STUDIES ON THE EFFECT OF CELL CYCLE ARREST ON CENTRAL METABOLISM IN THE DIATOM *PHAEODACTYLUM TRICORNUTUM*, USING PHYSIOLOGICAL AND SYSTEMS BIOLOGY APPROACHES

By

JOOMI KIM

A dissertation submitted to the

Graduate School- New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Oceanography

Written under the direction of

Paul G. Falkowski

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2016

ABSTRACT OF THE DISSERTATION

Studies on the effect of cell cycle arrest on central metabolism on the diatom Phaeodactylum tricornutum, using physiological and systems biology approaches

By JOOMI KIM

Dissertation Director:
Paul G. Falkowski

Diatoms (Bacillarophyceae) are photosynthetic unicellular microalgae that have risen to ecological prominence in the modern oceans over the past 30 million years. They are excellent candidates for biodiesel feedstocks. Global climate change has led to an interest in algal triacylglycerols (TAGs) as feedstocks for sustainable biodiesel, and diatoms are attractive candidates for TAG production as one of the most productive and environmentally flexible algae in the contemporary oceans. For Chapter 2, a genome-scale metabolic model was constructed to calculate intracellular fluxes of a diatom under different growth conditions. The model identified enzymes that may be relevant to increasing lipid synthesis, explored how transporters affect flux outputs, and explored unusual features of diatoms, including the Entner-Douderoff and phosphoketolase pathways, and glycolytic enzymes in

their mitochondria. Chapter 3 discusses how cell cycle arrest via cyclin-dependent kinase (Cdk) inhibition, can increase accumulation of TAGs, and shift metabolism away from protein synthesis. For Chapter 4, transcriptome analysis of cells under cell cycle arrest was performed to show that the pattern of gene expression was fundamentally different from nitrogen stress. Most of the genes related to fatty acid and TAG synthesis were up-regulated. The gene expression pattern for light harvesting complexes was similar to cells stressed by high light, suggesting that arrested cells have smaller sinks for photosynthetically generated electrons.

Acknowledgements

I would like to express my sincere gratitude to my advisor Paul G. Falkowski for his support of my Ph.D. studies; for his mentorship and knowledge. I am deeply grateful to Desmond S. Lun for his guidance and encouragement, especially during the final stages of this Ph.D. I would like to thank my committee members Debashish Bhattacharya, Kay D. Bidle, and Angela Falciatore, for their time, guidance, and expertise. I am grateful to all the collaborators and colleagues who provided technical support and many stimulating discussions: Chris M. Brown, Min Kyung Kim, James Kelley, Marie Huysman, Lieven De Veylder, Hilde Van den Daele, Michele Fabris, Gino Baart, Ilse M. Remmers, Alain Goossens, Wim Vyverman, Kuhn Ip, Harish Swaminathan, Xiao Qian, Miguel J. Frada, L. Tiago Guerra, Jorge Dinamarca, Orly Levitan, Kim Thamatrakoln, Udi Zelzion, Elizabeth H. Burrows, Kevin Wyman, Liti Haramaty, Benjamin Bailleul, Charles G. Dismukes, Joseph L. Dixon, Sang Hoon Lee, Aniça Amini, Rong Di, Michael Lawton, Maxim Gorbunov, Tali Mass, Frank Natale, Fedor Kuzminov, Raymond M. Remus, Douglas Campbell, Sandrine Ruchaud, Blandine Baratte, Stéphane Bach and Olivier Lozach. I thank Oscar Schofield, John Wilkin, Beatrice Birrer and Eric Lam for their various forms of support during my graduate studies. I thank my fellow (current and former) labmates for their camaraderie and support: John W. Harrold, Jeana Drake, John Kim, Ananya Agarwal and Ben Jelen. I am grateful to my family, for supporting me in all my pursuits. I would like to thank George T. Evageliou for his continuous support and encouragement throughout the final stages of this Ph.D. I also wish to acknowledge financial support from the Rutgers Department of Marine and Coastal Sciences,

EMBO, NSF IRES Grant, and the Rutgers-NSF IGERT Project, 'Solutions for Renewable and Sustainable Fuels for the 21st Century', NSF DGE O903675, Eric Lam, PI.

TABLE OF CONTENTS

List of Tables

List of illustrations

CHAPTER TWO

CHAPTER THREE

Chapter One

Introduction

Diatoms are unicellular, eukaryotic algae that first entered the fossil record about 150 million years ago, and rose to ecological prominence about 34 million years ago in the Oligocene (Falkowski et al. 2004). Over geologic time, they contributed significantly to the formation of petroleum deposits in shallow seas and on continental margins, and are among the most productive and environmentally flexible algae in the contemporary oceans. They are responsible for roughly 20% of global primary productivity, with greater carbon fixation ability than other co-existing microalgae, as measured in productivity per unit of crop carbon (Thomas et al. 1978). They are prominent in the ocean's food webs, responsible for supporting massive fisheries, krill, and ultimately large marine mammals (Berger et al. 2007).

Recently, global climate change has led to an interest in algal triacylglycerols (TAGs) as feedstocks for sustainable biodiesel. TAGs consist of three long chains of fatty acids attached to a glycerol backbone. They are one of the most energy-rich forms of reduced carbon available from nature, easily converted to fatty acid methyl esters (FAMEs), which can be used in modern diesel engines (Durrett et al. 2008). TAGS are one of the major carbon storage products in diatoms; under normal growth conditions, between 15 and 25% of their biomass is composed of lipids, though some strains can produce up to 60% of their ash-free dry weight as TAGs under the appropriate conditions (Sheehan et al. 1998). Fatty acids stored as TAGs are either

used for energy production through β–oxidation, or as substrates for phospholipid or membrane lipid synthesis. Diacylglycerols (DAGs) released by TAG degradation can also be used for synthesizing other lipid products, or act as messengers in signal transduction pathways (Becker et al. 2005). Theoretically, diatoms can serve as a biofuels production platform to satisfy the total oil consumption in the U.S., using between 3 and 5% of its land area (Levitan et al. 2014).

Though the depletion of phosphate triggers lipid accumulation (Valenzuela et al 2012), of the conditions that lead to increased TAG accumulation, nitrogen (N) starvation has been studied the most extensively. All algae have a requirement for fixed N, a necessary component of amino acids, nucleotides, and other cellular components. N deprivation leads to increased lipid accumulation, with a higher proportion as TAGs (Fogg et al. 1956, Badour and Gergis 1965, Opute 1974a, Shifrin and Chisholm 1981, Lynn et al. 2000). It also interferes with fundamental functions of the cell, such as the maintenance and repair of proteins, and leads to decreased protein and chlorophyll levels, reduced photosynthetic rates, changes in intermediate metabolite levels, and cell cycle arrest (Sheehan et al. 1998, Guerra et al. 2013, Levitan et al. 2015). These multiple effects make it unclear which changes lead to increased lipid accumulation.

Under nutrient replete conditions, most fixed carbon flows to pyruvate, which is decarboxylated to form acetyl-CoA (AcCoA). This AcCoA enters the tricarboxylic (TCA) cycle to form intermediate metabolites for anabolic pathways like amino acid

synthesis. Thus, under optimal conditions, carbon is preferentially allocated to proteins, and other components required for producing a duplicate cell. Under N deprivation however, the cell has a smaller sink for photosynthetically produced electrons, and excess electrons that accumulate in the photosynthetic electron transport chain induce over-production of reactive oxygen species, inhibiting photosynthesis and damaging membrane lipids and proteins. Under these conditions, N-deficient compounds, such as lipids or carbohydrates, serve to remove large amounts of photosynthetically-produced ATP and NAD(P)H. The formation of fatty acids consumes approximately 24 NADPH derived from the electron transport chain, which is twice that required for synthesis of a carbohydrate or protein molecule of the same mass.

More generally, evidence suggests that cell cycle arrest alone is sufficient for increasing TAG accumulation in algae. The cell cycle in all eukaryotes can be divided into four phases: G1, S, G2, and M. G1 and G2 refer to "gaps" in the cycle, while S refers to the phase during which DNA is replicated before mitosis (M) and cell division. Most diatoms have a requirement for silica (Si), the main component of their frustule or cell wall, and Si deprivation was found to lead to arrest at G1 or G2+M (Hildebrand et al. 2007), and an increase in the accumulation of total lipids, as well as the proportion of TAGs, with less physiological side-effects as N deprivation. Other examples of cell cycle arrest include early studies on batch cultures of the freshwater diatom, *Nitzchia palea,* in which it was reported that the presence of an autotoxin, secreted into the medium by aging cultures, blocked cell cycle

progression and resulted in increased lipid content (Opute 1974). In the green alga *Chlorella* CHLOR1, at high pH, autospore release (the stage after DNA replication) was suppressed, which increased the time for cell cycle completion and led to more TAG accumulation (Guckert et al. 1990). More recently, Hildebrand et al. 2012 found that treating diatoms with a microtubule inhibitor (nocodazole) was sufficient for them to increase TAG accumulation.

All methods of cell cycle arrest (nutrient deprivation, autotoxins, high pH, or microtubule inhibitors) lead to side effects that make it difficult to discern the precise causes of increased TAG accumulation in the cell. To refine our understanding of the mechanisms by which cell cycle arrest leads to increased TAG accumulation, Chapter 3 of this thesis examines the effects of cell cycle arrest by inhibiting a cyclin-dependent kinase, a protein involved in cell cycle processes; this avoids directly inhibiting an enzyme of metabolism or interfering with microtubule polymerization. Cyclin-dependent kinases, or Cdks, are serine-threonine kinases involved in regulating cell cycle progression, transcription, or mRNA processing, as well as differentiation in multicellular organisms. They regulate cell cycle progression by phosphorylating or activating other proteins involved in cell cycle progression or transcription. The study of Cdk inhibition could identify some of the connections between cell cycle regulation and metabolism; particularly how the cell partitions energy and “decides” how to allocate carbon skeletons under arrested vs. actively growing conditions.

The study of the "carbon decision tree," or how the cell allocates carbon skeletons under different conditions, would ideally benefit from knowledge of metabolic reaction fluxes. This requires studies using ^{13}C-labelled compounds, however, which are tedious and technically difficult. Alternatively, metabolic reaction fluxes can be approximated by a combination of different approaches, such as proteomic, transcriptomic, metabolomic studies, as well as metabolic modeling. Genomic sequencing has made it possible to reconstruct genome-scale metabolic networks of model organisms, or organisms of interest for biotechnological purposes. Flux balance analysis (FBA) is a widely used approach for studying metabolic networks. This approach takes all the known metabolic reactions in an organism, which can be based on genomic information or biochemical data, and calculates the fluxes of metabolites through this network using linear programming, knowledge of reaction stoichiometries, biomass composition, and additional constraints, such as limits on nutrient uptake or metabolite excretion rates. It can be used to predict growth rates, theoretical yields of metabolites under different conditions, and the viability of knockouts. FBA is used to guide the rational engineering of microorganisms; it can identify ways to engineer a microorganism to redistribute fluxes in its metabolism towards a desired end product or goal. This method has led to numerous successes in designing *E. coli* strains for the overproduction of a diverse range of metabolites, including ethanol (Pharkya et al., 2006), various amino acids (Pharkya et al., 2003, Lee et al., 2007), hydrogen, and vanillin (Pharkya et al., 2004), and *S. cerevisiae* strains to overproduce TCA metabolites (Zelle et al., 2008).

For Chapter 2, an FBA model was constructed for *P. tricornutum*. The metabolic network is based on the *P. tricornutum* genome, biochemical literature, online bioinformatic databases (Karp et al., 2000, Fabris et al., 2012), and extensive manual curation. It simulates global growth behaviors under autotrophic, mixotrophic, and heterotrophic conditions. The results from *in silico* predictions of the autotrophic and mixotrophic cases were validated with experimental evidence and data from previous studies. The goal was to provide further knowledge of this system, and serve as an interactive platform for integrating high-throughput data in the future, which could provide insights into the optimal way to engineer diatoms for bioproducts.

FBA is constrained by the stoichiometric coefficients of reactions; these constrain the flow of metabolites throughout the network. Every reaction can also be given upper or lower bounds that define the maximum or minimum allowable fluxes for each reaction. These constrain the allowable flux distributions of the system, or the rates at which metabolites are consumed or produced by each reaction (Orth et al. 2010). The metabolic state predicted by FBA is usually not unique for a given set of conditions- there is more than one optimal solution. Thus there are "missing constraints" in the model, and the more biologically plausible constraints that can be placed on the model, the smaller the set of optimal solutions, or the closer the model is to predicting accurate flux distributions. An attractive approach to constraining models is to use gene transcripts, since transcriptomic data is relatively easy to obtain. Though gene expression and metabolic fluxes do not always run in parallel, a

strong correlation is reported in microorganisms (Akesson et al. 2004, Famili et al. 2003).

In Chapter 4, transcripts from *P. tricornutum* cells were taken at different phases of the cell cycle, under normal growth conditions, and under the influence of a cell cycle inhibitor. This transcriptome data was also combined with the FBA metabolic network; this approach can yield cues for the likelihood that an enzyme in question supports metabolic flux in an associated reaction. These cues are then integrated to form a global, consistent metabolic behavior of the cell under these different conditions. Moreover, this process identifies those enzymes that do not fit into a consistent pattern, indicating that they are post-transcriptionally regulated (Shlomi et al. 2008).

CHAPTER TWO

Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*

Diatoms (Bacillarophyceae) are photosynthetic eukaryotic microalgae that constitute the most abundant and diverse group of marine eukaryotic phytoplankton. They account for roughly half of the ocean's primary productivity, or roughly 20% of the world's productivity (Nelson et al. 1995, Field et al. 1998). They are excellent candidates for sustainable biofuels feedstock production, not only because many strains are hardy, versatile, and efficient in carbon fixation, but also because their major storage products are lipids (Hildebrand et al. 2012, Levitan et al. 2015). In particular, interest in harnessing them for lipids as feedstocks for biodiesel or nutritional supplements has grown over the past decade, because they potentially can synthesize orders of magnitude more oil per hectare compared to terrestrial plants (Lebeau et al. 2003, Chisti 2007, Hu et al. 2008, Graham et al. 2011, Levitan et al. 2014). They also do not have to compete for agricultural land, and can be particularly environmentally beneficial when combined to treat wastewater (Hoffman 1998, Mehta et al. 2005, Clarens et al. 2010).

Sequencing has made it possible to reconstruct genome-scale metabolic networks of model organisms, or organisms of interest for biotechnological purposes. Flux balance analysis (FBA) is a widely used approach for studying metabolic networks. This approach takes all the known metabolic reactions in an organism, which can be

based on genomic information or biochemical data, and calculates the fluxes of metabolites through this network using linear programming, knowledge of reaction stoichiometries, biomass composition, and additional constraints, such as limits on nutrient uptake or metabolite excretion rates. It can be used to predict growth rates, theoretical yields of metabolites under different conditions, and the viability of knock-out or knock-in transformants. Furthermore, FBA can be used to guide the rational engineering of microorganisms; it can identify pathways in microbes that lead to efficient redistribution of metabolites towards a desired end product or goal. This method has been successfully applied to designing *E. coli* strains for the overproduction of a diverse range of metabolites, including ethanol (Pharkya et al. 2006), various amino acids (Pharkya et al. 2003, Lee et al. 2007), hydrogen, and vanillin (Pharkya et al. 2004), and in *S. cerevisiae* strains to overproduce TCA metabolites (Zelle et al. 2008). It has also been used to determine optimal gene knockouts to maximally increase respiration rates in *Geobacter sulfurreducens* in order to increase its bioremediation capabilities (Izallalen et al. 2008).

FBA has been used to model photosynthetic metabolism in the cyanobacteria, *Synechocystis* and *Arthrospira platensis*, and green algae, *Chlamydomonas reinhardtii* (Boyle et al. 2009, Klanchui et al. 2012, Knoop et al. 2013) and *Chlorella* sp. FC2 IITG (Muthuraj et al. 2013). However, to date only one FBA model of a stramenopile exists (Prigent et al. 2014). Stramenopiles, which include diatoms, chrysophytes (golden algae), brown algae, as well as some non-photosynthetic microorganisms, differ from green algae and cyanobacteria in cell structure and metabolism,

pigmentation, nutrient uptake and recycling, mixotrophy, the presence of a urea cycle, and mechanisms for carbon concentration (Kooistra et al. 2007). Diatoms in particular have some highly unusual features, including the presence of the Entner-Doudoroff pathway, connected to the lower half of glycolysis in the mitochondrion (Kroth et al. 2008, Fabris et al. 2012). Major questions remain concerning the origin and function of mitochondrial glycolysis. Traditionally, cytosolic glycolysis has been viewed as the original or ancestral form of eukaryotic energy metabolism (Whatley, John, and Whatley 1979, Blackstone 1995). Alternatively, it has been hypothesized that cytosolic glycolysis originated from the eubacteria that would eventually become mitochondria, and that mitochondrial glycolysis is an ancient eukaryotic trait (Martin and Muller 1998). Evidence for the latter is supported by the observations that (1) diatoms and non-photosynthetic oomycetes have a mitochondrial glycolytic pathway, and (2) all glycolytic enzymes of the eukaryotic cytosol analyzed to date are more similar to eubacterial homologs than to homologs in archaebacteria- the lineage that contains ancestral eukaryotes (Martin et al. 1993, Keeling et al. 1997, Liaud et al. 2000).

Given that diatoms (1) possess highly unusual features that may be relics of ancient eukaryotic metabolism, which have bearing on questions regarding the origins of eukaryotic glycolysis, and (2) are excellent candidates for biodiesel feedstocks, the absence of a genome-scale model of a diatom represents a large gap in the knowledge base. This study presents a genome-scale metabolic model of the pennate diatom *P. tricornutum*. This organism was chosen because it is sequenced,

and data on biomass composition are available (Bowler et al. 2008; Frada et al. 2012; Guerra et al. 2013). The metabolic network is based on the *P. tricornutum* genome, biochemical literature, online bioinformatic databases (Karp et al. 2000, Fabris et al. 2012), and extensive manual curation. The model simulates global growth behaviors under autotrophic, mixotrophic, and heterotrophic conditions, as well as knockout conditions that explore the role of lower glycolysis in the mitochondrion. Although *P. tricornutum* is not normally capable of growing solely on glucose in the dark, this condition was simulated as a way to understand the organism either under genetic modification (Zaslavskaia et al. 2001) or night-time metabolism. Mixotrophic growth conditions are of special interest for biofuel production (Wang et al. 2012). The results from *in silico* predictions of the autotrophic case were validated with experimental evidence from previous studies, and results for the mixotrophic case were compared with data from a recent isotope labeling study (Zheng et al. 2013). The goal of the model was to provide further knowledge of this system, as well as an interactive platform for integrating high-throughput data in the future, which can provide insights into the optimal way to engineer diatoms for bioproducts.

Results and Discussion

A stoichiometric model of metabolism was constructed for *P. tricornutum* from genomic databases, and the literature. The network includes glycolysis, the TCA cycle, oxidative and reductive pentose phosphate pathways, the phosphoketolase

pathway, amino acid, nucleotide, chlorophyll, chrysolaminarin, and lipid synthesis pathways.

Network Construction

The metabolic network of *P. tricornutum* consists of 587 metabolites and 849 metabolic reactions (not including the biomass equation). The complete model can be found in Appendix A:5. Abbreviations for metabolites are in Appendix A:6. For many of the pathways in *P. tricornutum*, most of the enzymes of the pathway were annotated, but with "gaps" (missing enzymes). In these cases, the "missing enzymes" were included in the model, on the assumption that the alga had the complete pathway, but the enzymes were missed in annotation. In the course of making the model, several missing enzymes were identified and given a "confidence score" of 1, based on the convention of Thiele et al. 2010. A confidence score of 1 indicates that there is no genomic evidence for the enzyme, but the enzyme is included in the model so that known pathways could be completely included.

Localization of enzymes and metabolites

The reactions of the network were localized into four compartments: cytosol, mitochondria (matrix), chloroplast, and peroxisome. The chloroplast was further divided into the stroma and lumen. Though the mitrochondrion is composed of an intermembrane space and matrix, the intermembrane space is chemically equivalent to the cytosol; "mitochondrion" therefore refers to the matrix here. Subcellular localization was obtained using software from TargetP (Emanuelsson et

al. 2000) or Mitoprot (Claros et al. 1996). When no signaling peptide was predicted, the recommendation in Thiele et al. 2010 was followed, and the enzyme was designated as cytosolic. In some cases, when these programs did not predict an enzyme to have a signal peptide but all the enzymes of its substrates and products were known to be in a certain compartment, the enzyme was placed in that compartment.

Biomass formation equations

The macromolecular composition of cells from each growth condition was based on experiments performed in chemostats, as well as data from the literature (Table 1). It consisted of amino acids, nucleotides, lipids, chlorophyll, and carbohydrates.

Table 1. Biomass formation: standing stocks of cell constituents

	moles/kg biomass
DNA	0.0009
RNA	0.001
Carbohydrate	0.15
Chlorophyll A	0.01
Growth Associated ATP	29.89
Non-Growth Associated ATP	0.36-7.60
Monogalactosyl-diacylglycerols	0.18
Triacylglycerols	0.14
Digalactosyldiacylglycerols	0.07
Sulphoquinovosyldiacylglycerols	0.10
alanine	0.15
arginine	0.07
aspartic acid	0.12
cysteine	0.01
glutamic acid	0.14
glycine	0.15
histidine	0.02

isoleucine	0.07
leucine	0.11
lysine	0.07
methionine	0.02
phenylalanine	0.08
proline	0.10
serine	0.11
threonine	0.09
tyrosine	0.04
valine	0.09
tryptophan	0.01
ornithine	0.02
asparagine	0.08
glutamine	0.51

Table 1. The coefficients for the equation describing the formation of biomass are given in units of moles per kg biomass (dry weight). Total amino acid, carbohydrate, lipid, and chlorophyll were measured. DNA was calculated from knowledge of genome size (Bowler et al., 2008). Total RNA was from Fernandez-Reiriz et al., 1989. The components of total lipid were taken from Yongmanitchai et al., 1992. The overall amino acid content was based on measurements, but the amino acid composition was taken from Marsot et al., 1991 and Brown et al., 1991. Growth-associated ATP maintenance was estimated using a *C. reinhardtii* model (Boyle et al., 2009).

Model Constraints

Flux maps for growth under autotrophic, mixotrophic, and heterotrophic conditions were calculated using the network reconstruction and FBA (Figures 1a-d). During autotrophic growth, *P. tricornutum* fixes carbon dioxide and bicarbonate, and uses water as the source of reductant, light as the source of energy, and micronutrients as essential components of the catalytic reactions. In order to constrain the model in

this condition, a maximum uptake rate for CO_2 was imposed using the value 1.3 mmol CO_2/g dry wt/h, and the HCO_3 uptake rate of 0.65 mmol/g dry wt/h, or a maximum uptake of 1.95 mmol/g dry wt/h for total inorganic carbon (Ci); these values were based on measurements of uptake rates in *P. tricornutum* cells during steady state photosynthesis (Burkhardt et al. 2001). The mixotrophic conditions, where both photoautotrophy and glucose consumption occur, were as follows: 1. growth with glucose in the light, with no CO_2 or bicarbonate (Ci), and 2. growth with glucose in the light, with Ci. A heterotrophic condition was also modeled: growth on glucose in the dark. In all cases, the input of carbon (number of carbons) was equal to the input of carbons of the autotrophic case to facilitate direct comparisons between conditions; the maximum glucose uptake rate for mixotrophic case 1 and the heterotrophic case was therefore set to 0.325 mmol/g dry weight/h, which is the Ci uptake rate of the autotrophic case (1.95 mmol/g dry weight/h) divided by the number of carbons (six) in glucose. For mixotrophic case 2, half the carbon came from inorganic carbon and half from glucose.

Simulations of heterotrophic or mixotrophic growth can be seen as tools to assess the metabolic capabilities of this organism with different genetic backgrounds. Some wild-type *P. tricornutum* can utilize glucose when grown in the light (Zheng et al. 2013), though different strains have different efficiencies of glucose metabolism, while some cannot use glucose at all (Huang et al. 2015). Wild type *P. tricornutum* cannot grow on the sugar heterotrophically. By inserting a human glucose transporter, Zaslavskaia et al. 2001 engineered a *P. tricornutum* strain that was able

to grow in the dark as the sole carbon source. The model simulations for the heterotrophic condition can therefore be viewed as either a simulation of night-time metabolism, or growth in a different genetic background.

The biomass composition was set to be the same across all conditions (based on the autotrophic case). However, since biomass likely differs for mixotrophic and heterotrophic conditions in actual cultures, a sensitivity analysis was performed whereby different versions of the mixotrophic and heterotrophic conditions were tested, using different biomass values of protein, carbohydrate, lipid and chlorophyll A, and RNA (Appendix A:7). DNA was kept the same. In total, three versions of the mixotrophic biomass, and three versions of the heterotrophic biomass were tested. Where available in the literature, data for *P. tricornutum* was used; if this was not available, data from other diatoms was used. For the mixotrophic cases, data for protein and carbohydrate was not available from diatom studies; for these cases, *C. reinhardtii* data was used. The three versions of heterotrophic biomass were based on diatom data.

Several versions of mixotrophic and heterotrophic cases were tested to cover multiple possibilities for biomass and address various discrepancies in the literature; for example, the studies by Liu et al. 2009 and Ceron Garcia et al. 2005 reported that mixotrophic *P. tricornutum* had about 14% and 64% less chlorophyll, respectively, compared to autotrophically grown cultures. A study with the diatom *Chaetoceros* sp. (Cheirsilp et al. 2011) reported 4% less lipid, while a study with the

diatom *Cyclotella cryptica* (Pahl et al. 2011) reported 38% more lipid in heterotrophically grown cultures, compared to autotrophic cultures. Versions 3 for mixotrophy and heterotrophy were the same as versions 1 except that RNA was doubled; this was tested because there was no available data for changes in RNA under mixotrophy or heterotrophy.

The sensitivity analysis revealed that changes in relative amounts of lipid, carbohydrate, chlorophyll, protein, or RNA did not change the overall flux pattern (the reactions that carried flux, or the directions of reversible reactions), but did change the amount of flux through those reactions slightly (see S3). The overall growth rate also varied, depending on which biomass equation was used, though the difference never varied by more than 16% of the original. As a result, it was expected that the model would reliably predict the overall flux distribution pattern, though it is not expected to predict the specific quantitative values of individual fluxes.

Autotrophic Simulation Results

Under autotrophic growth, the model predicts that the doubling time of the cell is 13.9 h, which is within the range of our experimental observations: 11.9 - 17.7 h (Appendix A:1). The fluxes with the largest magnitudes are found in the reactions of the Calvin cycle and photophosphorylation (Fig.1a). Photophosphorylation provides the energy required for regenerating glyceraldehyde-3-phosphate (g3p) from glycerate-3-phosphate (gly3p). Since there are no known direct NAD(P)H

transporters between the plastid and the cytosol, reducing equivalents in the form of glyceraldehyde-3-phosphate are transferred from the plastid to meet the demands for NAD(P)H in the cytosol. In contrast, ATP in the cytosol can be obtained through ATP/ADP translocators between the cytosol and mitochondrion, as well as through ATP-producing reactions in the cytosol.

Fig. 1a-

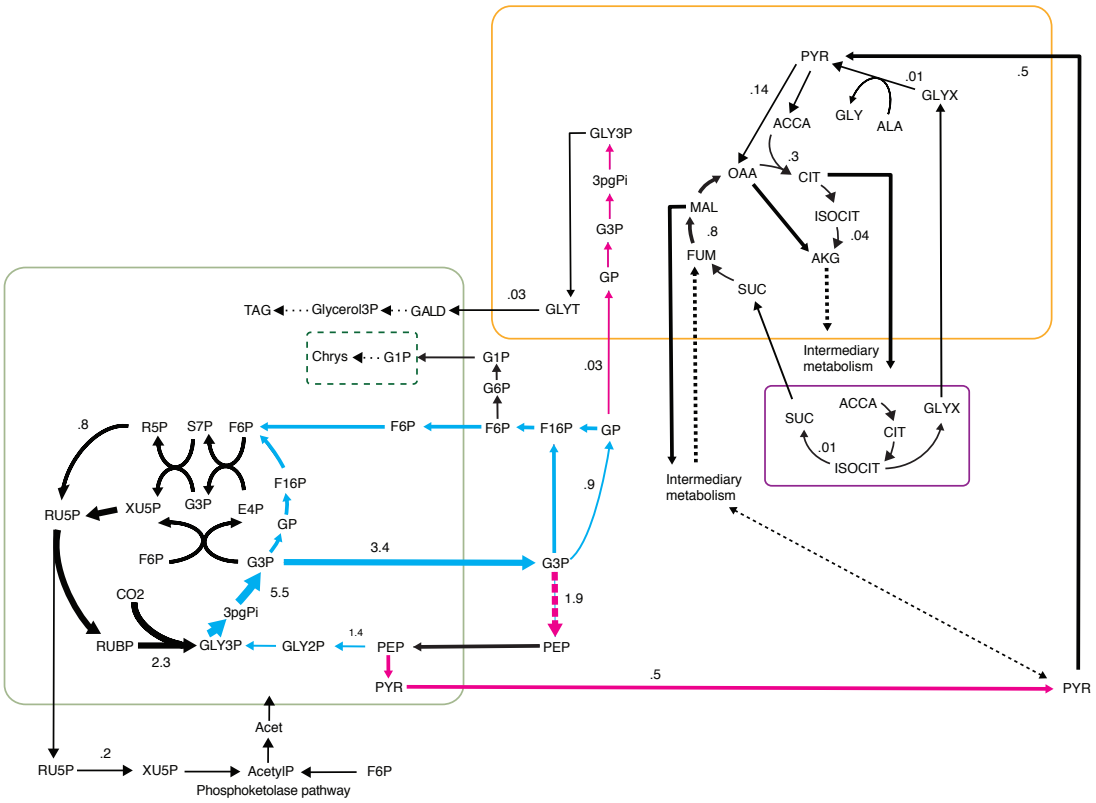


Fig. 1. Metabolic flux map for *P. tricornutum* under a. autotrophic growth conditions. The green rectangle represents the chloroplast, the orange rectangle represents the mitochondrion, the purple rectangle represents the peroxisome, and the dashed green line in the plastid represents a separate compartment for chrysolaminarin. Blue arrows are reactions that run in the direction of gluconeogenesis, and pink reactions run in the

direction of glycolysis. The thickness of the arrows correspond to relative amount of flux. Abbreviations for metabolites are in Appendix A:6.

Light drives linear electron flow (LEF) from water to NADPH by the two photosystems (PSI and PSII), and pumps protons to generate an electrochemical proton gradient across the thylakoid membrane, which in turn generates ATP in the plastid. If linear electron flow were the only process to produce ATP and NADPH, this would generate an ATP/NADPH ratio that is insufficient to fuel CO_2 assimilation by the Calvin cycle (Allen 2002). In Viridiplantae (green algae and land plants), cyclic electron flow (CEF) around PS I plays a large role in making up the shortfall of ATP. In diatoms, however, the PSII to PSI ratio is two or more, and CEF around PSI is not easily accommodated.

Growth and maintenance ATP requirements were included to account for the unknown ATP requirements associated with transport, biosynthesis, polymerization, as well as cell maintenance, homeostasis, and repair. In *C. reinhardtii* growth-associated maintenance is estimated to be about 29.89 mmol ATP/g dry wt/h (Boyle et al. 2009). Non-growth associated ATP requirements (maintenance ATP) can range from 0.36 to 7.60 mmol ATP/g dry wt/h for *Lactobacillus plantarum* and *E. coli* respectively, and in *C. reinhardtii* it is reported to be about 1.50 mmol ATP/g dry wt/h (Reed et al. 2003, Boyle et al. 2009). Since these values are not known for *P. tricornutum*, the default growth and non-growth associated maintenance requirements were set to be the same as that of *C.*

reinhardtii, or 29.89 mmol ATP/g dry wt/h and 1.50 mmol ATP/g dry wt/h, respectively. With this ATP maintenance requirement, the model predicts that in *P. tricornutum*, ATP is synthesized in the plastid entirely through linear electron flow. That is, there is no CEF around PS I. This is in agreement with the experimental results of Bailleul et al. 2015, which reported that CEF played a negligible role in regulating ATP/NADPH levels in diatoms. More generally, they found that reducing equivalents generated in the plastid were exported to the mitochondria to generate additional ATP; a partial re-routing of photosynthate flowed towards mitochondrial respiration to optimize the ATP/NADPH ratio in the plastid. This is in agreement with the model, which predicts that the flux through mitochondrion ATP synthase is about 1.2 times higher than through ATP synthase in the plastid.

In order to investigate how changes in ATP maintenance would change the predicted flux distributions, as well as flux through CEF, the autotrophic condition was tested using two extremes for the non-growth ATP maintenance value, 0.36 and 7.60 mmol ATP/g dry wt/h (the ATP maintenance values for *Lactobacillus plantarum* and *E. coli* respectively), while keeping the growth-associated maintenance as 29.89 mmol ATP/g dry wt/h. The results show that changing the ATP maintenance value causes the amount of flux for individual reactions to change, though the reactions carrying flux, and the overall flux distribution, do not change. However, at the highest non-growth associated ATP maintenance value of 7.6 mmol ATP/g dry wt/h, there is a small flux through CEF. More generally, flux through CEF

appears once the non-growth ATP maintenance is set to about 7.4 mmol ATP/g dry wt/h or higher (Table 2). Therefore, *P. tricornutum* is predicted to have zero CEF only so long as the total ATP maintenance (growth and non-growth associated) is lower than 29.89 + 7.4 mmol ATP/g dry wt/h.

