Computational and experimental models of diatom-bacteria interaction

Helena M. van Tol

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

E. Virginia Armbrust, Chair

Anitra Ingalls

Robert Morris

Program Authorized to Offer Degree:
Oceanography

University of Washington

Abstract

Computational and experimental models of diatom-bacteria interaction

Helena M. van Tol

Chair of the Supervisory Committee:
Professor E. Virginia Armbrust
School of Oceanography

Microbial interactions structure ecosystems and fuel biogeochemical cycling. The metabolic activities operating in the ocean are critical to the entire planet. In this work, I focused on interactions between diatoms and heterotrophic bacteria. Diatoms are a group of unicellular brown algae with frustules composed of silica. They form the base of coastal and polar marine food webs and contribute one fifth of global primary productivity. The inorganic nutrients fixed by oxygenic photosynthesis fuel secondary productivity by marine bacteria. Marine bacteria and diatoms have a range of different interaction strategies; many are still being elucidated.

In Chapter 1, I studied the antagonistic effects of a flavobacterium on diatom cell division. *Croceibacter atlanticus* inhibits cytokinesis in many species, causing the cells to elongate, become mutlinucleated, and filled with plastids.

In Chapter 2, I created a metabolic model of the diatom *Thalassiosira pseudonana* using the genome and physiological data from the literature. Simulations of diatom growth using Flux Balance Analysis revealed a role for nitrate and sulfate assimilation in dissipating reductants from the plastid. Changing redox and nutrient conditions causes the cell to secrete metabolites including organic carbon, nitrogen, and sulfur.

In Chapter 3, I created a metabolic model of the B_{12}-producing alphaproteobacterium *Ruegeria pomeroyi*. Previous work has demonstrated that *R. pomeroyi* will provide cobal-

amin to *T. pseudonana* in B_{12}-starvation conditions in exchange for organic sulfur and nitrogen. I constrained the metabolic models with transcriptomic data of *T. pseudonana* and *R. pomeroyi* in co-culture and simulated their interaction.

The distinct character of metabolites produced by diatoms likely fuels interactions with bacteria capable of utilizing those molecules. Bacteria influence diatom metabolism by interfering with the cell cycle, through nutrient-limitation, by altering redox conditions, and providing the cofactors required for growth. In this work, I have contributed to the literature exploring the complexity of diatom-bacteria interactions, where chemical or peptide cues, signals, and antagonists underlie the dynamics of microbial interactions. I have also created a framework for exploring more general metabolic exchanges between diatoms and bacteria. Genome-scale metabolic modeling of interactions between distinct marine microbial communities may be key to accurately predicting the character of dissolve organic matter in the ocean.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank Ginger for her faith in my abilities, and for her unwavering enthusiasm. Her ideas and optimism always manage to inspire me when I get stuck.

I would like to thank my committee members: Jody, Bob, Anitra, and Elhanan. Your insight and criticism have made me a better scientist. I know your comments on these last two chapters will help me get them published.

To Shady and Bryn, your ideas on diatom-bacteria interactions have inspired so much of my research.

I have had many wonderful office mates over the years. Thanks for the lunch dates and coffee breaks and thanks for being my friend. Thank you to everyone at the Center for Environmental Genomics (CEG) for your scientific rigour and friendly ways.

To all of my friends, new and old, in Canada and the US, thanks for your support. Thank you for reaching out when I am too busy with work. Thanks for making me laugh and relieving stress. Thanks for being excellent travel buddies, climbing partners, and karateka.

To my best Seattle friend, my office mate, and fellow grad student, Vega. Together we have learned so much about science and life over these last few years. I could not have finished this Ph.D. without you. I have missed you so much this year and I can't wait to be done so we can hang out again.

Thanks Mom. You know what a Ph.D. is, you know what to say and what not to say. You know how long it takes and how hard it is. Thanks for coming to watch my defense and for supporting my somewhat vague career ambitions. I wish I could thank my dad and tell him about my research. He was my first scientific mentor. He helped me with science fair

projects and brought me to his lab where we picked nematodes. Thanks for all the support from the rest of my family, especially Luke, Nick, Rob, and both Omas.

And thanks to my wonderful boyfriend! Tony your emotional and technical support got me through this. You fixed my computer, you answer my programming questions, and you let me cry on your shoulder. I love you.

DEDICATION

to Drs. Marianne and Monica Seger (Oma & Mom)

INTRODUCTION

Interactions between phototrophic and heterotrophic microbes form the base of the modern oceanic food web [1]. Microscopic algae capture light energy in the upper layers of the ocean and transform inorganic nutrients into biomass and and biomolecules. Bacteria in the surrounding waters recycle photosynthetically-derived organic matter by consuming dead phytoplankton or the metabolites leaked from living cells.

Diatoms are a particularly important group of marine phototrophs in nutrient-rich waters because their heavy silicified frustules can cause them to sink, transporting assimilated carbon, nutrients, and metals away from the atmosphere and the ocean surface [2]. Diatoms are a genetically and morphologically diverse group of phytoplankton, encompassing an estimated 100,000 different species [3]. They range in scale from micrometer to millimeter, and may be described as pennate or centric, chain-forming or unicellular [4]. Diatoms occupy a wide range of habitats across the globe: planktonic and benthic, marine and freshwater, polar and subpolar sea ice. There are few limits on diatom dispersal, many species have a global distribution [5, 6].

The evolutionary history of diatoms has been shaped by endosymbiotic transfer events, where genes of bacterial or algal origin were integrated into the ancestral diatom genome [7, 8]. The diatom mitochondrion contains genes originating from alphaproteobacteria while the plastid contains genes from cyanobacteria. The plastid genome was inherited through a secondary endosymbiotic event where a red algal cell was engulfed by the heterotrophic ancestor of diatoms. Additionally, there is evidence that a chlamydial symbiont contributed genetically to the progenitor of photosynthetic eukaryotes [9, 4]. The genome sequences of the pennate diatom *Phaeodactylum tricornutum* and the centric diatom *Thalassiosira pseudonana* suggest extensive ancient and more recent lateral gene transfer between prokaryotes

and photosynthetic heterokonts after the divergence with oomycetes. In the genome of *P. tricornutum*, 587 (or 7.5% of) gene models cluster with prokaryotic clades or form a sister group to those clades. Three hundred and twenty-three of those genes are shared with the more basal centric diatom *T. pseudonana*, while only 73 genes are shared with the more ancestral oomycete *Phytophthora* spp. Fifty-nine of those genes are also shared with *T. pseudonana*, suggesting that many of the prokaryotic genes in diatoms were transferred after the divergence with oomycetes. Many laterally transferred genes are part of novel functions and structures, including organic carbon and nitrogen utilization, the urea cycle, polyamine metabolism for cell was silicification, novel cell wall components, putative photoreceptors, and unorthodox mechanisms of DNA repair, replication, and recombination [10].

Diatoms account for about 20% of global primary productivity [11] and for up to 90% of regional productivity [12]. Phytoplankton release organic matter into the environment, stimulating the growth of heterotrophic bacterioplankton [13]. Epibiotic bacteria are known to attach to cell features that release organic matter. In the diatom genus *Pseudo-nitzschia*, diverse bacterial morphotypes can be found underneath the cingulae, along the raphe, and above the poroids of striae of cells using various attachment structures including pili, fimbria, holdfasts, mucous, or exopolymers [14].

Heterotrophic bacterioplankton that associate with diatoms typically belong to three different clades: *Alphaproteobacteria*, *Gammaproteobacteria*, and *Bacteroidetes* [15, 16]. These different groups of bacteria specialize in degrading distinct types of organic matter produced by phytoplankton, resulting in substrate-dependent succession over the course of a diatom bloom [17]. During a spring diatom bloom in the North Sea, *Bacteroidetes* abundance increased five-fold while the abundance of *Alpha-* and *Gammaproteobacteria* doubled in the first week. Proteomics revealed that *Alphaproteobacteria* specialize on the transport and utilization of low molecular weight organic matter, while *Bacteroidetes* and *Gammaproteobacteria* specialize in the degradation of high molecular weight carbohydrates [17]. *Bacteroidetes* are often found attached to particles and are some of the first to profit from algal lysis, while *Gammaproteobacteria* dominate the end of the bloom and contributed to algal decay. Al-

though heterotrophic bacteria can occupy different niches provided by the distinct types of organic matter produced by diatoms, there is steep competition for phosphate between diatoms and the different groups of bacteria which have distinct strategies for acquiring phosphate. Phosphate limitation likely contributed to bloom termination as well as increased exudation of organic carbon by diatoms [17].

Some of the first studies on diatom-bacteria interactions focused on the apparent paradox that diatoms would hasten their own demise during phosphate limiting conditions [18] due to increased production of polysaccharides that contributed to bacterial growth [19]. The nature of secreted metabolites can also be impacted by the metabolic properties of different bacteria. When *Phaeodactylum tricornutum* and *Donghicola* sp. were grown on monomethylamine as the sole source of nitrogen, *Donghicola* sp. secreted NH_4^+ that allowed the diatom to grow and secrete organic carbon for the bacteria – an interaction that could be characterized as mutualistic. In contrast, *Methylophaga* sp. uses monomethylamine as a source of both carbon and nitrogen, and *P. tricornutum* received NH_4^+ for free in an interaction that could be characterized as commensal [20]. More recent work [21, 22, 23, 24] has leveraged transcriptomics and metabolomics to discover the variety of metabolites exchanged between mutualistic diatoms and bacteria. There is increasing recognition that the exchange of vitamins and cofactors may be a key benefit for phytoplankton. The picoeukaryotic alga *Ostreococcus tauri* can grow in co-culture with the alphaproteobacterium *Dinoroseobacter shibae*; *O. tauri* receives cobalamin and thiamine from the bacterium and *D. shibae* receives niacin, biotin, and *p*-aminobenzoic acid from the alga [25]. In my work I have explored the question: 'How do heterotrophic bacteria impact diatom physiology?'

In Chapter 1, I uncovered a novel type of interaction between the flavobacterium *Croceibacter atlanticus* and diatoms. *C. atlanticus* is a globally-distributed species that has been isolated from multiple diatom cultures. The bacterium inhibits diatom cell division, while still allowing the proliferation of plastids and nuclei. This work has been published in the ISME Journal in 2016 [26]. Work by our collaborators has shown that *C. atlanticus* secretes a peptide inhibitor that causes multiple different diatom species to become multi-nucleated

and filled with plastids. These changes are likely to have a profound impact on diatom metabolism and sinking rates.

In Chapter 2, I constructed a genome-scale metabolic model of the diatom *Thalassiosira pseudonana* CCMP 1335. The model was constrained with realistic nutrient uptake rates, rates of O_2 production, and respiration. I simulated *T. pseudonana* growth under a range of different nutrient conditions and predicted the excretion of different metabolites. I found that nitrate and sulfate uptake play an important role in dissipating reductants from the plastid. The balance of nutrient uptake and changing redox conditions results in the secretion of organic nitrogen and sulfur. In this way diatom physiology can shape interactions with bacteria.

In Chapter 3, I studied the mechanism of metabolic interaction between *T. pseudonana* and the roseobacterium *Ruegeria pomeroyi* DSS-3. *R. pomeroyi* provides vitamin B_{12} to the diatom in cobalamin-starvation conditions in exchange for organic nitrogen and sulfur. I constructed a genome-scale model of the roseobacterium *Ruegeria pomeroyi* DSS-3. Transcriptomics data of *R. pomeroyi* and *T. pseudonana* growth in co-culture was re-analyzed in the context of these metabolic networks, and I found that *T. pseudonana* showed signs of vitamin B_{12} limitation. A simulation of interaction between *R. pomeroyi* and *T. pseudonana* using genome-scale metabolic models constrained by transcriptomics data indicated that aspects of the interaction can be attributed to the physiological effects of B_{12} limitation.

References

[1] A Buchan et al. "Master recyclers: features and functions of bacteria associated with phytoplankton blooms." In: *Nature reviews. Microbiology* 12.10 (Oct. 2014), pp. 686–98. ISSN: 1740-1534. DOI: 10.1038/nrmicro3326.

[2] ZV Finkel et al. "Phytoplankton in a changing world: cell size and elemental stoichiometry." In: *Journal of Plankton Research* 32.1 (Jan. 2010), pp. 119–37. ISSN: 0142-7873. DOI: 10.1093/plankt/fbp098.

[3] DG Mann and P Vanormelingen. "An inordinate fondness? the number, distributions, and origins of diatom species." In: *Journal of Eukaryotic Microbiology* 60.4 (May 2013), pp. 414–20. ISSN: 10665234. DOI: 10.1111/jeu.12047.

