or documenting resilience in the Anthropocene? *Elementa: Science of the Anthropocene, 000030*(1). doi:10.12952/journal.elementa.000026

Drescher, B., Dillaman, R. M., & Taylor, A. R. (2012). Coccolithogenesis In Scyphosphaera apsteinii (Prymnesiophyceae). *Journal of phycology, 48*(6), 1343-1361. doi:10.1111/j.1529-8817.2012.01227.x

Fukuda, I., Ooki, S., Fujita, T., Murayama, E., Nagasawa, H., Isa, Y., & Watanabe, T. (2003). Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton. *Biochemical and biophysical research communications, 304*(1), 11-17. doi:10.1016/S0006-291X(03)00527-8

Gal, A., Wirth, R., Kopka, J., Fratzl, P., Faivre, D., & Scheffel, A. (2016). Macromolecular recognition directs calcium ions to coccolith mineralization sites. *Science (American Association for the Advancement of Science), 353*(6299), 590-593. doi:10.1126/science.aaf7889

Geisen, M., Young, J., Probert, I., Sáez, A., Baumann, K.-H., Bollmann, J., Cros Miguel, L., C, D., Medlin, L., & Sprengel, C. (2004). Species level variation in coccolithophores. In *Coccolithophores: Molecular processes to global impac* (pp. 327-366): Springer.

Gotliv, B.-A., Addadi, L., & Weiner, S. (2003). Mollusk Shell Acidic Proteins: In Search of Individual Functions. *Chembiochem : a European journal of chemical biology, 4*(6), 522-529. doi:10.1002/cbic.200200548

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research, 36*(10), 3420-3435. doi:10.1093/nar/gkn176

Guillard, R. R. (1975). Culture of phytoplankton for feeding marine invertebrates. In *Culture of marine invertebrate animals* (pp. 29-60): Springer.

Guillard, R. R., & Ryther, J. H. (1962). Studies of marine planktonic diatoms: I. Cyclotella nana Hustedt, and Detonula confervacea (Cleve) Gran. *Canadian journal of microbiology, 8*(2), 229-239.

Gupta, N., Bandeira, N., Keich, U., & Pevzner, P. A. (2011). Target-Decoy Approach and False Discovery Rate: When Things May Go Wrong. *Journal of the American Society for Mass Spectrometry, 22*(7), 1111-1120. doi:10.1007/s13361-011-0139-3

Hang, K., Ye, C., Chen, E., Zhang, W., Xue, D., & Pan, Z. (2018). Role of the heat shock protein family in bone metabolism. *Cell stress & chaperones, 23*(6), 1153-1164. doi:10.1007/s12192-018-0932-z

Hara, E. S., Oozawa, S., Nagaoka, N., Okada, M., Kasahara, S., & Matsumoto, T. (2021). Nanostructural analysis of distinct nucleation sites in pathological mineralization. *Materials advances, 2*(13), 4423-4431. doi:10.1039/D1MA00273B

Harper, Jr., & Knoll, A. H. (1975). Silica, diatoms, and Cenozoic radiolarian evolution. *Geology (Boulder), 3*(4), 175-177. doi:10.1130/0091-7613(1975)3<175:SDACRE>2.0.CO;2

Hassenkam, T., Johnsson, A., Bechgaard, K., & Stipp, S. L. S. (2011). Tracking single coccolith dissolution with picogram resolution and implications for CO_2 sequestration and ocean acidification. *Proceedings of the National Academy of Sciences - PNAS, 108*(21), 8571-8576. doi:10.1073/pnas.1009447108

Hay W.W. (2004) Carbonate fluxes and calcareous nannoplankton. In: Thierstein H.R., Young J.R. (eds) Coccolithophores. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-06278-4_19

Herman, A., Addadi, L., & Weiner, S. (1988). Interactions of sea-urchin skeleton macromolecules with growing calcite crystals— a study of intracrystalline proteins. *Nature, 331*(6156), 546. doi:10.1038/331546a0