Table 2. The effect of changes in ATP maintenance

	Non-growth Associated ATP Maintenance (mmol/g dry wt/h)				
	0.36	1.5	7.4	7.6	1.5 (minimum PSII)
PSII	6.60	6.85	8.74	8.79	5.33
PSI	6.60	6.85	8.75	8.84	19.13
CEF	0.00	0.00	0.01	0.05	13.80

Table 2. The effect of changes in ATP maintenance on PSI, PSII, and CEF flux (mmol/gdry wt/h).

The results of the model simulations also show that as ATP maintenance is increased, the cell absorbs more light, and O_2 consumption for aerobic respiration increases linearly with O_2 production by PSII (Fig. 2). This is in agreement with the study by Bailleul et al. 2015, which found that O_2 consumption increased linearly with O_2 production; a constant proportion of the electron flow from photosynthesis was re-routed for O_2 consumption, regardless of light intensity.

Fig. 2-

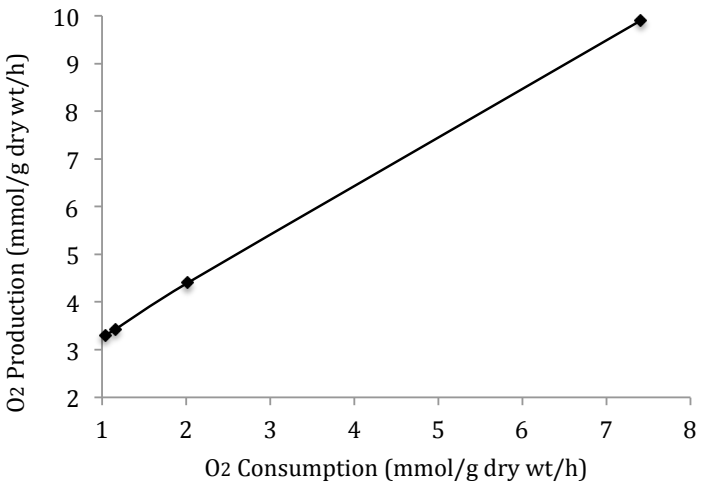


Fig. 2. O_2 consumption for aerobic respiration increases linearly with O_2 production by PSII. The points are (from left to right) for ATP maintenance values 0.36, 1.5, 7.6 and 50 mmol/g dry wt/h.

Additionally, mitochondrial respiration is directly involved in optimizing photosynthesis. Maintaining an optimal ATP/NADPH ratio should therefore be contingent on having sufficient communication between the plastid and mitochondria. The model results are in agreement with this: If the mitochondrial ADP/ATP translocator (which exports ATP out of the mitochondria in exchange for

ADP) is removed, there is flux through CEF, and the flux through ATP synthase in the plastid increases by about 1.7 times (Table 3). Moreover, in agreement with experimental results of Bailleul et al. 2015, which found that mitochondrial inhibition negatively impacted photosynthesis, the model predicts that deletion of one of the enzymes of the electron transport chain in the mitochondria (such as cytochrome C oxidase or coenzyme Q-cytochrome C reductase), lowers flux through PSII by 22%. If the plastid ADP/ATP translocator (49533), which can directly import ATP into the plastid, is deleted, there is flux through CEF, and the flux through ATP synthase in the plastid increases by about 109%. If the majority of transporters that transfer metabolites (transporters for glyceraldehyde-3-phosphate, pyruvate, aspartate and oxaloacetate) in or out of the plastid are also deleted, the model predicts that flux through CEF increases further, and ATP synthesis in the mitochondrion decreases (Table 3).

Table 3. Knockouts in the mitochondria and plastid affect flux through PSI, PSII, CEF, and ATP synthases.

	Wild Type	No mitochondrial ATP/ADP translocator	No mitochondrial complex IV	No mitochondrial complex III	No plastid ATP/ADP translocator	No plastid transporters, or plastid ATP/ADP translocator
PSII	6.9	5.5	5.3	5.3	6.4	6.2
PSI	6.9	18.3	19.1	19.1	14.3	16.0
ATPS_P	3.4	9.2	9.6	9.6	7.2	8.0
CEF	0.0	12.8	13.8	13.8	7.9	9.8
ATPS_M	7.7	0.7	0.1	0.1	3.3	2.1

Table 3. Predicted fluxes (mmol/gdry wt/h) for hypothetical knockouts in the mitochondria and plastids, on PSII, PSI, ATP synthase in the plastid (ATPS_P), cyclic electron flow (CEF), and ATP synthase in the mitochondria (ATPS_M).

In the model, transporters connect the plastid to the cytosol, which is connected to the mitochondrion by its own set of transporters. The plastidic transporters in this organism remain to be determined, but *P. tricornutum* is likely to at least have a triose phosphate translocator (Weber et al. 2006). The model contains a transporter for glyceraldehyde-3-phosphate, as well as transporters that are known in plants (Fischer et al. 2011). With these transporters, pyruvate and glyceraldehyde-3-phosphate are sent out of the plastid; the glyceraldehyde-3-phosphate is converted to dihydroxyacetone in the cytosol, and sent into the mitochondrion (Fig. 1a). Pyruvate, fumarate and oxaloacetate enter the mitochondrion from the cytosol. The mitochondrion then releases 2-oxoglutarate and malate, as well as a small amount of glycerate, which makes its way to the plastid. The flux distribution of metabolites is contingent on which transporters exist, but even in a highly conservative approach, where only the transporters for glyceraldehyde-3-phosphate, glutamate, and 2-oxoglutarate are included, the flux distribution shows a partial re-routing of dihydroxyacetone flowing towards the mitochondria. Given that it is highly likely that the majority of transporters in this organism remain to be discovered, the actual flux pattern likely looks closer to the flux distribution produced with the greater number of transporters.

The flux distribution solution from FBA is not unique; there are multiple solutions or flux distributions that can maximize biomass production. In order to constrain the model output, an optimization function, Euclidean norm minimization (discussed in *Calculation of Flux Distributions*), was used in conjunction with FBA to choose the (unique) flux vector with the lowest overall magnitude that still produces maximum biomass flux. The rationale for this optimization function is based on the assumption that fluxes for metabolic reactions in organisms are not larger than necessary for biomass production. The predictions for CEF are partially a result of this optimization function. Indeed, flux variability analysis (FVA), which calculates the maximum and minimum allowable values of each flux while still achieving maximal biomass flux, shows that with ATP maintenance of 1.5 mmol/g dry wt/h, CEF can vary over a relatively large range: 0 - 489 mmol/g dry wt/h (S3). To a degree, LEF can trade off with CEF. If an artificial limit is placed on the LEF by limiting PSII to 5.33 mmol /g dry wt/h, which is the minimum allowable PSII flux (obtained via FVA), CEF flux increases from 0 to 13.8 mmol/g dry wt/h, for an ATP maintenance value of 1.5 mmol ATP/g dry wt/h (Table 2). More generally, when ATP maintenance is lower than 7.4 mmol ATP/g dry wt/h, increased demand for ATP is met though higher flux through LEF. For values higher than 7.4 mmol ATP/g dry wt/h, increased ATP demand is met by increased flux through both LEF and CEF (Table 2).

This is only part of the explanation for the CEF results, however. Since the optimization function also picks solutions with the shortest pathways, the question

remains why such a large proportion of the plastid's ATP is synthesized in mitochondria, which would require a longer pathway than simply using CEF to optimize NADPH/ATP levels in the plastid (which would require absorbing more photons, but is not in limited supply in the model). The reason for this may lie in the fact that sending reductant into the mitochondrion is a way to provide carbon skeletons in the mitochondrion for use in anabolic pathways, in addition to facilitating mitochondrial ATP production. As such, it is more "efficient" than simply using CEF to meet ATP demands, at least when ATP maintenance (growth and non-growth associated) is lower than 29.89 + 7.4 mmol ATP/g dry wt/h. Additionally, photosynthetically-derived dihydroxyacetone can be sent into the mitochondria, where the lower half of glycolysis is present in *P. tricornutum*; this produces NADH and ATP in the mitochondria independent of the TCA cycle, possibly allowing for more precision in optimizing fluxes for specific metabolite composition. CEF also plays a large role in adjusting the ATP/NAD(P)H ratio in plants and green algae; mitochondrial ATP synthesis is sometimes "preferred" over CEF. For example, in pea leaves, mitochondrial ATP synthesis is "chosen" over CEF, which mainly takes effect only at high irradiances combined with very low CO_2 concentrations (Harbinson et al. 1991). More generally, the results are in agreement with numerous studies that demonstrate a highly interdependent relationship between plastids and mitochondria in plants and green algae (Hoefnagel et al. 1998, Matsuo et al. 2006). *P. tricornutum* is no exception; in fact, electron micrographs show that plastids and mitochondria in this organism are tightly associated physically (Bailleul et al. 2015).

Heterotrophic and Mixotrophic Simulation Results

Under heterotrophy, the doubling time of the cell is predicted to be about 21 hours, or about 54% longer than for the autotrophic case. In cultures, algae grown heterotrophically can sometimes have higher growth rates than under autotrophy, since light shading effects do not matter, and cells can grow to greater density. When normalized to the carbon input, however, heterotrophic growth is less efficient, and models for several different autotrophs predict this (Boyle et al.,2009, Muthuraj et al. 2013, Shastri and Morgan 2005). Under heterotrophy, some carbon is lost due to respiration; there is net release of CO_2, so less carbon is available for biomass production than under autotrophy. Under autotrophy, on the other hand, CO_2 produced from respiration can be recycled.

The flux distribution for the heterotrophic case starts with the uptake of glucose, which is converted to glucose-6-P in the cytosol (Fig. 1b). In *P. tricornutum*, the complete set of glycolytic reactions is found in both the plastid and cytosol, and the lower half of these reactions is found in the mitochondrion. Under heterotrophic growth, these reactions run in the direction of ATP or NAD(P)H formation in the plastid (pink reactions in Fig.1b); this provides energy required for lipid synthesis in the plastid. In general, the highest fluxes under heterotrophy are found in the electron transport chain and ATP synthase in the mitochondrion.

Fig. 1b-

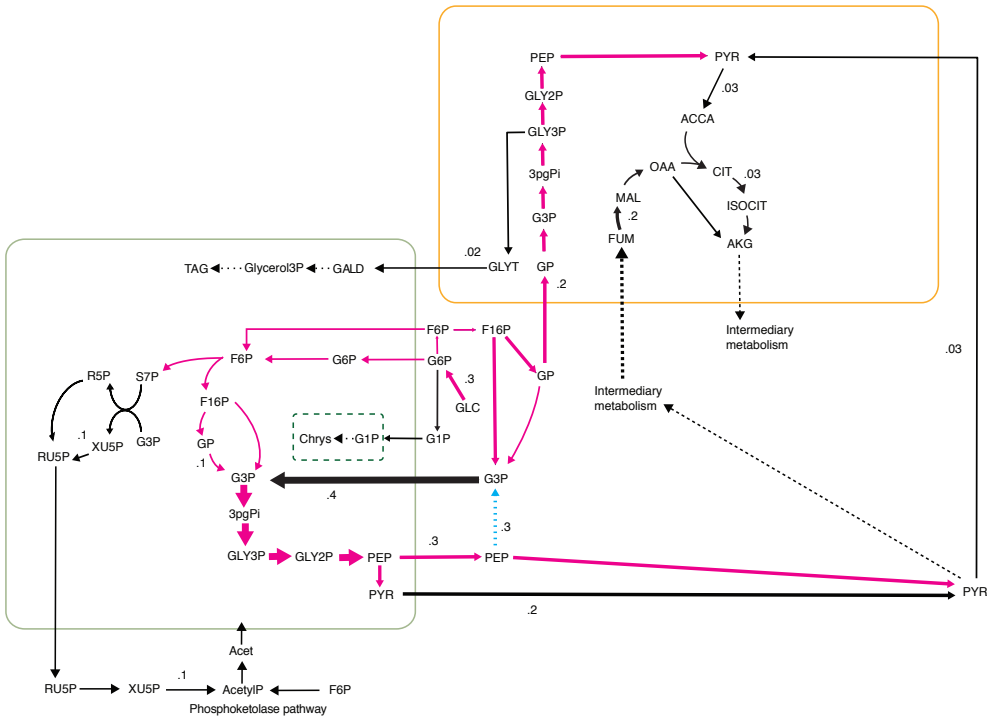

PEP
PYR
.03
GLY2P
ACCA
GLY3P
OAA
CIT
.03
3pgPi
MAL
ISOCIT
G3P
.2
FUM
AKG
.02
GLYT
GP
.2
TAG
Glycerol3P
GALD
Intermediary metabolism
F6P
F16P
.03
F6P
G6P
G6P
.3
GP
R5P
S7P
GLC
F16P
Intermediary metabolism
.1
XU5P
G3P
GP
Chrys
G1P
RU5P
.1
G3P
G3P
.4
3pgPi
.3
.3
GLY3P
GLY2P
PEP
PEP
PYR
PYR
.2
Acet
RU5P
XU5P
.1
AcetylP
F6P
Phosphoketolase pathway

Fig. 1c-

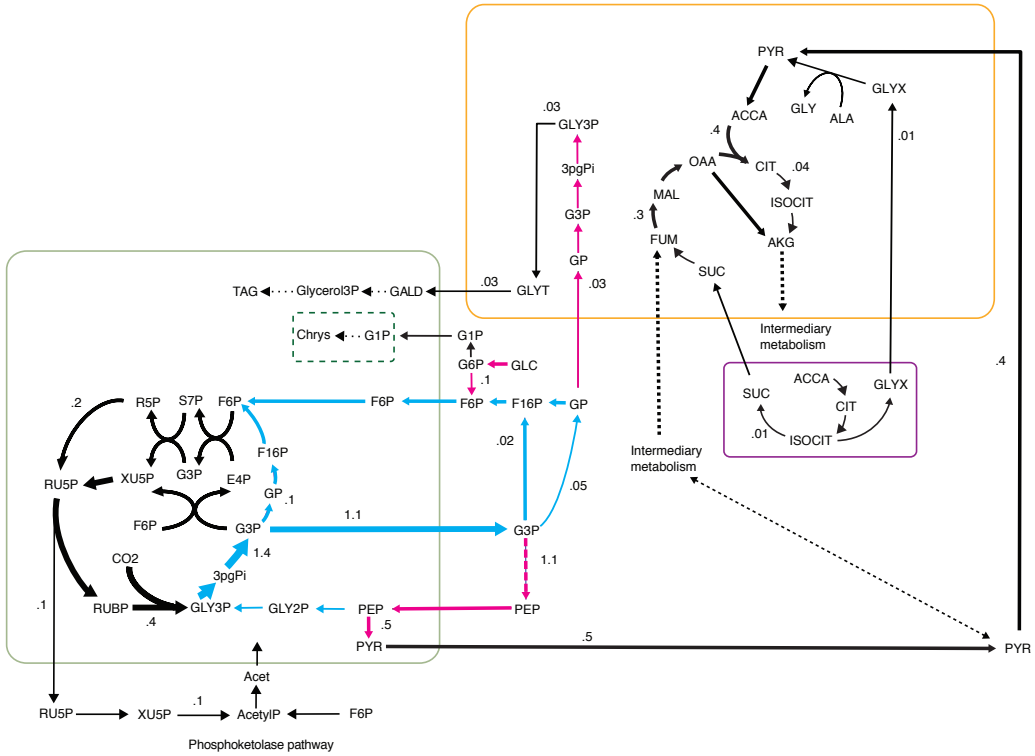

PYR
GLYX
.03
GLY3P
GLY
ACCA
ALA
.01
.4
OAA
CIT
.04
3pgPi
MAL
ISOCIT
G3P
.3
FUM
AKG
GP
SUC
.03
.03
GLYT
TAG
Glycerol3P
GALD
Intermediary metabolism
Chrys
G1P
G1P
G6P
GLC
.4
.1
SUC
ACCA
GLYX
.2
R5P
S7P
F6P
F6P
F6P
F16P
GP
CIT
.02
.01
ISOCIT
RU5P
XU5P
G3P
E4P
F16P
Intermediary metabolism
.05
GP
.1
F6P
G3P
1.1
G3P
CO2
1.4
1.1
3pgPi
.1
RUBP
GLY3P
GLY2P
PEP
PEP
.4
.5
PYR
.5
PYR
Acet
RU5P
XU5P
.1
AcetylP
F6P
Phosphoketolase pathway

Fig. 1d-

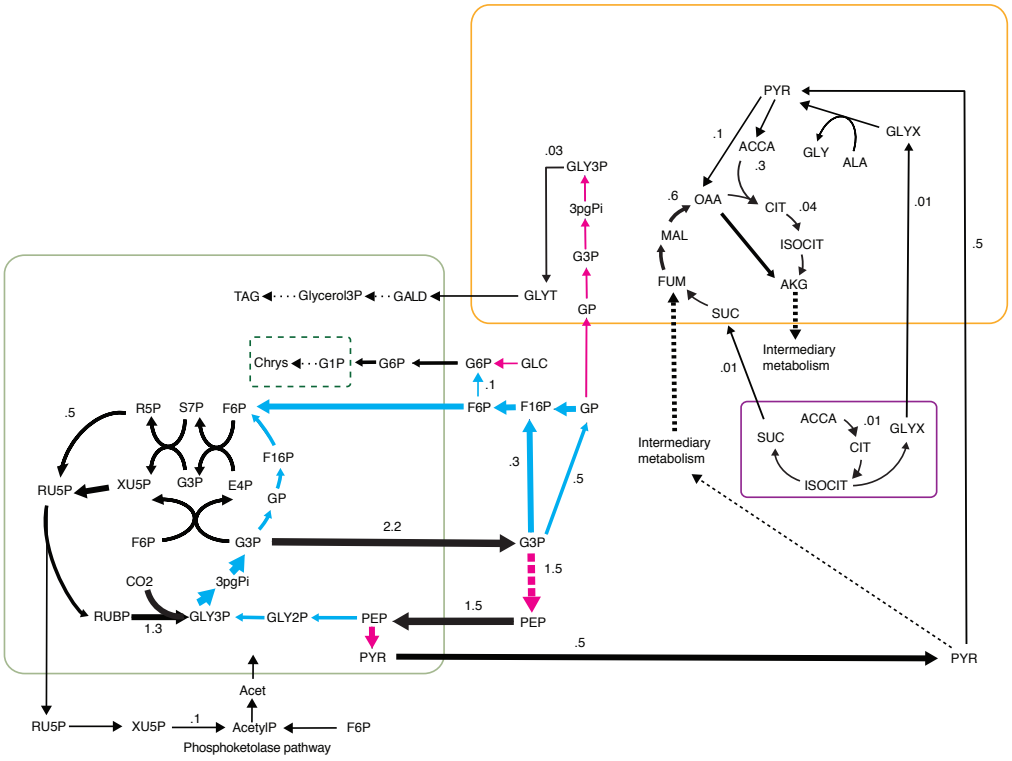


Fig. 1. Metabolic flux map for *P. tricornutum* under b. heterotrophic (dark) conditions, c. mixotrophic (glucose and light) conditions, and d. mixotrophic (glucose, light, and inorganic carbon) conditions. The green rectangle represents the chloroplast, the orange rectangle represents the mitochondrion, the purple rectangle represents the peroxisome, and the dashed green line in the plastid represents a separate compartment for chrysolaminarin. Abbreviations for metabolites are in Appendix A:6.

If a mixotrophic case is run with the same amount of Ci as the autotrophic case (1.95 mmol/gdry wt/h), plus extra carbon in the form of glucose, biomass production increases (or the doubling time decreases). If 0.325 mmol/gdry wt/h glucose is supplemented to the 1.95 mmol/gdry wt/h Ci, the doubling time is 6.9 hours, or

roughly half the doubling time of the autotrophic case. When mixotrophic conditions are set so that the carbon input is the same as in the autotrophic case, whether it is 1-growth with glucose in the light, without Ci (Fig. 1c), or 2- growth with glucose in the light with Ci (Fig. 1d), the doubling times are the same as in the autotrophic condition. The flux distributions for these two versions of mixotrophy are similar, though the flux through the Calvin cycle in case 1 is lower, because more of the initial carbon is in organic form. Additionally, the fluxes are opposite for glucose-6-phosphate isomerase (E.C. 5.3.1.9), since they have differing glucose inputs. In general, it is unsurprising that the flux distributions are similar for these two cases, since CO_2 produced in the mitochondrion can be recycled and moved into the plastid to fuel RuBisCO and carbon fixation.

P. tricornutum is predicted to have the components of the glycerol-3-phosphate shuttle. The shuttle adjusts the NADH/NAD+ ratio in the cytosol since it converts cytosolic NADH into NAD+, and contributes to oxidative phosphorylation in the mitochondrion. The model results predict that this shuttle is active in the mixotrophic conditions, possibly due to the fact that a larger proportion of cytosolic glyceraldehyde-3-phosphate (G3P) is turned into glycerate-1,3-diphosphate, via glyceraldehyde-3-phosphate dehydrogenase (which produces NADH and consumes NAD+), compared to the autotrophic case. In many organisms, mechanisms like the malate-aspartate shuttle also function to regenerate NAD+ in the cytosol, but this requires a malate dehydrogenase in the cytosol, which has not been found in *P. tricornutum*.

All the light conditions (autotrophy and mixotrophy) also had flux through either the mitochondrial or plastidic pyruvate carboxylase (6.4.1.1), which catalyzes the carboxylation of pyruvate to form oxaloacetate. This agreed with labeling experiments in *P. tricornutum* grown in the light, on media with inorganic carbon supplemented with U-^{13}C glucose (Zheng et al. 2013). The labeling experiments revealed that carbon originating from U-^{13}C glucose was uniformly distributed across the metabolic network and the majority of amino acids showed nearly uniform ^{13}C enrichments, but some fragments derived from oxaloacetate and 2-oxoglutarate, that contained carbon fixed through anaplerotic reactions, were less enriched, implying that the carbon assimilated originated from dissolved inorganic carbon.

Other Predictions

The TCA cycle is not predicted to run in a complete circle for all conditions, and the highest fluxes were between fumarate, malate, and oxaloacetate. Under autotrophy, the majority of this oxaloacetate (about 85%) is then converted to 2-oxoglutarate and aspartate via aspartate aminotransferase (EC 2.6.1.1). The TCA cycle not only fuels ATP synthesis in the mitochondrion, but also provides carbon skeletons for biosynthetic processes. Moreover, enzymes in the cytosol can bypass many of the steps in the TCA cycle. The non-cyclic flux pattern predicted by the model is consistent with findings from both labeling analyses and models in plants (Sweetlove et al. 2010), in which non-cyclic flux modes were found to occur in both

illuminated leaves and non-autotrophic tissues (seeds). The results for the heterotrophic condition were also in agreement with a flux-balance model of heterotrophic *Arabidopsis* metabolism, which predicted that the reactions between 2-oxoglutarate and fumarate carried no flux when the model was constrained solely by the need to synthesize biomass components in the correct proportions (Poolman et al. 2009). A combination of a partial TCA cycle, which can support a low rate of mitochondrial oxidative phosphorylation, and substrate level phosphorylation by glycolytic reactions, is sufficient to meet ATP demands.

The oxidative pentose phosphate pathway (OPP) produces NADPH and pentose-phosphates for biosynthesis of nucleotides, amino acids, and fatty acids in the dark by decarboxylating glucose-6-phosphate. In contrast to higher plants, there is no complete OPP in the plastid of *P. tricornutum*, although the complete OPP appears to be found in the cytosol (Kroth et al. 2008). In the results for the heterotrophic (dark) condition, there was no flux through the OPP because regeneration of pentose-phosphates can occur by other means; by phosphopentose epimerase (EC 5.1.3.1) in the cytosol or by reactions in the plastid, for example. If future experimental work reveals that OPP in the cytosol is in fact active, it will indicate that this pathway is active for reasons beyond maximizing biomass; that is, the OPP is part of a "suboptimal" flux solution and not the metabolically favored pathway, unless other constraints are introduced.

The Entner-Doudoroff pathway and glycolysis in the mitochondrion

Diatoms are unusual in that they possess the complete second half of glycolysis in the mitochondria (Kroth et al. 2008), as well as the Entner-Doudoroff pathway (EDP) — an ancient glycolytic pathway predominantly restricted to prokaryotes (Fabris et al. 2012). Mitochondrial glycolysis is currently believed to occur only in diatoms and non-photosynthetic oomycetes (Ginger et al. 2010), and the purpose of this pathway in mitochondria is unknown. In Smith et al. (2012), the possibility that these enzymes could run in the direction of gluconeogenesis and that glycerate feeds into these reactions, was suggested. FBA results predict that this is a possibility, though with the optimization function, these enzymes are predicted to run in the direction of glycolysis: from dihydroxyacetone (GP) to glycerate-3-phosphate(GLY3P), for all the growth conditions. These results depend on the presence of at least one transporter that can allow glycolysis metabolites either into or out of the mitochondrion; in the model a transporter for dihydroxyacetone phosphate is included. Without it (or some other transporter), the glycolysis/gluconeogenesis pathway in the mitochondria becomes a metabolic "dead end." The inclusion of the transporter was therefore based on the assumption that the glycolysis/gluconeogenesis pathway in the mitochondrion was connected to the rest of the network.

In order to explore the role of glycolysis in the mitochondrion, different versions of the model were run, each with varying numbers of transporters between the mitochondrion and cytosol. Figures 1a-d were based on a "liberal" version of the model, where multiple transporters were included to allow transfer of all the TCA

metabolites between the mitochondrion and cytosol. The assumption for this version was that *P. tricornutum* had all the transporters commonly found in plants or yeast. In a second version of the model, the transporters were limited to only those with some degree of genomic evidence. This version (version 2) had only four putative transporters: an aspartate/glutamate shuttle, a malate/2-oxoglutarate shuttle, a citrate/malate shuttle, and a fumarate/succinate shuttle.

The flux results for version 2 were similar to the original version, but with a few differences: (1) CEF was active, (2) the flux through the mitochondrial ATP synthase decreased by about 56%, since there was less connection between the mitochondrion and the rest of the cell, (3) pyruvate in the mitochondrion is converted into glycerate-3-phosphate (GLY3P) via the action of hydroxypyruvate reductase (E.C. 1.1.1.29) and serine-pyruvate transaminase (E.C. 2.6.1.51), (4) GLY3P is converted into GP in the mitochondrion, and GP is transported out of the mitochondrion (Fig. 3). The results underscore the importance of transporters for accurate flux predictions; the glycolysis/gluconeogenesis enzymes in the mitochondrion can go in either direction, depending on the transporters included. Given that it is likely that the number of transporters are underrepresented in this organism, the actual flux distribution is likely closer to the results of the original versions of the model.

Fig. 3-

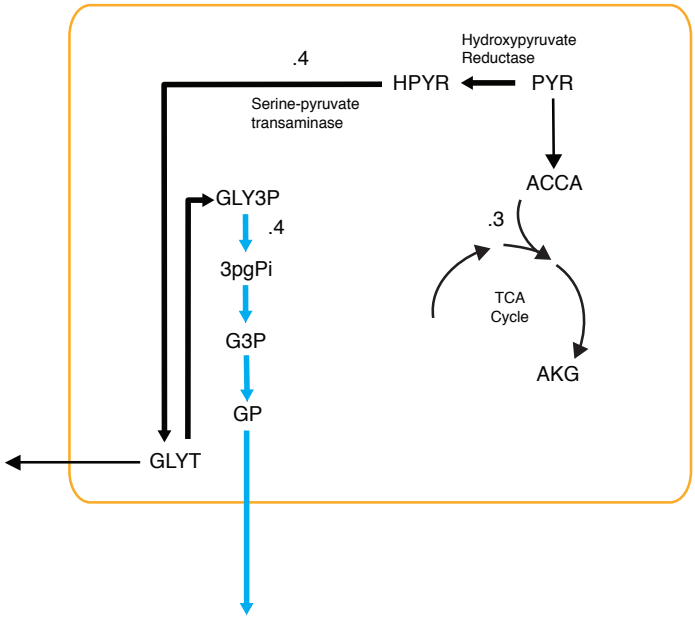


Fig. 3. Metabolic flux map for *P. tricornutum* under autotrophic growth conditions with limited transporters for TCA metabolites. The orange rectangle represents the mitochondrion. Abbreviations for metabolites are in Appendix A:6.

The purpose of the EDP in diatoms remains unclear, but it may be a way to quickly adjust to environmental changes, enabling fast metabolic responses, despite the fact that it produces less energy per molecule of glucose than the Embden-Meyerhof-Parnas (EMP) Pathway (Fabris et al. 2012). FBA predicts that there is no flux through the EDP- this is unsurprising given that EDP is less efficient than the EMP pathway. When the entire glycolysis pathway in the mitochondrion is deleted, FVA results show that there is no allowable flux through the EDP; since there is no outlet for the formation of glyceraldehyde-3-phosphate, it becomes a metabolic "dead-end."

Comparing model results with isotopomer analysis of mixotrophic glucose metabolism

In the study by Zheng et al. (2013), *P. tricornutum* grown under mixotrophic glucose conditions was found to produce glycine and serine primarily from glyoxylate via the photorespiratory reactions alanine-glyoxylate aminotransferase (AGAT), E.C. 2.6.1.44, and serine hydroxymethyltransferase (SHMT), E.C. 2.1.2.1, rather than from the more "conventional" pathway through 3-phosphoglycerate dehydrogenase. Model results also showed these pathways to be active in forming glycine and serine, with no flux through the more conventional pathway (Fig. 4). For the mixotrophic growth condition (with glucose and Ci), the model predicted that 99% of the serine was produced from SHMT in the mitochondrion, with the rest produced from cystathione beta-synthase (CBS) in the cytosol. About 98.8% of the total glycine was produced by SHMT in the cytosol and 1.3% produced by AGAT in the mitochondrion. There was no flux through the more conventional pathway from 3-phosphoglycerate (which utilizes 3-phosphoglycerate dehydrogenase, phosphoserine aminotransferase, and phosphoserine phosphatase).

Fig. 4-

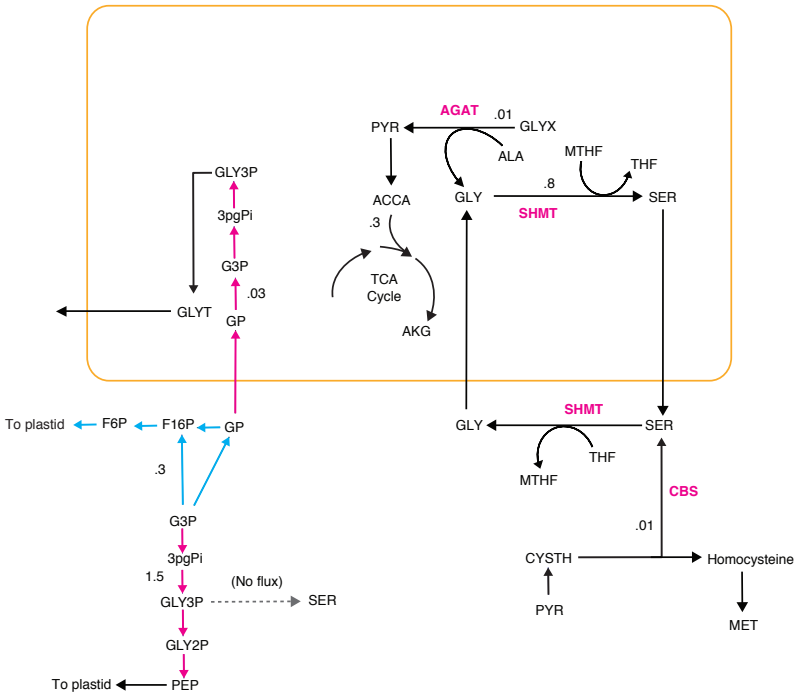


Fig. 4. Overview of glycine and serine synthesis in *P. tricornutum* under mixotrophy (growth with glucose and inorganic carbon in the light), and related pathways. The gray dashed arrow (no flux) represents the "conventional" pathway for serine synthesis. The orange rectangle represents the mitochondrion. Abbreviations for metabolites and enzymes are in the Appendix A:6.

When phosphoglycerate mutase (PGAM), which competes for 3-phosphoglycerate in the cytosol, is deleted, this does not raise flux through 3-phosphoglycerate, indicating that competition for 3-phosphoglycerate is not the reason for the alternative route. Deleting SHMT in the cytosol and mitochondrion, as well as CBS, is also not sufficient to send flux through the conventional pathway. However, when

all of these enzymes are deleted, along with AGAT2 (E.C. 2.6.1.51) there is flux through the pathway from 3-phosphoglycerate (if all SHMT enzymes, CBS, and AGAT are deleted, there is no biomass produced). The optimization function may pick the SHMT, CBS, and AGAT route for serine and glycine synthesis over the conventional pathway because these pathways also lead to other amino acids synthesized concomitantly. Flux through CBS is also linked to methionine synthesis, for example, since the action of CBS would lead to the production of homocysteine, which can be used for methionine production via methionine synthase (E.C. 2.1.1.13), and the AGAT enzymes lead to alanine production.