[4] EV Armbrust. "The life of diatoms in the world's oceans." In: *Nature* 459.7244 (May 2009), pp. 185–92. ISSN: 0028-0836. DOI: 10.1038/nature08057.

[5] WHCF Kooistra et al. "Global diversity and biogeography of *Skeletonema* species (Bacillariophyta)." In: *Protist* 159.2 (Apr. 2008), pp. 177–93. ISSN: 14344610. DOI: 10.1016/j.protis.2007.09.004.

[6] VL Trainer et al. "*Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health". In: *Harmful Algae* 14 (Feb. 2012), pp. 271–300. ISSN: 15689883. DOI: 10.1016/j.hal.2011.10.025.

[7] PG Falkowski. "The evolution of modern eukaryotic phytoplankton." In: *Science* 305.5682 (July 2004), pp. 354–60. ISSN: 0036-8075. DOI: 10.1126/science.1095964.

[8] EV Armbrust et al. "The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism." In: *Science* 306.5693 (Oct. 2004), pp. 79–86. ISSN: 0036-8075. DOI: 10.1126/science.1101156.

[9] B Becker, K Hoef-Emden, and M Melkonian. "Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes." In: *BMC Evolutionary Biology* 8.203 (July 2008). ISSN: 1471-2148. DOI: 10.1186/1471-2148-8-203.

[10] C Bowler et al. "The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes." In: *Nature* 456.7219 (Nov. 2008), pp. 239–44. ISSN: 0028-0836. DOI: 10.1038/nature07410.

[11] CB Field et al. "Primary production of the biosphere: integrating terrestrial and oceanic components". In: *Science* 281.5374 (July 1998), pp. 237–40. ISSN: 1095-9203. DOI: 10.1126/science.281.5374.237.

[12] DM. Nelson et al. "Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation." In: *Global Biogeochemical Cycles* 9.3 (Sept. 1995), pp. 359–72. ISSN: 08866236. DOI: 10.1029/95GB01070.

[13] LM Jensen. "Phytoplankton release of extracellular organic carbon, molecular weight composition, and bacterial assimilation". In: *Marine Ecology Progress Series* 11 (Feb. 1983), pp. 39–48. ISSN: 0171-8630. DOI: 10.3354/meps011039.

[14] I Kaczmarska et al. "Diversity and distribution of epibiotic bacteria on *Pseudo-nitzschia multiseries* (Bacillariophyceae) in culture, and comparison with those on diatoms in native seawater." In: *Harmful Algae* 4.4 (June 2005), pp. 725–41. ISSN: 15689883. DOI: 10.1016/j.hal.2004.10.001.

[15] SA Amin, MS Parker, and EV Armbrust. "Interactions between diatoms and bacteria." In: *Microbiology and Molecular Biology Reviews* 76.3 (Sept. 2012), pp. 667–84. ISSN: 1092-2172. DOI: 10.1128/MMBR.00007-12.

[16] G Behringer et al. "Bacterial communities of diatoms display strong conservation across strains and time." In: *Frontiers in Microbiology* 9 (Apr. 2018), pp. 1–15. ISSN: 1664302X. DOI: 10.3389/fmicb.2018.00659.

[17] H Teeling et al. "Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom." In: *Science* 336.6081 (May 2012), pp. 608–11. ISSN: 0036-8075. DOI: 10.1126/science.1218344.

[18] G Bratbak and TF Thingstad. "Phytoplankton-bacteria interactions: an apparent paradox? Analysis of a model system with both competition and commensalism." In: *Marine Ecology Progress Series* 25 (Aug. 1985), pp. 23–30.

[19] F Guerrini et al. "Bacterial-algal interactions in polysaccharide production". In: *Aquatic Microbial Ecology* 15 (Aug. 1998), pp. 247–53.

[20] M Suleiman et al. "Interkingdom cross-feeding of ammonium from marine methylamine-degrading bacteria to the diatom *Phaeodactylum tricornutum*". In: *Applied and Environmental Microbiology* 82.24 (Dec. 2016), pp. 7113–22. ISSN: 0099-2240. DOI: 10.1128/AEM.01642-16.

[21] SA Amin et al. "Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria." In: *Nature* 522.7554 (May 2015), pp. 98–101. ISSN: 0028-0836. DOI: 10.1038/nature14488.

[22] BP Durham et al. "Cryptic carbon and sulfur cycling between surface ocean plankton." In: *Proceedings of the National Academy of Sciences* 112.2 (Jan. 2015), pp. 453–7. ISSN: 0027-8424. DOI: 10.1073/pnas.1413137112.

[23] BP Durham et al. "Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system." In: *Environmental Microbiology* 19.9 (Sept. 2017), pp. 3500–13. ISSN: 14622912. DOI: 10.1111/1462-2920.13834.

[24] M Landa et al. "Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom." In: *The ISME Journal* 11 (July 2017), pp. 2677–90. ISSN: 1751-7362. DOI: 10.1038/ismej.2017.117.

[25] MB Cooper et al. "Cross-exchange of B-vitamins underpins a mutualistic interaction between *Ostreococcus tauri* and *Dinoroseobacter shibae*." In: *The ISME Journal* 13 (Sept. 2018), pp. 334–45. ISSN: 17517370. DOI: 10.1038/s41396-018-0274-y.

[26] HM Van Tol, SA Amin, and EV Armbrust. "Ubiquitous marine bacterium inhibits diatom cell division". In: *The ISME Journal* 11 (Sept. 2016), pp. 31–42. ISSN: 1751-7362. DOI: 10.1038/ismej.2016.112.

Chapter 1

UBIQUITOUS MARINE BACTERIUM INHIBITS DIATOM CELL DIVISION

Van Tol, H.M., Amin, S.A., and Armbrust, E.V. (2017). Ubiquitous marine bacterium inhibits diatom cell division. *The ISME Journal*. 11: 31-42. doi:10.1038/ismej.2016.112.

1.1 Abstract

Intricate relationships between microorganisms structure the exchange of molecules between taxa, driving their physiology and evolution. On a global scale, this molecular trade is an integral component of biogeochemical cycling. As important microorganisms in the world's oceans, diatoms and bacteria have a large impact on marine biogeochemistry. Here, I describe antagonistic effects of the globally distributed flavobacterium *Croceibacter atlanticus* on a phylogenetically diverse group of diatoms. I used the model diatom *Thalassiosira pseudonana* to study the antagonistic impact in more detail. In co-culture, *C. atlanticus* attaches to *T. pseudonana* and inhibits cell division, inducing diatom cells to become larger and increase in chlorophyll *a* fluorescence. These changes could be explained by an absence of cytokinesis that causes individual *T. pseudonana* cells to elongate, accumulate more plastids and become polyploid. These morphological changes could benefit *C. atlanticus* by augmenting the colonizable surface area of the diatom, its photosynthetic capabilities and possibly its metabolic secretions.

1.2 Introduction

The ocean is a dilute environment with a microbial density of $\sim 10^6$ bacteria per milliliter of seawater [1] and 10–1000 times fewer phytoplankton. To survive low nutrient concentrations,

marine microbes commonly form close associations with one another to facilitate metabolic and nutrient exchanges. Consequently, interacting networks of phytoplankton and bacteria recycle most fresh organic carbon in the sunlit layers of the upper ocean, while a relatively small fraction of photosynthetic products remains dissolved in the water column or sinks to the sea floor [2, 3].

Marine microbes experience a complex and heterogeneous chemical milieu that is unlike the macro-environment described by bulk measurements [4]. A thin boundary layer of fluid termed 'the phycosphere' surrounds algal cells [5]. In the phycosphere, molecular gradients are governed by diffusion [6] and altered by the activities of bacteria that accumulate through chemotaxis [7, 8] and reproduction. As the thickness of diffusive boundary layers scales with surface area [9], larger phytoplankton may harbor more bacteria than smaller phytoplankton. Bacteria modify the phycosphere by excreting metabolites, consuming metabolites produced by the algae or other bacteria, consuming and regenerating nutrients or making nutrients more bioavailable. As a result, bacteria turn the phycosphere into a dynamic micro-niche by shaping the environmental signals perceived by algae. Over time, co-evolution between algae and bacteria is likely to have resulted in interactions between these two groups.

Diatoms are an ecologically important group of phytoplankton that have a major role in supporting fisheries and driving the marine carbon cycle [10, 11]. Specific bacterial phyla, mainly the Proteobacteria and Bacteroidetes [12], are commonly associated with diatoms. Beneficial and detrimental bacteria can be isolated from phytoplankton cultures, but their roles in the ocean and their mechanisms of antagonistic activity remain under-studied [13, 12]. Beneficial bacteria typically provide a useful service to the diatom in exchange for organic nutrients; for example, *Ruegeria pomeroyi* provides vitamin B_{12} to *Thalassiosira pseudonana* in exchange for the organic sulfur compound 2,3-dihydroxypropane-1-sulfonate [14]. Many detrimental bacteria are from the phylum Bacteroidetes or the class Gamma-proteobacteria [13] and are thought to harm phytoplankton by secreting proteases or other algaecides [15, 16, 17].

Previously, bacteria belonging to the *Sulfitobacter*, *Hyphomonas*, *Limnobacter*, *Mari-*

nobacter and *Croceibacter* genera were isolated and identified from several cultures of the coastal diatom *Pseudo-nitzschia multiseries* originating from the Atlantic and Pacific Oceans. To probe interactions between the diatom and these bacteria, *P. multiseries* cultures were made axenic and a series of co-culture experiments were performed with the different bacterial isolates. Some isolates increased the growth, whereas others inhibited the growth of *P. multiseries* [18]. Here I report on the activities of *Croceibacter atlanticus*, an antagonistic flavobacterium from the *P. multiseries* consortium, which impacts diatom DNA content, cell division and cell morphology. This type of antagonistic effect has not, to my knowledge, been previously reported in the marine environment and could have important implications for understanding bloom dynamics and organic matter fluxes in the ocean.

1.3 Materials and methods

1.3.1 Culture acquisition and maintenance

Nitzschia sp. RCC 80 was acquired from the Roscoff Culture Collection (Roscoff, France); *Thalassiosira weissfloggii* CCMP 1052, *T. oceanica* CCMP 1003 and *T. pseudonana* CCMP 1335 from the National Center for Marine Algae and Microbiota (NCMA, East Boothbay, ME, USA); *Pseudo-nitzschia multiseries* CLNN-17 and *P. multiseries* CLN-47 from Stephen Bates and Claude Léger of Fisheries and Oceans Canada (Moncton, NB, Canada). *P. multiseries* CLN-47 is the product of a cross performed *ca* 2001 between environmental strains CL-143 and CL-147 collected from the Bay of Fundy, Canada; *P. multiseries* CLNN-17 is the product of a cross between CLN-47's siblings CLN-35 and CLN-48 performed at the end of 2006. *P. multiseries* PC9 and GGA2, and *P. fraudulenta* OC1 were isolated from Penn Cove WA, in 2010, from Golden Gardens WA in 2010, and from Newport Beach CA in 2011 by Michael Carlson, and *P. multiseries* IOES-1 from East Sound WA in 2010 by Irina Oleinikov. *Pseudo-nitzschia* species were identified using PCR primers to amplify and sequence a polymorphic region of the internal transcribed spacer 1 [19].

All strains were maintained as semicontinuous batch cultures in f/2 medium [20] at 13°C

in a 16-h light/8-h dark diurnal cycle ($\sim 120\ \mu$E m^{-2} s^{-1}) except RCC 80 and CCMP 1003, which were maintained at 20°C in 24-h light conditions ($\sim 120\ \mu$E m^{-2} s^{-1}). For experimental work, *Pseudo-nitzschia* strains and CCMP 1335 were acclimated to the synthetic seawater medium Aquil [21]. To monitor diatom division rates, *in vivo* chlorophyll *a* fluorescence was measured daily with a 10-AU fluorometer (Turner Designs, Sunnyvale, CA, USA) $1-2\,h$ before the end of the dark period.

The bacteria *Marinobacter* sp. SA14 and *C. atlanticus* SA60 were isolated from *P. multiseries* CLNN-17 in 2010; *C. atlanticus* HV2 was isolated from *P. multiseries* GGA2 in 2012. These strains were identified by their 16S ribosomal RNA gene sequence [18] (GenBank accession no.: KM033246, KM033280, KM033233). I acquired the *C. atlanticus* type strain HTCC 2559 from the American Type Culture Collection (ATCC, Manassas, VA, USA). All bacterial isolates were stored in 15% glycerol stocks at −80°C.