Hu, M. Y., Petersen, I., Chang, W. W., Blurton, C., & Stumpp, M. (2020). Cellular bicarbonate accumulation and vesicular proton transport promote calcification in the sea urchin larva. *Proceedings of the Royal Society. B, Biological sciences, 287*(1934), 20201506-20201506. doi:10.1098/rspb.2020.1506

Huang, J., Zhang, C., Ma, Z., Xie, L., & Zhang, R. (2007). A novel extracellular EF-hand protein involved in the shell formation of pearl oyster. *Biochimica et biophysica acta. General subjects, 1770*(7), 1037-1044. doi:10.1016/j.bbagen.2007.03.006

Huxley, T. (1858). Deep Sea Soundings in the North Atlantic Ocean Between Ireland and Newfoundland: Made in HMS Cyclops, Lieut.-Commander Joseph Dayman, in June and July 1857. In: Eyre and Spottiswoode.

Ikura, M. (1996). Calcium binding and conformational response in EF-hand proteins. *Trends in biochemical sciences (Amsterdam. Regular ed.), 21*(1), 14-17. doi:10.1016/S0968-0004(06)80021-6

Ip, Y. K., Hiong, K. C., Lim, L. J. Y., Choo, C. Y. L., Boo, M. V., Wong, W. P., Neo, M. L., & Chew, S. F. (2018). Molecular characterization, light-dependent expression, and cellular localization of a host vacuolar-type H+-ATPase (VHA) subunit A in the giant clam, Tridacna squamosa, indicate the involvement of the host VHA in the uptake of inorganic carbon and its supply to the symbiotic zooxanthellae. *Gene, 659*, 137-148. doi:10.1016/j.gene.2018.03.054

Jackson, D. J., Macis, L., Reitner, J., Degnan, B. M., & Wörheide, G. (2007). Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science (New York, N.Y.), 316*(5833), 1893-1895. doi:10.1126/science.1141560

Karakostis, K., Zanella-Cléon, I., Immel, F., Guichard, N., Dru, P., Lepage, T., Plasseraud, L., Matranga, V., & Marin, F. (2016). A minimal molecular toolkit for mineral deposition? Biochemistry and proteomics of the test matrix of adult specimens of the sea urchin Paracentrotus lividus. *Journal of proteomics, 136*, 133-144. doi:10.1016/j.jprot.2016.01.001

Klaas, C., & Archer, D. E. (2002). Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global biogeochemical cycles, 16*(4), 63-61-63-14. doi:10.1029/2001GB001765

Knoll, A. H. (2003). Biomineralization and Evolutionary History. *Reviews in Mineralogy and Geochemistry, 54*(1), 329-356. doi:10.2113/0540329

Kretsinger, R. H., & Nockolds, C. E. (1973). Carp muscle calcium-binding protein. II. Structure determination and general description. *The Journal of biological chemistry, 248*(9), 3313-3326.

Le Roy, N., Jackson, D. J., Marie, B., Ramos-Silva, P., & Marin, F. (2014). The evolution of metazoan [alpha]-carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in Zoology, 11*(1). doi:10.1186/s12983-014-0075-8

Lee, R. B. Y., Mavridou, D. A. I., Papadakos, G., McClelland, H. L. O., & Rickaby, R. E. M. (2016). The uronic acid content of coccolith-associated polysaccharides provides insight into coccolithogenesis and past climate. *Nature communications, 7*(1), 13144-13144. doi:10.1038/ncomms13144.

Lowenstam, H. A. (1989). On biomineralization. New York: Oxford University Press.

Lowenstam, H. A., & Weiner, S. (1985). Transformation of amorphous calcium phosphate to crystalline dahllite in the radular teeth of chitons. *Science, 227*, 51.