The Zheng et al. (2013) study also showed that there was flux through the Entner-Doudoroff (ED) and phosphoketolase (PPK) pathways. Although the model did not predict flux through the ED pathway, the PPK pathway, which converts xylulose-5-phosphate to glyceraldehyde-3-phosphate and acetyl-phosphate or fructose-6-phosphate to erythrose-4-phosphate and acetyl-phosphate, was active for all growth conditions (Figure 1a-d). The acetyl-phosphate that is formed from this pathway is converted to acetate and transferred into the plastid, where it is used to synthesize acetyl-CoA, which feeds fatty acid synthesis.

Comparison to other model photosynthetic organisms

Flux distributions of *P. tricornutum* were compared to two other photosynthetic microbes for which FBA models exist: the green alga, *Chlamydomonas reinhardtii*, and the cyanobacterium, *Synechocystis sp* PCC 6803 (Shastri and Morgan, 2005). As

discussed earlier, *P. tricornutum* did not utilize cyclic electron flow to generate ATP in the plastid, and therefore had the highest calculated flux through the non-cyclic (i.e., linear) electron transport chain, or ETC, normalized to CO_2 assimilated (Table 4). The output of O_2 was therefore slightly higher than for the other two organisms. However, as mentioned earlier, this was based on the assumption that *P. tricornutum* has ATP maintenance requirements that are not vastly different from those of *C. reinhardtii*.

Table 4: Comparison of selected fluxes with *Synechocystis* and *C. reinhardtii*

Autotrophic growth fluxes (moles/mole CO_2)

	P. tricornutum	*Synechocystis*	*C. reinhardtii*
Cyclic ET	0	0.5	1.9
Non cyclic ET	5.5	4.8	3.5
Photons	5.5	5.3	5.4
O_2 Output	0.7	0.6	0.4

Table 4. Comparison of selected fluxes (moles photons/mole CO_2) with *Synechocystis* and *C. reinhardtii.*

In contrast to results from a *Synechocystis* model, malic enzyme was not found to be essential for heterotrophic growth (Shastri and Morgan, 2005); in fact, none of the conditions had flux through this enzyme. Additionally, phosphoenolpyruvate carboxylase, which converts bicarbonate and phosphoenolpyruvate to oxaloacetate, exists in *P. tricornutum*, but was not found to be necessary for maximizing biomass.

This is in contrast to the chlorophyte alga, *Chlorella*, where the flux through phosphoenolpyruvate carboxylase was high, especially under autotrophic conditions (Muthuraj et al. 2013). This reaction has been suggested to play a role in adapting to low CO_2 conditions (Reinfelder et al. 2000), but its flux was not found to increase as the CO_2 or bicarbonate uptake rate was set lower (a third of the default case of 1.95 mmol/g dry wt/hr) in the model. This is consistent with the view that this reaction may be part of a way to dissipate excess light energy in the diatom, or generate amino acid precursors, rather than play some role in maximizing biomass output under low CO_2. Further constraints on the model would be needed in order to explore this effect.

In the plastids of land plants and green algae, some of the enzymes of the Calvin cycle, such as phosphoribulokinase, glyceraldehyde-3-phosphate dehydrogenase, fructose-1,6-bisphosphatase, and seduheptulose-1,7-bisphosphatase, become inactive in the dark via a ferredoxin/thioredoxin system (Buchanan et al. 1967, Holmgren et al. 1977, Wolosiuk and Buchanan 1977). In the current *P. tricornutum* model, deleting all of these enzymes for the dark condition is permissible (non-lethal), except in the case of glyceraldehyde-3-phosphate dehydrogenase; deletion of this enzyme in the plastid results in no biomass. Whether this is due to a key difference between diatoms and land plants/green algae, or the result of missing reactions or transporters in the model, remains to be elucidated. It is possible that light/dark regulation of plastidic enzymes simply differs in diatoms; in green algae and land plants, for example, the plastidic enzyme phosphoribulokinase (EC

2.7.1.19) is light-regulated via reversible reduction by reduced thioredoxin, while in the diatom *Odontella sinensis*, this enzyme is active under both light and dark conditions (Michels et al. 2005).

Conclusions and Further Work

The most recent *C. reinhardtii* model (Chang et al. 2011) integrated biological and optical data to devise a model that predicts growth for different light sources, and experimentally verified transcripts accounted for in the network. This provides higher predictive capacity and confidence in the model; it is a natural next step for the progression of the *P. tricornutum* model. Additionally, knowledge of ATP requirements and transporters between compartments should increase the accuracy of predictions for the glycolysis/gluconeogenesis pathway in the mitochondria, the direction of flux for these reactions, the role this pathway plays in carbon partitioning, and the percentage of ATP produced in the mitochondrion. The FBA model confirms recent findings that show that there is energetic coupling between plastids and mitochondria. The model also predicts that glycerate produced in the mitochondria is fed by the glycolysis pathway in the mitochondrion. Changes in these enzymes could therefore affect lipid synthesis, since glycerate can enter the plastid where it can potentially be shunted into glycerol, and ultimately into triacylglycerols. Glycerate also plays a role in photorespiration, though a plastidic glycerate kinase has yet to be identified in *P. tricornutum*. Since the model also predicts flux through the phosphoketolase pathway, overexpression of this pathway, or of acetate kinase, which converts acetyl-phosphate to acetate, could

increase lipid synthesis. As genome annotation for this organism improves, and further constraints for the metabolic network are identified, the model can be an invaluable tool for determining the suitability of diatoms as a source of lipid for biodiesel. Gene expression studies will also be vital for modeling flux through pathways that are active outside of conditions that biomass optimization, such as the Entner Doudoroff pathway. Knowledge of transporters between organelles will also play an invaluable role in elucidating how diatoms have managed to be such successful contributors to primary productivity.

Experimental Procedures

Network construction

A stoichiometric model of metabolism for *P. tricornutum* was constructed using genomic databases, DiatomCyc (Fabris et al. 2012), and archival journal articles. The reconstruction began with annotations from KEGG (Kanehisa et al. 2000) and DiatomCyc, The network includes the TCA cycle, oxidative and reductive pentose phosphate pathways, the Embden-Meyerhof-Parnas and Entner-Doudoroff pathways, phosphoketolase pathway, amino acid, nucleotide, chlorophyll, starch, and lipid pathways. The reversibility of reactions was assigned using a combination of KEGG, Metacyc (Caspi et al. 2012), and BRENDA (Schomburg et al. 2013) databases. More specifically: the reversibility of reactions was taken from DiatomCyc, which is based on KEGG and MetaCyc. In most cases, there was agreement between all three databases regarding the reversibility of reactions.

However, in some cases, there were discrepancies between the databases; in these cases, BRENDA was used to determine reaction reversibility. The complete model is in S2.

Gaps in the network were addressed by first searching for the missing enzymes in other organisms, then using BLAST (Altschul et al. 1990) to search for similar amino acid sequences against the *P. tricornutum* database. Transporters were also searched for in this way. This led to the inclusion of a malate/2-oxoglutarate antiporter (8990). The sequence was checked on NCBI by running a BLAST in the Swissprot database to ensure that hits gave back the same result; the first hit had an E value of 6E-62.

In some cases, "gap" enzymes that were still not found in databases were included; this was done in cases where the majority of the enzymes in a network were present, but the network was incomplete without the inclusion of these missing enzymes. One example is the plastidic sedoheptulose-bisphosphatase (Reaction 3.1.3.37). Both Sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase (which generates sedoheptulose 1,7-bisphosphate), and transketolase, (which uses sedoheptulose-7P), are found in the plastid. Sedoheptulose-bisphosphatase connects these two since it converts sedoheptulose 1,7-bisphosphate to sedoheptulose-7P. Another example in which a gap was filled was in the case of NAD biosynthesis from aspartate. A reaction that produces nicotinate ribonucleotide, and a separate reaction that consumes nicotinate adenine dinucleotide exist. Nicotinate-

mononucleotide adenyltransferase (reaction 2.7.7.18), converts nicotinate ribonucleotide to nicotinate adenine dinucleotide, and fills the gap in this pathway. A list of all the gap-filling reactions is in the Appendix A:3, "Gaps." These have a confidence score of 1. In total, 34 such reactions were included to fill in gaps (this does not include transporter reactions).

Transporters between compartments

Though little is known about the shuttles that transport metabolites out of the plastid in this organism, several were assumed since demands for both ATP and NAD(P)H outside the chloroplast must be met for biosynthetic reactions (Kramer et al. 2011). Triose phosphate translocators were included based on genomic information (Weber et al. 2006). Isolated plastids from two marine diatoms were able to reduce added 3-phosphoglycerate, and inhibited by external phosphate, which suggests that they have phosphate translocators (Wittpoth et al. 1998). Other translocators were included as a way to bridge reactions that would otherwise be disconnected. For example, a UDP glucose pyrophosphorylase (2.7.7.9) was encoded in the genome (23639); this catalyzes the conversion of UTP and glucose-1-phosphate to UDP-glucose and pyrophosphate, and is predicted to be in the plastid, but 1,3-beta-glucan synthase, which generates beta-glucan from UDP-glucose, is outside of the plastid. In this case, a transporter reaction was included to export UDP glucose out of the plastid. Other transporter reactions were included for similar reasons (export of amino acids, cofactors, sugars, etc). *P. tricornutum* also

has nucleotide transporters that shuttle nucleotides (synthesized in the cytosol) into the plastid (Ast et al. 2009).

The mitochondrial inner membrane is impermeable to NADH, so reducing equivalents must cross the membrane. An ATP-ADP translocase (which would take ATP generated by oxidative phosphorylation and counter exchange with ADP) is also needed, since ATP and ADP do not diffuse freely across the inner mitochondrial membrane. Though little is known about the specifics of the transporters across the inner membrane of the mitochondria of *P. tricornutum*, certain shuttles exist across a wide range of organisms, such as fungi, plants, animals, and algae; these were included in the model on the assumption that similar mechanisms exist in the diatom. Some mitochondrial transporters were also included based on genomic evidence.

Calculation of flux distributions

The calculation of flux distributions was performed using flux-balance analysis (FBA) coupled with Euclidean norm minimization (Bonarius et al. 1996), or flux variability analysis (Mahadevan et al. 2003). Briefly, using the network reconstruction, a stoichiometric matrix *S* with *m* rows and *n* columns is generated, where *m* is the number of metabolites and *n* is the number of reactions. Metabolites in different compartments of the cell are represented as distinct metabolites, i.e. separate metabolic entities. The *ij*th element S_{ij} of S is the stoichiometric coefficient

of the *i*th metabolite in the *j*th reaction. To find an FBA solution, the following is solved

$$\begin{aligned} z^* &= \max v_{biomass} \\ \text{subject to } Sv &= 0 \\ a \le v &\le b \end{aligned} \qquad (1)$$

where $v_{biomass}$ denotes the biomass flux component of the flux distribution vector v, a and b are lower and upper bounds, respectively, on the fluxes, and ≤ indicates component-wise inequality. The optimal solution z^* calculated by (1) is the predicted growth rate.

The optimal solution for (1) is in general not unique, allowing infinitely-many possibilities for the intracellular flux distribution. A unique flux distribution is found through Euclidean norm minimization, which has been shown to give good predictions of intracellular fluxes in a variety of systems from microbial to mammalian (Blank et al. 2005, Bonarius et al. 1996, Schuetz et al. 2007). In particular, FBA coupled with Euclidean norm minimization gives predicted fluxes that achieve an average uncentered Pearson correlation of 0.88 with fluxes determined by ^{13}C metabolic flux analysis in a dataset of 9 experiments conducted on *Saccharomyces cerevisiae* (Appendix A:4). Thus, after computing (1), the following is solved

$$\begin{aligned} &\min \Sigma_i v_i^2 \\ &\text{subject to } Sv = 0, \\ &a \le v \le b, \\ &v_{biomass} = z^*. \end{aligned} \tag{2}$$

The optimal solution v^* of problem (2) is the predicted flux distribution. This optimal solution vector is unique, meaning that if there are two vectors v_1 and v_2 that satisfy the constraints of problem (2) and achieve its optimum, then $v_1 = v_2$. To prove this, it can be noted that the objective function of problem (2) is strictly convex and its feasible set is convex. Because the optimal solution vector to the minimization of a strictly convex function over a convex set is unique (see, e.g., Bertsekas et al. 2003), the optimal solution vector to problem (2) is unique.
All model calculations were performed using the MOST (Metabolic Optimization and Simulation) software (Kelley et al. 2014) with the Gurobi solver.

Biomass composition

The biomass composition was formulated using a variety of measurements and sources in the literature. Small differences in biomass composition that were the result of different experimental conditions were not found to significantly affect flux distributions. See Appendix A:2 for the biomass composition.

Culture setup and chlorophyll, carbohydrate measurements

Axenic cultures of *P. tricornutum* strain CCAP 1055/1 cultures were grown in filtered and autoclaved seawater supplemented with F/2 medium (Guillard, 1983). Cultures were bubbled with sterilized air and maintained in 18ºC under approximately 150-200 μE/m^2/s light intensity. 10-20mls of culture were collected by vacuum filtration onto Whatman 25 mm GF/F filters. Chlorophyll was extracted using 90 % acetone, and measured spectrophotometrically (SLM-Aminco DW-2000 spectrophotometer). Chlorophyll concentrations were determined according to Jeffrey and Humphrey 1975.

Carbohydrate quantification was based on a colorimetric assay adapted from Dubois et al. 1956. Roughly 10^7 cells were filtered onto polycarbonate filters and sonicated 3x(20 sec each) in 300μl milliQ water, while kept on ice. For each sample, 10μl of cell lysate was transferred to a microplate with 2μl of 90% phenol solution, 70μl milliQ water, and 200μl sulfuric acid. Plates were incubated for 30 min. at room temperature, and the absorbance was read at 485nm, and compared with glucose standards.

Photobioreactor setup for protein quantification and growth rate determination

Growth rate data came from two separate culture setups: (1) Axenic cultures of *P. tricornutum* strain CCAP 1055/1 were cultivated in enriched artificial seawater (ESAW) medium (Berges et al. 2001) in two closed, fully controlled photobioreactors (PBRs), designed with custom specifications on the existing Labfors 5 model (Infors-HT, http://www.infors-ht.com). The working volume of the

PBRs was 3 l, illuminated by cool-white LEDs (Philips) at a light intensity of 233-308 µE. Mass transfer was ensured by internal constant stirring with 4 paddles at 400rpm; aeration was ensured by constant bubbling of compressed air through a metal sparger located at the bottom of the vessel. Temperature was kept at 21ºC and pH was automatically buffered at 7.9 (+/- 0.1) through occasional CO_2 pulses. The software Iris 6 (Infors-HT, http://www.infors-ht.com) was used as online control system to continuously monitor the following parameter, besides the ones mentioned above: dissolved oxygen concentration (pO2), stirring, balances, medium feed-in and culture feed-out pumps. Growth rate was monitored by optical density determination of cell density at 405 nm of daily samples. (2) Additionally, cultures were grown in filtered and autoclaved seawater supplemented with F/2 medium (Guillard, 1983), bubbled with air filtered through 0.2µm filters and maintained in 18ºC under continuous light (150 to 200 µE). Growth rate was determined with a Coulter counter multisizer 3 (Beckman Coulter Inc, Fullerton, CA, USA).

Protein content was analyzed according to the DC protein assay (Biorad), with modifications. 12 mg of freeze-dried biomass was dissolved in lysis buffer, containing 60 mM TRIS and 2% SDS. Cells were disrupted using a bead beater (Precellys 24, Bertin Technologies, France) at 6500 rpm for 3 cycles of 60 s with 120 s breaks on ice. Subsequently, samples were incubated for 30 min at 100 °C and centrifuged at 3500 rpm for 10 min. Supernatant was collected and diluted with lysis buffer. Samples were analyzed in triplicate.

CHAPTER THREE

Cell cycle arrest leads to increased lipid accumulation and a shift in metabolism in *P. tricornutum*

Recently, global climate change has led to an interest in algal triacylglycerols (TAGs) as feedstocks for sustainable biodiesel. TAGs consist of three long chains of fatty acids attached to a glycerol backbone. They are one of the most energy-rich forms of reduced carbon available from nature, easily converted to fatty acid methyl esters (FAMEs), which can be used in modern diesel engines (Durrett et al. 2008). TAGs are one of the major carbon storage product in diatoms; under normal growth conditions, between 15 and 25% of their biomass is composed of lipids, though some strains can produce up to 60% of their ash-free dry weight as TAGs under the appropriate conditions (Sheehan et al. 1998). Theoretically, diatoms can serve as a biofuels production platform to satisfy the total oil consumption in the U.S., using between 3 and 5% of its land area (Levitan et al. 2014).

In diatoms, ~40% of photosynthetically fixed carbon is incorporated into proteins (Parsons et al. 1961, Badour and Gergis 1965). Under optimal conditions of growth, carbon is preferentially allocated to proteins, and other components required for producing a duplicate cell. The percentage fixed into TAGs increases under certain conditions, such as nutrient starvation. Of the conditions that lead to increased TAG accumulation, nitrogen (N) starvation has been studied the most extensively. All

algae have a requirement for fixed N, a necessary component of amino acids, nucleotides, and other cellular components. N deprivation leads to increased lipid accumulation, with a higher proportion as TAGs (Fogg et al. 1956, Badour and Gergis 1965, Opute 1974a, Shifrin and Chisholm 1981, Lynn et al. 2000). It also interferes with fundamental functions of the cell, such as the maintenance and repair of proteins, and leads to decreased protein and chlorophyll levels, reduced photosynthetic rates, changes in intermediate metabolite levels, and cell cycle arrest (Sheehan et al. 1998, Guerra et al. 2013, Levitan et al. 2015). These multiple effects make it unclear which specifically lead to increased lipid accumulation.

To refine our understanding of the mechanisms that lead to increased TAG accumulation, this study examines the effects of cell cycle arrest by inhibiting a cyclin-dependent kinase, a protein involved in cell cycle processes. In all eukaryotes, cell cycle progression is regulated by serine-threonine cyclin-dependent kinases (CDKs). CDKs interact with cyclins, a diverse family of proteins ranging in size from 35 to 90 kDa, whose levels are controlled by transcription and degradation, in a cell cycle dependent manner (King et al. 1996). Progression through the cell cycle is linear; that is, it can only move forward, though it can be paused at various "checkpoints" if conditions for the cell cease to be favorable.

Since a cell must sense the physiological state of the cell in order to "decide" whether conditions are favorable for dividing, CDKs, along with cyclins, are thought to integrate metabolic cues from a network of signaling molecules in order to

determine whether conditions are favorable for dividing. Conversely, their activity should affect the metabolism of the cell; if a cell arrests in the cell cycle, this should result in physiological or metabolic changes, since the needs of a cell in an arrested state differ from those preparing to divide. It was hypothesized that cell cycle arrest via chemical inhibition of CDKs would lead to cell cycle arrest and lipid accumulation, while avoiding some of the adverse side effects (such as lower photosynthetic capacity) of nutrient deprivation. Cdk inhibition was achieved by employing a pharmacological inhibitor of Cdks, which are purine analogs that target the ATP-binding pocket (catalytic-site) of these kinases; by interfering with a CDK's ability to bind to ATP, they render the CDK inactive.

Results

Selection of the CDK inhibitor and effect on growth rate and the cell cycle

Six purine analogs were tested on *P. tricornutum* to select those that induced cell cycle arrest (and concomitant decrease in growth rate), with minimal disruption of basic physiological functions. The drugs tested included roscovitine, purvalanol A, purvalanol B, AT7519, Pha 79387, and NU 2058 (6-(Cyclohexylmethoxy)-9*H*-purin-2-amine). Of the inhibitors tested, all but Pha 79387 affected the cell cycle and decreased growth rate, but only NU 2058 led to higher fatty acid accumulation per cell. This inhibitor, which inhibits human CDK1 and CDK2 with K_i values 5 and 12 μM respectively (Arris et al. 2000), did not lead to any obvious adverse side effects, such as chlorosis or morphological abnormalities. NU 2058 has been shown to bind to the ATP binding site of Cdks in a different orientation from that of roscovitine or

the purvalanols (Arris et al. 2000). The two purvalanols led to obvious chlorosis. In the case of roscovitine, at 17 μM ($\mu = 0.11 d^{-1}$), a slight increase in fatty acid per cell was observed, but within the margin of error. *P. tricornutum* cultures recovered from roscovitine quickly; after one day of roscovitine application, cultures divided normally. This may be due to the fact that roscovitine is a more complex molecule compared to NU 2058, with more side chains that can be hydroxylated or methylated. Roscovitine can also dimerize, which would render it too bulky to fit into the Cdk ATP pocket (Karady et al. 2011).

The growth rate of *P. tricornutum* decreased with addition of NU 2058 in a dose dependent manner, up to about 7μM (Fig. 1). Twelve hours after NU 2058 had been administered, roughly 70-75% of cells with NU 2058 were in G1 phase, compared to 26-29% in control cells (Fig. 2a and 2b). At this time point, control cells had roughly 21% in S phase, and 52% in G2/M phase, while treated cells had about between 14-20% in S phase, and 10% in G2/M phase. Some cells with NU 2058 were therefore able to progress into S or G2 phases, but cultures with NU 2058 did not fully recover (grow at a rate comparable to untreated cells) until at least three days after NU 2058 had been administered. Cultures with NU 2058 for three days or more did not exhibit a consistent pattern of cell cycle synchronization, though the majority of cells (between 70-89%) were in G1.

Fig. 1-

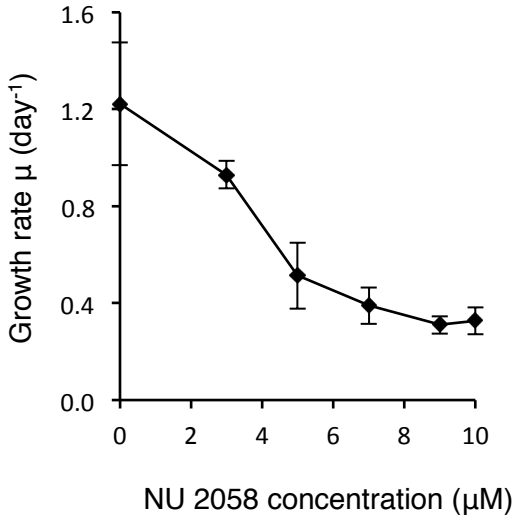


Fig. 1. The growth rate of *P. tricornutum* cells after one day of treatment with NU 2058. Growth rate decreased with addition of NU 2058 in a dose dependent manner, up to about 7μM.

Fig. 2a-

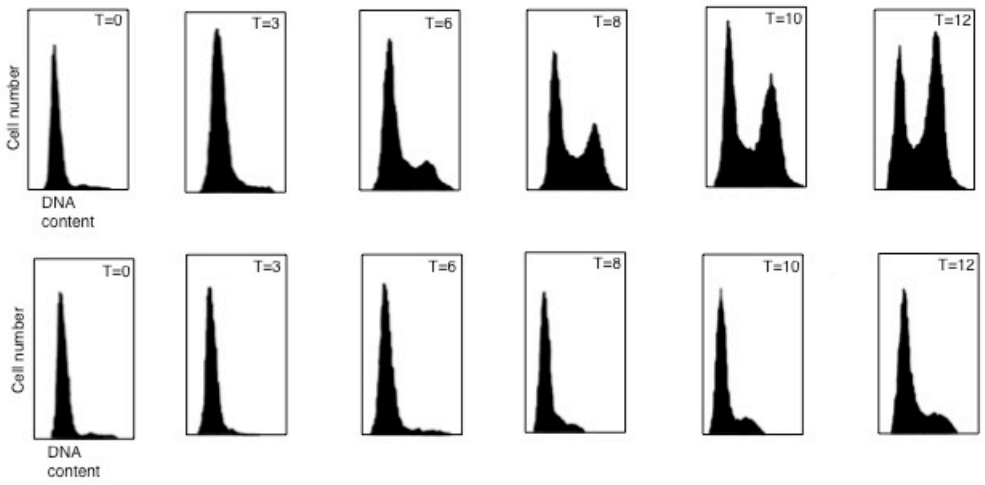


Fig. 2b-

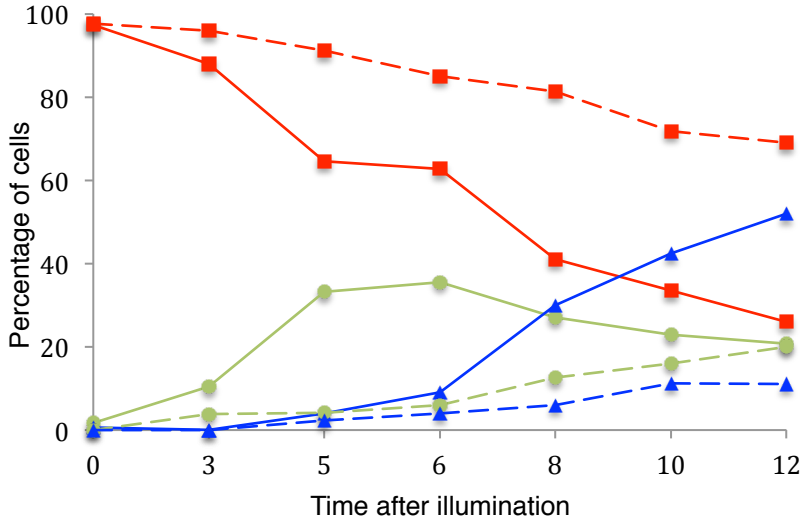


Fig. 2a. Cell cycle analysis. Cells synchronized in G1 were placed in the light and fixed in methanol at different time points. T = time, in hours after illumination. The top and bottom rows show the control and treated cells, respectively. Samples were analyzed on the flow cytometer. DNA was stained with propidium iodide (emission at 562-588nm). 2b. Percentage of cells in G1 (red), S (green), and G2/M (blue), in control (solid lines) and treated (dotted lines) cells.

NU 2058 is efficacious at different phases of the cell cycle; if it is administered to synchronized cells at 3, 5, and 6 hours after illumination, it is still efficacious in halting growth. This can be explained by the fact that NU 2058 can bind to both CDKA1 and CDKA2, which are active at different times in the cell cycle (G1-to-S and G2/M). Affinity chromatography and mass spectrometry were employed to isolate the target(s) of NU 2058 in *P. tricornutum*, and co-immunoprecipitation (COIP) with a PSTAIR antibody was performed; "PSTAIRE" is a "classic" cyclin-binding site for Cdks. In *P. tricornutum*, CDKA1 and CDKA2 have PSTAIRE and PSTALRE cyclin-

binding sites, respectively (Fig. 3a). Based on amino acid sequence similarities in the ATP binding site of CDK1 and CDK2 in humans, it was hypothesized that NU 2058 would bind to the orthologues of these proteins in *P. tricornutum* (CDKA1 and CDKA2). The PSTAIR COIP picked up CDKA1, CDKA2, and a small amount of hCDK3, a hypothetical CDK. The NU 2058 matrix did not pick up any other CDKs besides CDKA1 and CDKA2.

Figure 3a-

CDKA1 (*P. tricornutum*)
MERYQRMEKIGEGTYGVVYKAKDRVTGEIIALKKIRLEAEDEGIPSTAIREISLLKELQHPNIVRLYDVV

CDKA2 (*P.tricornutum*)
MERYHKIEKPGSNLGEGTYGVVYKALDRQTDEIVALKRIRLEVEDEGIPSTALREISLLRELSHPNIVDL

CDK1 (*Homo sapiens*)
MEDYTKIEKIGEGTYGVVYKGRHKTTGQVVAMKKIRLESEEEGVPSTAIREISLLKELRHPNIVSLQDVL

CDK2 (*Homo sapiens*)
MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRXDTETEGVPSTAIREISLLKELNHPNIVKLLDVI

Fig. 3b-

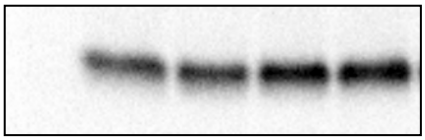


Fig. 3a. Partial sequences of CDKA1 and CDKA2 (*P. tricornutum*) and CDK1 and CDK2 (*Homo Sapiens*). Regions highlighted in yellow are ATP-binding pockets. Regions highlighted in blue are cyclin-binding regions. b. Western blot against monoclonal anti-PSTAIR (sigma

P7962) at 1:4000 dilution. C: control cells. N: Cells with NU 2058. CDKA1 and CKA2 are 33.11 and 34.12kDA respectively.

Though the function of CDKA1 is unknown, transcript data suggest that CDKA1 plays a role in the G1-to-S transition (Huysman et al. 2010). CDKA2 has been shown to play a role in G2/M; it localizes to the cell division place in pre-cytokinetic cells, and its overexpression prolongs mitosis (Huysman et al 2015). Western blot analysis with the PSTAIR antibody revealed that cells synchronized in G1 phase by prolonged darkness (for 24 hours) increased their levels of CDKA1 and/or CDKA2 (33.11 and 34.12kDa respectively) with time under illumination (Fig. 3b). There were no significant differences in CDKA1 and/or CKDA2 levels between treated and untreated cells.

Physiological characteristics

FAME content was compared in cultures grown in 0, 5, or 7 μM NU 2058 for one day under continuous illumination. FAME per cell increased with higher concentrations of NU 2058, or inversely with growth rate (Fig. 4a). Cultures with 7 μM exhibited a 2.5-fold increase in FAMEs compared to control cells. Concentrations higher than 7 μM NU 2058 did not result in significant decreases in growth rate or increases in fatty acid content; this concentration was therefore chosen as a working dosage for several experiments. At various time points, synchronized cells were collected for cell cycle analysis, macromolecules, and metabolites. Cells had the lowest levels of protein, carbohydrate, and triacylglycerols (TAGs) at dark arrest. All three

macromolecules gradually increased with time under illumination. Untreated cells, however, had roughly 12.5% and 17% higher protein after 8 and 10 hours under illumination, respectively (Fig. 4b). The activity of one of the key enzymes of nitrate assimilation, nitrate reductase (NR), decreased by 80% in treated cells after 24 hours of illumination (Fig. 4c). This effect was not clearly apparent before 12 hours of treatment with NU 2058; treated cells had NR activity that was comparable to those of control cells before 12 hours. Carbohydrate per cell, using glucose standards, was measured in synchronized cells at various time points after illumination; diatoms store polysaccharide in the form of chrysolaminarin, a linear polymer of β(1→3) and β(1→6) linked glucose units. Treated cells had roughly 39% and 30% higher levels by hours 8 and 10 respectively, compared to untreated cells (Fig. 4d). TAG accumulation was the lowest at T_0. After time in the light, both untreated and treated cells accumulated TAGs, but at hours 8 and 10, treated cells had 92% and 93% higher TAG levels respectively, compared to untreated cells (Fig. 5). Cell cycle arrest also led to 2.8 times less intact polar lipids (IPLs)- lipids that are primarily associated with membranes (Table 1). This was comparable to the threefold decrease in IPLs exhibited in nitrogen-stressed cells (Levitan et al. 2015). Most of the IPLs, including monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), and sulfoquinovosyldiacylglycerol (SQDG), showed comparable decreases compared to nitrogen-stressed cells, though the glycerophosphoglycerols (PGs) were 50% higher, and phosphatidylcholine about 3 times lower. These differences were within the margin of error, however. In general,

the total carbon content of treated cells was about 52% compared to control cells, and nitrogen content was 45% lower (Table 2).

