DISSERTATION

INVESTIGATING THE BIOCHEMISTRY AND GENETICS OF CHRYSOLAMINARIN METABOLISM IN A MODEL MARINE DIATOM

Submitted by

Michael Adan Caballero

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Graham Peers

John Belisle
Daniel Bush
Ashok Prasad

ABSTRACT

INVESTIGATING THE BIOCHEMISTRY AND GENETICS OF CHRYSOLAMINARIN METABOLISM IN A MODEL MARINE DIATOM

Diatoms are ecologically significant marine algae as they contribute approximately 20 % of Earth's photosynthetic productivity. The evolutionary history of diatoms is complex, and their chimeric genomes encode unexpected biochemistries. Diatoms are proposed to be an excellent candidate to produce sustainable biofuels and/or bioplastics; however, a limited functional understanding of their cellular biochemistries represents a critical challenge that limits the use of diatoms in industrial settings. Notably, diatoms make the unusual storage sugar chrysolaminarin, instead of glycogen or starch. The carbon stored in this carbohydrate pool may be exploited and redirected towards products of interest, such as the accumulation of triacylglycerol as a precursor metabolite for biofuel production. Optimizing carbon flow to biofuel or bioplastic precursors requires a fundamental understanding of how chrysolaminarin is made, modified, and consumed. Therefore, I have designed and executed a three-part research strategy to investigate the biochemistry and genetics of chrysolaminarin metabolism in the model diatom *Phaeodactylum tricornutum.* This research addresses key challenges in chrysolaminarin biology, including quantification of the polysaccharide, elucidating the first enzymatic step of chrysolaminarin biosynthesis, and identification of novel proteins associated with the chrysolaminarin metabolism.

First, I set out to improve the state of the art of chrysolaminarin quantification, and this work is documented in Chapter 2. The 3-methyl-2-benzothiazolinone hydrazone (MBTH) assay

affords considerably better accuracy than the phenol-sulfuric acid assay[1, 2]. This assay, when combined with a warm water extraction, permits parallel quantification of soluble non-reducing carbohydrates (chrysolaminarin), soluble reducing carbohydrates, and insoluble carbohydrates. This method was applied to an investigation of carbon partitioning in *Phaeodactylum* during nitrogen starvation, a treatment that has been shown to increase chrysolaminarin content in other diatoms[3]. Interestingly, *Phaeodacylum* did not increase chrysolaminarin content relative to total organic carbon during nitrogen starvation, challenging the assumption that diatoms universally repartition their carbon into chrysolaminarin reserves during nitrogen starvation. The outcome of this research was an accurate and reliable method to quantify carbohydrate fractions, especially chrysolaminarin, from diatom extracts.

Second, I set out to systematically characterize diatom enzymes capable of synthesizing UDP-glucose, which is documented in Chapter 3. The *E. coli* strain *DEV6* lacks a functional UDP-glucose diphosphorylase, conferring an inability to grow when galactose is the sole carbon source[4]. This deficiency has been used as a complementation assay to demonstrate the ability of a recombinant, eukaryotic protein to catalyze the synthesis of UDP-glucose [5]. I have adapted this strategy to facilitate a functional screen of a *Phaeodactylum* cDNA Expression Library. In this manner, I have identified a novel *Phaeodactylum* enzyme capable of synthesizing UDP-glucose, increasing the number of potential enzymes involved in the first step of chrysolaminarin biosynthesis to three. All three gene targets (*ugp1*, *ugp2*, and *ugp3*) were knocked out using a CRISPR/Cas9 approach and were quantified for chrysolaminarin content. Additionally, UDP-glucose can play other unique roles in the metabolism of photoautotrophs, such as sulfolipid biosynthesis[6], which was further explored. This research systematically explored the potential contribution of three UDP-glucose synthesizing enzymes to chrysolaminarin biosynthesis. The

comprehensive suite of CRISPR/Cas9 knockouts revealed unique phenotypic patterns associated for each disrupted gene. Knocking out *ugp1* confers mutants which accumulate less chrysolaminarin and more triacylglycerol at dusk. Knocking out *ugp2* yields mutants that accumulate not only less chrysolaminarin but also less triacylglycerol at dusk. These *ugp2* knockouts also exhibit reduced fitness both in exponential growth rates and maximum quantum yield of Photosystem II. Knocking out *ugp3* does not yield an obvious phenotype relative to WT for the carbon partitioning or fitness parameters examined. The *ugp1* knockouts exhibit phenotypes in good agreement with previous reverse genetics studies[7, 8]. These findings help resolve and predict distinct metabolic roles for these three enzymes, a critical detail when interpreting systems biology datasets or establishing genome-wide metabolic models.

Third, I set out to identify novel, β-1,3 glucan binding proteins from the *Phaeodactylum* proteome, and these efforts are documented in Chapter 4. I implemented a 2D-affinity electrophoresis approach to resolve proteins based on their interaction with an embedded substrate. This strategy has resolved several starch binding proteins[9]. I adapted this approach by substituting β-1,3 glucan in the gel, and separating *Phaeodactylum* soluble protein extracts. A gel spot was resolved in this manner, and unique peptides were identified from the hypothetical protein Phatrdraft_47612. RNAi-mediated knockdowns of Phatrdraft_47612 were generated to investigate a potential chrysolaminarin phenotype. However, later analyses of protein migration after 2D-affinity electrophoresis suggested that the gel spot I observed might be an artifact of protein overloading. While this research effort did not meet the goal of identifying novel chrysolaminarin-related proteins, I have documented several considerations for future efforts to identify β-1,3 glucan-binding proteins by reflecting on shortcomings in the execution of this study.

The following research chapters in this thesis aim to expand our understanding of chrysolaminarin metabolism in the model diatom *Phaeodactylum tricornutum*. By emphasizing biochemical approaches, novel insights in diatom biology have been realized. Better resolution of chrysolaminarin content from cell extracts provides not only an opportunity to study diatom carbon partitioning in more detail, but also the ability to specifically identify and characterize mutants with disrupted chrysolaminarin accumulation. Leveraging the latter point, I also studied a suite of gene knockouts and collected observations that indicate that Phatrdraft_50444 is the primary contributor to chrysolaminarin biosynthesis. Finally, I attempted to establish evidence for novel chrysolaminarin-related proteins by protein-carbohydrate affinity, which provided insights for future research strategies to explore these interactions. As such, the following chapters advance our understanding of chrysolaminarin metabolism and the general biology of the biotechnologically relevant diatom *Phaeodactylum tricornutum*.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Graham Peers, for his ongoing mentorship over the last six years. Graham empowered me with a valuable opportunity to explore and learn from my mistakes along the way. He has helped me stay the course during my PhD studies. I will never forget his support during my particularly difficult Spring 2014 semester.

I would also like to thank my committee, John Belisle, Dan Bush, and Ashok Prasad for their feedback during committee meetings, attending my CMB seminars, and clearly showing genuine interest in my academic development. I always appreciated bumping into you on campus.

I am humbled to thank many individuals for their time and effort in various collaborations. Naoki Sato kindly welcomed me as a visiting summer graduate student at the University of Tokyo in 2012 and trained me how to analyze cyanobacteria lipid profiles using thin layer chromatography and GC. Mam Sherman in the Protein Expression and Purification Facility at CSU trained me on FPLC for protein purification, and helped me troubleshoot issues with recombinant protein expression. Libin Shi in the Delphi Chatterjee lab trained me on how to prepare and analyze GC/MS carbohydrate profiles. Chris Rithner in the Central Instrument Facility ran the ^{1}H-NMR analysis of chrysolaminarin, which we interpreted together. Denis Jallet trained me on the Walz PAM fluorometer and programmed scripts to investigate the photophysiology of *Phaeodactylum* as well as quantify triacylglycerol via thin layer chromatography. Denis Jallet also facilitated the transfer of the CRISPR genome editing methods for *Phaeodactylum* from our collaborators at JCVI, Andrew Allen and Mark Moosburner. David Xing assembled the *Phaeodactylum* cDNA expression library, which I was fortunately able to

apply to my project. Corey Brockling ran the targeted lipidomics experiment, and guided me through interpreting the data. I am grateful for this tremendous collective effort.

I am grateful to the National Science Foundation which supported me financially during most of graduate school through the Graduate Research Fellowship and the Integrative Graduate Education and Research Traineeship – Multiple Approaches to Sustainable Bioenergy programs. Financial support for this research also came from the Department of Energy.

The Peers lab has been my second home for the last six years, and it goes without saying that I could not have accomplished this work without the support of Denis Jallet, Michael Cantrell, Matt Youngblood, Alexandra Gallina, and David Xing. It has been a pleasure to share lab space with Nate Sindt, Alex Hughes, Bjoern Andersson, and Annah Holmberg. I want to particularly acknowledge my three undergraduate mentees: Elena Michel, Kate Austin, and Leo Zhang. You have inspired me on this PhD journey and it was a privilege to work with each of you.

Finally, I would like to thank my family for their love and support during this chapter of my life. Mom and Dad, your unwavering belief in me carried me through difficult times. Ana, your work ethic and compassion inspired me to carry on. Mark, your dedication is contagious and helped me push myself farther than I thought I could. Matt, you encouraged me to find harmony wherever I was. Ben, your creative talent motivates me to look at the world around me in different ways. I am also incredibly grateful for the love and support from Aimee Jalkanen. I also thank Rob Curl, Craig Williams, and David Prange for their supportive phone calls and ongoing friendships.

DEDICATION

In memoriam.

Edith Rodriguez de Caballero

Maria Trazer Axt

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

An introduction to diatoms

The evolutionary history of diatoms

Diatoms are placed within the stramenopiles (heterokonts) by phylogenetic analysis, and are adjacent to alevolates and rhizarians in the "SAR" (Stramenopile-Alevolate-Rhizaria) supergroup of eukaryotes. Algae are well-represented within the stramenopiles, which includes diatoms, brown algae (Phaeophyceae), and golden algae (Chrysophyceae). More recently, 18S rRNA tools have identified picoeukaryotic algae as multiple lineages within the stramenopiles[10]. However, not all stramenopiles are photoautotrophs. A notable example of a heterotrophic stramenopile is the oomycete *Phytopthera*[11], which represents an ongoing challenge as an agricultural pathogen. Photosynthetic stramenopiles, such as diatoms, evolved through two independent endosymbiotic events, which entailed a heterotrophic cell engulfing a red algal cell[12, 13].

Phylogenetic analysis further specifies diatoms. Two classes of diatoms have been identified, called centric and pennate. Recent analysis has suggested the addition of an additional class, which divides the pennate class into araphid pennate and raphid pennate classes[14]. Centric and pennate groups have been estimated to diverge between 183 and 238 million years ago (Mya) using a molecular clock model[14]. The earliest fossil record of diatoms is 180 Mya[15]. Therefore, diatoms are an ancient species of algae with a complex evolutionary history.

The ecological significance of diatoms

Diatoms are ubiquitous and ecologically successful microalgae. Diatoms are most abundant in polar regions where blooms occur annually[16], but they can also be found in

diverse environments, including fresh water[17], soils[18], and arctic ponds[19, 20]. In addition to their bloom forming capacities and ubiquitous nature, diatoms are extremely productive photosynthetic organisms and contribute to global cycles of carbon and silica.

Net primary productivity is the amount of inorganic carbon fixed at a given scale (e.g. organism, population, ecosystem) after correcting for respiratory processes. Annual net primary productivity of the planet is estimated at 104.7 petagrams (10^{15}) of carbon each year, and diatoms contribute about 20 petagrams to this figure, representing nearly 20 % of Earth's annual net primary productivity[21]. As a reference, the contribution of all tropical rainforests to net primary productivity is estimated to be 17.8 petagrams and the contribution of all savannahs to net primary productivity is estimated to be 16.8 petagrams. Thus, diatom productivity in the marine environment is comparable to the most productive terrestrial ecosystems. Diatom blooms play a major role in the marine food web and consequently support the health of fisheries[22], which are relied on as a major food source.

A unique feature of diatoms is the synthesis of a silica-based cell wall (frustule) which is briefly discussed later in the introduction. Given the ubiquitous nature of diatoms, it is no surprise that diatoms represent a critical biological component of global silica cycles[23]. Diatoms serve as the primary driver of the 200 – 280 teramole (10^{12}) silica biologically produced each year in the ocean [24]. More than half of this silica is recycled or dissolved in the photic zone through biotic grazing and abiotic degradation [24]. The remainder of the biologically produced silica is exported deeper in the water column, below the photic zone [22]. Frustules may settle on ocean floors for hundreds of years, but silica buried deeper in the sediments may remain over geological timescales (between 10^6 to 10^9 years) [23].

Industrial uses of diatoms

Currently, two major strategies exploit the use of diatoms at an industrial scale. First, diatom silica frustules can be extracted from sedimentary deposits [25]. Extracted sedimentary deposits, also known as diatomaceous earth, have been used for inexpensive filtration and soil supplementation strategies. Second, diatoms synthesize the essential ω-3 fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). These high value lipids have been used as a supplement for animal feed[26].

There also exists a potential applied use of diatoms in the biofuel industry. The promise of biofuels lies in the potential to establish an open carbon cycle in which photosynthetic organisms are capable of not only providing feedstocks for the downstream production of fuels, but also recycling CO_2 emitted from combustion of that fuel[27]. First generation biofuel crops established this paradigm by fermenting ethanol from grains, but did not lead to reductions in greenhouse gas emissions[28], and struggled with an ethical dilemma of "food versus fuel." Second generation biofuels improved reductions in greenhouse gas emissions by fermenting agricultural waste biomass instead of grains[28]. However, ethanol remains a poor fuel choice due to its relatively low energy density and corrosive properties[29].

Chemical conversion of lipids into biodiesel represents an alternative strategy for biofuel production. Biodiesel precursors can be generated from a transesterification reaction which catalyzes the production of fatty acid methyl esters and glycerol from triacylglycerol (TAG)[29]. Diatom biomass has been specifically proposed as a platform for biodiesel production because they already maintain a significant amount of their carbon as TAG[30]. Algal biofuels are a promising technology due to their significant improvements over land plants for both lower arable land requirements[31] and lipid productivities[32]. However, biodiesel remains

econometrically unfavorable and has higher nitrogen oxide emissions when compared to conventional diesel[29]. One strategy to improve the econometrics of biodiesel production is to further increase triacylglycerol content through the manipulation of central carbon metabolism[33]. Therefore, understanding central carbon metabolism in diatoms may benefit the bioengineering of diatom strains for biodiesel production.

Several unique features of diatom physiology

Key physiological differences between diatoms and green algae has previously been reviewed[34]. Below, I describe several of these atypical features in additional detail.

As mentioned earlier, diatoms have a glass-like silica cell wall known as a frustule. In an effort to study the biological process of silica deposition, compositional analysis of frustules was performed to identify frustule-associated sugars and proteins. Hecky (1973) discovered an enrichment of serine residues from such preparations, and hypothesized that a covalent bond between these serine residues and the silica frustule assisted in the biosilification process [35]. This realization initiated further studies that identified frustule-associated silica-forming proteins using biochemical[36], proteomics[37], and transcriptomics[38] approaches. In this manner, four protein families have been identified: frustulins, pleuralins, p150 proteins, and silaffins [39]. While a frustulin was determined to be the first frustule-associated protein[36], these proteins appear to accumulate later in the process of silica cell-wall formation[40]. Nevertheless, the process of biological silica deposition is an area of active study, and a combination of biochemistry, comparative genetics, and bioinformatics may further clarify this ecologically significant pathway[39]. These studies were initiated by an effort to identify frustule-associated proteins, which represents a strategy to generate novel insights into specific substrate-protein interactions.

One of the challenges aquatic photosynthetic organisms face is the ability to concentrate inorganic, gaseous carbon dioxide in high concentrations in the chloroplast relative to the marine environment[41]. This is a shared challenge for some land plants, which exhibit a process known as C4 metabolism. These plants counter the process of photorespiration, an incidental and potentially wasteful oxygenation reaction during carbon fixation, by increasing carbon concentrations in bundle-sheath cells. Interestingly, diatoms exhibit carbon-concentrating mechanisms that may function like a C4-type metabolism in a single cell[42, 43]. The potential contribution of C4-related enzymes to diatom carbon fixation has been demonstrated with short term carbon labeling, which identified both C3 and C4 intermediates[44], although some diatom species did not accumulate C4 species. Bicarbonate pumps, such as the family of SLC4 transporters, have been identified and characterized in diatoms[45]. However, the role of C4 metabolism in diatoms is controversial. Knocking down an essential gene in C4 metabolism, pyruvate phosphate dikinase (PPDK), did not significantly disrupt carbon fixation[46], although parallel mechanisms to facilitate carbon fixation in diatoms might compensate for this disruption. Diatoms exhibit unique carbon concentrating mechanisms which are still being elucidated.

Bioinformatic analysis of diatom genomes predict unusual biochemistries

Diatoms have a radically different blueprint for cellular metabolism and biochemistry than other model organisms, due to their complex evolutionary history, and diatom genome sequencing efforts have underscored this complexity. Both pennate (*Phaeodactylum tricornutum)* and centric (*Thalassiosira pseudonana*) diatoms have been cultivated in laboratory settings and have sequenced genomes[47, 48]. Comparative genomic analysis between these two diatoms and a pool of eukaryotic organisms observed a striking finding[48]. While *Phaeodactylum* and *Thalassiosira* share 3,176 gene families in common, they also have 3,710

and 3,455 unique gene families, respectively. This suggests that a significant opportunity exists to characterize the genetics of diatom-unique biochemistries.

Phylogenetic analysis also suggested that as much as 5 % of diatom genes are a result of recent horizontal gene transfer from prokaryotes[48]. However, this value might be inflated by the lack of annotated genomes throughout the SAR supergroup[49]. A recent published diatom genome investigated the potential contribution of horizontal gene transfer to the *Cyclotella cryptica* genome[50], where they found that about 1.5 % gene families may have arisen from horizontal gene transfer. More diatom genome sequencing projects are in development[51], and may help resolve the extent of horizontal gene transfer events in diatom genomes. Nevertheless, recent horizontal gene transfer events contribute to the chimeric composition of diatom genomes.

Genomic mining strategies have identified unexpected metabolic pathways in diatoms. For example, the discovery that diatom genomes contained a suite of genes encoding a functional urea cycle was surprising[47, 48, 52], as a functional urea cycle was previously thought to be exclusive to metazoans. In metazoans, the urea cycle serves to remove excess nitrogen. However, in diatoms, the urea cycle serves as a key redistributor of nitrogen[53]. This was suggested from observation of slow recovery during the re-addition of nitrogen sources in an ornithine-urea cycle mutant relative to WT[53]. Likewise, discovery of the gene components of the Entner-Doudoroff pathway, an ancient glycolytic pathway thought to be mostly exclusive to prokaryotes, was found using an orthology-based approach[54]. However, a recent study has suggested that the Entner-Doudoroff pathway has been broadly overlooked in eukaryotes and cyanobacteria, showing that recombinant barley 2-keto-3-deoxygluconate-6-phosphate aldolase, a unique enzymatic step in Entner-Doudoroff pathway, has *in vitro* activity[55]. Nevertheless, *in silico* discovery of pathways have identified unique diatom biochemistries for further study.

Both these *in silico* discoveries were further characterized using functional genetics. Mitochondrial carbamoyl phosphate synthase, the driver of the urea cycle in metazoans, was knocked down in *Phaeodactylum* using a RNAi-mediated approach[53]. These knockdowns accumulated less mitochondrial carbamoyl phosphate synthase as determined by immunoblot, and exhibited a slower growth and an impaired metabolic response to nitrogen supplementation. Similarly, an *E. coli* strain with a knocked out Entner-Dourdoff pathway was generated, and this strain was complemented with the putative homologous genes from *Phaeodactylum*[54]. The base strain was not viable when grown on minimal media with gluconate as a carbon source, but was viable when complemented with the diatom genes. These additional biochemistry efforts are essential for creating genome-wide metabolic models, which rely on a fine-tuned understanding of enzymes and the stoichiometry of reactions they catalyze[56].

The storage of carbohydrates represents another unique physiological aspect of diatoms and will be introduced at the end of the next section. First, polysaccharide metabolism will be briefly described, providing context for potential comparisons to diatom polysaccharides.

Polysaccharide metabolism

Photosynthetic organisms convert light energy into chemical energy via photosynthesis. The two energetic co-factors produced by photosynthesis are ATP and NAPDH. Most of these cofactrors go towards energy intensive processes, such as nitrogen or carbon fixation. These are linked processes, as nitrogen fixation requires carbon skeletons and carbon fixation requires nitrogen in the form of Calvin-Benson-Bassham cycle enzymes[57]. Nevertheless, the intermediate metabolite 3-phosphoglyceric acid connects carbon fixation (Calvin-Benson-Bassham cycle) to central carbon metabolism (glycolysis/gluconeogenesis). Converting reduced carbon intermediate metabolites into longer-term storage compounds can be important for the

overall fitness of photosynthetic organisms, as mutants with disrupted storage polysaccharide metabolism can exhibit detrimental growth phenotypes[33, 58]. The following section briefly introduces the diversity of polysaccharide synthesis by highlighting the biochemical features and genetic studies of each polysaccharide species.

Polysaccharide synthesis requires activated sugars

Polysaccharides serve as a major form of carbon skeleton and energy storage in cellular metabolism, and they are synthesized from "activated" nucleotide sugars. Free sugars (e.g. glucose) are phosphorylated (glucose 1-phosphate), and then converted into a nucleotide sugar by a nucleotransferase[59]. Uridine diphosphate glucose (UDP-glucose) and adenine diphosphate glucose (ADP-glucose) are the most common nucleotide sugars, although as many as 30 have been identified in plants[59]. Nucleotransferases are reversible reactions, with the direction of the reaction depending on substrate concentrations[60, 61]. Nucleotide sugars like UDP-glucose are then used by glycotransferases in polysaccharide synthesis. While glucose is the predominant sugar species in many polysaccharides, the polysaccharide can be synthesized from several kinds of nucleotide sugars.

Callose

Callose is a β-1,3 polysaccharide with limited β-1,6 branching, and is entirely composed of glucose. Interestingly, molecules structurally and compositionally similar to callose are found in yeast, bacteria, and diatoms; however, the molecular weight and solubility varies by species[62]. Callose is a structural polysaccharide typically found in plant cell walls, but is also synthesized in response to cell wounding and cell plate formation[63]. Notably, callose deposition is important for normal pollen development[64]. Callose is produced by the membrane-bound protein callose synthase[63] which forms a complex with a UDP-glucose

transferase[65], suggesting that UDP-glucose serves as the activated sugar monomer for callose production. β-glucanases are the primary enzymes involved in callose degradation[66].

Starch

Starch is a non-linear α-1,4 polysaccharide with extensive α-1,6 branching, and is composed of glucose. Starch contains two glucans defined by the incidence of α-1,6 branching: amylose (infrequent branching in a polymer under 10,000 glucose units) and amylopectin (frequent branching in a polymer greater than 10,000 glucose units)[67]. The latter is particularly important to produce insoluble starch granules. Starch serves as a storage polysaccharide, providing an important pool of carbon skeletons and energy for cellular metabolism, although only select tissues store starch and most carbon and energy serves to fuel plant growth. Nevertheless, the diel accumulation of starch in the light and depletion in the dark is a particularly pronounced pattern in leaf tissue of the model plant *Arabidopsis thaliana*[68].

Some starch accumulates in leaf chloroplasts, but relatively large stores of starch accumulate in sink tissue amyloplasts. The metabolites involved in the process of transporting sugar from leaves to sink tissues is well characterized[67], and briefly summarized. Sucrose is a soluble disaccharide composed from fructose and glucose, and is produced in the cytoplasm. This disaccharide is then transported into the cytoplasm of sink cells where sucrose synthase or cell wall invertases cleave the disaccharide into fructose and UDP-glucose[69]. Increasing sucrose synthase expression has been shown to alter carbon partitioning in poplar[70] and plant height growth in tobacco[71]. UDP-glucose diphosphorylase (reversibly) catalyzes the production of glucose 1-phosphate and uridine triphosphate. Glucose 1-phosphate is utilized by ADP-glucose diphosphorylase to yield ADP-glucose. ADP-glucose is used as the monomer for starch synthesis, relying on a large suite of enzymes including, but not limited to, starch

synthases[72], starch branching and debranching enzymes, starch phosphorylases[73], and phosphoglucan phosphatases[74, 75]. Interestingly, starch-binding domains are found on many proteins actively involved in starch metabolism[76], underscoring the role of protein-carbohydrate interactions in starch metabolism. It is important to consider that several important starch-related proteins are involved in the regulation, rather than the direct synthesis, of the polysaccharide.

It is also important to consider studies of starch metabolism in the green alga *Chlamydomonas reinhardtii. Chlamydomonas* has been a platform for the study of starch metabolism in microalgae. Gene disruption studies of ADP-glucose diphosphorylase[77] and the debranching enzyme isoamylase[33] have shown that microalgae can substantially modify their cellular carbon partitioning away from carbohydrate reserves into lipid reserves. This realization has been exploited in bioengineering strategies to favor further lipid accumulation for biofuel production.

Cellulose

Cellulose is an insoluble β-1,4 glucan and is an integral component of the plant cell wall structure. In *Arabidopsis*, there are 39 cellulose synthase / cellulose synthase-like genes[78]. Cellulose is synthesized from UDP-glucose activated sugars, but some cellulose synthase-like genes encode a product capable of synthesizing polymers using other activated sugars such as guanosine diphosphate (GDP)-glucose and GDP-mannose[79]. Knocking out certain cellulose synthases can substantially alter the cell wall composition[80], emphasizing the role cellulose plays as a structural, cell wall polysaccharide.

Glycogen

Glycogen is a soluble α-1,4 glucan with extensive α-1,6 branching. Interestingly, glycogen is synthesized by animals, yeast, cyanobacteria, and other bacteria. The role of glycogen in metabolism is broadly considered to be a carbon and energy reserve. However, the biochemistry and genetics associated with glycogen synthesis differs between eukaryotes and prokaryotes.

In eukaryotes, glycogen is synthesized from UDP-glucose[81]. Glycogenin and glycogen synthase contribute to the synthesis of the α-1,4 glucan, while branching is mediated by branching enzyme. Degradation of glycogen is mediated by glycogen phosphorylase and debranching enzymes, yielding glucose 1-phosphate for catabolism. Deficiencies in these enzyme activities can cause severe phenotypes in humans. For example, Andersen disease (glycogen storage disease, type IV) results from deficiencies in glycogen branching enzyme and often causes premature death[82]. Dysregulated glycogen phosphorylation yields insoluble glycogen granules and has been studied in investigations exploring the mechanisms of Lafora disease [83].

In comparison, prokaryotes synthesize glycogen from ADP-glucose[84]. Glycogenenin homologs have not been identified in prokaryotes, and bacterial glycogen synthase alone synthesizes the α-1,4 glucan[85]. Prokaryotic glycogen catabolism is also mediated by glycogen phosphorylase and debranching enzymes. Disrupting the bacterial debranching enzyme in the cyanobacteria *Synechocystis elongatus* PCC 7942 exhibited a small reduction in the amount of glycogen present in a culture volume as well as a reduction in glycogen chain length[86].

Chrysolaminarin – the unique β-1,3 storage polysaccharide of diatoms

Historical perspective of chrysolaminarin metabolism

Leucosin (chrysolaminarin) was described by Klebs in 1893 in a chrysophyte by microscopic observation of refractive vacuoles[62]. Leucosin was later renamed "chrysolaminarin" by Beattie et al. (1961) to reflect its structural similarity to laminarin extracted from *Laminaria* species, noting that the only significant difference between the polysaccharides was that some laminarin contained a small quantity of mannose[87]. A unique characteristic of the stramenopiles is that both photosynthetic[87, 88] and nonphotosynthetic[89] species synthesize β-1,3 glucan as a food reserve. The composition and structure of glucans extracted from the diatom *Phaeodactylum tricornutum* was studied by Ford and Perceval (1965) [88]. They resolved two major polysaccharide fractions: a soluble, extractable β-1,3 glucan (chrysolaminarin) and an insoluble, cell-wall associated mannan. Cellular chrysolaminarin content presumably serves as a storage polysaccharide, as its accumulation and consumption has been linked to day:night cycles[3].

The seminal study by Roessler (1988) began to investigate the biochemistry of chrysolaminarin metabolism[90]. Roessler made two important observations in this study. First, he found that UDP-glucose is the most abundant activated sugar in diatom extracts. Second, providing radiolabeled UDP-glucose to soluble protein extracts yielded a radiolabeled soluble glucan – a fraction presumably containing chrysolaminarin. The major conclusion from this study is that UDP-glucose is the activated sugar contributing to chrysolaminarin synthesis. However, the specific gene(s) encoding this activity would not be predicted for several decades, until diatom genomes were sequenced and annotated several decades later, as discussed below.

Bioinformatics-enabled predictions of chrysolaminarin metabolic pathways

Comparative genomics efforts relying on orthologous pathways to identify genes has successfully been applied to explore unique aspects of diatom biology such as the urea cycle[53] and the Entner-Dourdoff pathway[54]. Similar comparative genomics approaches have been used to explore chrysolaminarin metabolism[91].

However, neither callose nor starch can serve as truly homologous pathways to predict chrysolaminarin synthesis. Callose, as a β-1,3 glucan with few β-1,6 branches, is structurally comparable to chrysolaminarin, but the biological role of callose is that of a structural carbohydrate. In contrast, chrysolaminarin is a food reserve for the cell, localized within a chrysolaminarin vacuole in the cytosol as determined by staining with aniline blue (Sirofluor), a β-1,3 glucan fluorophore[92, 93], as well as immunolabeling with an anti-β-1,3 glucan antibody[94]. Although, some β-1,3 glucan has been associated with the frustule[95], so it is possible to speculate that some fraction of β-1,3 glucan produced by diatoms can play a structural role. On the other hand, starch fulfills a comparable biological role as chrysolaminarin, but the enzymes involved in polysaccharide synthesis are not directly comparable, as their biochemistries differ. With these limitations in mind, multiple models of diatom carbon metabolism have been proposed, including some putative chrysolaminarin-related targets, in the diatom *Phaeodactylum tricornutum*[91, 96, 97]. This analysis identified two putative genes whose products are predicted to synthesize UDP-glucose (Phatrdraft_50444 and Phatrdraft_23639, hereafter called *ugp1* and *ugp2*, respectively), and a putative beta-glucan synthase (BGS1) based on homology to callose synthase in plants. Interestingly, Kroth et al. (2006) extended their comparative genomics analyses to include yeast cell wall-related proteins to identify β-1,6 branching enzymes[91]. Despite the major differences between other

polysaccharides and chrysolaminarin, these bioinformatics efforts created a starting point to explore the otherwise uncharacterized biochemistry and genetics of chrysolaminarin biology[98].

These assignments have since been explored in systems biology studies. Of particular interest to chrysolaminarin metabolism, Chauton et al. (2012) explored transcriptional changes over the course of a day:night light regime[99]. Since chrysolaminarin serves as reserve carbon skeletons and energy, and accumulates during the day and is consumed at night, identifying significant transcriptional changes could lend insight to transcriptional regulation of chrysolaminarin metabolism. Chauton et al. (2012) concluded that the *ugp2* (Phatrdraft_23639) gene product represented the first enzymatic step for chrysolaminarin biosynthesis, based on the diel accumulation pattern of its mRNA, maximally accumulating during the day and minimally at night.

Functional demonstrations of chrysolaminarin-related gene products

Multiple publications have recently supported the comparative genomics approaches of Kroth et al. (2008) [91] despite the significant evolutionary distance between diatoms and the respective genomes in comparison. These studies, described below, yield a putative chrysolaminarin biosynthetic pathway (Figure 1.1). These studies have been supported by sustained improvements in the diatom genetics molecular toolbox. RNAi-mediated knockdowns[100] and TALEN-based knockouts[7] have been the standard approaches in diatom genetics, until the recent development of CRISPR/Cas9 gene editing[101].

It has been hypothesized that UDP-glucose diphosphorylase represents the first step of chrysolaminarin biosynthesis[90, 91]. Both Daboussi et al. (2014) [7] and Zhu et al. (2015)[8] provide evidence supporting this claim, as they characterized *ugp1* (Phatrdraft_50444) knockouts and knockdowns, respectively. They concluded that the *ugp1* gene product was important for

chrysolaminarin biosynthesis. These mutants exhibited significantly different carbon partitioning patterns, most notably an increased accumulation in the neutral lipid triacylglycerol, the other major storage metabolite in diatoms. However, these studies are incomplete, as neither study quantified chrysolaminarin directly, and other potential enzymes predicted to catalyze UDP-glucose synthesis were not tested. It is also important to highlight that the UDP-glucose diphosphorylase characterized in these studies is not the same as the transcriptomics-predicted target[99].

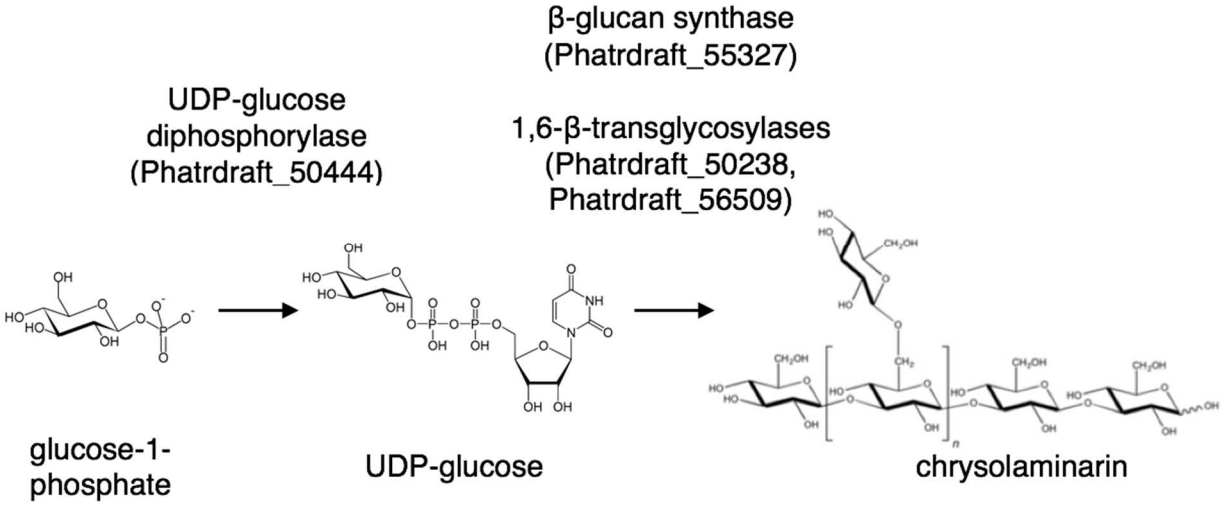

Figure 1.1: Establishing the chrysolaminarin biosynthetic pathway using functional genetics.

Roessler (1987) hypothesized that UDP-glucose is the building block for chrysolaminarin synthesis[90]. Daboussi et al. (2014) disrupted UDP-glucose diphosphorylase using a TALEN-based genome editing approach[7]. Hildebrand et al. (2017) knocked down β-glucan synthase [102] and Huang et al. (2016) observed 1,6-β-transglycosylases complementing yeast cell wall mutants[103]. Together, these studies provide a framework for the chrysolaminarin biosynthesis pathway.

Several genes predicted to encode β-1,6 branching enzymes[91] had their branching activity functionally demonstrated in yeast by Huang et al. (2016) [103]. Yeast deficient in cell-wall polysaccharide synthesis are sensitive to Calcofluor White and are unable to grow[104]. Huang et al. rescued growth in yeast cell-wall mutants grown with Calcofluor White by functional complementation with diatom β-1,6 transglycosylases[103]. However, knockdowns or

knockouts of these targets were not generated in diatoms, so the *in vivo* roles of these enzymes remain uncharacterized in diatoms, so it is possible that branching activity is associated with cell wall carbohydrates. This study suggests that comparative genomics can provide insights for diatom functional genetics studies across vast evolutionary distances. Furthermore, this study demonstrated that recombinant diatom proteins can complement metabolic deficiencies *in vitro*.

Hildebrand et al. (2017) recently published the characterization of β-1,3 glucan synthase knockdowns in *Thalassiosira pseudonana*[102].While they conclude that these knockdowns accumulate less chrysolaminarin than WT after 24 h of silica deprivation, several details limit confidence in this conclusion. Neither protein accumulation nor transcript levels associated with the knockdown target were quantified, so the RNAi strategy is simply presumed to be effective. They observe a reduction in the intensity of the β-1,3 glucan-binding fluorophore, Aniline Blue [92, 93], after 24 h of silica deprivation for both WT and the RNAi strains, suggesting a reduction in chrysolaminarin content. However, this is partially contradicted by their observation that soluble carbohydrate content per cell increases during this starvation timeframe. These carbohydrates were quantified using the phenol-sulfuric acid method[105], which does not differentiate between polysaccharides and monosaccharides, so it is possible that the increase of soluble carbohydrate content is due to a simultaneous increase in monosaccharides and a decrease in chrysolaminarin content per cell. While the callose synthase-like β-1,3 glucan synthase in diatoms represents an interesting target in chrysolaminarin metabolism, more thorough characterization is merited.

These studies highlight recent efforts to characterize of chrysolaminarin metabolism. However, several key knowledge gaps remain and aims to address them are discussed below.

Current challenges and opportunities in chrysolaminarin biology

Characterization of the chrysolaminarin metabolism in diatoms is limited, and represents a significant challenge independent of an applied goal of manipulating cellular chrysolaminarin content. Initial targets relied on comparative genomics which assumes similar enzymatic function over large evolutionary distances. However, studies investigating these early targets described in the previous section have generally supported this bioinformatics approach (Figure 1.1)[7, 8, 102, 103].

However, relying on bioinformatics alone has led to conflicting conclusions in the literature. For example, Kroth et al. identified two UDP-glucose diphosphorylases as potential enzymes catalyzing the first step of chrysolaminarin biosynthesis[91]. To date, only *ugp1* has been targeted for characterization via reverse genetics[7, 8], while the other *ugp2* appears to be transcriptionally regulated over a day:night cycle[99], in parallel with chrysolaminarin accumulation patterns. Subsequently, genome-wide modeling efforts have described chrysolaminarin synthesis with either the former[96], or the latter[99], or both enzymes[54]. Thus, additional characterization of chrysolaminarin biology is required to resolve this ambiguity.

To increase our knowledge of chrysolaminarin biology, I set out to improve the biochemical characterization of chrysolaminarin metabolism. This was achieved through three research aims, highlighted below.

Aim 1: Develop an accurate and reliable method to quantify chrysolaminarin

One immediate challenge of screening for chrysolaminarin metabolism mutants is that no standard kit or protocol exists for quantifying this β-1,3-glucan. Starch quantification strategies rely on the activity of characterized glucan hydrolases coupled with quantification of liberated

glucose[106]. Many putative diatom β-1,3-glucan hydrolases have been identified bioinformatically[91] and β-1,3-glucan hydrolase activity has been demonstrated with crude diatom extracts[107], but no individual diatom enzyme has been characterized as such. Instead, most studies quantifying carbohydrate content of diatoms have relied on total carbohydrate assays such as anthrone[108] or phenol-sulfuric acid[105] protocols. However, these methods can be inaccurate when quantifying total carbohydrates as different sugars yield different absorbances. This is problematic for accurate quantification of carbohydrates from diatom extracts. Therefore, Chapter 2 addresses the challenge of chrysolaminarin quantification.

Aim 2: Functional identification and characterization of *Phaeodactylum tricornutum* UGPases

UDP-glucose serves as the activated sugar for chrysolaminarin synthesis[90]. Two UDP-glucose diphosphorylases have been bioinformatically predicted in the *Phaeodactylum* genome[91]. However, UDP-glucose diphosphorylase is not the only enzyme capable of synthesizing UDP-glucose[109]. While several reverse genetics studies have assumed that Phatrdraft_50444 is the primary contributor to chrysolaminarin synthesis[7, 8], other enzymes potentially synthesizing UDP-glucose in diatoms have not been studied for their potential contribution to chrysolaminarin biosynthesis. Furthermore, directly relying on the current gene assignments may miss novel enzymes with UDP-glucose diphosphorylase activity. I hypothesized that if an enzyme had the capacity to synthesize UDP-glucose, and it was knocked out using genome editing methods, then it would decrease the amount of chrysolaminarin accumulation at the end of the light period.

Aim 3: Identify novel chrysolaminarin-related proteins by protein-carbohydrate affinity

Identifying proteins based on their association with a related substrate has been a successful strategy to identify enzymes involved in silica deposition[35]. Similarly, I wanted to

address the lack of biochemical evidence for chrysolaminarin-related proteins. I reasoned that chrysolaminarin-binding proteins could be identified based on a presumed interaction with β-1,3 glucan. Carbohydrate-binding domains are found on many active carbohydrate enzymes[76], but none have been experimentally demonstrated in *Phaeodactylum*[110]. I hypothesized that β-1,3 glucan binding proteins exist in the *Phaeodactylum* proteome, and as a corollary, that they are important for β-1,3 glucan metabolism.

Applications of an improved understanding of chrysolaminarin biology

It may be desirable to produce significant quantities of a β-glucan for a variety of applications. First, β-glucans can complement anti-fungal vaccine development[111, 112]. β-glucans have been demonstrated to evoke immunological responses in mammalian[113] and insect cells[114]. This response relies on a protein-carbohydrate interaction with the β-glucan, potentially dependent on the chain length of the macromolecule[111]. The number of glucose units, or degree of polymerization, of chrysolaminarin can vary significantly by diatom species[115-117]. Better understanding of both chrysolaminarin genetics and structural details of the glucan itself might facilitate tailored production of chrysolaminarin. For example, in the case of anti-fungal vaccine development, shorter polysaccharides are desired. Replacing the chrysolaminarin biosynthetic pathway from an industrial diatom strain with that of *Thalassiosira weissflogii* might yield chrysolaminarin with as few as 5 glucose residues[116].

Linear β-1,3 glucans represent relatively non-complex sugar compounds, and could potentially serve as a labile substrate for biofuel fermentations. Biomass recalcitrance is a problem facing cellulosic ethanol fermentation[118], and is partially due to plant components such as lignin and other insoluble sugars. As an alternative strategy, pairing a β-1,3 glucan producer with a β-1,3 glucan fermenting microbe could be a strategy to exploit the sugars

produced by stramenopiles[119]. Initial bioengineering strategies to produce more chrysolaminarin could rely on overexpressing chrysolaminarin synthesis-related targets shown in Figure 1.1. This strategy could work around the challenge of recalcitrant biomass, and potentially maximize the uses of algal biomass in an industrial setting.

Conversely, producing less storage sugars can be desirable to redirect carbon partitioning towards a metabolite of interest. Disrupting carbohydrate metabolism to increase lipid content for biodiesel production has been previously described in the model green algae *Chlamydomonas reinhardtii* [33]. Replicating this strategy in diatoms may be particularly promising, as diatoms have previously been identified for their potential as an industrial platform for lipid production[30, 120]. The additional conversion of carbon stored as chrysolaminarin into triacylglycerols could further enhance the econometrics of biofuel production in diatoms.

Preferential accumulation of lipids can also be desirable for non-biofuel purposes as well. *Phaeodactylum* has specifically been suggested to serve as a good industrial platform for the production of EPA[121], and has already been engineered to produce higher levels of DHA[122]. Redirecting carbon metabolism in diatoms away from carbohydrates and towards lipids could be a way to enhance the cellular content of either EPA or DHA. Strategies to produce value-added products in conjunction with biofuel metabolites represents an ongoing challenge to establish favorable economic conditions for biofuel production[123], and directly increasing cellular accumulation of EPA or DHA could benefit the co-production of high-value fatty acids with biofuel precursors.

Outline of research chapters

The following research chapters aim to expand our understanding of chrysolaminarin metabolism in the model diatom *Phaeodactylum tricornutum*. By emphasizing biochemical

approaches, novel insights in diatom biology have been realized. Chapter 2 addresses the challenge of chrysolaminarin quantification. The methodology developed in this chapter has been published[124] and used in subsequent research chapters. Chapter 3 systematically tests the contributions of UDP-glucose synthesizing enzymes to chrysolaminarin biosynthesis. Chapter 4 attempts to identify proteins capable of binding β-1,3 glucan. The following chapters advance our understanding of chrysolaminarin metabolism and the general biology of the biotechnologically relevant diatom *Phaeodactylum tricornutum*.

CHAPTER 2: DEVELOPING A CHRYSOLAMINARIN QUANTIFICATION ASSAY[1]

Preface

One of the existing challenges in diatom biology is the reliable quantification of the storage polysaccharide, chrysolaminarin. Previous methods have relied on total carbohydrate assays which do not differentiate between chrysolaminarin and other carbohydrates. This chapter introduces a method which enables more specific quantification of carbohydrate fractions extracted from *Phaeodactylum*.

Introduction

Diatoms are important contributors to the biogeochemical cycling of carbon, contributing as much as one-fifth of annual net primary productivity [21]. Diatoms allocate organic carbon into two primary storage metabolites: the neutral lipid triacylglycerol (TAG) and the storage polysaccharide chrysolaminarin [34, 47]. TAG is a major carbon reserve, especially during nutrient stress [125]. TAG productivity is of interest for biofuel production, as these neutral lipids are an important precursor for the production of biodiesel [126]. Diatoms have been identified as a promising platform for biodiesel production because of their relatively high biomass and lipid productivities relative to other algal species [120]. Simultaneously, diatoms synthesize chrysolaminarin as a reserve carbohydrate. Chrysolaminarin has been characterized as soluble polymer consisting of glucose monomers linked by a β-1,3 bond with limited β-1,6 branching [88]. The average molecular weight and number of branches of chrysolaminarin can

[1] Reprinted from *Algal Research*, 20, Caballero, M. A., Jallet, D., Shi, L., Rithner, C., Zhang, Y., & Peers, G., Quantification of chrysolaminarin from the model diatom Phaeodactylum tricornutum, 180-188, Copyright 2016, with permission from Elsevier. Figures and tables have been adjusted for formatting specific to this chapter. Preface and afterward have been added to tie the reprinted document into the thesis.

vary significantly between species [62]. The structure and composition of chrysolaminarin is known; however, the enzymes responsible for its creation and consumption remain broadly uncharacterized [62]. Furthermore, the lack of an established quantification method limits our understanding of this reserve polysaccharide as a fraction of total cellular carbon.

Several studies have addressed the challenge of chrysolaminarin quantification by using a warm-water extraction that quantifies insoluble carbohydrates and soluble carbohydrates in parallel [94, 127]. The diversity of insoluble and soluble carbohydrates produced by diatoms has recently been reviewed [128]. Briefly, insoluble carbohydrates contain structural carbohydrates associated with the frustule such as mannans [129], callose [95], and, in some species, chitin [130]. Soluble carbohydrates contain chrysolaminarin [88], exopolysaccharides [131], and free sugars in metabolism. These insoluble and soluble fractions are typically quantified with total carbohydrate assays such as phenol-sulfuric acid or anthrone methods [105, 108]. These quantification methods do not differentiate between monosaccharides and polysaccharides. Several recent studies have opted to observe diatom carbohydrate biochemistry by measuring total carbohydrates with these assays [99, 132, 133]. However, quantifying soluble monosaccharides, soluble polysaccharides (chrysolaminarin), and insoluble polysaccharides in parallel with a reducing sugar assay [134] will provide more insight into carbon partitioning following genetic manipulation or physiological conditions.

Metabolic engineering strategies in many photoautotrophs seek to increase the flux of carbon and energy towards molecules of interest [126, 135]. The proportions of carbon found in the two major carbon reserves of algae, neutral lipids and storage polysaccharides, can be manipulated through changing physiological status or by disruption of biosynthetic pathways. Nitrogen starvation is one of the best-characterized examples of manipulating metabolism:

increased TAG accumulation has been observed in a broad range of organisms from diatoms [57, 136-139] to green algae [139-142] when nitrogen is removed from the medium.