Luo, G., Ducy, P., McKee, M. D., Pinero, G. J., Loyer, E., Behringer, R. R., & Karsenty, G. (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature (London), 386*(6620), 78-81. doi:10.1038/386078a0

Mann, K., Edsinger-Gonzales, E., & Mann, M. (2012). In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet Lottia gigantea. *Proteome science, 10*(1), 28-28. doi:10.1186/1477-5956-10-28

Manton, I. (1967). Further observations on the fine structure of Chrysochromulina chiton with special reference to the haptonema, 'peculiar' golgi structure and scale production. *Journal of cell science, 2*(2), 265-272.

Marin, F., Luquet, G., (2007). Unusually acidic proteins in biomineralization. Handbook of Biomineralization. Wiley VCH Verlag GmbH, Weinheim, Germany, pp.273-290.

Marin, F., Smith, M., Isa, Y., Muyzer, G., & Westbroek, P. (1996). Skeletal matrices, muci, and the origin of invertebrate calcification. *Proceedings of the National Academy of Sciences of the United States of America, 93*(4), 1554-1559. doi:10.1073/pnas.93.4.1554

Marsh, M. E., Chang, D. K., & King, G. C. (1992). Isolation and characterization of a novel acidic polysaccharide containing tartrate and glyoxylate residues from the mineralized scales of a unicellular coccolithophorid alga Pleurochrysis carterae. *The Journal of biological chemistry, 267*(28), 20507-20512. doi:10.1016/S0021-9258(19)88731-8

Marsh, M. E., & Dickinson, D. P. (1997). Polyanion-mediated mineralization — mineralization in coccolithophore (Pleurochrysis carterae) variants which do not express PS2, the most abundant and acidic mineral-associated polyanion in wild-type cells. *Protoplasma, 199*(1), 9-17. doi:10.1007/BF02539801

Mass, T., Drake, Jeana L., Haramaty, L., Kim, J. D., Zelzion, E., Bhattacharya, D., & Falkowski, Paul G. (2013). Cloning and Characterization of Four Novel Coral Acid-Rich Proteins that Precipitate Carbonates In Vitro. *Current Biology, 23*(12), 1126-1131. doi:10.1016/j.cub.2013.05.007

Michenfelder, M., Fu, G., Lawrence, C., Weaver, J. C., Wustman, B. A., Taranto, L., Evans, J. S., & Morse, D. E. (2003). Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains. *Biopolymers, 70*(4), 522-533. doi:10.1002/bip.10536

Milliman, J. D. (1993). Production and accumulation of calcium carbonate in the ocean: Budget of a nonsteady state. *Global biogeochemical cycles, 7*(4), 927-957. doi:10.1029/93GB02524

Mizuhashi, K., Chaya, T., Kanamoto, T., Omori, Y., & Furukawa, T. (2015). Obif, a Transmembrane Protein, Is Required for Bone Mineralization and Spermatogenesis in Mice. *PLoS One, 10*(7), e0133704-e0133704. doi:10.1371/journal.pone.0133704

Moffatt, P., Gaumond, M. H., Salois, P., Sellin, K., Bessette, M. C., Godin, É., de Oliveira, P. T., Atkins, G. J., Nanci, A., & Thomas, G. (2008). Bril: A Novel Bone-Specific Modulator of Mineralization. *Journal of bone and mineral research, 23*(9), 1497-1508. doi:10.1359/jbmr.080412

Monteiro, F. M., Bach, L. T., Brownlee, C., Bown, P., Rickaby, R. E. M., Poulton, A. J., Tyrrell, T., Beaufort, L., Dutkiewicz, S., Gibbs, S., Gutowska, M. A., Lee, R., Riebesell, U., Young, J., & Ridgwell, A. (2016). Why marine phytoplankton calcify. *Science Advances, 2*(7), e1501822-e1501822. doi:10.1126/sciadv.1501822

Nagasawa, H. (2004). Macromolecules in Biominerals of Aquatic Organisms. *Thalassas : revista de ciencias del mar, 20*(1), 15-24.