Fig. 4a-

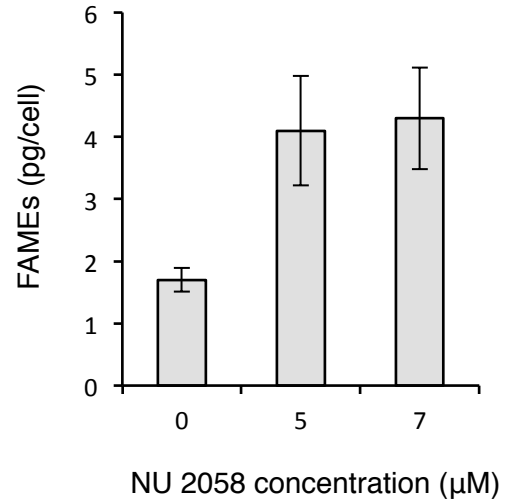


Fig. 4b-

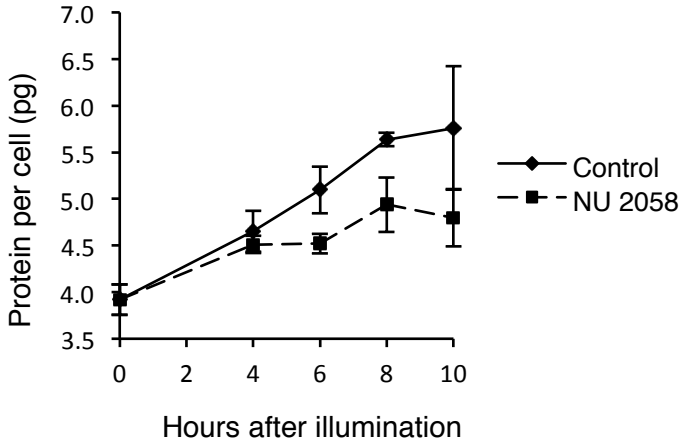


Fig. 4c-

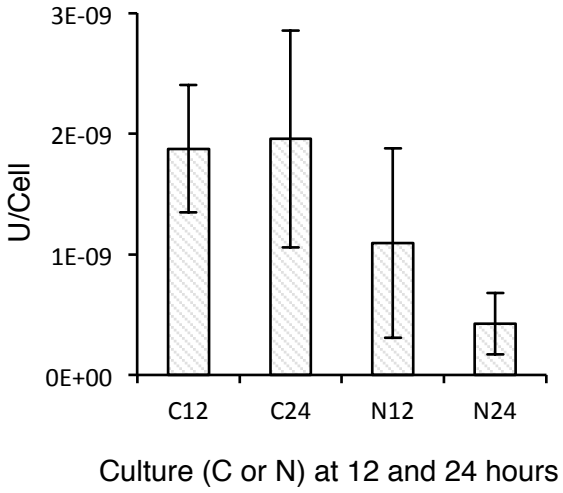


Fig. 4d-

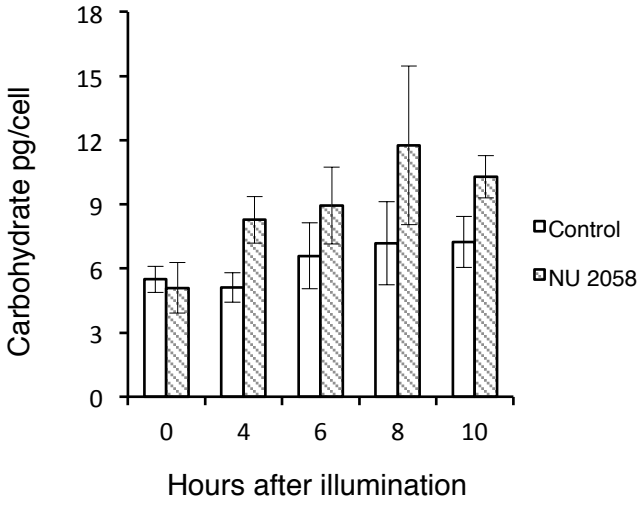


Fig. 4. Physiological characteristics of *P. tricornutum*, with and without NU 2058. a. FAME (fatty acid methyl ester) content in cultures grown in 0, 5, or 7μM NU 2058 for one day, under continuous light. b. Total protein per cell in synchronized cells measured at various time points after illumination. c. Nitrate reductase activity for C and N (control and treated cells) at 12 and 24 hours after illumination. U/cell = μmol $No2^{-}*min^{-1}*cell^{-1}$. d. Carbohydrate

per cell using glucose standards, measured in synchronized cells at various time points after illumination.

Fig. 5-

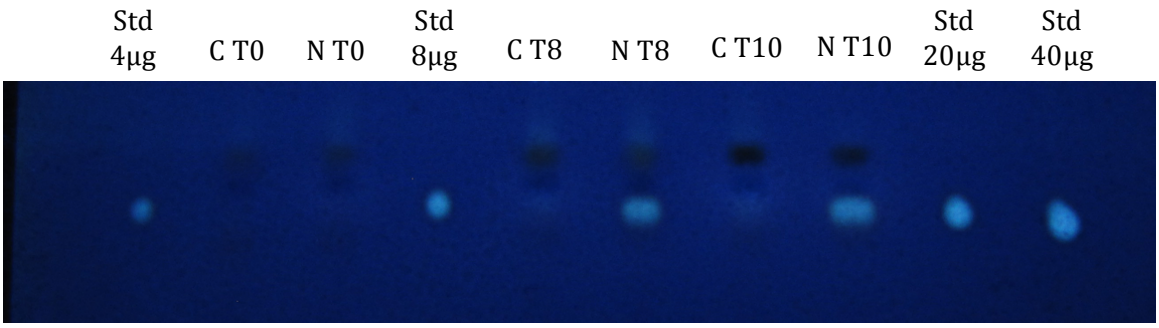


Fig. 5. Thin layer chromatography showing triacyglycerol in synchronized cells at 0, 8, and 10 hours after illumination. C = control; N = 7µM NU 2058; Std = standard.

Table 1. Intact polar lipids in control and treated cells (Fmol/cell).

	Control	Treated
MGDG	0.37 ± 0.42	0.05 ± 0.03
DGDG	0.55 ± 0.10	0.18 ± 0.03
SQDG	0.48 ± .15	0.46 ± 0.05
PG	0.37 ± 0.10	0.03 ± 0.02
PE	0.04 ± 0.01	0.00 ± 0.00
PC	0.20 ± 0.02	0.01 ± 0.01

Table 1. MGDG- monogalactosyldiacylglycerol; DGDG- digalactosyldiacylglycerol; SQDG- sulfoquinovosyldiacylglycerol; PG- glycerophosphoglycerols; PE- phosphatidylethanolamine; PC- phosphatidylcholine.

Table 2. Physiological characteristics of cells with and without NU 2058 in synchronized cells 10 hours after illumination.

	Control	NU 2058	NU2058/Control (%)	NR21/WT (%)
C (pg/cell)	9.2±0.8	14.01±0.07	152	132
N (pg/cell)	1.68±0.20	0.92±0.21	55	105
C:N (mol:mol)	6.39	17.76	278	128
Protein (pg/cell)	5.76±0.66	4.80±0.31	83	110
Carbohydrate (pg/cell)	7.24±1.19	10.30±1.00	142	112
FAME (pg/cell)	1.7±0.16	4.3±0.97	253	143

Table 2. NR21/WT- The ratio between *P.tricornutum* with nitrate reductase knocked down vs. the wild-type, as reported by Levitan et al. 2015; C- carbon; N- nitrogen; C:N- the carbon to nitrogen ratio; FAME- fatty acid methyl esters.

Cells with NU 2058 had no differences in maximum photochemical quantum yields of PSII, determined as the ratio of variable fluorescence (F_V) to maximal fluorescence (F_M, F_V/F_M), or chlorophyll *a* per cell, or the in vivo absorption cross-section normalized to chlorophyll *a* (a*), which describes the probability of a photon being absorbed within the defined light regime (Table 3). There were also no significant differences in levels of PsbD, the D2 subunit of PSII (Fig. 6a), or RbcL, the large subunit of RuBisCO (Fig. 6b). Electron transport rate (at 400 μE), however, was 56% lower (Fig. 6c), and non-photochemical quenching 33% higher (Fig. 6d), after one day of treatment. Since a* and chl *a* levels were comparable in treated and untreated cells, the quantum requirement 1/φFA for fatty acid biosynthesis in

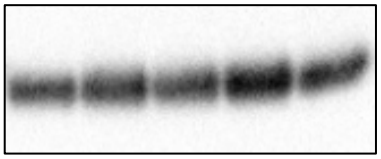
treated cells was 61% lower compared to untreated cells, or treated cells were 2.6 times more efficient at fatty acid biosynthesis over a period of one day (Table 2).

Table 3. Photosynthetic attributes of *P. tricornutum*.

NU 2058 (μM)	Fv/Fm	Chl a (pg/cell)	a* (m^2/mg Chl a)	1/φFA (quanta/C)
0	0.58	0.55 ± 0.02	0.006 ± 0.001	217 ± 13
5	0.60	0.60 ± 0.01	0.006 ± 0.002	150 ± 14
7	0.57	0.54 ± 0.03	0.006 ± 0.001	147 ± 14

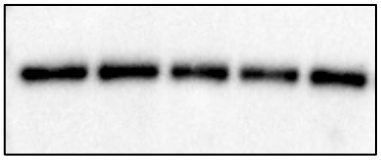
Table 3. Fv/Fm- maximum quantum efficiency of photochemistry; Chla a- chlorophyll a; a*- the optical absorption cross section of PSII normalized to chlorophyll a; 1/φFA- the photon energetic requirement for incorporation of carbon into fatty acid.

Fig. 6a. PsbD-



0h C:4h N:4h C:8h N:8h

Fig. 6b. RbcL-



0h C:4h N:4h C:8h N:8h

Fig. 6c-

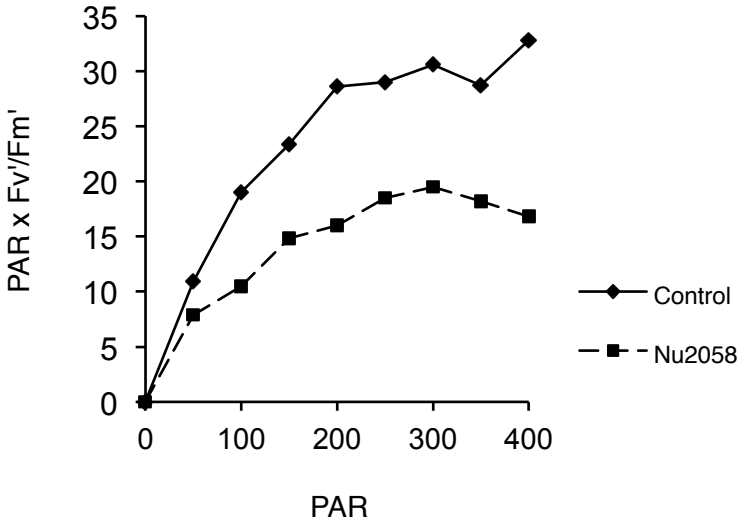


Fig. 6d-

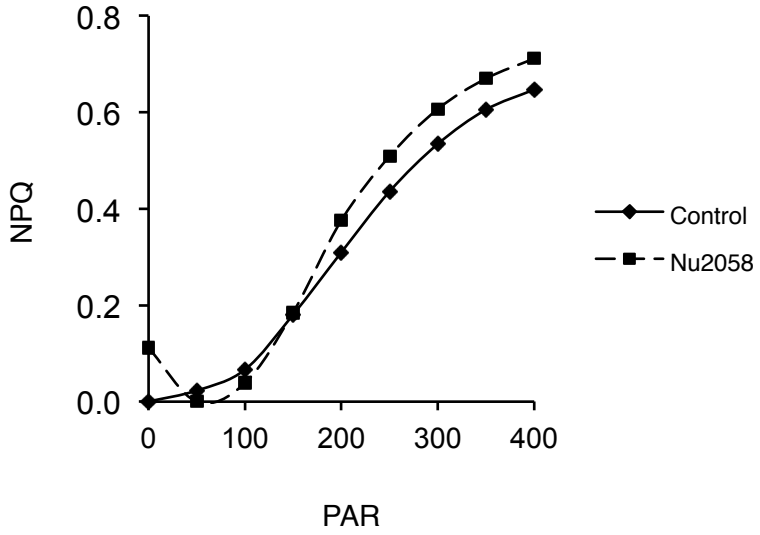


Fig. 6, a. PsbD, and b. RbcL (top and bottom, respectively), using extracts from synchronized cells, 0, 4, and 8 hours after illumination. C: control cells. N: Cells with NU 2058. c. Electron tansport rates (PAR x Fv'/Fm') and d. non-photochemical quenching (NPQ) in cells treated and untreated with NU 2058 after one day under continuous illumination. PAR = photosynthetically active radiation.

Metabolite profiles

Synchronized cells exposed to light and sampled at several intervals for metabolites, showed that the largest differences in metabolites between treated and untreated cells were found in those involved in the pentose phosphate pathway or glycolysis (Fig. 7). Some of these showed an initial increase compared to control cells (pyruvate, phosphoenolpyruvate, 3-phosphoglycerate, fructose bis-phosphate), suggesting an initial buildup of recently fixed photosynthate through carbon-fixation and/or glycolysis. At the ten-hour time point, however, treated cells had much lower pools of these metabolites; 3-phosphoglycerate, fructose bis-phosphate, fructose 6-phosphate, glucose 6-phosphate, and glucose 1-phosphate levels were roughly half of the levels found in control cells. Only five metabolites were found to be higher in treated cells: malate, citrate, glutamine, malonyl-CoA, and NADH. Malate and citrate levels were more than 50% higher in treated cells, though other TCA metabolites were measured to be equal or slightly higher in treated cells. Glutamine levels in both treated and untreated increased with time in illumination, but were 50% higher in treated cells by hour 10. Levels of glutamate, however, were equal to or slightly lower than in treated cells, compared to untreated cells. Malonyl-CoA, which provides 2-carbon units to fatty acids during fatty acid elongation, was about 50% higher in treated cells at hour 10. Acetyl-CoA levels, however, were measured to be the same or slightly lower than those in untreated cells.

Other large differences were in AMP, which was 50% less in treated cells, compared to untreated cells. NADP+ levels in treated cells were about half of those in control cells. NADH was about 50% higher in treated cells. Most metabolites increased within four hours upon illumination, in both control and treated cells (category A). Three metabolites (oxaloacetate, dihydroxyacetone-phosphate, and 6-phospho-gluconate) decreased within four hours of illumination (category B). In this category, there was no large difference between control and treated cells. Only one measured metabolite (malate) tended to rise in treated cells, and fall in control cells. Of the glycolysis/pentose phosphate metabolites, a few (phosphoenolpyruvate, 3-phosphoglycerate, glyceraldehyde-3-phosphate, fructose-bis-phosphate, dihydroxyacetone-phosphate) increased initially in arrested cells, but gradually decreased with time.

Fig. 7-

	Hours after illumination				
	0.3	4	8	10	Category
Glutamine	1.0	1.1	1.4	1.5	A
Glutamate	0.9	1.0	1.0	0.8	A
Malate	0.9	1.0	1.9	1.9	D
Pyruvate	2.1	0.9	1.0	1.1	A
Acetyl-CoA	1.0	0.9	0.9	0.9	A
Oxaloacetate	1.0	1.0	1.0	1.2	B
Succinate	0.8	0.9	1.1	1.1	A
Citrate	1.4	2.1	1.9	1.6	A
α-Ketoglutarate	1.1	1.1	1.1	1.1	A
PEP	1.2	1.3	1.3	1.1	A
3-P-Glycerate	1.4	1.1	0.6	0.6	A
Glyceraldehyde-P	1.1	1.0	0.8	0.7	C
DHAP	1.2	1.1	1.0		B
Fructose bis-P	1.5	1.0	0.6	0.6	A

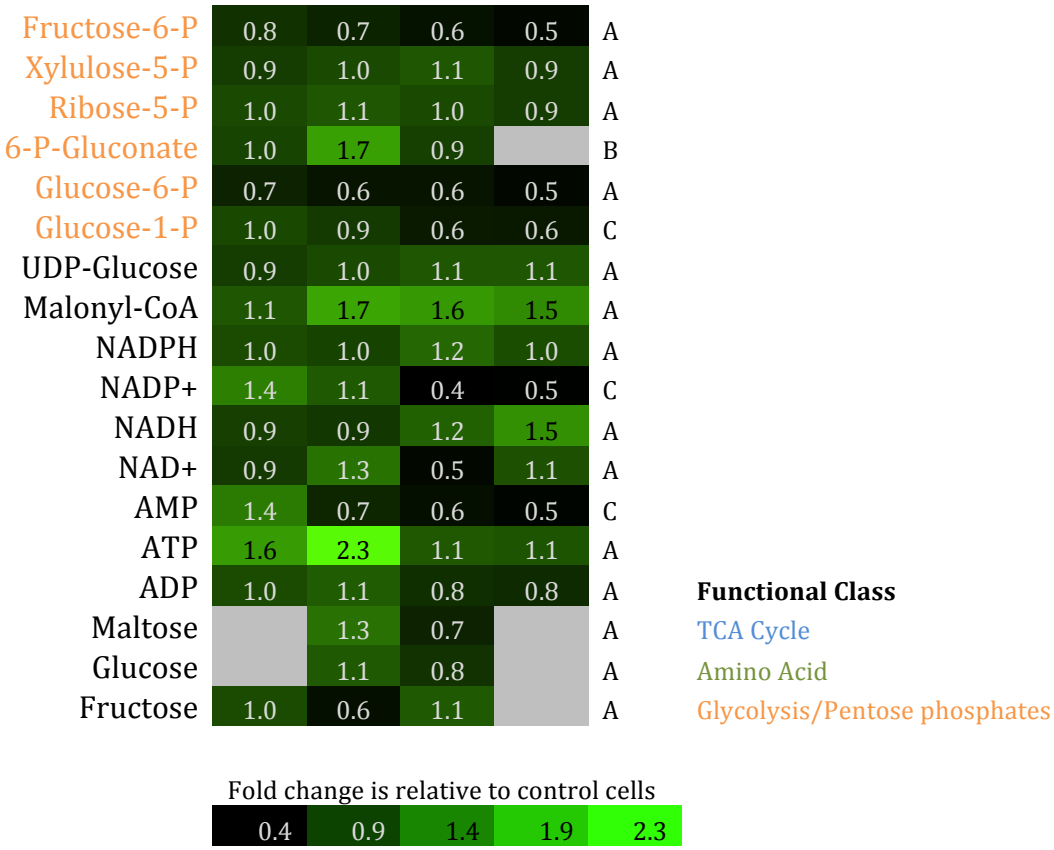


Fig. 7. Fold change in metabolites. Synchronized cells were exposed to continuous light and sampled at several intervals for metabolites, in both 7μM NU 2058 treated and untreated control cells. Category A: metabolites that rise with hours of illumination, in both control and treated cells. Category B: metabolites that fall with hours of illumination, in both control and treated cells. Category C: metabolites that rise in control cells, and fall or stay the same in treated cells. Category D: metabolites that rise in treated cells, and fall in control cells.

NU 2058 specificity

In order to gauge the specificity of NU 2058 towards different kinases, its effect on the activities of several human or rat kinases was measured (Table 4). It was least effective towards CDK2, CDK5, and GSK3 (glycogen synthase kinase 3, a serine/threonine kinase that phosphorylates tumor suppressor proteins, cyclin D1,

and a large variety of other substrates in humans). It was most effective in inhibiting human CDK1 (orthologue of CDKA1 in *P. tricornutum*), human CLK1 (CDC-like kinase 1), and rat DYRK1A, a member of the dual-specificity tyrosine phosphorylation-regulated kinase (DYRK) family, which may be involved in signaling pathways that regulate cell proliferation. A protein BLAST of CLK1 in the *P. tricornutum* genome database shows that the top hit is an un-annotated protein (JGI ID 14003) most similar to PK12, a dual-specificity protein kinase of the LAMMER family of protein kinases, related to the cell cycle-dependent CDC2-type kinases (Sessa et al. 1996). The top hit for DYRK1A in the *P. tricornutum* database is a kinase most similar to a DYRK1A kinase in *D. renrio*.

Table 4. Kinase activity assays: IC 50 and % residual activity for several human or rat kinases.

	CDK1	CDK2	CDK5	CLK1	DYRK1A	GSK3
IC 50 (μM)	< 10	>10	>10	<10	<10	>10
RA	44%	92%	70%	8%	43%	101%

Table 4. CDK1- Human cyclin-dependent kinase 1; CDK2- Human cyclin-dependent kinase 2; CDK5- Human cyclin-dependent kinase 5; CLK1- CDC-like kinase 1; DYRK1A rat dual specificity tyrosine-phosphorylation-regulated kinase 1A; GSK- Human glycogen synthase kinase 3. IC50- The half maximal inhibitory concentration; RA-% residual activity at 10 μM NU 2058.

Interacting partners

Studies in yeast have shown that the connection between lipid metabolism and the cell cycle can be direct; lipolysis, or the deacylation of TAG by lipases, is directly

regulated by Cdk1/Cdc28. The major TAG lipase Tgl4 is phosphorylated and activated directly by this cdk (Kurat et al. 2009). Activated Tgl4 breaks down TAGs, and free fatty acids released by lipolysis can be used to generate energy in the absence of other carbon sources, and serve as lipid precursors required for membrane lipid synthesis. To determine whether there was a similar link between lipid enzymes and Cdks in *P. tricornutum*, the results of the affinity chromatography with NU 2058 and the CO-IP with the PSTAIR antibody were examined for possible lipases or other lipid metabolism enzymes that were co-bound to CDKA1 or CDKA2. Additionally, affinity chromatography with the human P9 protein, orthologous to the Cks protein, was performed. The complete list of proteins sequenced is in Appendix A:8. Lipid metabolism enzymes, especially with Cdk substrate recognition sequences (S/T)PX(K/R), were selected for further study (Table 5). In order to confirm interactions, pairwise yeast two-hybrid (Y2H) assays were performed with these putative substrates as prey, and full-length CDKA1 and CDK2 sequences fused to the GAL4 DNA-binding domain used as bait. However, none of these enzymes showed interaction with either CDKA1 or CDKA2 according to the pairwise Y2H assays.

Table 5. Subset of protein sequencing: Results of affinity chromatography and PSTAIR antibody co-immunoprecipitation

JGI Phatr2 ID	Cell Cycle Protein	Count of Sequence PSTAIR COIP	NU 2058	P9
20262	CDKA1	63	5	3
51279	CDKA2	66	11	22
34165	hCDK3	2		63
46095	CYCB1			26
29283	CYCD1			5
46245	Rb-related protein			2

JGI Phatr2 ID	Lipid Metabolism Protein	PSTAIR COIP	NU 2058	P9	CDK Substrate Motif
28797	Fatty acid desaturase		2		
20143	Acyl-CoA synthetase	32	27	8	TPPK, TPCR
43025	Esterase/lipase		3		TPSR
43352	Esterase/lipase	3	14	3	TP
10718	Esterase/lipase		6		SP, TP

Table 5. PSTAIR COIP- PSTAIR antibody co-immunoprecipitation; NU 2058- affinity chromatography with NU 2058 matrix, P9- affinity chromatography with the P9 (Cks) protein.

Discussion

When conditions are favorable and nutrients plentiful, cells synthesize proteins and other macromolecules to focus their metabolic capacity on dividing. Under nutrient replete conditions, most fixed carbon flows to pyruvate, which is decarboxylated to form acetyl-CoA (AcCoA). This AcCoA enters the tricarboxylic (TCA) cycle to form intermediate metabolites for anabolic pathways like amino acid synthesis. Thus, under optimal conditions, carbon is preferentially allocated to proteins, and other

components required for producing a duplicate cell. A condition of "balanced growth" is achieved when cells are growing under constant conditions and have reached a steady-state biochemical composition; that is, the rate of growth of any cell constituent, such as protein, lipid, or DNA, remains the same if averaged over a generation (Falkowski and Raven 2007). Under these optimal conditions for growth, fatty acids are likely synthesized primarily for esterification into membrane lipids, and as an energy reserve. Under conditions of stress such as nitrogen deprivation, however, nitrogen-consuming constituents decrease, and nitrogen-poor constituents such as lipids increase; cells are forced into a state of unbalanced growth whereby photosynthetic products no longer needed for cell division are redirected into alternate, nitrogen-poor sinks. Additionally, the cell has a smaller sink for photosynthetically produced electrons, and excess electrons that accumulate in the photosynthetic electron transport chain induce over-production of reactive oxygen species, inhibiting photosynthesis and damaging membrane lipids and proteins. Absorption of light that exceeds a cell's capacity for CO_2 fixation also activates mechanisms to quench excess energy, such as non-photochemical quenching. Under these conditions, N-deficient compounds, such as lipids or carbohydrates, serve to remove large amounts of photosynthetically-produced ATP and NAD(P)H. The formation of fatty acids consumes approximately 24 NADPH derived from the electron transport chain, which is twice that required for synthesis of a carbohydrate or protein molecule of the same mass.

In arrested cells, the lower electron transport rates and higher non-photochemical quenching suggest that they have a smaller sink for electrons compared to actively dividing cells, since arrested cells have lower protein and nucleotide requirements, compared to cells progressing through the cell cycle. In arrested cells, photosynthetically generated products are instead shunted into neutral lipids and chrysolaminarin. Since division requires energy, membrane synthesis, and secretory vesicles, actively dividing cells may also tap into any available storage lipids, which would prevent triacylglycerols from accumulating.

Both cell cycle arrest and the inability to make nitrogen-rich metabolites lead to increased neutral lipid content. Levitan et al. 2015 also showed that a strain with nitrate reductase knocked down (NR21) led to increased lipid accumulation compared to the wild-type. This study shows that cell cycle arrest alone is sufficient for increasing lipid content, without significantly affecting the cell's photosynthetic capacity, or ability to produce amino acids. Moreover, the ratio of fatty acid content in arrested:control cells is 1.8 times higher than that of NR21:the wild-type, suggesting that cell cycle arrest alone, without disruption to nitrogen metabolism, is a more efficient method for maximizing lipid production.

In arrested cells, lower protein requirements also lead to lower nitrate assimilation rates, or NR activity. Interestingly, nitrogen-replete diatoms have been known to use nitrate and nitrite reduction, and subsequent NO_2- or NH_4+ release, as a way to dissipate excess energy (Collos, 1998, Lomas et al. 2000). Arrested cells, despite

having excess energy to dissipate, do not "choose" this mechanism in the long term (after 12 hours of arrest), suggesting that this mechanism is more of a short-term response to excess irradiance. Alternatively, the build-up of photosynthetic products leads to lower nitrate reductase activity. Vergara et al. 1998 showed that changes in NR activity and cellular carbon content were strongly negatively correlated in the diatom *T. weissflogii* with a time lag of 6 h, suggesting a feedback between cellular carbon content and NR.

A key branch point in lipid biosynthesis in *P. tricornutum* is in the GS/GOGAT/GDH (glutamine synthetase/glutamine oxoglutarate aminotransferase/glutamate dehydrogenase) branch point (Guerra et al. 2013, Levitan et al. 2015). Glutamine is a precursor for other amino acids and provides nitrogen for the synthesis of nucleotide bases. *P. tricornutum* cells under nitrogen stress (or with NR knocked down) increase their GLU/GLN ratios 12-fold or more (Flynn et al. 1988, Frada et al. 2013, Guerra et al. 2013, Levitan et al. 2015). In general, a higher GLU/GLN ratio has consistently been associated with N-deprivation in a number of different algae, and has been proposed as a biomarker for N-deplete cells (Flynn et al. 1989). In contrast, cells treated with NU 2058 lower the GLU/GLN ratio by 41% compared to control cells. The GLU/GLN ratio in N-stressed and arrested cells therefore show opposite tendencies. This may have to do with the fact that N-stressed cells have higher catabolism of enzymes that are part of the photosynthetic machinery, releasing intracellular nitrogen that can be redirected into enzymes that can confer a selective advantages under nitrogen stress, such as enzymes associated with nitrogen uptake

and assimilation: nitrate reductase, glutamate dehydrogenase, enzymes associated with the urea and the TCA cycles (Levitan et al. 2015). More generally, the GLU/GLN ratio may not just be a biomarker for N-stress, but an indicator for how much nitrogen is available relative to the nitrogen demand. So control cells, though not nitrogen-stressed, have a higher nitrogen demand than arrested cells, causing them to have a higher GLU/GLN ratio compared to arrested cells. In rat hepatomas, glutamine content also negatively correlates with tumor growth rates (Sebolt et al. 1983). Though multicellular systems are vastly different from algae, it is worth exploring whether generalizations can be made between GLU/GLN ratios and nitrogen demand.

Acetyl-CoA is used in many biochemical reactions, such as the TCA cycle, glyoxylate cycle, synthesis of fatty acids, malonyl-CoA, and isoprenoids. It is also a source of acetyl groups for posttranslational modification acetylation, and has many sources: pyruvate dehydrogenase, fatty acid or amino acid breakdown. Its metabolite levels therefore cannot be a proxy for acetyl-CoA carboxylase (ACCase) activity, the enzyme responsible for forming malonyl-CoA from acetyl-CoA. However, the fact that both citrate and malonyl-CoA levels were higher in arrested cells is intriguing, since ACCase is known to be allosterically activated by citrate (Martin et al. 1962).

The results of the Y2H pairwise assays could not confirm the putative substrates suggested by the results of affinity chromatography; it is possible that the affinity chromatography assay results led to false positives, or that the Y2H method could

not confirm interactions because the yeast system lacks cyclins that are close enough to the diatom's. This would preclude interactions since correct cyclin-binding is required to (1) activate Cdks and (2) confer substrate specificity. Future studies may have to employ yeast three-hybrid systems to confirm Cdk-substrate interactions in the diatom. Trentacoste et al. 2013 showed that knockdown of a lipase in the diatom *T. pseudonana* increased lipid yields; it will certainly be worth investigating the differences between Cdk and lipase inhibition, as well as possible links between Cdks and lipases.

Conclusions

Cell cycle arrest via Cdk inhibition leads to a state of unbalanced growth whereby photosynthetic products no longer needed for cell division are redirected into carbohydrates and neutral lipids. This leads to lower protein and higher lipid accumulation without severe disruption of photosynthetic machinery. In contrast to nitrogen-deprived cells, arrested cells lower nitrate reductase activity and the ratio of glutamate to glutamine. Lipid accumulation via cell cycle arrest is a result of fundamentally different processes from lipid accumulation via nitrogen-deprivation. Although the precise mechanisms behind both processes remain to be elucidated, the results of this study clearly establish cell cycle arrest as a promising means towards lipid accumulation in a model diatom. This has implications for biofuels research, given the recent interest in diatoms as sustainable sources of feedstocks for biodiesel (Levitan et al. 2014).

Methods

Experimental conditions

Axenic *Phaeodactylum tricornutum* strain CCAP 1055/1 cultures were grown in filtered and autoclaved seawater supplemented with F/2 medium (Guillard and Ryther, 1962). Cultures were aerated through 0.2 µm filters and maintained in 18ºC under continuous light (~90 µmol photons•m^2•s^{-1}). All purine analogs were administered with DMSO (final concentration 1µl/ml) to facilitate their dissolving in seawater media. Control cultures also had a final concentration of 1µl/ml DMSO.

Analytical methods

Growth and cell abundance were determined using a Beckman Coulter Multisizer 2 (Beckman Coulter Inc, Fullerton, CD, USA). For experiments that required synchronized cultures, *P. tricornutum* cells were arrested in G1 phase by prolonged darkness (24 hours). Synchronization was validated by flow cytometric analysis with an InFLux Model 209S Mariner Flow Cytometer (BD Biosciences) equipped with a 488 nm laser, on cells fixed in methanol, washed three times with 1x phosphate buffered saline, treated with RNase A for one hour, and stained with propidium iodide, or Sybr Green I. For each sample, roughly 10,000 cells were processed. Gating and data analyses were performed using FlowJo analytical software (Tree Star Inc.).

Cellular carbon and nitrogen content was measured from cells collected onto pre-combusted 13 mm A/E filters (Pall Gelman) and analyzed on a CHN analyzer (Na

1500 series 2, Carlo Erba Instruments, CE Elantech Inc., Lakewood, NJ, USA). Total fatty acid extraction (FA) and methylation was based on the method by Rodriguez-Ruiz et al. 1998. $5x10^7$ cells collected onto 25 mm Whatman GF/F filters were extracted in 10 ml glass-vials with 2ml of 1:20 acetyl chloride: methanol mixture and 1ml of hexane, in the presence of 25µg of heptadecanoic acid (Sigma-Aldrich), which was used as an internal standard for fatty acid (FA) methyl esters quantification. The vials were boiled in a water-bath for 1h. After cooling, 1ml of milliQ water and the hexane phases were transferred with glass-pipettes into GC vials with Teflon-lined caps (Fisher Scientific). FA analysis was performed with a gas-chromatographer (GC 2010 Shimadzu, Japan) equipped with a TR-FAME (0.25µm x 60 m) column (Thermo electron corporation, USA) and a flame ionization detector. FAs were identified by reference to authentic standards (Supelco™ 37 component FAME, Sigma-Aldrich) processed and analyzed under the same conditions. TAGS were separated using thin layer chromatography (TLC) as described in Guerra et al. 2013. IPLs were separated using stepwise elutions based on the method by Yongmanitchai et al. 1992. All lipids were compared to known amounts of commercial standards from Sigma-Aldrich and Avanti Polar Lipids, Inc. Carbohydrate quantification was based on a colorimetric assay adapted from Dubois et al. 1956. 10^7 cells were filtered onto polycarbonate filters and sonicated 3x(20 sec each) in 300µl milliQ water, while kept on ice. For each sample, 10µl of cell lysate was transferred to a microplate with 2µl of 90% phenol solution, 70µl milliQ water, and 200µl sulfuric acid. Plates were incubated for 30 min. at room temperature, and the absorbance was read at 485nm, and compared with glucose standards.

For protein quantification, $1x10^8$ cells were filtered onto polycarbonate filters, and protein was extracted with 300 µl of extraction buffer: 2 % LDS, 100 mM Tris Base, 105mM Tris HCl, 0.5mM EDTA, 10% glycerol and 15 µl of protease inhibitor cocktail solution (Sigma-Aldrich, Corp., St. Louis, MO, USA). Cell suspensions were sonicated three times in between liquid nitrogen freezes (Microson set at power 2), then centrifuged (16,000 g, 5min). Protein concentrations were determined with the Bio-Rad *DC* Proteins Assay. For Western blots, protein extracts were boiled for 5 min with 30 µl of 1M DTT and 300 µl of a solution containing of 4 % SDS, 15% glycerol and 0.05 % bromothymol blue. 2 µg of protein was loaded onto a pre-cast 4- 20 % Tris-HCl gel (BIORAD, USA), ran for 1h (100 mV) and transferred electrophoretically to a PVDF membrane. NR activity was determined colorimetrically and calculated as µmol NO_2^- produced per min^{-1} per $cell^{-1}$ as described in Berges and Harrison 1995. NU 2058 was added to control cells shortly before harvesting for NR activity assays (at a final concentration of 7µM), in order to determine whether NU 2058 had a direct effect on the NR enzyme.