Carbon partitioning also can be altered by genetic engineering. It is known that some proportion of cellular organic carbon is allocated as carbon reserves (both TAG and storage polysaccharides). It was reasoned that disrupting polysaccharide metabolism might further increase the accumulation of TAG, as the displaced storage carbon from polysaccharides had to be partitioned somewhere in metabolism. For instance, mixotrophically grown starchless *Chlamydomonas* mutants accumulated even more TAG than wild type during nitrogen starvation [33, 143]. Therefore, it is rational to suggest that manipulating chrysolaminarin metabolism may enhance TAG productivity in diatoms. Several groups have investigated this possibility in diatoms with a reverse genetics approach by knocking down or knocking out putative enzymes in chrysolaminarin biosynthesis [7, 8], or by blocking β-1,3 glucanase activity [144]. All of these studies reported an increase in TAG content; however, they could only infer a reduction in chrysolaminarin accumulation. Conclusions about carbon repartitioning from chrysolaminarin to other carbon sinks have broadly been limited by the inability to specifically quantify chrysolaminarin from other major carbohydrate pools.

Our primary objective with this study was to establish a method to quantify chrysolaminarin and thereby improve the biochemical toolkit of diatoms to study central carbon metabolism. The method described here is based off of the sensitive and accurate 3-methyl-2-benzothiazolinone hydrazone (MBTH) assay [134, 145]. The MBTH assay has already been applied to a range of carbohydrate quantification applications including field samples [146], glycolytic assays [147, 148], and total carbohydrate content of algae [2]. It provides considerably greater accuracy and precision compared to the commonly used anthrone and phenol-sulfuric

acid carbohydrate assays [2]. This assay is coupled to a workflow that quantifies three major carbohydrate pools from *Phaeodactylum tricornutum,* including a soluble, nonreducing fraction that contains chrysolaminarin. We then applied this workflow to quantify the changes in carbon partitioning to storage metabolites during nitrate starvation on a light/dark cycle. Our data indicate that, during nitrate starvation, chrysolaminarin per cell decreases and that TAG represents the dominant reserve carbon molecule.

Methods

Extraction and analytical chemistry of carbohydrates

Culturing conditions

Batch cultures of axenic *Phaeodactylum tricornutum* CCAP 1055/1 (hereafter, *Phaeodactylum*) were grown in 1 L Roux flasks with 0.8 L of artificial seawater (Instant Ocean, 35 ‰ salinity). Nutrients were added per the stoichiometry described by Guillard [149], but at a 2.3-fold higher concentration to mitigate any nutrient limitation. Silicon was omitted from the medium, as it is not required for *Phaeodactylum* growth. Light was provided in a 12:12 day/night cycle, where daytime irradiance was set at 420 ± 20 μmol photons m^{-2} s^{-1} (LI-COR light meter, LI-250A, 2 π sensor). Five replicate cultures were grown at 18 °C and were mixed by sparging with 1 L min^{-1} air. Cultures were maintained in exponential growth for at least two days prior to the start of the experiment with serial dilutions.

Extraction and purification of soluble carbohydrate

The 4 L volume was harvested at dusk during exponential growth and cells were pelleted by centrifugation (3220 × g, 10 min, 18 °C) in 50 mL conical tubes. All harvested cells were combined into a single pellet, resuspended with 10 mL deionized water, and incubated at 50 °C for 30 min. The suspension was spun (3220 × g, 10 min, 18 °C) and the 10 mL warm water

extract was transferred to a fresh 50 mL conical tube. Then, 40 mL of 95 % ethanol was added to the 10 mL warm water extract, inverted, and allowed to form a precipitate overnight at -20 °C [88]. Soluble, total carbohydrate content was monitored throughout the purification procedure (described below, 2.2.3. – 2.2.4.). The precipitate was pelleted by centrifugation (3220 × g, 10 min, 4 °C), supernatant discarded, and the pellet was thouroughly resuspended in 10 mL 95% ethanol by vigrous pipetting. This process was repeated once more to wash the pellet with 95 % ethanol twice in total. After the second ethanol wash, the pellet was resolubilized with 5 mL deionized water. The resolubilized precipitate was enzymatically treated: a simultaneous DNase (Thermo Scientific, #EN0521) and RNase (Thermo Scientific, #EN0201) digest was performed for 1 h at 4 °C, followed by a 90 μg mL^{-1} Proteinase K treatment (Fisher Scientific, BP1700-100) for 2 h at 4 °C. Then, 4 volumes of 95 % ethanol were added to this aqueous solution and allowed to precipitate overnight at -20 °C. Two 95% ethanol washes were performed the following day, and the pellet was resuspended with deionized water to a final volume of 3 mL. This 3 mL solution transferred into a 2K MWCO Slide-A-Lyzer dialysis cassette (Thermo Scientific, #87718) and was dialzyed against deionized water for 24 h at 18 °C with continuous stir bar mixing. The dialysis volume was 3 × 0.6 L, where deionized water was replaced at 2 h and 4 h. The dialyzed retentate was passed through a 5 mL DEAE FF column (GE Healthcare, #17-515-01) on an ÄKTA start FPLC (GE Healthcare). Protein content was monitored during elution using an A280 detector on the FPLC. A 0 – 1 M NaCl gradient was applied using a 1 M NaCl buffer and 5 mL fractions were collected. Carbohydrate-containing fractions were pooled, frozen in liquid nitrogen, and lyophilized overnight. A polishing step took place by resuspending the lyophilized powder with deionized water to a final volume of 0.5 mL and applied to a Superdex 75 10/300 GL (GE Healthcare, #17-5174-01). One mL fractions were eluted using a

flow rate of 0.5 mL min^{-1} and carbohydrate enriched fractions pooled, frozen, and lyophilized as described above, which yielded a purified polysaccharide sample used for compositional and structural analyses. Laminarin from *Laminaria digitata* was purchased from Sigma (#L-9634) and was directly used as a control for compositional and structural analyses.

Compositional analysis with alditol acetate derivatization and GC/MS

Compositional analysis of purified carbohydrate samples was performed using an established GC/MS approach [150]. Briefly, 1.0 µg of the internal standard, 3-*O*-methylglucose, was added to each sample: laminarin, the purified soluble glucan from *Phaeodactylum,* and the neutral sugar standard (rhamnose, arabinose, ribose, fucose, myo-inositol, mannose, galactose and glucose; 5 µg each). Dried samples were hydrolyzed with 250 µL 2 M trifluoroacetic acid (TFA) for 2 h at 120 °C. After cooling to room temperature (RT), samples were dried under nitrogen gas (N_2). The dried, hydrolyzed samples were reduced with 200 µL 10 mg mL^{-1} sodium borodeuteride in 1:1 of 1 M ammonium hydroxide and 95 % ethanol for 2 h at RT, which was then terminated by adding 4 - 5 drops of glacial acetic acid. After drying the reduced samples under N_2, 200 uL 10% glacial acetic acid in methanol were added, and then were dried under N_2. This process was repeated 4 - 5 times. To the completely dried samples, 100 µL acetic anhydride was added, incubated at 100 °C for 1 h, and then were gently dried under N_2. Two mL chloroform and 1 mL deionized water (v/v, 2:1) were added to the acetylated samples to extract the derivatized products by centrifugation (1000 × g, 1 min, RT). After removing the upper aqueous layers, 1 mL deionized water was added to the lower organic phase to repeat the extraction 4 more times. The organic phase was gently dried under N_2 and resuspended in 50 µL chloroform for gas chromatography/mass spectrometry (GC/MS) analysis. GC/MS analyses were carried out using a CP 3800 gas chromatograph (Varian) equipped with an MS320 mass

spectrometer. Helium was used as the carrier gas with a flow rate of 1 mL min^{-1}. The samples were run on a DB 5 column (30 m × 0.20 mm i.d.). The oven temperature was held at 50 °C for 1 min and programmed at 30 °C min^{-1} to 150 °C and then programmed at 5 °C min^{-1} to 275 °C. Chromatographs were exported to Sigmaplot for analysis.

Structural analysis and interpretation with ^{1}H-NMR

Proton NMR spectroscopy at 500 MHz of the glucan dissolved in DMSO-d_6 was done at 80 °C with a Varian-Inova NMR Spectrometer by using a switchable, 5mm broadband probe. The spectral window was 8kHz and the acquisition time was 2.0 seconds. 32 signal averaging transients were applied with a 22.5 degree tip angle in a simple one-pulse-acquire sequence. There was a relaxation delay of 10 seconds between transients. The free induction decay, FID, was Fourier transformed and the resulting spectral data resolution was 0.5 Hz per point. The DMSO-d_5H_1 impurity was used as an internal spectral reference set to 2.5ppm. Peaks resolved in the spectra were annotated based on their chemical shift and the area under the annotated peaks was determined to infer structural properties of the β-1,3 glucan as previously described [151].

Diel light regime experiment

Growth conditions, medium, and cell harvesting

Phaeodactylum was grown in medium as described in 2.1.1 but instead maintained in 125 mL Erlenmeyer flasks. Triplicate cultures were grown at 18 °C and were constantly mixed on a shaking platform at 120 rpm. Light was provided in a 16:8 day/night cycle, where daytime irradiance was maintained at 325 ± 15 μmol photons m^{-2} s^{-1} (LI-COR light meter, LI-250A, 2 π sensor). *Phaeodactylum* cells (~1.0 x 10^8) were harvested by centrifugation (3220 × g, 10 min, 18 °C) in 50 mL conical tubes. The supernatant was discarded and the pellet was stored at -80 °C until warm water extraction.

Warm water extraction of soluble carbohydrates

We followed the warm water extraction described by Chiovitti et al. [94], but at a higher incubation temperature. Frozen cell pellets were thawed on the day of analysis, suspended in 1.5 mL deionized water (18.2 mΩ), placed in microcentrifuge tubes, and incubated at 50 °C for 15 min to extract soluble carbohydrates. Insoluble carbohydrates and soluble carbohydrates were separated by centrifugation (20,000 × g, 5 min, RT), and the 1.5 mL soluble fraction was transferred to a new microcentrifuge tube.

Sulfuric acid hydrolysis of carbohydrate fractions

Sulfuric acid hydrolysis and quantification of carbohydrates were performed as described by Van Wychen and Laurens [2], with several changes to the protocol to enable quantification of multiple carbohydrate fractions. Immediately after the warm water extraction described above, the insoluble pellet was resuspended with 1 mL deionized water and transferred to a labeled, glass microwave vial (VWR, #89079-404). A 1 mL aliquot of the soluble fraction was also transferred to a labeled, glass microwave vial. The remaining 0.5 mL soluble carbohydrate sample was reserved at 4 °C until quantification. 250 µL sulfuric acid (72% (w/w); Fluka, #00647) was added to each sample, capped (Wheaton, #224100-203), vortexed, and incubated at RT for 1 h. After the incubation, the tubes were uncapped and 6 mL deionized water was added to each sample, bringing the final volume to 7.25 mL. The vials were recapped, crimp-sealed with 20 mm aluminum caps (Wheaton, #224193-01), and placed in an autoclavable rack in a pressure cooker for 1 h (121 °C, 83 kPa). After the pressure equalized, vials were removed from the pressure cooker and cooled to RT. A 2 mL aliquot was taken from each acidified sample and placed in a labeled 50 mL conical tube and was neutralized (pH 6 – pH 8) with calcium carbonate (Sigma-Aldrich, #C6763), verified by placing a few drops of solution on pH indicator

strips (MColorpHast, EMD Millipore, #109535). Neutralized solution was then filtered through a nylon filter (0.2 μm pore size; Fisher, #09-719-006) to remove precipates. Neutralized, filtered solutions were stored at 4 °C and used for quantification within 24 h.

Quantification of carbohydrate fractions

Quantification of carbohydrates using the MBTH method was performed as described previously [2], with some modifications outlined below. A 250 μL sample was added to a glass vial (Pyrex, #9820-10), to which 250 μL 0.5 M NaOH and 500 μL of a freshly prepared 1:1 (v/v) solution of 3 mg mL^{-1} MBTH:1 mg mL^{-1} dithiothreitol (Sigma-Aldrich, #129739; Sigma-Aldrich #D0632) were added, in that order. Glass vials were capped, vortexed, and incubated at 80 °C for 15 min. 500 μL ferric solution (0.5 % (w/v) ammonium iron (III) sulfate (Acros Organics, #205880500), 0.5 % (w/v) sulfamic acid (Sigma-Aldrich, #242772) in 0.25 M HCl) was added to each sample on the heat block. Samples were vortexed, removed from the heat block and allowed to cool for 15 min at RT. These samples were diluted with 1250 μL deionized water, and 1 mL was transferred to 1 cm pathlength acrylic cuvettes to determine absorbance at 620 nm. Carbohydrate concentrations were determined from a standard curve of glucose concentrations ranging from (0 to 0.050) mg mL^{-1} prepared in parallel with each set of experimental samples.

The soluble fraction that was hydrolyzed, neutralized, and filtered generated a value for total soluble carbohydrates. Similarly, the insoluble fraction that was hydrolyzed, neutralized, and filtered generated a value for total insoluble carbohydrates. The soluble fraction that was reserved after warm water extraction generated a value for soluble, reducing carbohydrates only. Subtracting the concentration of soluble, reducing carbohydrates from soluble, total carbohydrates yielded soluble, nonreducing carbohydrates. Dividing these concentrations by the total number of cells harvested for the respective experimental sample yields carbohydrate

fractions on a per cell basis. All carbohydrate fractions are reported as average pg glucose equivalents (g.e.) per cell.

Nitrate depletion experiment

Growth conditions and media

Phaeodactylum was grown in 1 L Roux flasks in 0.8 L of of artificial seawater (Instant Ocean, 35 ‰ salinity). Nutrients were added as described by Guillard [149], except for the nitrate deplete media where nitrate was omitted. Silica was omitted from the media. Light was provided in a 12:12 day/night cycle, where daytime irradiance was set at 420 ± 20 μmol photons m^{-2} s^{-1} (LI-COR light meter, LI-250A, 2 π sensor). Five replicate cultures were grown at 18 °C and were mixed by sparging with 1 L min^{-1} air. Cultures were maintained in exponential growth for at least two days prior to the start of the experiment with serial dilutions. The experiment was carried out for 4 days. Harvests extracted appropriate volumes (described below) to quantify: carbohydrates, total organic carbon and total nitrogen, chlorophyll, and triacylglycerol. Nitrate replete cultures were harvested at dusk and dawn at 0 and 0.5 days. Culture media was exchanged at 0.6 days by spinning down 400 mL of culture (3220 × g, 10 min, 18 °C), washing pellets twice with filtered artificial seawater, and resuspending the algae pellets in 400 mL nitrate deplete media. Nitrate deplete cultures were harvested at dusk and dawn at 2.5, 3.0, 3.5, and 4.0 days. Flow cytometry and chorophyll *a* fluorescence measurements were taken every 12 h, corresponding to dusk and dawn during the entire nitrate depletion experiment.

Flow cytometry: cell counts, chlorophyll autofluorescence and forward scattering

We used an Accuri C6 flow cytometer (BD) to determine cell density, chlorophyll autofluorescence per cell, and forward scattering per cell. 2 mL of culture were passed through a 30 μm filter (Myltenyi Biotec, #130-101-812) to remove any debris prior to analysis and diluted

in filtered artificial seawater to maintain less than 2000 detected events sec^{-1}. The flow rate was 35 $\mu L\ min^{-1}$, and the core size was 16 μm. Relative chlorophyll autofluoresence was determined by excitation with a 488 nm laser (emission > 670 nm). *Phaeodactylum* cell counts μL^{-1} were determined by gating for particles with chlorophyll autofluoresence.

Nitrate depletion experiment cell harvesting

Cells harvested during nitrate depletion did not adequately pellet in 50 mL conical tubes; therefore, we collected cells by filtration. About 2.0×10^8 *Phaeodactylum* cells were harvested by filtering a culture volume (ranging from 50 to 150 mL, dependent on cell density) onto a polycarbonate filter (EMD Millipore, #ATTP02500). Filters were then placed in a beaker and cells were removed from the filter by repeated pipetting with 2 mL filtered artificial seawater. The 2 mL suspension was transferred into two 1 mL aliquots in microcentrifuge tubes. These resuspended cells were pelleted by centrifugation (20,000 × g, 10 min, 18 °C) and the supernatant was meticulously removed. These pellets were stored at -80 °C until further processing. One pellet was used for quantification of carbohydrate fractions (described above) and the other was used for triacylglycerol quantification (described below).

Total organic carbon and total nitrogen quantification

Total organic carbon (TOC) and total nitrogen (TN) were determined using a Shimadzu TOC-L Laboratory TOC Analyzer. TOC was determined by the difference method, where inorganic carbon (IC) was subtracted from total carbon (TC) to yield TOC. Ten mL of culture was harvested at dusk and dawn during the nitrate depletion experiment and placed into 50 mL conical tubes. These samples were pelleted by centrifugation (12000 × g, 15 min, 18 °C) and washed once with filtered artificial seawater, then resuspsended in 10 mL deionized water. A 1 mL aliquot was taken to determine cell density by flow cytometry. The remaining 9 mL was

frozen at -80 °C until the day of analysis. All glass and plasticware was acid-washed with 10 % HCl prior to use to remove any residual organic carbon. On the day of analysis, the 9 mL samples were thawed to room temperature. Eight mL was transferred into TOC glass vials and were diluted to a final volume of 40 mL with deionized water. The injection volume was 100 µL, and the instrument perfomed technical replicates of the injection such that duplicate values of less than 10% difference were used to determine TC, IC and TN for each sample. TC, IC, and TN concentrations were then divided by the number of cells in the 8 mL sample to determine TOC and TN per cell.

Chlorophyll quantification

10 mL of *Phaeodactylum* cells were collected and stored at -80 °C. Samples were thawed the day of analysis and filtered through a glass fiber filter (EMD Millipore, AP2001300). Pigments were extracted from glass filters by resuspension with 1 mL methanol, vortexed for 15 s, and incubated in the dark for 15 min. Cell debris and residual glass fibers were removed by centrifugation (15000 × g, 10 min, 4 °C) and the pigment-containing supernatants were pipetted into disposable acrylic cuvettes for analysis. No visible color remained on the filters. Chlorophyll *a* and chlorophyll $c_1 + c_2$ were determined using established spectrophotometric methods [152]. Chlorophyll *a* per cell was determined by dividing the chlorophyll *a* concentration by the total number of cells in the harvested sample, calculated from cell density and the volume extracted.

Monitoring neutral lipids with BODIPY and triacylglycerol quantification

Neutral lipid accumulation patterns were observed during the nitration depletion experiment using the fluorescent dye BODIPY (Ex = 503 nm, Em = 512 nm, Life Technologies, #D-2184). A single solution of 100 µg mL^{-1} BODIPY in dimethyl sulfoxide was prepared and divided into single-use aliquots stored at -80 °C for all timepoints during the nitrate depletion

experiment to minimize variability of the dye from handling and freeze-thaw cycles. Samples were prepared for flow cytometry as described above. A BODIPY aliquot was thawed immediately before analysis and 6.7 μL of the BODIPY solution was added to a 1 mL filtered culture sample, yielding a final concentration of 0.067 μg mL^{-1} [153]. Samples were vortexed for 3 s, incubated in the dark for 4 min, then briefly vortexed again prior to analysis by flow cytometry. BODIPY fluorescence was determined by excitation with a 488 nm laser and collected by a 533/30 nm bandpass filter, enabling quantification of mean BODIPY fluoresence per cell.

Triacylglycerols (TAG) were quantified from a neutral lipid extraction, as described previously [154]. Briefly, cell pellets were prepared for analysis and neutral lipids from these pellets were extracted via solid phase extraction, as described previously [155]. Eluted TAG samples were placed in a random order on a silica plate for thin layer chromatography and quantification. TAG was quantified from a standard curve of tripalmitin (1,2,3-trihexadecanoyl-glycerol, TAG 16:0 16:0 16:0, Avanti Polar Lipids, #111000) using ImageJ, version 1.48 (http://imageJ.nih.gov/ij/). Values reported as tripalmitin equivalents (t.e.) per cell.

Chlorophyll a fluorescence measurements

Five mL of cell culture was harvested at dawn and dusk timepoints during the nitrate depletion experiment for variable chlorophyll fluorescence measurements to determine F_v/F_m with a Walz DUAL-PAM 100 fluorometer. Samples were acclimated at 9 μmol photons m^{-2} s^{-1} for 30 min to relax nonphotochemical quenching prior to F_v/F_m measurement [156]. Samples were filtered onto a glass fiber filter (EMD Millipore, AP2001300) for measurement. Measuring light intensities were adjusted so that the baseline fluoresence was approximately 0.20 V prior to the saturating light pulse.

Statistical analysis

A paired student's t-test ($\alpha = 0.05$) was used to compare quantities of carbohydrate pools between dawn and dusk. For the nitrate depletion timecourse, a repeated measures one-way ANOVA ($\alpha = 0.05$) followed by a Tukey's HSD post-hoc test was performed to describe statistically different groups. All statistical analysis was performed using Sigmaplot (Systat Software).

Results

Analytical chemistry of soluble polysaccharide extracted from *Phaeodactylum*

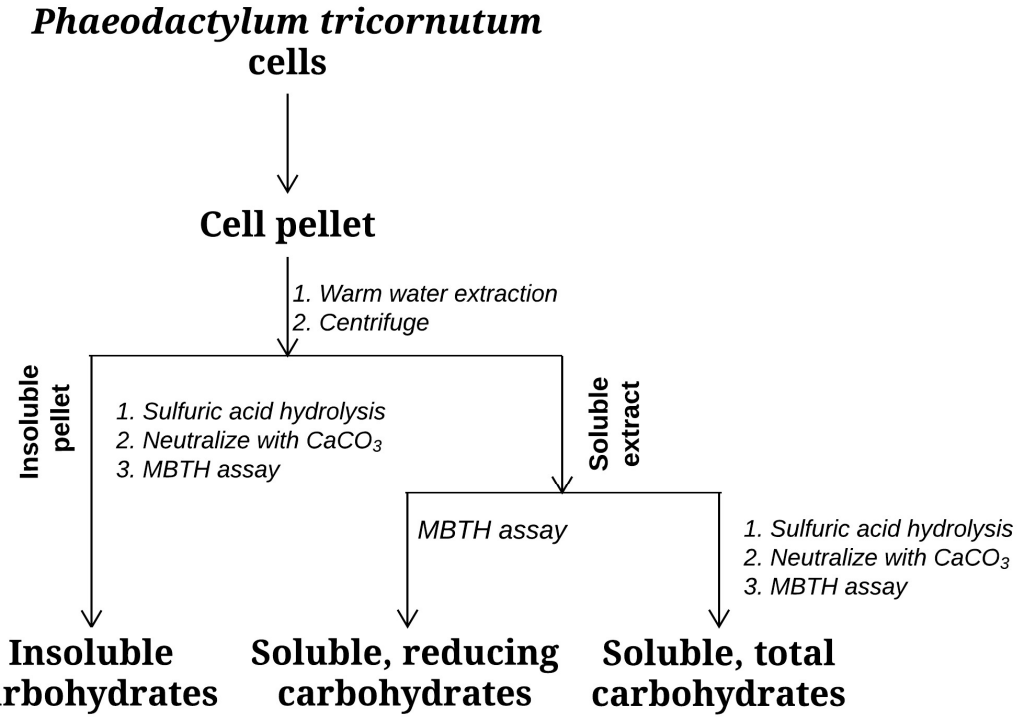

Figure 2.1: Overview of carbohydrate pools resolved by an integrative MBTH method. Insoluble sugars; Soluble, reducing sugars; and Soluble, total sugars are resolved by a warm water extraction, selective hydrolysis, and quantification with the MBTH reducing sugar assay. Compositionally, these three fractions represent structural carbohydrates, monosaccharides, and monosaccharides plus polysaccharides, respectively. The polysaccharide fraction (soluble, nonreducing carbohydrates) is determined by subtracting the soluble, reducing concentration from the soluble, total concentration.

Three major carbohydrate fractions were resolved with our method: insoluble; soluble, reducing; and soluble, total carbohydrates (Figure 2.1). A critical methodological assumption is

that chrysolaminarin is present in the soluble, nonreducing carbohydrate determined from subtracting soluble, reducing sugars from soluble, total carbohydrates. To test this assumption, we set out to analyze the soluble, nonreducing carbohydrate. We tested a range of warm water temperatures in order to extract soluble carbohydrates from *Phaeodactylum*. We found that the greatest values of soluble carbohydrate were found at extraction temperatures above 50 °C, which therefore represents an appropriate soluble carbohydrate extraction temperature for *Phaeodactylum* (Figure 2.2).

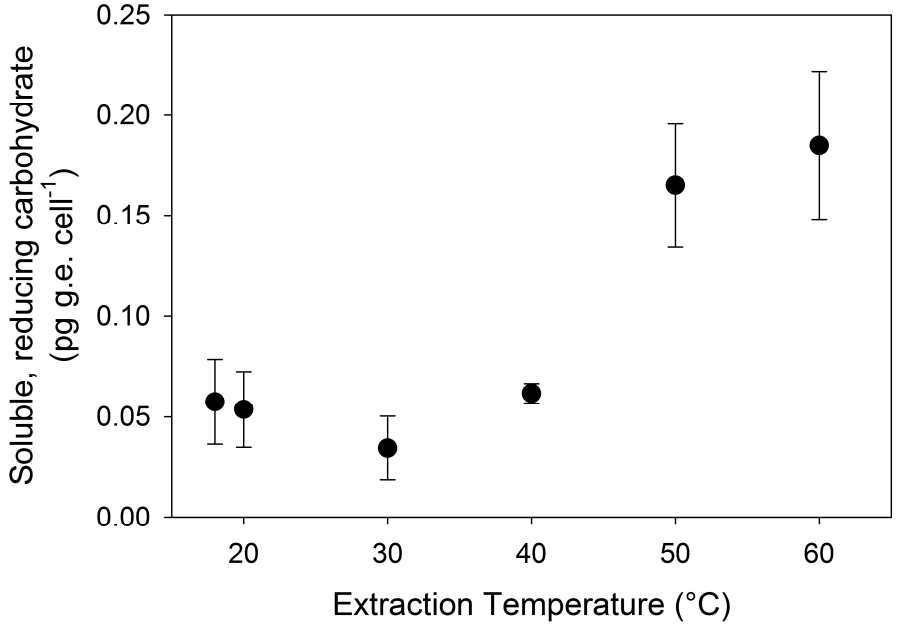

Figure 2.2: Soluble, reducing carbohydrate content extracted from *Phaeodactylum tricornutum* across a temperature gradient.
(n = 2-3, error = range or standard deviation, error bars may be smaller than symbol).

We enriched a *Phaeodactylum* polysaccharide from this soluble extract for compositional and structural analysis. The first round of precipitation contained 72 mg glucose equivalents (g.e.) carbohydrate from the combined 4 L batch culture as determined by MBTH quantification of soluble, total carbohydrate. 57 mg g.e. of soluble, total carbohydrate remained after enzymatic digests, ethanol precipitation, and dialysis, and 45 mg g.e. remained after two rounds of FPLC purification (Figure 2.3). FPLC fractions were pooled, dialyzed and lyophilized, yielding a dry

weight of 42.1 mg of material that was 97.6 % pure carbohydrate by dry weight. The final yield of purified polysaccharide was 58 % of the initial soluble extract quantity, with most of the losses occurring during the early ethanol precipitation steps prior to FPLC purification.

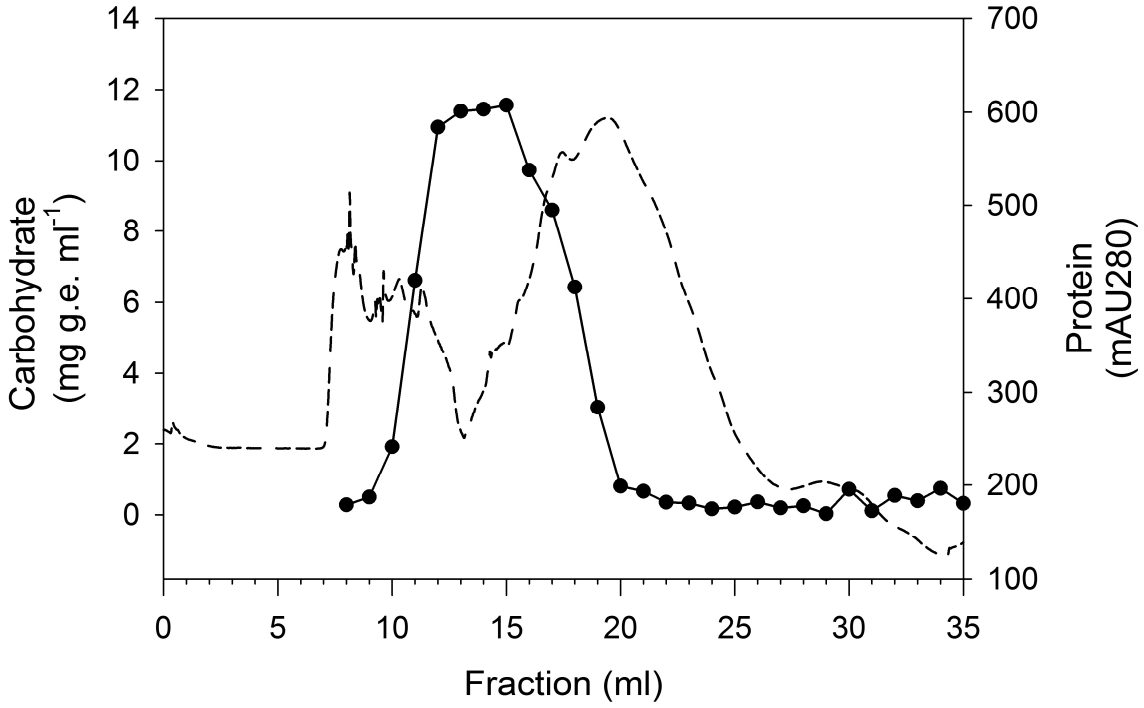

Figure 2.3: Size fractionation of carbohydrate and protein from a *Phaeodactylum* dialysis retentate on a Superdex 75 column.
Carbohydrate (•) and protein (---) content are presented by fraction number. An aliquot from each fraction enabled carbohydrate quantification with the MBTH method, while protein was monitored in line on the FPLC. Fractions 12-15 were pooled, lyophilized, and used for compositional and structural analysis.

Monomer analysis of the purified polysaccharide was determined via GC/MS of alditol acetate sugar derivatives (Figure 2.4, A). Our positive control, laminarin, was comprised of glucose (Figure 2.4, B). The *Phaeodactylum* purified polysaccharide was composed of glucose (Figure 2.4, C). The internal standard, 3-O-methylglucose, was observed in all samples.

We also performed structural analysis of the purified polysaccharide using ^{1}H-NMR. This analysis relied on existing annotations and chemical shift assignments [151]. A model β-glucan with individually labeled anhydrous glucose units (AGU) is shown as a structural reference

(Figure 2.5, A). Protons from each of these unique anhydrous glucose units are resolved in the β-glucan ^{1}H-NMR spectra as a doublet at an expected chemical shift. The spectra of laminarin (Figure 2.5, B) and *Phaeodactylum* purified polysaccharide (Figure 2.5, C) are presented.

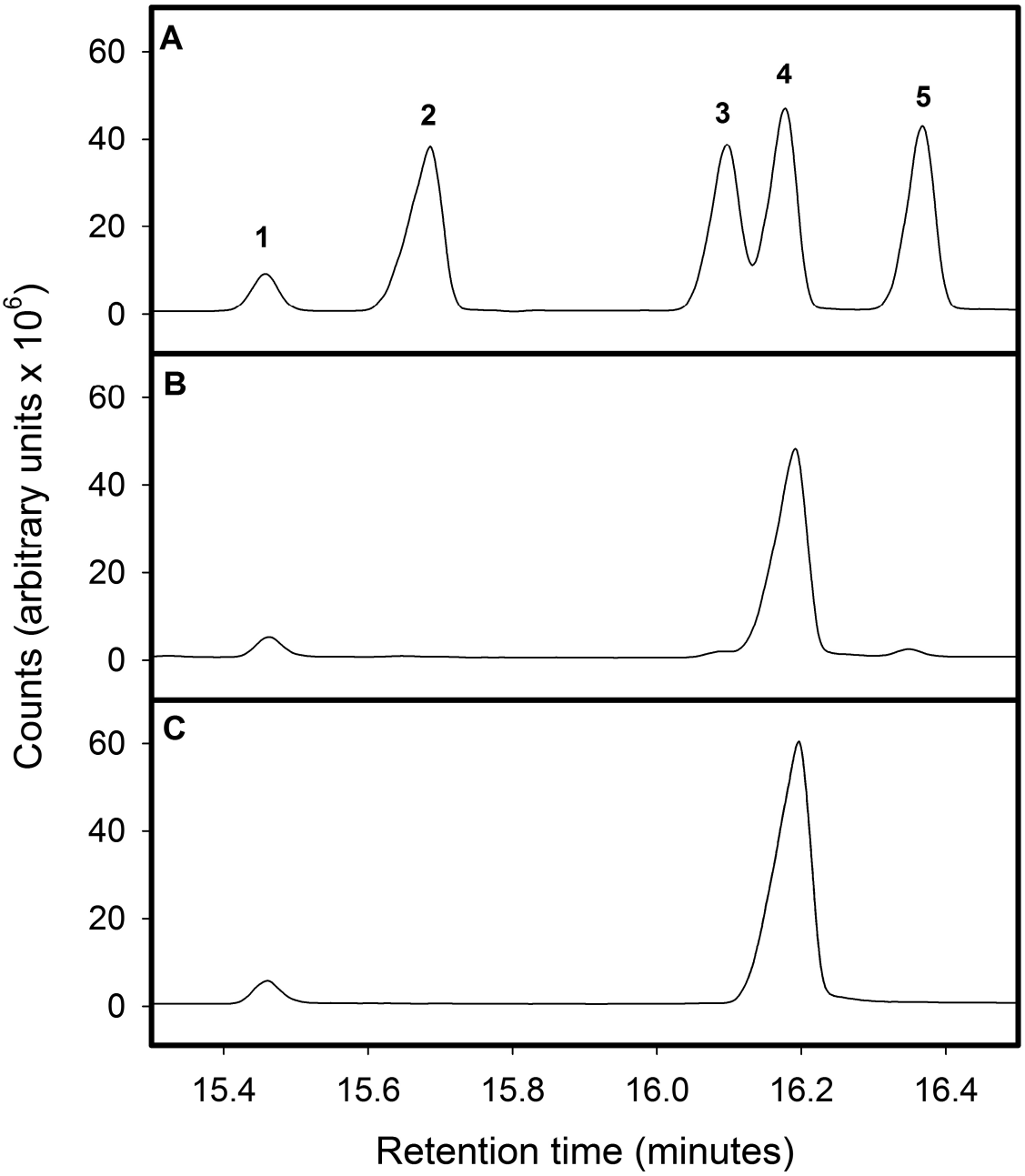

Figure 2.4: Hexose composition analyses of laminarin and soluble glucan extracted from *Phaeodactylum tricornutum*.

Gas chromatography traces of the neutral sugar standard (A), laminarin (B), and soluble glucan extracted from *Phaeodactylum tricornutum* (C). The neutral sugar standard contains the internal standard 3-*O*-methylglucose (1), myo-inositol (2), mannose (3), glucose (4), and galactose (5).

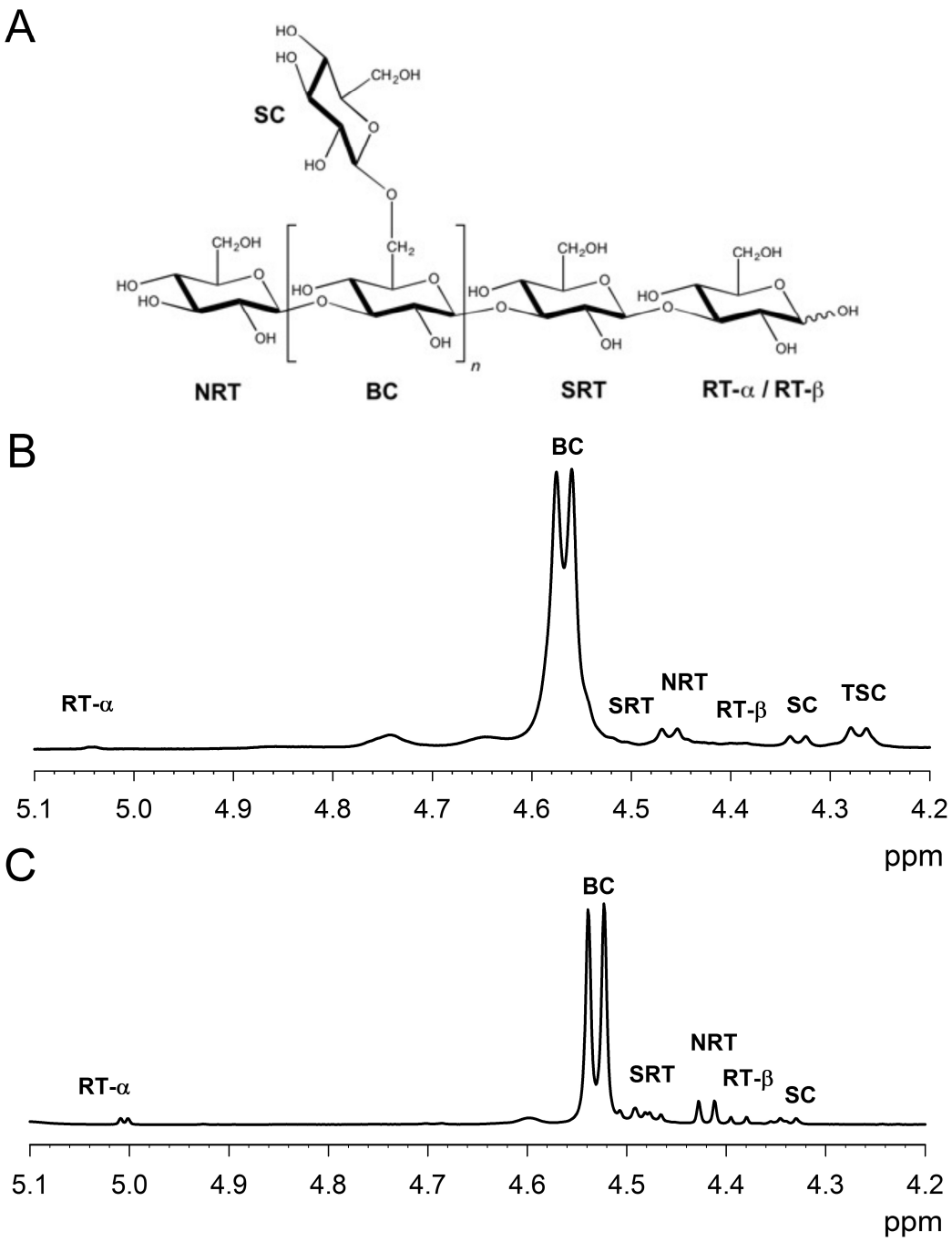

Figure 2.5: 1H-NMR structural analysis of soluble polysaccharide extracted from *Phaeodactylum tricornutum.*

A reference model of a β-glucan with labeled anhydrous glucose units (AGU) whose unique ^{1}H correspond to previously annotated chemical shifts (A). The ^{1}H-NMR spectra of laminarin (B) and the purified, soluble polysaccharide from *Phaeodactylum tricornutum* (C) in DMSO-d_6 at 80 °C are also presented. The major glucose doublet is the backbone chain (BC) and minor peaks include reducing terminal (RT, either α or β anomer), second AGU adjacent to the reducing terminal (SRT), nonreducing terminal (NRT), side chain (SC), and the terminal AGU of the SC (TSC).

The largest annotated doublet corresponded to the backbone chain (BC) at ~4.54 ppm in both spectra. A doublet for the terminal side chain AGU is not observed in the *Phaeodactylum* soluble polysaccharide, indicating a single AGU for the side chain. Analysis of the integrals of the doublets permits structural inference of the glucans (Table 2.1). Laminarin has 28 glucose residues as determined by degree of polymerization (DP) with a degree of branching (DB) of 0.063. These values are higher than the *Phaeodactylum* purified polysaccharide with 17 glucose residues and a DB of 0.015. The number of branches per each molecule can be estimated by multiplying DP and DB. Laminarin has about 2 branches per polysaccharide while only about one in four of the *Phaeodactylum* purified polysaccharide has a side chain.

Table 2.1: Structural analysis of laminarin and soluble polysaccharide extracted from *Phaeodactylum tricornutum* using annotated β-glucan signatures.
[a]DP – degree of polymerization, [b]DB – degree of branching, [c]number of branches per molecule

	Kim et al. (2000)	**This study**	
substrate	*laminarin*	*laminarin*	*soluble glucan (P.t.)*
DP[a]	33	28	17
DB[b]	0.07	0.063	0.015
DP × DB[c]	2.3	1.7	0.26

Accumulation of carbohydrate fractions differs in day/night cycles

We compared our three experimentally resolved carbohydrate fractions at dawn and dusk (Figure 2.6). There was no significant change of insoluble carbohydrates per cell between dawn and dusk ($p = 0.80$). A significant decrease of soluble, reducing carbohydrates was observed between dawn and dusk ($p < 0.05$). However, the largest relative and absolute differences

observed were for soluble, nonreducing carbohydrates. This component dropped from 1.56 ± 0.25 pg g.e. per cell at dusk to 0.06 ± 0.03 pg g.e. per cell at dawn (n = 3, $p < 0.05$).

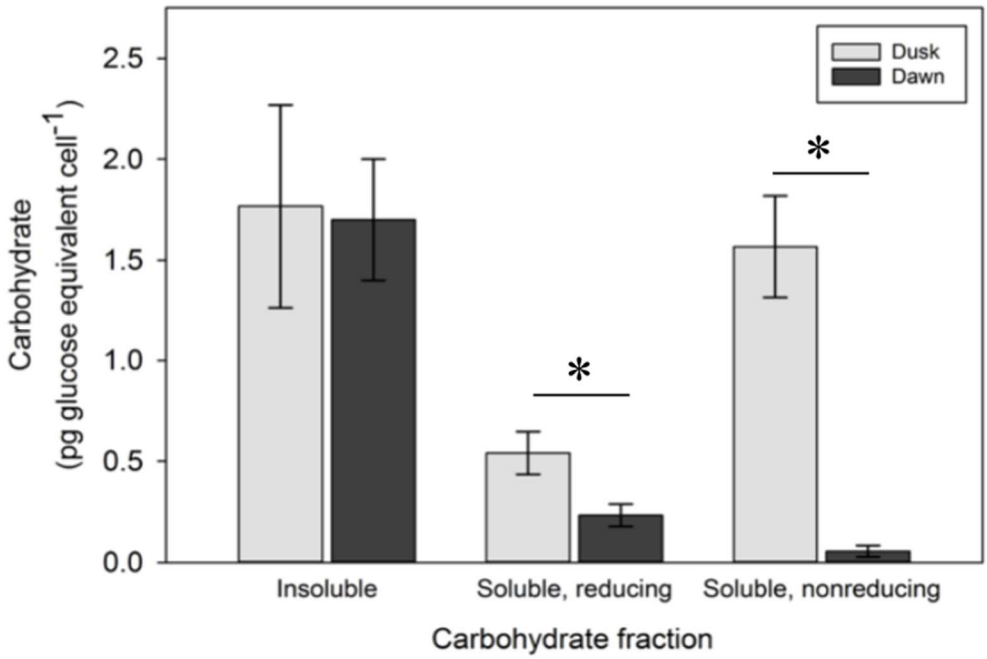

Figure 2.6: Average carbohydrate fraction quantities extracted from *Phaeodactylum tricornutum* at dawn and dusk.
(n = 3, error = standard deviation; Paired student's T-test, * = $p < 0.05$).

Nitrate depletion cell physiology and carbon partitioning into carbohydrate fractions

We investigated cell physiology during the first 3 days of nitrate depletion. Cell densities increased in nitrate replete medium, but the increase of cell densities stopped 2 days following transfer into nitrate deplete medium (Figure 2.7, A). The forward scatter (FSC) measured by flow cytometry is a proxy for cell size, since the relationship between FSC and cell size is not linearly proportional for pennate diatoms [157]. The FSC of *Phaeodacylum* oscillated in nitrate replete medium, with a maximum at dusk and a minimum at dawn (Figure 2.7, B). The oscillation pattern ceased after 2 days in nitrate deplete medium.

TOC per cell significantly decreased in the dark under nitrate replete conditions but the daily increase in TOC was not observed in nitrate deplete conditions (Figure 2.7, C, $p < 0.05$).

The decrease of TN per cell in nitrogen replete medium at night was small but statistically significant (Figure 5D, $p < 0.05$). TN per cell decreased during nitrogen deprivation to about 0.44 pg TN per cell ($p < 0.05$).

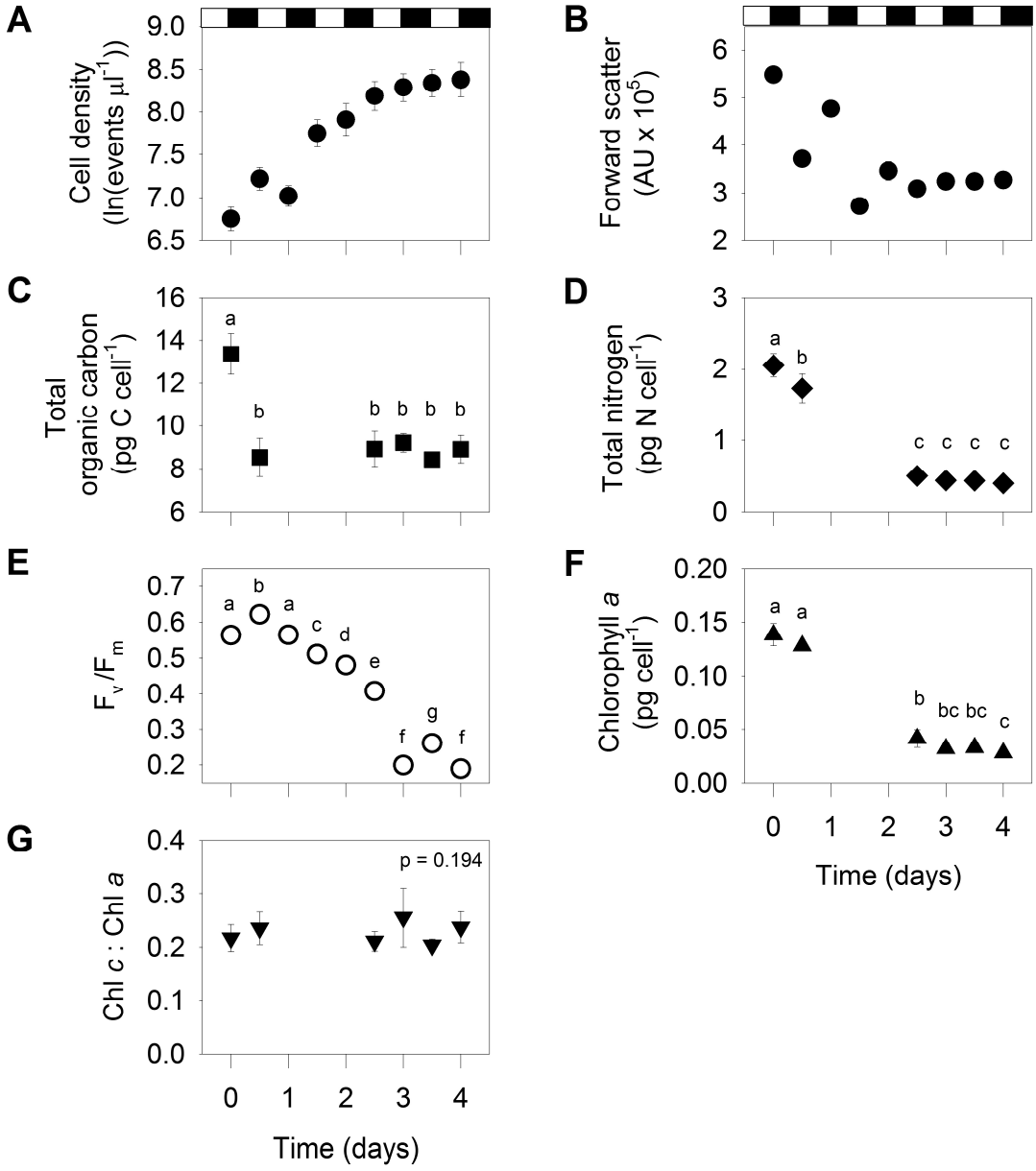

Figure 2.7: Cell physiology changes during nitrate deprivation.
Cell density (A), cell size approximated by mean forward scatter per cell (B), total organic carbon (C), total nitrogen (D), F_v/F_m (E), chlorophyll *a* (F), and the chlorophyll *c* : chlorophyll *a* ratio (G) are depicted above. Black and white bars at the top of the figure represents day and night, respectively. Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a repeated measures one way ANOVA (n = 5, error = standard deviation, error bars may be smaller than symbol).