Pseudo-nitzschia strains IOES-1, PC9, GGA2 and OC1 were 'cured' of associated bacteria by adapting a protocol from ref. [22], treating cultures with the detergent Triton X-100 (20 μg ml^{-1}) and an antibiotic cocktail (per ml: 50 μg streptomycin, 67 μg gentamycin, 20 μg ciprofloxacin, 2.2 μg chloramphenicol and 100 μg ampicillin) as described in ref. [18]. CCMP 1335 is maintained bacteria-free by the National Center for Marine Algae and Microbiota (NCMA). The absence of bacteria in these cultures was verified periodically by examining SYBR green-stained cultures with epifluorescence microscopy and by checking for bacterial growth in marine broth (MB; per l: 5 g peptone, 0.5 g yeast extract and 750 ml 0.2 μm filtered seawater; [23]).

1.3.2 Co-culture experiments

To assess the range of possible hosts for *C. atlanticus*, I performed a series of co-culturing experiments between different diatoms and the three *C. atlanticus* strains. Before each experiment, bacterial glycerol stocks were plated at room temperature on marine agar (MB with 1.5% w/v agar). Single colonies were inoculated into MB and grown overnight with shaking at 30°C. Triplicate co-cultures were initiated by inoculating an exponentially growing

diatom culture (initial relative chlorophyll *a* fluorescence units ~ 0.5) with $\sim 4\times10^5$ bacteria ml^{-1} (initial bacterial concentration). This concentration of *C. atlanticus* was chosen because it is within the range of concentrations certain flavobacteria reach in the euphotic zone during a diatom bloom [24]. Control cultures were treated in each experiment with 0.01–0.20% (v/v) MB to match the initial concentrations of growth media added with bacteria.

More comprehensive experiments were conducted with *T. pseudonana* CCMP 1335 and *C. atlanticus* SA60 with the goal of honing in on the mechanism of antagonistic activity. *T. pseudonana* cells were grown exponentially in the organic carbon-free artificial medium Aquil [21] to $\sim 10^5$ cells ml^{-1} and diluted to an initial concentration of 2500 cells ml^{-1} for the experiment. Overnight cultures of *C. atlanticus* were grown in MB to a density of $\sim 10^9$ cells ml^{-1} and either added directly to the diluted *T. pseudonana* cultures or added after washing 2× with either Aquil, fresh MB, or the spent MB from an overnight grown culture after the bacteria were removed by centrifugation and the media filtered through a 0.2 μm filter. Washing consisted of centrifuging 1 ml cell cultures at 3220 $\times g$ and 10°C for 5 min and replacing the supernatant with 1 ml media. These experiments were designed to determine whether *C. atlanticus* secretes a growth inhibitor into MB in the absence of diatoms, and whether *C. atlanticus* can grow on diatom-produced metabolites. As *C. atlanticus* is non-motile except for surface gliding [25, 26], the initial concentration of *C. atlanticus* added to each co-culture was increased to $\sim 2\times10^6$ cells ml^{-1} for an initial ratio of about 800 bacteria per diatom cell, enhancing the encounter rate with diatoms and increasing the likelihood of early attachment. For diatom control treatments, 0.2% v/v of fresh MB, fresh Aquil media, or 0.2% v/v of the sterile-filtered spent overnight MB media was added to cultures of *T. pseudonana* growing in Aquil.

Samples $(1-2ml)$ for flow cytometry were collected each day 0.5–1.5 h after the start of the light period and preserved in 1% buffered paraformaldehyde and 0.05% glutaraldehyde, incubated 20–30 min, flash frozen in liquid nitrogen and stored at −80°C [27]. Samples (1–2 ml) for epifluorescence microscopy were taken from a single replicate of each sample and were preserved in 8–9ml Aquil with 2% glutaraldehyde and stored in the dark at 4°C.

Specific growth rates (μ, day^{-1}) were estimated from *in vivo* chlorophyll *a* fluorescence (relative fluorescence units) or cell counts (flow cytometry) measured during the exponential growth phase. Specific growth rate is defined as the natural log cell number or fluorescence at time t_2 over the cell number or fluorescence at time t_1 divided by the difference between t_2 and t_1. Percent growth inhibition for the co-cultures was defined as the difference between $\mu_{control}$ and $\mu_{co-culture}$ divided by $\mu_{control}$ multiplied by 100. Final cell concentrations (C, cells ml^{-1}) were defined as the maximum value for each replicate growth curve measured by flow cytometry cell counts. Means were calculated from triplicate cultures unless otherwise indicated.

In a separate experiment, the hypothesis that *C. atlanticus* requires amino acids as a supplementary nutrient in co-culture was tested. *T. pseudonana* was grown exponentially in Aquil to $\sim 3 \times 10^5$ cells ml^{-1} and diluted to an initial concentration of ∼2000 cells ml^{-1}. Overnight cultures of *C. atlanticus* were grown in MB to a density of $\sim 10^9$ cells ml^{-1}, washed 2× with Aquil and added to *T. pseudonana* at an initial concentration of $\sim 2 \times 10^6$ cells for an initial ratio of 1000 bacteria per diatom cell. Each co-culture was amended with MEM Amino Acids solution (1:200 dilution, product no.: M5550, Sigma-Aldrich Inc., St Louis, MO, USA), MEM non-essential amino-acid solution (1:400 dilution, product no.: M7145, Sigma-Aldrich Inc.), a mix of both solutions, or nothing (duplicate samples each). After a 5-day incubation period, flow cytometry samples were collected as described previously.

1.3.3 Epifluorescence microscopy

Microscopy was used to provide phenotypic evidence of the interaction and to interpret flow cytometry results. Polyvinyl alcohol basal solution (Mowiol 4-88, Sigma-Aldrich Inc.) was prepared as described by ref. [28] and stored as 200 μl aliquots in amber microcentrifuge tubes at −20°C. Fresh staining solution was prepared by adding 2 μl of freshly prepared 1 M ascorbic acid and 4 μl SYBR Green I (Invitrogen, Waltham MA, USA) to 200 μl basal solution. Samples preserved in 2% glutaraldehyde were gently filtered onto black polycarbonate membrane filters (0.2 μm pore size, 25 mm diameter, Poretics, Osmonics Inc.,

Minnetonka, MN, USA). The filters were air dried on glass slides cleaned with 70% ethanol and then attached to the slide with 5 μl of staining solution. An additional 14 μl of staining solution was pipetted onto a cover slip that was flipped onto the filter. Air bubbles were squeezed out and the cover slip was sealed in place with clear nail polish. The samples were examined by epifluorescence microscopy with a 470/40 nm blue excitation 515 nm long-pass emission filter set ($\times$100 magnification, Eclipse 80i, Nikon, Tokyo, Japan; 11001v2, Chroma Corp., Rockingham, VT, USA) on the day of slide preparation and sections of the slide were imaged (MicroPublisher 3.3 RTV, QImaging, Surrey, British Columbia, Canada). Slides were stored in the dark at -20°C. All microscopy images can be accessed from the figshare online digital repository (doi:10.6084/m9.figshare.3471488).

ImageJ (Rasband, 2016), image analysis software, was used to measure diatom length and diameter (or just diameter depending on the cell orientation), plastid surface area, and to count nuclei from control and co-culture treatments at $t = 2$ days and $t = 6$ days. Each image was split into three channels Red Green Blue (RGB) and measurements were made on cells that were in focus. The wand tool was used in the red channel to measure the surface area covered by plastids for each diatom cell. The number of nuclei per cell was counted in the green channel. Length and diameter measurements (or minimum and maximum lengths) were also taken with the straight-line selection tool in the green channel, using nonspecific SYBR Green staining or attached bacteria as a guideline for the cell boundaries.

1.3.4 Flow cytometry

Flow cytometry was used both to count *T. pseudonana* and unattached *C. atlanticus cells* as well as to compare the light scatter and emission properties of the *T. pseudonana* population in co-culture and mono-culture. Flow cytometry measurements were made on an influx flow cytometer (BD/Cytopeia Inc., Seattle, WA, USA) equipped with two 10% neutral density filters and a 488 nm laser; data collection was triggered by forward light scatter. Samples were stained with SYBR Green I (1:10 000 dilution) and incubated in the dark on ice for at least 20 min. To minimize coincidence and improve population resolution, flow rates and

sample concentrations were adjusted to achieve event rates below 1000 s^{-1}. Two-micrometer beads were added to each sample as an internal standard (PolySciences Inc., Warring- ton, PA, USA). Samples were run for 5 min before data collection to allow the stain to equilibrate. Samples were run until at least 200 *T. pseudonana* cells were counted. Forward light scatter, side scatter and emission at 692 nm (40 nm band pass) (chlorophyll *a* fluorescence) were measured on a log scale. A flow meter was used to measure the volume sampled.

Flow cytometry standard files were processed with the flowCore and flowStats libraries in *R* [29, 30]. Each population of cells was semiautomatically gated with the lymphGate function from the flowStats library after a rectangular pre-selection. *T. pseudonana* was gated in the forward scatter and 692/40 nm emission dimensions, whereas *C. atlanticus* and the 2 μm beads were gated in the forward scatter and 530/40 nm emission dimensions. These gating parameters were chosen to effectively separate the population of chlorophyll *a*-containing particles from the unattached bacteria. The accuracy of gated populations was visualized with scatter plots from the flowViz library [31]. *T. pseudonana* distributions were normalized to the 2 μm beads before analysis. Beads were inadvertently not added to a *T. pseudonana* (t = 3 days) replicate and a *T. pseudonana*+fresh MB (t = 5 days) replicate before data collection; these samples were therefore removed from further analysis. Flow cytometry data can be accessed from the figshare online digital repository (doi:10.6084/m9.figshare.3471455).

1.3.5 Statistical analyses

R was used to plot *in vivo* chlorophyll *a* fluorescence and flow cytometry data, and the *R* stats package (https://stat.ethz.ch/R-manual/R-devel/library/stats/DESCRIPTION) was used to perform all statistical analyses (v. 3.22.6). Two sample t-tests with a significance cut-off value of 0.05 were used to compare growth rates and final cell concentrations (μ, C) of different treatments as well as microscopy measurements made with ImageJ. A chi-squared test with a significance cutoff value of 0.05 was used to compare the distributions of nuclei counted in different treatments.

When discussing the co-culture treatments, *T. pseudonana* cells with attached bacteria is

sometimes refer to as 'colonized cells'. The flow cytometry data were not normally distributed according to Shapiro-Wilk test results. Q-Q plots showed that control populations tended to have a slightly positive skew while colonized populations skewed increasingly negative over the course of the experiment, making data transformations intractable. I therefore used one-tail Wilcoxon rank sum tests with a significance cutoff of 0.05 to test for significant differences between the forward scatter, side scatter, and red fluorescence per cell of the *T. pseudonana* populations over time. Given that nonparametric tests with large sample sizes can be sensitive to small statistical differences with little biological meaning, I designed the test to detect a significant shift for a given time point if the colonized *T. pseudonana* population mean was greater than the control population mean by as much as the upper quartile of all samples from the control population, a relatively consistent metric over time. Replicate samples were merged to compare control and colonized treatments over time.

1.3.6 C. atlanticus genome

The sequenced genomes of *C. atlanticus* was used to look for evidence of metabolic deficiencies in the bacterium that could be supplemented by metabolites from diatoms. *C. atlanticus* strain Pm1 was sequenced by the Joint Genome Institute as part of the *P. multiseries* genome project. The *C. atlanticus* strain Pm1 genome corresponds to the first scaffold in the prokaryotic assembly file (http://genome.jgi-psf.org/Psemu1/download/Psemu1_prokaryotic_scaffolds.fasta.gz). The *C. atlanticus* type strain genome sequence (HTCC 2559) is publically available from GenBank (accession number CP002046). Strain SA60 was sequenced at the Center for Environmental Genomics (Seattle, WA, USA). DNA from SA60 was extracted using a DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA), sheared into 1–2-kb fragments (Hydroshear, Applied Biosystems, Waltham, MA, USA), and a 50 bp×50 bp mate-paired library was generated (Genomic Solutions Ltd, Cambridgeshire, UK). Emulsion PCR was used to attach the library to beads (Life Technologies, Research and Development Unit, Carlsbad, CA, USA); approximately 56 million beads were deposited and sequenced with a SOLiD 4 Next Generation Sequencer (Life Technologies). Sequencing generated 98,959,690 reads (or

49,479,845 mate pairs), which resulted in 74,025,907 reads (12,275,915 singlets and 30,874,996 mate pairs) after quality control. The short-read de Bruijn graph assembler Velvet was used for *de novo* contig assembly [32] and five scaffolds were generated with SEAStAR graph_ops [33]. Scaffold order was determined by alignment to the HTCC 2559 genome.