Paasche, E. (1962). Coccolith Formation. *Nature (London), 193*(4820), 1094-1095. doi:10.1038/1931094b0

Pearson, W. R., & Lipman, D. J. (1988). Improved Tools for Biological Sequence Comparison. *Proceedings of the National Academy of Sciences - PNAS, 85*(8), 2444-2448. doi:10.1073/pnas.85.8.2444

Politi, Y., Mahamid, J., Goldberg, H., Weiner, S., & Addadi, L. (2007). Asprich mollusk shell protein: in vitro experiments aimed at elucidating function in CaCO3 crystallization. *CrystEngComm, 9*(12), 1171-1177. doi:10.1039/B709749B

Puverel, S., Tambutté, E., Pereira-Mouriès, L., Zoccola, D., Allemand, D., & Tambutté, S. (2005). Soluble organic matrix of two scleractinian corals: partial and comparative analysis. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology, 141B (4)*, 480-487.

Quinn, P., Bowers, R. M., Xiaoyu, Z., Wahlund, T. M., Fanelli, M. A., Olszova, D., & Read, B. A. (2006). cDNA Microarrays as a Tool for Identification of Biomineralization Proteins in the Coccolithophorid Emiliania huxleyi (Haptophyta). *Applied and Environmental Microbiology, 72*(8), 5512-5526. doi:10.1128/AEM.00343-06

Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A., Young, J., Aguilar, M., Claverie, J.-M., Frickenhaus, S., Gonzalez, K., Herman, E. K., Lin, Y.-C., Napier, J., Ogata, H., Sarno, A. F., Shmutz, J., Schroeder, D., de Vargas, C., Verret, F., von Dassow, P., Valentin, K., Van de Peer, Y., Wheeler, G., Dacks, J. B., Delwiche, C. F., Dyhrman, S. T., Glöckner, G., John, U., Richards, T., Worden, A. Z., Zhang, X., & Grigoriev, I. V. (2013). Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature (London), 499*(7457), 209-213. doi:10.1038/nature12221

Rodriguez-Navarro, C., Cizer, Z., Kudacz, K., Ibaez-Velasco, A., Ruiz-Agudo, C., Elert, K., Burgos-Cara, A., & Ruiz-Agudo, E. (2019). The multiple roles of carbonic

anhydrase in calcium carbonate mineralization. *CrystEngComm, 21*(48), 7407-7423. doi:10.1039/c9ce01544b

Rost, B., & Riebesell, U. (2004). Coccolithophores and the biological pump: responses to environmental changes. In *Coccolithophores* (pp. 99-125): Springer.

Schiebel, R. (2002). Planktic foraminiferal sedimentation and the marine calcite budget. *Global biogeochemical cycles, 16*(4), 3-1-3-21. doi:10.1029/2001GB001459

Shashikant, T., Khor, J. M., & Ettensohn, C. A. (2018). From genome to anatomy: The architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. *Genesis (New York, N.Y. : 2000), 56*(10), e23253-n/a. doi:10.1002/dvg.23253

Simkiss, K., & Wilbur, K. M. (1989). *Biomineralization : cell biology and mineral deposition*. San Diego: Academic Press.

Smyth, T. J., Tyrrell, T., & Tarrant, B. (2004). Time series of coccolithophore activity in the Barents Sea, from twenty years of satellite imagery: COCCOLITHOPHORE TIME SERIES. *Geophysical research letters, 31*(11), n/a. doi:10.1029/2004GL019735

Takano, H., Jeon, J., Burgess, J. G., Manabe, E., Izumi, Y., Okazaki, M., & Matsunaga, T. (1994). Continuous production of extracellular ultrafine calcite particles by the marine coccolithophorid alga Pleurochrysis carterae. *Applied microbiology and biotechnology, 40*(6), 946-950.