F_v/F_m, a measurement of the quantum efficiency of photosystem II [158], dropped from 0.6 to 0.2 after two days of nitrate deprivation (Figure 2.7, E, $p < 0.05$). Chlorophyll *a* was maintained between 0.13 – 0.14 pg per cell in nitrate replete media but decreased after 2 – 3 days without nitrate (Figure 2.7, F, $p < 0.05$). No significant difference was observed for the chlorophyll $c_1 + c_2$: chlorophyll *a* ratio at any timepoint during the experiment (Figure 2.7, G, $p = 0.19$).

We observed the partitioning of carbon within different carbohydrate fractions and neutral lipids during nitrate depletion. Soluble, nonreducing carbohydrate dropped significantly in the dark from 2.32 pg g.e. per cell to 0.36 pg g.e. per cell but remained between 1.04 and 1.34 pg g.e. per cell 2 – 3 days after nitrate removal (Figure 2.8, A, $p < 0.05$). Soluble, reducing sugars also significantly decreased at night in nitrate replete medium (Figure 2.8, B, $p < 0.05$). Soluble, reducing sugars accumulated at slightly lower quantities after 2 – 3 days in nitrate deplete medium than the maximum in nitrate replete medium. A small, significant drop of insoluble carbohydrate was observed in nitrogen replete growth between dawn and dusk, but also remained at a steady level in nitrate deplete conditions, between 0.82 and 0.86 pg g.e. per cell (Figure 2.8, C, $p < 0.05$). The amount of soluble, reducing sugars did not change throughout nitrogen starvation (Figure 2.8, D, $p < 0.05$).

BODIPY fluorescence per cell, which is commonly used as an indicator of neutral lipid content, oscillated with maxima at dusk and minima at dawn for the entire experiment (Figure 2.8, E, $p < 0.05$). However, quantification of TAG per cell suggested that the oscillating neutral lipid abundance is limited to nitrate replete conditions (Figure 2.8, F, $p < 0.05$). During nitrate starvation, TAG per cell accumulated to about 6 pg t.e. per cell and did not differ between dawn and dusk.

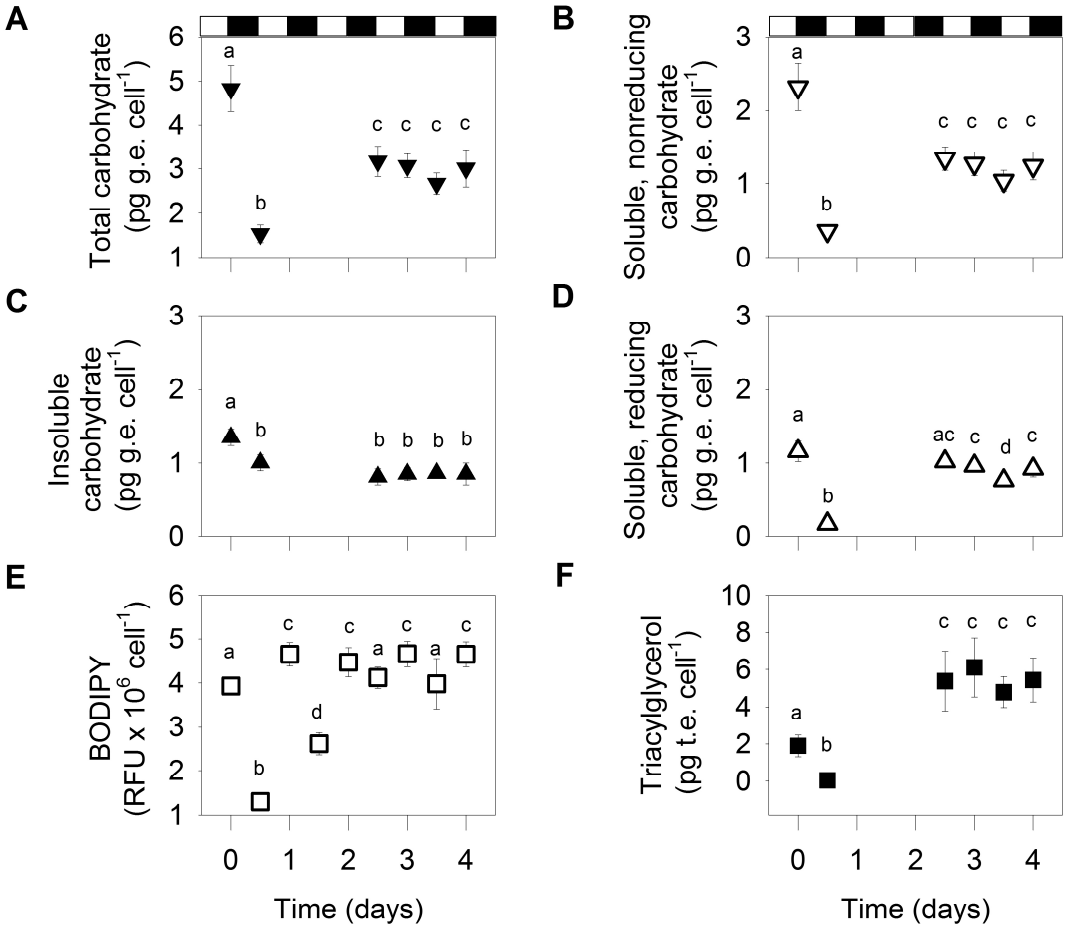

Figure 2.8: Quantifying carbohydrate fractions and neutral lipid during nitrate deprivation.

The sum of all quantified carbohydrate fractions (A) is shown alongside soluble, nonreducing (B), insoluble (C), and soluble, reducing (D), carboydrates and are reported as glucose equivalents (g.e.). Mean BODIPY fluorescence (E) and triacylglycerol (F) in tripalmitin equivalents (t.e.) are also reported. Black and white bars at the top of the figure represents day and night, respectively. Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a repeated measures one way ANOVA (n = 3-5, error = standard deviation, error bars may be smaller than symbol).

Soluble, nonreducing sugars and TAG were normalized for their mass-fraction of carbon, 0.4 units (glucose, $C_6H_{12}O_6$) and 0.76 units (tripalmitin, $C_{51}H_{98}O_6$) respectively (Figure 2.9). Carbon partitioning into soluble, nonreducing sugar was comparable to triacylglycerols at dusk under nitrate replete conditions, accounting for 6.9 % and 10 %, respectively. However, after 2 – 3 days of nitrate removal, soluble, nonreducing carbohydrate only represents about 5.5 % of TOC whereas TAG dominates carbon partitioning, accounting for about 45 % of TOC.

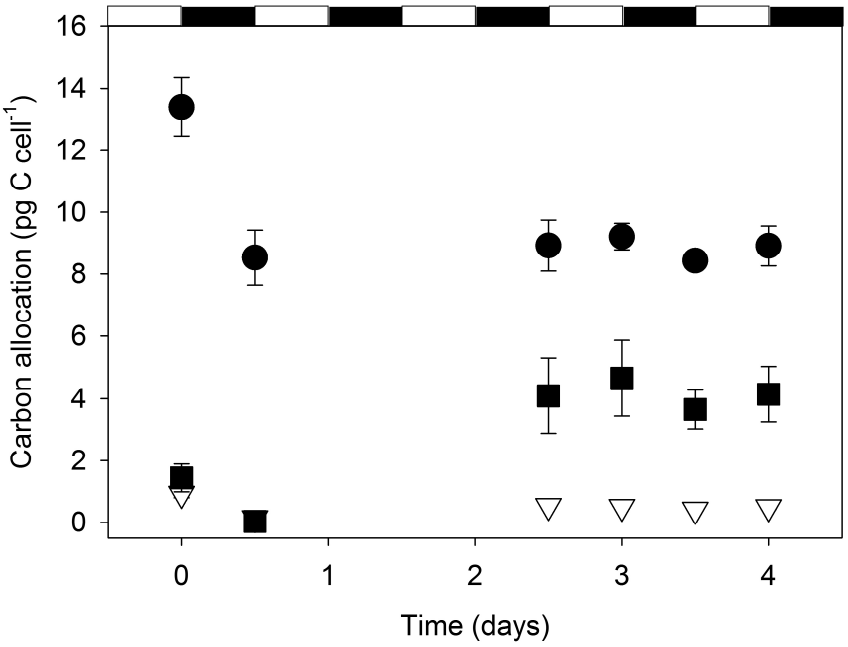

Figure 2.9: Carbon partitioning into storage metabolites during nitrate deprivation. Black and white bars at the top of the figure represents day and night, respectively. Total carbon per cell is included as a reference (circles) for the total amount of carbon contributed by the storage metabolites triacyglycerol (squares) and soluble, nonreducing carbohydrate (inverted triangles). Mean triacylglycerol and soluble, nonreducing carbohydrate per cell were multiplied by carbon mass fractions for tripalmitin ($C_{51}H_{98}O_6$; 0.76) and glucose ($C_6H_{12}O_6$; 0.4) (n = 3-5, error = standard deviation, error bars may be smaller than symbol).

Discussion

Soluble, nonreducing carbohydrate fraction contains chrysolaminarin

Chrysolaminarin was originally identified as a soluble polysaccharide, entirely composed of glucose, with limited branching [88]. Our extraction workflow began by extracting saccharides from whole cells at 50 ºC. This bulk extract does not contain any signature components of *Phaeodactylum* extracellular polysaccharides, which are composed of mannose and glucose with complex linkages and branching [159]. Instead, our isolated glucan from the soluble extract was entirely composed of glucose (Figure 2.4), agreeing with past studies of chrysolaminarin [88, 116]. The purified glucan exhibited ^{1}H-NMR signatures characteristic of β-1,3 glucans (Figure 2.5). However, purification accounted for a loss of 15 mg g.e of.

uncharacterized glucan from dialyzed retentate to purified glucan, or about 26 %. It is possible that this lost fraction contains some alternative soluble glucose polymer whose identity is unknown.

The ^{1}H-NMR spectrum also permits inference of structural characteristics of the glucan. The integrated, annotated peaks can be used to determine the degree of polymerization (DP) and the degree of branching (DB) [151]. Structural features significantly vary by diatom species, even between closely related organisms [115-117, 128]. We found that *Phaeodactylum's* chrysolaminarin is smaller and lightly branched compared to laminarin, which is consistent with earlier observations. [88] (Table 2.1). The structure of the β-1,3/β-1,6 glucan means that for an average molecule of chrysolaminarin, 16 of 17 glucose units are non-reducing sugars (94 %) (Table 2.1). Compositional and structural analysis of the *Phaeodactylum* glucan supports our assumption that the soluble, nonreducing carbohydrate fraction contains chrysolaminarin. We recommend similar analytical steps be taken when applying this method to species other than *Phaeodactylum*.

Chrysolaminarin, is thought to serve as a respiratory substrate in the dark [154]. Our results demonstrate that soluble, nonreducing carbohydrate dramatically decreases between dusk and dawn, especially relative to the other carbohydrate pools (Figure 2.6). We also found this fraction significantly increased with growth irradiance (Figure 2.10). These observations further support our assertion that the soluble, nonreducing fraction is mostly chrysolaminarin. We note that our values are lower than those found by Chauton et al. [99] from *Phaeodactylum* cells entrained to day/night cycles. The extraction methods used by Chauton et al. [99], which contains dilute sulfuric acid, have been shown to extract other non-glucose containing carbohydrates in addition to chrysolaminarin [117] and the colorimetric assay utilized to quantify

reducing sugars (phenol-sulfuric acid) provides poor precision for a complex mixture of sugars [160]. In light of our analytical chemistry and physiology results for the soluble, nonreducing carbohydrate, we suggest that this fraction represents a good proxy for quantifying chrysolaminarin and hereafter refer to this as chrysolaminarin.

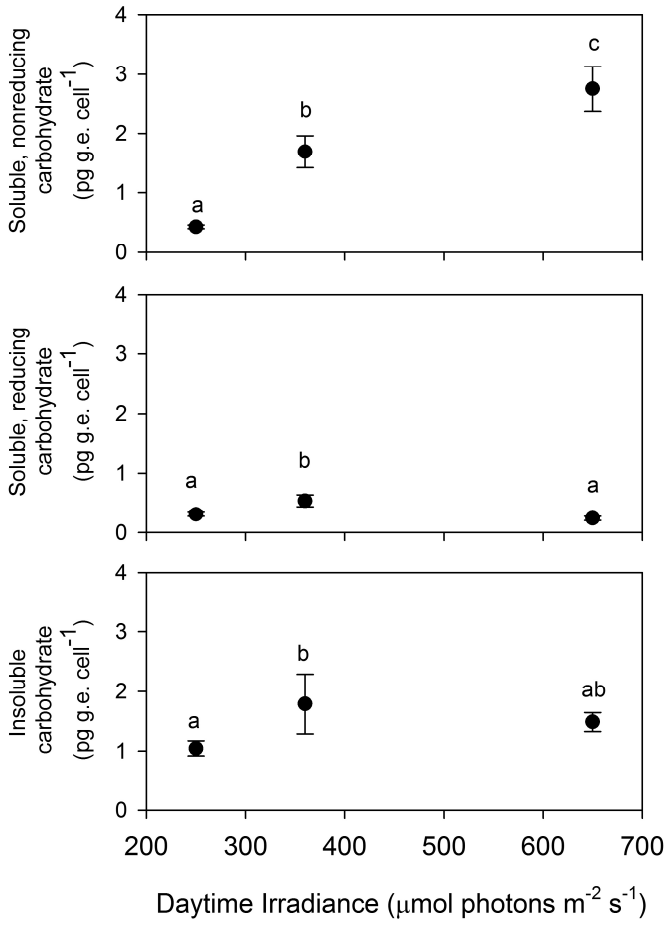

Figure 2.10: Comparing major carbohydrate fractions grown under different daytime irradiances.
Cells were grown in 16:8 day:night cycles and all pellets were harvested at dusk. Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a repeated measures one way ANOVA (n = 3, error = standard deviation, error bars may be smaller than symbol).

Repartitioning carbohydrate pools during nitrogen starvation.

We sought to understand the dynamics of chrysolaminarin storage in response to nitrogen starvation. Our biochemistry and photophysiology data show a daily oscillation in

Phaeodactylum cell composition and size prior to exchange into nitrogen free media. Cell density increased (Figure 2.7, A) and cell size (assumed to be proportional to forward scatter) oscillated depending on time of day (Figure 2.7, B). The quantum efficiency of photosystem II remained high (Figure 2.7, E), as did chlorophyll *a* per cell (Figure 2.7, F). We, and others, have previously observed synchronized cell division in *Phaeodactylum* growing in a day/night cycle [99, 154]. An increase of cell TOC during the day, particularly as chrysolaminarin and TAG, (Figure 2.8) provides carbon skeletons that support cell division and heterotrophic metabolism at night during nutrient replete growth. The consumption of these substrates in the dark to fuel heterotrophic metabolism has been described previously [154, 161].

The daily oscillation observed in nutrient replete conditions ceased in nitrogen-deplete conditions. This was observed after 2 – 4 days in nitrate-free medium. Increases in cell concentration slowed (Figure 2.7, A) and oscillations in cell size stopped (Figure 2.7, B). While TN (Figure 2.7, D), F_v/F_m (Figure 2.7, E), and chlorophyll *a* (Figure 2.7, F) dramatically decreased, the chlorophyll $c_1 + c_2$: chlorophyll *a* ratio did not change (Figure 2.7, G), in agreement with previous nitrate starvation observations [162].

The patterns of carbon partitioning between carbohydrate fractions and TAG during nitrogen starvation were unforeseen. The dusk/dawn cellular content of carbohydrate did not change during nitrogen starvation and did not increase relative to nutrient replete growth (Figure 2.8). This was also observed for each experimentally resolved carbohydrate pool, including the cellular content of chrysolaminarin (Figure 2.8, B, timepoints 4-6) which only represented 5.5 % of total cellular carbon during nitrogen limitation (Figure 2.9). TAG content reached its highest values on a per cell basis (Figure 2.8, F) and also as a total fraction of the TOC (Figure 2.9).

These data show that TAG, not chrysolaminarin, is the primary repository of storage carbon for *Phaeodactylum* during nitrate starvation in day:night cycles.

While it has been determined that total carbohydrate in *Phaeodactylum* decreases during nitrate starvation [132, 163], the dynamics of individual carbohydrate pools have previously only been inferred through proteomics and transcriptomics. For instance, Hockin et al. [57] performed a comparative proteomics experiment at the onset of nitrogen starvation, where they observed a variety of enzymes predicted to participate in carbohydrate metabolism significantly changed. Some of the proteins that increased in relative abundance were glycolysis-specific, suggesting a net shift of carbon from chrysolaminarin reserves into free sugars for metabolism. Transcripts for glycolytic enzymes also increase in abundance during nitrate depletion [132, 138, 144]. Mus et al. [164] also noted an increased abundance of transcripts that are bioinformatically predicted to be related to chrysolaminarin metabolism during nitrate starvation. We note that all experimentally resolved carbohydrate fractions decreased during N-starvation in our study so we posit that some of these transcripts might be associated with the metabolism of insoluble carbohydrates.

One proposed strategy to enhance lipid productivity in algae is to interfere with polysaccharide synthesis [126]. Indeed, photoheterotrophically grown *Chlamydomonas* strains that are unable to produce starch greatly increase the amount of cellular carbon in TAG [33, 143]. This change in carbon partitioning may be further enhanced by nitrogen deprivation, when TAG is significantly accumulated [125]. We found that *Phaeodactylum* only accumulates 4.9 – 6.0 % of cellular carbon as chrysolaminarin during nitrate starvation, a relatively small fraction compared to the TAG accumulation of 43 – 50 % of cellular organic carbon (Figure 2.9). We note that a large proportion of cellular carbon is used in protein and structural lipids [154], but

we did not quantify those pools in this study. Therefore, the rational engineering strategy of manipulating storage carbohydrate metabolism to increase TAG productivity in *Phaeodactylum* might not be appropriate. However, it is important to note that the accumulation of soluble carbohydrate varies significantly by algal species [165]. For instance, *Skelotonema costatum* accumulated soluble carbohydrate during nitrate depletion in a diel light regime [3]. Additionally, N-starved *Isochysis sp.* accumulated total carbohydrate without daily oscillations compared to N-replete cultures [137], but it is unclear how much of the observed carbohydrate was chrysolaminarin in either of these studies. We caution that careful attention to species-specific physiology and growth conditions must be made when predicting the potential yield changes associated with the engineering of metabolism.

Conclusions

We report a method that enables parallel characterization of insoluble; soluble, reducing; and soluble, nonreducing carbohydrates, representing an improvement over existing total carbohydrate assays. We have established that the soluble, nonreducing carbohydrate resolved by our MBTH-based method contains chrysolaminarin. Quantification of chrysolaminarin in parallel with other carbohydrate pools is a valuable addition to the biochemical toolkit available for investigating diatom metabolism. We found that high accumulation of chrysolaminarin in *Phaeodactylum* requires nutrient replete conditions, high irradiance, and occurs at dusk in a diel light regime. We illustrated the utility of this method by showing that: 1) chrysolaminarin accumulates at the end of the day and is consumed by dawn and 2) chrysolaminarin is not preferentially accumulated during nitrogen limitation. So, we stress that estimates of carbon partitioning to biofuel precursors should be made on cultures grown in day/night cycles. We also

suggest that rerouting of carbon flux away from chrysolaminarin reserves is not the most prudent biological engineering strategy for maximizing TAG productivity in this organism.

Afterword

This afterword addresses two discussion points during the dissertation defense from the publication above.

First, extraction temperatures were tested for as high as 60 °C, but no higher. This parameter was determined based off previous extraction protocols for *Phaeodactylum*. Chiovitti et al. (2004) performed a warm water extraction at 30 °C for 1 h [94], but tested for membrane disruption of *Phaeodactylum* cells up to 60 °C. Meanwhile, Granum and Myklestad (2002) performed a dilute sulfuric acid extraction at 60 °C for 10 min [127]. I relied on an extraction temperature of 50 °C for 15 min. It is possible that additional carbohydrate could be extracted from *Phaeodactylum* cells using incubation temperatures above 60 °C. However, it is possible that these extraction parameters might extract non-chrysolaminarin carbohydrate fractions, as described in the introduction above. Changes in the absolute amount of extracted carbohydrate would have to be complemented with verification of a chrysolaminarin-exclusive polysaccharide extraction, such as compositional analysis by GC/MS to ensure that extracted polysaccharides are entirely composed of glucose.

Second, the nitrogen starvation experiment halted the growth and decreased the quantum yield of photosynthesis of *Phaeodactylum* cells. This combination implies that limited amounts of new carbon were being fixed. This observation is significant from a carbon partitioning standpoint. In the green algae *Chlamydomonas reinhardtii*, it has been observed that TAG accumulation can be explained through the degradation of chloroplast lipids[166]. A similar phenomenon may be taking place in *Phaeodactylum* cells during nitrogen starvation. In

Phaeodactylum lipid profiles during nitrogen starvation, the chloroplast membrane lipids monogalactosyldiacylglycerol and phosphatidylglycerol decrease while TAG increases[125]. In other words, it is possible that TAG accumulation in *Phaeodactylum* might be partially explained by the recycling of chloroplast membranes during nitrogen starvation, rather than *de novo* TAG biosynthesis.

CHAPTER 3: ROLES OF DIATOM UDP-GLUCOSE SYNTHESIZING ENZYMES[2]

Preface

The previous chapter introduced a method which enabled chrysolaminarin quantification from *Phaeodactylum* cell extracts. One application of this methodology is to characterize chrysolaminarin accumulation mutants. The genetic understanding of chrysolaminarin biosynthesis is essentially unknown. This chapter set out to identify and characterize UDP-glucose synthesizing enzymes, thought to represent the first step of chrysolaminarin synthesis. These enzymes were knocked out using a gene editing approach and phenotyped, including the quantification of chrysolaminarin content. This chapter resolves ambiguities associated with the roles of these enzymes in chrysolaminarin synthesis.

Introduction

Stramenopiles are an incredibly diverse group of eukaryotes, which include not only photosynthetic species, such as diatoms, kelps, and chrysophytes, but also non-photosynthetic groups, such as the pathogenic oomycetes. Photosynthetic stramenopiles evolved through a secondary endosymbiotic event, in which a heterotrophic eukaryotic cell engulfed a red algae[12]. This evolutionary feature confers a substantially different genetic blueprint for photosynthetic stramenopiles when compared to Archaeplastidae such as plants, chlorophyte algae, or rhodophytes.

Stramenopiles store carbohydrates as an energy and carbon reserve. Diatoms synthesize a carbohydrate that is a water-soluble polysaccharide called chrysolaminarin. It is defined by its β-

[2] A version of this chapter is anticipated to be revised and submitted to *Algal Research*. "We" refers to anticipated co-authors: Michael Caballero, Denis Jallet, David Xing, Mark Moosburner, Andrew Allen, and Graham Peers.

1,3 linkages, limited β-1,6 branching, and is entirely composed of glucose[88, 124]. Kelps synthesize laminarin, which contains more glucose residues than chrysolaminarin[124], and it contains a small percent of mannose[87]. Oomycetes synthesize mycolaminarin, structurally and compositionally comparable to chrysolaminarin, except for evidence of its phosphorylation during different life cycle stages[89, 167]. In comparison, plants and chlorophyte algae synthesize the storage polysaccharide composed of glucose called starch[68]. It is a water-insoluble glucan with α-1,4 linkages and α-1,6 branching. These polysaccharide reserves play similar metabolic roles despite these differences.

Storage polysaccharides serve as a reserve of carbon skeletons and energy for metabolism in the dark. Diatoms synthesize both triacylglycerol and chrysolaminarin as carbon reserves. These reserves accumulate during the light period and are depleted over the course of the night. The diel phenomena of storage polysaccharide accumulation and depletion has been demonstrated in a variety of diatom species [3, 124]. Like other polysaccharides, chrysolaminarin appears to be made from activated nucleotide sugars, specifically UDP-glucose[90]. These physiological and biochemical studies provide a basis for the metabolic role and biosynthetic pathway of chrysolaminarin. However, studies investigating the genetic components that encode chrysolaminarin biosynthesis are limited.

We do not know the genes associated with chrysolaminarin biosynthesis in diatoms [62]. This is partially due to the evolutionary distance between these organisms and relatively well-characterized land plants. While the transient starch accumulation in plant leaves follows a comparable physiological pattern as algae[58], the biochemistry of starch synthesis is fundamentally different as it relies on an entirely different cohort of enzymes for its metabolism[168]. For example, starch is synthesized from ADP-glucose, not UDP-glucose.

Plants and green algae also synthesize callose, a β-1,3 glucan. However, this polysaccharide is a structural polymer associated with pollen development and wounding responses[169], rather than serving as an energy reserve. Nevertheless, functional genetic studies of starch and callose biology have enabled suggestions of chrysolaminarin-related genes in diatoms[91].

Two genes encoding enzymes capable of synthesizing UDP-glucose have been bioinformatically predicted in diatoms: *ugp1* (Phatrdraft_50444) and *ugp2* (Phatrdraft_23639)[91]. UGP1 appears to be a protein fusion of UDP-glucose diphosphorylase (UGPase) and phosphoglucomutase[91], and has been the target of several reverse genetics studies. While gene expression of *ugp2* has been shown to be coordinated in a diel manner, maximally expressed at the beginning and minimally expressed at the end of the light period[99], no genetic studies disrupting *ugp2* have been reported. Knockouts of *ugp1* contain significantly higher cellular triacylglycerol content, inferring a disruption of chrysolaminarin synthesis by the redirection of carbon away from chrysolaminarin and into triacylglycerols[7]. This phenotype was anticipated based on studies of green algae that displayed an increased accumulation of triacylglycercols when starch metabolism was disrupted [33]. More recently, a study of a suite of RNAi knockdowns targeting *ugp1* was reported by Zhu et al. (2015) [8]. While the characterized strains in this study were grown under low light (75 μmol photons m^{-2} s^{-1}) and harvested in stationary phase, Zhu et al. (2015) also observed an increase in lipid content as measured by total lipids as a percentage of dry weight, although dry weights were not reported. Importantly, Zhu et al. (2015) also quantified soluble carbohydrates, and observed as much as a 62 % reduction of soluble carbohydrates when normalized by dry weight. However, at relatively low daytime growth irradiances, the proportion of soluble carbohydrate as chrysolaminarin is expected to be

low[124]. Neither of the *ugp1* disruption studies quantified chrysolaminarin directly, so the contribution of the *ugp1* gene product to chrysolaminarin synthesis is still inferred.

These *ugp1* disruption studies investigated the potential contribution of UDP-glucose to chrysolaminarin metabolism. However, it is known that UDP-glucose plays multiple metabolic roles in a photosynthetic organisms[59]. For example, UDP-glucose is necessary for the synthesis of the sulfolipid sulfoquinovodiacylglycerol (SQDG)[6], which is an important structural lipid in the thylakoid membranes of the chloroplast. UDP-glucose also plays an essential role in nucleotide sugar metabolism for housekeeping metabolic roles, potentially feeding into cell wall polysaccharides[170] and N-glycan biosynthesis for the production of glycoproteins[171]. Overall, the contributions of other enzymes that can synthesize UDP-glucose should be considered for their potential roles in chrysolaminarin and/or in other unique metabolic pathways that require UDP-glucose.

Daboussi et al. (2014)[7] and Zhu et al. (2015)[8] assumed that the UGPase encoded by *ugp1* is the only contributor to the synthesis of UDP-glucose. UGPase catalyzes the production of UDP-glucose: [UTP + glucose-1-phosphate → UDP-glucose + PPi] (EC: 2.7.7.9). However, it is known that other enzymes closely related to UGPase can also synthesize UDP-glucose, such as UDP sugar diphosphorylases[61, 109] (EC: 2.7.7.64). Therefore, it is possible that additional diatom proteins may also contribute to chrysolaminarin synthesis.

The combination of genes specific to the *Phaeodactylum tricornutum* genome or specific to sequenced diatom genomes account for more than 57 % of the predicted genes in this organism [48], limiting our ability to elucidate unique diatom biochemical pathways such as chrysolaminarin synthesis. A strategy to functionally identify and characterize enzymes with the

ability to synthesize UDP-glucose may identify novel enzymes that could play a role in chrysolaminarin synthesis or in other areas of diatom metabolism.

It is important to resolve the roles of UDP-glucose in diatom metabolism to strengthen the body of evidence for the chrysolaminarin biosynthesis pathway. Genome-wide metabolic models of chrysolaminarin synthesis employ either the gene product of *ugp1*[96], *ugp2*[99], or both[54]. These conflicting metabolic models underscore the limited functional evidence for chrysolaminarin biosynthesis. Experimental evidence of the functional contributions of each of these gene products will help resolve this controversy. Systems biology studies would benefit from elucidating the roles of diatom UGPases at a gene-by-gene level as they rely on evidence of protein function to resolve metabolic pathways or patterns.

This work sets out to functionally identify enzymes with UGPase activity from the model diatom *Phaeodactylum tricornutum* and characterize them. *Phaeodactylum* is a model system for diatom genetics with a sequenced genome[48] and an array of established molecular tools[172] including genome editing with CRISPR/Cas9[101]. We adapted a functional selection for UGPase activity in *E. coli* using a *Phaeodactylum* cDNA expression library. We identified a novel *Phaeodactylum* gene encoding UGPase activity, hereafter called *ugp3* (Phatrdraft_54493). The proposed assignments of *ugp1*, *ugp2*, and *ugp3* were reexamined and updated using a phylogenetics approach. Knockouts of *ugp1*, *ugp2*, and *ugp3* were generated using a CRISPR/Cas9 approach. We observed significantly different phenotypic patterns between the *ugp1*, *ugp2*, and *ugp3* gene knockouts. These data provided functional evidence to contextualize the roles of UGP1, UGP2, and UGP3 in diatoms, informing not only genome-wide modeling efforts but also clarifying the chrysolaminarin biosynthesis pathway in diatoms.

Methods

Identification and characterization of *Phaeodactylum* UGPases

Culture conditions

Axenic *Phaeodactylum tricornutum* (strain CCAP 1055/1) was gifted from Andrew E. Allen (JCVI). Cells were grown in artificial seawater (35 ‰ salinity, Instant Ocean) plus nutrients at a 2.3-fold higher than f/2[149], to mitigate any nutrient limitation. Silica was omitted from the media as it is not required for *Phaeodactylum* growth. Two light regimes were used in this study. 1) A 12 h:12 h light:dark regime where the daytime light intensity was 200 ± 20 µmol photons m^{-2} s^{-1}, mixed on a shaking platform at a rate of 130 rpm and 2) continuous light with an irradiance of 150 ± 15 µmol photons m^{-2} s^{-1} in stationary cultures. All *Phaeodactylum* cultures were maintained at a temperature of 18 °C.

The *E. coli* strain DEV6 was acquired from the Coli Genetic Stock Center (Yale University, strain #5957). DEV6 was cultured in 1 x M9 salts [22 mM KH_2PO_4, 22 mM Na_2HPO_4, 85 mM NaCl, and 1 mM $MgSO_4$; Sigma, #M6030] + 0.1 % thiamine + 0.2 % carbon source (glucose or galactose, w/v); hereafter M9gal or M9glu. DEV6 was cultured at room temperature (25 °C).

TOP10 competent cells (ThermoFisher, #C404003) were used for plasmid construction of Gateway vectors. TOP10 cells were grown in LB media at 37 °C.

Construction of a Phaeodactylum cDNA library.

RNA was extracted from *Phaeodactylum* cells grown in four conditions to broadly capture the diatom's transcriptome. 1) *Phaeodactylum* cells harvested in exponential growth phase or 2) cells harvested in stationary growth under continuous light and 3) *Phaeodactylum* cells harvested 2 h after dawn or 4) harvested 2 h after dusk in the day:night light regime. 50 mL

of harvested cells were concentrated by centrifugation (2000 x g, 5 min, room temperature), transferred as a 1 mL volume into an Eppendorf tube, and pelleted (2000 x g, 5 min, room temperature). After discarding the supernatant, pellets were flash frozen in liquid N_2 and a liquid N_2 cooled metal bead (McMaster Carr, 1/8") was added to the frozen pellet. Cells were broken with a Tissue Lyser II (Quiagen; 1 min, 30 Hz). The algal powder was resuspended with 1 mL Trizol reagent (ThermoScientific, #15596026) and total RNA was extracted following the manufacturer's protocol. The quantity, purity, and integrity of total RNA was determined with a TapeStation RNA screen tape (Agilent, #5067-5576) on an Agilent 4200 TapeStation. Equal amounts of total RNA from each condition were pooled together and used for poly(A) RNA purification using Magnosphere^{TM} UltraPure mRNA Purification Kit (Takara, #9186) following the manufacturer's protocol. The purified poly(A) RNA was further concentrated using RNeasy MiniElute kit (Qiagen, #74204). The *Phaeodactylum* cDNA expression library was then constructed using CloneMiner^{TM} II cDNA Library Construction Kit (Invitrogen, #A11180) per the manufacturer's instruction. The first BP reaction (Fisher, #11789020) recombined an attB-flanked PCR product with a donor plasmid (pDONR221). This recombination displaced the ccdB toxin from the donor plasmid which yielded an entry clone with an attL-flanked PCR product, which was selected for on LB + 50 μg mL^{-1} kanamycin. The second LR reaction (Fisher, #11791020) recombined the entry clone with a destination vector (pDEST-527, gift of Dominic Esposito, Addgene plasmid #11518). The pDEST-527 vector had an IPTG-inducible promoter to drive expression of the Gateway-inserted sequence. This recombination displaced the ccdB toxin from the destination vector which yielded an expression clone containing the PCR product, again flanked by attB sequences, which was selected for on LB + 100 μg mL^{-1} ampicillin. In this study, the initial attB-flanked PCR product was generated from pooled

Phaeodactylum cDNA, so this process ultimately yielded a *Phaeodactylum* cDNA expression library.

Functional selection for UGPase activity in E. coli from a Phaeodactylum cDNA expression library

Rescuing *E. coli* strains with known metabolic deficiencies using a cDNA library from a species of interest has been demonstrated previously through the cloning of functional homologues of auxotrophic markers for leucine and tryptophan biosynthesis from plants[173]. DEV6 cannot metabolize galactose as a carbon source due to a mutation in its *galU* loci, which confers an inability to synthesize UDP-glucose[4]. This deficiency has previously been exploited in a targeted way demonstrating UDP-glucose diphosphorylase activity from recombinant eukaryotic gene products[60]. DEV6 electrocompetent cells were prepared by harvesting 0.8 L of exponentially growing cells (OD600 = 0.6; 3220 x g, 3220 x g, 15 m, 4 °C), chilled on ice for 20 min, washed 4 times with 10 % glycerol (v/v; centrifugation as above), and 50 µL aliquots were flash frozen in a dry ice/ethanol slurry for 5 min and stored at -80 °C until transformation. The *Phaeodactylum* cDNA library was transformed into DEV6 by electroporation (GenePulser Xcell, BioRad; 2.2 kV, 200 Ω, 25 µF, 1 mm cuvette). After 1 h recovery in SOC (shaking at 220 rpm, 37 °C), cells were washed with 1x M9 (salts only) and spread on selective plates (M9gal + 100 µg mL^{-1} ampicillin + 1 mM IPTG; 1 % agar). In parallel, transformed cells were spread on control plates without IPTG (M9gal + 100 µg mL^{-1} ampicillin; 1 % agar). Viable DEV6 colonies were observed on galactose plates after growth for 2 d at 25 °C when transformed with the cDNA library. Colonies were restreaked on selective plates (M9gal + 100 µg mL^{-1} ampicillin + 1 mM IPTG; 1 % agar) to verify growth. Plasmids were extracted from a 4 mL overnight culture using the GenElute Plasmid Miniprep Kit (Sigma, #PLN350). Sequencing of purified plasmids

was performed with M13F(-21) and M13R universal primers (Genewiz) and the transcript was identified by aligning sequencing results against the *Phaeodactylum* RefSeq with nBLAST (NCBI).

Recombinant expression of Phaeodactylum UGPases

The coding sequences of *ugp1, ugp2, and ugp3* were amplified from an independent cDNA preparation from wild-type (WT) *Phaeodactylum* grown in continuous light. This overlap extension PCR added attB recombination sites to facilitate Gateway cloning and primers were designed per the manufacturer's manual (Thermo Scientific, #MAN0000282). A list of primers used to amplify *galU, ugp1*, *ugp2*, and *ugp3* is included (Table 3.2, end of chapter). The *galU* coding sequence was amplified from DH5α genomic DNA using MAC115 and MAC116. Using a *Phaeodactylum* cDNA template, the *ugp1* coding sequence was amplified with MAC073 and MAC074. Similarly, the *ugp2* coding sequence was amplified with MAC077 and MAC078, while the *ugp3* coding sequence was amplified using MAC123 and MAC124. These same primers were used to validate the presence of these genes in genomic DNA, the *Phaeodactylum* cDNA Expression Library, and independent preparation of *Phaeodactylum* cDNA. 1 µL of the BP reaction was transformed into TOP10 competent cells and spread on LB + 50 µg mL^{-1} kanamycin; 1 % agar plates. Viable colonies carried pENTR plasmids with the gene of interest. Then, these pENTR plasmids were recombined into pDEST-527 with a LR reaction. 1 µL of the LR reaction was transformed into TOP10 competent cells and spread on LB + 100 µg mL^{-1} ampicillin; 1 % agar plates. These IPTG inducible UGPases were re-transformed into DEV6 to assess UGPase activity by quantifying growth rates in liquid M9gal or M9glu (+ 100 µg mL^{-1} ampicillin + 1 mM IPTG). Growth rates were determined from log-transformed OD600 values

measured from 3 h to 9 h after initial inoculation from a M9glu (+ 100 μg mL^{-1} ampicillin) starter culture.

Phylogenetic analysis of diatom UGPases

A previously constructed phylogenetic tree[6] was expanded to include *Phaeodactylum* UGP1, UGP2, and UGP3 protein sequences. Homologues from *Thalassiosira oceanica, Thalassiosira pseudonana*, and *Nannochloropsis gaditana* were identified by pBLAST queries of *Phaeodactylum* UGP1, UGP2, or UGP3 against these species respective non-redundant protein sequences databases (NCBI). Whole protein sequences were collected from UniProt (Table 3.3, end of chapter) and aligned with MUSCLE using default settings[174]. The subsequent alignment was used to construct an unrooted maximum-likelihood tree in MEGA7[175] (100 bootstrap replicates, WAG model, uniform rates among sites, use all sites, ML Hueristic Method SPR level 5, no branch swap filter). The result was visualized in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Constructing and genotyping UGPase knockouts.

Gene editing strategies such as CRISPR/Cas9 generate a double-strand break in a specific DNA sequence. Two endogenous DNA repair mechanisms can compete to repair this double-strand break: Non-homologous end joining (NHEJ) and homology directed repair (HDR). The former is error prone and tends to leave small insertions or deletions of nucleotides[101]. The latter benefits from a relatively straightforward screening process by genotyping PCR screens for an insert at the targeted DNA sequence[176]. Nymark et al. (2016) previously described CRISPR/Cas9 mediated gene editing for diatoms[101], and their methods are followed with several key changes detailed below. HDR has been previously demonstrated in *Phaeodactylum* [176], but relied on a TALEN-based approach rather than CRISPR/Cas9 to generate the double-

strand DNA break. We integrate CRISPR/Cas9 and HDR strategies this study. In addition, two unique guide RNAs were designed for each target to reduce the probability of drawing conclusions from unpredicted cutsites.

Assembling gRNA and homology donor vectors to knockout ugp1 and ugp2

Two gRNA vectors denoted "a" and "b" were designed and assembled for both *ugp1* (Phatrdraft_50444) and *ugp2* (Phatrdraft_23639). CRISPR spacers were identified by querying the respective coding sequences for the motif [5' – $GN^{20}NGG$ – 3']. The N^{20} sequence was queried against the *Phaeodactylum* genome (nBLAST, NCBI, taxonomic identifier: 556484), and discarded if multiple off-target loci were found. A 100 base pair guide-RNA containing the CRISPR spacer of interest (N^{20}) was assembled from forming a duplex with the forward [5' – $ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGN^{20}$ – 3'] and reverse [5' – $GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACN^{20}$ – 3'] primers. The duplex was formed in a 25 µL 1 x Phusion reaction following the manufacturer's instructions (NEB, #M0530) with 1 µL forward primer (100 µM) and 1 µL reverse primer (100 µM). A single cycle of denaturation (98 °C, 60 s), annealing (60 °C, 30 s), extension (72 °C, 60 s), and hold (4 °C) produced the gRNA duplex. A list of all primers used to construct gRNA vectors is included (Table 3.2). The gRNA-a duplex targeting *ugp1* was formed from DJ_UGP1_Fa and DJ_UGP1_Ra while the gRNA-b duplex was formed from DJ_UGP1_Fb and DJ_UGP1_Rb. The gRNA-a duplex targeting *ugp2* was formed from DJ_UGP2_Fa and DJ_UGP2_Ra while the gRNA-b duplex was formed from DJ_UGP2_Fb and DJ_UGP2_Rb. The gRNA backbone was amplified from a previously assembled JCVI urease gRNA plasmid (gift of A. Allen, JCVI) with the DJ058 and DJ059 primers. The gRNA vector was assembled from the gRNA duplex and the gRNA backbone using the SLIC method[177].

One homology donor was designed per each gene targeted for knocking out. A 150 base pair sequence that encompassed both gRNA target sites was identified, and 1 kb sequences upstream and downstream of this targeted region were defined as the upper and lower homology arms. The homology donor incorporates a Zeocin resistance cassette (a stop codon, FcpB promoter, and *sh ble*[178, 179]) in between the upper and lower homology arms. A list of all primers used to construct homology donors is included (Table 3.2). SLIC assembly of the *ugp1* homology donor vector entailed amplification of the upper homology arm with DJ_UGP1_HOF1 and DJ_UGP1_HOR1 and the lower homology arm with DJ_UGP1_HOF3 and DJ_UGP1_HOR3 from *Phaeodactylum* genomic DNA, while it entailed amplification of the Zeocin resistance cassette with DJ_UGP1_HOF2 and DJ_UGP1_HOR2 and the homology donor backbone with DJ062 and DJ063 from a previously assembled urease homology donor plasmid (gift of A. Allen, JCVI). SLIC assembly of the *ugp2* homology donor vector entailed amplification of the upper homology arm with DJ_UGP2_HOF1 and DJ_UGP2_HOR1 and the lower homology arm with DJ_UGP2_HOF3 and DJ_UGP2_HOR3 from *Phaeodactylum* genomic DNA, while it entailed amplification of the Zeocin resistance cassette with DJ_UGP2_HOF2 and DJ_UGP2_HOR2 and the homology donor backbone with DJ062 and DJ063 from the JCVI urease homology donor plasmid.

Assembling gRNA and homology donor vectors to knockout ugp3

Two gRNA vectors targeting *ugp3* (Phatrdraft_54493) were designed by identification of CRISPR spacers and assembled as described above. The gRNA-a duplex targeting *ugp3* was formed from MAC126 and MAC127 while the gRNA-b duplex was formed from MAC130 and MAC131. However, we designed a novel vector, pUC57-HD-UGP3 (Genewiz), to facilitate easier assembly of future SLIC assemblies of homology donors for *Phaeodactylum* knockout

projects (data not shown). The Zeocin resistance cassette was slightly modified from the JCVI vector by adding two additional stop codons to ensure a stop codon in each reading frame. The pUC57-HD-UGP3 vector was used as a homology donor to generate *ugp3* knockouts. pUC57-HD-UGP3 was digested with SacI (ThermoFisher, #FD1133) and BamHI (ThermoFisher, #FD0055) restriction endonucleases and the products were separated by agarose electrophoresis. The bands corresponding to the expected sizes of the pUC57 backbone and the Zeocin resistance cassette were excised, extracted from the agarose gel slices using the GenElute Gel Extraction kit (Sigma, #NA1111), and ligated with T4 ligase (NEB, #M0202) to yield the pUC57-HD vector. This vector enables either a 4-piece SLIC assembly or a restriction endonuclease and ligation approach to create homology donors for future CRISPR/Cas9 knockout projects in *Phaeodactylum*, and will be submitted to Addgene for distribution.

Biolistic transformation and selection

Exponentially growing *Phaeodactylum* cultures in continuous light were harvested at a cell density of approximately 2 x 10^6 cells mL^{-1}, and concentrated to a density of 5 x 10^8 cells mL^{-1}. 100 µL aliquots from this concentrate were plated in an approximately 25 mm diameter circle on f/2 + 1 % agar plates, plating approximately 5 x 10^7 cells. These plates were kept in a laminar flow hood until dry, then returned to the continuous light incubator until performing biolistics the following morning. The biolistics components were prepared as previously described[180]. Briefly, a previously prepared 30 µg aliquot of Tungsten M-17 microcarrier, 1.1 µm (BioRad, #1652267) was vortexed at room temperature for 5 min. Then, transformation components were added in the following order: 6 µg of plasmid (2 µg pPsaB_Cas9, 2 µg gRNA vector, 2 µg homology donor vector), 50 µL 2.5 M $CaCl_2$, 20 µL 0.1 M spermidine. This mixture was vortexed for 3 min. Particles were allowed to settle for 1 min prior to a 2 s spin in a

microcentrifuge. The supernatant was carefully discarded. 140 µL 70 % ethanol was slowly added to the side of the microcentrifuge tube. The 70 % ethanol was immediately removed, and the rinse was repeated once more with 100 % ethanol. Then, 48 µL 100 % ethanol was added. The pellet was carefully resuspended by tapping the side of the microcentrifuge tube and then vortexing for 3 s. 6 µL of this mixture was plated on a macrocarrier (BioRad, #1652335). Biolistics was performed with the Biolistic PDS-1000/He Particle Delivery System (BioRad), using 1550 PSI rupture disks (BioRad, #1652331). Bombarded plates were immediately protected from light, and allowed to recover in the dark for 12 - 16 h. The following morning, cells were transferred to selective plates (f/2 + 100 µg mL^{-1} Zeocin (ThermoFisher, #R25001); 1 % agar). Colonies were picked after 14 - 21 days of growth on selective media in continuous light, and maintained in 96-well plates containing f/2 + 100 µg mL^{-1} Zeocin media in continuous light.

Genotyping screens to identify knockout strains

Zeocin resistant colonies of recombinant *Phaeodactylum* strains were screened by genotyping through a two-step process. The first step was a 96-well plate colony PCR, using 1 µL of the culture volume as template in a 1x 25 µL REDTaq ReadyMix (Sigma, #R2523) reaction, using primers that flanked the 150 base pair sequence. The primer pairs used for screening of *ugp1* mutants were MAC132 and MAC133, for *ugp2* mutants were MAC134 and MAC135, and for *ugp3* mutants were MAC136 and MAC137. A list of all genotyping primers sequences that were used are included in Table 3.2. This typically yielded either a 1 kb product (expected for wild-type) or no product at all. Colonies that did not yield a 1 kb product were marked as potential knockouts and transferred into a 24 well plate containing f/2 + 100 µg mL^{-1} Zeocin media grown in continuous light. After 7 d – 10 d of growth, 0.5 mL of this culture was

pelleted and cells were lysed using the TE boil method[181]. Briefly, the pellet was resuspended in 50 µL of 10 mM Tris, 0.1 mM EDTA, pH 8.0, heated at 100 °C for 5 min, cooled on ice for 2 min, then spun (10000 x g, 1 min, 4 °C). 1 µL of the supernatant was used as template for a 1x Phusion reaction (NEB, #M0530) using the same primers from the first step. These parameters yielded either a 1 kb product or a 2.4 kb product. In parallel, a second 1x Phusion reaction was set up where the forward primer was outside the 150 base pair region, and the reverse primer was inside the Zeocin cassette. The primer pairs used for this secondary PCR screening of *ugp1* mutants were MAC144 and MAC138, for *ugp2* mutants were MAC146 and MAC140, and for *ugp3* mutants were MAC136 and MAC142. These parameters yielded either no product or a product, with the latter suggesting a targeted insertion at the locus of interest. Colonies exhibiting both a 2.4 kb product for the flanking PCR and a product for the flanking/inside Zeocin cassette PCR were identified as complete knockouts. This process was repeated until 2 – 3 strains were identified for each gRNA vector.