All three genomes were submitted to the Integrated Microbial Genomes Expert Review (IMG ER) gene prediction and annotation pipeline [34]. The genome sequences and their annotations are publically available through IMG (http://img.jgi.doe.gov submissions: 15010, 11665, 11666). LASTZ was used to align the Pm1 and SA60 genomes to the type strain [35]; aligned segments were plotted [36] in *R* (v. 3.22.6). Overall genome percent identity was calculated from a bl2seq alignment as the number of matches $\times 100$ divided by the alignment length between each genome and HTCC 2559.

The BioCyc organism-specific Pathway/Genome Databases (PGDBs; ref. [37] were used to compare the amino-acid biosynthesis, vitamin biosynthesis and inorganic nutrient utilization pathways of *T. pseudonana* CCMP 1335 and *C. atlanticus* HTCC 2559. The HTCC 2559 PGDB is not curated and the *T. pseudonana* PGDB is missing some genes, likely due to lack of homology. For this reason, the results were verified in the SA60 genome and partial pathways were sometimes considered to be evidence that the whole pathway is present in *T. pseudonana*.

1.4 Results and discussion

1.4.1 C. atlanticus is commonly associated with P. multiseries in culture

Three strains of *C. atlanticus* from three different *P. multiseries* cultures were identified based on 16S ribosomal DNA sequence identity with the *C. atlanticus* type strain HTCC 2559 isolated from the Sargasso Sea [25]. The *P. multiseries* isolates were collected from different locales and maintained in culture for varying lengths of time. *C. atlanticus* HV2 was isolated after 2 years in culture with *P. multiseries* GGA2. *C. atlanticus* strain SA60 was isolated from *P. multiseries* strain CLNN-17, although the bacterial consortium was likely derived

from the original parental strains (CL-143, CL-147), both isolated from the Bay of Fundy ∼10 years prior. The genome sequence of *C. atlanticus* strain Pm1 was identified in sequences derived from the *P. multiseries* strain CLN-47 genome project. CLN-47 is an F1 generation cross produced *ca* 2001 from parental strains CL-143 and CL-147 and sequenced after 5 years in culture (Supplementary Table 1.1); CLN-47 has since been lost and no available isolate is associated with the *C. atlanticus* Pm1 sequence. The detection of *C. atlanticus* in several diatom cultures with different origins and culturing time in the laboratory indicates that it is a common, long-term member of the bacterial consortia associated with *P. multiseries* in culture.

1.4.2 Antagonistic impact of C. atlanticus on diatoms

When co-cultured with *P. multiseries* IOES-1, *C. atlanticus* strains SA60, HV2 and HTCC 2559 all displayed a similarly negative impact on the growth of the diatom. Within hours of inoculating *P. multiseries* with the different strains of *C. atlanticus*, *in vivo* chlorophyll *a* fluorescence of the co-culture began to decline, leveling off within about 6 days (Figure 1.1). Similar results were obtained with the closely related diatom *P. fraudulenta* (Table 1.1). Epifluorescence microscopy revealed that *C. atlanticus* strain HV2 proliferated on the cell surface of both *P. multiseries* and *P. fraudulenta* (Supplementary Figure A.1). These results differ significantly from co-culturing with other bacteria from the consortium. For example, in previous work *Sulfitobacter* sp. SA11 enhanced the growth of *P. multiseries* [18], whereas *Marinobacter* sp. SA14 did not influence diatom growth (Figure 1.1). The antagonistic effect of *C. atlanticus* was unexpected given the observed long-term co-habitation of this bacterium with various *P. multiseries* strains in culture.

All three *C. atlanticus* strains negatively impacted a variety of diatoms to varying degrees of severity (Table 1.1). Growth inhibition by *C. atlanticus* was not restricted to diatoms nor did it correlate with diatom phylogeny. Strong inhibition was observed with the closely related *P. fraudulenta* but not with *Nitzschia* sp. RCC 80; moderate inhibition was observed with two distantly related diatoms *T. pseudonana* and *T. oceanica* and no inhibition was

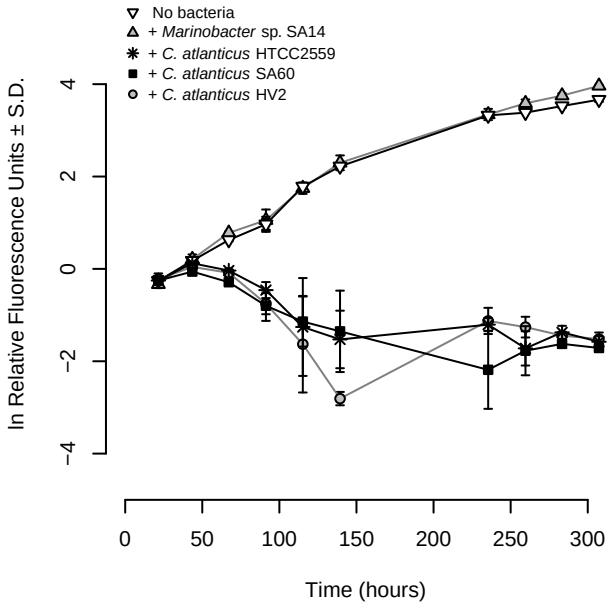


Figure 1.1: Effect of three different *C. atlanticus* isolates on the growth rates of *P. multiseries* IOES-1. The growth rates of *P. multiseries* are indistinguishable from each other when 0.05% v/v of either fresh MB media ($n = 2$) or an overnight grown culture of *Marinobacter* sp. (strain SA14) at $\sim 4 \times 10^5$ bacteria ml^{-1} ($n = 3$) is added to diatom cultures at $t = 0$ h. Growth rates and cell yield are significantly inhibited when overnight grown *C. atlanticus* isolates ($n = 3$ each; Table 1.1, Bonferonni-adjusted $p \leq 0.05$) are added to the diatom culture at a final concentration of $\sim 4 \times 10^5$ bacteria ml^{-1}. Error bars indicate ± 1 s.d. or the range (for $n = 3$ or 2, respectively) and are shown when larger than symbol size.

Heterokont ID	Axenic	Croceibacter strains	Marine broth concentration (% v/v)	$\mu^{a}_{control}$±s.d. (day^{-1})	$\mu_{co-culture}$±s.d. (day^{-1})	% Growth inhibition
Pseudo-nitzschia multiseries PC9	Yes	HTCC 2559	0.10	0.68±0.02	0.18±0.10	73
		SA60	0.10		0.32±0.04	53
Pseudo-nitzschia multiseries IOES-1	Yes	HTCC 2559	0.04	0.50±0.02	-0.26±0.14	151
		SA60	0.05		-0.22±0.18	144
		HV2	0.03		-0.53±0.03	204
Pseudo-nitzschia fraudulenta OC1	Yes	HTCC 2559	0.01	0.63±0.03	0.44±0.02	30
		SA60	0.01		0.41±0.06	34
		HV2	0.01		0.45±0.02	29
			0.02	0.82±0.04	0.00±0.03	100
Nitzschia sp. RCC 80	No	SA60	0.02	1.37±0.04	1.38±0.04	0
Thalassiosira weissfloggii CCMP 1052	No	SA60	0.02	1.57±0.03	1.58±0.03	0
Thalassiosira oceanica CCMP 1003	No	SA60	0.02	0.92±0.01	0.78±0.00	15
Thalassiosira pseudonana CCMP 1335	Yes	SA60	0.01	0.93±0.02	0.63±0.02	33
Aureococcus anophagefferens CCMP 1784[b]	No	Isolate from Great South Bay, New York	0.20	ND	ND	>0

Abbreviation: ND, no data.

[a] $\mu_{control}$ is the diatom growth rate in the absence of additional *C. atlanticus*.

[b] From ref. [38].

Table 1.1: *Croceibacter atlanticus* impacts on selected heterokonts

observed with *T. weissflogii* (Table 1.1). Ref. [38] found that the pelagophyte *Aureococcus anophagefferens* was also inhibited by an isolate of *C. atlanticus* from Great South Bay, New York, although the extent of the inhibitory effect was not reported. Two of the seven tested diatoms were not impacted by the addition of *C. atlanticus*; neither of these diatom strains was axenic, raising the possibility that members of the pre-existing bacterial consortia from those diatoms mitigated the impacts of *C. atlanticus*. This possibility may explain how *C. atlanticus* remained a consistent member of the *P. multiseries* consortium without completely arresting diatom growth. Complex host-independent patterns of interaction have been observed in many host-dependent microbiomes (for example, ref. [39]), and likely exist in the diatom microenvironment as well.

1.4.3 Mechanism of antagonistic effect of C. atlanticus on T. pseudonana

I chose to focus on the diatom-associated strain SA60 given the similarity in antagonistic phenotypes of the three *C. atlanticus* strains and their high genome sequence identities – the SA60 genome is 98% identical and the Pm1 genome is 99% identical to HTCC 2559 at the nucleotide level and both are missing a genome segment involved in alginate metabolism (Supplementary Figure A.2). The interaction between strain SA60, and the model diatom *T. pseudonana* CCMP 1335 was examined in more detail because the annotated whole genome of this diatom is available [40], with complementary metabolomic data [41], and because the cells are small enough to analyze with flow cytometry.

Epifluorescence microscopy revealed that in co-culture, *C. atlanticus* cells colonized the cell surface of *T. pseudonana* while a number of *C. atlanticus* cells remained unattached to the diatoms. As *C. atlanticus* is non-motile except for surface gliding [25, 26], association with the diatom likely proceeds by random encounter followed by attachment and replication on the cell surface of the diatom. In acclimated non-axenic cultures and in nature, it may be that vertical transmission from parent to daughter cells is a more important mechanism of maintaining associations over multiple generations. The *C. atlanticus* population counted by flow cytometry consists of the unattached portion of the bacterial community–those cells

that did not encounter *T. pseudonana* and any cells that detached from the diatom either naturally or during sample processing.

In the previous co-culture experiments (Figure 1.1, Table 1.1), an aliquot of overnight grown cultures of *C. atlanticus* cells in spent MB was added directly to the different diatom cultures. Although in each instance the added MB media was diluted 500–10 000-fold (Table 1.1), variable amounts of undefined organic matter, as well as any growth inhibitor potentially released by *C. atlanticus* into the MB, were also added to the co-cultures. I therefore carried out a series of experiments to determine potential sources of the antagonistic effects on diatoms.

First, I tested whether *C. atlanticus* constitutively released a diatom growth inhibitor into the MB during overnight growth in the absence of diatoms. An aliquot of filter-sterilized MB media from an overnight grown *C. atlanticus* culture was added to a *T. pseudonana* culture and did not significantly, decrease the growth rate ($\mu_{1335+freshMB} = 1.01 \pm 0.02$ day^{-1}, $\mu_{1335+spentMB} = 1.08 \pm 0.05$ day^{-1}, $p = 0.13$) or final cell concentration ($C_{1335+freshMB} = 7.76 \pm 0.57$ ($\times 10^5$) cells ml^{-1}, $C_{1335+spentMB} = 7.71 \pm 0.80$ ($\times 10^5$) cells ml^{-1}, $p = 0.94$) of the diatom (Figure 1.2a). These results suggest that *C. atlanticus* does not constitutively release a growth inhibitor into MB under these conditions, at least at an effective concentration. By contrast, another study showed that the flavobacterium *Kordia algicida* constitutively secretes a protease that negatively impacts several diatoms [16]. One diatom, *Chaetoceros didymus*, appears to deactivate the algicidal activity of *K. algicida* through production of its own counter-acting protease-degrading enzymes [42]. It is not clear whether *T. pseudonana* can similarly degrade low concentrations of the *C. atlanticus* growth inhibitor.

Second, I asked whether *C. atlanticus* required a component of the organic matter carried over in the MB inoculum to elicit the negative effect on the diatom. When *C. atlanticus* cells were washed with the organic carbon-free diatom medium Aquil before addition to the *T. pseudonana* culture, the bacteria did not divide in co-culture with *T. pseudonana*, did not attach to the *T. pseudonana* cells nor did they impact the growth rate or final cell concentration of the diatom ($\mu_{1335} = 0.93 \pm 0.05$ day^{-1}, $\mu_{1335+SA60} = 0.97 \pm 0.04$ day^{-1},

$p = 0.40$; $C_{1335} = 6.26 \pm 0.65$ ($\times 10^5$) cells ml^{-1}, $C_{1335+SA60} = 7.70 \pm 0.03$ ($\times 10^5$) cells ml^{-1}, $p = 0.06$; Table 1.2, Figure 1.2b). In contrast, a *C. atlanticus* overnight culture washed with spent MB before addition to the co-culture remained antagonistic indicating that the washing step itself did not impact the *C. atlanticus* phenotype ($\mu_{1335+SA60} = 0.39 \pm 0.07$ day^{-1}, $\mu_{1335+SA60centrifuged} = 0.44 \pm 0.03$ day^{-1}, $p = 0.29$; $C_{1335+SA60} = 4.15 \pm 0.1$ ($\times 10^4$) cells ml^{-1}, $C_{1335+SA60centrifuged} = 4.07 \pm 1.42$ ($\times 10^4$) cells ml^{-1}, $p = 0.9$; Supplementary Figure A.3). This result suggests that specific nutrients present in MB and required in co-culture were not supplied by the diatom, at least at the start of the experiment.