Tambutté, S., Tambutté, E., Zoccola, D., & Allemand, D. (2007). Organic Matrix and Biomineralization of Scleractinian Corals. In *Handbook of Biomineralization* (Vol. 1, pp. 243-259): WILEY-VCH.

Taylor, A. R., Russell, M. A., Harper, G. M., Collins, T. f. T., & Brownlee, C. (2007). Dynamics of formation and secretion of heterococcoliths by Coccolithus pelagicus ssp. braarudii. *European journal of phycology, 42*(2), 125-136. doi:10.1080/09670260601159346

Termine, J. D., Kleinman, H. K., Whitson, S. W., Conn, K. M., McGarvey, M. L., & Martin, G. R. (1981). Osteonectin, a bone-specific protein linking mineral to collagen. *Cell, 26*(1), 99-105. doi:10.1016/0092-8674(81)90037-4

van der Wal, P., de Jong, E. W., Westbroek, P., de Bruijn, W. C., & Mulder-Stapel, A. A. (1983). Polysaccharide localization, coccolith formation, and golgi dynamics in the coccolithophorid Hymenomonas carterae. *Journal of ultrastructure research, 85*(2), 139-158. doi:10.1016/S0022-5320(83)90103-X

Verret, F., Wheeler, G., Taylor, A. R., Farnham, G., & Brownlee, C. (2010). Calcium channels in photosynthetic eukaryotes: implications for evolution of calcium-

based signalling. *The New phytologist, 187*(1), 23-43. doi:10.1111/j.1469-8137.2010.03271.x

Wang, Y., Nizkorodov, A., Riemenschneider, K., Lee, C. S. D., Olivares-Navarrete, R., Schwartz, Z., & Boyan, B. D. (2014). Impaired bone formation in Pdia3 deficient mice. *PLoS One, 9*(11), e112708-e112708. doi:10.1371/journal.pone.0112708

Weiner, S., & Addadi, L. (2011). Crystallization Pathways in Biomineralization. *Annual Review of Materials Research, 41*(1), 21-40. doi:10.1146/annurev-matsci-062910-095803

Weiner, S., & Dove, P. M. (2003). An Overview of Biomineralization Processes and the Problem of the Vital Effect. *Reviews in Mineralogy and Geochemistry, 54*(1), 1-29. doi:10.2113/0540001

Weiner, S., & Hood, L. (1975). Soluble Protein of the Organic Matrix of Mollusk Shells: A Potential Template for Shell Formation. *Science (American Association for the Advancement of Science), 190*(4218), 987-989. doi:10.1126/science.1188379

Westbroek, P., De Jong, E., Dam, W., & Bosch, L. (1973). Soluble intracrystalline polysaccharides from coccoliths of Coccolithus huxleyi (Lohmann) Kamptner (I). *Calcified tissue research, 12*(1), 227-238.

Westbroek, P., De Jong, E. W., Van Der Wal, P., Borman, A. H., De Vrind, J. P. M., Kok, D., De Bruijn, W. C., & Parker, S. B. (1984). Mineral phases in biology - Mechanism of calcification in the marine alga Emiliania huxleyi. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 304*(1121), 435-444. doi:10.1098/rstb.1984.0037

Young, J. R., & Henriksen, K. (2003). Biomineralization within vesicles; the calcite of coccoliths. *Reviews in Mineralogy and Geochemistry, 54*(1), 189-215. doi:10.2113/0540189

Young, S. D. (1971). Organic material from scleractinian coral skeletons—I. Variation in composition between several species. *Comparative biochemistry and physiology. B, Comparative biochemistry, 40*(1), 113-120. doi:10.1016/0305-0491(71)90067-8

Ziegler, A., Weihrauch, D., Hagedorn, M., Towle, D. W., & Bleher, R. (2004). Expression and polarity reversal of V-type H+-ATPase during the mineralization-demineralization cycle in Porcellio scaber sternal epithelial cells. *Journal of experimental biology, 207*(Pt 10), 1749-1756. doi:10.1242/jeb.00953