Physiology of *Phaeodactylum* UGPase knockouts

Growth rates

Growth rates were determined from events μL^{-1} on an Accuri C6 flow cytometer (BD), as previously described[124]. Timepoints were taken at the end of the light period each day for 3 to 5 days and the growth rate was calculated from the linear slope after natural log transformation of the events μL^{-1} values.

Variable chlorophyll fluorescence measurements.

The maximum quantum efficiency of Photosystem II (F_v/F_m) [158] was determined after harvesting 2 mL of cells one h prior to the light to dark transition. These volumes were placed in labeled 15 mL conical tubes and allowed to acclimate in 9 µmol photons m^{-2} s^{-1} ± 1 µmol

photons m^{-2} s^{-1} for 15 min to relax non-photochemical quenching[156]. Samples were filtered onto glass fiber disks and analyzed on a WALZ PAM Fluorometer as previously described[124].

BODIPY staining and observation

During exponential growth, a 0.5 mL sample was harvested from the cultures at the transition from light to dark. Samples were filtered and BODIPY was added to a final concentration of 0.067 μg mL^{-1}, as previously described[153].

Carbohydrate quantification

During exponential growth, 50 mL of cell culture was harvested at an approximate cell density of 2 x 10^6 cells mL^{-1}. Carbohydrates were quantified using a modified 3-methyl-2-benzothiazolinone hydrazone reducing sugar assay as described previously[124]. Chrysolaminarin was determined by subtracting soluble, reducing carbohydrates from soluble, total carbohydrates.

Determining UGPase activity in Phaeodactylum soluble protein extracts

50 mL of culture was harvested at an approximate cell density of 2 x 10^6 cells mL^{-1}, centrifuged (3200 x g, 10 m, 18 °C), and the supernatant was discarded. The pellet was resuspended with 1 mL of artificial seawater, transferred to a microcentrifuge tube, and re-pelleted (10000 x g, 10 m, 18 °C). The supernatant was carefully discarded and the microcentrifuge tube was frozen in liquid nitrogen. A metal bead (McMaster Carr, 1/8") was cooled in liquid nitrogen and added to the frozen pellet. Microcentrifuge tubes containing both the pellet and metal bead were kept at -80 °C until cell disruption (Tissue Lyser II, Quiagen; 1 min, 30 Hz). The algal powder was resuspended with 1 mL ice-cold extraction buffer (10 mM Tris, pH 8.0, 1 mM EDTA) and vortexed for 30 s at room temperature. The suspension was centrifuged (10000 x g, 10 m, 18 °C) to remove cell debris, and the supernatant was transferred

to a fresh, labeled microcentrifuge tube. This supernatant was centrifuged once more (10000 x g, 10 m, 18 °C). Total soluble protein was quantified using the Pierce BCA Protein Quantification Kit (Thermo, #23225) according to the manufacturer's protocol. UGPase activity was assayed as before[182], with slight modifications. 30 µg of soluble protein was used to observe the continuous release of pyrophosphate with the EnzChek Pyrophosphate Assay Kit (Thermo, E6645) as described in the manufacturer's manual. In this reaction, 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) is catalyzed by purine nucleoside phosphorylase in the presence of inorganic phosphate into the products ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine, shifting the maximum absorbance peak from 330 nm to 360 nm. A 1 mL reaction volume contained the following: 1x reaction buffer (50 mM TrisHCl, 1mM $MgCl_2$, pH 7.5), 1 mM DTT, 0.2 mM MESG, 1 U purine nucleoside phosphorylase (1 unit = phosphorolysis of 1 µmol of inosine to hypoxanthine and ribose 1-phosphate per minute, pH 7.4 at 25 °C), 0.1 U inorganic pyrophosphorylase (1 unit = liberation of 1 µmol of inorganic orthophosphate per minute, pH 7.2 at 25 °C), 30 µg soluble protein extract and either the substrate (1 mM UTP, 1 mM glucose 1-phosphate) or substrate control (1 mM glucose 1-phosphate). The assembled reaction, minus the substrate or substrate control, was incubated for 10 min at 18 °C to remove contaminating phosphate prior to addition of the substrate or substrate control, as described in the EnzChek manual. The reaction was incubated at 18 °C, and timepoints were taken every 5 min for 35 min after the addition of the substrate or substrate control. The stoichiometry of the UGPase reaction yields products diphosphate (PPi) and UDP-glucose at a 1 to 1 ratio (EC: 2.7.7.9 and EC: 2.7.7.64). UGPase activity was directly inferred from the rate of PPi synthesis. A standard curve of 1 µM – 50 µM inorganic pyrophosphate was prepared to convert substrate control-corrected absorbance to PPi concentrations.

Total Organic Carbon analysis

Total Organic Carbon (TOC) analyses were performed on a Shimadzu TOC-L Laboratory TOC Analyzer as previously described[124].

Rapid light curves with oxygen evolution

At midday (6 h into the 12 h light period), 20 mL of exponentially growing cells were harvested. Cells were concentrated by centrifugation (3000 x g, 18 °C, 10 minutes in a 50 mL conical tube) to a volume of 2 mL using 18 °C f/2 supplemented with 20 mM HEPES buffer (pH 7.5). From this 2 mL concentrate, a 0.1 mL aliquot was diluted with filtered sea water and used to determine cell density by flow cytometry (see above). A 0.2 mL aliquot was transferred from the concentrate to a 1.5 mL micro-centrifuge tube and pelleted (10000 x g, 18 °C, 10 min). Supernatant was discarded and the pellet was resuspended in 1 mL methanol in order to extract pigments and determine chlorophyll concentrations as previously described[152]. A 1 mL aliquot from the concentrate was used to perform photosynthesis vs irradiance curves in a cuvette holder (Walz, #ED-101US/MD). A custom cuvette adapter enabled simultaneous measurement of O_2 evolution (FireSting, #OXROB10) and fluorescence parameters on a Walz DUAL-PAM 100 fluorometer, as described previously[154], including a 15 min far red light adaptation step prior to executing the light curve script. Gross oxygen evolution rates were determined from the sum of net oxygen evolution rates and respiration rates, and enabled calculation of maximum photosynthesis rate (P_{max}), light limited slope (α), and saturation irradiance (E_k) as previously described[183].

Targeted lipidomics pilot experiment at CSU Proteomics and Metabolomics Facility

Exponentially growing WT and *ugp2* knockout strain A cells (approximately 9 x 10^7) were harvested in biological triplicate at the end of the light period on methanol-rinsed 21 mm

glass filters (GE Healthcare, #1820-021) via vacuum filtration, flash frozen in liquid nitrogen, and lyophilized. A blank filter with no algal biomass was also prepared as a negative control. Lipids were extracted following the methyl *tert*-butyl ether MBTE method as previously described[184]. Purified SQDG was acquired and used as a positive control (Avanti, #840525). Samples were analyzed on a Waters UPLC system, and lipids were separated on a C18 column. The column was connected to a Waters TQ-S tandem quad mass spectrometer, where data was collected in both positive and negative ion modes. Identification of an abundant diatom SQDG species (32:1, [125]) was performed manually using a reference structure database (LIPID MAPS, #LMGL05010008). *Phaeodactylum* SQDG (32:1) was identified as *m/z* 791.5 at the retention time of 12.7 min. Peak intensity was integrated and normalized by total peak intensity in the positive mode for relative quantification of SQDG (32:1) in these lipid extracts.

Statistical analysis

A paired student's t-test ($\alpha = 0.05$) was used for the comparisons of total organic carbon and total nitrogen between dawn and dusk and the targeted analysis of SQDG (32:1). Otherwise, a one-way ANOVA ($\alpha = 0.05$) followed by a Tukey's HSD post-hoc test was performed to define statistically different groups. All statistical analysis was performed using SAS University Edition (SAS).

Results

Functional identification of *ugp3*

The DEV6 strain of *E. coli* was used to select for functional UGPase activity from a *Phaeodactylum* cDNA expression library. DEV6 *E. coli* have a mutation in their *galU* gene, conferring an inability to metabolize galactose as a carbon source[4]. DEV6 strains have been successfully complemented with other eukaryotic cDNAs that encode a gene product with

UGPase activity[60, 185], restoring their ability to grow on galactose. Complementing DEV6 with an inducible *Phaeodactylum* cDNA expression library yielded 8 viable colonies after 2 days of growth at 25 °C. Plasmids were extracted from all eight CFUs, and all 8 cDNA sequences encoded a sequence that aligned to *ugp3* (Phatrdraft_54493, Table 3.4, end of chapter). No colonies were observed when transformed DEV6 was plated on control plates that omitted IPTG. The coding sequences for *ugp1*, *ugp2*, and *ugp3* were all verified to be present in the *Phaeodactylum* cDNA Expression Library by PCR (Figure 3.1).

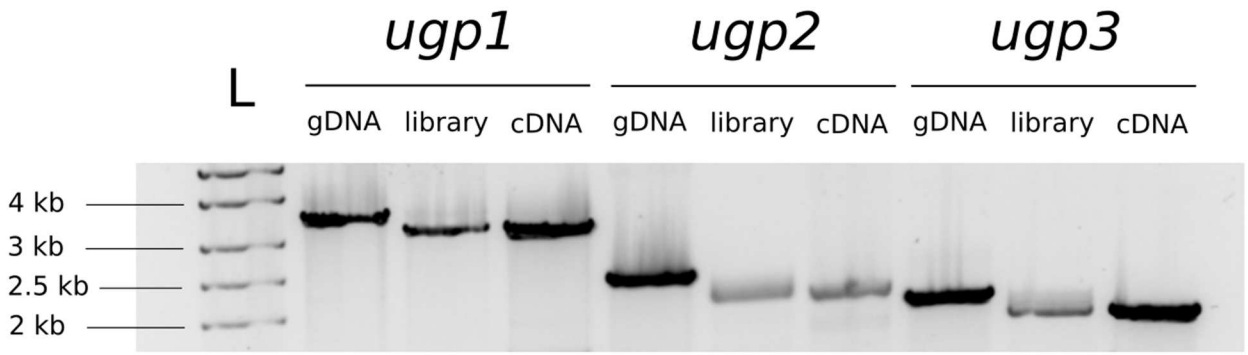

Figure 3.1: Amplification of *ugp1*, *ugp2*, and *ugp3* coding sequences from genomic DNA, cDNA, and the *Phaeodactylum* cDNA Expression Library.
L = 1 kb ladder. Primers used for the PCR were the attB-extension primers for each gene as indicated above the line and template used for the PCR is designated below the line. The first lane is a 1 kb ladder (L).

DEV6 was then directly complemented with either a positive control (functional *galU* amplified from DH5α), a negative control (β-galactosidase, "*gus*," a selectable marker that encodes for a protein with no UGPase activity[186]), *ugp1*, *ugp2*, or *ugp3* (Figure 3.2). Growth rates at 25 °C were determined by monitoring OD600 hourly. All strains grew when glucose was provided as a carbon source, ranging from 0.066 ± 0.002 hr^{-1} to 0.104 ± 0.008 hr^{-1} (Figure 3.2). However, when galactose was provided as a carbon source, only *galU* and *ugp3* complemented DEV6 exhibited growth rates at 0.051 ± 0.002 hr^{-1} and 0.047 ± 0.002 hr^{-1}, respectively. *Gus*,

ugp1, and *ugp2* complemented DEV6 did not demonstrate growth on galactose (-0.001 ± 0.001 hr^{-1}, Figure 3.2).

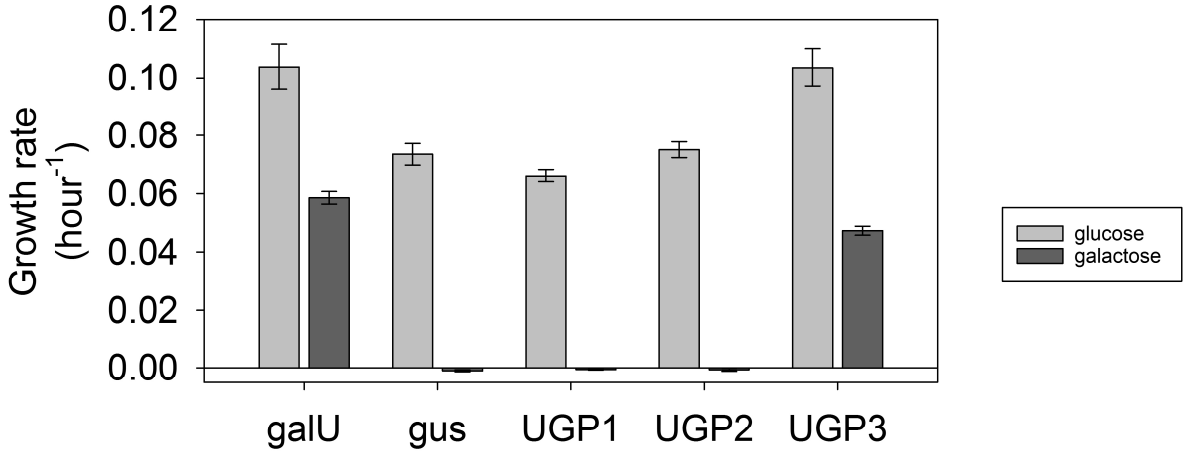

Figure 3.2: Exponential growth rates of DEV6 complemented with an inducible expression vector.

The *E.coli* strain DEV6 has a non-functiona UGPase and cannot synthesize UDP-glucose, disrupting its ability to metabolize galactose. DEV6 is rescued when complemented with a functional *E. coli* UGPase, *galU*, but is not rescued when complemented with *gus*. Complementing DEV6 with the coding sequences of three putative UGPases from *Phaeodactylum* resolved that only complementation with the *ugp3* coding sequence rescues DEV6 on galactose as a carbon source with UGP3. OD600 was measured as a proxy for cell density, and exponential rates were determined from timepoints between 3 h and 9 h after transfer from M9glu into either M9glu or M9gal (n = 3, ± standard deviation).

Predicting roles for diatom UDP-hexose diphosphorylases based on homology searches

As described previously [6], phylogenetic analysis of previously described UDP-hexose diphosphorylases resolved four major clusters: UDP-glucose diphosphorylase, Type A; UDP-glucose diphosphorylase, Type B; UDP-sugar diphosphorylase; and UDP-N-acetylglucosamine diphosphorylase (Figure 3.3). By expanding the phylogenetic analysis to include diatom species, we observed that *ugp1* groups with the UDP-glucose diphosphorylase, Type A cluster, while both *ugp2* and *ugp3* are both found in the UDP-sugar diphosphorylase cluster. Both UDP-glucose diphosphorylase and UDP-sugar disphosphorylase synthesize UDP-glucose.

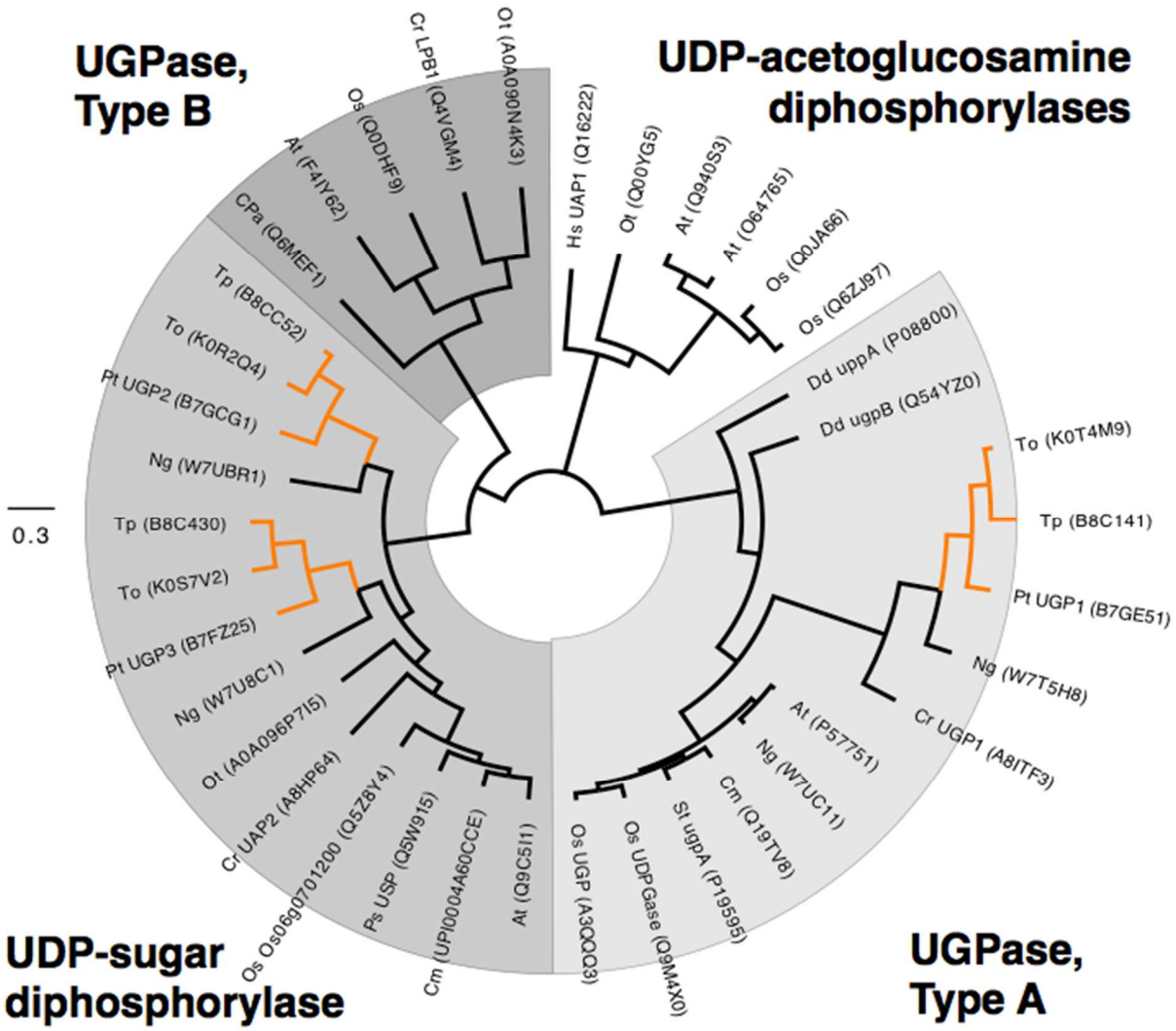

Figure 3.3: Bioinformatically predicted and functionally identified diatom UDP-glucose diphosphorylases overlaid onto a phylogenetic tree of UDP-hexose diphosphorylases. A previous study resolved four major UGPase clusters[6], and this tree expands on that experiment by including several stramenopile sequences highlighted in orange (Pt = *Phaeodactylum tricornutum*, Tp = *Thalassiosira pseudonana*, To = *Thalassiosira oceanica*, Ng = *Nannochloropsis gaditana*) and two Amoebozoa sequences with functional evidence (Dd = *Dictyostelium discoideum*). Protein sequences were taken from UniProt (accession number) and aligned with the MUSCLE program. Bar indicates substitutions per site. A Maximum Likelihood Tree was generated in MEGA7 with 100 replicates.

Generating and genotyping CRISPR/Cas9 knockouts of *ugp1*, *ugp2*, and *ugp3*

A genotyping screen using primers that flank the predicted cutsite enabled us to identify multiple HDR knockouts for each guide RNA. For each gRNA used, 2 – 3 HDR knockouts were selected for phenotyping. Several Zeocin-resistant lines that did not exhibit HDR or any

insertion/deletion at the anticipated cutsite (data not shown) were maintained as transformation controls to account for stochastic effects from random nuclear integration.

In the process of developing these mutants, pUC57-HD was designed and assembled (Figure 3.4). The pUC57-HD plasmid was derived from the synthesized vector pUC57-HD-UGP3 (Genewiz). The pUC57-HD vector enables a 4-piece SLIC assembly that only requires PCR amplification of the homology arms, eliminating amplifications of the Zeocin resistance cassette and the vector backbone. pUC57-HD will be submitted to Addgene.

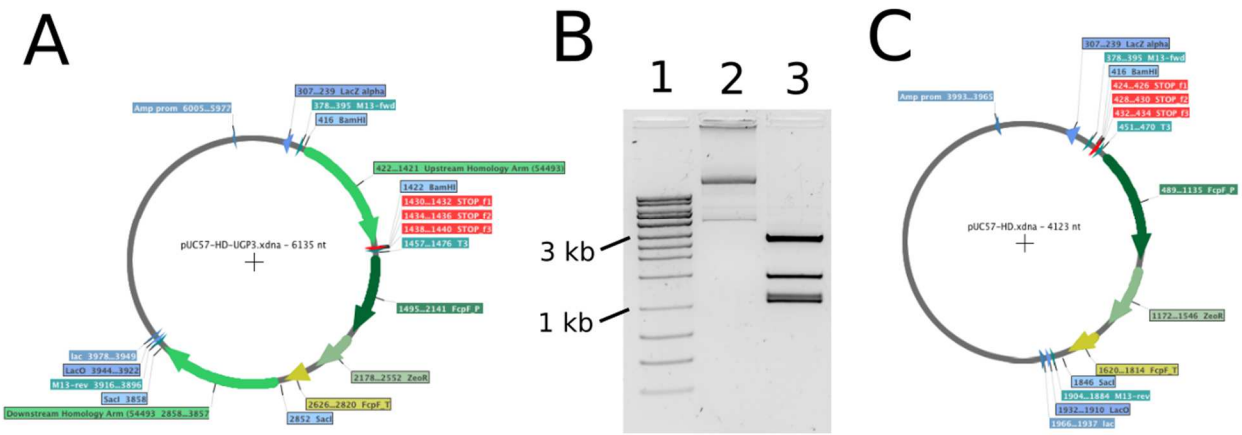

Figure 3.4: Assembly of pUC57-HD, a new vector benefiting homology donor construction for CRISPR HDR strategies in *Phaeodactylum*.
The pUC57-HD-UGP3 was synthesized (A, Genewiz). This vector was double digested with BamHI and SacI, then run on a 1 % agarose gel (B). Lane 1 = 1 kb ladder, lane 2 = undigested pUC57-HD-UGP3, lane 3 = BamHI/SacI digested pUC57-HD-UGP3. The pUC57 backbone (2.7 kb) and the Zeocin resistance cassete (1.4 kb) were separated from the 2 homology arms (1 kb doublet). The pUC57 backbone and the Zeocin resistance cassette bands were excised, gel purified, and ligated to yield pUC57-HD (C).

PCR amplifications of WT, the *ugp1* transformation control (TC_ugp1), and mutants ugp1-a and ugp1-b were performed (Figure 3.5, B). PCR using flanking primers MAC132 and MAC133 yielded an amplicon of about 1 kb for WT and TC_ugp1 but yielded an amplicon of about 2.4 kb for mutants ugp1-a and ugp1-b. PCR amplification using a primer outside the Zeocin cassette (MAC144) and in the Zeocin cassette (MAC138) did not yield a product for WT or TC_ugp1, but did yield a 1.8 kb amplicon for mutants ugp1-a and ugp1-b.

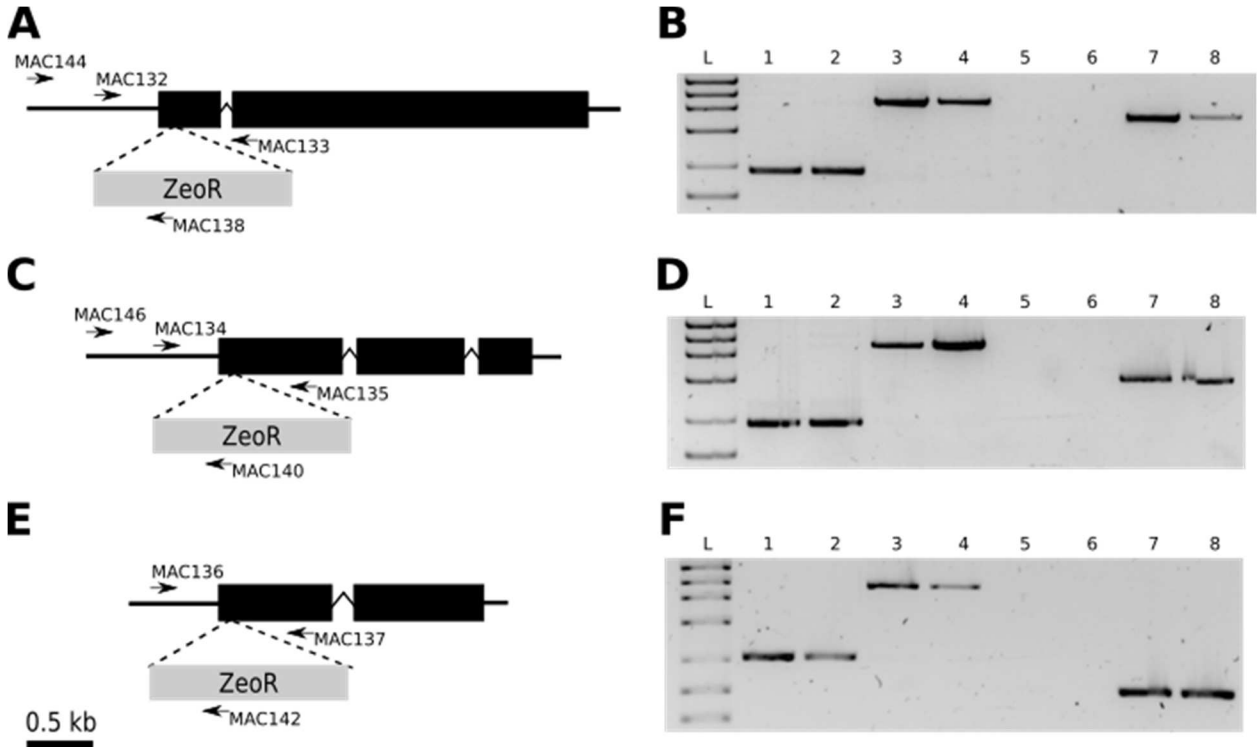

Figure 3.5: Genotyping *ugp1*, *ugp2*, and *ugp3* knockouts.
Gene models of *ugp1*(A), *ugp2* (C), and *ugp3* (E) are displayed with primers used for genotyping. Bar represents 500 base pairs. Genotyping PCRs investigating gene insertion are also presented for *ugp1* (B), *ugp2* (D), and *ugp3* (F) targets. Templates were the following: WT (lanes 1 and 5), respective transformation control (lanes 2 and 6), knockout "a" (lanes 3 and 7), and knockout "b" (lanes 4 and 8). Lanes 1-4 were amplicons produced with primers immediately flanking the cut site, while lanes 5-8 were generated with an outside primer and another primer inside the Zeocin resistance cassette (ZeoR), described in additional detail in the methods.

PCR amplifications of WT, the *ugp2* transformation control (TC_ugp2), and mutants ugp2-a and ugp2-b were performed (Figure 3.5, D). PCR using flanking primers MAC134 and MAC135 yielded an amplicon of about 1 kb for WT and TC_ugp1 but yielded an amplicon of about 2.4 kb for mutants ugp2-a and ugp2-b. PCR amplification using a primer outside the Zeocin cassette (MAC146) and in the Zeocin cassette (MAC140) did not yield a product for WT or TC_ugp2, but did yield a 1.5 kb amplicon for mutants ugp1-a and ugp1-b.

PCR amplifications of WT, the *ugp3* transformation control (TC_ugp3), and mutants ugp3-a and ugp3-b were performed (Figure 3.5, F). PCR using flanking primers MAC136 and MAC137 yielded an amplicon of about 1 kb for WT and TC_ugp3 but yielded an amplicon of

about 2.4 kb for mutants ugp3-a and ugp3-b. PCR amplification using a primer outside the Zeocin cassette (MAC136) and in the Zeocin cassette (MAC142) did not yield a product for WT or TC_ugp1, but did yield a 0.75 kb amplicon for mutants ugp3-a and ugp3-b.

Phenotyping CRISPR/Cas9 knockouts of *ugp1*, *ugp2*, and *ugp3*

Genotype-verified knockouts were characterized for fitness (growth rate and the maximum quantum yield of Photosystem II, Fv/Fm) (Figure 3.6, A and B), BODIPY fluorescence, a proxy for neutral lipids (Figure 3.6, C), and chrysolaminarin content (Figure 3.6, D) for the representative strains indicated above. However, all physiology data is also presented for each of the 16 identified knockouts (Table 3.5, end of chapter).

Growth rates of ugp1-a and ugp1-b were ($0.82 \pm 0.12\ d^{-1}$) and ($0.84 \pm 0.13\ d^{-1}$), respectively. While they appear to be lower than their transformation control TC_ugp1 ($1.06 \pm 0.04\ d^{-1}$), this reduction was not statistically significant (Figure 3.6, A, $p > 0.05$). On the other hand, ugp2-a and ugp2-b mutants had growth rates of ($0.51 \pm 0.03\ d^{-1}$) and ($0.65 \pm 0.14\ d^{-1}$), respectively, and were significantly different from each other as well as their transformation control TC_ugp2 ($1.06 \pm 0.08\ d^{-1}$) ($p < 0.05$). The ugp3-a mutant grew significantly slower ($0.85 \pm 0.02\ d^{-1}$, $p < 0.05$) than its transformation control TC_ugp3 ($0.91 \pm 0.03\ d^{-1}$), but ugp3-b did not ($0.93 \pm 0.01\ d^{-1}$, $p > 0.05$).

The maximum quantum efficiency of photosystem II[158], F_v/F_m, was statistically similar for ugp1-a, ugp1-b, ugp3-a and ugp3-b knockouts relative to their respective transformation controls (Figure 3.6, B, $p > 0.05$). The F_v/F_m of ugp2-a was 0.448 ± 0.020 and ugp2-b was 0.526 ± 0.022. These values were significantly different from each other, and both were lower than TC_ugp2 (0.601 ± 0.008, $p < 0.05$). While ugp3-a exhibited a statistically

significant increase in its F_v/F_m relative to TC_ugp3 ($p < 0.05$), the ugp3-b mutant strain B did not ($p > 0.05$).

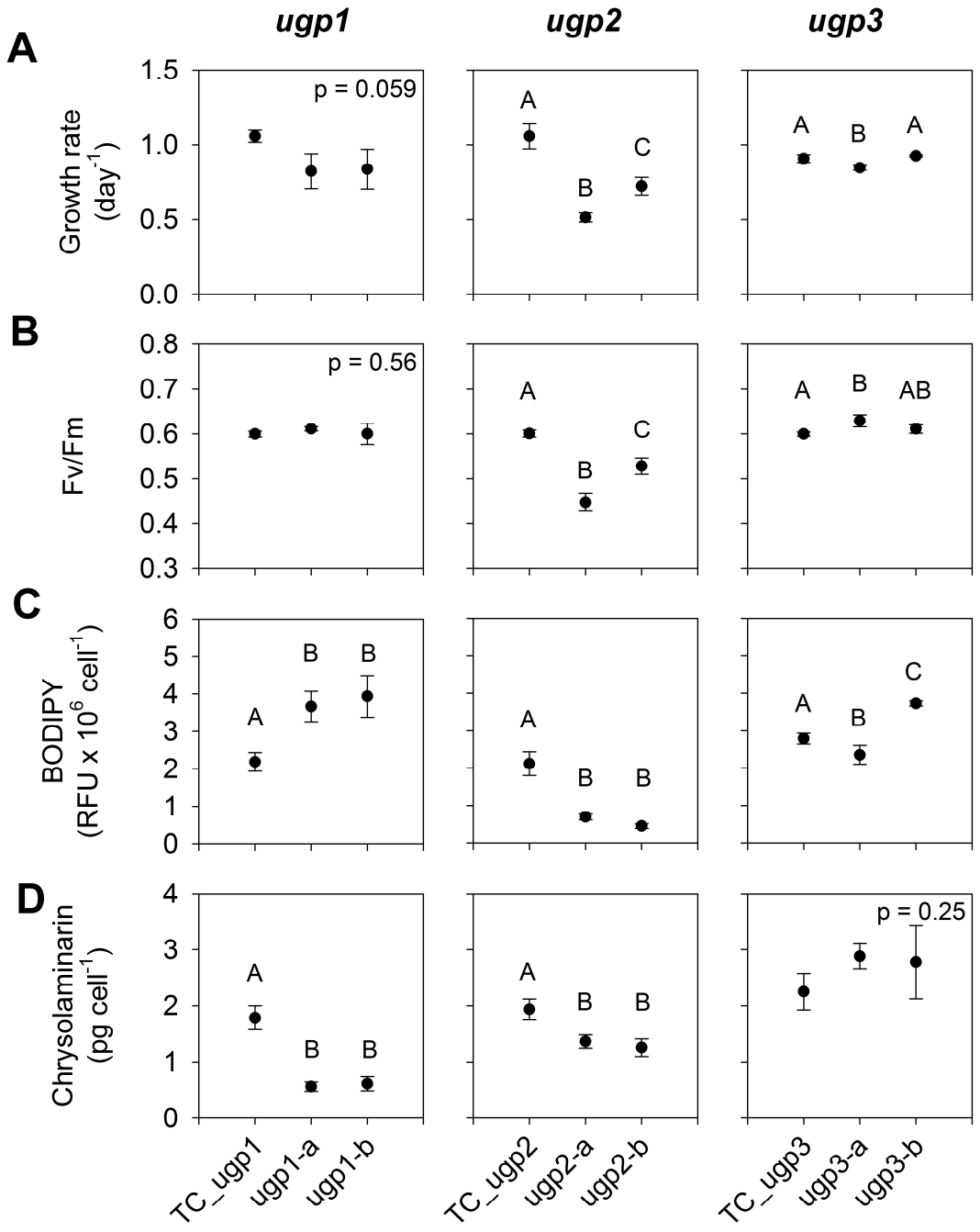

Figure 3.6: Comparative physiology of *ugp1, ugp2,* and *ugp3* gene knockouts. Exponentially growing cultures were assayed for growth rate (A), maximum quantum efficiency of Photosystem II (F_v/F_m, B), BODIPY RFU per cell, a proxy for neutral lipid accumulation (C), and chrysolaminarin content per cell (D). (n = 3, ± standard deviation). Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a one way ANOVA ($p < 0.05$).

BODIPY fluorescence per cell, a proxy for neutral lipid content per cell, was quantified from exponentially growing cultures at the end of the light period when BODIPY fluorescence per cell is maximized[154]. Significantly higher BODIPY fluorescence per cell was observed for both ugp1-a ($3.67 \pm 0.41 \times 10^6$ RFU) and ugp1-b ($3.94 \pm 0.56 \times 10^6$ RFU) relative to TC_ugp1 ($2.19 \pm 0.25 \times 10^6$ RFU, Figure 3.6, C, $p < 0.05$). Both ugp2-a and ugp2-b had significantly lower BODIPY fluorescence per cell ($6.9 \pm 0.4 \times 10^5$ RFU, $1.05 \pm 0.37 \times 10^6$ RFU, respectively) compared to TC_ugp2 ($2.12 \pm 0.32 \times 10^6$ RFU) ($p < 0.05$). While the ugp3-a mutant had significantly lower BODIPY fluorescence per cell compared to TC_ugp3 ($p < 0.05$), the ugp3-b mutant had significantly higher BODIPY fluorescence per cell relative to TC_ugp3 ($p < 0.05$).

Chrysolaminarin content per cell was determined by harvesting cell extracts from exponentially growing cells at the end of the light period, when chrysolaminarin accumulation is at its maximum[124]. Chrysolaminarin content was significantly lower in both ugp1-a (0.56 ± 0.08 pg glucose equivalents (g.e.) cell^{-1}) and ugp1-b (0.61 ± 0.13 pg g.e. cell^{-1}) relative to TC_ugp1 (1.79 ± 0.21 pg g.e. cell^{-1}, Figure 3.6, D, $p < 0.05$). There was a modest but significant reduction of chrysolaminarin content per cell in both ugp2-a (1.36 ± 0.12 pg g.e. cell^{-1}) and ugp2-b (0.85 ± 0.09 pg g.e. cell^{-1}) relative to TC_ugp2 (1.94 ± 0.18 pg g.e. cell^{-1}, $p < 0.05$). There was no statistical change in chrysolaminarin content for the *ugp3* knockout strains compared to their transformation control ($p > 0.05$).

The ugp1-a, ugp2-a, and ugp3-a strains were selected for further characterization of *in vitro* UGPase activity (Figure 3.7) and total organic carbon content per cell (Figure 3.8).

UGPase activity was determined for wild-type (WT) and the ugp1-a, ugp2-a, and ugp3-a strains. The *in vitro* rate of UGPase synthesis in WT *Phaeodactylum* was 40.2 ± 8.1 nmol UDP-

glucose min^{-1} mg soluble $protein^{-1}$ (Figure 3.7). The ugp1-a mutant demonstrated substantially lower UGPase activity (2.8 ± 0.8 nmol UDP-glucose min^{-1} mg soluble $protein^{-1}$) than WT ($p < 0.05$). The ugp2-a mutant had significantly more UGPase activity (98.3 ± 10.2 nmol UDP-glucose min^{-1} mg soluble $protein^{-1}$) than WT ($p < 0.05$). The UGPase activity of ugp3-a was statistically identical to WT ($p > 0.05$).

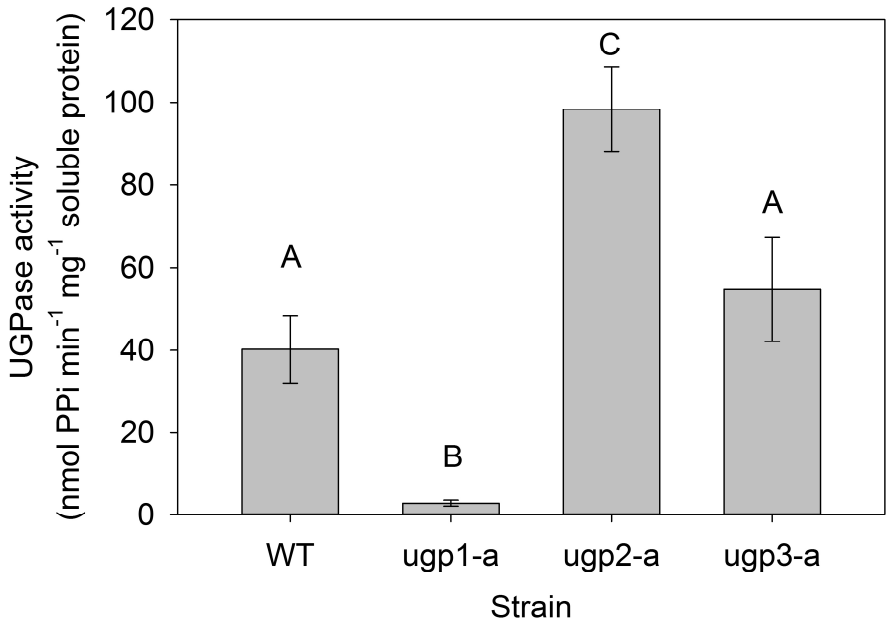

Figure 3.7: UDP-glucose diphopsphorylase activity in wild-type *Phaeodactylum* and ugp1-a, ugp2-a and ugp3-a mutants.
The kinetics of inorganic pyrophosphate synthesis were determined using the EnzCheck Pyrophosphate assay kit. Since synthesis of UDP-glucose yields 1 molecule of UDP-glucose and 1 molecule of pyrophosphate, UDP-glucose diphosphorylase activity was directly inferred from this rate. (WT = wild-type, n = 3, ± standard deviation). Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a one way ANOVA ($p < 0.05$).

WT *Phaeodactylum* cells contain significantly more total organic carbon at the end of the light period (22.6 ± 4.4 pg total organic carbon $cell^{-1}$) versus the end of the dark period (14.5 ± 3.0 pg total organic carbon $cell^{-1}$, Figure 3.8, $p < 0.05$). This pattern is not observed for either ugp1-a or ugp2-a mutants, as total organic carbon per cell was statistically identical between dusk and dawn ($p > 0.05$). The ugp3-a mutant, like WT, contained significantly more total organic carbon at dusk compared to dawn ($p < 0.05$).

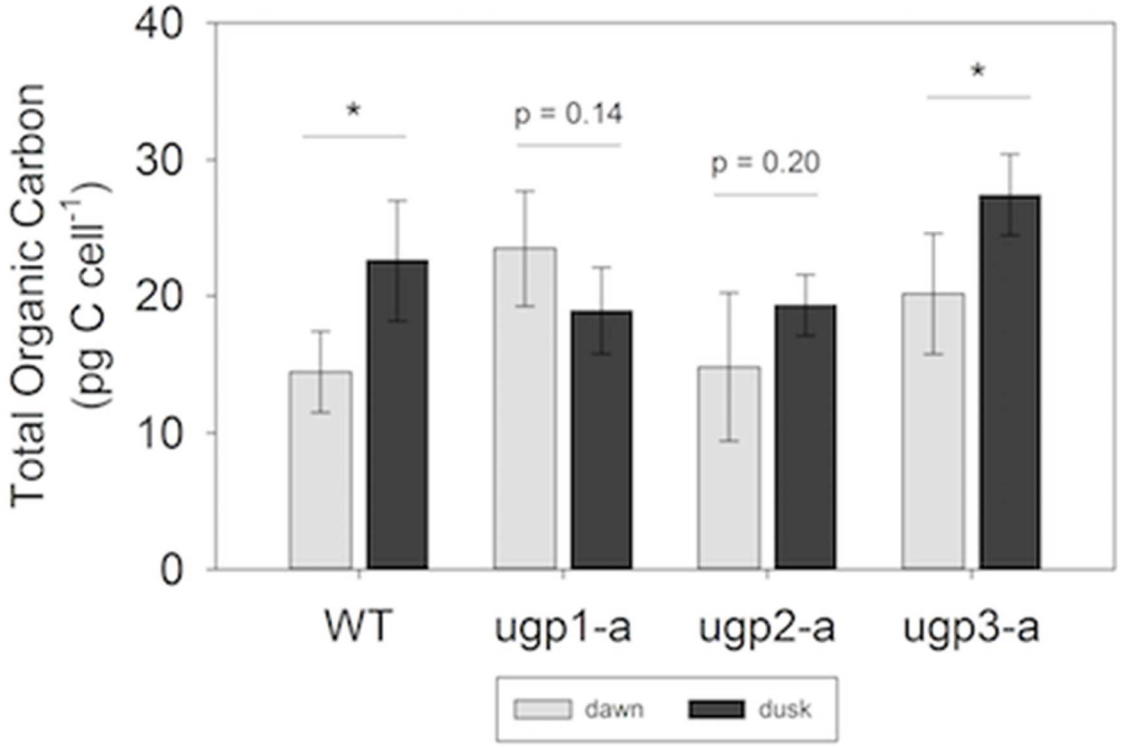

Figure 3.8: Determination of total organic carbon at dawn and dusk in wild-type *Phaeodactylum* and ugp1-a, ugp2-a and ugp3-a mutants.
(WT = wild-type, n = 3, ± standard deviation). Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a one way ANOVA ($p < 0.05$).

Low F_v/F_m in both ugp2-a and ugp2-b suggests an impairment of photosynthetic capacity which was further investigated with photosynthesis vs. irradiance curves (Figure 3.9). Fitting these curves enabled determination of the maximum capacity for photosynthesis (P_{max}), the light limited slope (α), and saturation irradiance (E_k) (Table 3.1). The ugp2-a mutant exhibited significantly lower P_{max} (405 ± 43 nmol O_2 µg chl a^{-1} h^{-1}) and α (1.2 ± 0.2 nmol O_2 µg chl a^{-1} h^{-1} µmol photons^{-1} m^2 s) versus TC_ugp2 (636 ± 110 nmol O_2 µg chl a^{-1} h^{-1}, 1.7 ± 0.2 nmol O_2 µg chl a^{-1} h^{-1} µmol photons^{-1} m^2 s, $p < 0.05$). The ugp2-b had intermediate values for P_{max} (489 ± 89 nmol O^2 µg chl a^{-1} h^{-1}) and α (1.4 ± 0.2 nmol O_2 µg chl a^{-1} h^{-1} µmol photons^{-1} m^2 s) that were not significantly different than either ugp2-a or TC_ugp2 ($p > 0.05$). E_k and dark respiration rates were statistically similar for all strains ($p > 0.05$).

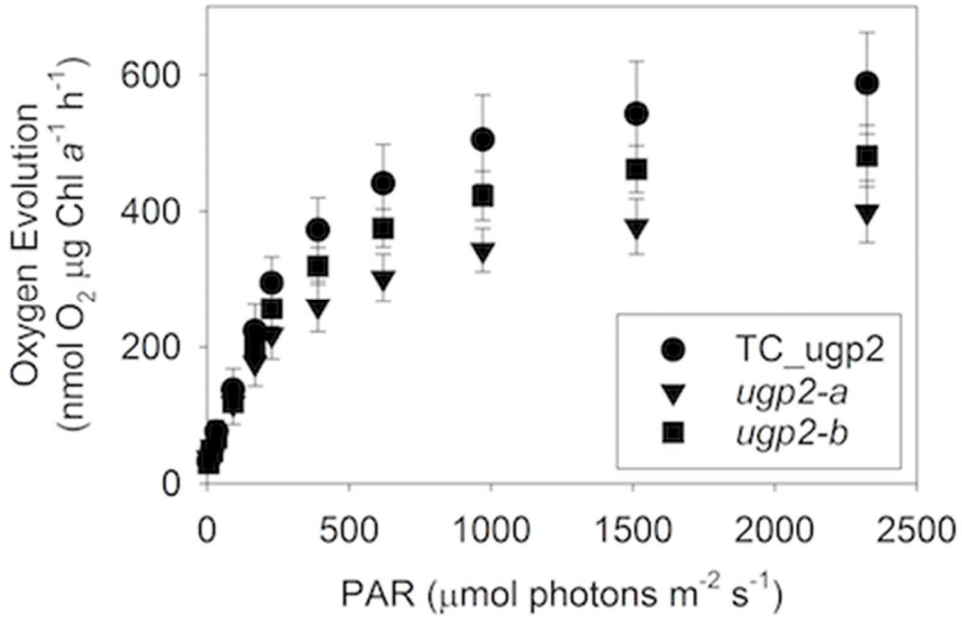

Figure 3.9: Gross photosynthesis versus irradiance curves for *ugp2* gene knockouts ugp2-a and ugp2-b and their transformation control TC_ugp2.
Measurements were recorded on cells harvested at the end of the light period. Oxygen evolution or consumption rates were measured using a custom cuvette setup that facilitated exposure to increasing actinic light intensities on a Pulse Amplitude Modulated fluorometer (n = 4, ± standard deviation).

Table 3.1: Photosynthetic performance parameters for *ugp2* gene knockouts and their transformation control (TC_ugp2).
Maximum photosynthesis rate (P_{max}; nmol O^2 µg chl a^{-1} h^{-1}), light limited slope (α, nmol O_2 µg chl a^{-1} h^{-1} µmol photons^{-1} m^2 s), saturation irradiance (E_k, µmol photons m^{-2} s^{-1}), and dark respiration rate (nmol O_2 $cell^{-1}$ h^{-1}) were determined. n = 4, ± standard deviation. Different letters represent statistically different groups from a Tukey's HSD post-hoc test after a one way ANOVA ($p < 0.05$).