To rule out that washing the cells in Aquil was toxic to the bacterium and inhibited its antagonistic activity, amino acids were added to Aquil-washed *C. atlanticus* in co-culture with *T. pseudonana* and bacterial growth was rescued (Supplementary Figure A.4). The amino-acid formulations used in this experiment include two mixtures of amino acids (named essential and non-essential based on cultured mammalian cell nutrient requirements). *C. atlanticus* requires both mixtures for growth, implying that there are amino acids in both solutions that are required by *C. atlanticus*.

The extent of the negative impact when supplemented with MB appeared to be related to *C. atlanticus* growth rate in co-culture. When *C. atlanticus* cells were washed with spent MB rather than fresh MB, the unattached *C. atlanticus* cells displayed a lower growth rate in co-culture, ($\mu_{spent} = 1.81 \pm 0.08$ day^{-1}, $\mu_{fresh} = 2.26 \pm 0.11$ day^{-1}, $p = 0.01$) and reached a lower final cell concentration ($C_{spent} = 2.57 \pm 0.06$ ($\times 10^7$) cells ml^{-1}, $C_{fresh} = 3.01 \pm 0.01$ ($\times 10^7$) cells ml^{-1} , $p = 0.01$). This lower *C. atlanticus* growth rate translated into a significantly reduced impact on the *T. pseudonana* growth rate (59% vs 71% inhibition, Table 1.2). These results reinforce the idea that the observed detrimental effect is driven by proliferation of *C. atlanticus* whose growth is supported by nutrients present in MB.

These co-culturing experiments show that the Aquil-washing step eliminates nutrients required for the growth of *C. atlanticus* in co-culture, and that early exponential phase cultures of *T. pseudonana* do not provide sufficient organic matter to support *C. atlanticus* cell division. *C. atlanticus* growth can, however, be rescued by the addition of amino acids

Media supplement	$\mu^a_{control}$ ± s.d. (day^{-1})	$\mu_{co-culture}$ ± s.d. (day^{-1})	% Growth inhibition
None (Aquil)	0.93 ± 0.05	0.97 ± 0.04	0
0.2% v/v fresh marine broth	1.01 ± 0.02	0.29 ± 0.03	71
0.2% v/v spent marine broth	1.08 ± 0.05	0.44 ± 0.03	59

[a] $\mu_{control}$ is the diatom growth rate in the absence of bacteria.

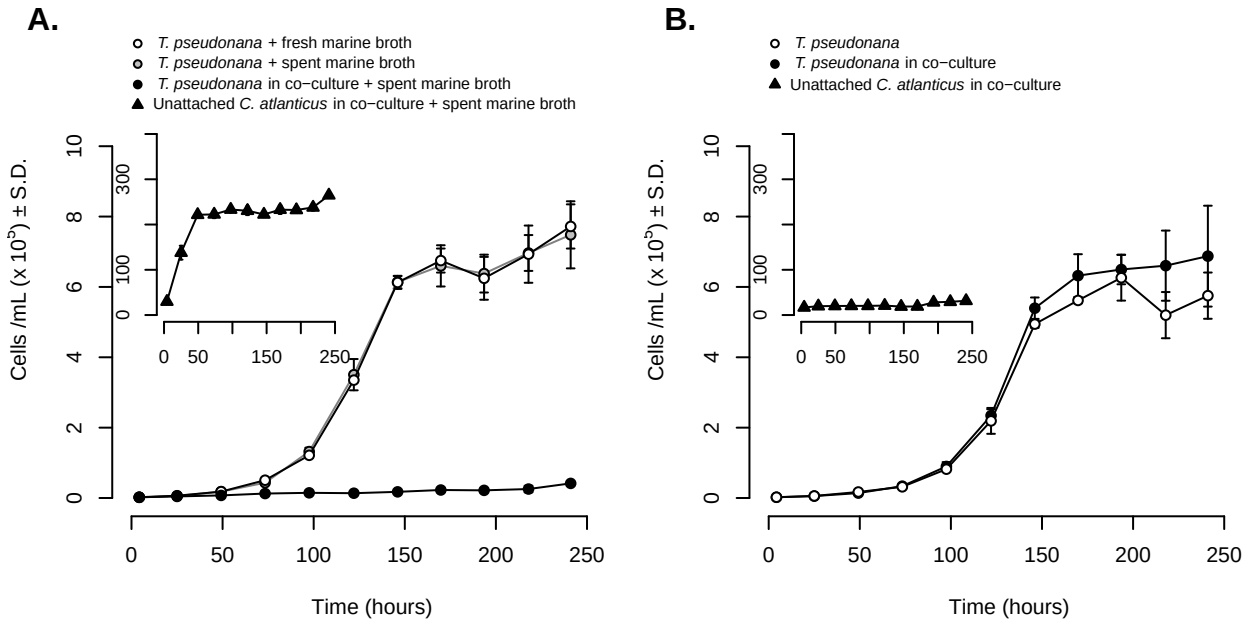
Table 1.2: Impact of *Croceibacter atlanticus* on the growth rate of *Thalassiosira pseudonana*

or MB to co-culture. In the absence of nutrients, *C. atlanticus* does not attach to the diatom cells and the antagonistic phenotype is not induced. In contrast, when *C. atlanticus* cells are added to the diatom cultures along with sufficient organic nutrients, the unattached cells continue to divide, presumably attaining a cell density that provides a sufficiently high encounter rate with the diatom cells. *C. atlanticus* likely has a mechanism to recognize and attach to diatom cells; perhaps algal exudates serve as infochemicals that induce colonization after an encounter. Once *C. atlanticus* cells attach to a diatom, they appear to proliferate on the cell surface, likely due to the enhanced flux of diatom-released nutrients at the cell surface not seen in the bulk media.

1.4.4 Effects of C. atlanticus on the cell size, morphology and DNA content of T. pseudonana

Microscopy and flow cytometry measurements of diatoms in mono-culture and co-culture indicated that over the course of a few days, *T. pseudonana* cells with attached bacteria elongate (Figures 1.3a and b, Supplementary Figure A.5), are filled with more plastids (Figure 1.3c), and typically possess more nuclei (Figure 1.3d) than cells without attached bacteria (Figure 1.4). In treatments where both *C. atlanticus* and MB were added, the vast majority of *T. pseudonana* cells hosted attached bacteria by the second day of co-culture (Figure 1.4a).

In examining both the means and distributions of flow cytometry parameters and mi-



(a) Filter-sterilized spent MB from an overnight culture of *C. atlanticus* has an indistinguishable effect relative to fresh MB on the growth of *T. pseudonana*, whereas a direct inoculate of overnight grown *C. atlanticus* drastically inhibits the growth of *T. pseudonana*; unattached *C. atlanticus* cells in co-culture display growth (inset).

(b) *C. atlanticus* cells that were washed with Aquil do not grow in co-culture with *T. pseudonana* (inset) and do not impact the growth of *T. pseudonana* in co-culture.

Figure 1.2: Effects of *C. atlanticus* SA60 and its growth media on the growth of *T. pseudonana* CCMP 1335. Symbols mark the mean of diatom and unattached bacterial cell concentration (circles, triangles) when grown alone, in the presence of *C. atlanticus* waste products, or in co-culture (white, gray, black; $n = 3$ biological replicates ± 1 s.d. and are shown when larger than symbol size).

croscopy measurements, a time course of the impact of *C. atlanticus* on *T. pseudonana* was apparent. Within 2 days of the addition of *C. atlanticus* cells, the growth rate of *T. pseudonana* in co-culture decreased significantly relative to the control (Figure 1.2a), whereas the chlorophyll *a* fluorescence (692/50nm emission) per cell in colonized populations increased significantly ($p = 0.001$) (Figure 1.4b, Supplementary Figure A.6). Microscopy measurements of the cellular surface area covered by plastids suggest that elevated chlorophyll *a* fluorescence is correlated with an increase in the number of plastids per cell (Figures 1.3c and 1.4). By day 3, the forward scatter of colonized *T. pseudonana* populations increased ($p = 0.001$) (Figure 1.4b, Supplementary Figure A.6). Increased forward scatter is often correlated to increased cell size [43], although this relationship is sometimes affected by other refractive properties of the cell [44]. However, microscopy measurements of *T. pseudonana* diameter and length (Figures 1.3a and b) leads to the conclusion that increased forward scatter of colonized cells can be attributed to their increased cell size. By day 4, side scatter values also increase significantly ($p = 0.001$) (Figure 1.4b, Supplementary Figure A.6). The reason for these shifts are not known, but changes in side scatter distributions may be related to cell shape, cytoplasmic granularity and differences in internal cell structures [44]. I also observed that *T. pseudonana* sank to the bottom of the culture tubes in treatments where the cells were colonized, possibly as a result of morphological changes. The similar pattern between cell length measurements and cell area covered by plastids suggests that the number of plastids scales with cell size.

Epifluorescence microscopy indicated that by day 6 the larger, SYBR Green-stained colonized *T. pseudonana* cells commonly possessed two or more nuclei rather than a single nucleus as observed in control cultures (Figure 1.3d). A significant proportion of cells without a nucleus were also observed on day 6 using microscopy (Figure 1.3d). This is suggestive of nuclear disintegration and DNA fragmentation, which accompanies apoptosis [45]. *C. atlanticus* thus appears to manipulate the cell cycle of the diatom, stimulating both polyploidy and nuclear disintegration.

Increased DNA content is often correlated with enhanced cell size [46, 47] as well as in-

creased metabolic activity [48, 49]. I hypothesize that manipulation of the diatom cell cycle could therefore be an effective strategy for *C. atlanticus* to induce morphological and physiological changes in its host that enhance the release of required metabolites. For example, *C. atlanticus* is unable to synthesize the amino acids isoleucine, valine, leucine and arginine, all of which *T. pseudonana* produces within its plastids and mitochondria, which may be more abundant now that the diatom is larger (Supplementary Table A.2, [40]). In addition, the *C. atlanticus* genome encodes a variety of peptidases and carbohydrate active enzymes (CAZYmes) and *C. atlanticus* can grow in defined media with amino acids as its sole source of organic carbon and nitrogen (data not shown). Secretion of carbohydrates and peptides by *T. pseudonana* [41] could also fulfill *C. atlanticus* nutritional requirements.

Polyploidy occurs naturally in a variety of plant, mammalian and insect cell types [49]. It has also been observed at nutrient exchange points between fungal or bacterial pathogens or symbionts and their host cells [48]. As ploidy correlates linearly with gene expression and metabolic flux, it has been hypothesized that increased DNA content may be a common biological mechanism to cope with increased metabolic demands [48]. In particular, primary metabolism (glycolysis, respiration and fermentation), transport, and plastid-localized production of metabolites are sensitive to increases in ploidy [48]. In mammals, pathogenic bacteria are known to target cell cycle progression of their host cells through production of effectors known as cyclomodulins [50]. These effectors regulate the expression of cyclins and cyclin-dependent kinases, required for cell cycle progression. Analogously, growth-promoting and pathogenic bacteria can interfere with the plant cell cycle hormonal balance to modulate host cyclin and cyclin-dependent kinase activity [51]. Whether *C. atlanticus* manipulates *T. pseudonana* cell division through a similar disruption of cyclin-mediated cell cycle progression requires further study.