The maximum photochemical quantum yields of PSII (Fv/Fm) were measured on dark-adapted cell suspensions with a fluorescence induction and relaxation system (Satlantic Inc., Halifax, NS, Canada; (Gorbunov and Falkowski 2004). For chlorophyll measurements, 10-20mls of culture were collected by vacuum filtration onto Whatman 25 mm GF/F filters. Chlorophyll was extracted using 90 % acetone, and measured spectrophotometrically (SLM-Aminco DW-2000 spectrophotometer).

Chlorophyll concentrations were determined according to Jeffrey and Humphrey 1975. The *in vivo* absorption spectra were measured with an SLM-Aminco DW-2000 spectrophotometer, using live cell suspensions. The values obtained were used to calculate the wavelength-specific cross-section of optical absorption, normalized to chlorophyll a (a*) (Dubinsky et al. 1986, Falkowski et al. 1985).

For the affinity chromatography, *P. tricornutum* protein extracts loaded over NU 2058 immobilized to an agarose matrix were sent for sequencing. The extracts contained protein from a mixture of cells in G1, S, and G2 phases (roughly 40%, 25%, and 35% respectively). Preparation of NU 2058 resins were performed as described in Bach et al. 2005. For the affinity chromatography, 1.5mg of diatom protein extracts were used for 15µl packed beads. Protein kinase assays were performed as described in Bach et al. 2005.

Yeast two-hybrid bait plasmids were generated with recombinational GATEWAY cloning (Invitrogen). For the bait ENTRY clones, full-length open reading frames of the *P. tricornutum CDKA1* and *CDKA2* were isolated and cloned into the pENTR-D-TOPO vector. The ENTRY clones were then recombined in the pDEST32 (bait) vector (Invitrogen) by attL × attR recombination, resulting in translational fusions between the CDKA1 or CKDA2 proteins and the GAL4 DNA-binding domains. Yeast PJ694-alpha cells (MATalpha; trp1-901, leu2-3,112, ura3-52, his3-200, gal4D, gal80D, LYS2::GAL1-HIS3, GAL2-ADE2, met2GAL7-lacZ) were co-transformed with bait (CDKA1 or CDKA2) and prey plasmid by the LiAc method (Gietz et al. 1992). Co-

transformation was analyzed on medium without leucine and tryptophan (-L-T). Co-transformants were tested for their ability to activate the histidine marker gene by testing for yeast growth on medium without leucine, tryptophan and histidine (-L-T-H) and for their ability to activate the *LacZ* reporter gene (X-Gal). The *GUS* gene was used as a negative control. dsCYC2 and CKS1, which have been shown to interact with CDKA1 and CDKA2 respectively, were used as positive controls. For each combination three independent colonies were screened.

Metabolite profiles were obtained via the method described by Bennette et al. (2011). Cells were quenched via rapid filtration of 15 mL of cells followed by immediate transfer of filters to 1.8 mL of 80:20 methanol: water at -20^{o}C in 35x10 mm Petri dishes. Filters were incubated at -20^{o}C for 15 minutes and scraped clean of cells. The 1.8 mL of extraction solvent and cells was transferred to 2 mL eppendorf tubes, followed by a 0.25 mL 80:20 methanol: water at -20^{o}C wash of the filter to collect remaining quenched cell material. The eppendorf tubes were spun at 4^{o}C, and supernatant was collected. To the cell pellet, 100 µL 80:20 methanol: water at -20^{o}C was added, vortexed, incubated at -20^{o}C for 15 minutes, and pelletted. The 100 µL was combined with the original supernatant, and 100 µL evaporated in a speed vac and resuspended in 33 µL of water, which was transferred to GC vials. The analysis of the metabolites was achieved in one 35-minute LC-MS/MS run on a 1200-series LC with a 6410 QQQ MS employing reversed-phase ion pairing chromatography (Agilent Technologies, Santa Clara, CA, USA). Standards for each metabolite, run with sample background, were used for quantification. The column

used was a Synergi 2.5µm hydro RP 100A, size 100x2 mm (Phenomonex, Torrance, CA, USA).

Calculation of the quantum requirement for fatty acid biosynthesis. The efficiency of transduction of light energy at a given irradiance to fatty acid was calculated with a modified version of the model proposed by Falkowski et al. 1985, and expressed as the quantum requirement for fatty acid biosynthesis ($1/\phi_{FA}$):

$$1/\phi_{FA} = (a^* \times (\text{chl}a/\text{C}) \times I_\mu \times 1040)/\ \mu \text{ (mol quanta. mol C}^{-1}) \quad (3)$$

Where ϕ_{FA} represents the quantum yield for fatty acid biosynthesis (mol C^{-1}); a^* the in vivo absorption cross-section (m^2. mg^{-1} Chl a); Chl *a* and C represent the biomass (mg) of chlorophyll *a* and carbon (we calculated the C stored as fatty acid (FA), assuming oleic acid [C18: 1, MW= 282,46 g. mol^{-1}] as the average fatty acid molecule); I_μ represents the growth irradiance (µmol quanta. m^{-2}. s^{-1}); µ represents the specific growth rate (d^{-1}); and the constant 1040, was used to convert units from µmol quanta m^{-2} s^{-1} to mol quanta m^{-2} d^{-1} and from mg C to mol C. Higher $1/\phi_{FA}$ values indicate a decrease in efficiency of light energy transduction into fatty acid production, and lower values indicate an increase in efficiency.

CHAPTER FOUR

Gene regulation and flux predictions for *P. tricornutum* at different stages of the cell cycle: an integrated analysis.

The results from Chapter 3 suggested that arrested cells shift energy away from protein synthesis, and towards storage products such as lipid and chrysolaminarin. For Chapter 4, the transcriptomes of synchronized cells, with and without NU 2058, were analyzed at different points in the cell cycle. Other gene regulation studies with *P. tricornutum* were compared in order to compare regulatory behaviors between cells under cell cycle arrest, nitrogen-limitation, and high illumination. These conditions share similarities: arrested and nitrogen-limited cells both accumulate lipids, while all three conditions must dissipate excess light. A comparison between these three conditions can therefore reveal the fine-tuned responses to these conditions.

Gene transcripts for enzymes of metabolism were also correlated with the FBA model of *P. tricornutum* to predict metabolic fluxes that "best fit" the available transcript data. In Chapter 2, a simple objective function (maximizing biomass production rate) was used with FBA to constrain the solution space and predict flux distributions. Here, the model was constrained with transcriptomic data. As discussed in Chapter 2, the information contained in a stoichiometric model results in an underdetermined linear equation system; it is not enough to calculate a unique flux distribution. Physiologically meaningful flux solutions need to be narrowed

down from all the possible flux distributions by imposing additional constraints on the system. Stoichiometric models must therefore be combined with assumptions or experimental data, in order to yield a well-defined flux map. Transcripts can be used as cues that enzymes support metabolic fluxes in certain reactions, and network reconstruction can be used to combine these cues into a global, consistent metabolic behavior. Transcriptomic data can be seen as regulatory constraints that are self-imposed by the organism, presumably as a result of an optimal evolutionary process. Though gene expression and protein activity do not necessarily correlate, and examples can be given where fluxes increase while transcription of corresponding genes decrease or remain unchanged (Yang et al 2002), studies have shown that the integration of expression data improves the prediction performance of genome-scale models through defining flux bounds, objective functions, or both (Covert et al. 2002, Palsson, 2002, Famili et al. 2003, Akesson et al 2004).

E-Flux2 is a method for predicting intracellular fluxes using transcriptomic data and genome-scale models, when a suitable biological objective is available (Kim et al. 2015). It is an extension of FBA; the main objective is to maximize biomass, but mRNA levels are used as approximate upper bounds on the maximum amounts of correlating enzymes. The mRNA levels therefore serve as bounds on reaction rates. Additionally, as in Chapter 2, Euclidean norm minimization is employed as a further constraint, after maximizing the biomass flux. The method has been validated in *E. coli* and *S. cerevisiae,* and outperforms competing methods.

Results and Discussion

Transcriptome

To understand the effect of cell cycle arrest, *P. tricornutum* cells were arrested in G1 phase by prolonged darkness (24 hours), and separated into cultures with or without NU 2058. At the time of release from G1 by reillumination (t0), cells (100% in G1 phase) were collected for transcriptome and cell cycle analysis. After four hours of illumination, cells were again collected for mRNA and cell cycle analysis. Cells at the 4 hour timepoint were called "t4" for cultures without the inhibitor, and "N" for cultures with the inhibitor. At this time point, roughly 79% and 20% of t4 cells were in G1 and S phases, respectively, and 95% and 5% of N cells were in G1 and S phases, respectively, according to cell cycle analysis. Additionally, a culture with the inhibitor was kept alive for another 20 hours in order to have mRNA for arrested cells at 24 hours after illumination ("ND"). This culture had 70% of cells in G1 phase, 18% in S phase, and 12% in G2 phase. A 24-hour time point for control cells was not taken, because at 24 hours cells would be desynchronized, and transcriptome data would have been difficult to interpret.

N cells exhibited the most differential gene regulation compared to t0 and t4 cells. 5930 genes were differentially expressed (false detection rate (FDR) <5%), with 48% up-regulated and 52% down-regulated, when compared to t4 cells, and 5207 were differentially expressed compared to t0 cells. In comparison, 3030 genes were

differentially expressed in t4 cells compared to t0 cells, and 3409 genes differentially expressed in ND cells compared to N cells.

Cell cycle genes

Using the study by Huysman et al. 2010, cell cycle genes could be roughly categorized by when peak transcription occurs: (1) The time immediately following the onset of light (early G1), or 0 – 1 hrs after illumination, (2) G1, or 1 – 4 hrs after illumination, (3) late G1 & S phase, or 4 – 8 hrs after illumination, and (4) G2/M, or 8 – 12 hrs after illumination. The general pattern of transcription for t4 cells showed that genes in the first category were down-regulated, and genes of category 4 were up-regulated, compared to cells at t0 (Fig. 1a). This was expected, since t0 cells would be relatively enriched for category 1 transcripts, and t4 cells should be relatively enriched for later category transcripts. However, two category 3 genes (dsCYC4 and dsCYC8) were not consistent with this general trend, since they were down-regulated at t4 compared to t0.

Category 1 and 2 cyclins (active earlier in the cell cycle) were upregulated in arrested (N) cells compared to both t0 and t4 cells, since N cells had elongated G1 phases (Fig. 1b and 1c). These included all the P-type cyclins, and most of the diatom-specific cyclins (dsCYCs), which are cyclins that are highly divergent and found only in diatoms. Of the P-type cyclins, three (CYCP1, CYCP4, CYCP5) were highly up-regulated in N cells, and the rest were slightly up-regulated. P-type cyclins are normally expressed in G1. In plants, P-type cyclins (and in *S. cerevisiae*, its

homolog PHO80), were shown to be involved in phosphate availability signaling (Kaffman et al. 1994, Torres et al. 2004). CYCP6 was identified as a possible interactor of both CDKA1 and CDKA2, and CYCP1 was identified as an interactor of CDKA1 (Huysman et al. 2015). This is consistent with findings in plants; all the P-type cyclins were found to associate with CDKA in plants (Acosta et al. 2004). A transcriptome study of nitrogen-starved cells also exhibited up-regulation of P-type cyclins, possibly because nitrogen-starved cells have elongated G1 phases (Levitan et al. 2015). Of the eleven dsCYCs, six were up-regulated (dsCYC2, dsCYC4, dsCYC5, dsCYC6, dsCYC7, and dsCYC9), and only one down-regulated (dsCYC11), compared to t4 cells. Most of these are, under normal growth conditions, expressed early in the cell cycle (Huysman et al. 2010), and possibly function as early signal integrators controlled by light or mitogens; dsCYC7 and dsCYC2 are likely interactors of CDKA1 (Huysman et al. 2013 and 2015).

Fig. 1. Regulation of cell cycle genes.

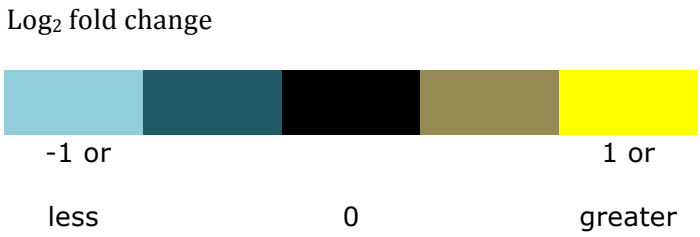


Fig. 1a-

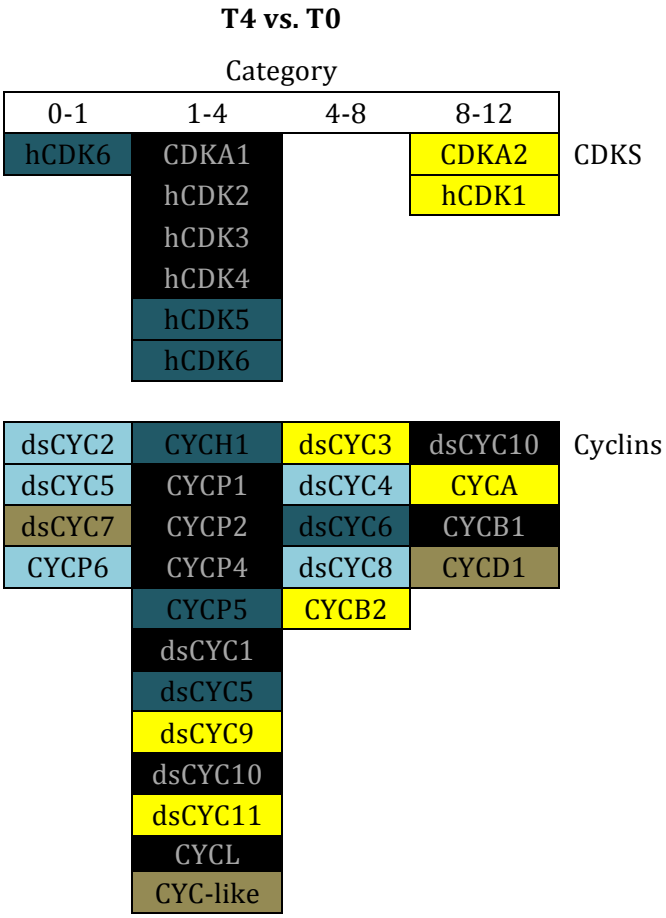


Fig. 1b-

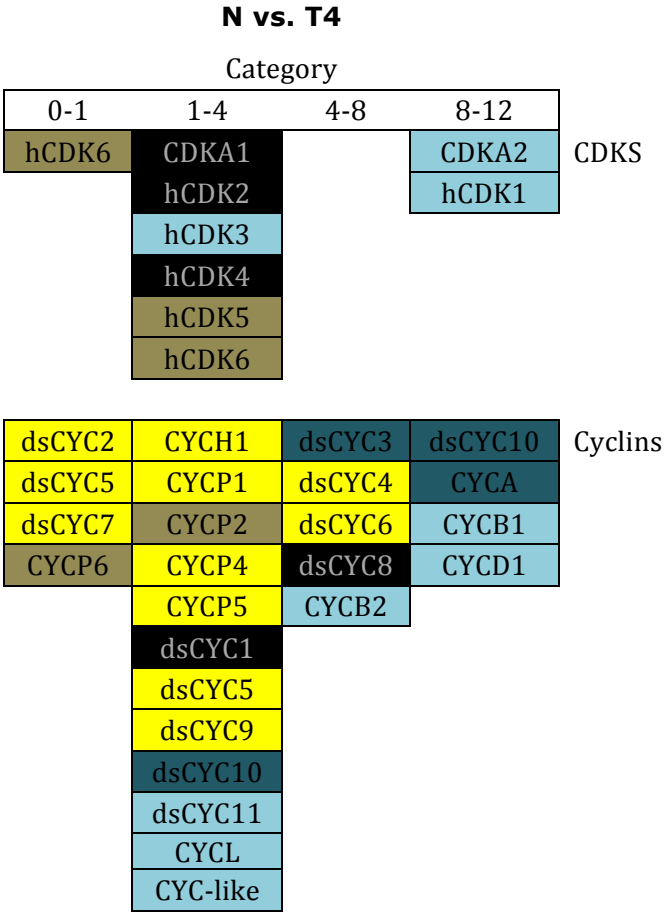

N vs. T4
Category
0-1
1-4
4-8
8-12
hCDK6
CDKA1
hCDK2
hCDK3
hCDK4
hCDK5
hCDK6
CDKA2
hCDK1
CDKS
dsCYC2
dsCYC5
dsCYC7
CYCP6
CYCH1
CYCP1
CYCP2
CYCP4
CYCP5
dsCYC1
dsCYC5
dsCYC9
dsCYC10
dsCYC11
CYCL
CYC-like
dsCYC3
dsCYC4
dsCYC6
dsCYC8
CYCB2
dsCYC10
CYCA
CYCB1
CYCD1
Cyclins

Fig. 1c-

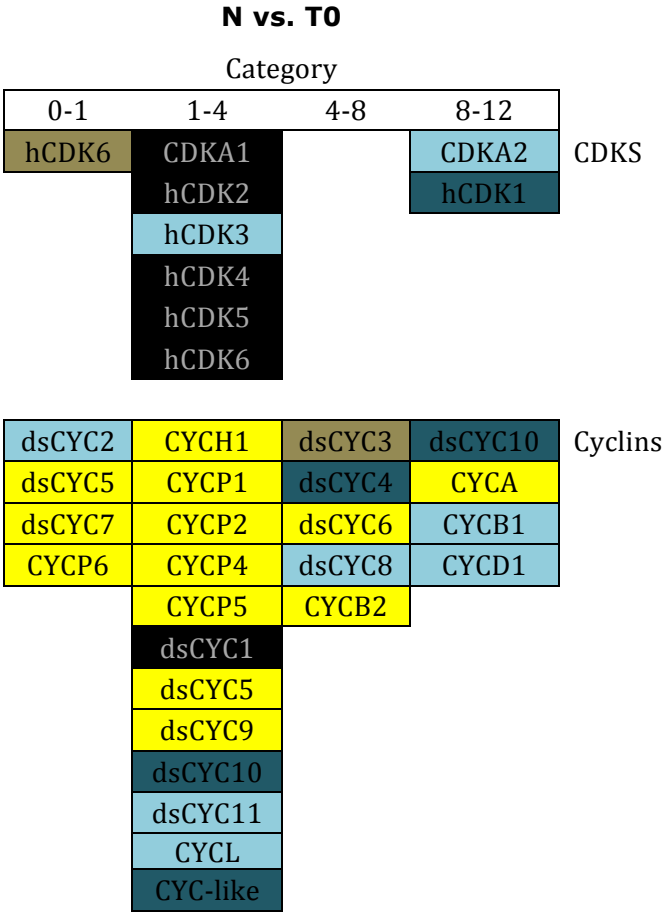

N vs. T0
Category
0-1
1-4
4-8
8-12
hCDK6
CDKA1
hCDK2
hCDK3
hCDK4
hCDK5
hCDK6
CDKA2
hCDK1
CDKS
dsCYC2
dsCYC5
dsCYC7
CYCP6
CYCH1
CYCP1
CYCP2
CYCP4
CYCP5
dsCYC1
dsCYC5
dsCYC9
dsCYC10
dsCYC11
CYCL
CYC-like
dsCYC3
dsCYC4
dsCYC6
dsCYC8
CYCB2
dsCYC10
CYCA
CYCB1
CYCD1
Cyclins

Fig. 1d-

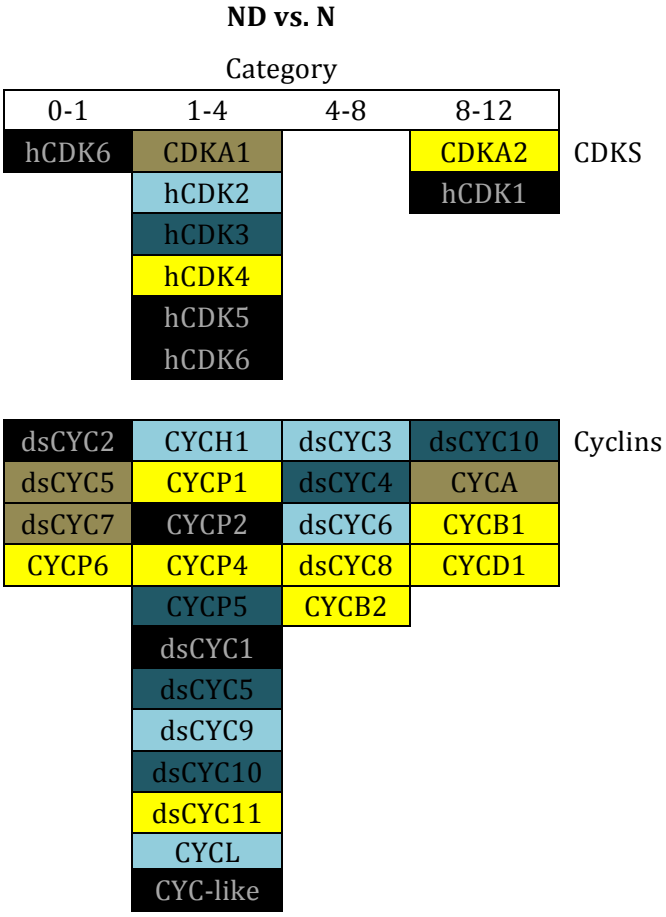


Fig. 1. Cell cycle genes were categorized by when peak transcription occurs in synchronized cells based on the study by Huysman et al. 2010: (1) The time immediately following the onset of light (early G1), or 0 – 1 hrs after illumination, (2) G1, or 1 – 4 hrs after illumination, (3) late G1 & S phase, or 4 – 8 hrs after illumination, and (4) G2/M, or 8 – 12 hrs after illumination.

In *S. cerevisiae,* various studies have revealed some of the specific mechanisms behind cell cycle progression. The concentration of Cdk is in excess of the total

amount of cyclins in the cell, and different cyclins can bind to the same Cdk (reviewed in Udvardy 1996). The actual cyclin partner of a Cdk is determined by the availability of different cyclins, which is determined by the rates of synthesis and degradation of these cyclins. In vitro experiments have revealed that after a cyclin binds to a Cdk, it is hyperphosphorylated and multiubiquinated by a ubiquitin-conjugating enzyme. This targets the cyclin for degradation. The Cdk-cyclin activity in a cell is therefore a result of (1) different cyclins competing for the Cdks, and (2) the rate at which the different cyclins are synthesized and degraded. Once a cyclin binds to a Cdk, this allows for ATP-binding to the Cdk-cyclin complex; activated Cdk can then phosphorylate its substrates, including transcription factors that initiate transcription of other cyclins or cell cycle proteins. Some Cdk-cyclin complexes also autophosphorylate. This autophosphorylation leads to more transcription of cyclins, causing a positive feedback loop that leads to a build-up of Cdk activity in the cell. A threshold level of Cdk-cyclin activity usually has to be reached in order for the next stage of the cell cycle to occur.

Immediately after the onset of light exposure, diatom-specific cyclin dsCYC2 is the most strongly and rapidly expressed cyclin; its transcript levels rise after only 5 minutes of illumination (peaking at 15 minutes after the onset of light), and dsCYC2 protein reaches high levels at 30 to 60 minutes (Huysman et al. 2013). This cyclin is likely the first cyclin to complex with CDKA1. Moreover, cyclin-binding is necessary to expose the ATP pocket of a Cdk; it changes the conformation of the Cdk such that its flexible T loop no longer blocks the ATP binding cleft. This would allow ATP or

any ATP analogues, like NU 2058, to bind to the ATP cleft. In all four conditions, CDKA1 was expressed, but NDE. This means that in N cells, though NU 2058 binds to CDKA1, it does not cause a reaction to upregulate CDKA1 to compensate for the lack of CDKA1 activity. However, in N cells, some of the later Cdks were differentially expressed compared to both t0 and t4 cells. CDKA2, hCDK1 (both category 4 genes), hCDK3 (category 2) and hCDK7 (no clear category) were down-regulated compared to t4 cells. The only Cdks that were up-regulated in N cells (compared to either t0 or t4 cells) were the C-type Cdks, which may not play a role in regulating the cell cycle. Transcript studies of C-type Cdks show no discernible cell cycle pattern, and in plants, C-type Cdks associate with transcription initiation complexes (Barroco et al. 2003).

CDKA2 and hCDK1 play a role in G2 or M phase; according to the study by Huysman et al. 2010, transcripts for these two Cdks normally peak between 8 - 12 hrs after illumination. Moreover, CDKA2 plays a role in mitosis- specifically, in the formation of the cleavage furrow or in recruiting components for the cell division plane. It localizes to the cell division plane in pre-cytokinetic cells, and its overexpression has been found to prolong mitosis (Huysman et al. 2015). Since ND cells are a later time point compared to N cells, more of the later category genes were upregulated, including that of CDKA2 (Fig. 1d).

In every eukaryotic cell, ubiquitination of targets for proteolysis is mediated by ubiquitin-activating enzyme El (which activates ubiquitin), a ubiquitin-conjugating

enzyme E2 (which transfers the ubiquitin to the protein selected for degradation by a protease), and a ubiquitin ligase E3, as well as two principal multiprotein ubiquitin ligases: Skp/Cullin/F-box protein (SCF) complex and the anaphase-promoting complex, also called cyclosome (APC/C) (Harper et al. 2002 and Vodermaier, 2004). The SCF complexes mainly control the G1-to-S and G2-to-M transitions, while the APC/C is responsible for targeting of anaphase inhibitors (securins) and mitotic cyclins (Baker et al. 2007 and Morgan, 1999). The SCF complex consists of three primary subunits RBX1, cdc53/CUL1 and SKP1, as well as a few different F-box proteins (Deshaies, 1999). F box proteins bridge the interactions between the target and E2; it determines the SCF specificity. In *P. tricornutum*, three *Skp1* homologs, three *Cul1* homologs, and one *Rbx1* homolog have been identified (Huysman et al. 2014). The APC/C also consists of subunits, most of which are present in the *P. tricornutum* genome in a single copy, except APC4, APC9, Cdc26, APC13/Swim1 and Mnd2, which could not be found. APC9 has only been identified in budding yeast, and appears to be absent in many organisms, while Cdc26, APC13/Swim1 and Mnd2 are three small conserved subunits that have been shown to be nonessential (Thornton and Toczyski, 2006).

In a transcriptional analysis of SCF and APC/C subunit genes in *P. tricornutum*, it was reported that transcription of the SKP1, CUL1 and RBX1 subunits of the SCF complex was highest at the G1-to-S transition (Huysman et al. 2014), which is consistent with what is known about its role in G1-to-S transitions. In the current study, most of these subunits were either down-regulated or NDE in N cells, while

all were NDE in t4 cells (Appendix B:2a and 2b). Given that the APC/C marks M phase cyclins for degradation, which triggers transition from metaphase to anaphase, it might be expected that it would be down-regulated in N cells. However, of the subunits of the APC/C that were expressed in the N cells, three were up-regulated and the rest NDE compared to t4 cells, while none were differentially expressed in t4 cells, compared to cells at t0.

Huysman et al. 2014 made *in silico* predictions of cyclin APC/C targets through D-box analysis. The D-boxes (destruction boxes) or the KEN boxes (Glotzer et al. 1991 and Pfleger and Kirschner, 2000), specific sequences that get recognized by APC/C, could be found in almost all G2/M cyclins, except CYCB1 and dsCYC10. Additionally, several cyclins expressed during G1/S also show putative D-box elements, including dsCYC2, suggesting that APC/C may play a role in bringing dsCYC2 levels down after G1.

The sequence of events during the cell cycle of *P. tricornutum* (with and without NU 2058) is most consistent with the results of the transcriptome as follows: (1) With the onset of light, G1 cyclins dsCYC2, dsCYC5, dsCYC7, and CYCP6 are expressed. The first cyclin to be expressed is dsCYC2; this cyclin binds to CDKA1 first. (2) Once dsCYC2 binds to CDKA1 in N cells, NU 2058 can bind to CDKA1. (3) With increasing time in the light, CDKA1-cyclin complexes increase Cdk activity in t4 cells. This Cdk activity leads to activation of substrates, and transcription of other cyclins. In N cells, there is no CDKA1-cyclin activity, which prevents activation of various

substrates. (4) dsCYC2 that is bound to CDKA1 may still be targeted for degradation by APC/C, but because of low Cdk activity, signals that would normally cause the cell to down-regulate transcription of dsCYC2 and other G1 cyclins, are not fully active in N cells, which may explain why (5) other putative cyclin partners of CDKA1, which include dsCYC7, CYCP1 and CYCP6, are upregulated in N cells.

Other cell cycle genes

Other interactors of Cdks, besides cyclins, include CKIs (Cdk inhibitors), CKS proteins (Cdk subunit proteins), and various kinases or phosphatases that serve to inhibit or activate Cdks. Two important cell cycle regulators are the Rb proteins and E2F transcriptional regulators. All eukaryotes except fungi share a common pathway for entry into S phase; phosphorylation of RB (retinoblastoma) by CDK activity causes RB to no longer bind to E2F transcriptional regulators (consisting of a heterodimer of E2F and DP proteins), and this allows the G1-to-S transition (Weinberg 1995, Claudio et al. 2002). *P. tricornutum* has an Rb-related protein (46245) that is upregulated in N cells, compared to both t0 and t4 cells (Appendix B:1). There is no clear trend or pattern with E2F or DP gene expression.

WEE1/MYT1/MIK1 kinases inhibit cell cycle progression by phosphorylating Cdks. They preferentially phosphorylate and inactivate Cdk1/CDC2, which negatively regulates G2/M transitions. They are highly regulated during the cell cycle. The study by Huysman et al. 2010 reported that the expression pattern of MYT1 was highly variable and not associated with a specific cell cycle phase. In the current

study, however, it was highly up-regulated in N cells compared to t0 and t4 cells (Appendix B:1). Additionally it was upregulated in ND cells. MYT1 may be involved in stress responses during the cell cycle (Zhou et al. 2000). The expression levels of MAD3, a mitotic checkpoint kinase, were consistent with the fact that a significant portion of t4 cells were in G2; it was highly up-regulated in t4 cells, compared to t0 cells, and down-regulated in N cells, compared to t4 cells.

Analysis of gene expression of light receptors and circadian rhythm

Circadian rhythm is the periodicity of cell processes observed in cultures that are entrained by light/dark cycles. It is endogenous, or driven by a circadian clock, but adjustable (entrained) by the local environment by external cues, such as light. It is regulated by cryptochrome/photolyase family (CPF) proteins, which act as photoreceptors. In *P. tricornutum*, cryptochrome PtCPF1 has DNA repair activity, and acts as a transcriptional repressor of the circadian clock. Coesel et al. 2008 found that PtCPF1 protein levels showed modest diurnal fluctuation, presumably because of high protein stability, and was expressed in the dark, in cells grown under continuous light, though higher levels were induced with the onset of blue light. The transcriptome results showed that PtCPF1 was down-regulated in t4 cells compared to t0 cells; cells at the onset of light would have higher levels of transcripts for this gene (Appendix B:1). In Coesel et al. 2008, the maximum protein level of CPF1 was at 3 hours after the onset of light, compared to cells that had just experienced light- the fact that t4 cells in our study (4 hours after the onset of light)

showed less mRNA for PtCPF1 compared to t0 cells reflects a lag between the up-regulation of the gene and the maximal protein level.

PtCPF1 was slightly up-regulated in N cells compared to t0 cells, and highly up-regulated compared to t4 cells (Appendix B:1). This may indicate that a signal to down-regulate PtCPF1, which would have occurred in cells growing normally (t4 cells), was missing in N cells. Since N cells were still in G1, up-regulation of PtCPF1 might have reflected the delayed expression of a gene that is connected indirectly to the cell cycle; that is, PtCPF1 expression may be connected to signals other than light. In fact, Coesel et al. 2008 found that PtCPF1 overexpression resulted in changes in genes involved in cell-cycle progression (dsCYC2 was highly up-regulated in overexpression lines, for example). The results of the current study show that the reverse (changes in cell-cycle progression affect PtCPF1 expression) is also true. ND cells slightly up-regulated PtCPF1 compared to N cells; this may be the result of a continuous rise in PtCPF1 gene expression between hours 4 and 24 in N cells, though it is possible that transcript levels decreased sometime between 4 and 24 hours, in a semi-diurnal matter. Cells had been in continuous light between hours 4 and 24, and cultured in continuous light prior to dark-arrest; this could certainly have an effect on expression levels of this gene.

Expression of PtCPF2, a putative cryptochrome-DASH photoreceptor, was enhanced in PtCPF1 overexpression (Coesel et al. 2008). In the current study, it was up-regulated in t4 cells compared to t0 cells, and down-regulated in N cells (compared

to both t0 and t4 cells); regulation of this photoreceptor therefore follows the opposite pattern of expression from PtCPF1.

Analysis of gene expression of light-harvesting complexes and ROS scavenging systems

P. tricornutum is predicted to have at least 40 genes that encode for light-harvesting complexes (LHCs). These include the LHCFs, encoding the major fucoxanthin Chl a/c proteins, the red algal-like LHCRs, and the LI818-like LHCXs. The majority of LHCs in t4 cells (about 76% of LHC genes) were up-regulated compared to cells at t0. In N cells however, almost all light harvesting complexes were down-regulated, compared to t0 and t4 cells. Close to 80% of LHC genes were down-regulated compared to t4 cells, though LHCR8 and LHCX3 were up-regulated.