Strain	P_{max}	α	E_k	Dark respiration rate
TC_ugp2	636 ± 110^{a}	1.7 ± 0.2^{a}	370 ± 63^{a}	15.5 ± 3.7^{a}
ugp2-a	405 ± 43^{b}	1.2 ± 0.2^{b}	333 ± 61^{a}	17.7 ± 4.6^{a}
ugp2-b	489 ± 89^{ab}	1.4 ± 0.2^{ab}	330 ± 10^{a}	15.4 ± 3.9^{a}

The 32:1 SQDG species is the most abundant chain length and desaturation combination for this sulfolipid in *Phaeodactylum*[125]. A targeted metabolomics experiment determined that

lipid extracts from the ugp2-a contains slightly less 32:1 SQDG than WT (Figure 3.10, n = 3, p < 0.05).

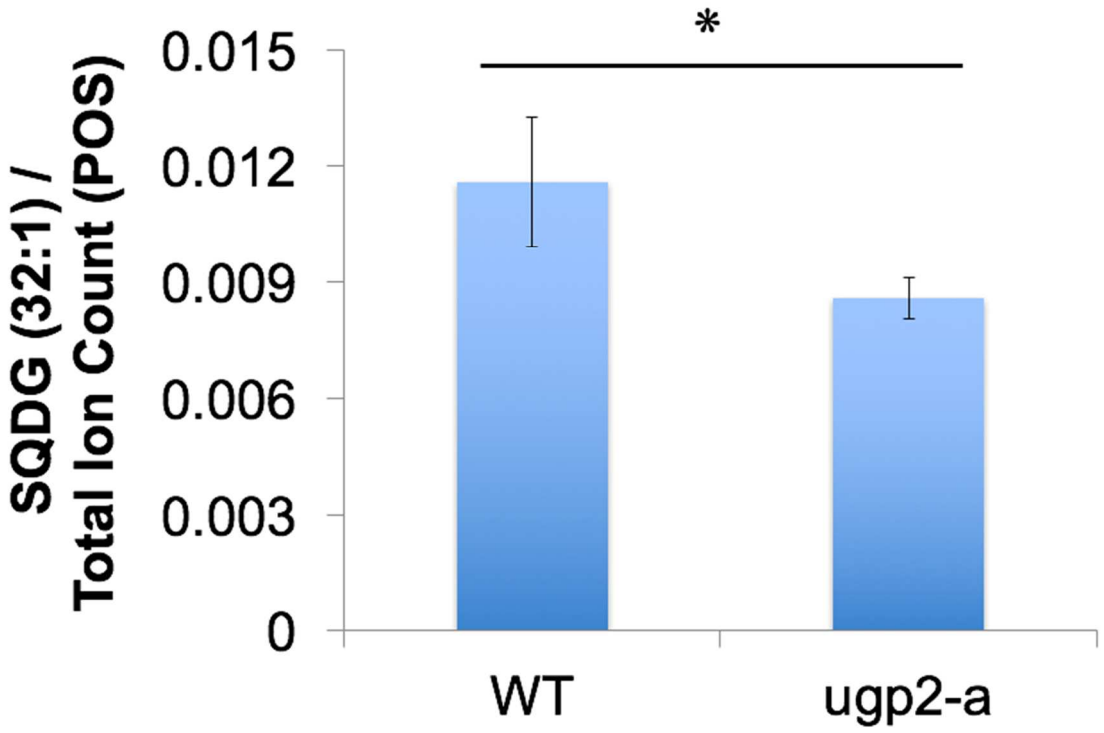

Figure 3.10: Targeted metabolomics experiment investigating SQDG (32:1) content in wild-type Phaeodactylum and the ugp2-a knockout.
The intensity of the SQDG (32:1) signal (m/z = 791.5, retention time = 12.7 min) was normalized by the total ion count in the positive phase (POS) (n = 3, error = standard deviation; Paired student's t- test, * = p < 0.05).

Discussion

Identification of a novel UGPase using a functional complementation screen

Complementing DEV6 with a *Phaeodactylum* cDNA expression library yielded clones that contained plasmids encoding *ugp3* (Table 3.4). Surprisingly, the cDNA library screen did not identify either of the bioinformatically predicted UGPases in *Phaeodactylum*: UGP1, Phatrdraft_50444; UGP2, Phatrdraft_23639. Full-length cDNA sequences of all three UGPases could be amplified from the cDNA library (Figure 3.1), so limited library coverage does not account for the inability of UGP1 or UGP2 to rescue DEV6 cells. Furthermore, DEV6 directly

complemented with *ugp1* or *ugp2* were not viable when grown with galactose as a carbon source (Figure 3.2). However, there are some key limitations that limit our ability to confidently state that the *ugp1* or *ugp2* gene products cannot complement the DEV6 strain of *E. coli*. Namely, there was a lack of verification of transcriptional activity. Because no data was collected regarding *ugp1* or *ugp2* transcriptional accumulation, the lack of a rescue phenotype could be explained because the mRNA for these complement vectors were not expressed properly. This could have been determined using targeted transcript analysis using qRT-PCR. Similarly, it is possible that some defect of transcriptional activity could explain the lack of a rescue phenotype, since there were no there were also no translation controls. It is possible that *ugp1* or *ugp2* gene products successfully expressed their gene products, but formed inclusion bodies instead of functional, soluble protein. Ascertaining protein expression could have been performed by immunoblot using a His-tag antibody, and tracking for protein accumulation after IPTG induction. This immunoblot could be performed in parallel with microscopy to ensure that inclusion bodies were not being formed using these conditions.

Alternatively, assuming successful transcription and translation, some common protein folding challenges may explain the lack of activity of recombinant UGP1 and UGP2, as proteins larger than 60 kDa can be difficult to express properly in *E. coli*[187]. UGP1 is a large fusion protein (~115 kDa), encoding both UGPase and a phosphoglucomutase[91]. Protein folding can be improved at lower induction temperatures[188], but growth at 18 °C did not alter our cDNA rescue results (data not shown). UGP2 is also large protein (~76 kDa) and is predicted to be localized in the chloroplast[91, 189]. The cytoplasmic environment in *E. coli* may not adequate for protein folding and activity of UGP2.

We found that DEV6 complemented with our *ugp1* and *ugp2* rescue vectors did not confer a rescue phenotype when challenged with galactose as a sole carbon source, under our induction conditions. This could be explained by faulty transcription or translation of these gene targets from the expression vector. Nevertheless, this does not invalidate the annotation of *ugp1* or *ugp2* as UGPases in diatoms. Instead, these results suggest that the *upg3* gene product, in addition to the bioinformatically predicted *ugp1* or *ugp2* gene products, could contribute to chrysolaminarin biosynthesis. We turned to phylogenetic approaches to generate hypotheses about the roles of these three targets.

Bioinformatics complemented with biochemical studies enables predictions of metabolic roles for diatom UGPases

The *ugp1*, *ugp2,* and *ugp3* genes have been previously annotated in the NCBI Gene database for *Phaeodactylum*. The *ugp1* gene was previously assigned as a UGPase (EC 2.7.7.9). While *ugp2* was initially assigned as a UDP sugar diphosphorylase (EC 2.7.7.64)[91], it was manually annotated as a UGPase (NCBI). The *ugp3* gene was annotated as a UDP-acetoglucosamine diphosphorylase (EC 2.7.7.23) (NCBI). These assignments were reexamined by replicating a previously published clustering method[6]. This phylogenetic analysis was expanded to include the gene products of *ugp1*, *ugp2*, *ugp3*, and their homologs from closely related stramenopiles, such as *Thalassiosira pseudonana* and *Nannochloropsis gaditana*. Notably, one *Nannochloropsis* protein sequence identified in this manner was manually annotated as a phosphoglucomutase (Uniprot #W7T5H8) but we included it under the assumption that it was misannotated. UppA and UgpB protein sequences from the Amoebozoa *Dictyostelium discoideum* were included, as *uppA* and *ugpB* knockouts significantly reduce the

accumulation of β-1,3 glucan in this species[190], suggesting that homologus enzymes play a role in β-1,3 glucan biosynthesis.

Our clustering analysis indicates that *ugp1* was correctly annotated as UGPase (Figure 3.3). It clusters with the UGPase, Type A group which includes the *Dictyostelium* sequences that have been shown to be important for β-1,3 glucan synthesis for that organism. UGPase, Type A sequences appear to be important for the synthesis of *de novo* nucleotide sugars which is important for callose synthesis[64]. The unique *Nannochloropis* protein sequence annotated as a phosphoglucomutase also clustered with this group, suggesting that it, like *ugp1*, is a protein fusion of phosphoglucomutase and a UGPase. Together, this bioinformatic evidence suggests that *ugp1* is important for β-1,3 glucan synthesis in diatoms. While chrysolaminarin is expected to be the major β-1,3 glucan species, some structural β-1,3 glucans may also exist[95].

The gene products of *ugp2* and *ugp3* both clustered with the UDP sugar disphosphorylases (Figure 3.3). UDP sugar diphosphorylases can catalyze the synthesis of UDP-glucose from UTP and glucose 1-phosphate, but are also capable of synthesizing a range of other UDP sugars, including galactose and mannose[109]. Galactans and mannans are potentially important polysaccharides for protein glycosylation[191] and diatom cell wall polysaccharides[129]. Plant UDP sugar diphosphorylases have limited characterization but have been suggested to contribute to salvage pathways in nucleotide metabolism [59].

Surprisingly, neither of the gene products from *ugp2* or *ugp3* clustered with the UGPase, Type B group, which has been shown to be critical for the synthesis sulfoquinovodiacylglycerol (SQDG), a chloroplast suflolipid. Knocking out the UGPase, Type B in *Arabidopsis* conferred a lipid profile that contained no detectable SQDG[6]. Similarly, knocking out the *Chlamydomonas* homologue abolished SQDG accumulation[192]. SQDG is the second most abundant

glycerolipid class in *Phaeodactylum*[125], so it was surprising that no diatom UGPases clustered into this group. By process of elimination, we propose that either UGP2 or UGP3 could be involved in SQDG biosynthesis. This also represents a potentially novel function of a subgroup of UDP sugar diphosphorylases if either UGP2 or UGP3 plays a role in SQDG biosynthesis.

Therefore, we hypothesized that UGP1 represents the primary contributor to chrysolaminarin synthesis, not UGP2 or UGP3. As a corollary, UGP2 or UGP3 may be involved in other cellular processes that require UDP-glucose, such as SQDG synthesis or nucleotide salvage metabolism. To test these hypotheses, *ugp1*, *ugp2*, and *ugp3* were knocked out using a CRISPR/Cas9 approach.

Distinct phenotypes for each UGPase knockout suggests specialized metabolic roles

We found that *ugp1* knockout mutants exhibit increased BODIPY RFU per cell, as observed previously[7, 8]. In contrast to these studies, we also quantified chrysolaminarin[124]. *Phaeodactylum* maximally accumulates chrysolaminarin at the end of the day period, and we found that *ugp1* mutants only accumulate about 33 % chrysolaminarin relative to their transformation control. While this represents a dramatic decrease in chrysolaminarin per cell at the end of the light period, the fact that *ugp1* mutants accumulate some chrysolaminarin suggests that there exist alternative means of producing some UDP-glucose *in vivo* for β-1,3 glucan synthesis.

A proposed phenotype of disrupting chrysolaminarin metabolism is the repartitioning of cellular carbon into the other major carbon storage molecule in diatoms, triacylglycerol (TAG). TAG content per cell may be semi-quantitatively estimated using the dye BODIPY. TAG, like chrysolaminarin, accumulates at the end of the light period in a diel light regime[124, 154]. The *ugp1* knockouts exhibited 168 % - 180 % BODIPY RFU per cell relative to their transformation

control (Figure 3.6, C). The combined observation of lower chrysolaminarin and higher BODIPY RFU per cell suggests that *Phaeodactylum ugp1* mutants are capable of dramatically repartitioning intracellular carbon reserves from carbohydrate into neutral lipids in a facile manner. Therefore, our findings agree with previous studies concluding that UGP1 is a major contributor to chrysolaminarin biosynthesis.

Chrysolaminarin is a key storage metabolite and its disruption affects nominal cellular carbon partitioning, which may reduce cellular fitness. However, *ugp1* mutants did not have a statistically significant reduction of growth rate (Figure 3.6, A) and we did not observe any changes to the maximum quantum yield of photosystem II, F_v/F_m (Figure 3.6, B). Previous strategies to increase lipid content in diatoms have also observed the lack of a growth phenotype [193]. However, we observed that total organic carbon per cell in the ugp1-a knockout was statistically similar between dawn and dusk (Figure 3.8), suggesting that some carbon might be trapped in metabolic bottlenecks in a diel light regime. Nevertheless, these findings suggest that diatoms can be tolerant to major biochemical modifications, a potential benefit for bioengineering strategies to produce biodiesel from algal lipids.

The *ugp2* knockout mutants exhibited a different phenotypic pattern than the *ugp1* knockout mutants. While chrysolaminarin content was reduced between 44 - 70 % relative to TC_ugp2 (Figure 3.6, D), substantial reductions in BODIPY RFU per cell were also observed, ranging from 33 - 50 % of TC_ugp2 (Figure 3.6, C). The *ugp2* knockouts were also the only mutants to consistently demonstrate lower growth rates (Figure 3.6, A) and Fv/Fm values (Figure 3.6, B). The ugp2-a mutant also has a significantly lower α and P_{max} (Figure 3.9, Table 3.1). The phenotypic patterns of the *ugp2* knockouts are substantially different than that of the *ugp1*

knockouts, and repartitioning of storage metabolites from chrysolaminarin to triacylgycerols was not observed. Therefore, UGP2 is unlikely to contribute to chrysolaminarin synthesis.

Lower photosynthetic efficiencies could provide an explanation for both reduced growth rates and less accumulation of major storage metabolites. If the maximum quantum yield of photosynthesis is reduced, then fewer ATP and NADPH cofactors are generated to drive the Calvin-Benson-Bassham cycle and consequently less carbon is fixed. Since α and P_{max} are reduced and respiration rates are comparable for ugp2-a relative to TC_ugp2 (Table 3.1), net productivity is reduced. Lower net productivity implies that less carbon is available to partition into storage metabolites, explaining the reduced accumulation of both chrysolaminarin and triacylglycerol. Finally, growth rates are negatively affected as fewer carbon and energy reserves are available to fuel metabolism. Overall, the carbon and photosynthetic phenotypes of *ugp2* knockouts are very different than that of *ugp1* knockouts.

Instead, it is possible that UGP2 plays a role in SQDG biosynthesis. UGP2 has a transit peptide (SignalP 3.0)[194] with predicted localization to the chloroplast[189]. As a predicted chloroplastic UGPase, UGP2 may be essential for production of the sulfolipid SQDG, since sulfolipid biosynthesis occurs in the chloroplast[195]. Our preliminary targeted metabolomics experiment suggested that there is a small but significant reduction of the abundant 32:1 SQDG species in the ugp2-a mutant compared to WT (Figure 3.10). However, a key limitation of this pilot study is that other mutants were not tested for SQDG content, as the pilot experiment anticipated a presence/absence phenotyupe of SQDG in the ugp2-a mutant. Knocking out UGPase has been shown to eliminate all detectable SQDG content in *Arabidopsis*[6], *Chlamydomona*s[196], and *Synechocystis*[197]. While *Phaeodactylum ugp2* knockouts do not appear to lack SQDG like their counterparts in other organisms, they do share a pattern of

reduced photosynthetic efficiencies, as *Synechocystis* strains deficient in SQDG confer a slight reduction in P_{max}[197]. SQDG interacts with photosynthetic assemblies[198], playing an important role as a charged, structural lipid. This phenomenon has also been observed for other lipid species abundant in the chloroplast, such as digalactosyldiacylglycerol[199]. A more comprehensive analysis of lipid profiles in all mutants is required to ascertain whether the gene product of *ugp1*, *ugp2*, or *ugp3* play a role in diatom sulfolipid biosynthesis.

Surprisingly, ugp2-a exhibits more than double the UGPase activity compared to WT (Figure 3.7). It is possible that another enzyme with UGPase activity accumulates in response to the absence of UGP2. Isolated chloroplasts can take up exogenous UDP-glucose[200], so it is possible that increased production of UDP-glucose in the cytoplasm can compensate a UDP-glucose deficiency in the chloroplast despite different subcellular localizations. Additional studies investigating potential compensatory mechanisms of UGP1 or UGP3 are required to better understand the unexpected increase of UGPase activity in the ugp2-a mutant. Since antibodies targeting UGPases are not predicted to be reactive for the diatom proteins (Agrisera), relative transcriptional abundances of *ugp1*, *ugp2,* and *ugp3* could be determined in parallel using qRT-PCR, which could help determine if a compensatory mechanism exists at the level of transcription.

The *ugp3* knockout mutants did not demonstrate a phenotype comparable to either *ugp1* or *ugp2* mutants. While BODIPY RFU per cell were significantly different for the mutants relative to TC_ugp3, differences were in opposite directions (Figure 3.6, C). Statistically significant differences of growth rate (Figure 3.6, A) and F_v/F_m (Figure 3.6, B) were small in magnitude and might not be biologically significant. Other patterns such as chrysolaminarin content at the end of the day (Figure 3.6, D), UGPase activity (Figure 3.7), and total organic

carbon patterns (Figure 3.8) were comparable to controls. Only minor, if any, phenotypic changes were observed in the *ugp3* mutants.

The lack of a strong phenotype suggests that UGP3 plays a minor housekeeping role. An alternative explanation is that other UGPases readily compensate UDP-glucose synthesis in its absence, although this reasoning is speculative as no targeted or untargeted transcriptional analysis was performed. Again, targeted transcriptional abundances of *ugp1*, *ugp2,* and *ugp3* could be performed in parallel using qRT-PCR, which could help elucidate potential compensation at the level of transcription. The phylogenetic clustering of *Phaeodactylum* UGP3 indicates that it is a UDP sugar phosphorylase, which suggests that it may have more functional substrates beyond glucose. The production of UDP-mannose is important for mannan synthesis, potentially a component of diatom cell wall polysaccharides[129]. UDP-mannose and other UDP sugars are also important for protein glycosylation[191]. Interestingly, homozygous knockouts of UDP sugar phosphorylase are not viable in plants[170, 201]. Perhaps *Phaeodactylum* knockouts of *ugp3* might provide some insight into the metabolic roles of UDP sugar phosphorylase, but additional characterization of carbohydrate profiles from these mutants is required.

While CRISPR knockouts are a powerful new addition to the diatom genetics toolkit, other experiments could further strengthen our understanding of the roles of the *ugp1, ugp2*, and *ugp3* genes and their gene products. For example, complementing these knockouts would provide even more confidence for the roles of these potential UGPases. This could be performed by tweaking the coding sequence and replacing the targeted gRNA sequence with degenerate codons which encode the same amino acid but do not match the gRNA sequence. Similarly, UGPase overexpression studies have demonstrated changes in carbon partitioning by increasing

carbohydrate content[202], and might be of interest for strategies aiming to increase the total organic carbon content per cell, such as hydrothermal liquefaction[203].

Conclusions

This study set out to complement existing bioinformatics approaches to understand the potential contributions of different UGPases to diatom cell biology. We conclude that the *ugp1* gene product is the primary contributor to chrysolaminarin accumulation based on a significant reduction of chrysolaminarin content per cell and a significant increase of TAG in *ugp1* knockouts, as measured by proxy with BODIPY staining. This conclusion agrees with previous reports that did not directly measure chrysolaminarin. Furthermore, we discount the potential roles of UGP2 or UGP3, whose corresponding mutant phenotypes were not comparable to the *ugp1* knockouts. Instead, the *ugp2* gene product might play a role in SQDG biosynthesis. However, SQDG appears to accumulate in this mutant, unlike SQDG-related UGPase knockouts in *Arabidopsis*, *Chlamydomonas*, and *Synechocystis*. The role of the *ugp3* gene product is still uncertain, but knocking out this gene did not confer substantial, consistent changes for the phenotypes tested. The role of UDP-glucose and UGPase in *E. coli* is that of intermediate sugar metabolism, so it is possible that the *ugp3* gene product fulfills a similar biological role. However, a lack of verifying transcriptional accumulation or translational activity of recombinant *ugp1* and *ugp2* genes and gene products limits the conclusions that these do not have the capacity to rescue DEV6 *E. coli*. While additional characterization is required to clarify the roles of *ugp2* and *ugp3* gene products in diatom metabolism, this study does suggest new metabolic roles for these genes. Gene-by-gene functional characterizations inform metabolic models in diatoms, and improves our understanding of UDP-glucose metabolism throughout the stramenopiles.

Table 3.2: List of primers used in this study.

Project	Primer Name	Description	Sequence
Gateway cloning for DEV6 rescue	MAC073	50444-cds-fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCCTTCTTTCGATCCCAT
	MAC074	50444-cds-rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTACGTAATTACGGTTGGTT
	MAC077	23639-cds-fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGACTAGCCATTGCTGT
	MAC078	23639-cds-rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTACTCTGCCGCCTTTTGCA
	MAC115	attB1+galU_fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTGCCATTAAT
	MAC116	attB2+galU_rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTTCTTAATGCC
	MAC123	54493-cds-fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGGATTTGCAAAGGTC
	MAC124	54493-cds-rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTACAAAATTGCACAAGAAT
CRISPR spacer duplex	DJ_UGP1_Fa	50444 gRNA A fwd	ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGACGGTAGTTCTCAAACTCAA
	DJ_UGP1_Ra	50444 gRNA A rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTGAGTTTGAGAACTACCGT
	DJ_UGP1_Fb	50444 gRNA B fwd	ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGCACGGGTATGGGTCTGGACA
	DJ_UGP1_Rb	50444 gRNA B rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTGTCCAGACCCATACCCGTG
	DJ_UGP2_Fa	23639 gRNA A fwd	ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGTTTGCGTCGTTCGGTCTTCG
	DJ_UGP2_Ra	23639 gRNA A rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCGAAGACCGAACGACGCAAA
	DJ_UGP2_Fb	23639 gRNA B fwd	ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGAACGCCATCAACCTTGACCT
	DJ_UGP2_Rb	23639 gRNA B rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTCAAGGTTGATGGCGTT
	MAC126	54493 gRNA A fwd	ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGGTTCGCAAACTTTGCTCGCT
	MAC127	54493 gRNA A rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGCGAGCAAAGTTTGCGAAC
	MAC130	54493 gRNA B fwd	ACTCTTTCTATTTATATATTTGCAGAAAATCATAGTTTTGCTCGACGAGGCCTACGCAGA
	MAC131	54493 gRNA B rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCTGCGTAGGCCTCGTCGAG
	DJ058	FsgRNAbackbone	AATATATAAATAGAAAGAGTATACCTATAC
	DJ059	RsgRNAbackbone	AAGTTAAAATAAGGCTAGTCCGTTA
HD vector	DJ_UGP1_HOF1	50444 SLIC HA up fwd	AGTCACGACGTTGTAAAACGACGGCCAGTGCTGAAGAAAAAAAAAGAAAAATGAGGCAAA
	DJ_UGP1_HOR1	50444 SLIC HA up rev	GGTTAATTTCGAGCTTGGCGTAATCATGGTTCCAGCTGGGGGACGGGCGCAATAGAGTCT
	DJ_UGP1_HOF2	50444 SLIC ZeoR fwd	AGACTCTATTGCGCCCGTCCCCCAGCTGGAACCATGATTACGCCAAGCTCGAAATTAACC

	DJ_UGP1_HOR2	50444 SLIC ZeoR rev	CTTACGCATTTGAATGACTTGTTTGGCGGTTGAAGACGAGC TAGTGTTATTCCTGACTGT
	DJ_UGP1_HOF3	50444 SLIC HA down fwd	ACAGTCAGGAATAACACTAGCTCGTCTTCAACCGCCAAACA AGTCATTCAAATGCGTAAG
	DJ_UGP1_HOR3	50444 SLIC HA down rev	AGGAAACAGCTATGACCATGATTACGCCAATTCTTCCAAGG CGCCGACCAGCTTGTATAG
	DJ062	Backbone SLIC fwd	GTTTTACAACGTCGTGACTGG
	DJ063	Backbone SLIC rev	CATGGTCATAGCTGTTTCCTG
	DJ_UGP2_HOF1	23639 SLIC HA up fwd	AGTCACGACGTTGTAAAACGACGGCCAGTGCAAGACTTTG GGCTGATCAAAAGCGGTCCC
	DJ_UGP2_HOR1	23639 SLIC HA up rev	GGTTAATTTCGAGCTTGGCGTAATCATGGTGCTCTCCAGGA TGAACGATCACTTTGGTTT
	DJ_UGP2_HOF2	23639 SLIC ZeoR fwd	AAACCAAAGTGATCGTTCATCCTGGAGAGCACCATGATTAC GCCAAGCTCGAAATTAACC
	DJ_UGP2_HOR2	23639 SLIC ZeoR rev	GTCCTCGTGGCGACAAAAGGGCCAGATTCTTGAAGACGAG CTAGTGTTATTCCTGACTGT
	DJ_UGP2_HOF3	23639 SLIC HA down fwd	ACAGTCAGGAATAACACTAGCTCGTCTTCAAGAATCTGGCC CTTTTGTCGCCACGAGGAC
	DJ_UGP2_HOR3	23639 SLIC HA down rev	AGGAAACAGCTATGACCATGATTACGCCAAGCCGCGGGTG ACACTGACACCCAAGGCCGG
Genotyping CRISPR lines	MAC132	50444 flank cut site fwd	GTAAAGTACCACGTACACAAGCA
	MAC133	50444 flank cut site rev	CGTTCGCAGCACTCCATCAA
	MAC134	23639 flank cut site fwd	GGAACCGATTTTCGAGGGGT
	MAC135	23639 flank cut site rev	GGCAACGAGGACAAAAACGG
	MAC136	54493 flank cut site fwd	AGTCGCTTATCTGCCCCTTT
	MAC137	54493 flank cut site rev	GGACGATGGTGATTTGCGAC
	MAC138	50444 inside Cas9 rev	AGAGCACTGTTCCAAGTCCG
	MAC140	23639 inside Cas9 rev	AATCAGCACCGCATGTCGAA
	MAC142	54493 inside Cas9 rev	CAGCACCGCATGTCGAAAAT
	MAC144	50444 flank fwd	GGCGTTGGAATCTGTGTAGA
	MAC146	23639 flank fwd	GCTGTAACCAACCCGCAATC

Table 3.3: List of protein sequences used for phylogenetic analysis.

Pt = *Phaeodactylum tricornutum*, Tp = *Thalassiosira pseudonana*, To = *Thalassiosira oceanica*, Ng = *Nannochloropsis gaditana*, Dd = *Dictyostelium discoideum*, At = *Arabidopsis thaliana*, Cm = *Cucumis melo*, Cr = *Chlamydomonas reinhardtii*, Hs = *Homo sapiens*, Os = *Oryza satvia*, CPa = *Protochlamydia amoebophilia.*

#	**Protein (UniProt)**
1	At AT5G17310 (P57751)
2	At AT5G52560 (Q9C5I1)
3	At AT1G31070 (Q940S3)
4	Ng (W7UC11)
5	At AT3G56040 (F4IY62)
6	At AT2G35020 (O64765)
7	Cm NP 001315373.1 (UPI0004A60CCE)
8	Cm NP 001284438 (Q19TV8)
9	CPa WP 011174874 (*Q6MEF1*)
10	Cr UAP2 (A8HP64)
11	Cr UGP1 (A8ITF3)
12	Cr LPB1 (Q4VGM4)
13	Dd ugpB (Q54YZ0)
14	Dd uppA (P08800)
15	Hs UAP1 (Q16222)
16	Ng Naga 100065g10 (W7T5H8)
17	Ng Naga 100071g9 (W7U8C1)
18	Tp THAPSDRAFT 42326 (B8CC52)
19	Ng Naga 100003g178 (W7UBR1)
20	Os Os04g0613700 (Q0JA66)
21	Os UDPGase (Q9M4X0)
22	Os UGP (A3QQQ3)
23	Os Os05g0468600 (Q0DHF9)
24	Os Os06g0701200 (Q5Z8Y4)
25	Os Os08g0206900 (Q6ZJ97)
26	Ot ostta11g03190 (Q00YG5)
27	Ot ostta13g00320 (A0A090N4K3)
28	Ot ostta13g02030 (A0A096P7I5)
29	Ps USP (Q5W915)
30	Pt PHATRDRAFT 50444 (B7GE51)
31	Pt PHATRDRAFT 23639 (B7GCG1)
32	Pt PHATRDRAFT 54493 (B7FZ25)
33	St ugpA (P19595)
34	To THAOC 34900 (K0R2Q4)
35	To THAOC 17238 (K0S7V2)
36	To THAOC 06412 (K0T4M9)
37	Tp THAPSDRAFT 262059 (B8C141)
38	Tp THAPSDRAFT 13482 (B8C430)

Table 3.4: Sequences identified from plasmids carried by DEV6 clones viable with galactose as a carbon source.

DEV6 cells were transformed with the *Phaeodactylum* cDNA expression library, then plated on M9gal + IPTG + Ampicillin plates. Viable clones were visible after 2 d growth. Colonies were restreaked on M9gal + IPTG + Ampicillin plates to ensure a stable phenotype. Plasmids were then prepared from these colonies, and the inserted cDNA sequence was determined by Sanger sequencing using a commercially supplied T7 primer (Genewiz). Sequences were then queried against the *Phaeodactylum* nucleotide collection via nBLAST (NCBI) and the subsequent gene alignment was reported.

Colony #	Sequence	Gene alignment in Phatr2 database
1	CNNNNNNNNNNNNNNNCNNCTAGANATAATTTTGTTTAACTTTAAGAAG GAGATATACATATGAGATCTGGATCCCNCCNTCACCATCACCATAGATCTG ATATCACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCCCT CGATTTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGAT GGAGGATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCATT GGTTCGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGCC CTGCAACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGGA GCTCGACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAACG CGAAGCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGGC TGGGAGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAGA GTTCGAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGCAAGGTTGGAT TTGTGCTTGTAGCNNNAGGTCTTGGTGAACGTCTCGGCTACAGTAGCATCA AAGTCGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGTACT ACATTGAATATATCTTAGCGGTCCAGGTCAAGTATGGCGAAGGCAAACGTT TGCCGCTATGTATCATGACTTCNGGCGACACGAACGAGAAGACAGCNAGC TACTCCGCAAAACAATTATTTTGGTATGCAGAAGTCGCAAATCANCATCGT CCAGCANGGGCAGGAGNNCCCTGNNCTCATGGATAANANNNCNNNGNGC TGNNANATGATTCCNTCANATCGTCANNAANCNTNNNNGACACGGTGATG TNNNNNCNTACTATANNCNNNNNGGAGTGNCNAACGNNGNNANNNNNGN NNTGANNGNNACNCNNTTCNNNNNNCAANNGTCNTGCNTTNNCNNNNNN NGNNGCNANGNNNNNNNNNNGACNGNNTNNNNNNNNNGNCNNNNNNNN NNNNNNNNNNATNNGGNNTNNCNNNTGNNNNNTNNNNNNNNNANNNNAN AN	Phatrdraft_54493

2	NNNNNNNNNNNNNNNNNNNNNTCTNNNNNNNTTTTGTTTAACTTTAAGA AGGAGATATACATATGAGATCTGGATCCCACCATCACCATCACCATAGATC TGATATCACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCC CTCGATTTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGA TGGAGGATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCAT TGGTTCGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGC CCTGCAACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGG AGCTCGACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAAC GCGAAGCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGG CTGGGAGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAG AGTTCGAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGCAAGGTTGGA TTTGTGCTTGTAGCGGGAGGTCTTGGTGAACGTCTCGGCTACAGTAGCATC AAAGTCGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGTAC TACATTGAATATATCTTANCGGTCCAGGTCAAGTATGGCGAANGCAAACGT TTGCCGCTATGTATCATGACTTCGGGCGACACGAACGAGAAGACAGCAAA GCTACTCCGCAAAAACAATTATTTTGGTATGCAGAAGTCGCAAATCACCAT CGTCCAGCAAGGGCAGGAGTCCCTGNNCTCATGGNTAACACGCNAAAATG GTGCTGGANANAATGATTCCTCCAGATCGTCACGAAACCTCNCGGANNNG NNGATGTTCACGCNTACTANNNNCTCNNGGAGTGGCTAAACGNNGGNNAN NGACGGNNNTGANNNNACGNTNNNTTCANGNNNNNNANGGTCNNGCATT TCNNNNNNNTNCNTTGNATGCTANGNGNTTNNAAAANNNNCTGANNNNNA ATNNNNNNTGNNCNNGAANNNAACNGNANTNNNNGNNTGNCNACNGNAN NCNTNNNNNNNGNNNNN	Phatrdraft_ 54493
3	NNNNNNNNNNNNNNNNTNCCNCNCNNNNNTNNTNTGNTTAACTTTAANAA GGAGAGATACATATGAGATCTGGATCCCNCNNTNNNCATCACCATAGATC TGATATCACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCC CTCGATTTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGA TGGAGGATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCAT TGGTTCGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGC CCTGCAACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGG AGCTCGACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAAC GCGAAGCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGG CTGGGAGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAG AGTTCGAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGNAAGGTTGGA NTTTGTGCTTGTAGCGGGAGGTCTTGGTGAACGTCTCTGCTTACNGTAGCA TCAAAGTCGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGT ACTACATTTGAATATATCTTAGCGGTCCAGGTCAAGTATGGCCGAANGCNA ACGTTTNGCCGCTATGTATCNTGACTTCGGGCGACNCNAAACGANAAGAN AGCNAAGCTACTCCGCAAAACAATTATTTNGGNATGCANACTCGCAANAT CACCATCGTCCAGCAAGGGCAGGGAGTCCCTGCNCTCATGNATAACANCG CCNAANNGNNGNTNGAANAANATGAANTCCTCCAAGNNCGNNCACGAAA CCTCNCGGANANNGGGGNN	Phatrdraft_ 54493

4	NNNNNNNNNNNNNNNNNCTAGAATAATTTTGTTTAACTTTAAGAAGGAG ATATACATATGAGATCTGGATCCCACCATCACCATCACCATAGATCTGATA TCACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCCCTCGA TTTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGATGGA GGATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCATTGGT TCGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGCCCTG CAACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGGAGCT CGACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAACGCGA AGCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGGCTGG GAGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAGAGTT CGAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGCAAGGTTGGATTTG TGCTTGTAGCGGGAGGTCTTGGTGAACGTCTCGGCTACAGTAGCATCAAAG TCGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGTACTACA TTGAATATATCTTAGCGGTCCAGGTCAAGTATGGCGAANGCAAACGTTTGC CGCTATGTATCATGACTTNNGGCGACACGAACGAGAAGACAGCNAAGCTA CTCCGCAAAAACAATTATTTTGGTATGCANAAGTCGCAAATCACCATCGTC CAGCAAGGGNAGGGNGTCCNTGCACTCATNGGATAACAACGCNNAATGGT GCTGGAANANAATGATTCCTCCAGATCGNNCGAANCCNTCACGGACNCGG NGATGNTCANNNCTTACTATACNNTNNNGGANTGGNTNAAACNGGGGNTA AGCNACNGNNTTGAANGNGTTAACGCTNTTNNNAGANNNNANNCNGNN	Phatrdraft_ 54493
5	NNNNNNNNNNNNNNNNTCTANNNNNATTTTGTTTAACTTTAAGAAGGAGA TATACATATGAGATCTGGATCCCACCATCACCATCACCATAGATCTGATAT CACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCCCTCGAT TTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGATGGAG GATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCATTGGTT CGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGCCCTGC AACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGGAGCTC GACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAACGCGAA GCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGGCTGGG AGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAGAGTTC GAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGCAAGGTTGGATTTGT GCTTGTAGCGGGAGGTCTTGGTGAACGTCTCGGCTACAGTANCATCAAAGT CGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGTACTACAT TGAATATATCTTAGCGGTCCAGGTCAAGTATGGCGAANGCAAACGTTTGCC GCTATGTATCATGACTTCNGGCGACACGAACGAGAAGACAGCAAAGCTAC TCCGCAAAACAATTATTTTGGTATGCANAAGTCGCAAATCACCATCGTCCA GCAAGGGCAGGGAGTCCCTGCACTCATGGATAACAACGCCAAAATGGTGC TGGAANANAATGATTCCTCCNNATCGTCACNAANNCACGGACACGGTGAT GTCACGNCTTACTATNCACTCNNNGNNNGNTAAACGNGNTTANCNNNNGT NNNNNNGNTTANGCTATTNCNNNNNNNNNNGNNNGNNNNCACACGNNTC NTTGGANGNNNNNNGNNCAANNNNTNGNCNGATNNNATNNTGNNNNNNC TCNNAANNNNANNNNNNANTGGNGNANNNNNNNN	Phatrdraft_ 54493

6	NNNNNNNNNNNNNNNNCTAGAATAATTTTGTTTAACTTTAAGAAGGAGA TATACATATGAGATCTGGATCCCACCATCACCATCACCATAGATCTGATAT CACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCCCTCGAT TTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGATGGAG GATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCATTGGTT CGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGCCCTGC AACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGGAGCTC GACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAACGCGAA GCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGGCTGGG AGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAGAGTTC GAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGCAAGGTTGGATTTGT GCTTGTAGCGGGAGGTCTTGGTGAACGTCTCGGCTACAGTAGCATCAAAGT CGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGTACTACAT TGAATATATCTTAGCGGTCCAGGTCAAGTATGGCGAAGGCAAACGTTTGCC GCTATGTATCATGACTTCGGGCGACACGAACGAGAAGACAGCAAAGCTAC TCCGCAAAAACAATTATTTTGGTATGCAGAAGTCGCAAATCACCATCGTCC AGCAAGGGCAGGGAGTCCCTGCACTCATGGATAACAACGCCAAAATGGTG CTGGAAGAGAATGATTCCTCCAAGATCGTCACGAAACCTCACGGACACGG TGATGTTCACGCCTTACTATACACTCACGGAGTGGCTAAACGGTGGTTAAG CGACGGTATTGAATGGNNAACGCTATTTCAAGACACAAACGGTCTTGCATT TCACACACTTTCCTTTGATGCTNNNGTTTCCAAAAAGTTNGGNNCNGATTA TNNNNCNNNNNTGTGCCTCNNANNCNANCAGNAATNGGNNNNTTGCCAAN NNNAGCATCAAACAANNGGGNNNNTNNANNNNN	Phatrdraft_ 54493
7	NNNNNNNNNNNNNNNNNNNNNNNNNTCCCTCTNNNNNNNTTTTGTNNNA CTTTAAGAAGGAGATATACATATGAGATCTGGATCCCACCATCACCATCAC CATAGATCTGATATCACAAGTTTGTACAAAAAAGTTGGCGCATTCTTACCC TGGCTCCCTCGATTTCGGACGACTCCCATCATGACGAGCAGTAGCGATAAC ATTAGGATGGAGGATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAA AAAGCATTGGTTCGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCC AAGTGGCCCTGCAACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAG CTTCAGGAGCTCGACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACAT AAATAACGCGAAGCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGC TCGAAGGCTGGGAGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGA ACGAAAGATTTCCAGGCANCGGAGTCGACNGGACGGGCCAGAACTTGGNT AGGTTGGAGTTGAGCTTGCNCNANNAGTACTTANGCNAAGTNTNCNTTAC GGNTGTTTATAGTNGATTGTCCATCGANNN	Phatrdraft_ 54493

8	NNNNNNNNNNNNNNNNNNCTANNNNNNTTTTGTTTAACTTTAAGAAGGA GATATACATATGAGATCTGGATCCCACCATCACCATCACCATAGATCTGAT ATCACAAGTTTGTACAAAAAAGTTGGACGCATTCTTACCCTGGCTCCCTCG ATTTCGGACGACTCCCATCATGACGAGCAGTAGCGATAACATTAGGATGG AGGATTTGCAAAGGTCTCTGACGGTCCTTTCTGATCCTCAAAAAGCATTGG TTCGCAAACTTTGCTCGCTGGGACAGTCACACTTGTTTGCCAAGTGGCCCT GCAACACGTCTCCCACCTCTAGGCGGAAACTCGCTACGCAGCTTCAGGAGC TCGACGAGGCCTACGCAGATGGAGGTCTTGAAGGCTACATAAATAACGCG AAGCGATTGTTGTCGGACAGCCGAGAAGGAGTTAATCCGCTCGAAGGCTG GGAGCCAAGCGTTCCTGAAGGGGAGCGTTTCGACCTCGGAACGAAAGAGT TCGAGGAAACAGAGTCGACTGGACGTCCAGAACTCGGCAAGGTTGGATTT GTGCTTGTAGCGGGAGGTCTTGGTGAACGTCTCGGCTACAGTAGCATCAAA GTCGGTTTACCCACAGAAATGGCTACTGAAACATGCTACCTCCAGTACTAC ATTGAATATATCTTAGCGGTCCAGGTCAAGTATGGCGAANGCAAACGTTTG CCGCTATGTATCATGACTTCNGGCGACACGAACGAGAAGACAGCAAAGCT ACTCCGCAAAAACAATTATTTTGGTATGCAGAAGTCGCAAATCACCATCGT CCAGCAAGGGCAGGGNGTCCCTGCACTCATGGATAACAACGCCAAATGGT GCTGGAAGAGAATGATTCCTCCAAGATCGTCACGAAACCTCACGGACACG GNGATGTTCACGCCNTACTATACACTCACGGAGNNGNTAAACGGNNGNNA AGCGANGGNNNNTGAATNNNNACNNCTATTTCANGANNNNNAACGNCTG CATTTCNNNNNANNNCTTTGATGCTAGNNNNTTNCAAAAGTNGNANCTGA TNATGAANTCNNNGNTGNGCNNNNGNN	Phatrdraft_ 54493

Table 3.5: Summary of all phenotypes for all UGPase knockouts.
Exponentially growing cultures were assayed for growth rate, maximum quantum efficiency of Photosystem II (F_v/F_m), BODIPY RFU per cell, a proxy for neutral lipid accumulation, and chrysolaminarin content per cell. Results for insoluble carbohydrates and total carbohydrates are also presented. Mutant strains visualized in Figure 4 are in bold. (n=3, ± standard deviation).

Strain	BODIPY (RFU cell^{-1}1)	Growth Rate (day^{-1})	Fv/Fm	Chrysolaminarin (pg g.e. cell^{-1})	Insoluble Carbohydrates (pg g.e. cell^{-1})	Total Carbohydrates (pg g.e. cell^{-1})
WT	2319482 ± 62436	1.02 ± 0.06	0.601 ± 0.013	2.18 ± 0.25	1.24 ± 0.19	4.35 ± 0.38
TC_U1	2186592 ± 248264	1.06 ± 0.04	0.600 ± 0.007	1.79 ± 0.21	0.91 ± 0.26	3.54 ± 0.51
U1_a_C4	3645488 ± 221148	0.80 ± 0.09	0.587 ± 0.013	0.48 ± 0.03	1.01 ± 0.18	1.79 ± 0.25
U1_a_E3 (ugp1-a)	**3665600 ± 406548**	**0.82 ± 0.12**	**0.611 ± 0.005**	**0.56 ± 0.08**	**0.98 ± 0.16**	**1.91 ± 0.22**
U1_a_G3	4294777 ± 250407	0.81 ± 0.14	0.621 ± 0.013	0.42 ± 0.09	1.28 ± 0.21	2.05 ± 0.30
U1_b_C6	3428870 ± 329351	0.88 ± 0.13	0.599 ± 0.004	0.46 ± 0.12	1.23 ± 0.05	2.04 ± 0.12
U1_b_E9 (ugp1-b)	**3937720 ± 558675**	**0.84 ± 0.13**	**0.600 ± 0.024**	**0.61 ± 0.13**	**1.35 ± 0.15**	**2.34 ± 0.31**
U1_b_F9	3510048 ± 106074	0.79 ± 0.16	0.614 ± 0.008	0.52 ± 0.19	1.25 ± 0.25	2.13 ± 0.44
TC_U2	2117353 ± 318437	1.06 ± 0.08	0.601 ± 0.008	1.94 ± 0.18	1.01 ± 0.21	3.85 ± 0.37
U2_a_A4	784329 ± 35170	0.45 ± 0.01	0.439 ± 0.009	1.60 ± 0.06	1.66 ± 0.17	3.92 ± 0.14
U2_a_D1 (ugp2-a)	**692136 ± 77436**	**0.51 ± 0.03**	**0.448 ± 0.020**	**1.36 ± 0.12**	**1.33 ± 0.05**	**3.39 ± 0.22**
U2_a_F1	1114277 ± 221263	0.73 ± 0.04	0.537 ± 0.012	1.20 ± 0.13	1.49 ± 0.03	3.40 ± 0.16
U2_b_F1	1136146 ± 302808	0.65 ± 0.15	0.533 ± 0.039	1.79 ± 0.04	1.15 ± 0.06	3.61 ± 0.04
U2_b_D7 (ugp2-b)	**297342 ± 82606**	**0.64 ± 0.11**	**0.512 ± 0.028**	**2.05 ± 0.12**	**1.29 ± 0.04**	**4.20 ± 0.17**
U2_b_C6	881345 ± 669730	0.63 ± 0.16	0.511 ± 0.034	1.67 ± 0.08	1.12 ± 0.19	3.48 ± 0.30
TC_U3	2785273 ± 141344	0.95 ± 0.03	0.604 ± 0.013	2.25 ± 0.33	1.37 ± 0.22	5.53 ± 0.66
U3_a_F2	2713501 ± 455805	1.12 ± 0.14	0.586 ± 0.014	1.91 ± 0.34	1.10 ± 0.13	4.12 ± 0.78
U3_a_C4 (ugp3-a)	**1950726 ± 250432**	**0.85 ± 0.02**	**0.597 ± 0.026**	**2.88 ± 0.22**	**1.45 ± 0.10**	**5.84 ± 0.08**
U3_b_A2	3295835 ± 523730	0.99 ± 0.13	0.634 ± 0.013	2.35 ± 0.53	1.74 ± 0.03	5.78 ± 0.65
U3_b_B2 (ugp3-b)	**3726592 ± 66478**	**0.93 ± 0.01**	**0.611 ± 0.010**	**2.78 ± 0.66**	**1.91 ± 0.19**	**6.42 ± 1.06**

CHAPTER 4: IDENTIFYING NOVEL DIATOM β-GLUCAN BINDING PROTEINS

Preface

The previous chapter characterized three enzymes based on their catalytic activities potentially contributing to chrysolaminarin synthesis. However, biochemical evidence associated with chrysolaminarin metabolism is negligible beyond this first biosynthetic step. I set out to expand our understanding of chrysolaminarin-related proteins using a protein-carbohydrate affinity technique. In this chapter, I describe the strategy implemented to discover these potentially novel proteins. Additional analysis of the strategy suggested that the experimental design was suboptimal, and I suggest that the resolved spots were a technical artifact, although several critical controls are missing to confidently come to this conclusion. Nevertheless, I conclude by describing alternative experimental strategies for future pursuit of chrysolaminarin-binding proteins.

Introduction

Diatoms are stramenopile algae, taxonomically assigned to the Stramenopile-Alevolate-Rhizaria (SAR) supergroup of eukaryotes. During their evolutionary history, diatoms underwent two distinct endosymbiotic events[204] and significant horizontal gene transfer with prokaryotes[48]. Consequently, diatom genomes are as dissimilar to plant and animal genomes as plant and animal genomes are to each other. Of 7166 diatom gene families predicted in the model diatom *Phaeodactylum tricornutum*, 4721 (66 %) are exclusive to the diatoms, when compared to plants, red algae, and other eukaryotes[48]. It is reasonable to suggest that the genetic blueprint for many unique diatom biochemistries[34] are encoded within these diatom-exclusive gene families.

An ongoing challenge in diatom biology is the functionally accurate assignment of genes to their roles in orphan biochemical pathways. Orphan pathways are those which produce a known metabolite, but the genes and enzymes responsible for synthesizing that metabolite remain unknown. One strategy to address the metabolic gaps from orphan metabolism relies on comparative genetics and modeling[205]. This is performed by finding sequence similarities between genes from a well-characterized reference and a query. Ideally, the reference's gene product has been experimentally verified for its enzymatic activity. Another approach is gene inactivation and phenotyping[206]. After disrupting expression of a gene of interest, phenotyping evidence can be collected for the gene product's direct contribution to orphan metabolism. However, these approaches remain challenging for identifying *Phaeodactylum* orphan metabolism targets, as most of its predicted gene families are not shared with other eukaryotes.