1.5 Conclusion

Flavobacteria such as *C. atlanticus* have a reputation as fast-growing *r*-strategists that specialize on initiating organic matter degradation near the end of algal blooms [24]. These

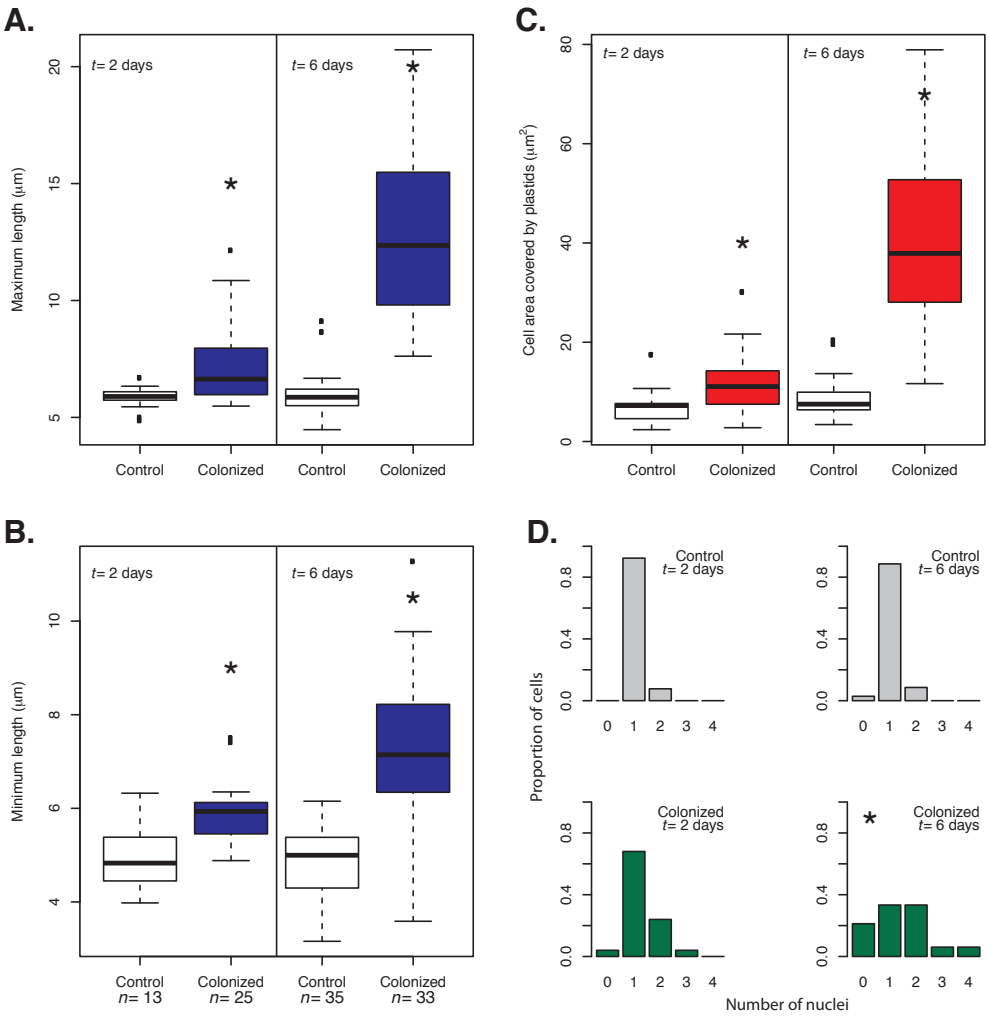


Figure 1.3: Microscopy measurements comparing colonized and control diatom cells after 2 and 6 days in culture. Box-and-whiskers plots of the (a) maximum and (b) minimum lengths and (c) the surface area per cell covered by plastids show the distribution of data including the median, quartiles, and range. N represents the number of cell measurements made. (d) Histograms of nucleus count per cell for colonized and control diatoms after 2 and 6 days in culture. Asterisks represent treatments that are statistically different from the control for the same incubation period ($p = 0.05$).

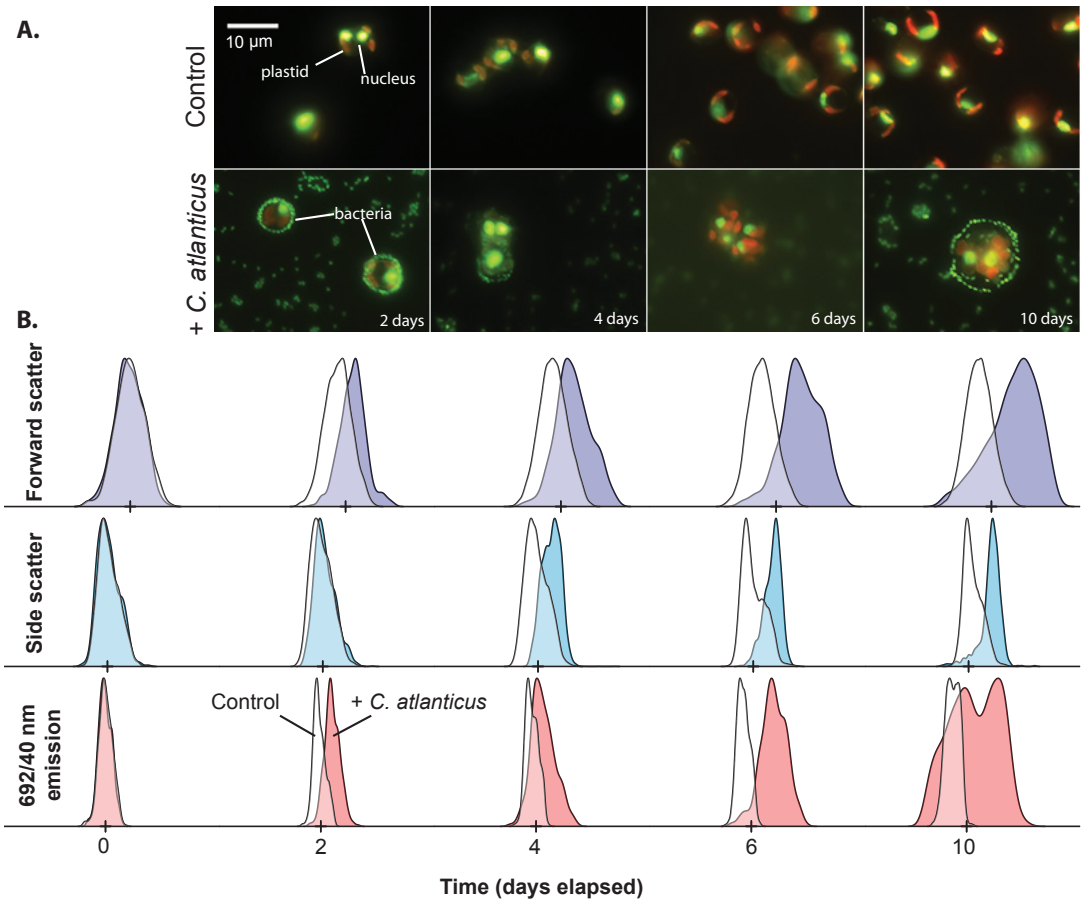


Figure 1.4: Attachment of *C. atlanticus* to *T. pseudonana* and impacts on diatom morphology. (a) Representative epifluorescence micrographs from the experiment portrayed in Figure 1.2 show *C. atlanticus* attachment to *T. pseudonana* and morphological changes over the course of the experiment. Diatom plastids autofluoresce red, SYBR Green-stained DNA from the diatom nucleus and the bacteria fluoresce green. The co-culture micrographs for day 4 and 6 focus on the diatoms rather than the attached bacteria to show examples of polyploidy. Scale bar = 10 μm. (b) Increased forward and side scatter values (blue, turquoise) over time show how *C. atlanticus* impacts the size and shape of *T. pseudonana*. Shifts in the 692/40 nm emission distribution (red) over time indicate that *C. atlanticus* impacts the amount of chlorophyll per cell. Colored polygons are representative distributions of *T. pseudonana* population scatter and fluorescence characteristics in co-culture; white polygons show the same for a control population. Distributions represent the fractions of total cells in a population.

results raise the possibility that some flavobacteria could contribute to diatom bloom decline if they, like *C. atlanticus*, are capable of halting cell division. The colonized diatom population appears to survive until the late stationary phase when there is some indication of nuclear disintegration. As *C. atlanticus* is only capable of gliding motility, it cannot actively seek out a new host and must continue to colonize an arrested diatom cell. By causing *T. pseudonana* to elongate, *C. atlanticus* increases the surface area of its niche. In addition, *C. atlanticus* benefits from colonizing a live diatom that continues to synthesize photosynthetic products. However, once nutrients are depleted in solution and the diatom stops its metabolic activities, it would be more beneficial for *C. atlanticus* to consume all available organics from the failing diatom and detach from the cell before it sinks.

The mechanisms of diatom-bacteria interactions can reveal new insights into the dynamics of marine microbial ecology. Although it is not yet known whether bacteria affect the cell cycle of diatoms in nature, I provide evidence that *C. atlanticus* can affect the DNA content of *T. pseudonana* and prior evidence that *Sulfitobacter* sp. can upregulate G1-S phase cyclins of *P. multiseries* in culture [18]. It may be that the potential for cell cycle manipulations is widespread in the marine environment and is a common ecological strategy. *C. atlanticus* sequences are prevalent across the global ocean [52], and the interaction appears to lack specificity. If *C. atlanticus* is capable of antagonizing a variety of phytoplankton in nature, then this microbe could have an important role in the remineralization of organic matter in the ocean. Interactions between *C. atlanticus* and other algal-associated bacteria may also be involved in mediating these antagonistic interactions.

1.6 Acknowledgements

I would like to thank Michael Carlson, Irina Oleinikov, Lyndsey Swanson and Megan Schatz for supplying and maintaining diatom cultures; Chris Berthiaume and Rhonda Morales for doing the genome sequencing and assembly work; Rhonda Morales, Jarred Swalwell and Francois Ribalet for advice on flow cytometry data collection and processing. This work was supported by Gordon and Betty Moore Foundation grant GBMF3776 awarded to E. Virginia

Armbrust. Helena van Tol was partially supported by a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship Master's program award (NSERC PGS M 489334).

References

[1] WB Whitman, DC Coleman, and WJ Wiebe. "Prokaryotes: The unseen majority." In: *Proceedings of the National Academy of Sciences* 95.12 (June 1998), pp. 6578–83. ISSN: 0027-8424. DOI: 10.1073/pnas.95.12.6578.

[2] F Azam et al. "Bacteria-organic matter coupling and its significance for oceanic carbon cycling." In: *Microbial Ecology* 28.2 (Sept. 1994), pp. 167–79. ISSN: 0095-3628. DOI: 10.1007/BF00166806.

[3] F Azam. "Microbial control of oceanic carbon flux: the plot thickens." In: *Science*. New Series 280.5364 (May 1998), pp. 694–6. ISSN: 00368075. DOI: 10.1126/science.280.5364.694.

[4] R Stocker. "Marine microbes see a sea of gradients." In: *Science* 338.6107 (Nov. 2012), pp. 628–33. ISSN: 0036-8075. DOI: 10.1126/science.1208929.

[5] W Bell and R Mitchell. "Chemotactic and growth responses of marine bacteria to algal extracellular products." In: *Biological Bulletin* 143.2 (Oct. 1972), pp. 265–77. ISSN: 00063185. DOI: 10.2307/2383716.

[6] JRN Lazier and KH Mann. "Turbulence and the diffusive layers around small organisms." In: *Deep Sea Research Part A. Oceanographic Research Papers* 36.11 (Nov. 1989), pp. 1721–33. ISSN: 01980149. DOI: 10.1016/0198-0149(89)90068-X.

[7] R Stocker and JR Seymour. "Ecology and physics of bacterial chemotaxis in the ocean." In: *Microbiology and Molecular Biology Reviews* 76.4 (Nov. 2012), pp. 792–812. ISSN: 1092-2172. DOI: 10.1128/MMBR.00029-12.

[8] S Smriga et al. “Chemotaxis toward phytoplankton drives organic matter partitioning among marine bacteria.” In: *Proceedings of the National Academy of Sciences* 113.6 (Feb. 2016), pp. 1576–81. ISSN: 0027-8424. DOI: 10.1073/pnas.1512307113.

[9] L Karp-Boss, E Boss, and PA Jumars. “Nutrient fluxes to planktonic osmotrophs in the presence of fluid motion.” In: *Oceanography And Marine Biology An Annual Review* 34 (1996), pp. 71–107. ISSN: 00243590.

[10] EV Armbrust. “The life of diatoms in the world’s oceans.” In: *Nature* 459.7244 (May 2009), pp. 185–92. ISSN: 0028-0836. DOI: 10.1038/nature08057.

[11] S Agusti et al. “Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump.” In: *Nature Communications* 6.May (July 2015), pp. 1–8. ISSN: 2041-1723. DOI: 10.1038/ncomms8608.

[12] SA Amin, MS Parker, and EV Armbrust. “Interactions between diatoms and bacteria.” In: *Microbiology and Molecular Biology Reviews* 76.3 (Sept. 2012), pp. 667–84. ISSN: 1092-2172. DOI: 10.1128/MMBR.00007-12.

[13] X Mayali et al. “Algicidal bacteria in the sea and their impact on algal blooms.” In: *The Journal of Eukaryotic Microbiology* 51.2 (Mar. 2004), pp. 139–44. ISSN: 1066-5234. DOI: 10.1111/j.1550-7408.2004.tb00538.x.