The general pattern of LHC expression was most similar to cells stressed by high light, as reported by Nymark et al. 2009 (Appendix B:3). In the Nymark study, cells that had been subjected to high light stress for three hours down-regulated most LHCs, except for LHCX2, LHCF15, LHCR8, and LHCX3, which were highly up-regulated. These LHCs were also up-regulated in N cells, compared to t4 cells, except LHCX2. This LHC was also found to be up-regulated under high light in a separate study in *P. tricornutum* (Bailleul et al. 2010). Similarly, nitrogen-stressed cells, as reported by Levitan et al. 2015, up-regulated LHCX3. This LHC, and LHCX4, were the only two LHCs that were significantly up-regulated in nitrogen-stressed cells. The LHCs that were up-regulated in all these conditions (nitrogen stress, high light, and cell cycle arrest) likely play roles in photoprotection. In the case of arrested cells,

cells have smaller “sinks” for photosynthetically generated electrons, and in nitrogen-stressed cells, photosynthetic machinery is compromised.

Reactive oxygen species (ROS) are produced in the chloroplast when light intensity is in excess of carbon fixation. Mechanisms to scavenge ROS have evolved to detoxify ROS and reduce oxidative stress. Of the genes that were involved in ROS scavenging, about 67% were up-regulated and 22% down-regulated in N cells compared to t0 cells, and 67% were up-regulated and 25% down-regulated in N cells compared to t4 cells, indicating that N cells were under greater oxidative stress than t0 or t4 cells (Appendix B:3). Some of these genes were also up-regulated in ND cells compared to N cells. This suggests that although ND cells have had more time to adapt to light conditions than N cells, they are still affected by excess light because they have a smaller “sink” for photosynthetically generated electrons, possibly due to the accumulation of storage products.

Phosphatidylinositol signaling genes

Phosphatidylinositols (PIs) are glycerophospholipids that form a component of eukaryotic cell membranes. When phosphorylated, they are called phosphoinositides, and are used as signals in the AKT/mTOR/p70S6K1 signaling pathway, which is involved in cell growth, proliferation, and membrane trafficking. In multicellular organisms, this pathway regulates apoptosis, and is overactive in tumor cells. Aberrant activation of this pathway has been identified in a wide range of cancers. Activated phosphatidylinositol 3-kinase (PI 3-kinase), for example, is an

enzyme that is sufficient to promote entry into S phase by Cdk activation (Klippel et al. 1998). Though this enzyme was only slightly up-regulated in N cells compared to t4 cells, many of the genes for other putative proteins involved in PI signaling were up-regulated in N cells, compared to both t0 and t4 cells (Appendix B:4). In contrast, most of these genes were NDE in t4 and ND cells. This could indicate that many of the genes for PI signaling are differentially expressed as cells initially adjust to cell cycle arrest.

Analysis of gene expression of Calvin cycle, pentose phosphate, lipid and chrysolaminarin pathways

Compared to t4 cells, N-cells down-regulated most genes of the Calvin cycle, including plastidic phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase, triosephosphate isomerase, and fructose 1,6-bisphosphate aldolase, by more than 3-fold (Appendix B:10). These genes were also down-regulated compared to t0 cells. N cells also down-regulated most genes for the pentose phosphate pathway in the cytosol, compared to both t0 and t4 cells (Appendix B:14). This is consistent with metabolite profiles in Chapter 3, which showed that immediately after the onset of light, there was a build up of most pentose phosphates in N cells compared to control cells, but with increasing time in the light, levels of these metabolites were lower than those of control cells.

Nitrogen-stressed cells accumulate neutral lipids, but down-regulate most genes involved in lipid synthesis (Levitan et al. 2015). In contrast, N cells up-regulated

most genes involved in fatty acid synthesis, compared to t4 cells (Fig. 2a). The genes involved in TAG metabolism are not well annotated, but of the five known diacylglycerol acyltransferases (DGATs) in *P. tricornutum*, four were upregulated in N cells (43469, 49462, 49544, 9794), and one was down-regulated (31662), compared to t4 cells (Fig. 2a). Acetyl-CoA carboxylase (55209, 54926), which performs the first committed step in fatty acid synthesis, was also up-regulated. These results point to a fundamental difference in the causes behind lipid accumulation in arrested cells, compared to nitrogen-starved cells. Both cases lead to cell cycle arrest or a lengthening of the cell cycle, as well as the redirection of photosynthetic output towards lipids, but nitrogen-starvation leads to breakdown of core photosynthetic proteins, chlorosis, and cannibalization or remobilization of plastid proteins and polar lipids toward energy storage in the form of TAGs (Levitan et al. 2013). The accumulation of TAGs in nitrogen-stressed cells is a consequence of carbon and reductant reallocation, rather than an up-regulation of lipid biosynthesis genes.

In N cells, most genes involved in fatty acid synthesis are not differentially expressed compared to t0 cells, and some are even down-regulated, indicating that many of these fatty acid synthesis genes are the most highly upregulated immediately at the onset of light (Fig. 2b). Many of the lipid synthesis genes up-regulated in N cells, were down-regulated, or NDE in ND cells (Fig. 2c). These include ACC1 (54926), ACC2 (55209), four DGATs (43469, 49462, 49544, 9794), and most of the genes for fatty acid synthesis. A large accumulation of neutral lipids

in ND cells may send a signal to down-regulate lipid synthesis, suggesting a natural limit for lipid accumulation in cells.

Fig. 2a-

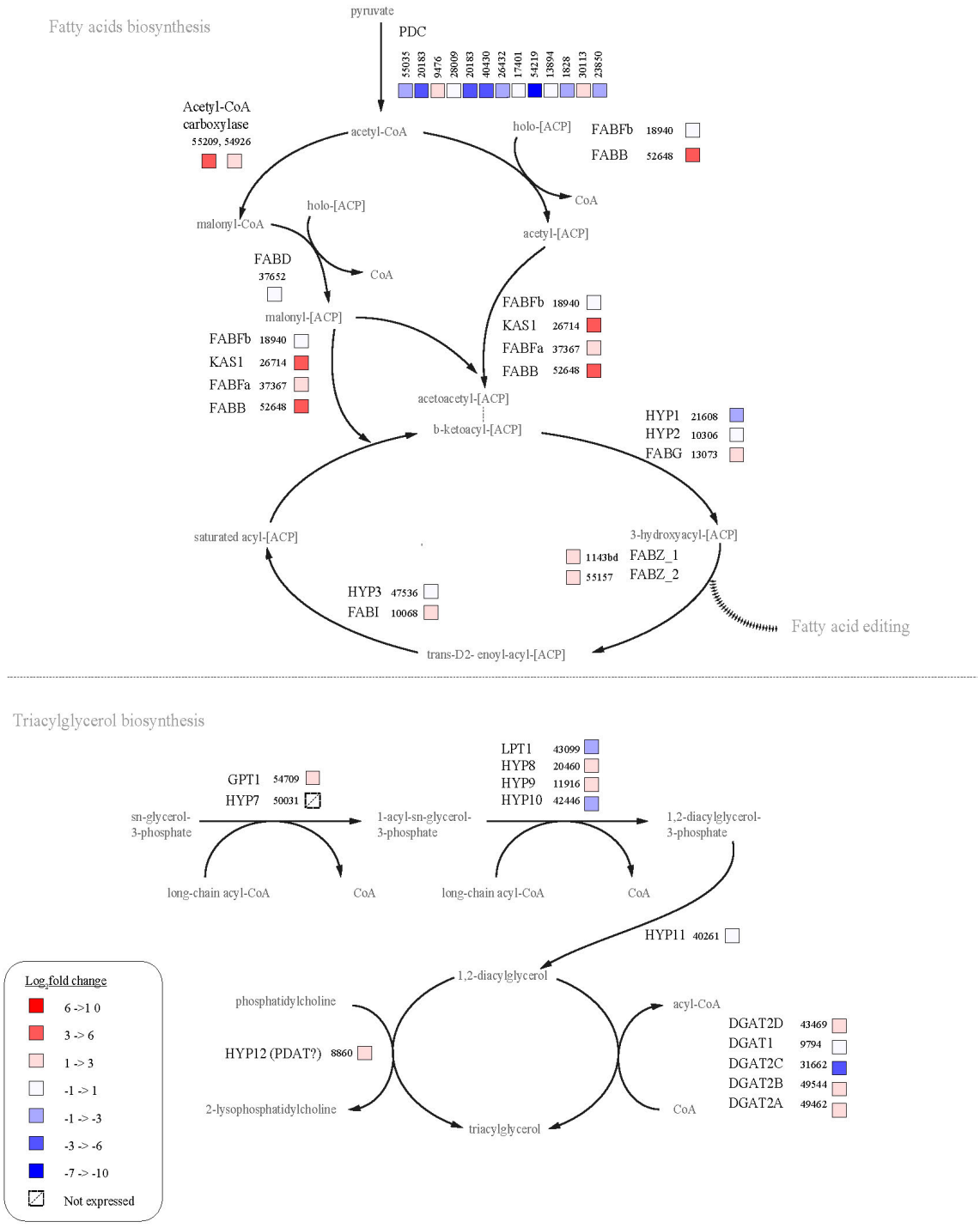

Fatty acids biosynthesis
pyruvate
PDC
55035 20183 9476 28009 20183 40430 26432 17401 54219 13894 1828 30113 23850
Acetyl-CoA carboxylase
55209, 54926
acetyl-CoA
holo-[ACP]
FABFb 18940
FABB 52648
CoA
malonyl-CoA
holo-[ACP]
acetyl-[ACP]
FABD
37652
CoA
malonyl-[ACP]
FABFb 18940
KAS1 26714
FABFa 37367
FABB 52648
acetoacetyl-[ACP]
b-ketoacyl-[ACP]
HYP1 21608
HYP2 10306
FABG 13073
saturated acyl-[ACP]
3-hydroxyacyl-[ACP]
1143bd FABZ_1
55157 FABZ_2
HYP3 47536
FABI 10068
Fatty acid editing
trans-D2- enoyl-acyl-[ACP]
Triacylglycerol biosynthesis
GPT1 54709
HYP7 50031
LPT1 43099
HYP8 20460
HYP9 11916
HYP10 42446
sn-glycerol-3-phosphate
1-acyl-sn-glycerol-3-phosphate
1,2-diacylglycerol-3-phosphate
long-chain acyl-CoA
CoA
long-chain acyl-CoA
CoA
HYP11 40261
1,2-diacylglycerol
phosphatidylcholine
HYP12 (PDAT?) 8860
2-lysophosphatidylcholine
triacylglycerol
acyl-CoA
CoA
DGAT2D 43469
DGAT1 9794
DGAT2C 31662
DGAT2B 49544
DGAT2A 49462
Log2 fold change
6 ->1 0
3 -> 6
1 -> 3
-1 -> 1
-1 -> -3
-3 -> -6
-7 -> -10
Not expressed

Fig. 2b-

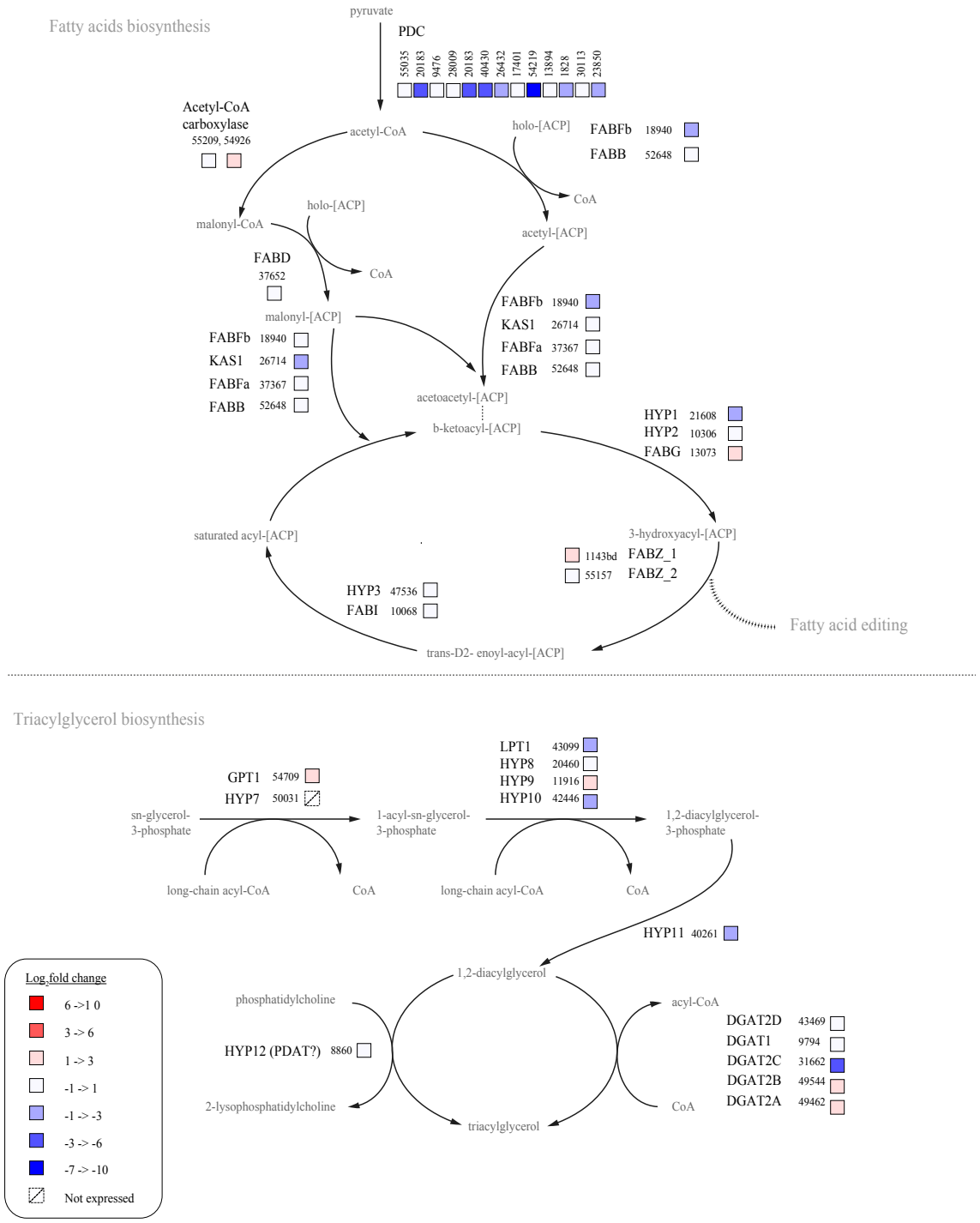

Fatty acids biosynthesis
pyruvate
PDC
55035 20183 9476 28009 20183 40430 26432 17401 54219 13894 1828 30113 23850
Acetyl-CoA carboxylase
55209, 54926
acetyl-CoA
holo-[ACP]
FABFb 18940
FABB 52648
CoA
malonyl-CoA
holo-[ACP]
acetyl-[ACP]
FABD
37652
CoA
malonyl-[ACP]
FABFb 18940
KAS1 26714
FABFa 37367
FABB 52648
FABFb 18940
KAS1 26714
FABFa 37367
FABB 52648
acetoacetyl-[ACP]
b-ketoacyl-[ACP]
HYP1 21608
HYP2 10306
FABG 13073
saturated acyl-[ACP]
3-hydroxyacyl-[ACP]
1143bd FABZ_1
55157 FABZ_2
HYP3 47536
FABI 10068
Fatty acid editing
trans-D2- enoyl-acyl-[ACP]
Triacylglycerol biosynthesis
GPT1 54709
HYP7 50031
LPT1 43099
HYP8 20460
HYP9 11916
HYP10 42446
sn-glycerol-3-phosphate
1-acyl-sn-glycerol-3-phosphate
1,2-diacylglycerol-3-phosphate
long-chain acyl-CoA
CoA
long-chain acyl-CoA
CoA
HYP11 40261
1,2-diacylglycerol
phosphatidylcholine
HYP12 (PDAT?) 8860
2-lysophosphatidylcholine
acyl-CoA
CoA
triacylglycerol
DGAT2D 43469
DGAT1 9794
DGAT2C 31662
DGAT2B 49544
DGAT2A 49462
Log2 fold change
6 ->1 0
3 -> 6
1 -> 3
-1 -> 1
-1 -> -3
-3 -> -6
-7 -> -10
Not expressed

Fig. 2c-

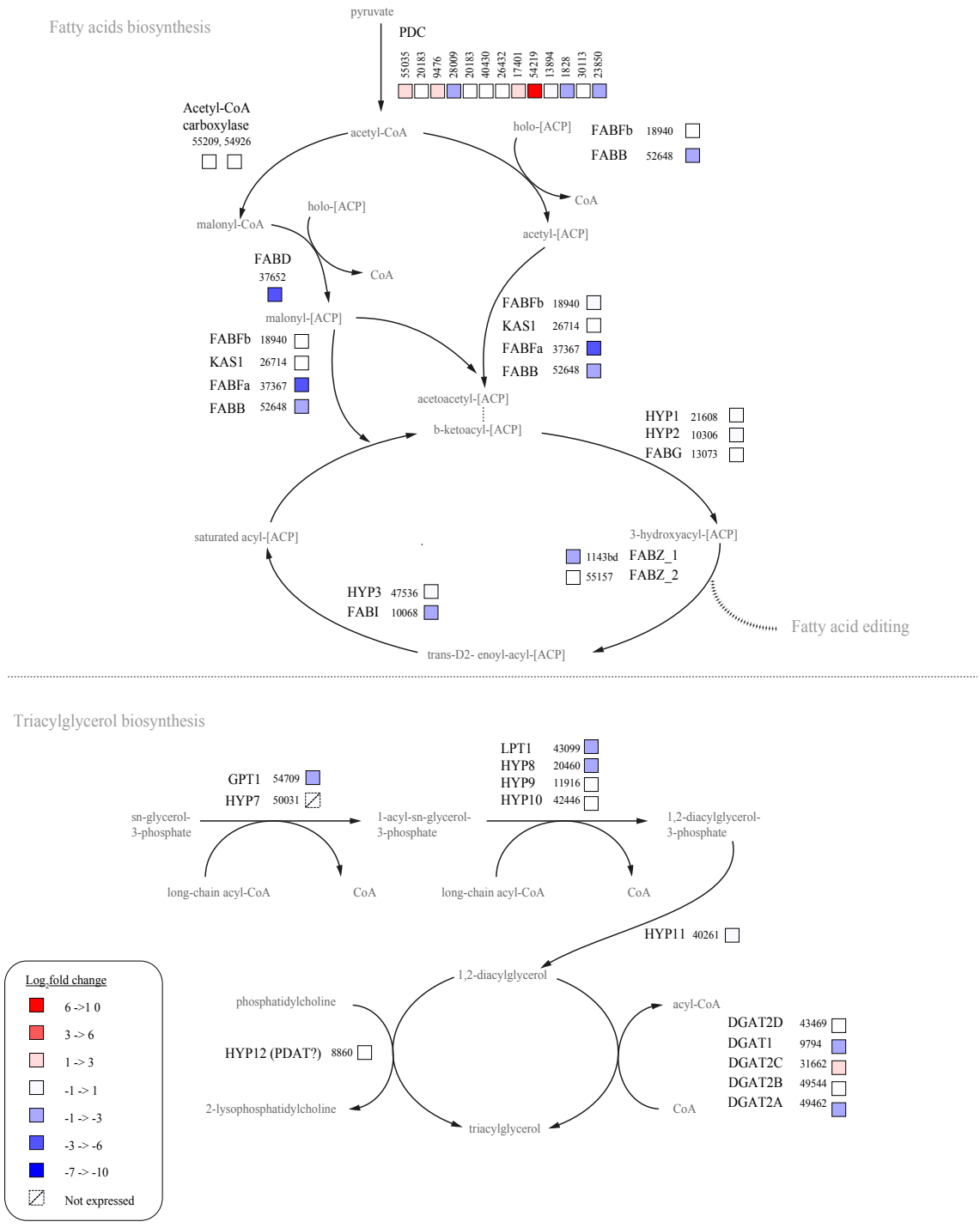

Fatty acids biosynthesis
pyruvate
PDC
55035 20183 9476 28009 20183 40430 26432 17401 54219 13894 1828 30113 23850
Acetyl-CoA carboxylase
55209, 54926
acetyl-CoA
holo-[ACP]
FABFb 18940
FABB 52648
CoA
malonyl-CoA
holo-[ACP]
acetyl-[ACP]
FABD
37652
CoA
malonyl-[ACP]
FABFb 18940
KAS1 26714
FABFa 37367
FABB 52648
FABFb 18940
KAS1 26714
FABFa 37367
FABB 52648
acetoacetyl-[ACP]
b-ketoacyl-[ACP]
HYP1 21608
HYP2 10306
FABG 13073
saturated acyl-[ACP]
3-hydroxyacyl-[ACP]
1143bd FABZ_1
55157 FABZ_2
HYP3 47536
FABI 10068
Fatty acid editing
trans-D2- enoyl-acyl-[ACP]
Triacylglycerol biosynthesis
GPT1 54709
HYP7 50031
LPT1 43099
HYP8 20460
HYP9 11916
HYP10 42446
sn-glycerol-3-phosphate
1-acyl-sn-glycerol-3-phosphate
1,2-diacylglycerol-3-phosphate
long-chain acyl-CoA
CoA
long-chain acyl-CoA
CoA
HYP11 40261
1,2-diacylglycerol
phosphatidylcholine
HYP12 (PDAT?) 8860
2-lysophosphatidylcholine
triacylglycerol
acyl-CoA
CoA
DGAT2D 43469
DGAT1 9794
DGAT2C 31662
DGAT2B 49544
DGAT2A 49462
Log2 fold change
6 ->1 0
3 -> 6
1 -> 3
-1 -> 1
-1 -> -3
-3 -> -6
-7 -> -10
Not expressed

Fig. 2- Differential expression of lipid synthesis genes (a) N vs t4, (b) N vs t0 and (c) ND vs N. Figure adapted from Levitan et al. 2015.

About 55% of genes involved in lipid degradation were down-regulated in N cells compared to t4 cells (Appendix B:6). In contrast, of these 38 genes, about 24% were up-regulated, 8% down-regulated, and the rest NDE, in t4 cells compared to t0 cells; t4 cells likely tapped into stored lipids for membranes and other components required for cell duplication. Interestingly, ND cells up-regulated more of these lipid-degradation genes compared to N cells; about 50% of these were up-regulated, and about 23% down-regulated compared to N cells. Lipid-degradation genes may have been up-regulated as a result of lipid accumulation in ND cells.

Genes involved in the synthesis of intact polar lipids (IPLs), which are associated with membranes, were down-regulated in N cells. These included enzymes for synthesizing monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), and sulfoquinovosyldiacylglycerol (SQDG). This was similar to nitrogen-stressed cells; both cases involved cell cycle arrest and a decreased need for IPLs.

In contrast to lipid synthesis genes, beta-glucan synthesis genes (50238, 48300) involved in chrysolaminarin synthesis were down-regulated in N cells, even though N cells had higher accumulation of chrysolaminarin, compared to t4 cells. Either transcription of chrysolaminarin synthesis genes does not follow activity, or both chrysolaminarin synthesis and degradation were higher in t4 cells compared to N cells.

Analysis of gene expression of nitrogen metabolism

In nitrogen-stressed cells, genes encoding enzymes for the TCA cycle are up-regulated, but the central hub involving pyruvate, phosphoenolpyruvate, and oxaloacetate is down-regulated. This intermediate pathway, which ultimately leads to the formation of acetyl-CoA, is unlikely to be a significant source of carbon for lipid synthesis. Interestingly, in N cells most of these enzymes (phosphoenolpyruvate carboxykinase, pyruvate kinase, and pyruvate dehydrogenase) are also down-regulated compared to t4 cells, except acetyl-CoA carboxylase (Fig. 3a). This implies that the acetyl-CoA that is fed into increased fatty acid synthesis in N cells does not come from this oxaloacetate/phosphoenolpyruvate/pyruvate hub.

Fig. 3a-

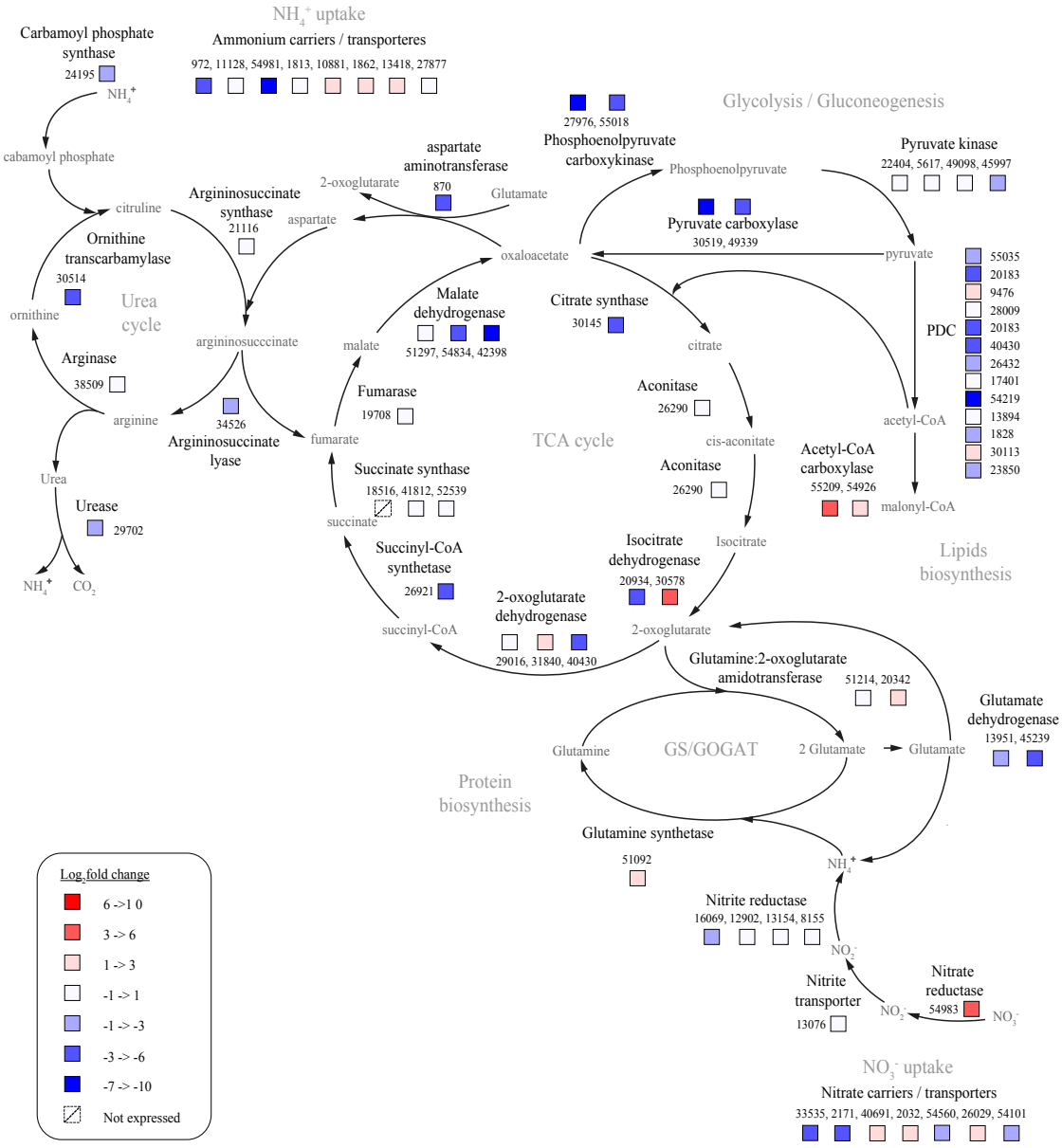

NH4+ uptake
Ammonium carriers / transporteres
972, 11128, 54981, 1813, 10881, 1862, 13418, 27877
Carbamoyl phosphate synthase
24195
NH4+
cabamoyl phosphate
citruline
Argininosuccinate synthase
21116
aspartate
2-oxoglutarate
aspartate aminotransferase
870
Glutamate
Ornithine transcarbamylase
30514
ornithine
Urea cycle
argininosuccinate
malate
Malate dehydrogenase
51297, 54834, 42398
oxaloacetate
27976, 55018
Phosphoenolpyruvate carboxykinase
Phosphoenolpyruvate
Glycolysis / Gluconeogenesis
Pyruvate kinase
22404, 5617, 49098, 45997
Pyruvate carboxylase
30519, 49339
pyruvate
Citrate synthase
30145
citrate
PDC
55035
20183
9476
28009
20183
40430
26432
17401
54219
13894
1828
30113
23850
Arginase
38509
arginine
34526
Argininosuccinate lyase
fumarate
Fumarase
19708
Aconitase
26290
cis-aconitate
TCA cycle
acetyl-CoA
Acetyl-CoA carboxylase
55209, 54926
malonyl-CoA
Urea
Urease
29702
NH4+
CO2
Succinate synthase
18516, 41812, 52539
succinate
Aconitase
26290
Isocitrate
Isocitrate dehydrogenase
20934, 30578
Lipids biosynthesis
Succinyl-CoA synthetase
26921
succinyl-CoA
2-oxoglutarate dehydrogenase
29016, 31840, 40430
2-oxoglutarate
Glutamine:2-oxoglutarate amidotransferase
51214, 20342
Glutamate dehydrogenase
13951, 45239
Glutamine
GS/GOGAT
2 Glutamate
Glutamate
Protein biosynthesis
Glutamine synthetase
51092
NH4+
Nitrite reductase
16069, 12902, 13154, 8155
NO2-
Nitrite transporter
13076
Nitrate reductase
54983
NO2-
NO3-
NO3- uptake
Nitrate carriers / transporters
33535, 2171, 40691, 2032, 54560, 26029, 54101
Log2 fold change
6 ->1 0
3 -> 6
1 -> 3
-1 -> 1
-1 -> -3
-3 -> -6
-7 -> -10
Not expressed

Fig. 3b-

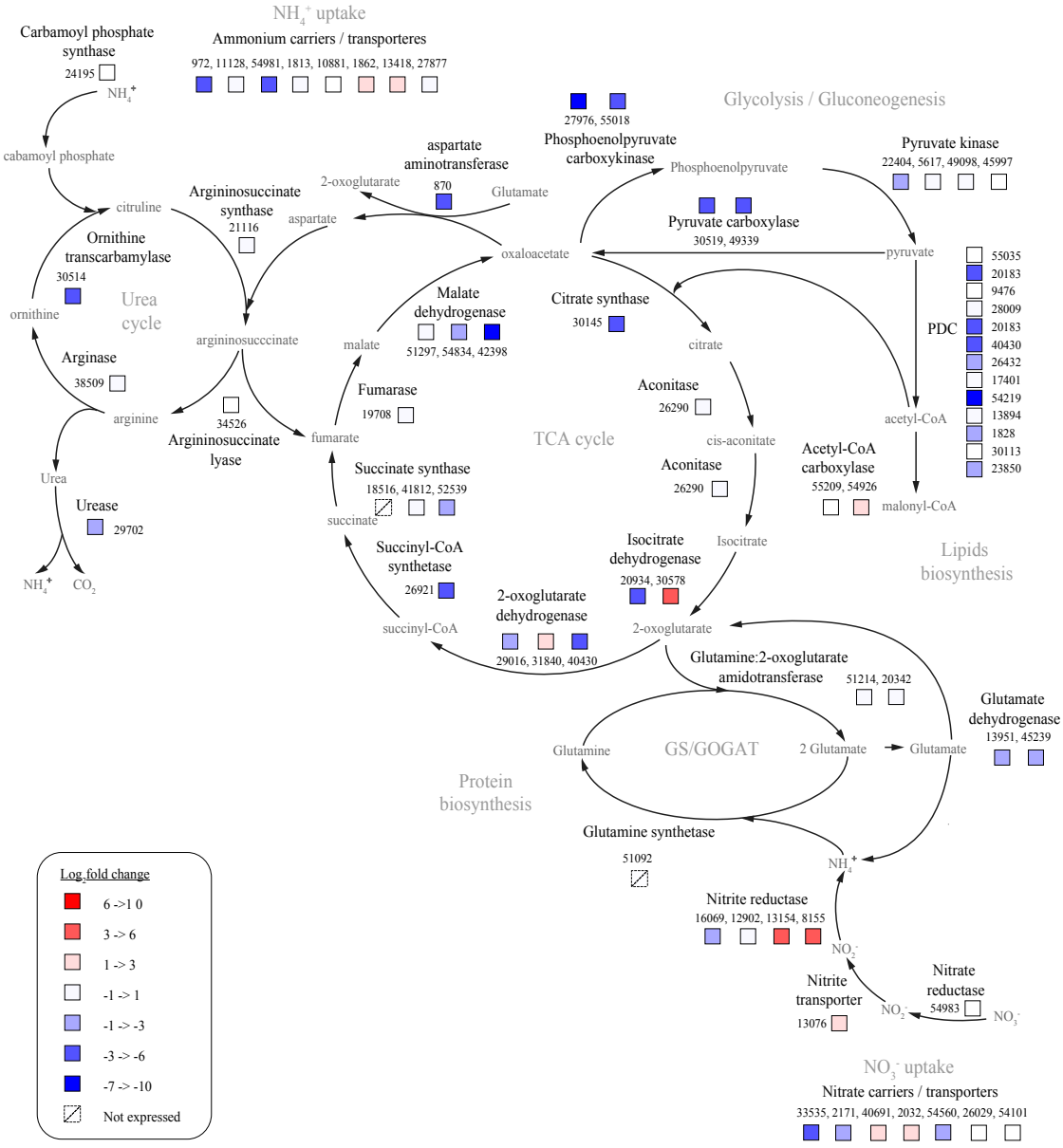

NH4+ uptake
Ammonium carriers / transporteres
972, 11128, 54981, 1813, 10881, 1862, 13418, 27877
Carbamoyl phosphate synthase
24195
NH4+
cabamoyl phosphate
citruline
Ornithine transcarbamylase
30514
ornithine
Urea cycle
Arginase
38509
arginine
Urea
Urease
29702
NH4+
CO2
Argininosuccinate synthase
21116
aspartate
argininosuccinate
34526
Argininosuccinate lyase
fumarate
Fumarase
19708
malate
Malate dehydrogenase
51297, 54834, 42398
2-oxoglutarate
aspartate aminotransferase
870
Glutamate
oxaloacetate
Glycolysis / Gluconeogenesis
27976, 55018
Phosphoenolpyruvate carboxykinase
Phosphoenolpyruvate
Pyruvate kinase
22404, 5617, 49098, 45997
Pyruvate carboxylase
30519, 49339
pyruvate
PDC
55035
20183
9476
28009
20183
40430
26432
17401
54219
13894
1828
30113
23850
Citrate synthase
30145
citrate
Aconitase
26290
cis-aconitate
Aconitase
26290
Isocitrate
TCA cycle
Isocitrate dehydrogenase
20934, 30578
2-oxoglutarate
2-oxoglutarate dehydrogenase
29016, 31840, 40430
succinyl-CoA
Succinyl-CoA synthetase
26921
succinate
Succinate synthase
18516, 41812, 52539
acetyl-CoA
Acetyl-CoA carboxylase
55209, 54926
malonyl-CoA
Lipids biosynthesis
Glutamine:2-oxoglutarate amidotransferase
51214, 20342
Glutamate dehydrogenase
13951, 45239
Glutamine
GS/GOGAT
2 Glutamate
Glutamate
Protein biosynthesis
Glutamine synthetase
51092
NH4+
Nitrite reductase
16069, 12902, 13154, 8155
NO2-
Nitrite transporter
13076
Nitrate reductase
54983
NO2-
NO3-
NO3- uptake
Nitrate carriers / transporters
33535, 2171, 40691, 2032, 54560, 26029, 54101
Log2fold change
6 ->1 0
3 -> 6
1 -> 3
-1 -> 1
-1 -> -3
-3 -> -6
-7 -> -10
Not expressed