I am particularly interested in identifying and characterizing enzymes involved in chrysolaminarin synthesis, the unique storage polysaccharide of diatoms[62]. While chrysolaminarin metabolism models have been drafted for diatoms[91], they are established on physiologically distinct and evolutionarily distant β-1,3-glucan-related enzymes such as yeast cell wall β-branching enzymes and plant callose synthases[91]. Chrysolaminarin serves as a carbon and energy reserve, while yeast cell wall glucans and callose are both structural carbohydrates. These represent fundamentally different metabolic roles despite catalyzing similar enzymatic reactions. This distinction suggests cautious interpretation of such models. Gene assignments are most robust when reference and target templates are similar, and both plants and yeast are evolutionarily distant from diatoms[12]. Nevertheless, this approach has generated hypotheses for proteins whose functions *in vitro* or *in vivo* might play a role in chrysolaminarin biosynthesis, such as branching[103] and β-glucan synthase[102] enzymes.

I propose that functional biochemistry approaches would strongly complement bioinformatics-assembled models[206], which could expand our understanding of chrysolaminarin metabolism. For example, enzyme activity assays indicated that more UDP-glucose was synthesized than any other nucleotide sugar in cell-free protein extracts of the diatom *Cyclotella cryptica*[90]. This evidence provided the basis for directed gene knockouts and knockdowns in *Phaeodactylum* to study the effect of disrupting UDP-glucose synthesis on central carbon metabolism[8]. Therefore, providing evidence for or against chrysolaminarin-related targets with a biochemical evidence may expand our understanding of chrysolaminarin-related proteins and diatom biology.

Classical biochemical strategies represent another strategy to identify novel genes and their encoded enzymes contributing to chrysolaminarin biosynthesis. It is known that UDP-glucose is an important substrate for chrysolaminarin biosynthesis from Chapter 3. It would be possible to incubate protein fractions with excess UDP-glucose and then assay for *in vitro* chrysolaminarin accumulation. However, the chrysolaminarin quantification method outlined in Chapter 2 had not yet been established at the time of the experimental design of this chapter. Because no direct assay for chrysolaminarin synthesis existed, I had to rely on other strategies to identify novel proteins involved in chrysolaminarin metabolism.

In the place of a classical biochemical approach, I developed a strategy to identify carbohydrate-binding proteins. Multiple enzymes important in storage polysaccharide metabolism bind their substrate. Starch and glycogen binding domains are well characterized and curated in the Carbohydrate Active enZYme database[110]. Overall, an opportunity exists to identify and characterize diatom β-1,3 glucan binding sequences and proteins. Functional

identification of these proteins could enable broader gene assignment efforts throughout the stramenopiles that synthesize a variety of β-1,3 glucans.

I hypothesized that (1) β-1,3-glucan binding proteins exist in the *Phaeodactylum* proteome and (2) that they are important for chrysolaminarin metabolism. This chapter highlights efforts to identify novel β-glucan binding proteins using a protein-carbohydrate affinity based approach. A strategy to characterize the role of that protein in chrysolaminarin biology was also implemented.

Identifying affinity-based interactions using an electrophoresis-based technique has been performed using a wide variety of techniques, as reviewed by Heegaard[207]. Generally, these techniques have been performed either in gels or using capillary electrophoresis. Both strategies have been used to characterize carbohydrate-protein interactions. First, an in gel strategy known as the 2D-Affinity Electrophoresis method (2D-AE, Figure 4.1) resolved starch-binding proteins from a barley endosperm crude protein extract[9]. Second, an affinity capillary electrophoresis approach observed differential migration of putative β-globulins from human serum by their interactions with a derivatized gentiobiose. I was interested in the 2D-AE method as it has been used to characterize a variety of protein-carbohydrate interactions[208].

I adapted the 2D-AE protocol to identify β-1,3-glucan binding proteins from diatom crude protein extracts by replacing starch with laminarin. Laminarin is a water-soluble β-1,3-glucan that is nearly identical to chrysolaminarin. Embedding water-soluble β-1,3 glucans has previously been used to identify cellulases in an *in situ* assay after electrophoresis[209], so I reasoned that these changes could allow us to identify novel β-1,3 glucan binding proteins that are important for chrysolaminarin biology.

Laminarin is composed of two polysaccharide species, the M-chain and the G-chain. These species refer to the terminal sugar residue (mannose and glucose, respectively) for an otherwise β-1,3 glycosidic backbone. The ratio of M-chains to G-chains can vary by species[210]. It is possible that the terminal mannose in laminarin could interfere with this approach, but the biochemical similarities of chrysolaminarin to laminarin has been emphasized in the literature. Moreover, the large-scale chrysolaminarin purification strategy described in Chapter 2 which enabled ^{1}H-NMR analysis had not yet been developed when the experimental design for this chapter was being developed. Therefore, the use of laminarin was a practical substrate for the development of the 2D-AE protocol to identify β-1,3-glucan binding proteins, although there are obvious biochemical limitations for using a biochemical substitute for chrysolaminarin, the polysaccharide of interest in this study.

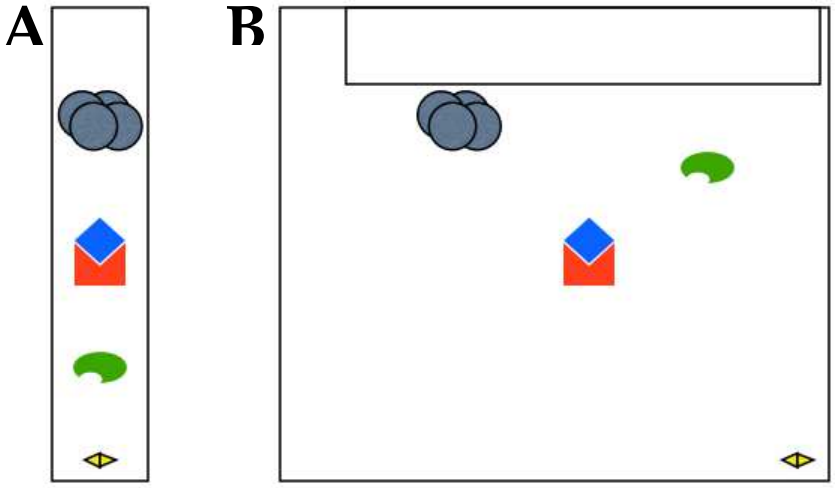

Figure 4.1: A model of the 2D-affinity electrophoresis method.
2D-affinity electrophoresis isolates proteins by in-gel binding affinity. Total protein is run by Native PAGE (A). An excised lane is transferred into a custom Native PAGE gel with an embedded ligand of interest (B). Most protein does not interact with the embedded ligand, yielding a diagonal line. Protein interacting with the ligand may migrate more slowly, resolving a gel spot relative to the diagonal line.

I resolved a protein spot using 2D-AE and identified a protein that migrated in this spot. RNAi knockdown strains were generated and characterized for growth rate and chrysolaminarin

accumulation. Surprisingly, these strains did not accumulate less protein. I was unable to purify expressed recombinant protein for additional characterization. Later experiments led to the conclusion that the gel spot migration appears to be an artifact of protein overloading. As several core assumptions in the experimental design of this study were challenged, the interpreted strength of our conclusions weakened. These limitations were carefully considered and discussed in order to benefit a revised experimental strategy to identify novel β-1,3 glucan binding proteins from the *Phaeodactylum* proteome.

Methods

Culture conditions

Phaeodactylum tricornutum was cultured axenically under constant illumination (200 μmol photons m^{-2} s^{-1}). Cells were provided with f/2 media as prescribed by Guillard et al.[149] but with 2.3 times the concentration of all nutrients to ensure that cells were nutrient replete. Silica was omitted from the medium.

Protein extraction and quantification

French press

Biomass was harvested from these cultures by centrifugation (3220 x g, 18 °C, 10 min). Harvested cells were suspended in a protein extraction buffer (50 mM Tris HCl, pH 7.5) with a small addition of Tween-X100 (0.1 %) to facilitate disruption. The cells were passed once through a French pressure cell press. Cell disruption was confirmed by examining the flow-through by microscopy. This flowthrough was further purified by centrifugation in microeppendorf tubes (10000 x g, 4 °C, 10 min) to separate the protein-containing supernatant from residual cellular debris. Supernatant was transferred to a fresh microeppendorf tube and later quantified, as described below. Protein was stored at -80 °C.

Hand-probe sonication

Biomass was harvested from these cultures by centrifugation (3220 x g, 18 °C, 10 min). Harvested cells were resuspended in a 2 mL volume of protein extraction buffer in a 15 mL Falcon tube on ice. The probe sonicator (QSonica, #Q55) was set to an intensity of 70 % and the suspension was sonicated manually in three intervals of 15 seconds on : 15 seconds off. The suspension was examined by microscopy to ensure cell disruption. The 2 mL volume was transferred to a microeppendorf tube to separate soluble protein from cellular debris (10000 x g, 4 °C, 10 min) and the supernatant was transferred to a fresh microppendorf tube and was stored at -80 °C until quantification and use at a later time.

Protein quantification with BCA assay

Protein quantification was performed per the manufacturer's instructions (Pierce BCA Protein Assay Kit, #23227, ThermoFisher). Briefly, a standard curve was prepared with provided albumin and quantified in parallel with samples of interest. 50 µL aliquots of protein samples reacted with 1 mL of prepared working reagent for 30 mins at 37 °C on a dry heat block. After this incubation, absorbances at 562 nm were measured Cary UV-60 Spectrophotometer (Agilent) using diH_2O as a blank. A 2nd degree parabolic curve function was used to fit the standard curve in order to quantify protein concentration for the samples tested.

<u>Identification of putative diatom β-glucan binding proteins with 2D-AE</u>

Protein separation by 2DAE

The approach of 2D-AE has been previously described[9]. Briefly, a protein sample (5 µg - 30 µg) was separated in the 1st dimension via Native-PAGE using a polyacrylamide concentration ranging from 6 – 12 % (Novex precast gels), which resulted in a 1-dimensional protein lane. Running conditions were 125 V for approximately 2 h at room temperature or 18

°C. Running buffer (25 mM Tris base, 192 mM Glycine, pH 8.3) was reused for 3-4 runs. Protein migration was tracked by both the protein ladder (Page Ruler Prestained protein ladder, ThermoFisher #26616) and chlorophyll in the loaded samples. The 1D protein lane was excised with a scalpel and carefully transferred into a custom-poured gel. The custom-poured separating gel was prepared with 2 g L^{-1} laminarin (Sigma, #L9634) and polyacrylamide at either 6 % or 12 % in 1.5 mm cassettes (ThermoFisher, #NC2015) with a 1.5 mm 2D well comb (ThermoFisher) as described by the manual (#MAN0001660), except that the stacking gel omitted sucrose. Fresh running buffer was provided for the 2^{nd} dimension and the running time was extended to 2.5 h to 8 h, depending on running temperature.

Visualizing protein gels

Imidazole-zinc staining provided a rapid visualization of protein content in a gel[211]. Imidazole-zinc staining required a preliminary incubation in the detergent sodium dodecyl sulfate (SDS) to facilitate staining, so it was not used during the 2D-AE method, as native protein would be denatured prior to the affinity-separation step. Imidazole-zinc staining was used after 2D-AE separation, where native gels were shaken in 0.1 % SDS for a 10 min incubation and then rinsed with diH_2O. The gel was then shaken for 10 min in 0.2 M imidazole. The imidazole was discarded after incubation and shaken for approximately 30 s with 0.3 M $ZnSO_4$ (substitute for $ZnCl_2$). Then, the zinc solution was discarded and the gel briefly rinsed with diH_2O. A white zinc-imidazole precipitate formed in areas where there was no protein, so protein was visualized as transparent bands/spots when placed on a dark background.

SYPRO-Ruby (ThermoFisher, #S12000) represented a more sensitive protein visualization approach. First the polyacrylamide gel was shaken for 15 min in a tupperware with 100 mL fixing solution (50 % v/v methanol, 7 % v/v acetic acid). The fixing solution was

discarded and fixation step was repeated with 100 mL fresh fixing solution. After discarding the fixing solution from the 2nd fixation step, the gel was suspended with 60 mL SYPRO Ruby stain, protected from light, and gently shaken overnight. The next day, the SYPRO Ruby stain was discarded and the gel was washed twice for 30 min with a washing solution (10 % v/v methanol, 7 % v/v acetic acid). After these washes, protein was visualized by excitation with UV light.

Protein sequencing using Mass Spectrometry

Visualized gel spots were submitted to the CSU Proteomics and Metabolomics Facility for protein sequencing, following their protocol as described here. Briefly, protein was extracted from the submitted gel spot and samples were digested with trypsin. Peptides were purified and concentrated (Thermo Scientific 5µm, 100 µm ID x 2cm C18 column). Purified peptides were separated by reverse phase chromatography (Thermo Scientific EASYnano-LC, 3µm, 75 µm ID x 100mm C18 column) with a 30 min linear gradient of 10 % - 30 % buffer B (100 % acetonitrile, 0.1 % formic acid) with a flow rate of 400 nL / min. Peptides were eluted directly into the mass spectrometer (Thermo Scientific Orbitrap Velos). MS/MS samples were analyzed using the Phatr2 Phaeodactylum protein database (NCBI) using Mascot (Matrix Science, London, UK; version 2.3.02) after processing the tandem mass spectra in ProteoWizard (MSConvert). MS/MS based peptide and protein identification was validated in Scaffold (version 4.4.1.1, Proteome Software Inc., Portland, OR). Peptide identifications were accepted if they had a false discovery rate of < 0.1 % from the Scaffold algorithm. Protein identifications were accepted if they had a false discovery rate of < 1.0 % from the Protein Prophet algorithm[212] as well as 2 identified peptides. A mixture of 6 trypsin-digested bovine proteins were used as quality control samples to validate instrument stability and data acquisition.

Immunoblots

Protein transfers relied on the XCell SureLock Mini-Cell Electrophoresis system with the XCell II blot module (ThermoFisher, #EI9051). Gels were extracted from the Novex precast cassettes and placed into a PVDF membrane sandwich as described (#MAN0000740). Tris-glycine transfer buffer (20 % methanol, 12 mM Tris Base, 96 mM glycine, pH 8.3) was used as all custom and precast gels used the Tris-Glycine buffer system. The transfer ran at 25 V for 1 h – 2 h, and the transfer was verified using the PageRuler Prestained Protein Ladder (ThermoFisher, #26616). After protein transfer, PVDF membranes were incubated overnight at 4 °C in blocking solution (1x TBS (50 mM Tris, pH 7.5 + 150 mM NaCl) + 5 % milk w/v). Membranes were washed the following morning 4 times, with each wash lasting 5 min with fresh 1 x TBS + 0.5 % milk w/v. Primary antibody concentrations were the following: anti-Phatrdraft_47612 (Genscript) at 1 : 636 (corresponding to 1 µg custom antibody mL^{-1}); anti-AtpB (Agrisera) at 1 : 5000; anti-urease at 1 : 5000. The anti-Phatrdraft_47612 antibody was generated by expression of a recombinant protein sequence synthesized by Genscript, corresponding to residues 19-436 of Phatrdraft_47612 (XP_002181876.1, NCBI), which was used to challenge rabbits for the PolyExpress Premium Antigen-Specific Affinity Purified polyclonal antibody package. The anti-urease antibody was kindly gifted by the Andrew Allen Lab (JCVI). Primary staining took place by shaking at room temperature for 1 h. After primary staining, membranes were washed 4 x 5 min with 1 x TBS + 0.5 % milk (w/v). The secondary antibody for all western blots was a donkey anti-rabbit antibody conjugated with horseradish peroxidase used at a final concentration of 1:50000, corresponding to 0.8 ng antibody mL^{-1} (ThermoFisher, #SA1-200). The secondary antibody was prepared and stored as a 10 x solution. Secondary staining took place by shaking at room temperature for one hour. Visualization of the

western blots took place with the SuperSignal West Femto Max Sensitivity substrate (ThermoFisher, #34095). 2 mL of prepared substrate was sufficient to coat a PVDF membrane. The substrate incubated with the PVDF membrane on saran wrap for 5 min. Excess reagent was blotted off prior to visualization using UV. Separate images were taken using white light to visualize the membrane or UV to view the illuminated protein spots. These images were superimposed using the Gel Doc software (Bio-Rad). Immunoblots are representative of at least 2 technical replicates, unless otherwise noted. Quantitative immunoblots were performed by calculating the integrated pixel densities of Phatrdraft_47612 bands visualized from a blot in ImageJ, and normalized to the integrated pixel densities of AtpB bands visualized from a second blot run in parallel.

Generating and screening Phatrdraft_47612 RNAi lines

Vector design and construction

The RNAi vector pMAC3 was designed and synthesized (Genscript, Figure 4.2). Expression of *sh ble* driven by the light-independent Histone 4-1β promoter confers resistance to Zeocin at a concentration of 100 μg mL^{-1} in diatoms[172, 179] (ThermoFisher, #R25001). A unique EcoRV restriction site facilitated blunt ligation of a sequence of interest immediately after the *sh ble* stop codon. The sequence of interest was defined as an approximately 250 base pair sequence unique to the targeted RNAi transcript[100]. Primers unique to the target were identified with primer-BLAST, and their respective amplicon was compared for target specificity relative to the *Phaeodactylum* reference RNA sequence database (taxonomic identifier: 556484) using nBLAST (NCBI). At least 10 sequences were analyzed in this manner and the two sequences that had the least potential for off-target annealing are identified as sequences "A" and "B." Generally, sequences closer to the start codon were preferred. This process yielded primer

pair MAC043 and MAC044 for generating the "A" amplicon and yielded primer pair MAC045 and MAC046 for the "B" amplicon targeting Phatrdraft_47612. These two amplicons were ligated into the unique EcoRV site in pMAC3 and antisense orientation was verified by Sanger sequencing.

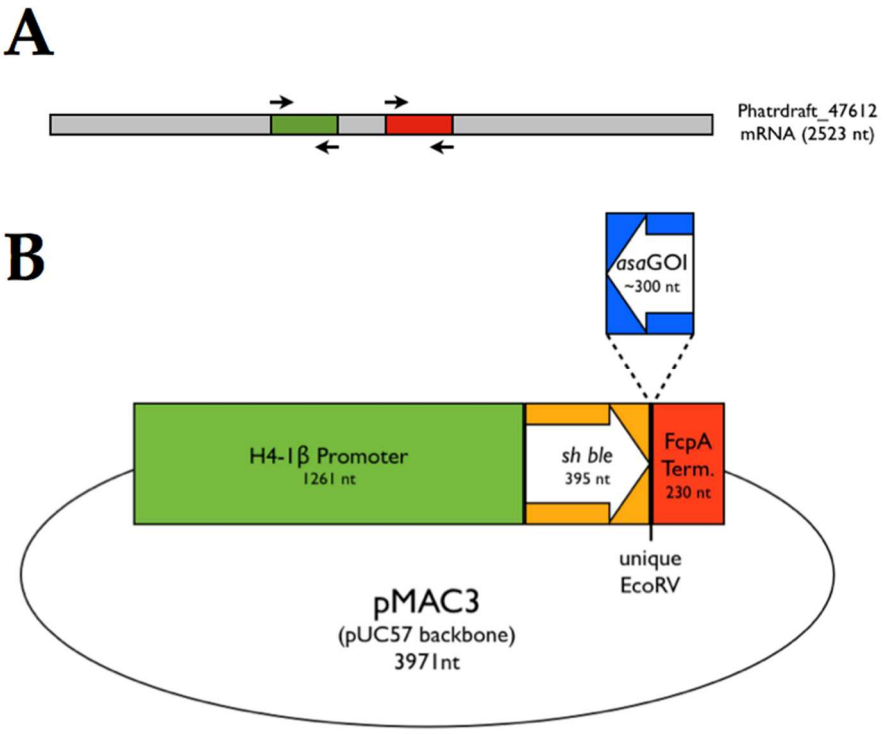

Figure 4.2: Overview of RNAi knockdown approach.
15+ primer pairs yielding a product of about 300 nucleotides were created with Primer-BLAST. Sequence specificity to Phatrdraft_47612 relative to the *Phaeodactylum tricornutum* transcriptome was tested with nBLAST. The two most specific, non-overlapping amplicons were selected to enable construction of two unique knockdown vectors (A). A custom RNAi vector, pMAC3, was previously designed and synthesized (B, Genscript). Expression of a codon optimized *sh ble* is driven by a light-independent histone promoter. Amplicons were ligated into the unique EcoRV site of pMAC3 and transformed into *E. coli.* Antisense orientation was verified by sequencing isolated plasmids. These constructs were then used for biolistic transformation.

Generating, screening, and characterization of putative RNAi lines

Sequence verified "A" and "B" knockdown vectors were prepared at midiprep volumes of 50 mL – 200 mL, following the manual's instructions (Sigma, #NA0200). Between 1 µg and 5 ug of plasmid were bombarded into plates of prepared *Phaeodactylum* as previously described in

Chapter 3. Zeocin-resistant colonies were restreaked on Zeocin plates to confirm resistance prior to screening. Our primary screen selected strains exhibiting lower growth rates than wild-type.

Transcript quantification with qRT-PCR

Approximately 1 x 10^8 cells were harvested for RNA extractions via centrifugation (3220 x g, 18 °C, 10 min). These cells were respuspended in 1 mL of f/2 media, transferred into 1.5 mL microcentrifuge tubes, and spun down (10000 x g, 18 °C, 10 min), discarding the supernatant. These cell pellets were stored at -80 °C until RNA extraction. RNA was extracted from these cell pellets with 1 mL TRIzol Reagent (ThermoFisher, #15596026), per the manufacturers' protocol for RNA extraction. Contaminating genomic DNA was removed from the RNA suspension using TURBO DNase following the manufacturer's protocol (ThermoFisher, #AM2238), rather than DNase I. RNA concentrations were determined either with the Qubit RNA kit (ThermoFisher, #Q10210) or on a Nanodrop 2000 (ThermoFisher). Approximately 1 μg of total RNA was used for first-strand cDNA synthesis in a mixture also containing SuperScript III reverse transcriptase (ThermoFisher, #18080093), 50 ng oligo dT, and 500 μM dNTPs. This final mixture was diluted to a concentration of about 40 ng total RNA μL^{-1}, from which 1 μL was used as template for qRT-PCR. A list of all primers designed and used in this study is provided (Table 4.3, end of chapter). Phatrdraft_47612 cDNA was amplified using MAC067 and MAC068. PCR primers amplifying the TATA Binding Protein (TBP, Phatrdraft_10199) were used to generate a control amplicon, as previously described[172]. Amplification efficiency was tested for these primer pairs (data not shown) and applied for relative quantification using the ΔΔCT method.

Determining cell biovolume

Cell volume was determined as described by Hillebrand[213]. The length and diameter of 30 – 100 diatom cells were measured using an ocular ruler eyepiece on Micromaster (Fisher, #125614B) at 100X with oil immersion.

Chrysolaminarin content

Chrysolaminarin content was determined in glucose equivalents as described in Chapter 2.

Expression and purification of recombinant Phatrdraft_47612

Vector design and construction

The Phatrdraft_47612 transcript was amplified from previously prepared *Phaeodactylum tricornutum* WT cDNA. The sequence was cloned into two different expression vectors: pHAT4 and pDEST-HisMBP[214]. The pHAT4 vector was a gift from Mam Scherman (Protein Expression/Purification Facility, Colorado State University). The pDEST-HisMBP vector was a gift from David Waugh (Addgene plasmid #11085)[214]. Phatrdraft_47612 was cloned into pHAT4 using a restriction endonuclease and ligation approach, which generated the pHAT4/Phatrdraft_47612 vector. Cloning into pDEST relied on the Gateway cloning approach that required in-frame addition of attB1 and attB2 sequences to Phatrdraft_47612 using cDNA as a template. The coding sequence of Phatrdraft_47612 was first amplified with MAC085 and MAC086, followed by a second overlap extension PCR for addition of the attB sites with MAC087 and MAC088. Gateway cloning was performed using the attB-flanked Phatrdraft_47612 amplicon, generating the pEXP/Phatrdraft_47612 vector. A list of all primers designed and used to construct these vectors is provided (Table 4.3, end of chapter). Sanger sequencing validated proper frame for these two expression vectors.

Protein expression and purification

Several expression strains were used during this study including BL21(DE3) + pLysS, CodonPlus + RIPL, and Rosetta2 + pLysS, gifted by Mam Scherman (Protein Expression/Purification Facility, Colorado State University). Chemically competent cells for each of these strains were prepared[215]. These strains were transformed with either pHAT4/Phatrdraft_47612 or pEXP/Phatrdraft_47612. These strains were grown on LB + 100 μg mL^{-1} ampicillin + 1 % glucose until an OD600 of approximately 0.6. Then, the media was supplemented with 0.1 mM – 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma, #I5502) was provided. 1 mL aliquots were harvested immediately prior to IPTG addition and at intervals during induction to track the accumulation of recombinant Phatrdraft_47612 over time. These cultures were grown with IPTG for up to 16 h post inoculation either at room temperature or at 18 °C. Pellets were harvested by centrifugation (3220 x g, 4 °C, 15 min) and stored at -20 °C until purification. Pellets were suspended in Buffer A (20 mM Tris, pH 7.5, 300 mM NaCl), with the addition of 1 mM phenylmethylsulfonyl fluoride to limit proteolysis. Cells were kept on ice and lysed by sonication (2 min in 10 s on:off pulse intervals). These mixtures were then clarified by centrifugation (10000 x g, 4 °C, 30 min) and the supernatant was loaded onto an AKTA FPLC chromatography system maintained in a 4 °C chamber. Recombinant protein was enriched from total protein using a HisTrap HP, 5mL (GE Healthcare Life Sciences, #17-5248-01). Elution was performed using a linear gradient from 10 % – 36 % buffer B (20 mM Tris, pH 7.5, 300 mM NaCl, 0.5 M imidazole), monitoring for protein content by tracking absorbance at 280 nm. Fractions were auto-collected and analyzed by SDS-PAGE.

Results

Identifying a putative β-1,3 glucan binding protein using 2D-AE

The 2D-AE method resolved a single gel spot (Figure 4.3). 30 µg of *Phaeodactylum tricornutum* protein extract was run using the amended 2D-AE method. The amount of protein and amount of embedded carbohydrate as ligand was comparable to the previously published method.[9] Visualization of a gel spot when separating crude protein extract in the 2D-AE gel was replicated with both SYPRO-Ruby (Figure 4.3, A) and imidazole-zinc (Figure 4.3, B) staining, although the former provided a more striking migration pattern. In both cases, most protein was resolved in a diagonal line, suggesting that most diatom proteins do not interact with embedded β-glucan in the custom 2D gel.

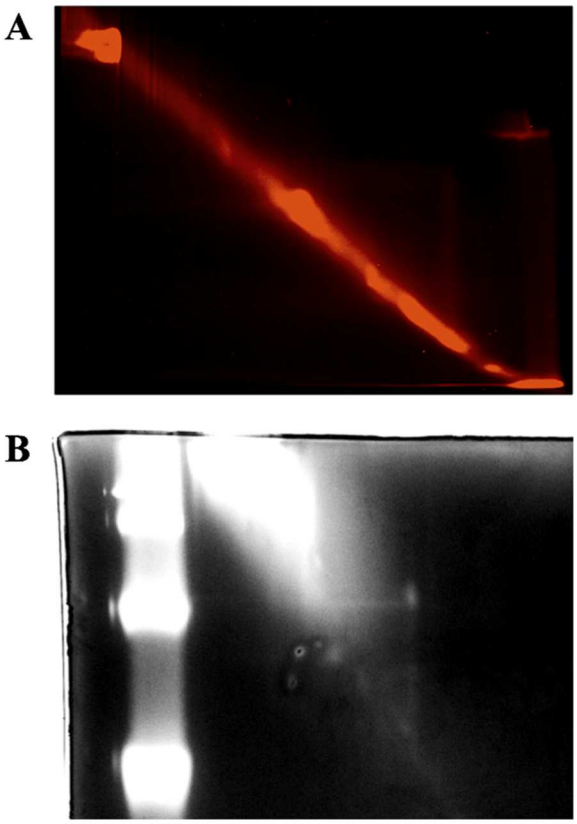

Figure 4.3: Visualizing 2D-affinity electrophoresis of *Phaeodactylum tricornutum* crude protein extract.
A gel spot is isolated from a crude protein extract using the 2D-AE approach. 30 µg of protein extract was run and visualized with SYPRO Ruby(A) or zinc-imidazole (B).

The isolated gel spot was excised and submitted for protein sequencing. Three proteins were identified in the spot (Table 4.1). Phatrdraft_47612 was identified from 13 exclusive unique peptides. Bioinformatic analysis of Phatrdraft_47612 reveals two predicted protein domains: WD-40 like and Phytase-like. Interestingly, Phatrdraft_47612 was independently identified in a parallel proteomics study in our lab (data not shown). Subsequently, characterization efforts focused on Phatrdraft_47612 due to these two independent observations of Phatrdraft_47612 in the 2D-AE gel and the proteomics study. The second protein identified in the gel spot was Phatr_46721, predicted to be a Rossman-fold NAD(P)(+)-binding protein. The third protein identified was Phatrdraft_42612, with a B-cell receptor-associated protein 31-like domain. All three proteins are annotated as hypothetical proteins.

Table 4.1: Sequencing results of excised gel spot from 2D-affinity electrophoresis. An isolated gel spot from *Phaeodactylum tricornutum* crude protein extract using the 2D-AE approach with laminarin as a ligand was excised and submitted for protein identification. Proteins in the gel spot were digested into peptides with trypsin. MS/MS spectra from these peptides enabled protein identification with Mascot and Scaffold software, summarized below.

Protein	Pfam predicted domains	Exclusive unique peptide count	Percent coverage of protein
Phatrdraft_47612	Phytase-like	13	24
Phatr_46721	NmrA-like	9	44
Phatrdraft_42612	B-cell protein 31-like	2	10

The migration of Phatrdraft_47612 in the gel spot under the 2D-AE conditions was verified with immunoblot. Proteins were transferred to a blotting membrane after protein separation using the 2D-AE method. Phatrdraft_47612 appears to co-localize with the migration of the spot (Figure 4.4) when compared to the SYPRO-Ruby visualization of total protein (Figure 4.3, A). However, migration of a control protein, such as AtpB or urease was not visualized in parallel immunoblots.

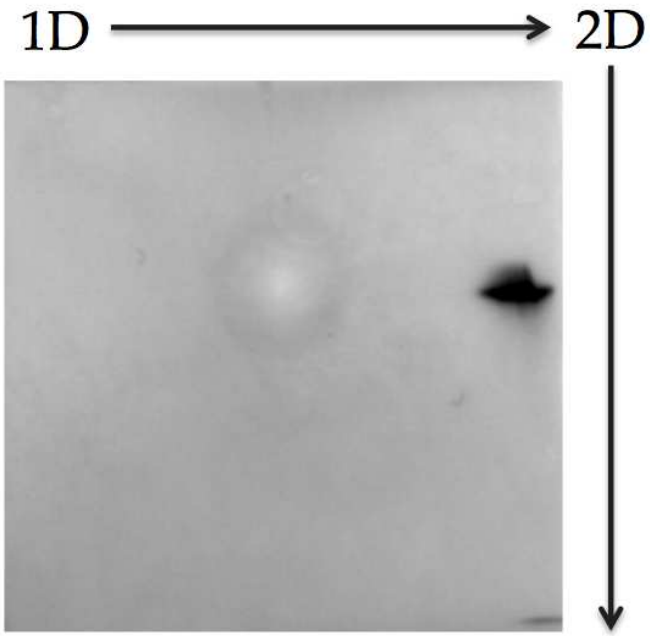

Figure 4.4: Gel spot migration of Phatrdraft_47612 in 2D-affinity electrophoresis. A custom Phatrdraft_47612 antibody was produced (Genscript). 10 µg of *Phaeodactylum tricornutum* crude protein extract was run through 2D-AE parameters as above. The gel was soaked in 0.1% SDS to permit transfer to a membrane for western blotting.

Characterizing Phatrdraft_47612 gene expression in putative knockdowns

RNAi strains putatively knocking down Phatrdraft_47612 were generated with a biolistics approach by bombarding *Phaeodactylum* with either the RNAi "A" or "B" vector. Zeocin-resistant colonies formed after approximately 14 days of growth on selective f/2 + 100 µg mL^{-1} Zeocin plates. These colonies were re-streaked on selective plates and viable colonies were transferred into liquid media without selection to screen for detrimental growth rates. RNAi lines A22 and B28 were identified by their significantly lower exponential growth rates of 0.74 d^{-1} and 0.63 d^{-1}, respectively, compared to the WT growth rate of 1.12 d^{-1} (Table 4.2). The A22 and B28 RNAi lines were further characterized for transcript accumulation and protein abundance of Phatrdraft_47612 as well as biovolume and chrysolaminarin content per cell relative to WT.

The pMAC3 RNAi vector utilizes a light-independent histone promoter[172], so I anticipated constitutively lower abundance of the Phatrdraft_47612 transcript, independent of time of day. However, Phatrdraft_47612 mRNA abundance in WT and RNAi lines varied by the time of day (Table 4.2). While there was a statistically significant difference between WT and B28 at the transition from dark to light, there was no difference between WT and either of the RNAi lines at the transition from light to dark. Both A22 and B28 lines accumulate more chrysolaminarin per cell volume, but this difference is not consistently significant on a per-cell basis (Table 4.2). Unexpectedly, an increase of Phatrdraft_47612 protein accumulation was detected in both A22 and B28 lines relative to WT (Figure 4.6). In summary, both putative RNAi lines had, relative to WT, lower growth rates, no constitutive reduction in Phatrdraft_47612 transcript abundance, and an increase of chrysolaminarin content per cell when normalized by cell volume, and surprisingly, more Phatrdraft_47612 protein accumulation.

Table 4.2: Preliminary identification and characterization of Phatrdraft_47612 RNAi lines. A preliminary screening strategy identified RNAi lines by reduced growth rate. These RNAi lines were also characterized for their transcript accumulation of Phatrdraft_47612 at dawn or dusk, cell volume, and chrysolaminarin content. n = 3, error = standard deviation, except for determining cell volume which was n = 30, error = SEM. Different letters denote statistically different groups as determined by one-way ANOVA and Tukey's HSD.

	WT	**A22**	**B28**
Growth rate (day^{-1})	1.12 ± 0.02^{A}	0.74 ± 0.02^{B}	0.63 ± 0.02^{C}
Transcript abundance (dawn)	1.00 ± 0.86^{A}	0.10 ± 0.095^{AB}	0.069 ± 0.036^{B}
Transcript abundance (dusk)	1.00 ± 0.83^{A}	1.37 ± 0.64^{A}	2.96 ± 1.41^{A}
Cell Volume (μm^3 $cell^{-1}$)	66.5 ± 4.0	48.7 ± 3.9	35.6 ± 3.8
Chrysolaminarin at dusk (pg $cell^{-1}$)	0.28 ± 0.03^{A}	0.45 ± 0.07^{B}	0.34 ± 0.07^{AB}
Chrysolaminarin at dusk (fg μm^{-3})	4.26 ± 0.46^{A}	9.24 ± 1.45^{B}	9.62 ± 2.10^{B}

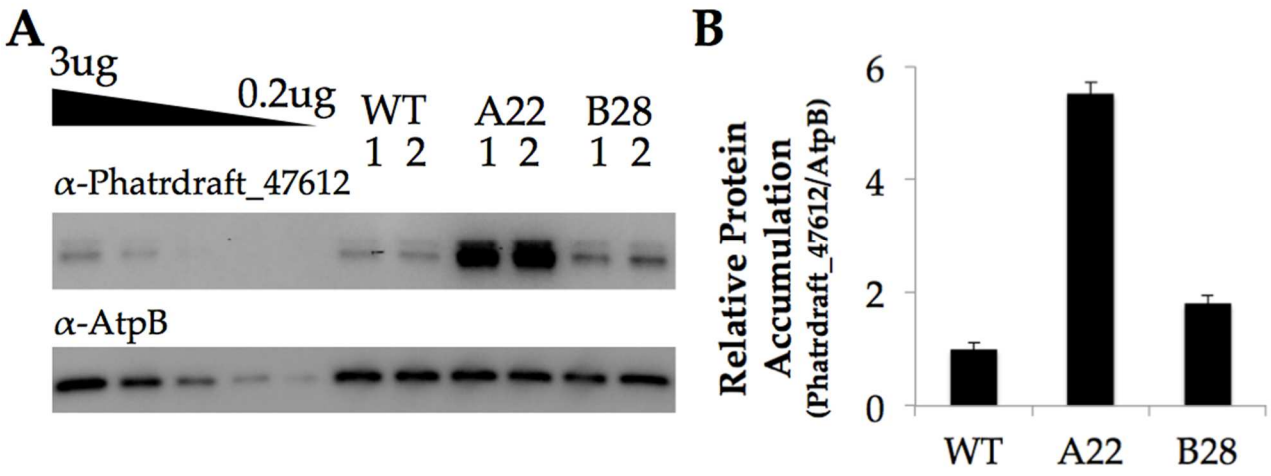

Figure 4.5: Protein quantification of Phatrdraft_47612 in putative RNAi lines. Representative western blot (A) and relative quantification (B) of Phatrdraft_47612 in WT and RNAi lines. RNAi lines appear to accumulate more, not less, of the protein relative to wild-type (n=2 biological replicates, error=range).

Induction and purification of recombinant Phatrdraft_47612

Phatrdraft_47612 was recombinantly expressed in *E. coli* and purified using FPLC. The approximate molecular weight of Phatrdraft_47612 was anticipated to be 120 kDa. Separating imidazole eluate fractions by SDS-PAGE and visualization with Imperial staining or immunoblot only resolved a faint band at 120 kDa (Figure 4.6). In addition, 4 more intense bands were visualized, all with apparent molecular weights lower than 100 kDa. This pattern remained despite altering the *E. coli* expression host and reducing the induction temperature from room temperature to 18 °C (data not shown). Therefore, I was unable to enrich Phatrdraft_47612 to an adequate purity for additional *in vitro* experiments.

Re-analysis of 2D-AE migration

The 2D-AE parameters were further tested to evaluate the consistency of the resolved gel spot. Immunoblot tracked migration of several proteins after 2D-AE migration, and are representative of 2 technical replicates. Previously followed 2D-AE parameters were included as a reference, and, as before, a gel spot was observed when anti-Phatrdraft_47612 was used as a

primary antibody stain (Figure 4.7, A). This same experiment was repeated but with a 2D-AE gel that omitted the β-1,3 ligand, and substantial vertical streaking was observed (Figure 4.7, B). I also investigated the migration of a protein not predicted to interact with β-1,3 glucans, urease. After 2D-AE separation, some proportion of urease exhibited differential migration (Figure 4.7, C). Finally, when I investigated the migration of 5 µg total protein, primary staining with anti-Phatrdraft_47612 revealed a single bright spot, which does not correspond with the top right-most gel spot (Figure 4.7, D) when compared to the reference (Figure 4.7, A). In summary, with the 2D-AE method, I observed 1) differential migration of Phatrdraft_47612 in the absence of β-1,3 glucan, and 2) differential migration of urease, a protein not expected to interact with β-1,3 glucan.

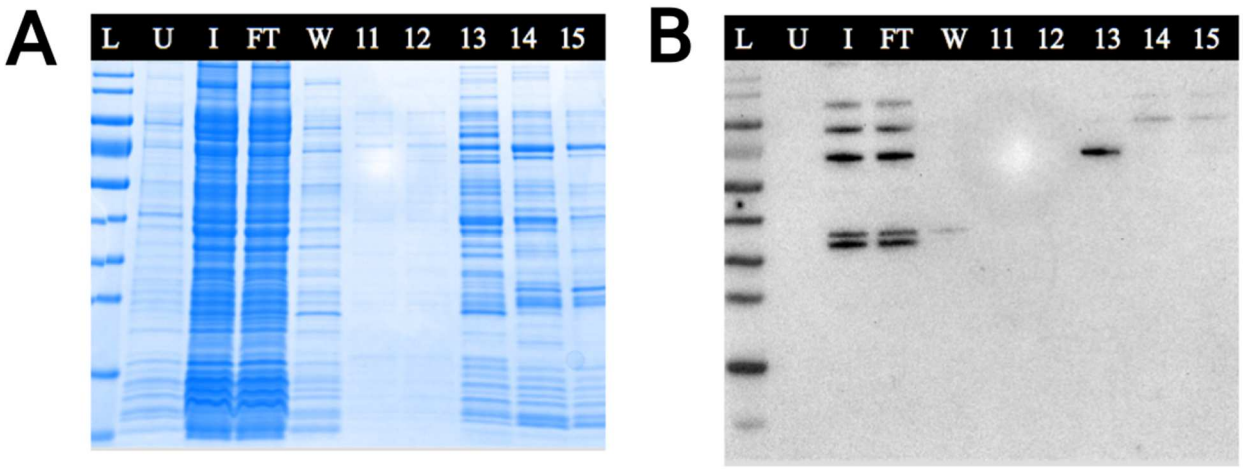

Figure 4.6: Recombinant protein expression of Phatrdraft_47612.
Both the Imperial Total protein stain (A) and immunoblot using the anti-Phatrdraft_47612 primary antibody (B) are shown. L = PageRuler Prestained Protein Ladder (10 kDa – 180 kDa), U = uninduced cells, I = induced cells, FT = flow-through, W = wash, # = fraction number from FPLC auto-fraction collector.

Discussion

I attempted to identify proteins from the *Phaeodactylum* proteome based on their proposed binding to β-1,3 glucan. I hypothesized that I could leverage the biochemical affinity of proposed, hypothetical β-1,3 glucan binding proteins from the *Phaeodactylum* proteome using

the 2D-AE method to identify them. I adapted the 2D-AE method to address this question by embedding the β-1,3 glucan laminarin as the ligand of interest. While I did observe a gel spot when I separated *Phaeodactylum* crude protein extract with the 2D-AE method, two assumptions in my experimental design were invalidated, and are discussed in additional detail below.

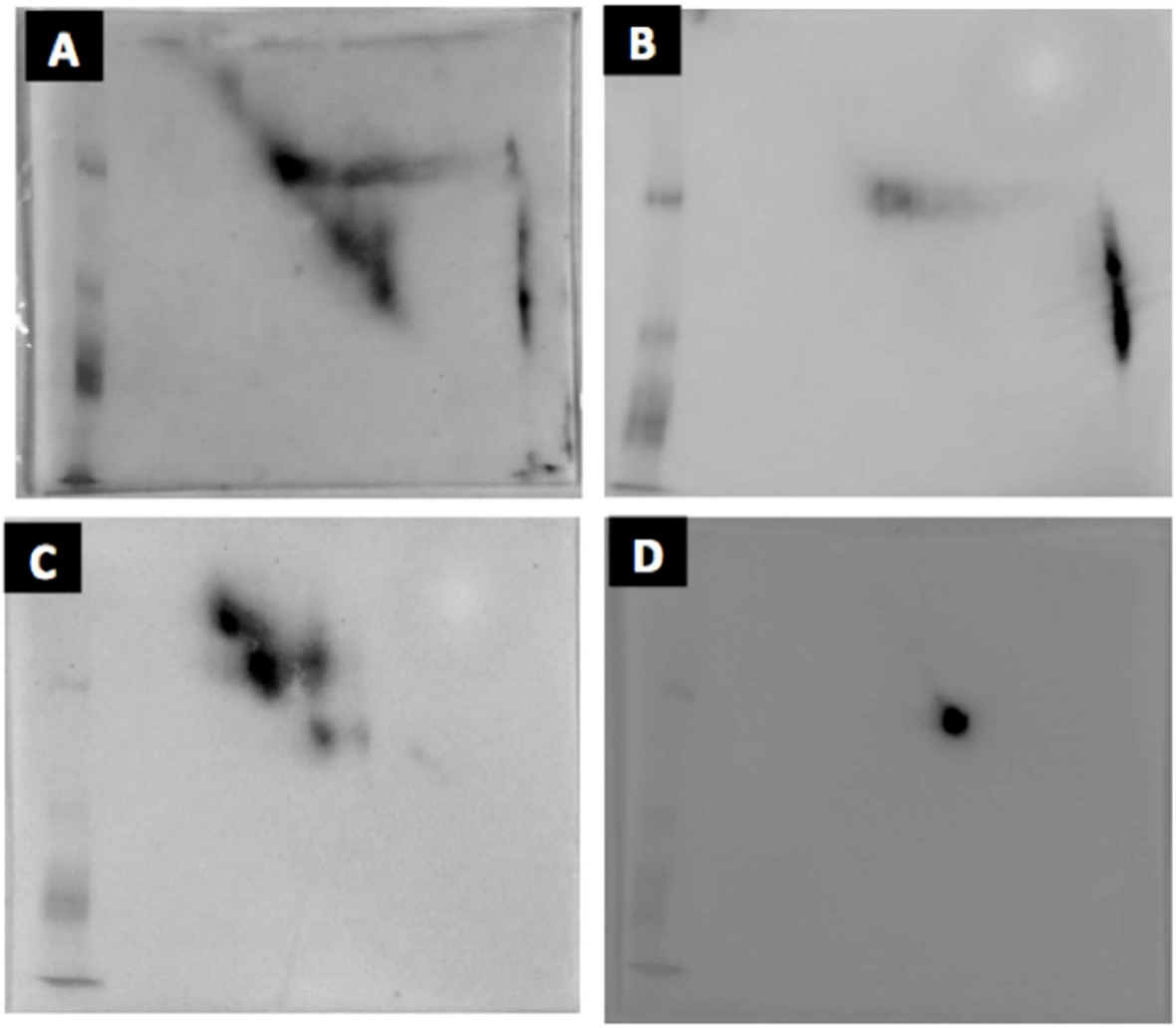

Figure 4.7: Analyzing 2D-affinity electrophoresis gel spot migration of Phatrdraft_47612 and urease with varying protein concentrations.
Several 2D-AE conditions were tested to challenge our assumption that vertical streaking was the result of protein affinity towards the embedded laminarin, and protein migrations were visualized using Western blots. The blots correspond to the following amounts of loaded protein, laminarin concentration in the gel, and primary antibody stain: A) 25 μg total protein, 2 μg L^{-1} laminarin, Phatrdraft_47612; B) 25 μg total protein, 0 μg L^{-1} laminarin, Phatrdraft_47612; C) 25 μg total protein, 2 μg L^{-1} laminarin, urease; and D) 5 μg total protein, 2 μg L^{-1} laminarin, Phatrdraft_47612.

Differential migration in the 2D-AE gel is not necessarily from affinity

A core assumption of the 2D-AE method is that that differential migration in the vertical dimension of the 2D gel is due to protein interacting with the embedded ligand. While I resolved differential migration of Phatrdraft_47612 in a 2D-AE gel (Figure 4.3; Figure 4.7, A), a significant vertical streak in the 2nd dimension was also observed for Phatrdraft_47612 in a 2D-AE gel without laminarin (Figure 4.7, B). Urease, a target unlikely to interact β-1,3 glucan, also appeared to have differential migration in the 2nd dimension (Figure 4.7, C). Together, these results suggest that differential migration in the 2nd dimension can occur in the absence of β-1,3 glucan and for enzymes that do not have a biochemical role with β-1,3 glucan, weakening a fundamental assumption of the 2D-AE approach.

However, reference of the migration of these spots to an expected diagonal migration are lacking in my results. This could have been resolved by loading a small aliquot of PageRuler ladder to the experimental sample to provide a reference for such a migration using the 2D-AE method. PageRuler proteins are expected to migrate in a diagonal pattern, and could be visualized in parallel with the immunoblot. This lack of a reference between immunoblots limits the comparison of protein migration, especially for Phatrdraft_47612, between immunoblots. Additionally, lack of consistent control immunoblots (e.g. AtpB or urease) further limits our ability to interpret the relative migration of Phatrdraft_47612. Therefore, any conclusions about protein migration using the 2D-AE approach described in this chapter are of limited confidence.