[14] BP Durham et al. “Cryptic carbon and sulfur cycling between surface ocean plankton.” In: *Proceedings of the National Academy of Sciences* 112.2 (Jan. 2015), pp. 453–7. ISSN: 0027-8424. DOI: 10.1073/pnas.1413137112.

[15] SO Lee et al. “Involvement of an extracellular protease in algicidal activity of the marine bacterium *Pseudoalteromonas* sp. strain A28.” In: *Applied and environmental microbiology* 66.10 (Oct. 2000), pp. 4334–9. ISSN: 0099-2240. DOI: 10.1128/AEM.66.10.4334-4339.2000.

[16] C Paul and G Pohnert. "Interactions of the algicidal bacterium Kordia algicida with diatoms: regulated protease excretion for specific algal lysis." In: *PloS one* 6.6 (Jan. 2011), e21032. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0021032.

[17] MR Seyedsayamdost et al. "The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*." In: *Nature Chemistry* 3.4 (Apr. 2011), pp. 331–5. ISSN: 1755-4330. DOI: 10.1038/nchem.1002.

[18] SA Amin et al. "Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria." In: *Nature* 522.7554 (May 2015), pp. 98–101. ISSN: 0028-0836. DOI: 10.1038/nature14488.

[19] KA Hubbard, G Rocap, and EV Armbrust. "Inter- and intraspecific community structure within the diatom genus *Pseudo-nitzschia* (Bacillariophyceae)." In: *Journal of Phycology* 44.3 (June 2008), pp. 637–49. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2008.00518.x.

[20] RRL Guillard. "Culture of phytoplankton for feeding marine invertebrates." In: *Culture of Marine Invertebrate Animals*. Ed. by WL Smith and MH Chanley. New York: Spring US, 1975, pp. 29–60.

[21] NM Price et al. "Preparation and chemistry of the artificial algal culture medium Aquil." In: *Biological Oceanography* 6 (1989), pp. 443–61. ISSN: 0196-5581. DOI: 10.1080/01965581.1988.10749544.

[22] SM Shishlyannikov et al. "A procedure for establishing an axenic culture of the diatom *Synedra acus* subsp. *radians* (Kütz.) Skabibitsch. from Lake Baikal." In: *Limnology and Oceanography: Methods* 9.10 (Oct. 2011), pp. 478–84. ISSN: 15415856. DOI: 10.4319/lom.2011.9.478.

[23] CE ZoBell. "Studies on marine bacteria. I. The cultural requirements of heterotrophic aerobes." In: *Journal of Marine Research1* 4 (1941), pp. 42–75.

[24] H Teeling et al. "Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom." In: *Science* 336.6081 (May 2012), pp. 608–11. ISSN: 0036-8075. DOI: 10.1126/science.1218344.

[25] J Cho and SJ Giovannoni. "*Croceibacter atlanticus* gen. nov., sp. nov., a novel marine bacterium in the family Flavobacteriaceae." In: *Systematic and Applied Microbiology* 26.1 (Jan. 2003), pp. 76–83. ISSN: 07232020. DOI: 10.1078/072320203322337344.

[26] MJ McBride and Y Zhu. "Gliding motility and Por secretion system genes are widespread among members of the phylum Bacteroidetes". In: *Journal of Bacteriology* 195.2 (Jan. 2013), pp. 270–8. ISSN: 1098-5530. DOI: 10.1128/JB.01962-12.

[27] D Marie et al. "Enumeration of phytoplankton, bacteria, and viruses in marine samples". In: *Current Protocols in Flow Cytometry*. Ed. by J P Robinson. New York: John Wiley & Sons, Inc., 1999, pp. 11.11.1–11.11.15.

[28] M Lunau et al. "An improved method for counting bacteria from sediments and turbid environments by epifluorescence microscopy." In: *Environmental Microbiology* 7.7 (July 2005), pp. 961–8. ISSN: 1462-2912. DOI: 10.1111/j.1462-2920.2005.00767.x.

[29] B Ellis et al. *flowCore: flowCore: basic structures for flow cytometry data.* 2015. DOI: 10.18129/B9.bioc.flowCore.

[30] F. Hahne et al. *flowStats: statistical methods for the analysis of flow cytometry data.* 2015. DOI: 10.18129/B9.bioc.flowStats.

[31] B Ellis et al. *flowViz: visualization for flow cytometry.* 2015. DOI: 10.18129/B9.bioc.flowViz.

[32] DR Zerbino and E Birney. "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs." In: *Genome Research* 18.5 (Feb. 2008), pp. 821–9. ISSN: 1088-9051. DOI: 10.1101/gr.074492.107. eprint: 0209100 (arXiv:quant-ph).

[33] V Iverson et al. "Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota." In: *Science* 335.6068 (Mar. 2012), pp. 587–90. ISSN: 1095-9203. DOI: 10.1126/science.1212665.

[34] VM Markowitz et al. "IMG ER: A system for microbial genome annotation expert review and curation." In: *Bioinformatics* 25.17 (Sept. 2009), pp. 2271–8. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp393.

[35] RS Harris. "Improved pairwise alignment of genomic DNA." PhD thesis. The Pennsylvania State University, 2007, p. 84. URL: http://www.bx.psu.edu/Brsharris/lastz/.

[36] L Guy, J Roat Kultima, and SGE Andersson. "genoPlotR: comparative gene and genome visualization in R." In: *Bioinformatics* 26.18 (Sept. 2010), pp. 2334–35. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq413.

[37] R Caspi et al. "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." In: *Nucleic Acids Research* 42.D1 (Jan. 2014), pp. 471–80. ISSN: 03051048. DOI: 10.1093/nar/gkt1103.

[38] AD Frazier et al. "Bacterial lysis of *Aureococcus anophagefferens* CCMP 1784 (Pelagophyceae)." In: *Journal of Phycology* 43.3 (June 2007), pp. 461–5. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2007.00353.x.

[39] P Trosvik et al. "Web of ecological interactions in an experimental gut microbiota." In: *Environmental Microbiology* 12.10 (June 2010), pp. 2677–87. ISSN: 14622912. DOI: 10.1111/j.1462-2920.2010.02236.x.

[40] EV Armbrust et al. "The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism." In: *Science* 306.5693 (Oct. 2004), pp. 79–86. ISSN: 0036-8075. DOI: 10.1126/science.1101156.

[41] K Longnecker, MC Kido Soule, and EB Kujawinski. "Dissolved organic matter produced by *Thalassiosira pseudonana*." In: *Marine Chemistry* 168 (Jan. 2015), pp. 114–23. ISSN: 0304-4203. DOI: 10.1016/j.marchem.2014.11.003.

[42] C Paul and G Pohnert. "Induction of protease release of the resistant diatom *Chaetoceros didymus* in response to lytic enzymes from an algicidal bacterium." In: *PLoS ONE* 8.3 (Jan. 2013), e57577. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0057577.

[43] RJ Olson, ER Zettler, and OK Anderson. "Discrimination of eukaryotic phytoplankton cell types from light scatter and autofluorescence properties measured by flow cytometry." In: *Cytometry* 10.5 (Sept. 1989), pp. 636–43. ISSN: 0196-4763. DOI: 10.1002/cyto.990100520.

[44] HM Shapiro. "Physical parameters and their uses." In: *Practical Flow Cytometry*. Ed. by HM Shapiro. New York, NY: John Wiley & Sons, Inc., 2003, pp. 273–85.

[45] X Huang et al. "Cytometric assessment of DNA damage in relation to cell cycle phase and apoptosis." In: *Cell Proliferation* 38.4 (Aug. 2005), pp. 223–43. ISSN: 0960-7722. DOI: 10.1111/j.1365-2184.2005.00344.x.

[46] P Von Dassow et al. "Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus *Thalassiosira* (Bacillariophyceae)." In: *Journal of Phycology* 44.2 (Apr. 2008), pp. 335–49. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2008.00476.x.

[47] JA Koester et al. "Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta)." In: *BMC Evolutionary Biology* 10.1 (2010), p. 1. ISSN: 1471-2148. DOI: 10.1186/1471-2148-10-1.

[48] MC Wildermuth. "Modulation of host nuclear ploidy: A common plant biotroph mechanism." In: *Current Opinion in Plant Biology* 13.4 (Aug. 2010), pp. 449–58. ISSN: 13695266. DOI: 10.1016/j.pbi.2010.05.005.

[49] BA Edgar, N Zielke, and C Gutierrez. "Endocycles: A recurrent evolutionary innovation for post-mitotic cell growth." In: *Nature Reviews Molecular Cell Biology* 15.3 (Feb. 2014), pp. 197–210. ISSN: 1471-0072. DOI: 10.1038/nrm3756.

[50] J Nougayrède et al. "Cyclomodulins: Bacterial effectors that modulate the eukaryotic cell cycle." In: *Trends in Microbiology* 13.3 (Mar. 2005), pp. 103–10. ISSN: 0966842X. DOI: 10.1016/j.tim.2005.01.002.

[51] E Stes et al. "A successful bacterial coup d'état: How *Rhodococcus fascians* redirects plant development." In: *Annual review of phytopathology* 49.1 (Sept. 2011), pp. 69–86. ISSN: 0066-4286. DOI: 10.1146/annurev-phyto-072910-095217.

[52] S Yooseph et al. "Genomic and functional adaptation in surface ocean planktonic prokaryotes." In: *Nature* 468.7320 (Nov. 2010), pp. 60–6. ISSN: 0028-0836. DOI: 10.1038/nature09530.

Chapter 2

GENOME-SCALE METABOLIC MODEL OF THE DIATOM *THALASSIOSIRA PSEUDONANA* HIGHLIGHTS THE IMPORTANCE OF DIATOM NITROGEN AND SULFUR METABOLISM IN REDOX BALANCE

2.1 Abstract

Thalassiosira pseudonana strain CCMP 1335 is a model for diatom molecular biology and physiology, with a 60-year history of studies. Diatoms are known to secrete organic matter that fuels secondary production, though our knowledge of how their physiology impacts the character of dissolved organic matter remains limited. Light energy provides reducing power to photosynthetic organisms, while creating the challenge of avoiding cellular damage during energy processing and balancing redox reactions. To better understand the interplay between redox balance and organic matter secretion, I reconstructed a genome-scale metabolic model of CCMP 1335 that simulates the metabolic activities of 1,426 genes via a network of 1,996 unique metabolites produced through 5,649 reactions distributed across six subcellular compartments. A mechanistic model of photon absorption and electron transfer via fucoxanthin chlorophyll *a/c* binding proteins was implemented, as well as experimentally determined photosynthetic constraints. Using dynamic Flux Balance Analysis, I simulated growth in batch culture by tracking changes in cell density, self-shading, nutrient concentration, extracellular metabolites, and the storage carbohydrate chrysolaminarin under different growth conditions including NO_3 versus NH_4, N-limitation, and P-limitation. The model was used to examine the dissipation of reductants in the plastid generated through light-dependent processes. Simulations indicate that nitrate assimilation is an important means of dissipating reductants. However, under nitrate-limiting conditions, sulfate uptake plays a similar role.

The use of either nitrate or sulfate uptake to balance redox reactions leads to the secretion of distinct organic nitrogen and sulfur compounds, which fuel secondary production in the surface ocean.

2.1.1 Importance

Diatoms are a diverse group of unicellular eukaryotic algae that evolved through secondary endosymbiosis. They thrive in dynamic environments – both freshwater and marine ecosystems – as planktonic and benthic organisms. Diatoms form the base of coastal food webs and are estimated to be responsible for ∼20% of carbon fixed globally. As such, diatom metabolism has an important influence on global biogeochemistry. Metabolic models of marine microorganisms link genes to ecosystems and may be key to integrating molecular data with models of ocean biogeochemistry. A model of the diatom *Thalassiosira pseudonana* provides a mechanistic explanation for the production of ecologically and climatologically relevant compounds.

2.2 Introduction

Diatoms are unicellular photosynthetic eukaryotes derived from a secondary endosymbiotic event, when a heterotrophic eukaryote engulfed a red algal cell and acquired a chloroplast [1]. They appeared in the fossil record ∼ 180 million years ago [2]. Through horizontal gene transfer, diatoms also acquired a diversity of bacterial genes [3] to create an organism with a combination of metabolic pathways distinct from other photosynthetic organisms. For example, diatoms have a complete urea cycle derived from both the exosymbiont and bacteria [4], photosynthesis from the algal endosymbiont [1], as well as xylanases, glucanases, amidohydrolase, carbon-nitrogen hydrolase, SAM-dependent decarboxylases and methyltransferases from bacteria [5]. Diatoms have a wide assortment of uncommon metabolic capabilities including silica precipitation, chitin biosynthesis, carbohydrate storage as the polymer chrysolaminarin, and biosynthesis of the pigments fucoxanthin, diadinoxanthin, and diatoxanthin. Novel aspects of diatom metabolism, such as the biosynthesis of 2,3-dihydroxypropane-1-

sulfonate (DHPS), are still being discovered [6]. It is a special challenge of systems biology to understand how this unique combination of interlocking pathways has allowed diatoms to thrive in the dynamic conditions of oceanic ecosystems.