APPENDIX

Table 2. 77 predicted proteins in *E. Huxleyi* coccolith SOM samples detected by LC-MS/MS and their bioinformatic analysis

Protein	Accession number	Description	% Acidic residues
P1	XP_005756361	histone H4	6.8
P2	XP_005758028	pentapeptide repeat-containing protein	10.2
P3	XP_005768681	core histone H2A/H2B/H3/H4	7.5
P4	XP_005786390	pentapeptide repeat-containing protein	12.9
P5	XP_005762404	pentapeptide repeat-containing protein	9.7
P6	XP_005785295	pentapeptide repeat-containing protein	8.2
P7	XP_005771877	pentapeptide repeat-containing protein	11.9
P8	XP_005773372	pentapeptide repeat-containing protein	11.6
P9	XP_005762848	pentapeptide repeat-containing protein	10.4
P10	XP_005762630	putative death-specific protein with Ca binding EF hand domain	17.5
P11	XP_005774988	DUF5011 domain-containing protein	8.8
P12	XP_005766741	pentapeptide repeat-containing protein	10.3
P13	XP_005771822	Esyt3	8.4
P14	XP_005759712	light harvesting protein	11.2
P15	XP_005757874	protein disulfide isomerase	15.2
P16	XP_005784644	pentapeptide repeat-containing protein	12.6
P17	XP_005761517	light harvesting protein	7.1
P18	XP_005794364	pentapeptide repeat-containing protein	7.8
P19	XP_005781697	pentapeptide repeat-containing protein	7.2
P20	XP_005761107	pentapeptide repeat-containing protein	12.6
P21	XP_005763896	peptidylprolyl isomerase	9.6
P22	XP_005766580	kinase UbiB	9.6
P23	XP_005772133	cystatin, putative	7.7
P24	XP_005785778	pentapeptide repeat-containing protein	11.3
P25	XP_005767324	pentapeptide repeat-containing protein	11.0
P26	XP_005756168	2OG-Fe(II) oxygenase	10.4
P27	XP_005768173	peptidyl-prolyl cis-trans isomerase CYP23	12.3
P28	XP_005760557	pentapeptide repeat-containing protein	9.2
P29	XP_005760022	histone 3	11.9
P30	XP_005780078	14-3-3 protein	17.9
P31	XP_005785519	protein disulfide-isomerase	12.8
P32	XP_005766370	DUF500-domain-containing protein	8.2
P33	XP_005783675	peptidyl-prolyl cis-trans isomerase CYP23	13.0
P34	XP_005772886	Cathepsin B-like cysteine proteinase 6	11.0