Fig. 3c-

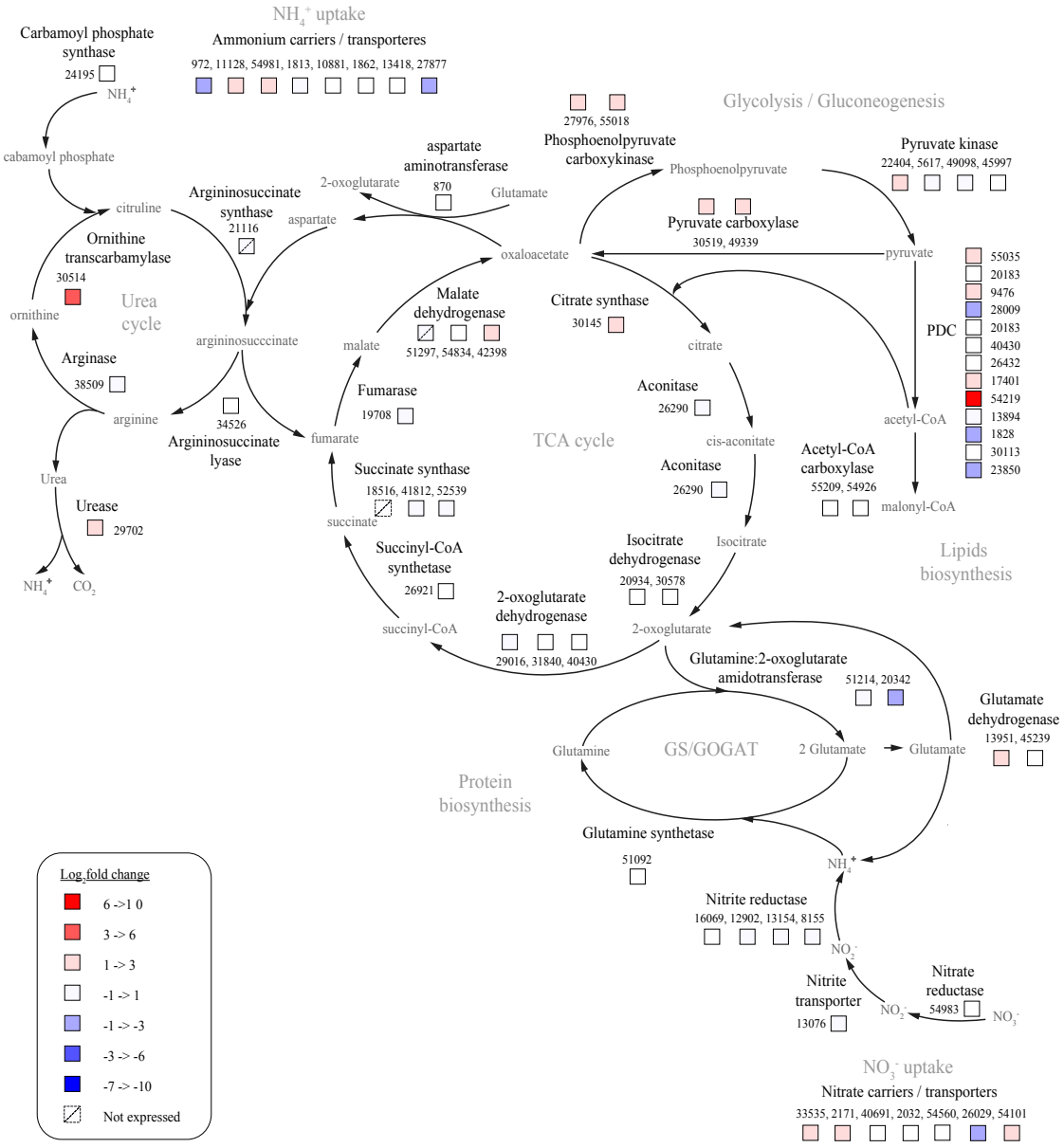


Fig. 3. Differential expression of TCA and nitrogen metabolism genes (a) N vs t4, (b) N vs t0 and (c) ND vs N. Figure adapted from Levitan et al. 2015.

N cells also down-regulated most genes involved in nitrogen metabolism, including those for the TCA cycle, urea cycle, and glutamate dehydrogenase (Fig. 3a). About 70% of genes involved in the TCA cycle, and roughly 60% of genes involved in urea metabolism were down-regulated in N cells (Appendix B:7). However, some of the enzymes of the GS/GOGAT hub (glutamine synthetase and glutamine 2_oxoglutarate amidotransferase) were up-regulated in N cells compared to t4 cells. These enzymes were down-regulated in ND cells, however, indicating that the up-regulation of these genes in N cells (at hour 4) is a temporary adjustment to conditions, perhaps as a response to upregulation of nitrate reductase, which was highly up-regulated in N cells.

The pattern of expression was fundamentally different from nitrogen-stressed cells, where a significant number of TCA genes were up-regulated, according to Levitan et al. 2015 (aspartate aminotransferase, citrate synthase, succinate dehydrogenase, aconitate hydratase, fumarase, succinyl-coA synthetase, and isocitrate dehydrogenase). These were all down-regulated in N cells, compared to both t0 (Fig. 3b) and t4 cells, with the exception of isocitrate dehydrogenase, which was up-regulated compared to t4 cells. The expression pattern for urea metabolism also differed significantly from nitrogen-stressed cells (Appendix B:11). A key difference between cells that are arrested because of nitrogen-stress vs. Cdk inhibition, is that nitrogen-stressed cells catabolize proteins in order to redirect intracellular nitrogen. Nitrogen-stressed cells are also likely to shunt their internal pools of

nitrogen toward nitrogen-assimilation machinery, so that they can quickly respond to changes in external nitrogen availability (Levitan et al. 2015).

Compared to t0 cells, t4 cells highly down-regulated NR (nitrate reductase), which catalyzes the rate-limiting step of nitrate assimilation- the two-electron transfer to reduce NO^{3-} to NO^{2-} using NAD(P)H as the reductant (Fig. 3b, Appendix B:8). Blue-light can induce NR expression (Ninneman, 1987, Lillo, 2006), which explains how at the onset of light (t0), NR transcripts are high, but have decreased by hour 4 (in t4 cells). As already mentioned, nitrate reductase was highly up-regulated in N cells compared to t4 cells, so in N cells, NR transcripts remained high after initial expression at the onset of light. Even though N cells would presumably have less nitrate requirement compared to t4 cells, N cells may have maintained higher NR expression as a short-term response, perhaps as a mechanism for dissipating excess reductant. Numerous studies have revealed that diatoms use NR as a way to dissipate overflow of electron energy; for example, at low temperatures, carbon uptake and metabolism are limited by temperature and unable to buffer sudden changes in the flow of electrons from the light reactions; in these cases, NR activity is a means to dissipate excess energy flow (Lomas and Gilbert, 1999).

ND cells, in contrast to N cells, down-regulated NR, indicating that either some other energy dissipation mechanism became more significant by hour 24, and/or light-harvesting complexes had been lowered to the point where NR was not utilized as heavily as an energy-dissipation mechanism (Fig. 3c).

NR activity assays (Chapter 3) were consistent with these results; they showed that cells with NU 2058 before 12 hours exhibited no significant difference in NR activity compared to control cells, but by 24 hours, NR activity had significantly decreased. That is, N cells had the same NR activity as control cells, even though they presumably had a lower nitrate assimilation requirement. The fact that both NR transcription and activity eventually went down in ND cells, supports the view that NR is used as a short-term energy dissipation mechanism. Alternatively, the buildup of metabolites that are end-products of nitrate assimilation (such as glutamine) may have also caused the eventual decline in NR activity in ND cells.

In Levitan et al. 2015, it was reported that an NR-knockdown strain (NR21) exhibited increased lipid accumulation without chlorosis or significant changes in photosynthetic capacity that normally accompany nitrogen-starvation. Transcriptome analysis of NR21 under nitrogen-replete conditions revealed that lipid accumulation occurred in this strain without up-regulation of glutamate and 2-oxoglutarate hub genes. It was further hypothesized that NR itself was a signal that could be involved in the allocation of reductant toward lipids, independent of protein or nitrogen shuffling (Levitan et al. 2015). This was consistent with the findings of the current study, since N cells exhibited higher NR expression, but most of the genes for the TCA cycle, urea cycle, glutamine synthetase, glutamate dehydrogenase, and glutamine 2-oxoglutarate amidotransferase were down-regulated or NDE.

Integration of gene expression data with the FBA model

The predicted fluxes from E-Flux2 are listed in Appendix A:10. All reaction and metabolite abbreviations are as described in Kim et al. 2015. Since the "biological objectives" were different for the conditions compared, biomass equations (objectives) were adjusted using experimental data (Appendix A:9). The biomass equation for t0 was set as the "average" cell as formulated for the autotrophic cell in Chapter 2. The t4 biomass equation was formulated relative to t0 cells; that is, the "goal" was the biomass of a cell at t0, minus the macromolecular components (DNA, amino acids, lipids, etc.) that were made between t0 and t4. For the N and ND conditions, the differences in macromolecules between the cells at t0 and the measured macromolecules at the respective time points for N and ND were taken to be the "goal" and these were called version 1 of N (N v. 1) and version 1 of ND (ND v. 1) respectively. Since the true "biological objectives" for N and ND conditions were unknown, a sensitivity analysis was performed whereby several simulations of the model using different biomass equations for these conditions, were compared. In total, two versions of N and four versions of ND were tested. Versions 2 for N and ND (N v. 2 and ND v. 2) were the same as N v. 1 and ND v. 1 except that the growth-associated ATP maintenance, the unknown ATP requirements associated with biosynthesis, polymerization and transport, was a third of the original (version 1), which was set to 29.89 mmol ATP/g dry wt/h based on the rationale discussed in Kim et al. 2015. Versions 2 for both N and ND were established to analyze flux

outputs when ATP requirements were drastically lower, since it is likely that growth arrest would lower growth-associated ATP-maintenance. The two other versions for ND (versions 3 and 4) were the same as versions 1 and 2 for ND respectively, except that the biomass was set to be the differences in macromolecules between the cell at N (instead of t0) and the measured biomass at the ND time point.

The sensitivity analysis revealed that the differences in these versions did not change the overall flux pattern (the reactions that carried flux, or the directions of reversible reactions), but did change the amount of flux through those reactions (Appendix A:11). The overall biomass output did not vary greatly, except versions 3 and 4 for ND had a higher biomass output than ND versions 1 and 2. The reason that changing the ATP growth requirement does not change the biomass output is because photon and carbon input fluxes are left unconstrained, so any increased ATP demands could be met by a combination of increased photon and/or carbon input to increase linear and/or cyclic electron flow. As a result, it is expected that the model will reliably predict the overall flux distribution pattern, though it is not expected to predict the specific quantitative values of individual fluxes.

Overall flux predictions

The overall flux distributions were similar to those predicted in Kim et al. 2015. There were some differences in some of the reactions of the mitochondria; for t4 and ND, these reactions serve as a source of glycerate that ultimately goes into the production of glycerol-3-phosphate, which is used for polar or neutral lipids. In the

t0 and N conditions, the enzymes of lower glycolysis run in the direction of gluconeogenesis, whereas these reactions run in the opposite direction for t4 and ND. In N and t0, the source of carbon intermediates to feed these reactions comes from serine-pyruvate transaminase and glycerate dehydrogenase, which convert pyruvate to hydroxypyruvate, and hydroxypyruvate to glycerate, respectively. The glycerate kinase converts glycerate to glycerate-3-phosphate. Mitochondrial gluconeogenesis then allows the mitochondria to export triose phosphates that can ultimately be sent into the plastid to support additional flux through the Calvin cycle. It also results in the conversion of pyruvate and serine into alanine and hydroxypyruvate, and supports the mitochondrial electron transport chain, since it results in the net removal of protons from the inner mitochondrial matrix. Interestingly, an examination of how the cell regulates transcription of enzymes involved in alanine production reveals that t4 cells synthesize most of their alanine using cytosolic alanine transaminase, while N cells rely more heavily on mitochondrial serine-pyruvate transaminase.

More generally, the model predicts that serine-pyruvate transaminase and glycerate dehydrogenase in the mitochondrion support triacylglycerol synthesis, by providing a source of glycerate. Moreover, whenever there is flux through serine-pyruvate transaminase, glycerate dehydrogenase and/or glycerate kinase in the mitochondria, the enzymes of glycolysis in the mitochondrion run in the direction of gluconeogenesis to export triose phosphates. This is because there is no other outlet for the glycerate-3-phosphate that is produced in the mitochondrion, unless the

organism has a transporter for glycerate-3-phosphate out of the mitochondrion. In the model, a transporter for glycerone-phosphate was included in order for the reactions of glycolysis/gluconeogenesis in the mitochondrion to not be a "dead end." Therefore, the model predictions may partially be a result of the transporters included in the model (which are not known). However, at the least, these enzymes are expressed and differentially regulated across these conditions, suggesting that these enzymes play a role in regulating carbon distribution. Moreover, transcripts and predicted fluxes are closely correlated for the lower glycolysis reactions in the mitochondrion for t4, N, and ND conditions (Appendix A:12). Cells at t0 did not express these genes, except for glyceraldehyde-3-phosphate dehydrogenase, which was lowly expressed.

Compared to t4 cells, N cells up-regulated most genes for fatty acid synthesis, as well as acetyl-CoA carboxylase, and four out of five known DGATs. When normalized to inorganic carbon input, the fluxes through fatty acid synthesis genes in N cells were 67% higher than in both t0 and t4 cells (Appendix A:10). In comparison, fluxes through the Calvin cycle enzymes were roughly comparable across conditions, when normalized to inorganic carbon input.

Transcripts involved in the glyoxylate cycle have shown good correlation with measured in vivo fluxes in *S. cerevisiae*; the genes for these pathways are known to be strongly transcriptionally regulated (Daran-Lapujade et al. 2004). The glyoxylate cycle is not well-annotated for *P. tricornutum*, however, and fluxes for two of the

reactions involved, malate synthase and isocitrate lyase (E.C. 2.3.3.9 and E.C. 4.1.3.1) were not consistent with transcripts.

Fig. 4a- Flux distribution predictions using E-Flux2 for t0.

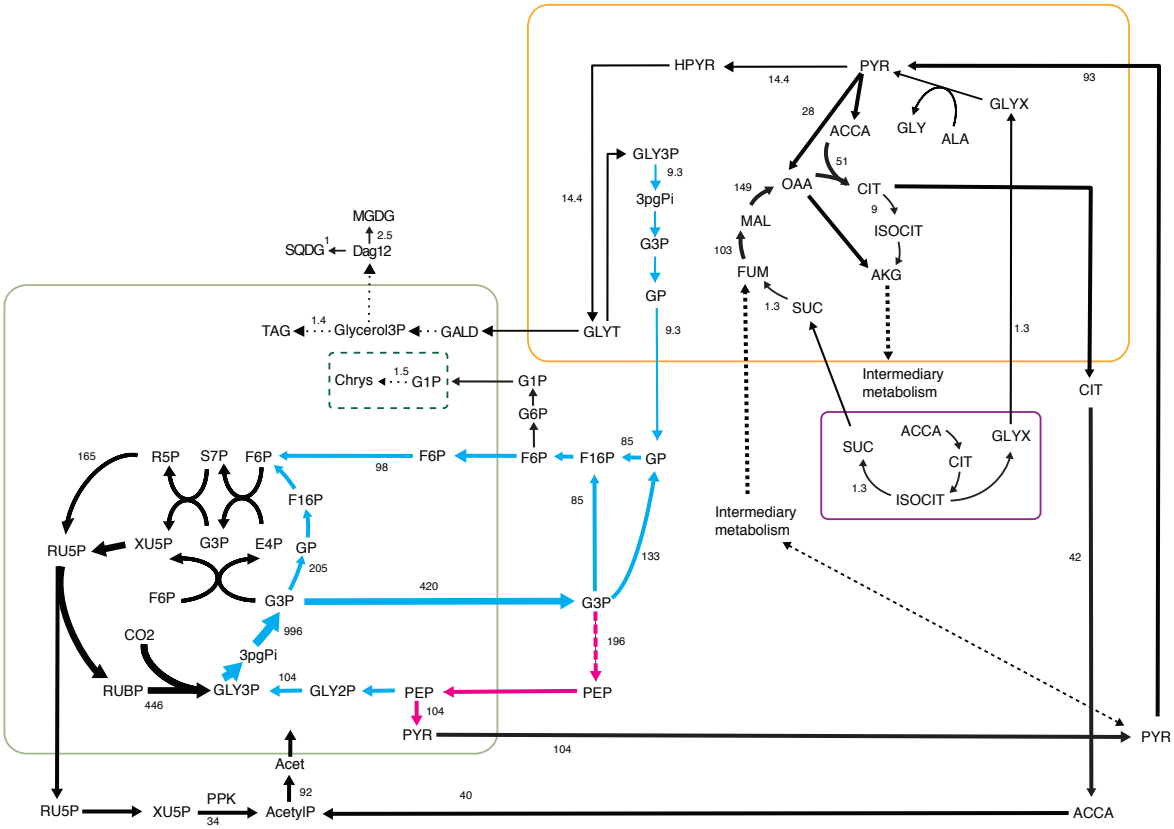


Fig. 4b- Flux distribution predictions using E-Flux2 for t4.

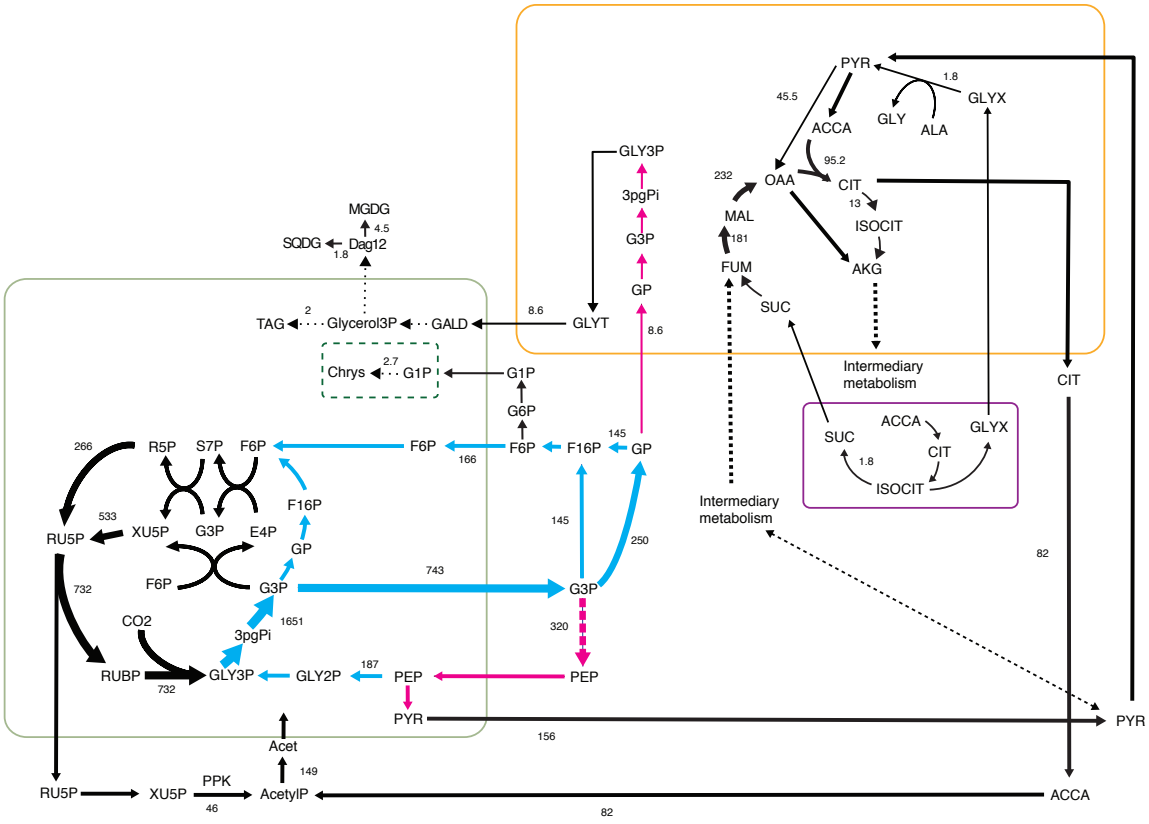


Fig. 4c- Flux distribution predictions using E-Flux2 for N v.1.

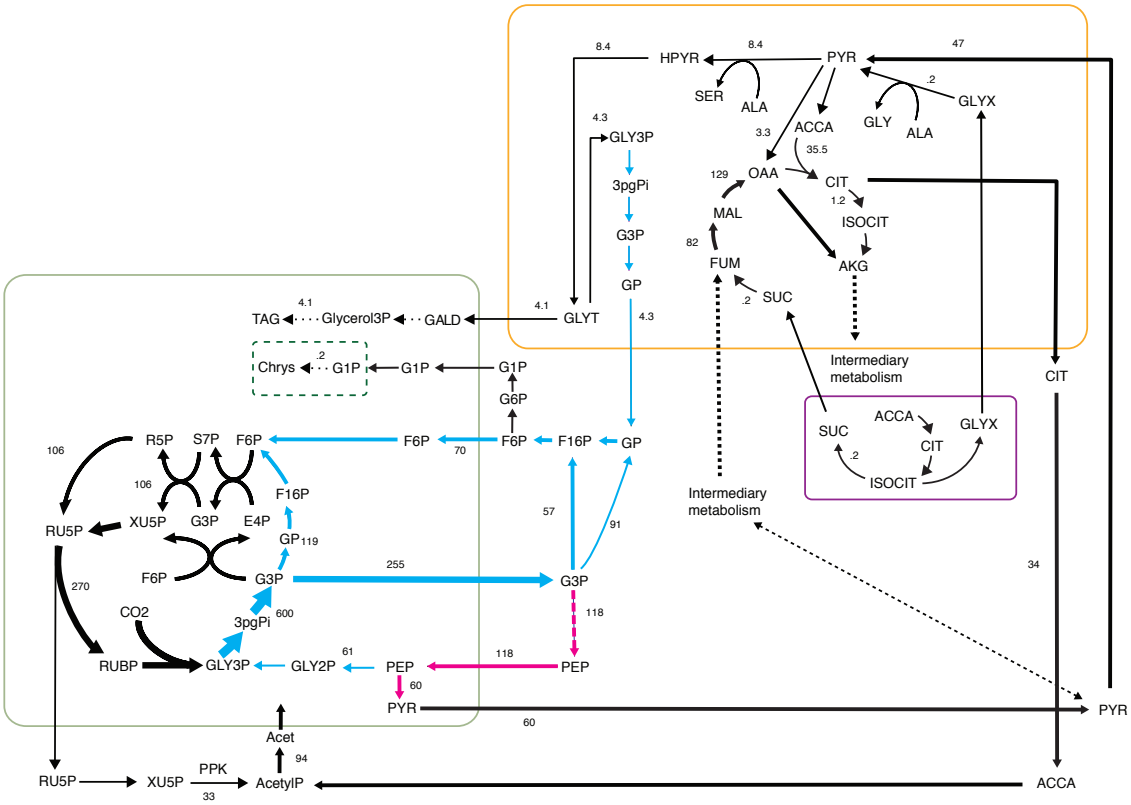


Fig. 4d- Flux distribution predictions using E-Flux2 for ND v. 1.

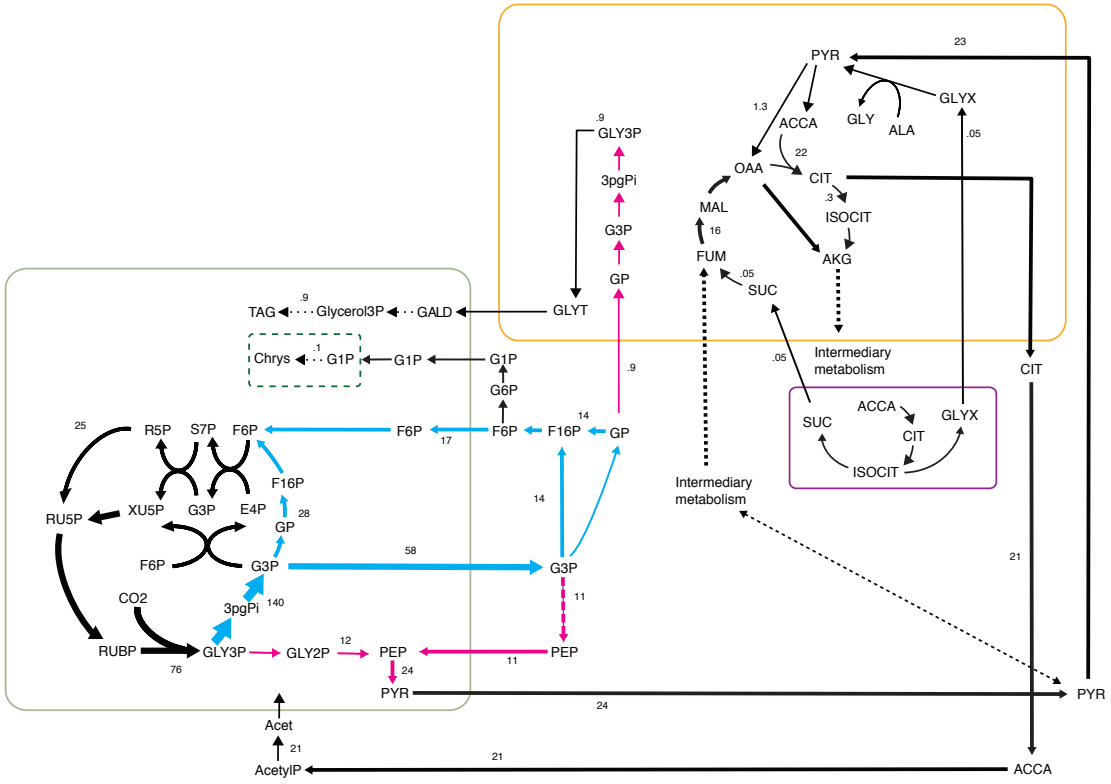


Fig. 4- Overall flux distribution predictions using E-Flux2 for (a) t0, (b) t4, (c) N v.1 and (d) ND v. 1. The thickness of the arrows corresponds to the relative amount of flux. Numbers for some fluxes were included. All metabolite abbreviations are from Kim et al. 2015.

Methods

RNA-Seq

P. tricornutum cultures synchronized in G1 phase (subjected to complete darkness for 24 hours) were split into cultures with and without NU 2058, placed in the light (~90 µmol photons•m^2•s^{-1}), and immediately collected for RNA. Samples for RNA-Seq from two independent biological replicates of each treatment were harvested by centrifuging cultures at 8000g using a Sorvall centrifuge at 4˚C, then transferring

them to RNase and DNase free microtubes that were immediately flash frozen in liquid N_2. RNA was extracted using an RLT lysis buffer and RNeasy Plant Mini Kit (Qiagen) following manufacturer's instructions, followed by removal of DNA contaminants with Ambion Turbo DNase (AM1907; Life Technologies). RNA quantification and quality assessment were made spectrophotometrically with a Nanodrop 1999 (Thermo Scientific). TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego) was used to prepare mRNA libraries for each of the samples according to the manufacturer's instructions. The 50-bp single-ended libraries were multiplexed and sequenced on an Illumina HiSeq2000 platform. The raw reads were trimmed for low-quality and adaptor sequences and aligned to *P. tricornutum*'s version 2.0 set of 10,402 filtered gene models using CLC Genomics Workbench (v6. 02) (Anders et al. 2010). Files were filtered to include uniquely aligned reads with no more than three mismatches. Gene counts (unique aligned reads per gene) were used for differentially expressed (DE) analysis carried out using the DESeq R/Bioconductor package (Anders et al. 2010), which infers DE based on the negative binomial distribution. A cutoff of 5% was used to control for false detection rate (false positives) and only genes that had a log two fold change greater than or equal to +/- 1 were considered DE.

Computation of Metabolic Flux Predictions

The *P. tricornutum* FBA model (Kim et al. 2015) was correlated with transcriptome data using E-Flux2, as described in Kim et al. 2015 (under review). In summary, with E-Flux2, mRNA levels are used as upper bounds on the maximum amount of

metabolic enzymes (reaction rates). If two (or more) genes are in OR relationship, the upper bound is calculated as the sum of those two genes' expression values. If they are in an AND relationship, the minimum (smaller) value between two expression level is chosen. The calculations performed had numerical tolerances such that there was a round-off error associated with all values. The lower bound on the uncertainty was 10^{-9}; and all flux values therefore have an uncertainty of at least 10^{-9} plus the uncertainties associated with the transcriptome measurements. Comparisons across conditions were made by first normalizing all fluxes to the flux of total incoming inorganic carbon.

Concluding Remarks

The inhibitor NU 2058 (6-(Cyclohexylmethoxy)-9*H*-purin-2-amine) binds to CDKA1 and CDKA2 in the diatom *Phaeodactylum tricornutum*, and leads to cell cycle arrest in G1 phase. Although cell cycle arrest does not lead to changes in photosynthetic attributes such as F_v/F_m, chlorophyll *a* levels, or levels of PsbD and RbcL (the subunits of PSII and RuBisCO enzymes, respectively), arrested cells exhibit gene expression patterns for LHC/ROS scavenging pathways similar to cells stressed by high light (Nymark et al. 2009); they down-regulate most LHC genes and up-regulate 67% of genes involved in ROS scavenging. They also have lower electron transport rates and higher non-photochemical quenching compared to actively growing cells, suggesting that they have a smaller sink for photosynthetically generated electrons. For this reason, arrested cells build up more pentose phosphates immediately upon the onset of light, compared to control cells. Pentose phosphates then gradually decrease with increasing time in the light.

Arrested cells have lower protein and nitrate reductase activity, and down-regulate most genes involved in nitrogen metabolism, compared to actively growing cells. Nitrogen-replete diatoms have been known to use nitrate reduction as a way to dissipate excess energy, but arrested cells do not use this mechanism over the long term (after twelve hours of arrest), suggesting that NR activity is utilized mainly as a short-term response to excess energy (Collos, 1998, Lomas et al. 2000). Alternatively, the build-up of photosynthetic products in arrested cells may lead to lower nitrate reductase activity; cellular carbon content and nitrate reductase

activity have been shown to be negatively correlated in the diatom *T. weissflogii* (Vergara et al. 1998).