However, investigating troubleshooting other 2D-based protein separation approaches, it is plausible that the unexpected vertical migration of the 2D gel was caused by protein overloading. Vertical streaking has been shown to be a common problem associated with protein overloading in other 2D gel strategies, such as IEF/SDS-PAGE[216]. Loading 30 µg of protein

may not have been appropriate for the mini-gel scale used in this approach. Furthermore, loading less protein in the 1st dimension yielded a western blot without vertical streaking of Phatrdraft_47612 (Figure 4.7, D). Because the first principle of the 2D-AE methodology was violated by demonstrating vertical streaking of Phatrdraft_47612 in the absence of laminarin, I cannot conclude from the 2D-AE approach that Phatrdraft_47612 interacts with β-1,3 glucan.

Screening for reduced growth rates might not select for chrysolaminarin metabolism mutants

Second, I assumed that disrupting chrysolaminarin metabolism would be disruptive to cellular fitness. Starch and chrysolaminarin are major sinks of photoassimilate and disrupting this pool of carbon and energy may be detrimental to respiration and maintaining cellular metabolism. This assumption was based on the report of reduced fitness in a starch excess *Arabidopsis* mutant[58]. Similarly, *Chlamydomonas sta7-10* mutants exhibit lower cell densities when grown either heterotrophically or photoautotrophically[33]. However, I have shown in Chapter 3 that CRISPR/Cas9 diatom mutants that accumulated less chrysolaminarin than their transformation control did not have a significantly lower growth rate relative to their transformation control. Thus, the biological rationale underpinning the RNAi selection strategy was not appropriate, since our primary selection relied on identifying RNAi lines with lower growth rates.

Therefore, I cannot discard the possibility that all the phenotypes observed (Table 4.3) are the result of pleotropic effects from transformation. Biolistics-based transformation relies on random nuclear integration of a vector of interest after bombardment. In plants, vectors delievered via biolistics have been shown to be co-transformed in nuclear and plastid genomes[217]. This indicates that biolistics yields nonspecific insertional mutagenesis. Random insertion of the RNAi vector could lead to unexpected, unintended phenotypes since the

transformation site is not controlled[218]. Moreover, a recent study has shown than most RNAi strains are not effective knockdowns[219]. However, Bielinski et al. (2017) used a different knockdown vector which could affect these results. Since my RNAi lines did not exhibit consistent knockdown patterns, it is even more likely that the phenotpyes observed in these lines are due to random, positional effects from random nuclear integration of the knockdown vector, rather than a knockdown phenotype.

Speculating on the role of Phatrdraft_47612

The central rationale for this study was to biochemically identify *Phaeodactylum* proteins that interact with β-1,3 glucans. Given the key limitations of our experimental approach outlined above, it is unlikely that Phatrdraft_47612 plays a role in chrysolaminarin biology. However, the RNAi lines appeared to accumulate more chrysolaminarin at the end of the light period, especially when normalized by cell size (Table 4.3). Despite the evidence to the contrary as discussed above, let us assume that Phatrdraft_47612 plays a role in chrysolaminarin biology. I propose two speculative explanations:

First, Phatrdraft_47612 contains two putative protein domains: WD40-like and Phytase-like. WD-40-like domains are best characterized for their protein-protein interactions as protein scaffolds[220]. Phytases are structurally similar to phosphorylases, and have been bioprospected for their applied use as an enzymatic preparation step of animal feed to increase biologically available phosphorous for animals[221]. If WD-40 is capable of binding chrysolaminarin and that the phytase-like domain is capable of dephosphorylation, it could be suggested that Phatrdraft_47612 serves as a β-1,3 glucan phosphatase. Glucan phosphatases have been characterized in vertebrates[222, 223] and plants[74, 75, 224, 225]. Glucan phosphatases bind and dephosphorylate their substrate, consequently affecting glucan catabolism. Altered

expression of these non-orthologous glucan phosphatases leads to a progressive accumulation of storage glucan, in agreement with our chrysolaminarin content measurements (Table 4.3). However, no evidence has been shown to suggest that WD-40 domains interact with carbohydrates[226], effectively disarming this interpretation.

Second, it is also possible to speculate that Phatrdraft_47612 plays a regulatory role as a posttranslational activator or repressor. Following the same assumptions as above, Phatrdraft_47612 might bind other proteins that are actively involved in chrysolaminarin biology. The increased accumulation of Phatrdraft_47612 protein in our RNAi lines may be inhibiting chrysolaminarin catabolism or activating chrysolaminarin synthesis. Unfortunately, the inability to prepare recombinant Phatrdraft_47612 limited the opportunity to test these hypotheses: *in vitro* experiments could determine dephosphorylation rates with p-nitrophenyl phosphate [227] and binding affinity to β-1,3 glucan[208].

Suggestions for future β-1,3 glucan-binding studies

While this research relied on some technical and biological assumptions that might not have been appropriate, the biochemical approach underpinning this chapter still has the potential to expand our understanding of chrysolaminarin biology. Therefore, I have outlined three suggestions to consider for future studies investigating β-1,3 glucan binding proteins from diatoms.

First, the protein sample prepared should be from a culture adapted to a light:dark growth regime. The chrysolaminarin quantification method outlined in Chapter 2 had not yet been developed at the onset of this study. I have since determined that cells grown in continuous light contain less chrysolaminarin per cell than cells at the end of the dark period in a light:dark growth regime. End of day and midday appear to be maxima for nucleotide sugar synthesis

activities in microalgae[228], a prerequisite for polysaccharide synthesis. Therefore, I would specifically target both the middle and the end of the light period for protein extraction to identify chrysolaminarin synthesis related proteins. The rise of *in vitro* enzymatic activity during that time of day does not necessarily correlate to increased protein accumulation but relevant enzymes for this aspect of metabolism should be present.

Second, identifying β-1,3 glucan binding proteins should be performed by exploting the chemistry of glycoprotein immobilization kits (e.g.: GlycoLink Immobilization Kit) to covalently bind chrysolaminarin. A benefit of this approach over the 2D-AE is that there would be a greater probability of carbohydrate-protein interaction in a bead and protein extract slurry, which may mean dramatically shorter co-incubation times than the 2D-AE approach, which required electrophoresis under native conditions for as many as 8 h. However, a limitation of this strategy could be bead fouling as any active degradation enzymes, such as endo- or exo-glucosidases, might degrade the bead-bound chrysolaminarin.

Finally, the molecular toolbox for diatom genetics has advanced significantly in the past 4 years, specifically for gene editing. Diatom CRISPR/Cas9 strategies[101] have been successfully implemented in our lab, and this technology could be applied to generate mutants more objectively using PCR genotyping, rather than phenotyping RNAi lines. Moreover, the roles of specific domains of a protein may be explored through these approaches. Fusion proteins are a common occurrence in the diatom genome and selective deletion through properly assembled homology donor vectors could enable careful dissection of annotated domains encoded by the coding sequence. In this manner, protein domains could be functionally characterized, such as assaying for *in vitro* binding affinity[208].

Conclusions

I set out to better understand chrysolaminarin metabolism by identifying novel targets with a protein biochemistry-focused approach: identifying novel diatom proteins for their ability to bind a β-1,3 glucan, laminarin. Capacity for binding to β-1,3 glucans might imply a role of the enzyme in β-1,3 glucan metabolism, as this is a consistent pattern in other important starch and glycogen-related proteins. In the absence of bioinformatically predicted targets, I adapted the 2D-AE method to attempt to identify β-1,3-glucan binding proteins from a diatom protein extract. While I was encouraged by early experiments, the observed gel spot is likely an artifact of protein overloading. Other critical flaws in our experimental design and assumptions were also noted and discussed. I maintain that identifying novel diatom proteins on their basis of binding to β-1,3 glucan is still a promising strategy to expand our understanding of chrysolaminarin metabolism and diatom biology. Towards this end, several alternative strategies have been provided for consideration in future research efforts.

Table 4.3: List of primers used in this study.
Primers for qRT-PCR were designed in primer-BLAST.

Primer name	Description	Sequence
MAC043	47612kdA-fwd	CCAACAGCACTCTTGTCAAC
MAC044	47612kdA-rev	CACCGTACTCAGCAATACCT
MAC045	47612kdB-fwd	TACGTCTTTAACGTCGACCC
MAC046	47612kdB-rev	GGGCTGAAAACGGTATAGGA
MAC055	TBP-qRT-fwd	AGAAGGAAATGGACCTGGCG
MAC056	TBP-qRT-rev	AACCCCGGGTACCGTAGTTA
MAC067	47612qRT-fwd	GACGACATGACGGTCAACCT
MAC068	47612qRT-rev	CACTGTACCTGCACCTTCGT
MAC085	47612-recomb1-fwd	ATGAATCTTCGTTGTATCCTTCCGT
MAC086	47612-recomb1-rev	TCACAATAAAGCAAAGAAGCCGAAG
MAC087	47612-recomb2-fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAATCTTC
MAC088	47612-recomb2-rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTCACAATAAAG
MAC107	47612 pHAT4 fwd	CCCCCCATGGGTATGAATCTTC
MAC108	47612 pHAT4 rev	CCCCGGGAATTCTCACAATAAA

CHAPTER 5: CONCLUSIONS

The research documented in the previous chapters sets out to explore chrysolaminarin biology in the model diatom *Phaeodactylum tricornutum*. There was an intentional emphasis on the use and development of biochemistry methods to facilitate novel discoveries in chrysolaminarin metabolism. Reliable quantification of chrysolaminarin represented a technical challenge in the field of diatom biology, which I addressed by developing and testing a new protocol that accurately quantifies soluble, non-reducing carbohydrates, which contain chrysolaminarin. UDP-glucose formation represents the first metabolic step in chrysolaminarin biosynthesis, and I identified a novel gene (Phatrdraft_54493) encoding a protein capable of synthesizing UDP-glucose *in vitro*. I also characterized a suite of UDP-glucose diphosphorylase mutants by disrupting Phatrdraft_50444, Phatrdraft_23639, and Phatrdraft_54493 with a CRISPR/Cas9 approach to investigate their potential contribution to chrysolaminarin accumulation and other metabolic routes that utilize UDP-glucose. Finally, I attempted to identify novel chrysolaminarin-related proteins based on their binding affinity to β-glucan. While Phatrdraft_47612 represented a false positive due to protein overloading, I have outlined strategies for future efforts to identify β-glucan binding proteins from diatoms. The significance of each chapter is briefly outlined below.

First, I developed an accurate and reliable chrysolaminarin quantification protocol, which is reported in Chapter 2. While phenol-sulfuric acid assays for total carbohydrate quantification have been the gold standard in diatom carbohydrate studies, it is limited by its accuracy when quantifying complex sugar mixtures, a problem for the dynamic carbohydrate profiles from cellular extracts. The method described in Chapter 2 is a reliable means to quantify

chrysolaminarin from cell extracts. This protocol enabled profiling of carbon partitioning in *Phaeodactylum* during nitrogen starvation. Unlike other diatoms, I found that *Phaeodactylum* does not expand its chrysolaminarin reserves during nitrogen starvation. This method also enables reliable identification of chrysolaminarin excess or deplete mutants for phenotyping studies.

Accurate quantification of carbon partitioning in diatoms is an essential prerequisite to develop rational bioengineering strategies such as the synthesis of triacylglycerols for biofuels. Findings in Chapter 2 indicated that manipulating chrysolaminarin content might not be an impactful strategy to enhance triacylglycerol content during nitrogen starvation, as chrysolaminarin represents only about 5.5 % of cellular total organic carbon under such conditions[124]. Moreover, it has been shown that protein, and not triacylglycerol or carbohydrates, may represent the largest pool of carbon in *Phaeodactylum tricornutum*[154]. Chapter 2 complements this previous research, suggesting that protein turnover might represent an unexpected reservoir of carbon for new bioengineering strategies in diatoms to maximize production of a metabolite of interest.

Second, I found that the *ugp1* (Phatrdraft_50444) gene product is the major contributor to chrysolaminarin biosynthesis. This conclusion agrees assumptions made in the literature [7, 8], which also revealed an increase in neutral storage lipid accumulation. However, these studies failed to investigate not only the alternative roles UDP-glucose plays in metabolism but also the potential contribution of other UDP-glucose synthesizing enzymes. Therefore, I set out to test the potential roles of these enzymes more comprehensively in Chapter 3. I challenged the assumption that bioinformatics predictions alone were sufficient to identify all UDP-glucose synthesis activity, and discovered a novel gene, *ugp3* (Phatrdraft_54493) encoding for a protein

with potential UGPase activity using a functional screen in *E. coli.* Phenotyping three putative targets (Phatrdraft_50444, Phatrdraft_23639, and Phatrdraft_54493) revealed distinct patterns for each CRISPR/Cas9 knockout. Phatrdraft_50444 knockouts were the only mutants that conferred a carbon partitioning phenotype of less chrysolaminarin and more neutral lipids. By discovering that neither knockouts of *ugp2* or *ugp3* share this phenotypic pattern, I can confidently conclude that Phatrdraft_50444 is the major contributor to chrysolaminarin metabolism.

This research has improved on the initial model of chrysolaminarin metabolism (Figure 1.1) through several observations, highlighted below (Figure 5.1). Changes in chrysolaminarin accumulation at dusk were small or not statistically significant for *ugp2* and *ugp3* disrupted mutants, relative to their transformation controls. Only the *ugp1*-disrupted mutants conferred a significant decrease in their accumulated chrysolaminarin reserves at the end of the light period.

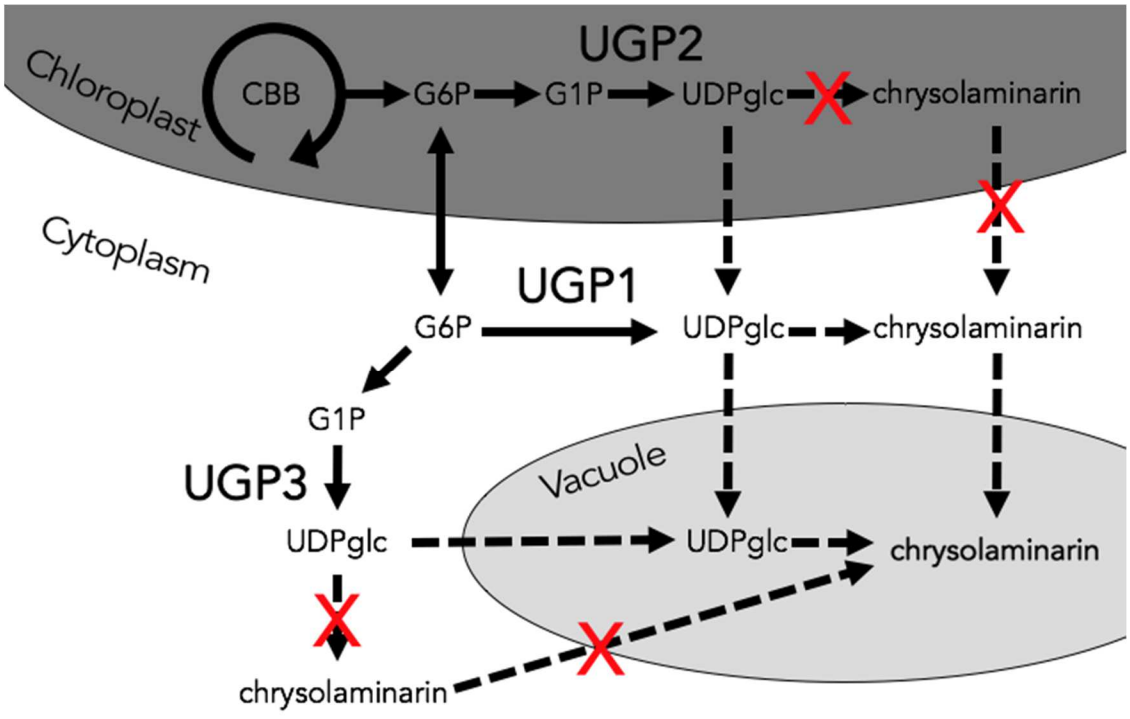

Figure 5.1: Updated model of UDP-glucose synthesizing enzymes and their contribution to chrysolaminarin accumulation at the end of the light period
Carbon is fixed by the Calvin-Bessham-Bassam (CBB) cycle in the chloroplast. Glucose 6-phosphate (G6P) and glucose 1-phosphate (G1P) are metabolites downstream of this cycle and

may be acted on by three potential enzymes to produce UDP-glucose (UDPglc): UGP1, UGP2, or UGP3. Therefore, enzymatic activity from any of these three proteins may contribute to the accumulation of cellular chrysolaminarin content. However, knocking out *ugp2* or *ugp3* did not dramatically affect chrysolaminarin content at the end of the light period, so their contribution to chrysolaminarin synthesis can be discounted, visualized by a red "X". Only knockouts of *ugp1* exhibited a carbon repartitioning phenotype where triacylglycerol content significantly increased and chrysolaminarin content significantly decreased. The metabolic significance of UDPglc produced by UGP2 or UGP3 merits further investigation.

Investigating targets using a gene-by-gene approach is essential for and complementary to systems biology studies and genome-wide modeling efforts. There exist ambiguities in the understanding of chrysolaminarin metabolism, which in turn has led to the assignment of multiple UGPases in modeling chrysolaminarin synthesis[54, 96]. A reductionist approach, such as a reverse genetics approach by knocking out each gene individually and characterizing their phenotype, is a thorough approach to support or reject the physiological roles of each gene product in parallel. Now that Chapter 3 suggests that the *ugp1* gene product (and not the other potential targets, *ugp2* or *ugp3*), this knowledge might be further manipulated for additional studies. For example, it is possible that the *ugp1* gene product is post-translationally regulated. ADP-glucose diphosphorylases from several plants have been shown to be redox regulated, partially through light-dependent activation by thioredoxins[229]. Additionally, yeast 2-hybrid approaches could discover other proteins interacting and potentially regulated UGP1. Findings from this chapter enables future studies to explore the biology of UGP1 in chrysolaminarin metabolism.

Third, I explored a strategy to identify novel β-1,3 glucan binding proteins from *Phaeodactylum* soluble protein extracts. Protein-substrate investigations have proven useful in the study of biosilica deposition[36] and starch metabolism[9]. I reasoned that a similar approach could benefit studies in chrysolaminarin metabolism. To date, all chrysolaminarin-related targets have been identified via bioinformatics. I adapted a biochemical approach, 2D-affinity

electrophoresis, that has resolved starch-binding proteins from a crude protein extract [9]. While early experiments looked promising, I was unable to conclude that the repeatable gel spot I observed was due to an affinity of Phatrdraft_47612 to the embedded carbohydrate laminarin. I also did not consider the potential migration of proteins identified in the gel spot other than Phatrdraft_47612, but this selection was primarily driven by the independent observation of this protein in a shotgun proteomics physiology study. However, the biochemical strategy to isolate proteins based on their affinity to carbohydrates is still a reasonable approach to identify novel chrysolaminarin-related proteins. Therefore, I outlined possible strategies to consider for future research, such as replacing the in gel approach with a bead based approach with covalently bound β-glucan.

Chapter 4 highlights my early efforts to identify novel proteins involved in chrysolaminarin biosynthesis. Chronologically, Chapter 4 took place before either Chapter 2 or 3. With the information from Chapter 2 and Chapter 3, other strategies emerge in the effort to discover novel chrysolaminarin-related proteins. Namely, the ability to quantify chrysolaminarin could enable an approach to isolate and purify proteins capable of synthesizing chrysolaminarin *in vitro*. Protein fractions separated by non-denaturing FPLC methods could be supplemented with excess UDP-glucose and assayed for the accumulation of soluble, non-reducing sugars. Fractions which exhibit significant accumulation of soluble, non-reducing sugars could then be analyzed by SDS-PAGE and protein sequencing. A control set of experiments could be run in parallel with ADP-glucose. Also, exploiting the knowledge that UGP1 is involved in the biosynthesis of chrysolaminarin, novel chrysolaminarin-related proteins might be identified based on protein-protein interactions with UGP1, such as a yeast-two hybrid assay. Nevertheless, the original investigative effort associated with Chapter 4 served as an important part of my

graduate studies as it provided an opportunity for me to struggle with experimental roadblocks and how to better think about and address such problems in future efforts.

This research has also uncovered potential opportunities for new avenues of research. I have generated mutants that accumulate significantly less chrysolaminarin. These mutants could be further characterized using a systems biology approach, such as transcriptomics. Investigating these changes could identify additional metabolic bottlenecks in diatom carbon metabolism. The knockouts generated in Chapter 3 also set a foundation for future studies in diatom sulfolipid biosynthesis and characterization of a homozygous UDP-sugar phosphorylase knockout, which has not been previously possible in plants [109]. Finally, alternative substrates (chrysolaminarin, UDP-glucose) could be used in place of laminarin to identify novel diatom protein-carbohydrate interactions.

In summary, these studies have contributed to our understanding of chrysolaminarin metabolism, the unique storage polysaccharide of diatoms. Developing a chrysolaminarin quantification tool enabled us to accurately explore carbon partitioning during nutrient starvation as well as phenotyping cellular chrysolaminarin content in CRISPR/Cas9 gene knockouts. Continued exploration of chrysolaminarin metabolism, whether through protein-carbohydrate interactions or otherwise, are critical contributions to the understanding of diatom carbon metabolism. Chrysolaminarin-like polysaccharides are synthesized by a diversity of species within the stramenopiles, and experimental evidence for a given diatom gene in carbon metabolism helps identify homologs and map out pathways for newly sequenced SAR genomes, such as recently sequenced kelp genomes[230]. Moreover, a better understanding of β-1,3 metabolism enables better bioengineering strategies that exploit diatoms in an industrial setting, such as the production of lipids for biodiesel (triacylglycerols) or food supplements (ω-3

essential fatty acids; eicosapentaenoic acid and *docosahexaenoic acid*). Diatoms are photosynthetically productive and broadly distributed over the globe, leading to their significant contribution to global silica cycles. By similar reasoning, chrysolaminarin represents a major polysaccharide species on this planet, and while this research made a clear contribution, much of chrysolaminarin biology remains to be understood.

REFERENCES

[1] S. Van Wychen, W. Long, S.K. Black, L.M.L. Laurens, MBTH: A novel approach to rapid, spectrophotometric quantitation of total algal carbohydrates, Analytical Biochemistry, 518 (2017) 90-93.

[2] S. Van Wychen, L.M. Laurens, Determination of total carbohydrates in algal biomass, 303 (2013) 275-3000.

[3] E. Granum, K. Ståle, S.M. Myklestad, Cellular and extracellular production of carbohydrates and amino acids by the marine diatom Skeletonema costatum: diel variations and effects of N depletion, Marine Ecology Progress Series, 242 (2002) 83-94.

[4] B. Remes, D. Elseviers, Adenosine 5'-triphosphate leakage does not cause abortive infection of bacteriophage T7 in male Escherichia coli, Journal of Bacteriology, 143 (1980) 1054-1056.

[5] S. Damerow, A.C. Lamerz, T. Haselhorst, J. Fuhring, P. Zarnovican, M. von Itzstein, F.H. Routier, Leishmania UDP-sugar Pyrophosphorylase: The missing link in galactose salvage?, Journal of Biological Chemistry, 285 (2010) 878-887.

[6] Y. Okazaki, M. Shimojima, Y. Sawada, K. Toyooka, T. Narisawa, K. Mochida, H. Tanaka, F. Matsuda, A. Hirai, M.Y. Hirai, H. Ohta, K. Saito, A chloroplastic UDP-glucose pyrophosphorylase from Arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis, The Plant Cell, 21 (2009) 892-909.

[7] F. Daboussi, S. Leduc, A. Maréchal, G. Dubois, V. Guyot, C. Perez-Michaut, A. Amato, A. Falciatore, A. Juillerat, M. Beurdeley, D.F. Voytas, L. Cavarec, P. Duchateau, Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology, Nature Communications, 5 (2014).

[8] B.-H. Zhu, H.-P. Shi, G.-P. Yang, N.-N. Lv, M. Yang, K.-H. Pan, Silencing UDP-glucose pyrophosphorylase gene in Phaeodactylum tricornutum affects carbon allocation, New Biotechnology, 33 (2015) 237-244.

[9] B. Kosar-Hashemi, J.A. Irwin, J. Higgins, S. Rahman, M.K. Morell, Isolation, identification and characterisation of starch-interacting proteins by 2-D affinity electrophoresis, Electrophoresis, 27 (2006) 1832-1839.

[10] R. Massana, J. del Campo, M.E. Sieracki, S. Audic, R. Logares, Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles, ISME Journal, 8 (2014) 854-866.

[11] B.M. Tyler, S. Tripathy, X. Zhang, P. Dehal, R.H.Y. Jiang, A. Aerts, F.D. Arredondo, L. Baxter, D. Bensasson, J.L. Beynon, J. Chapman, C.M.B. Damasceno, A.E. Dorrance, D. Dou, A.W. Dickerman, I.L. Dubchak, M. Garbelotto, M. Gijzen, S.G. Gordon, F. Govers, N.J. Grunwald, W. Huang, K.L. Ivors, R.W. Jones, S. Kamoun, K. Krampis, K.H. Lamour, M.-K.

Lee, W.H. McDonald, M. Medina, H.J.G. Meijer, E.K. Nordberg, D.J. Maclean, M.D. Ospina-Giraldo, P.F. Morris, V. Phuntumart, N.H. Putnam, S. Rash, J.K.C. Rose, Y. Sakihama, A.A. Salamov, A. Savidor, C.F. Scheuring, B.M. Smith, B.W.S. Sobral, A. Terry, T.A. Torto-Alalibo, J. Win, Z. Xu, H. Zhang, I.V. Grigoriev, D.S. Rokhsar, J.L. Boore, Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis, Science, 313 (2006) 1261-1266.

[12] P.J. Keeling, The number, speed, and impact of plastid endosymbioses in eukaryotic evolution, Annual Review of Plant Biology, 64 (2013) 583-607.

[13] J.M. Archibald, The puzzle of plastid evolution, Current Biology, 19 (2009).

[14] L.K. Medlin, Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence, Phycologia, 55 (2016) 79-103.

[15] P.A. Sims, D.G. Mann, L.K. Medlin, Evolution of the diatoms: insights from fossil, biological and molecular data, Phycologia, 45 (2006) 361-402.

[16] C. Bowler, A. Vardi, A.E. Allen, Oceanographic and Biogeochemical Insights from Diatom Genomes, Annual Review of Marine Science, 2 (2010) 333-365.

[17] A.J. Alverson, B. Beszteri, M.L. Julius, E.C. Theriot, The model marine diatom Thalassiosira pseudonana likely descended from a freshwater ancestor in the genus Cyclotella, BMC Evolutionary Biology, 11 (2011) 125.

[18] J.R. Johansen, Cryptogamic crusts of semiarid and arid lands of North America, Journal of Phycology, 29 (1993) 140-147.

[19] T. Van Oijen, M.A. Van Leeuwe, W.W.C. Gieskes, H.J.W. De Baar, Effects of iron limitation on photosynthesis and carbohydrate metabolism in the Antarctic diatom Chaetoceros brevis (Bacillariophyceae), Eur J Phycol, 39 (2004) 161-171.

[20] C. Von Quillfeldt, Common diatom species in Arctic spring blooms: their distribution and abundance, Botanica Marina, 43 (2000) 499-516.

[21] R.J. Geider, E.H. Delucia, P.G. Falkowski, A.C. Finzi, J.P. Grime, J. Grace, T.M. Kana, J. La Roche, S.P. Long, B.A. Osborne, T. Platt, I.C. Prentice, J.A. Raven, W.H. Schlesinger, V. Smetacek, V. Stuart, S. Sathyendranath, R.B. Thomas, T.C. Vogelmann, P. Williams, F.I. Woodward, Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats, Global Change Biology, 7 (2001) 849-882.

[22] L. Legendre, The significance of microalgal blooms for fisheries and for the export of particulate organic carbon in oceans, Journal of Plankton Research, 12 (1990) 681-699.

[23] P. Treguer, D.M. Nelson, A.J. Van Bennekom, D.J. DeMaster, A. Leynaert, B. Queguiner, The silica balance in the world ocean: A reestimate, Science, 268 (1995) 375.

[24] D.M. Nelson, P. Tréguer, M.A. Brzezinski, A. Leynaert, B. Quéguiner, Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation, Global Biogeochemical Cycles, 9 (1995) 359-372.

[25] Y. Wang, J. Cai, Y. Jiang, X. Jiang, D. Zhang, Preparation of biosilica structures from frustules of diatoms and their applications: current state and perspectives, Appl Microbiol Biotechnol, 97 (2013) 453-460.

[26] M.A. Borowitzka, Microalgae for aquaculture: Opportunities and constraints, J Appl Phycol, 9 (1997) 393.

[27] B. Borak, D.R. Ort, J.J. Burbaum, Energy and carbon accounting to compare bioenergy crops, Current Opinion in Biotechnology, 24 (2013) 369-375.

[28] J. Fargione, J. Hill, D. Tilman, S. Polasky, P. Hawthorne, Land Clearing and the Biofuel Carbon Debt, Science, 319 (2008) 1235-1238.

[29] T.P. Durrett, C. Benning, J. Ohlrogge, Plant triacylglycerols as feedstocks for the production of biofuels, The Plant Journal, 54 (2008) 593-607.

[30] J. Sheehan, T. Dunahay, J. Benemann, P. Roessler, A look back at the US Department of Energy's Aquatic Species Program: Biodiesel from algae, 328 (1998).

[31] G.C. Dismukes, D. Carrieri, N. Bennette, G.M. Ananyev, M.C. Posewitz, Aquatic phototrophs: efficient alternatives to land-based crops for biofuels, Current Opinion in Biotechnology, 19 (2008) 235-240.

[32] K.M. Weyer, D.R. Bush, A. Darzins, B.D. Willson, Theoretical Maximum Algal Oil Production, BioEnergy Research, 3 (2010) 204-213.

[33] V.H. Work, R. Radakovits, R.E. Jinkerson, J.E. Meuser, L.G. Elliott, D.J. Vinyard, L.M.L. Laurens, G.C. Dismukes, M.C. Posewitz, Increased Lipid Accumulation in the Chlamydomonas reinhardtii sta7-10 Starchless Isoamylase Mutant and Increased Carbohydrate Synthesis in Complemented Strains, Eukaryotic Cell, 9 (2010) 1251-1261.

[34] C. Wilhelm, C. Büchel, J. Fisahn, R. Goss, T. Jakob, J. LaRoche, J. Lavaud, M. Lohr, U. Riebesell, K. Stehfest, K. Valentin, P.G. Kroth, The Regulation of Carbon and Nutrient Assimilation in Diatoms is Significantly Different from Green Algae, Protist, 157 (2006) 91-124.

[35] R.E. Hecky, K. Mopper, P. Kilham, E.T. Degens, The amino acid and sugar composition of diatom cell-walls, Marine Biology, 19 (1973) 323-331.

[36] N. Kröger, G. Lehmann, R. Rachel, M. Sumper, Characterization of a 200-kDa Diatom Protein that is Specifically Associated with a Silica-Based Substructure of the Cell Wall, European Journal of Biochemistry, 250 (1997) 99-105.

[37] L.G. Frigeri, T.R. Radabaugh, P.A. Haynes, M. Hildebrand, Identification of Proteins from a Cell Wall Fraction of the Diatom Thalassiosira pseudonana : Insights into Silica Structure Formation, Molecular & Cellular Proteomics, 5 (2006) 182-193.

[38] T. Mock, M.P. Samanta, V. Iverson, C. Berthiaume, M. Robison, K. Holtermann, C. Durkin, S.S. BonDurant, K. Richmond, M. Rodesch, T. Kallas, E.L. Huttlin, F. Cerrina, M.R. Sussman, E.V. Armbrust, Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses, Proceedings of the National Academy of Sciences, 105 (2008) 1579-1584.

[39] N. Kröger, N. Poulsen, Diatoms—From Cell Wall Biogenesis to Nanotechnology, Annual review of genetics, 42 (2008) 83-107.

[40] W.H. van de Poll, E.G. Vrieling, W.W.C. Gieskes, Location and expression of frustulins in the pennate diatoms Cylindrotheca fusiformis, Navicula pelliculosa, and Navicula salinarum (Bacillariophyceae), Journal of Phycology, 35 (1999) 1044-1053.

[41] J.R. Reinfelder, A.M.L. Kraepiel, F.M.M. Morel, Unicellular C4 photosynthesis in a marine diatom, Nature, 407 (2000) 996-999.

[42] J.R. Reinfelder, A.J. Milligan, F.M.M. Morel, The Role of the C4 Pathway in Carbon Accumulation and Fixation in a Marine Diatom, Plant Physiology, 135 (2004) 2106-2111.

[43] K. Roberts, E. Granum, R.C. Leegood, J.A. Raven, Carbon acquisition by diatoms, Photosynthesis research, 93 (2007) 79-88.

[44] K. Roberts, E. Granum, R.C. Leegood, J.A. Raven, C3 and C4 Pathways of Photosynthetic Carbon Assimilation in Marine Diatoms Are under Genetic, Not Environmental, Control, Plant Physiology, 145 (2007) 230-235.

[45] K. Nakajima, A. Tanaka, Y. Matsuda, SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater, Proceedings of the National Academy of Sciences, 110 (2013) 1767-1772.

[46] M. Haimovich-Dayan, N. Garfinkel, D. Ewe, Y. Marcus, A. Gruber, H. Wagner, P.G. Kroth, A. Kaplan, The role of C4 metabolism in the marine diatom Phaeodactylum tricornutum, New Phytologist, 197 (2013) 177-185.

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

[48] C. Bowler, A.E. Allen, J.H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R.P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J.A. Berges, C. Brownlee, J.-P. Cadoret, A. Chiovitti, C.J. Choi, S. Coesel, A. De Martino, J.C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M.J.J. Huysman, B.D. Jenkins, K. Jiroutova, R.E. Jorgensen, Y. Joubert, A. Kaplan, N. Kroger, P.G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jezequel, P.J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L.K. Medlin, A. Montsant, M.-P.O.-L. Secq, C. Napoli, M. Obornik, M.S. Parker, J.-L. Petit, B.M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T.A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M.R. Sussman, A.R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L.S. Wyrwicz, D.S. Rokhsar, J. Weissenbach, E.V. Armbrust, B.R. Green, Y. Van de Peer, I.V. Grigoriev, The Phaeodactylum genome reveals the evolutionary history of diatom genomes, Nature, 456 (2008) 239-244.

[49] A.J. Alverson, Untangling the origin and diversification of diatoms, 2015.

[50] J.C. Traller, S.J. Cokus, D.A. Lopez, O. Gaidarenko, S.R. Smith, J.P. McCrow, S.D. Gallaher, S. Podell, M. Thompson, O. Cook, M. Morselli, A. Jaroszewicz, E.E. Allen, A.E. Allen, S.S. Merchant, M. Pellegrini, M. Hildebrand, Genome and methylome of the oleaginous diatom Cyclotella cryptica reveal genetic flexibility toward a high lipid phenotype, Biotechnology for Biofuels, 9 (2016) 258.

[51] A. Vardi, K. Thamatrakoln, K.D. Bidle, P.G. Falkowski, Diatom genomes come of age, Genome Biology, 9 (2008) 245-245.

[52] A.E. Allen, A. Vardi, C. Bowler, An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms, Current Opinion in Plant Biology, 9 (2006) 264-273.

[53] A.E. Allen, C.L. dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J.P. McCrow, H. Zheng, D.A. Johnson, H. Hu, A.R. Fernie, C. Bowler, Evolution and metabolic significance of the urea cycle in photosynthetic diatoms., Nature, 473 (2011) 203-207.

[54] M. Fabris, M. Matthijs, S. Rombauts, W. Vyverman, A. Goossens, G.J.E. Baart, The metabolic blueprint of Phaeodactylum tricornutum reveals a eukaryotic Entner-Doudoroff glycolytic pathway, The Plant Journal, 70 (2012) 1004-1014.

[55] X. Chen, K. Schreiber, J. Appel, A. Makowka, B. Fähnrich, M. Roettger, M.R. Hajirezaei, F.D. Sönnichsen, P. Schönheit, W.F. Martin, K. Gutekunst, The Entner–Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants, Proceedings of the National Academy of Sciences, 113 (2016) 5441-5446.

[56] M. Stitt, Y. Gibon, Why measure enzyme activities in the era of systems biology?, Trends in Plant Science, 19 (2014) 256-265.

[57] N.L. Hockin, T. Mock, F. Mulholland, S. Kopriva, G. Malin, The Response of Diatom Central Carbon Metabolism to Nitrogen Starvation Is Different from That of Green Algae and Higher Plants, Plant Physiology, 158 (2012) 299-312.

[58] S.C. Zeeman, F. Northrop, A.M. Smith, T.a. Rees, A starch-accumulating mutant ofArabidopsis thalianadeficient in a chloroplastic starch-hydrolysing enzyme, The Plant Journal, 15 (1998) 357-365.

[59] M. Bar-Peled, M.A. O'Neill, Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling, Annual Review of Plant Biology, 62 (2011) 127-155.

[60] A.-C. Lamerz, T. Haselhorst, A.K. Bergfeld, M. von Itzstein, R. Gerardy-Schahn, Molecular cloning of the Leishmania major UDP-glucose pyrophosphorylase, functional characterization, and ligand binding analyses using NMR spectroscopy, Journal of Biological Chemistry, 281 (2006) 16314-16322.

[61] T. Kotake, D. Yamaguchi, H. Ohzono, S. Hojo, S. Kaneko, H.-K. Ishida, Y. Tsumuraya, UDP-sugar pyrophosphorylase with broad substrate specificity toward various monosaccharide 1-phosphates from pea sprouts, Journal of Biological Chemistry, 279 (2004) 45728-45736.

[62] A. Bacic, G.B. Fincher, B.A. Stone, Chemistry, biochemistry, and biology of 1-3 beta glucans and related polysaccharides, Academic Press2009.

[63] D.P.S. Verma, Z. Hong, Plant callose synthase complexes, Plant Molecular Biology, 47 (2001) 693-701.

[64] R. Chen, X. Zhao, Z. Shao, Z. Wei, Y. Wang, L. Zhu, J. Zhao, M. Sun, R. He, G. He, Rice UDP-Glucose Pyrophosphorylase1 Is Essential for Pollen Callose Deposition and Its Cosuppression Results in a New Type of Thermosensitive Genic Male Sterility, The Plant Cell, 19 (2007) 847-861.

[65] Z. Hong, Z. Zhang, J.M. Olson, D.P.S. Verma, A Novel UDP-Glucose Transferase Is Part of the Callose Synthase Complex and Interacts with Phragmoplastin at the Forming Cell Plate, The Plant Cell, 13 (2001) 769-779.

[66] V.A. Iglesias, F. Meins, Movement of plant viruses is delayed in a beta-1,3-glucanase-deficient mutant showing a reduced plasmodesmatal size exclusion limit and enhanced callose deposition, The Plant Journal, 21 (2000) 157-166.

[67] P.L. Keeling, A.M. Myers, Biochemistry and Genetics of Starch Synthesis, Annual Review of Food Science and Technology, 1 (2010) 271-303.

[68] Samuel C. Zeeman, Steven M. Smith, Alison M. Smith, The diurnal metabolism of leaf starch, Biochemical Journal, 401 (2007) 13-28.

[69] A. Sturm, Invertases. Primary Structures, Functions, and Roles in Plant Development and Sucrose Partitioning, Plant Physiology, 121 (1999) 1-8.

[70] H.D. Coleman, J. Yan, S.D. Mansfield, Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure, Proceedings of the National Academy of Sciences, 106 (2009) 13118-13123.

[71] H.D. Coleman, D.D. Ellis, M. Gilbert, S.D. Mansfield, Up-regulation of sucrose synthase and UDP-glucose pyrophosphorylase impacts plant growth and metabolism, Plant Biotechnology Journal, 4 (2006) 87-101.

[72] K. Denyer, C.M. Hylton, C.F. Jenner, A.M. Smith, Identification of multiple isoforms of soluble and granule-bound starch synthase in developing wheat endosperm, Planta, 196 (1995) 256-265.

[73] C.S. Hanes, The Reversible Formation of Starch from Glucose-1-Phosphate Catalysed by Potato Phosphorylase, Proceedings of the Royal Society: Biological Sciences, 129 (1940) 174-208.

[74] L.N. Sokolov, J.R. Dominguez-Solis, A.-L. Allary, B.B. Buchanan, S. Luan, A redox-regulated chloroplast protein phosphatase binds to starch diurnally and functions in its accumulation, Proceedings of the National Academy of Sciences, 103 (2006) 9732-9737.

[75] S. Comparot-Moss, O. Kötting, M. Stettler, C. Edner, A. Graf, S.E. Weise, S. Streb, W.-L. Lue, D. MacLean, S. Mahlow, G. Ritte, M. Steup, J. Chen, S.C. Zeeman, A.M. Smith, A putative phosphatase, LSF1, is required for normal starch turnover in Arabidopsis leaves, Plant Physiology, 152 (2010) 685-697.

[76] D. Guillén, S. Sánchez, R. Rodríguez-Sanoja, Carbohydrate-binding domains: multiplicity of biological roles, Appl Microbiol Biotechnol, 85 (2010) 1241-1249.

[77] Y. Li, D. Han, G. Hu, D. Dauvillee, M. Sommerfeld, S. Ball, Q. Hu, Chlamydomonas starchless mutant defective in ADP-glucose pyrophosphorylase hyper-accumulates triacylglycerol, Metabolic Engineering, 12 (2010) 387-391.

[78] N.C. Carpita, Update on Mechanisms of Plant Cell Wall Biosynthesis: How Plants Make Cellulose and Other (1 -> 4)-beta-D-Glycans, Plant Physiology, 155 (2011) 171-184.

[79] A.H. Liepman, C.G. Wilkerson, K. Keegstra, Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases, Proceedings of the National Academy of Sciences, 102 (2005) 2221-2226.

[80] A. Caño-Delgado, S. Penfield, C. Smith, M. Catley, M. Bevan, Reduced cellulose synthesis invokes lignification and defense responses in Arabidopsis thaliana, The Plant Journal, 34 (2003) 351-362.

[81] P.J. Roach, A.A. DePaoli-Roach, T.D. Hurley, V.S. Tagliabracci, Glycogen and its metabolism: some new developments and old themes, Biochemical Journal, 441 (2012) 763-787.

[82] P.L. Magoulas, A.W. El-Hattab, Glycogen storage disease type IV, (2013).

[83] M.S. Gentry, R.H. Dowen, C.A. Worby, S. Mattoo, J.R. Ecker, J.E. Dixon, The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease, Journal of Biological Chemistry, 178 (2007) 477-488.

[84] M.A. Ballicora, A.A. Iglesias, J. Preiss, ADP-Glucose Pyrophosphorylase, a Regulatory Enzyme for Bacterial Glycogen Synthesis, Microbiology and Molecular Biology Reviews, 67 (2003) 213-225.

[85] W.A. Wilson, P.J. Roach, M. Montero, E. Baroja-Fernández, F.J. Muñoz, G. Eydallin, A.M. Viale, J. Pozueta-Romero, Regulation of glycogen metabolism in yeast and bacteria, FEMS Microbiology Reviews, 34 (2010) 952-985.

[86] E. Suzuki, K. Umeda, S. Nihei, K. Moriya, H. Ohkawa, S. Fujiwara, M. Tsuzuki, Y. Nakamura, Role of the GlgX protein in glycogen metabolism of the cyanobacterium, Synechococcus elongatus PCC 7942, Biochimica et Biophysica Acta (BBA) - General Subjects, 1770 (2007) 763-773.

[87] A. Beattie, E.L. Hirst, E. Percival, Studies on the metabolism of the Chrysophyceae. Comparative structural investigations on leucosin (chrysolaminarin) separated from diatoms and laminarin from the brown algae, Biochemical Journal, 79 (1961) 531-537.

[88] C. Ford, E. Percival, The carbohydrates of Phaeodactylum tricornutum. Part I. Preliminary examination of the organism, and characterisation of low molecular weight material and of a glucan, Journal of the Chemical Society (Resumed), (1965) 7035-7041.

[89] M.C. Wang, S. Bartnicki-Garcia, Mycolaminarans: storage (1→3)- β -D-glucans from the cytoplasm of the fungus phytophthora palmivora, Carbohydrate Research, 37 (1974) 331-338.

[90] P.G. Roessler, UDP-glucose pyrophosphorylase activity in the diatom Cyclotella cryptica. Pathway of chrysolaminarin biosynthesis, Journal of Phycology, 23 (1987) 494-498.

[91] P.G. Kroth, A. Chiovitti, A. Gruber, V. Martin-Jezequel, T. Mock, M.S. Parker, M.S. Stanley, A. Kaplan, L. Caron, T. Weber, U. Maheswari, E.V. Armbrust, C. Bowler, A model for carbohydrate metabolism in the diatom Phaeodactylum tricornutum deduced from comparative whole genome analysis, PLoS ONE, 3 (2008) e1426.

[92] B.A. Stone, N.A. Evans, I. Bonig, A.E. Clarke, The application of Sirofluor, a chemically defined fluorochrome from aniline blue for the histochemical detection of callose, Protoplasma, 122 (1984) 191-195.

[93] N.A. Evans, P.A. Hoyne, B.A. Stone, Characteristics and specificity of the interaction of a fluorochrome from aniline blue (sirofluor) with polysaccharides, Carbohydrate Polymers, 4 (1984) 215-230.

[94] A. Chiovitti, P. Molino, S.A. Crawford, R. Teng, T. Spurck, R. Wetherbee, The glucans extracted with warm water from diatoms are mainly derived from intracellular chrysolaminaran and not extracellular polysaccharides, Eur J Phycol, 39 (2004) 117-128.

[95] L. Waterkeyn, A. Bienfait, Localisation et rôle des β-1, 3-glucanes (callose et chrysolaminarine) dans le genre Pinnularia (diatomées), La Cellule, 74 (1987) 198-226.

[96] J. Levering, J. Broddrick, C.L. Dupont, G. Peers, K. Beeri, J. Mayers, A.A. Gallina, A.E. Allen, B.O. Palsson, K. Zengler, Genome-Scale Model Reveals Metabolic Basis of Biomass Partitioning in a Model Diatom, PLOS ONE, 11 (2016) e0155038.

[97] D. Singh, R. Carlson, D. Fell, M. Poolman, Modelling metabolism of the diatom Phaeodactylum tricornutum, Biochemical Society Transactions, 43 (2015) 1182-1186.

[98] T. Mock, Personal communication, 2015.

[99] M.S. Chauton, P. Winge, T. Brembu, O. Vadstein, A.M. Bones, Gene Regulation of Carbon Fixation, Storage, and Utilization in the Diatom Phaeodactylum tricornutum Acclimated to Light/Dark Cycles, Plant Physiology, 161 (2013) 1034-1048.

[100] V. De Riso, R. Raniello, F. Maumus, A. Rogato, C. Bowler, A. Falciatore, Gene silencing in the marine diatom Phaeodactylum tricornutum., Nucleic Acids Research, 37 (2009) e96.

[101] M. Nymark, A.K. Sharma, T. Sparstad, A.M. Bones, P. Winge, A CRISPR/Cas9 system adapted for gene editing in marine algae, Scientific Reports, 6 (2016) 24951.

[102] M. Hildebrand, K. Manandhar-Shrestha, R. Abbriano, Effects of chrysolaminarin synthase knockdown in the diatom Thalassiosira pseudonana: Implications of reduced carbohydrate storage relative to green algae, Algal Research, 23 (2017) 66-77.

[103] W. Huang, C. Río Bártulos, P.G. Kroth, Diatom Vacuolar 1,6-β-Transglycosylases can Functionally Complement the Respective Yeast Mutants, Journal of Eukaryotic Microbiology, 63 (2016) 536-546.

[104] A.F.J. Ram, A. Wolters, R.T. Hoopen, F.M. Klis, A new approach for isolating cell wall mutants in Saccharomyces cerevisiae by screening for hypersensitivity to calcofluor white, Yeast, 10 (1994) 1019-1030.

[105] M. DuBois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, Colorimetric Method for Determination of Sugars and Related Substances, Analytical Chemistry, 28 (1956) 350-356.

[106] A.M. Smith, S.C. Zeeman, Quantification of starch in plant tissues, Nature Protocols, 1 (2006) 1342-1345.