P35	XP_005762482	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase	11.1
P36	XP_005783192	heat shock cognate 70 kDa protein 2-like	15.5
P37	XP_005756898	Ctsz	9.7
P38	XP_005784200	luminal-binding protein 5	17.5
P39	XP_005761790	alkaline phosphatase	10.4
P40	XP_005767414	AP-2 complex subunit beta	12.9
P41	XP_005784095	BspA family leucine-rich repeat surface protein	9.9
P42	XP_005793203	clathrin heavy chain 1	13.3
P43	XP_005771708	Dockerin-containing protein	11.1
P44	XP_005763580	DUF805 domain-containing protein	11.3
P45	XP_005760612	FAD-binding oxidoreductase	10.4
P46	XP_005764796	heat shock protein 90	20.3
P47	XP_005760781	Lysosomal acid lipase/cholesteryl ester hydrolase	10.2
P48	XP_005760433	membrane attack complex/perforin domain-containing protein	8.8
P49	XP_005774125	membrane attack complex/perforin domain-containing protein	11.6
P50	XP_005757726	pentapeptide repeat-containing protein	12.0
P51	XP_005759281	pentapeptide repeat-containing protein	10.9
P52	XP_005760845	pentapeptide repeat-containing protein	10.2
P53	XP_005761277	pentapeptide repeat-containing protein	13.2
P54	XP_005762982	pentapeptide repeat-containing protein	10.9
P55	XP_005766276	pentapeptide repeat-containing protein	23.0
P56	XP_005766609	pentapeptide repeat-containing protein	13.5
P57	XP_005767209	pentapeptide repeat-containing protein	8.9
P58	XP_005768400	pentapeptide repeat-containing protein	6.6
P59	XP_005768763	pentapeptide repeat-containing protein	11.5
P60	XP_005770461	pentapeptide repeat-containing protein	9.2
P61	XP_005774719	pentapeptide repeat-containing protein	11.2
P62	XP_005782444	pentapeptide repeat-containing protein	9.8
P63	XP_005785161	pentapeptide repeat-containing protein	8.9
P64	XP_005785949	pentapeptide repeat-containing protein	11.4
P65	XP_005788457	pentapeptide repeat-containing protein	14.0
P66	XP_005790425	pentapeptide repeat-containing protein	11.8
P67	XP_005790778	pentapeptide repeat-containing protein	9.4
P68	XP_005793747	proliferation-associated protein 2G4	13.6
P69	XP_005766765	putative actin	13.0
P70	XP_005780283	putative actin	12.6
P71	XP_005785898	putative lipoprotein	7.9

P72	XP_005773365	putative membrane protein	14.6
P73	XP_005776240	S-antigen protein	8.9
P74	XP_005788418	SGNH/GDSL hydrolase family protein	8.7
P75	XP_005789186	S-layer homology domain-containing protein	10.2
P76	XP_005763039	transmembrane protein 87A-like	8.3
P77	XP_005778320	V-type proton ATPase subunit B, brain isoform	13.2

Returned sequences with e-values $\leq 10^{-10}$ are presented in order of decreasing e-value. "Description" is the associated annotation in Blast2GO.

Table 3. 23 hypothetical proteins in *E. Huxleyi* coccolith SOM samples detected by LC-MS/MS with no functional annotations

Protein	Accesion number	Description	% Acidic residues
P78	XP_005779826	EMIHUDRAFT_114669	9.2
P79	XP_005763069	EMIHUDRAFT_215387	14.1
P80	XP_005757876	EMIHUDRAFT_433204	7.1
P81	XP_005770234	EMIHUDRAFT_436302	12.8
P82	XP_005766307	EMIHUDRAFT_437054, partial	9.9
P83	XP_005762693	EMIHUDRAFT_215961	9.8
P84	XP_005760748	EMIHUDRAFT_453072	7.2
P85	XP_005770022	EMIHUDRAFT_458952, partial	7.1
P86	XP_005759268	EMIHUDRAFT_125120, partial	13.2
P87	XP_005771920	EMIHUDRAFT_209050	10.5
P88	XP_005756231	EMIHUDRAFT_372689, partial	9.4
P89	XP_005770935	EMIHUDRAFT_196821	9.6
P90	XP_005769861	EMIHUDRAFT_210142	14.8
P91	XP_005772107	EMIHUDRAFT_255620	7.8
P92	XP_005760775	EMIHUDRAFT_438298	10.2
P93	XP_005792924	EMIHUDRAFT_222578	8.4
P94	XP_005779424	EMIHUDRAFT_236194	13.3
P95	XP_005757050	EMIHUDRAFT_446637	11.1
P96	XP_005789004	EMIHUDRAFT_122496, partial	12.0
P97	XP_005759750	EMIHUDRAFT_198760	10.2
P98	XP_005765648	EMIHUDRAFT_247099	15.1
P99	XP_005791291	EMIHUDRAFT_251619	9.0
P100	XP_005782625	EMIHUDRAFT_113418, partial	11.1

Returned sequences with e-values $\leq 10^{-10}$ are presented in order of decreasing e-value. "Description" is the associated annotation in Blast2GO. P100 had no Blast hits