In general, a higher glutamate (GLU) to glutamine (GLN) ratio has consistently been associated with nitrogen-deprivation in a number of different algae, and has been proposed as a biomarker for nitrogen-deplete cells (Flynn et al. 1988, Flynn et al. 1989, Frada et al. 2013, Guerra et al. 2013, Levitan et al. 2015). The GLU/GLN ratios for nitrogen-deprived and arrested cells show opposite tendencies; arrested cells have a lower GLU/GLN ratio, compared to actively growing cells. More generally, the GLU/GLN ratio may not just be a biomarker for nitrogen-stress, but an indicator for how much nitrogen is available relative to the nitrogen demand.

Cell cycle arrest leads to unbalanced growth, whereby photosynthetic products no longer needed for cell division are redirected into carbohydrate and neutral lipids (TAGs). Arrested cells after four hours of illumination up-regulate most genes involved in fatty acid synthesis, including acetyl-CoA carboxylase, which yields malonyl-CoA. Arrested cells also have higher malonyl-CoA and citrate levels. Citrate has been known to allosterically activate acetyl-CoA carboxylase; this suggests that arrested cells may have higher acetyl-CoA carboxylase activity, which leads to the build up of malonyl-CoA. Arrested cells also up-regulate three out of five putative diglyceride acyltransferases (DGATs), the enzymes that catalyze TAG production. This is in contrast to nitrogen-stressed cells, which accumulate neutral lipids via carbon reallocation, but do not up-regulate most genes involved in lipid synthesis,

and only up-regulate one diglyceride acyltransferase, DGAT2D (Levitan et al. 2015). The question remains why nitrogen-stressed cells up-regulate only one DGAT. Future studies should explore the role of the other DGATs (DGAT2B and DGAT2A) that are up-regulated under cell cycle arrest, through knockdown or overexpression methods.

Although arrested cells initially up-regulate most lipid synthesis genes, after a longer period of illumination (24 hours), they down-regulate most of these genes, suggesting that the accumulation of lipids in the cell sends a signal to down-regulate lipid synthesis. This is likely due to a natural limit for lipid accumulation in cells. Future endeavors to increase lipid accumulation in algae will have to find ways to "milk" algal cells for their TAGs, in order to prevent eventual down-regulation of lipid synthesis pathways, and tap algae for their full potential.

Flux balance analysis (FBA) predicts that there is energetic coupling between the mitochondrion and plastid, when there is sufficient communication (transporters) between the organelles. Knowledge of transporters should lead to more accurate predictions of the role that mitochondria play in optimizing photosynthesis, and the role of glycolytic enzymes in the mitochondrion. The model predicts that mitochondrial enzymes generate glycerate that can be used for glycerolipid production. Results correlating the transcriptome and the FBA model predict that mitochondrial serine-pyruvate transaminase and hydroxypyruvate reductase yield glycerate that can be used for TAG production under cell cycle arrest. It also

predicts that flux through these enzymes causes the glycolytic enzymes in the mitochondrion to run in the direction of gluconeogenesis to export triose phosphates from the mitochondrion. Further work using ^{13}C labeled metabolites should verify these findings, and explore whether mitochondrial serine-pyruvate transaminase and hydroxypyruvate reductase affect glycerolipid synthesis.

E-Flux2, the method used to correlate transcripts with the FBA model, works under the assumption that there is a suitable biological objective- maximizing biomass. Since the biological objectives for arrested cells are essentially unknown, more accurate flux predictions may have to come from methods that attempt to maximize correlation between gene expression data and corresponding flux vectors, without trying to maximize an objective function such as biomass production. More generally, the question remains what biological objectives really correspond to in a cell. In a sense, it can be seen as the "goal" of the cell, but the deeper question remains: how does a cell "know" it has reached its goal? How does it sense that it has built up enough cellular constituents for it to proceed through the cell cycle and divide? For many pathways, the build up of end-product metabolites causes a negative feedback regulation, which may explain why there are natural limits for the accumulation of any metabolite in a cell. In come cases, cell cycle proteins directly regulate enzymes of metabolism, such as lipases (Kurat et al. 2009). Future studies should investigate this possibility in *P. tricornutum*, by uncovering the substrates of cell cycle proteins.

Although cell cycle arrest via chemical inhibition of Cdks is likely toxic and prohibitively expensive for use in large-scale cultures for biofuels production, it can be achieved by the knockdown of cell cycle proteins (Huysman et al. 2015). Ideally, knockdown can be under the control of an inducible promoter. For the purposes of biofuels production, the ideal system would first allow cultures a period of rapid growth, followed by the induction of cell cycle arrest. At this point, cells could be "milked" or harvested for their lipids. Cell cycle arrest via Cdk inhibition is a promising means towards lipid accumulation, and its application could apply to other storage products, such as carbohydrates. More generally, it is a way to vary the rate of synthesis of cellular components relative to each other, or manipulate the cell into a state of unbalanced growth.

Bibliography

Acosta JAT, Engler JA, Magyar JRZ, Inze RDG, Veylder LD (2004) Molecular characterization of Arabidopsis PHO80-like proteins, a novel class of CDK;1-interacting cyclins. *CMLS* 61: 1485-1497.

Akesson M, Forster J, Nielson J (2004) Integration of gene expression data into genome-scale metabolic models. *Met Eng* 6: 285-293.

Allen JF (2002) Photosynthesis of ATP-electrons, proton pumps, rotors, and poise. *Cell* 110**:** 273–276.

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Gen Biol* 11:R106.

Arris CD, Boyle FT, Calvert AH, Curtin NJ, Endicott JA, Garman EF, Gibson AE, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Lawrie AM, Newell DR, Noble ME, Sausville EA, Schulz R, Yu W (2000) Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J Med Chem* 43(15): 2797-804.

Ast M, Gruber A, Schmitz-Esser S, Neuhaus HE, Kroth PG, Horn M, and Haferkamp I (2009) Diatom plastids depend on nucleotide import from the cytosol. *P Natl Acad Sci* 106(9): 3621-3626

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Bach S, Knockaert M, Reinhardt J, Lozach O, Schmitt S, Baratte B, Koken M, Coburn SP, Tang L, Jiang T, Liang D, Galons H, Dierick JF, Pinna LA, Meggio F, Totzke F, Schachtele C, Lerman AS, Carnero A, Wan Y, Gray N, Meijer L (2005) Roscovitine targets, protein kinases and pyridoxal kinase. *J Biol Res* 280(35):31208-31219.

Badour SS, Gergis MS (1965) Cell division and fat accumulation in Nitzchia sp. grown in continuously illuminated mass cultures. *Archiv Mikrobiol* 51: 94-102**.**

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D, Krieger-Liszkay A, Santabarbara S, Rappaport F, Joliot P, Tirichine L, Falkowski PG, Cardol P, Bowler C, and Finazzi G (2015) Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* 524: 366-269.

Baker DJ, Dawlaty MM, Galardy P, van Deursen JM (2007) Mitotic regulation of the anaphase-promoting complex. Cell. *Mol Life Sci* 64: 589–600.

Barrôco RM, De Veylder L, Magyar Z, Engler G, Inzé D, Mironov V. (2003) Novel complexes of cyclin-dependent kinases and a cyclin-like protein from *Arabidopsis thaliana* with a function unrelated to cell division. *Cell Mol Life Sci* 60:401–412. doi: 10.1007/s000180300033.

Becker KP and Hannun YA (2005) Protein kinase C and phospholipase D: intimate interactions in intracellular signaling. *Cell Mol Life Sci* 62**:** 1448–1461

Bennette NB, Eng J and Dismukes GC (2011) An LC–MS-Based Chemical and Analytical Method for Targeted Metabolite Quantification in the Model Cyanobacterium *Synechococcus* sp. PCC 7002. *Anal Chem* 83(10): 3808-3816.

Berges J and Harrison PJ (1995) Nitrate reductase activity quantitatively predicts rate of nitrate incorporation under steady state light limitation on photosynthetic efficiency of photosystems I and II in microalgae. *Plant Physiol* 110:689-96.

Berges AJ, Franklin JD, and Harrison PJ (2001) Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water the last two decades. *J Phycol* 37: 1138–45.

Bertsekas DP (2003) *Convex analysis and optimization*. Athena Scientific

Blackstone NW (1995) A units-of-evolution perspective on the endosymbiotic theory of the origin of the mitochondrion. *Evolution* 49: 785-796.

Blank L, Kuepfer L, and Sauer U (2005) Large-scale ^{13}C-flux analysis reveals mechanistic principles network robustness to null mutations in yeast. *Genome Biol* 6:R49. doi:10.1186/gb-2005-6-6-r49

Bonarius HPJ, Vassily H, Meesters KPH, de Gooijer CD, Schmid G, and Tramper J (1996) Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnol Bioeng* 50(3): 299-318.

Bowler C, Allen AE, Badger, *et al*. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456(7219), 239-244.

Boyle NR, and Morgan JA (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Sys Biol* 3:4 doi:10.1186/1752-0509-3-4

Brown MR (1991) The amino-acid and sugar composition of 16 species of microalgae used in mariculture. *J Exp Mar Biol* 145: 79-99.

Buchanan BB, Kalberer PP and Arnon DI (1967) Ferredoxin-activated fructose diphosphatase in isolated chloroplasts. *Biochem Biophys Res Commun* 29**:** 74-79.

Burkhardt S, Amoroso G, Riebesell U, and Sultemeyer D (2001) CO_2 and HCO_3^- uptake in marine diatoms acclimated to different CO_2 concentrations. *Limnol Oceanogr* 46(6): 1378-1391.

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, and Karp PD (2012) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nuc Acids Res* 40(D1): D742-D753.

Cerón García MC, Sánchez Mirón A, Fernández Sevilla JM, Molina Grima E, García Camacho F (2005) Mixotrophic growth of the microalga *Phaeodactylum tricornutum*: Influence of different nitrogen and organic carbon sources on productivity and biomass composition. *Proc Biochem* 40(1): 297-305.

Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, Shen Y, Hao T, Palsson BO, Salehi-Ashtiani K, and Papin JA (2011) Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol Sys Biol* 7:518.

Cheirsilp B, Torpee S (2011) Enhanced growth and lipid production of microalgae under mixotrophic culture condition: Effect of light intensity, glucose concentration and fed-batch cultivation. *Bioresource Technol* 110:510-516.

Chisti Y (2007) Biodiesel from microalgae. *Biotech Adv* 25(3) 294-306.

Clarens AF, Resurreccion EP, White MA, and Colosi LM (2010) Environmental life cycle comparison of algae to other bioenergy feedstocks. *Environ Sci Technol* 44(5): 1813-1819.

Claros MG, and Vincens P (1996) Computation method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 241: 779-786.

Coesel S, Oborník M, Varela J, Falciatore A, Bowler C (2008) Evolutionary Origins and Functions of the Carotenoid Biosynthetic Pathway in Marine Diatoms. *PLoS ONE* 3(8): e2896. doi:10.1371/journal.pone.0002896

Collos, Y (1998) Nitrate uptake, nitrite release and uptake and new production estimates. *Mar Ecol Prog Ser* 171: 293-301.

Covert MW, Palsson BO (2002) Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *J Biol Chem* 277(31):28058-28064.

Daran-Lapujade P, Jansen MLA, Daran JM, van Gulik W, de Winde JH, Pronk JT (2004) Role of transcriptional regulation in controlling fluxes in central carbon metabolism in *Saccharomyces cerevisiae*. *J Biol Chem* 279(10): 9125-9138.

Deshaies RJ (1999) SCF and cullin/RING H2-based ubiquitin ligases. *Annu Rev Cell Dev Biol* 15: 435-467.

Dubinsky Z, Falkowski PG, and Wyman K (1986) Light harvesting and utilization by phytoplankton. *Plant cell Physiol* 27: 1335-49.

Dubois M, Gilles KA, Hamilton JK, Rebers PA, and Smith F (1956) Colorimetric method for determination of sugars and related substances. *Anal Chem* 28(3): 350-356.

Durrett TP, Bennine C, Ohlrogge J (2008) Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J* 54(4): 593-607.

Emanuelsson O, Nielsen H, Brunak S, and von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005-1016.

Fabris M, Matthijs M, Rombauts S, Vyverman V, Goossens A, and Baart GJE (2012) The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J* 70 (6): 1004-14.

Falkowski PG, Dubinsky Z, and Wyman K (1985) Growth-irradiance relationships in phytoplankton. *Limnol Oceanogr* 30(2): 311-321.

Falkowski PG, et al. (2004) The evolution of modern eukaryotic phytoplankton. *Science* 305(5682): 354-360.

Falkowski PG and Raven JA (2007) *Aquatic Photosynthesis*, 2nd ed. Princeton University Press, Princeton, NJ.

Famili I, Forster J, Nielsen J, Palsson BO (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *P Natl Acad Sci* 100(23):13134-13139.

Fernandez-Reiriz MJ, Perez-Camacho A, Ferreiro MJ, Blanco J, Planas M, Campos MJ, and Labarta U (1989) Biomass production and variation in the biochemical profile (total protein, carbohydrates, RNA, lipids, and fatty acids) of seven species of marine microalgae. *Aquaculture* 83: 17-37.

Field CB, Behrenfeld MJ, Randerson JT, and Falkowski PG (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281(5374): 237-240.

Fischer K (2011) The import and export business in plastids: Transport processes across the inner envelope membrane. *Plant Phys* 155: 1511-1519.

Flynn KJ, Al-Amoudi OA (1988) Effects of N deprivation and darkness on composition of free amino acid pool in and on amino acid release from diatom*Phaeodactylum tricornutum* Bohlin. *J Exp Mar Biol Ecol* 119(2):131-143.

Flynn KJ, Dickson DMJ, Al-Amoudi OA (1989) The ratio of glutamine:glutamate in microalgae: a biomarker for N-status suitable for use at natural cell densities. *J Plankton Res* 11(1): 165-170.

Fogg GE (1956) Photosynthesis and formation of fats in a diatom. *Ann Bot* 200(2): 265-285.

Frada MJ, Burrows EH, Wyman KD and Falkowski PG (2013) Quantum requirements for growth and fatty acid biosynthesis in the marine diatom *Phaeodactylum tricornutum* (Bacillariophyceae) in nitrogen replete and limited conditions. *J Phycol* 49:381-388.

Giege P, Heazlewood JL, Roessner-Tunali U, Millar A, Fernie AR, Leaver CJ *et al.* (2003) Enzymes of glycolysis are functionally associated with the mitochondrion in *Arabidopsis* cells. *Plant Cell* 15: 2140-2151.

Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. *Nuc Acids Res* 20:1425.

Ginger ML, McFadden GI, and Michels PAM (2010) Rewiring and regulation of cross-compartmentalized metabolism in protists. Phil Trans Roy Soc London. Series B, *Biol Sci* 365: 831-845.

Glotzer M, Murray AW, Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. *Nature* 349(6305):132-8.

Gorbunov MY and Falkowski PG (2004) Fluorescence induction and relaxation (FIRe) technique and instrumentation for monitoring photosynthetic processes and primary production in aquatic ecosystems. In van der Est A and Bruce D (Eds) *Photosynthesis: Fundamental Aspects to Global Perspectives*. Proceedings of the 13th International Congress of Photosynthesis. Allen Press, Oxford.

Graham JM, Graham LE, Zulkifly SB, Pfleger BF, Hoover SW, and Yoshitani J (2011) Freshwater diatoms as a source of lipids for biofuels. *J Ind Microbiol Biotechnol* 39(3): 419-28.

Guckert JB, Cooksey KE (1990) Triglyceride accumulation and fatty acid profile changes in *Chlorella* (Chlorophyta) during high pH-induced cell cycle inhibition. *J Phycol* 26: 72-79.

Guerra LT, Levitan O, Frada MJ, Sun JS, Falkowski PG, Dismukes GC (2013) Regulatory branch points affecting protein and lipid biosynthesis in the diatom *Phaeodactylum tricornutum*. *Biomass Bioenerg* 59: 306-315.

Guillard RR and Ryther JH (1962) Studies of marine plankton diatoms. I. *Cyclotella nana* (Husted) and *Detonula confervacea* (Cleve). *Can J Microbiol* 8: 229-239.

Harbinson J, Foyer CH (1991) Relationships between the efficiencies of photosystem I and II and stromal redox state in CO2-free air. *Plant Phys* 97(1):41-49.

Harper JW, Koepp X, Ye X, Jin JP, Elledge SJ (2002) Cell cycle control by the SCF ubiquitin ligase. *FASEB J* 16: A740.

Harrison PJ, Waters RE, and Taylor FJR (1980) A broad spectrum artificial seawater medium for coastal and open ocean phytoplankton. *J Phycol* 16: 28-35.

Hildebrand M, Frigeri LG, Davis AK (2007) Synchronized growth of *Thalassiosira pseudonana* (Bacillariophyceae) provides novel insights into cell-wall synthesis processes in relation to the cell cycle. *J Phycol* 43: 730-740.

Hildebrand M, Davis AK, Smith SR, Traller JC, and Abbriano R (2012) The place of diatoms in the biofuels industry. *Biofuels* 3(2): 221-240.

Hoefnagel MHN Atkin OK, Wiskich JT (1998) Interdependence between chloroplasts and mitochondria in the dark. *Biochim Biophys Acta* 1366: 235-255.

Hoffman JP (1998) Wastewater treatment with suspended and nonsuspended algae. *J Phycol* 34(5): 757-763.

Holmgren A, Buchanan BB, and Wolosiuk RA (1977) Photosynthetic regulatory protein from rabbit liver is identical with thioredoxin. *FEBS Lett* 82(2): 351 – 354.

Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J* 54(4):621-39.

Huang A, Liu L, Yang C, Wang G (2015) *Phaeodactylum tricornutum* photorespiration takes part in glycerol metabolism and is important for nitrogen-limited response. *Biotechnol Biofuels* 8:73

Huysman MJJ, Martens C, Vandepoele K, Gillard J, Rayko E, Heijde M, Bowler C, Inze D, Peer YV, De Veylder L, Vyverman W (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol* 11: R17 doi:10.1186/gb-2010-11-2-r17.

Huysman MJJ, Martens C, Vyverman W, De Veylder L (2014) Protein degradation during the diatom cell cycle" Annotation and transcriptional analysis of SCF and

APC/C ubiquitin ligase genes in *Phaeodactylum tricornutum. Mar Genomics* 14:39-46.

Huysman MJJ, Tanaka A, Bowler C, Vyverman W, De Veylder L, (2015) Functional characterization of the diatom cyclin-dependent kinase A2 as a mitotic regulator reveals plant-like properties in a non-green lineage. *BMC Plant Biol* 15:86.

Izallalen M, Mahadevan R, Burgard A, Postier B, Didonato R, Sun J, Schilling CH, and Lovley DR (2008) *Metab Eng* 10: 267-275.

Jeffrey SW, and Humphrey GF (1975) New spectrophotometric equations for determining chlorophylls a, b, c1, and c2 in higher plants, algae and natural phytoplankton. *Biochem Physiol Pflanz* 167: 191-194.

Kanehisa M, and Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27-30.

Karady M, Novak O, Horna A, Strnad M, Dolezal K (2011) High performance liquid chromatography-electrochemistry-electrospray ionization mass spectrometry (HPL/EC/ESI-MS) for detection and characterization of roscovitine oxidation products. *Electroanal*. 23(12):2898-2905.

Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, and Pellegrini-Toole A (2000) The EcoCyc and MetaCyc databases. *Nucleic Acids Res* 28**:**56-9

Keeling PJ, and Doolittle WF (1997) Evidence that eukaryotic triosephosphate isomerase is of alpha- proteobacterial origin. *P Natl Acad Sci* USA 94:1270–1275.

Kelley JJ, Lane A, Li X, Mutthoju B, Maor S, Egen D, and Lun D (2014) MOST: a software environment for constraint-based metabolic modeling and strain design. *Bioinformatics* 31(4): 610-611.

Kim J, Fabris M, Baart G, Kim MK, Goossens A, Vyverman W, Falkowski PG, Lun DS (2015) Flux balance analysis of primary metabolism in *Phaeodactylum tricornutum*. *Plant J.* 85(1):161-176.

Kim MK, Lane A, Kelley JJ, Lun DS (2015) E-Flux2 and SPOT: Validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. *Under Review.*

King RW, Deshaies RJ, Peters JM, Kirschner MW (1996) How proteolysis drives the cell cycle. *Science* 274: 1652-59.

Klanchui A, Khannapho C, Phodee A, Cheevadhanarak S, and Meechai A (2012) iAK692: A genome-scale metabolic model of *Spirulina platensis* C1. *BMC Sys Biol* 6:71.

Klippel A, Escobedo M, Wachowicz MS, Apell G, Brown TW, Giedlin MA, Kavanaugh WM, Williams LT (1998) Activation of Phosphatidylinositol 3-Kinase Is Sufficient for Cell Cycle Entry and Promotes Cellular Changes Characteristic of Oncogenic Transformation. *Mol Cell Biol* 18(10) 5699-5711.

Knoop H, Grundel M, Zilliges Y, Lehmann R, Hoffman S, Lockau W, and Steuer R (2013) Flux balance analysis of cyanobacterial metabolism: The metabolic network of *Synechocystis* sp. PCC 6803. *PLOS Comput Biol* 9(6): e1003081. doi:10.1371/journal.pcbi.1003081

Kooistra WHCF, Gersonde R, Medlin LK, and Mann DG (2007) The origin and evolution of the diatoms: their adaptation to a planktonic existence. In *Evolution of Primary Producers in the Sea*, ed. PG Falkowski, AH Knoll, Academic Press, Inc.

Kramer DM, and Evans JR (2011) The importance of energy balance in improving photosynthetic productivity. *Plant Phys* 155: 70-78.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust A, Bowler C (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* 3(1): e1426. doi: 10.1371/journal.pone.0001426

Kurat CF, Wolinski H, Petschnigg J, Andrews B, Natter K, Kohlwein SD (2009) Cdk1/Cdc28-dependent activation of the major triglyceride lipase Tgl4 in yeast links lipolysis to cell-cycle progression. *Mol Cell* 33: 53-63.

Lebeau T, Robert JM (2003) Diatom cultivation and biotechnologically relevant products. Part II: Current and putative products. *Appl Microbiol Biot* 60(6):624-632.

Lee KH, Park JH, Kim TY, Kim HU, and Lee SY (2007) Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Sys Biol* 3:149.

Levitan O, Dinamarca J, Hochman G, and Falkowski PG (2014) Diatoms: a fossil fuel of the future. *Biotech* 32(3): 117-124.

Levitan O, Dinamarca J, Zelzion E, Lun DS, Guerra LT, Kim MK, Kim J, Van Mooy BAS, Bhattacharya D, Falkowski PG (2015) Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. *P Natl Acad Sci* 112(2):412-417.

Levitan O, Dinamarca J, Zelzion E, Gorbunov, M, Falkowski PG (2015) An RNAi knock-down of nitrate reductase enhances lipid biosynthesis in the diatom *Phaeodactylum tricornutum*. *Plant J* doi: 10.1111/tpj.13052

Liaud M, Lichtle C, Apt K, Martin W, and Cerff R (2000) Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion

protein: Evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. *Mol Biol Evol* 17(2): 213-223.

Lillo C (2006) Light regulation of nitrate reductase in green leaves of higher plants. *Physiol Plantarum* 90(3): 616-620. DOI: 10.1111/j.1399-3054.1994.tb08822.x

Liu, X Duan, S, Li A, Xu N, Cai Z, and Hu Z (2009) Effects of organic carbon sources on growth, photosynthesis, and respiration of *Phaeodactylum tricornutum*. *J Appl Phycol* 21: 239-246.

Lomas MW, Glibert PM (1999) Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. *Limnol Oceanogr* 44(3): 556-572.

Lomas, MW, Rumbley, CJ, and Gilbert, PM (2000) Ammonium release by nitrogen sufficient diatoms in response to rapid increases in irradiance. *J Plankton Res* 22(12): 2351-2366.

Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, and Church GM (2009) Large-scale identification of genetic design strategies using local search. *Molec Sys Biol* 5:296.

Lynn SG, Kilham SS, Kreeger DA, Interlandi SJ (2000) Effect of nutrient availability on the biochemical and elemental stoichiometry in the freshwater diatom *Stephanodiscus minutulus* (Bacillariophyceae). *J Phycol* 36: 510-522.

Mahadevan R, and Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5(4): 264-276.

Marsot P, Cembella AD, and Colombo JC (1991) Intracellular and extracellular amino acid pools of the marine diatom *Phaeodactylum tricornutum* (Bacillariophyceae) grown on unenriched seawater in high-cell-density dialysis culture. *J Phycol* 27: 478-491.

Martin DB, Vagelos PR (1962) The mechanism of tricarboxylic acid cycle regulation of fatty acid synthesis. *J Biol Chem* 237(6): 1787-1792.

Martin W, Brinkmann H, Savona C, and Cerff R (1993) Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *P Natl Acad Sci* USA 90: 8692–8696.

Martin W, and Müller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392(6671):37-41.

Matsuo M, Obokata J (2006) Remote control of photosynthetic genes by the mitochondrial respiratory chain. *Plant J* 47: 873-882.

Mehta SK, and Gaur JP (2005) Use of algae for removing heavy metal ions from wastewater: progress and prospects. *Crit Rev Biotechnol* 25(3): 113-52.

Michels AK, Wedel N, and Kroth PG (2005) Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. *Plant Phys* 137(3): 911-920.

Morgan DO (1999) Regulation of the APC and the exit from mitosis. *Nat Cell Biol* 1: E47-E53.

Muthuraj M, Palabhanvi B, Misra S, Kumar V, Sivalingavasu K, and Das D (2013) Flux balance analysis of *Chlorella* sp. FC2 IITG under photoautotrophic and heterotrophic growth conditions. *Photosynth Res* 118: 167-179.

Nelson DM, Treguer P, Brzezinski MA, Leynaert A, and Queguiner B(1995) Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem Cy* 9(3): 359-372.

Ninnemann, H (1987) Photoregulation of eukaryotic nitrate reductase. In *Blue Light Responses: Phenomena and Occurrence in Plants and Microorganisms* (Senger, H., editor), 17-29. CRC Press, Boca Raton, FL.

Nymark M, Valle KC, Brembu T, Hancke K, Winge P, et al. (2009) An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom *Phaeodactylum tricornutum*. *PLoS ONE* 4(11): e7743. doi:10.1371/journal.pone.0007743

Opute FI (1974) Studies on fat accumulation in *Nitzschia palea* Kutz. *Bot* 38: 889-902.

Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nature Biotech* 28(3): 245-248.

Pahl SL, Lewis DM, Chen F, King KD (2010) Growth dynamics and the proximate biochemical composition and fatty acid profile of the heterotrophically grown diatom *Cyclotella cryptica. J Appl Phycol* 22(2): 165-171.

Palsson B (2002) In silico biology through 'omics.' *Nat Biotechnol* 20: 649-650.

Parsons TR, Stephens K, Strickland JDH (1961) On the chemical composition of eleven species of marine phytoplankters. *J Fisheries Res Board Canada* 18: 1001-1016.

Pfleger CM, Kirschner MW (2000) The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev* 14(6):655-65.

Pharkya P, Burgard AP, and Maranas CD (2003) Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol Bioeng* 84: 887-899.

Pharkya P, Burgard AP, and Maranas CD (2004) Optstrain: a computational framework for redesign of microbial production systems. *Genome Res* 14:2367-2376.

Pharkya P, and Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng* 8(1): 1-13.

Poolman MG, Miguet L, Sweetlove LJ, and Fell DA (2009) A Genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Phys* 151(3): 1570-1581. http://dx.doi.org/10.1104/pp.109.141267

Prigent S, Collet G, Dittami SM, Delage L, Ethis de Corny F, Dameron O, Eveillard D, Thiele S, Cambefort J, Boyen C, Siegel A, and Tonon T (2014) The genome-scale metabolic network of *Ectocarpus siliculosus* (EctoGEM): a resource to study brown algal physiology and beyond *Plant J* 80(2): 367-81. doi: 10.1111/tpj.12627

Reed JL, Vo TD, Schilling CH, and Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biol* 4:R54.

Reinfelder JR, Kraepiel AM, and Morel FMM (2000) Unicellular C4 photosynthesis in a marine diatom. *Nature* 407: 996-999.

Rodriguez-Ruiz J, Belarbi EH, Sanchez JLG, and Alonso DL (1998) Rapid simultaneous lipid extraction and transesterification for fatty acid analyses. *Biotech Tech* 12 (9): 689-691.

Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, Scheer M, and Schomburg D (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nuc Acids Res* 41: D764-772.

Schuetz R, Kuepfer L, and Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Sys Biol* 3:119. DOI 10.1038/msb4100162

Sebolt, JS, and Weber, G (1983) Negative correlation of L-glutamine concentration with proliferation rate in rat hepatomas. *Life Sci* 34: 301-306.

Sessa G, Raz V, Savaldi S, Fluhr R (1996) PK12, a plant dual-specificity protein kinase of the LAMMER family, is regulated by the hormone ethylene. *Plant Cell* 8(12):2223-2234. doi:10.1105/tpc.8.12.2223.

Shastri AA, and Morgan JA (2005) Flux balance analysis of photoautotrophic metabolism. *Biotechnol Prog* 21(6): 1617-1626.

Sheehan J, Dunahay T, Benemann J, Roessler P (1998) A look back at the U.S. Department of Energy's Aquatic Species program: biodiesel from algae. National Renewable Energy Laboratory, Report NREL/TP-580-24190.

Shifrin NS, Chisholm SW (1981) Phytoplankton lipids: Interspecific differences and effects of nitrate, silicate and light-dark cycles. *J Phycol* 17: 374-384.

Shlomi T, Cabili MN, Herrgard MJ, Palsson B, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. *Nature Biotech* 26(9): 1003 - 1010.

Smith SR, Abbriano RM, and Hildebrand M (2012) Comparative analysis of diatom genomes reveals substantial difference in the organization of carbon partitioning pathways. *Algal Res* 1(1): 2-16. doi:10.1016/j.algal.2012.04.003

Sweetlove LJ, Beard KFM, Nunes-Nesi A, Fernie AR, and Ratcliffe RG (2010) Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci* 15: 462-470.

Thiele I, and Palsson B (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5(1): 93-121.

Thomas WH, Dodson AN, Reid FMH (1978) Diatom productivity compared to other algae in natural phytoplankton assemblages. *J Phycol* 14(3): 250-253.

Thornton BR, Toczyski DP (2006) Precise destruction: an emerging picture of the APC. *Genes Dev* 20: 3069 – 3078.

Trentacoste EM, Shrestha RP, Smith SR, Gle C, Hartmann AC, Hildebrand M, Gerwick WH (2013) Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *P Natl Acad Sci* 11(49): 19748-19753.

Udvardy A (1996) The role of controlled proteolysis in cell cycle regulation. *EJB Rev* 1996:195-201.

Valenzuela J, Mazurie A, Carlson RP, Gerlach R, Cooksey KE, Peyton BM, Fields MW (2012) Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. *Biotech Biofuels* 5: 40.

Vergara JJ, Berges JA, Falkowski PG (1998) Diel periodicity of nitrate reductase activity and protein levels in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae). *J Phycol* 34(6): 952-961.

Vodermaier HC (2004) APC/C and SCF: controlling each other and the cell cycle. *Curr Biol* 14: R787-R796.

Wang H, Fu R, and Guofeng P (2012) A study on lipid production of the mixotrophic microalgae *Phaeodactylum tricornutum* on various carbon sources. *Afr J Microbiol Res* 6(5): 1041-1047.

Weber APM, Linka M, and Bhattacharya D (2006) Ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Euk Cell* Mar: 609-612.

Whatley, JM, John, P, and Whatley, FR (1979) From extracellular to intracellular: the establishment of mitochondria and chloroplasts. *P R Soc Lond B Bio* 204: 165-187.

Wilken S, Schuurmans JM, and Matthijs HC (2014) Do mixotrophs grow as photoheterotrophs? Photophysiological acclimation of the chrysophyte *Ochromonas danica* after feeding. *New Phytol* 204(4): 882-1.doi: 10.1111/nph.12975. Epub 2014 Aug 19.

Wittpoth C, Kroth PG, Weyrauch K, Kowallik KV, and Strotmann H (1998) Functional characterization of isolated plastids from two marine diatoms. *Planta* 206: 79-85.

Wolosiuk RA and Buchanan BB (1977) Thioredoxin and glutathione regulate photosynthesis in chloroplasts. *Nature* 266: 565-567.

Yang C, Hua Q, Shimizu K (2002) Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*. *Appl Microbiol Biotechnol* 58: 813-822.

Yongmanitchai W, and Ward OP (1992) Separation of lipid classes from *Phaeodactylum tricornutum* using silica cartridges. *Phytochemistry* 31(10): 3405-3408.

Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR, and Apt KE (2001) Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* 292(5524): 2073-2075.

Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, … van Maris AJA (2008). Malic Acid Production by *Saccharomyces cerevisiae*: Engineering of Pyruvate Carboxylation, Oxaloacetate Reduction, and Malate Export. *Appl Environ Microb 74*(9), 2766–2777. doi:10.1128/AEM02591-07

Zheng Y, Quinn AH, and Sriram G (2013) Experimental evidence and isotopomer analysis of mixotrophic glucose metabolism in the marine diatom *Phaeodactylum tricornutum*. *Microb Cell Fac* 12:109.

Zhou BS, Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. *Nature* 408, 433-439. doi:10.1038/35044005