[107] S. Myklestad, R. Djurhuus, A. Mohus, Demonstration of exo-(β-1,3)-d-glucanase activity in some planktonic diatoms, J Exp Mar Biol Ecol, 56 (1981) 205-211.

[108] T.A. Scott, E.H. Melvin, Determination of Dextran with Anthrone, Analytical Chemistry, 25 (1953) 1656-1661.

[109] L.A. Kleczkowski, D. Decker, M. Wilczynska, UDP-sugar pyrophosphorylase: a new old mechanism for sugar activation., Plant Physiology, 156 (2011) 3-10.

[110] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Research, 42 (2014) D490-495.

[111] G. Liao, Z. Zhou, S. Burgula, J. Liao, C. Yuan, Q. Wu, Z. Guo, Synthesis and Immunological Studies of Linear Oligosaccharides of β-Glucan as Antigens for Anti-Fungal Vaccine Development, Bioconjugate Chemistry, 26 (2015) 466-476.

[112] A. Torosantucci, C. Bromuro, P. Chiani, F. De Bernardis, F. Berti, C. Galli, F. Norelli, C. Bellucci, L. Polonelli, P. Costantino, R. Rappuoli, A. Cassone, A novel glyco-conjugate vaccine against fungal pathogens, The Journal of Experimental Medicine, 202 (2005) 597-606.

[113] E. Wakshull, D. Brunke-Reese, J. Lindermuth, L. Fisette, R.S. Nathans, J.J. Crowley, J.C. Tufts, J. Zimmerman, W. Mackin, D.S. Adams, PGG-Glucan, a soluble β-(1,3)-glucan, enhances the oxidative burst response, microbicidal activity, and activates an NF-κB-like factor in human PMN: Evidence for a glycosphingolipid β-(1,3)-glucan receptor, Immunopharmacology, 41 (1999) 89-107.

[114] D. Takahashi, H. Dai, Y. Hiromasa, R. Krishnamoorthi, M.R. Kanost, Self-association of an Insect β-1,3-Glucan Recognition Protein Upon Binding Laminarin Stimulates Prophenoloxidase Activation as an Innate Immune Response, Journal of Biological Chemistry, 289 (2014) 28399-28410.

[115] T.R. Størseth, S. Kirkvold, J. Skjermo, K.I. Reitan, A branched β-d-(1→3,1→6)-glucan from the marine diatom Chaetoceros debilis (Bacillariophyceae) characterized by NMR, Carbohydrate Research, 341 (2006) 2108-2114.

[116] T.R. Størseth, K. Hansen, K.I. Reitan, J. Skjermo, Structural characterization of β-d-(1→3)-glucans from different growth phases of the marine diatoms Chaetoceros mülleri and Thalassiosira weissflogii, Carbohydrate Research, 340 (2005) 1159-1164.

[117] S. Xia, B. Gao, A. Li, J. Xiong, Z. Ao, C. Zhang, Preliminary characterization, antioxidant properties and production of chrysolaminarin from marine diatom Odontella aurita, Marine Drugs, 12 (2014) 4883-4897.

[118] M.E. Himmel, S.-Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady, T.D. Foust, Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production, Science, 315 (2007) 804-807.

[119] S.J. Horn, I.M. Aasen, K. Østgaard, Ethanol production from seaweed extract, Journal of Industrial Microbiology and Biotechnology, 25 (2000) 249-254.

[120] M. Hildebrand, A.K. Davis, S.R. Smith, J.C. Traller, R. Abbriano, The place of diatoms in the biofuels industry, Biofuels, 3 (2012) 221-240.

[121] W. Yongmanitchai, O.P. Ward, Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions, Applied and Environmental Microbiology, 57 (1991) 419-425.

[122] M.L. Hamilton, R.P. Haslam, J.A. Napier, O. Sayanova, Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids, Metabolic Engineering, 22 (2014) 3-9.

[123] F. Yang, M.A. Hanna, R. Sun, Value-added uses for crude glycerol--a byproduct of biodiesel production, Biotechnology for Biofuels, 5 (2012) 13.

[124] M.A. Caballero, D. Jallet, L. Shi, C. Rithner, Y. Zhang, G. Peers, Quantification of chrysolaminarin from the model diatom Phaeodactylum tricornutum, Algal Research, 20 (2016) 180-188.

[125] H. Abida, L.-J. Dolch, C. Meï, V. Villanova, M. Conte, M.A. Block, G. Finazzi, O. Bastien, L. Tirichine, C. Bowler, F. Rébeillé, D. Petroutsos, J. Jouhet, E. Maréchal, Membrane Glycerolipid Remodeling Triggered by Nitrogen and Phosphorus Starvation in Phaeodactylum tricornutum, Plant Physiology, 167 (2015) 118-136.

[126] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, A. Darzins, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, The Plant Journal, 54 (2008) 621-639.

[127] E. Granum, S.M. Myklestad, A simple combined method for determination of β-1,3-glucan and cell wall polysaccharides in diatoms, Hydrobiologia, 477 (2002) 155-161.

[128] B. Gügi, T. Le Costaouec, C. Burel, P. Lerouge, W. Helbert, M. Bardor, Diatom-Specific Oligosaccharide and Polysaccharide Structures Help to Unravel Biosynthetic Capabilities in Diatoms, Marine Drugs, 13 (2015) 5993-6018.

[129] C. Ford, E. Percival, The carbohydrates of Phaeodactylum tricornutum. Part II. A sulphated glucuronomannan, Journal of the Chemical Society (Resumed), (1965) 7042-7046.

[130] J. Blackwell, K.D. Parker, K.M. Rudall, Chitin fibres of the diatoms Thalassiosira fluviatilis and Cyclotella cryptica, Journal of Molecular Biology, 28 (1967) 383-385.

[131] S.N. Aslam, T. Cresswell-Maynard, D.N. Thomas, G.J.C. Underwood, Production and Characterization of the Intra- and Extracellular Carbohydrates and Polymeric Substances (EPS) of Three Sea-Ice Diatom Species, and Evidence for a Cryoprotective Role for EPS, Journal of Phycology, 48 (2012) 1494-1509.

[132] L.T. Guerra, O. Levitan, M.J. Frada, J.S. Sun, P.G. Falkowski, G.C. Dismukes, Regulatory branch points affecting protein and lipid biosynthesis in the diatom Phaeodactylum tricornutum, Biomass Bioenerg, 59 (2013) 306-315.

[133] A.S. Mirón, M.C.C. García, A.C. Gómez, F.G.a. Camacho, E.M. Grima, Y. Chisti, Shear stress tolerance and biochemical characterization of Phaeodactylum tricornutum in quasi steady-state continuous culture in outdoor photobioreactors, Biochemical Engineering Journal, 16 (2003) 287-297.

[134] G.E. Anthon, D.M. Barrett, Determination of reducing sugars with 3-methyl-2-benzothiazolinonehydrazone, Analytical Biochemistry, 305 (2002) 287-289.

[135] R. Radakovits, R.E. Jinkerson, A. Darzins, M.C. Posewitz, Genetic Engineering of Algae for Enhanced Biofuel Production, Eukaryotic Cell, 9 (2010) 486-501.

[136] K.M. Vårum, S. Myklestad, Effects of light, salinity and nutrient limitation on the production of β-1,3-d-glucan and exo-d-glucanase activity in Skeletonema costatum (Grev) Cleve, J Exp Mar Biol Ecol, 83 (1984) 13-25.

[137] T. Lacour, A. Sciandra, A. Talec, P. Mayzaud, O. Bernard, Diel Variations of Carbohydrates and Neutral Lipids in Nitrogen-Sufficient and Nitrogen-Starved Cyclostat Cultures of Isochrysis Sp., Journal of Phycology, 48 (2012) 966-975.

[138] L. Alipanah, J. Rohloff, P. Winge, A.M. Bones, T. Brembu, Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum, Journal of Experimental Botany, 66 (2015) 6281-6296.

[139] G. Breuer, P.P. Lamers, D.E. Martens, R.B. Draaisma, R.H. Wijffels, The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains, Bioresource Technology, 124 (2012) 217-226.

[140] M.T. Juergens, R.R. Deshpande, B.F. Lucker, J.-J. Park, H. Wang, M. Gargouri, F.O. Holguin, B. Disbrow, T. Schaub, J.N. Skepper, D.M. Kramer, D.R. Gang, L.M. Hicks, Y. Shachar-Hill, The Regulation of Photosynthetic Structure and Function during Nitrogen Deprivation in Chlamydomonas reinhardtii, Plant Physiology, 167 (2015) 558-573.

[141] R. Tevatia, J. Allen, P. Blum, Y. Demirel, P. Black, Modeling of rhythmic behavior in neutral lipid production due to continuous supply of limited nitrogen: Mutual growth and lipid accumulation in microalgae, Bioresource Technology, 170 (2014) 152-159.

[142] M. Chen, H. Tang, H. Ma, T.C. Holland, K.Y.S. Ng, S.O. Salley, Effect of nutrients on growth and lipid accumulation in the green algae Dunaliella tertiolecta, Bioresource Technology, 102 (2011) 1649-1655.

[143] Z.T. Wang, N. Ullrich, S. Joo, S. Waffenschmidt, U. Goodenough, Algal Lipid Bodies: Stress Induction, Purification, and Biochemical Characterization in Wild-Type and Starchless Chlamydomonas reinhardtii, Eukaryotic Cell, 8 (2009) 1856-1868.

[144] Y. Zhang, Y. Liu, X. Cao, P. Gao, X. Liu, X. Wang, J. Zhang, J. Zhou, S. Xue, G. Xu, J. Tian, Free amino acids and small molecular acids profiling of marine microalga Isochrysis zhangjiangensis under nitrogen deficiency, Algal Research, 13 (2016) 207-217.

[145] T.R. Hauser, R.L. Cummins, Increasing Sensitivity of 3-Methyl-2-Benzothiazolone Hydrozone Test for Analysis of Aliphatic Aldehydes in Air, Analytical Chemistry, 36 (1964) 679-681.

[146] J. Dean Pakulski, R. Benner, An improved method for the hydrolysis and MBTH analysis of dissolved and particulate carbohydrates in seawater, Marine Chemistry, 40 (1992) 143-160.

[147] G.E. Anthon, D.M. Barrett, Kinetic Parameters for the Thermal Inactivation of Quality-Related Enzymes in Carrots and Potatoes, Journal of Agricultural and Food Chemistry, 50 (2002) 4119-4125.

[148] S. Jarle Horn, V.G.H. Eijsink, A reliable reducing end assay for chito-oligosaccharides, Carbohydrate Polymers, 56 (2004) 35-39.

[149] R.R.L. Guillard, Culture of Phytoplankton for Feeding Marine Invertebrates, in: W.L. Smith, M.H. Chanley (Eds.) Culture of Marine Invertebrate Animals: Proceedings — 1st Conference on Culture of Marine Invertebrate Animals Greenport, Springer US, Boston, MA, 1975, pp. 29-60.

[150] S. Bhamidi, L. Shi, D. Chatterjee, J.T. Belisle, D.C. Crick, M.R. McNeil, A bioanalytical method to determine the cell wall composition of Mycobacterium tuberculosis grown in vivo, Analytical Biochemistry, 421 (2012) 240-249.

[151] Y.-T. Kim, E.-H. Kim, C. Cheong, D.L. Williams, C.-W. Kim, S.-T. Lim, Structural characterization of β -D-(1→ 3, 1→ 6)-linked glucans using NMR spectroscopy, Carbohydrate Research, 328 (2000) 331-341.

[152] R.J. Ritchie, Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and total chlorophylls in natural assemblages of photosynthetic organisms using acetone, methanol, or ethanol solvents, Photosynthetica, 46 (2008) 115-126.

[153] T. Govender, L. Ramanna, I. Rawat, F. Bux, BODIPY staining, an alternative to the Nile Red fluorescence method for the evaluation of intracellular lipids in microalgae, Bioresource Technology, 114 (2012) 507-511.

[154] D. Jallet, M.A. Caballero, A.A. Gallina, M. Youngblood, G. Peers, Photosynthetic physiology and biomass partitioning in the model diatom Phaeodactylum tricornutum grown in a sinusoidal light regime, Algal Research, 18 (2016) 51-60.

[155] M.A. Danielewicz, L.A. Anderson, A.K. Franz, Triacylglycerol profiling of marine microalgae by mass spectrometry, Journal of Lipid Research, 52 (2011) 2101-2108.

[156] I. Grouneva, T. Jakob, C. Wilhelm, R. Goss, The regulation of xanthophyll cycle activity and of non-photochemical fluorescence quenching by two alternative electron flows in the diatoms Phaeodactylum tricornutum and Cyclotella meneghiniana, Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1787 (2009) 929-938.

[157] R.J. Olson, E.R. Zettler, O.K. Anderson, Discrimination of eukaryotic phytoplankton cell types from light scatter and autofluorescence properties measured by flow cytometry, Cytometry, 10 (1989) 636-643.

[158] N.R. Baker, Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo, Annual Review of Plant Biology, 59 (2008) 89-113.

[159] A.S. Abdullahi, G.J.C. Underwood, M.R. Gretz, Extracellular matrix assembly in diatoms (Bacillariophyceae). V. Environmental effects on polysaccharide synthesis in the model diatom Phaeodactylum tricornutum, Journal of Phycology, 42 (2006) 363-378.

[160] L.M.L. Laurens, T.A. Dempster, H.D.T. Jones, E.J. Wolfrum, S. Van Wychen, J.S.P. McAllister, M. Rencenberger, K.J. Parchert, L.M. Gloe, Algal Biomass Constituent Analysis: Method Uncertainties and Investigation of the Underlying Measuring Chemistries, Analytical Chemistry, 84 (2012) 1879-1887.

[161] K.M. Varum, K. Ostgaard, K. Grimsrud, Diurnal Rhythms in Carbohydrate-Metabolism of the Marine Diatom Skeletonema-Costatum (Grev) Cleve, J Exp Mar Biol Ecol, 102 (1986) 249-256.

[162] R.J. Geider, J. La Roche, R.M. Greene, M. Olaizola, Response of the photosynthetic apparatus of Phaeodactylum tricornutum (Bacillariophyceae) to nitrate, phosphate, or iron starvation, Journal of Phycology, 29 (1993) 755-766.

[163] M. Palmucci, S. Ratti, M. Giordano, Ecological and Evolutionary Implications of Carbon Allocation in Marine Phytoplankton as a Function of Nitrogen Availability: A Fourier Transform Infrared Spectroscopy Approach, Journal of Phycology, 47 (2011) 313-323.

[164] F. Mus, J.-P. Toussaint, K.E. Cooksey, M.W. Fields, R. Gerlach, B.M. Peyton, R.P. Carlson, Physiological and molecular analysis of carbon source supplementation and pH stress-induced lipid accumulation in the marine diatom Phaeodactylum tricornutum, Appl Microbiol Biotechnol, 97 (2013) 3625-3642.

[165] S. Myklestad, Production of carbohydrates by marine planktonic diatoms. I. Comparison of nine different species in culture, J Exp Mar Biol Ecol, 15 (1974) 261-274.

[166] X. Li, E.R. Moellering, B. Liu, C. Johnny, M. Fedewa, B.B. Sears, M.-H. Kuo, C. Benning, A Galactoglycerolipid Lipase Is Required for Triacylglycerol Accumulation and Survival Following Nitrogen Deprivation in Chlamydomonas reinhardtii, Plant Cell, 2012.

[167] M.-C. Wang, S. Bartnicki-Garcia, Distribution of mycolaminarans and cell wall β-glucans in the life cycle of Phytophthora, Experimental Mycology, 4 (1980) 269-280.

[168] M. Stitt, S.C. Zeeman, Starch turnover: pathways, regulation and role in growth, Current Opinion in Plant Biology, 15 (2012) 282-292.

[169] X.-Y. Chen, J.-Y. Kim, Callose synthesis in higher plants, Plant Signaling & Behavior, 4 (2009) 489-492.

[170] T. Kotake, S. Hojo, D. Yamaguchi, T. Aohara, T. Konishi, Y. Tsumuraya, Properties and Physiological Functions of UDP-Sugar Pyrophosphorylase in Arabidopsis, Bioscience, Biotechnology, and Biochemistry, 71 (2007) 761-771.

[171] B. Baïet, C. Burel, B. Saint-Jean, R. Louvet, L. Menu-Bouaouiche, M.-C. Kiefer-Meyer, E. Mathieu-Rivet, T. Lefebvre, H. Castel, A. Carlier, J.-P. Cadoret, P. Lerouge, M. Bardor, N-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme, Journal of Biological Chemistry, 286 (2011) 6152-6164.

[172] M. Siaut, M. Heijde, M. Mangogna, A. Montsant, S. Coesel, A. Allen, A. Manfredonia, A. Falciatore, C. Bowler, Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum, Gene, 406 (2007) 23-35.

[173] S.J. Elledge, J.T. Mulligan, S.W. Ramer, M. Spottswood, R.W. Davis, Lambda YES: a multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and Escherichia coli mutations, Proceedings of the National Academy of Sciences, 88 (1991) 1731-1735.

[174] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research, 32 (2004) 1792-1797.

[175] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, Molecular Biology and Evolution, 33 (2016) 1870-1874.

[176] P.D. Weyman, K. Beeri, S.C. Lefebvre, J. Rivera, J.K. McCarthy, A.L. Heuberger, G. Peers, A.E. Allen, C.L. Dupont, Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis, Plant Biotechnology Journal, 13 (2015) 460-470.

[177] M.Z. Li, S.J. Elledge, Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC, Nature Methods, 4 (2007) 251-256.

[178] L.A. Zaslavskaia, J.C. Lippmeier, P.G. Kroth, A.R. Grossman, K.E. Apt, Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes, Journal of Phycology, 36 (2000) 379-386.

[179] K.E. Apt, A.R. Grossman, P.G. Kroth-Pancic, Stable nuclear transformation of the diatom Phaeodactylum tricornutum, Molecular and General Genetics, 252 (1996) 572-579.

[180] J.C. Sanford, F.D. Smith, J.A. Russell, Optimizing the biolistic process for different biological applications, Methods in Enzymology, 217 (1993) 483-509.

[181] M. Wan, J.N. Rosenberg, J. Faruq, M.J. Betenbaugh, J. Xia, An improved colony PCR procedure for genetic screening of Chlorella and related microalgae, Biotechnology Letters, 33 (2011) 1615-1619.

[182] K. Maeda, R. Narikawa, M. Ikeuchi, CugP Is a Novel Ubiquitous Non-GalU-Type Bacterial UDP-Glucose Pyrophosphorylase Found in Cyanobacteria, Journal of Bacteriology, 196 (2014) 2348-2354.

[183] P.H.C. Eilers, J.C.H. Peeters, A model for the relationship between light intensity and the rate of photosynthesis in phytoplankton, Ecological Modelling, 42 (1988) 199-215.

[184] Y. Liang, Y. Maeda, T. Yoshino, M. Matsumoto, T. Tanaka, Profiling of Polar Lipids in Marine Oleaginous Diatom Fistulifera solaris JPCC DA0580: Prediction of the Potential Mechanism for Eicosapentaenoic Acid-Incorporation into Triacylglycerol, Marine Drugs, 12 (2014) 3218-3230.

[185] J.M. Daran, N. Dallies, D. Thines-Sempoux, V. Paquet, J. Francois, Genetic and Biochemical Characterization of the UGP1 Gene Encoding the UDP-Glucose Pyrophosphorylase from Saccharomyces cerevisiae, European Journal of Biochemistry, 233 (1995) 520-530.

[186] M. Joersbo, F.T. Okkels, A novel principle for selection of transgenic plant cells: positive selection, Plant Cell Reports, 16 (1996) 219-221.

[187] G.L. Rosano, E.A. Ceccarelli, Recombinant protein expression in Escherichia coli: advances and challenges, Frontiers in Microbiology, 5 (2014) 172.

[188] A. Vera, N. González-Montalbán, A. Arís, A. Villaverde, The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures, Biotechnology and Bioengineering, 96 (2007) 1101-1106.

[189] A. Gruber, G. Rocap, P.G. Kroth, E.V. Armbrust, T. Mock, Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage, The Plant Journal, 81 (2015) 519-528.

[190] J.D. Bishop, B.C. Moon, F. Harrow, D. Ratner, R.H. Gomer, R.P. Dottin, D.T. Brazill, A Second UDP-glucose Pyrophosphorylase Is Required for Differentiation and Development in Dictyostelium discoideum, Journal of Biological Chemistry, 277 (2002) 32430-32437.

[191] M. Peschke, D. Moog, A. Klingl, U.G. Maier, F. Hempel, Evidence for glycoprotein transport into complex plastids, Proceedings of the National Academy of Sciences, 110 (2013) 10860-10865.

[192] A. Sato, K. Sugimoto, M. Tsuzuki, N. Sato, UDP-Glucose Pyrophosphorylase Responsible for Sulfolipid Synthesis in a Green Alga Chlamydomonas Reinhardtii, Photosynthesis Research for Food, Fuel and the Future: 15th International Conference on Photosynthesis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 616-619.

[193] E.M. Trentacoste, R.P. Shrestha, S.R. Smith, C. Glé, A.C. Hartmann, M. Hildebrand, W.H. Gerwick, Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth, Proceedings of the National Academy of Sciences, 110 (2013) 19748-19753.

[194] J. Dyrløv Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved Prediction of Signal Peptides: SignalP 3.0, Journal of Molecular Biology, 340 (2004) 783-795.

[195] R. Haas, H.P. Siebertz, K. Wrage, E. Heinz, Localization of sulfolipid labeling within cells and chloroplasts, Planta, 148 (1980) 238-244.

[196] N. Sato, K. Sonoike, M. Tsuzuk, A. Kawaguchi, Impaired Photosystem II in a Mutant of Chlamydomonas Reinhardtii Defective in Sulfoquinovosyl Diacylglycerol, European Journal of Biochemistry, 234 (1995) 16-23.

[197] S. Güler, A. Seeliger, H. Härtel, G. Renger, C. Benning, A Null Mutant of Synechococcus sp. PCC7942 Deficient in the Sulfolipid Sulfoquinovosyl Diacylglycerol, Journal of Biological Chemistry, 271 (1996) 7501-7507.

[198] A. Minoda, N. Sato, H. Nozaki, K. Okada, H. Takahashi, K. Sonoike, M. Tsuzuki, Role of sulfoquinovosyl diacylglycerol for the maintenance of photosystem II in Chlamydomonas reinhardtii, European Journal of Biochemistry, 269 (2002) 2353-2358.

[199] I. Sakurai, N. Mizusawa, H. Wada, N. Sato, Digalactosyldiacylglycerol Is Required for Stabilization of the Oxygen-Evolving Complex in Photosystem II, Plant Physiology, 145 (2007) 1361-1370.

[200] C.E. Pugh, A.B. Roy, T. Hawkes, J.L. Harwood, A new pathway for the synthesis of the plant sulpholipid, sulphoquinovosyldiacylglycerol, Biochemical Journal, 309 (1995) 513-519.

[201] J.A. Schnurr, K.K. Storey, H.-J.G. Jung, D.A. Somers, J.W. Gronwald, UDP-sugar pyrophosphorylase is essential for pollen development in Arabidopsis, Planta, 224 (2006) 520-532.

[202] H.D. Coleman, T. Canam, K.Y. Kang, D.D. Ellis, S.D. Mansfield, Over-expression of UDP-glucose pyrophosphorylase in hybrid poplar affects carbon allocation, Journal of Experimental Botany, 58 (2007) 4257-4268.

[203] D.C. Elliott, T.R. Hart, A.J. Schmidt, G.G. Neuenschwander, L.J. Rotness, M.V. Olarte, A.H. Zacher, K.O. Albrecht, R.T. Hallen, J.E. Holladay, Process development for hydrothermal liquefaction of algae feedstocks in a continuous-flow reactor, Algal Research, 2 (2013) 445-454.

[204] A. Reyes-Prieto, A.P. Weber, D. Bhattacharya, The origin and establishment of the plastid in algae and plants, Annual review of genetics, 41 (2007) 147-168.

[205] L. Chen, D. Vitkup, Predicting genes for orphan metabolic activities using phylogenetic profiles, Genome Biology, 7 (2006) 1-13.

[206] H. Gross, Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects, Appl Microbiol Biotechnol, 75 (2007) 267-277.

[207] N.H.H. Heegaard, Affinity in Electrophoresis, Electrophoresis, 30 (2009) S229-S239.

[208] S. Moraïs, R. Lamed, E.A. Bayer, Affinity Electrophoresis as a Method for Determining Substrate-Binding Specificity of Carbohydrate-Active Enzymes for Soluble Polysaccharides, in:

M.E. Himmel (Ed.) Biomass Conversion: Methods and Protocols, Humana Press, Totowa, NJ, 2012, pp. 119-127.

[209] W.H. Schwarz, K. Bronnenmeier, F. Gräbnitz, W.L. Staudenbauer, Activity staining of cellulases in polyacrylamide gels containing mixed linkage β-glucans, Analytical Biochemistry, 164 (1987) 72-77.

[210] L.A. Elyakova, T.N. Zvyagintseva, A study of the laminarins of some far-eastern, brown seaweeds, Carbohydrate Research, 34 (1974) 241-248.

[211] C. Fernandez-Patron, L. Castellanos-Serra, P. Rodriguez, Reverse staining of sodium dodecyl sulfate polyacrylamide gels by imidazole-zinc salts: sensitive detection of unmodified proteins, BioTechniques, 12 (1992) 564-573.

[212] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry, Analytical Chemistry, 75 (2003) 4646-4658.

[213] H. Hillebrand, C.-D. Dürselen, D. Kirschtel, U. Pollingher, T. Zohary, Biovolume calculation for pelagic and benthic microalgae, Journal of Phycology, 35 (1999) 403-424.

[214] S. Nallamsetty, B.P. Austin, K.J. Penrose, D.S. Waugh, Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli, Protein Science, 14 (2005) 2964-2971.

[215] W.-T. Chan, Chandra S. Verma, David P. Lane, Samuel K.-E. Gan, A comparison and optimization of methods and factors affecting the transformation of *Escherichia coli*, Bioscience Reports, 33 (2013).

[216] M.G.B. Tom Berkelman, Haiyin Chang, Tim Cross, and William Strong, Tips to Prevent Streaking on 2-D Gels, BioRadiations, Bio-Rad Laboratories, 2005, pp. 56-57.

[217] Z. Elghabi, S. Ruf, R. Bock, Biolistic co-transformation of the nuclear and plastid genomes, The Plant Journal, 67 (2011) 941-948.

[218] B. Miki, A. Abdeen, Y. Manabe, P. MacDonald, Selectable marker genes and unintended changes to the plant transcriptome, Plant Biotechnology Journal, 7 (2009) 211-218.

[219] V.A. Bielinski, T.M. Bolt, C.L. Dupont, P.D. Weyman, Episomal tools for RNAi in the diatom Phaeodactylum tricornutum, PeerJ Preprints, 5 (2017) e2907v2901.

[220] C.U. Stirnimann, E. Petsalaki, R.B. Russell, C.W. Müller, WD40 proteins propel cellular networks, Trends in Biochemical Sciences, 35 (2010) 565-574.

[221] T.R. Shieh, J.H. Ware, Survey of Microorganisms for the Production of Extracellular Phytase, Applied Microbiology, 16 (1968) 1348-1351.

[222] J.-M. Girard, K.H.D. Lê, F. Lederer, Molecular characterization of laforin, a dual-specificity protein phosphatase implicated in Lafora disease, Biochimie, 88 (2006) 1961-1971.

[223] C.A. Worby, M.S. Gentry, J.E. Dixon, Laforin, a Dual Specificity Phosphatase That Dephosphorylates Complex Carbohydrates, Journal of Biological Chemistry, 281 (2006) 30412-30418.

[224] O. Kötting, D. Santelia, C. Edner, S. Eicke, T. Marthaler, M.S. Gentry, S. Comparot-Moss, J. Chen, A.M. Smith, M. Steup, G. Ritte, S.C. Zeeman, STARCH-EXCESS4 Is a Laforin-Like Phosphoglucan Phosphatase Required for Starch Degradation in Arabidopsis thaliana, The Plant Cell, 21 (2009) 334-346.

[225] T. Niittyla, S. Comparot-Moss, W.L. Lue, G. Messerli, M. Trevisan, M.D.J. Seymour, J.A. Gatehouse, D. Villadsen, S.M. Smith, J. Chen, S.C. Zeeman, A.M. Smith, Similar Protein Phosphatases Control Starch Metabolism in Plants and Glycogen Metabolism in Mammals, Journal of Biological Chemistry, 281 (2006) 11815-11818.

[226] C. Xu, J. Min, Structure and function of WD40 domain proteins, Protein & Cell, 2 (2011) 202-214.

[227] C. MacKintosh, Protein Phosphorylation: A Practical Approach, New York: IRL Press1993.

[228] C. Zabawinski, N. Van Den Koornhuyse, C. D'Hulst, R. Schlichting, C. Giersch, B. Delrue, J.M. Lacroix, J. Preiss, S. Ball, Starchless mutants of Chlamydomonas reinhardtii lack the small subunit of a heterotetrameric ADP-glucose pyrophosphorylase, Journal of Bacteriology, 183 (2001) 1069-1077.

[229] J.H.M. Hendriks, A. Kolbe, Y. Gibon, M. Stitt, P. Geigenberger, ADP-Glucose Pyrophosphorylase Is Activated by Posttranslational Redox-Modification in Response to Light and to Sugars in Leaves of Arabidopsis and Other Plant Species, Plant Physiology, 133 (2003) 838-849.

[230] N. Ye, X. Zhang, M. Miao, X. Fan, Y. Zheng, D. Xu, J. Wang, L. Zhou, D. Wang, Y. Gao, Y. Wang, W. Shi, P. Ji, D. Li, Z. Guan, C. Shao, Z. Zhuang, Z. Gao, J. Qi, F. Zhao, Saccharina genomes provide novel insight into kelp biology, Nature Communications, 6 (2015) 6986.

[231] Y. Chisti, Biodiesel from microalgae, Biotechnology Advances, 25 (2007) 294-306.

[232] R. Bosma, W.A. van Spronsen, J. Tramper, R.H. Wijffels, Ultrasound, a new separation technique to harvest microalgae, J Appl Phycol, 15 (2003) 143-153.

[233] H. Bruus, Acoustofluidics 10: Scaling laws in acoustophoresis, Lab on a Chip, 12 (2012) 1578-1586.

[234] J.S. Tebbutt, R.E. Challis, Ultrasonic wave propagation in colloidal suspensions and emulsions: a comparison of four models, Ultrasonics, 34 (1996) 363-368.

APPENDIX A: THE ACOUSTIC CONTRAST FACTOR OF PHAEODACTYLUM TRICORNUTUM IS REDUCED IN THE ABSENCE OF SILICA

Preface

Interdisciplinary rotation projects are a requirement of the NSF IGERT Multiple Approaches to Sustainable Bioenergy fellowship, which is reported in this appendix. I collaborated with former graduate student Esteban Hincapie (Mechanical Engineering) to investigate the potential use of hydroacoustics to prepare enriched volumes of diatom cell cultures grown in different media types. This project was designed and performed during the Summer 2013 semester.

Introduction

While algal photosynthetic processes have been coupled with the production of various carbon products to produce sustainable fuels and chemicals[231], separating algae from liquid media to collect the product of interest has proven to be a difficult problem. Conventional techniques include centrifugation and flocculation. These strategies are effective at the laboratory scale but can be energy intensive and financially prohibitive upon scaling up to industrially-relevant volumes. An acoustophoretic approach has the potential to separate algae while requiring less energy and lower capital costs than conventional strategies.

The properties of a standing wave enable the use of acoustophoresis to separate solid particles in a liquid media. Sound waves travel from a piezoelectric source from one side of a cuvette to the other, and at certain wavelengths, cause resonance[232] (Figure A.1). Acoustic force is exerted on particles in this scenario, as characterized in Equation 1[233]:

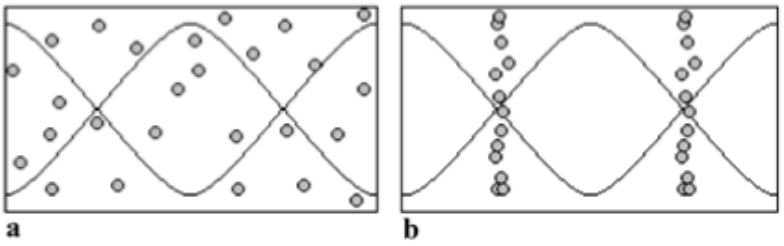

Figure A.1: Acoustophoresis principle diagram.
Resonance frequencies are applied to cells in suspension (a). The acoustic force applied in this scenario permits alignment of cells at the node (b).

$$F_{ac} = \frac{4\pi r^3}{3} \times E_{ac} \times \kappa \times \mathcal{F}$$

Equation A.1: The Acoustic Force Equation.

The acoustic contrast factor ($\mathcal{F}$) in Equation C.1 is a function of density and speed of sound of the cells and the surrounding media. $\mathcal{F}$ may be roughly calculated from the cellular composition of pure lipid, carbohydrate, and protein ($\mathcal{F}$= -0.2, 0.3, and 0.2 respectively). The acoustic contrast factor is further defined as: $\mathcal{F} = \frac{1}{3}\left[\frac{5\Lambda-2}{1+2\Lambda} - \frac{1}{\sigma^2\Lambda}\right]$ where Λ is the ratio of density of particle and the media ($\Lambda = \rho_p/\rho_m$) and σ the ratio of the speed of sound in the particle and in the media ($\sigma = c_p/c_m$). The ratio of $1/\sigma^2\lambda$ could be replaced for the ratio of particle and media compressibility (κ_p/κ_m) since $\kappa = 1/\rho c^2$.

$\mathcal{F}$ should be as large as possible to maximize the acoustophoretic effect, as $\mathcal{F}$ is a direct multiplier of acoustic force. Therefore, understanding $\mathcal{F}$ and maximizing it from a cell culture perspective is of great interest for acoustic separation strategies.

Determining the density and speed of sound of the cells and media is possible despite existing as a mixture. Urick observed the following relationships of speed of sound (c) and density (ρ) of particles (_c) and media (_m) in a mixture, where φ is the volumetric ratio (Equations A.2 and A.3 [234]). These equations have successfully been applied to determining

the acoustic properties and ultimately $\mathcal{F}$ of the algae species *Nannochloropsis* and *Chlamydomonas* (Hincapie and Marchese, 2013).

$$c = \varphi c_m + (1 - \varphi) c_c$$

Equation A.2: Urick's speed of sound equation for particles and media in a mixture.

$$\rho = \varphi \rho_m + (1 - \varphi) \rho_c$$

Equation A.3: Urick's density equation for particles and media in a mixture.

The diatom *Phaeodactylum tricornutum* has been identified as a promising platform to produce lipids for alage-based biodiesel. Most diatoms require a media supplementation of silica (Si) to generate a glass shell structure known as a frustule. Si is an inconvenient nutrient supplement, as it is comparatively expensive and it exhibits poor solubility. Interestingly, *Phaeodactylum* is the only model diatom known to be viable in the absence of Si. The frustule may increase the average density of each cell, and the perturbed metabolism may affect lipid, carbohydrate, and protein composition, all of which may alter $\mathcal{F}$. The primary goal of this short rotation was to determine if the absence of Si in the media would alter the acoustic properties of *Phaeodactylum,* specifically $\mathcal{F}$.

Methods

Culture conditions

Phaeodactylum was grown in f/2 media. Artificial seawater was prepared at 100 % saltwater concentration (Instant Ocean) and was filter sterilized (Nalgene Rapid-Flow bottle top filter, 0.2 μm aPES membrane). An excess of nutrients was provided by adding 300 μl L^{-1} of filter sterilized f/2 Algae Food Solutions A and B (Proline). The plus silica medium also supplemented 1 ml L^{-1} of filter sterilized sodium metasilicate nonahydrate (30.0 g L^{-1} diH2O, Fischer). Cultures were grown in 800 ml volumes in Roux flasks at 18 °C (Percival). Cultures

were continuously illuminated on both sides by T5 high output fluorescent bulbs (Phillips F54T5/841 HO) at a total intensity of 475 μmol photons m^{-2} s^{-1} ± 45 μmol photons m^{-2} s^{-1}. Cell densities were determined by flow cytometry (BD Accuri C6) using slow fluidics for 2 minutes.

Harvesting cells, determining acoustic properties, and calculating biovolume.

16 ml of 8 % (v/v) cells was prepared by centrifugation (3220 x g, 5 minutes) using a preliminary estimate of cell biovolume ($7.91*10^{-17}$ m^3). This stock was diluted with supernatant to prepare samples of the following concentrations: 100, 85, 70, 55, 40, 25, 10, and 0 %. These samples were processed using a DSA 5000 M density and sound velocity meter (Anton Parr) to calculate density and speed of sound of the emulsion, given Urick's equations (Equations A.2 & A.3). Cell density and speed of sound calculations for various biovolumes permitted calculations for cell compressibility and the acoustic contrast factor in Microsoft Excel. Photographs of cells were taken with an inverted microscope (Nikon TMZ, 40x) with a reference. Cell volumes were calculated assuming a prolate sphere as described [213]. determined by averaging calculated values as described by Hillebrand et al. (1999) and also using a microscope camera pictures (HITACHI KPD50) that were post-processed with AutoCAD v14.

Results

Phaeodactylum cultures were grown with and without silica in the media. There were no major differences in growth rate between medium conditions (Figure A.2), assuming the 10 % variation typically observed with biological replicates. The maximum quantum yield of photosystem II also was comparable between cultures during the experiment (Figure A.3). These observations agree with previous observations that *Phaeodactylum* is viable in the absence of silica.

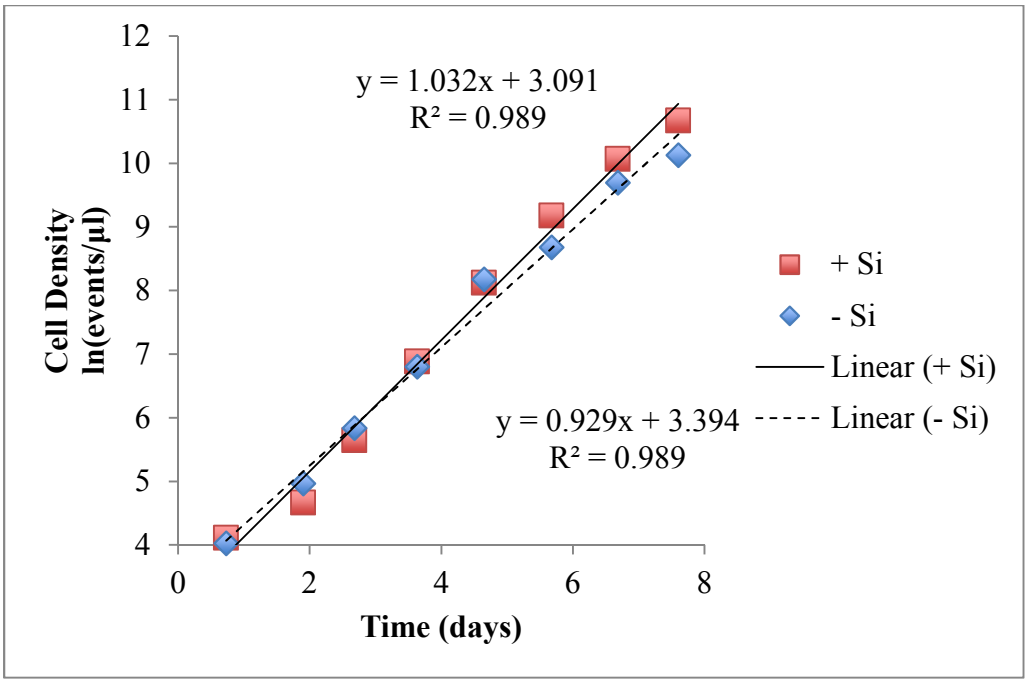

Figure A.2: Cell densities of *Phaeodactylum* during growth with or without silica.
Cell densities were determined by flow cytometry by gating for chlorophyll autofluorescence.

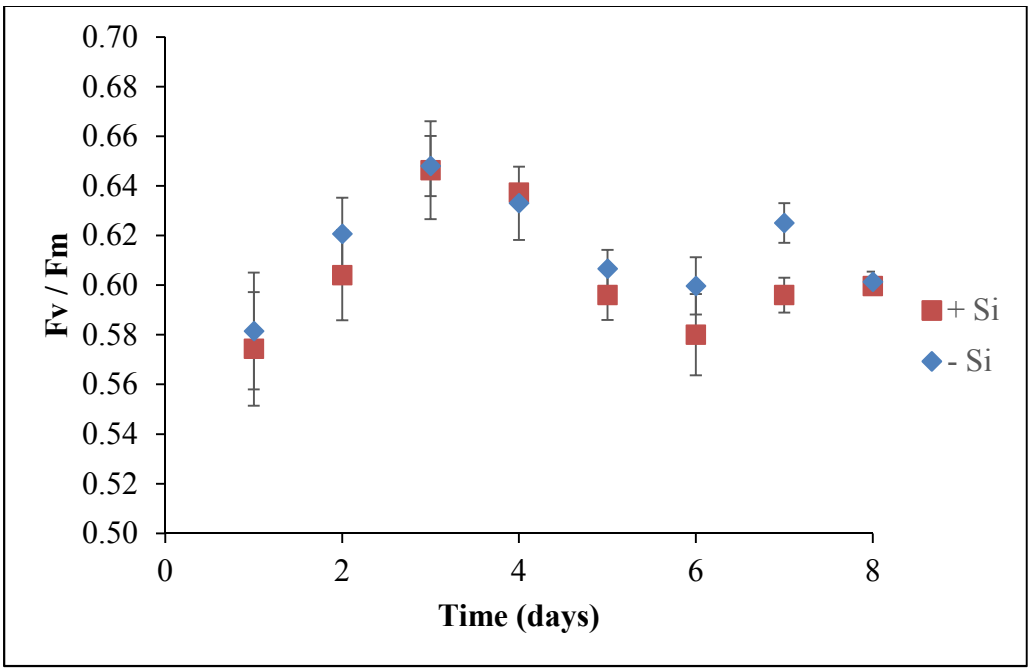

Figure A.3: Fv/Fm of *Phaeodactylum* during growth with or without silica.
Fv/Fm values were measured using FIRe fluorometry after a 15 minute dark incubation to relax non-photochemical quenching. n = 3 (technical replicates) and error = standard deviation.

However, *Phaeodactylum* cultures grown with silica had greater biovolume and cell compressibility than those grown without silica (Table A.1).

Table A.1 Hydroacoustic properties of *Phaeodactylum* cells grown with or without silica.
If values include (standard deviation), values represent the mean (n = 7, technical replicates).

Media:	+ Si	- Si
Cell Biovolume (m^3)	$1.71*10^{-16}$	$2.70*10^{-16}$
Cell Density (g/cm^3)	1.037810 (0.000515)	1.033308 (0.000126)
Cell Speed of Sound (m/s)	1530.71 (2.66)	1527.93 (0.07)
Cell Compressibility (1/Pa)	$4.11*10^{-10}$	$4.15*10^{-10}$
$\mathcal{F}$, Acoustic Contrast Factor	0.051	0.048

Alignments for determining cell density (Figure C.4) and speed of sound (Figure C.5) exhibited R^2 values greater than 0.99.

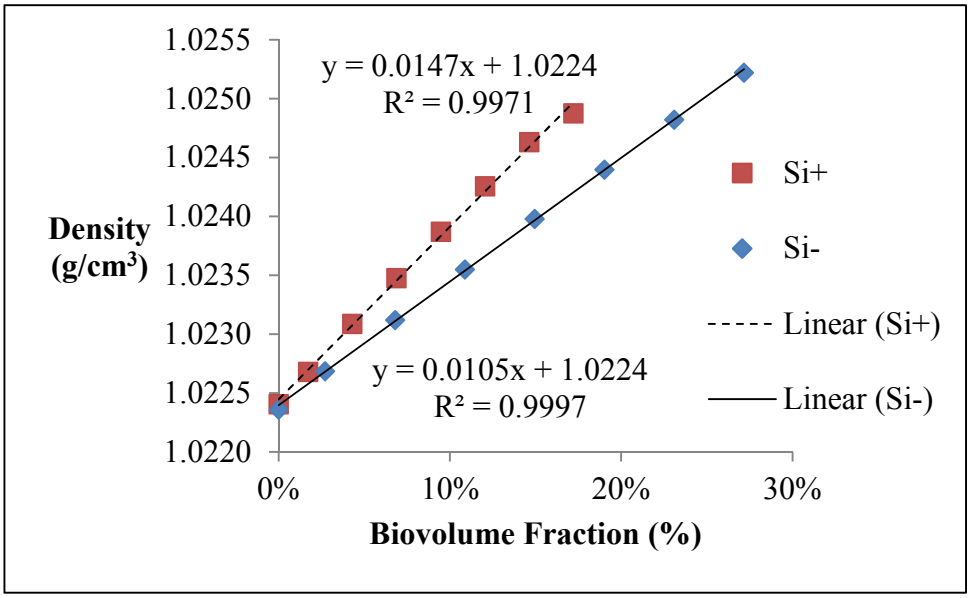

Figure A.4: Density vs. biovolume fraction of *Phaeodactylum* with or without silica.
Biovolumetric fractions represent volume/volume percentage of cells diluted in suspension. Cells grown in silica were denser than those grown without silica.

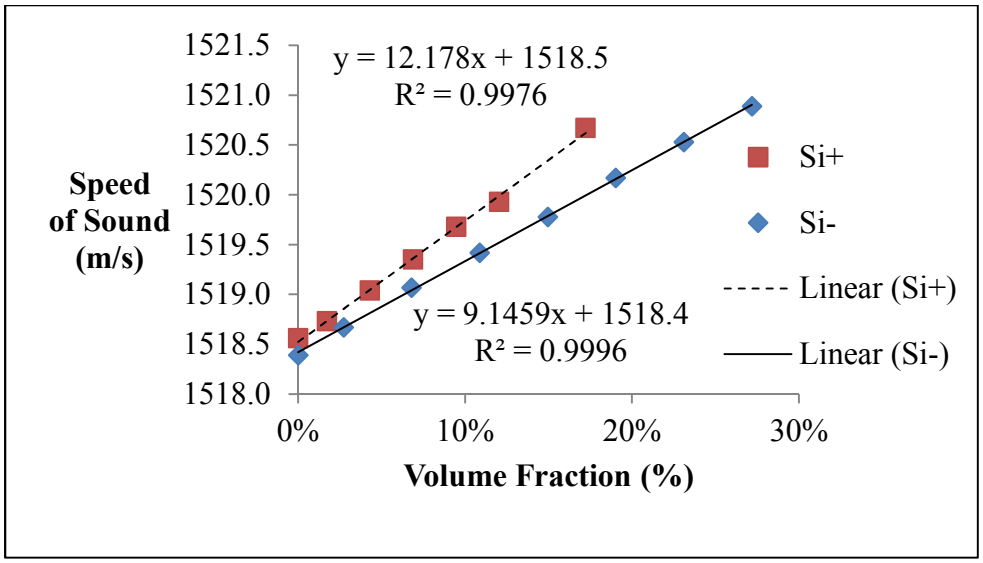

Figure A.5: Speed of sound vs. biovolume fraction of *Phaeodactylum* with or without silica. Biometric fractions represent volume/volume percentage of cells in suspension. The speed of sound of the emulsion of cells and media was greater in the f/2 media supplemented with silica.

Conclusions

- Cells grown with silica exhibited a reduced compressibility that permitted more of a mismatch between cells and the media and resulted in a larger acoustic contrast factor, $\mathcal{F}$.
- Given that $\mathcal{F}$ for exponentially growing *Chlamydomonas reinhardtii* is between 0.10-0.15 (E. Hincapie), 2-3 times more acoustic energy is required for *Phaeodactylum* to generate equivalent acoustophoretic outcomes, namely, linear alignment of cells at resonant nodes (Figure A.1, b).
- Biovolumes of nutrient replete cells are much greater than nutrient depleted cells (actual versus preliminary estimates of cell biovolumes).
- Modifications of cell culture media composition may indirectly change energy requirements to effectively perform acoustophoresis for a given algal culture.