A central need for photosynthetic organisms is to balance redox reactions, particularly when nutrient availability limits growth. Light energy drives linear electron flow from water split at photosystem II (PSII) to PSI, generating both reducing power (NADPH) and a proton gradient across the thylakoid membrane that drives ATP synthase. An ATP/NADPH ratio of $\sim$1.5 is required for CO_2 reduction by the Calvin-Benson-Bassham cycle [7]. Linear electron flow alone is unlikely to satisfy these ATP requirements, which may cause the chloroplast to become over-reduced, causing damage to the thylakoid membranes and subsequent photoinhibition [8]. In plants, NADPH-consuming pathways and alternative electron pathways that produce ATP without generating NADPH are thought to play an important role in balancing redox reactions in the plastid [9, 10]. Alternative electron fluxes include cyclic electron flow (CEF) around PSI, and water-to-water cycles where electrons from water oxidation at PSII are re-routed to an oxidation pathway – the Mehler reaction, plastoquinone terminal oxidase, or photorespiration [11]). In contrast to plants and green algae, diatoms instead preferentially regulate the ATP/NADPH ratio via energetic coupling between the plastid and the mitochondria [12], by which reduced metabolites are shuttled out from the plastid to fuel ATP generation in the mitochondria. Genome-scale metabolic modeling of the diatom *Phaeodactylum tricornutum* suggested that a malate shuttle, a glutamine-ornithine cycle [13], or a glycerol-3-phosphate/dihydroxyacetone phosphate shuttle [14] could be responsible for exchanging metabolites between these organelles.

Nitrate and sulfate assimilation are also involved in the dissipation of reducing equivalents in the plastid. Nitrite reductase consumes 3 NADPH to convert nitrite to ammonia, and the GS-GOGAT cycle (glutamine synthase – glutamine oxoglutarate aminotransferase) utilizes 2 reduced ferredoxin to assimilate ammonia, while only consuming 1 ATP. During sulfate assimilation, sulfate is converted to APS and either APS reductase or PAPS [(phospho)adenosine phosphosulfate] reductase utilizes a reduced thioredoxin to produce sulfite,

consuming the equivalent of 2 ATP and 1 NADPH. Sulfite reductase then consumes 6 reduced ferredoxin to produce sulfide for cysteine. Alternatively, sulfite can be utilized by UDP-sulfoquinovose synthase (which consumes no reductants while producing sulfolipids) or PAPS-mediated 2-aminoacrylate sulfotransferase (to produce L-cysteate, a hypothesized precursor for DHPS biosynthesis [15] which consumes 1 NADPH). Phosphate is assimilated by ATP synthase both in the plastid and the mitochondria. There is no genetic or experimental evidence that diatoms can reduce phosphates [16] to phosphite or phosphonate.

Metabolite production and secretion or storage can also help balance redox reactions or dissipate energy [17], particularly when biomass production is otherwise inhibited. Diatoms are known to secrete more metabolites in conditions of high light intensity [18], suboptimal temperature [19], and nutrient stress [20]. When nitrate or sulfate assimilation rates increase to balance redox reactions, resulting nutrient imbalances can be corrected through excess production of different types of organic nitrogen or sulfur compounds. And as carbon fixation continues under nutrient-limited conditions, excess carbon can be excreted as organic carbon or stored as chrysolaminarin and lipids [21].

Here I created a mechanistic model of metabolism for the diatom *Thalassiosira pseudonana* CCMP 1335. Light dependent respiration and nutrient uptake rates were constrained with experimental data, and growth was simulated in batch culture. These simulations demonstrate how the character of secreted metabolites depends on balancing cell redox status and nutrient availability.

2.3 Results

2.3.1 Reconstruction of the Thalassiosira pseudonana metabolic model

A genome-scale metabolic model of *T. pseudonana* CCMP 1335 (*i*Tps1426) was generated using as a framework the model *i*LB1027_lipid for the diatom *Phaeodactylum tricornutum* CCAP 1055/1 [13]. The *i*Tps1426 model has six compartments that represent the cytosol, mitochondria, chloroplast, thylakoid lumen, peroxisome as well as the extracellular environ-

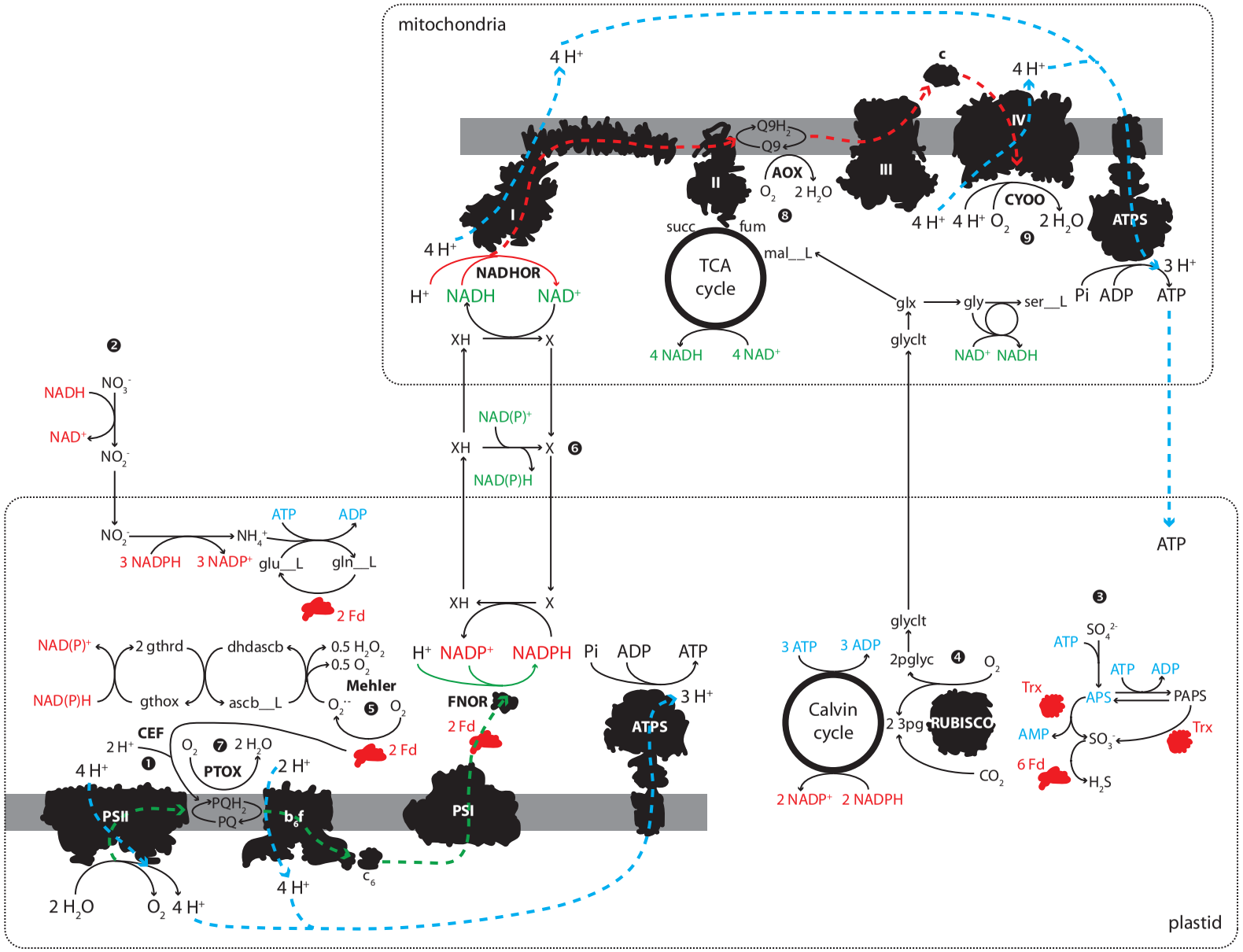


Figure 2.1: Diagram illustrating the principle reactions involved in generating ATP and balancing the ATP/NADPH ratio.

Figure 2.1: (Previous page.) Black shapes: enzymatic complexes, grays bars: plastidial or mitochondrial membranes, dotted lines: grouping of plastidial and mitochondrial reactions, green: reactions that produce NAD(P)H equivalents, red: reactions that consume NAD(P)H equivalents, blue: ATP producing and consuming reactions. Abbreviations: L-glutamate (glu__L), L-glutamine (gln__L), 2-phosphoglycolate (2pglyc), glycolate (glyclt), glyoxylate (glx), glycine (gly), L-serine (ser__L), L-malate (mal__L), fumarate (fum), succinate (succ), dehydroascorbate (dhdascb), L-ascorbate (ascb__L), oxidized glutathione disulfide (gthox), reduced glutathione (gthrd), reduced ferredoxin (Fd), reduced thioredoxin (Trx), plastoquinone (PQ), plastoquinol (PQH_2), ferredoxin:NADPH oxidoreductase (FNOR), NADH:ubiquinone oxidoreductase (NADHOR), ATP synthase (ATPS). Numbers indicate reaction references in the text including: cyclic electron flow (CEF, ❶), nitrate assimilation (❷), sulfate assimilation (❸), ribulose-1,5,-bisphosphate oxygenase (RUBISO, ❹), the Mehler reaction (❺), energetic coupling between the plastid and the mitochondria (❻), plastid terminal oxidase (PTOX, ❼), alternative oxidase (AOX, ❽), cytochrome *c* oxidase complex IV (CYOO, ❾).

ment. The *T. pseudonana* model contains 6,109 reactions that represent a network of 2,778 metabolites and 1,426 genes, approximately 10-12% of the *T. pseudonana* genome (Table 2.1, Supplementary Data Set B.1). A reaction representing the biomass composition of *T. pseudonana* was developed and used as an objective function during growth simulations. Additionally, several blocked reactions were resolved, and the number of dead-end metabolites in the network was reduced. Differences between the two models include changes in the subcellular localization of reactions based on protein targeting sequences, improvements in modeling light absorption, and the inclusion of known metabolic differences between the two diatoms.

Two major metabolic differences distinguish the diatom species. First, *T. pseudonana* has an absolute requirement for vitamin B_{12} as it possesses only the B_{12}-dependent methionine synthase gene METH [22]. Adenosylcobalamin, aquacobalamin, and methylcobalamin were

included in the biomass reaction as estimated mmol proportions of 1-gram dry weight ([23], Supplementary Data Set B.2). Additionally, the methionine synthase reaction ('METS_c') was modified to include dependence on methylcobalamin, the methionine synthase reductase reaction ('MTRR_c') was added, and the methylmalonyl-CoA reaction ('MMM_m') was modified to include dependence on adenosylcobalamin (Table 2.2). As a result, *i*Tps1426 requires a form of vitamin B_{12} to produce biomass. A second difference is that *T. pseudonana* utilizes cytochrome c_6 [24, 25] rather than the copper-containing plastocyanin to shuttle electrons between cytochrome b_6/f and photosystem I. Plastocyanin was therefore replaced with cytochrome c_6 in *i*Tps1426. Although this modification should include additional iron requirements for *T. pseudonana*, it had no impact on model results as neither *i*Tps1426 nor *i*LB1027_lipid includes the details of protein biosynthesis or metal requirements. The only metal required by the *i*Tps1426 model is magnesium for chlorophyll biosynthesis.

An orthoMCL comparison [26] of homologous genes between *T. pseudonana* and *P. tricornutum* was conducted to guide additional modifications to *i*Tps1426. These analyses confirmed information from the literature that *T. pseudonana* lacks the enzymes guanine deaminase, tryptophanase, ATP citrate synthase, and β-carbonic anhydrase [27], and these enzymes were removed from *i*Tps1426. In addition, citrate synthase was modified to deprotonate water rather than phosphorylate ADP. Differences between *T. pseudonana* and *P. tricornutum* in the localization of carbonic anhydrases [28] were incorporated into *i*Tps1426 and a cytoplasmic carbonic anhydrase reaction was added. Additional differences in the specific localizations of carbonic anhydrases within the plastid (endoplasmic reticulum, stroma, or pyrenoid) and the periplasmic space were not included as this level of cellular compartmentalization is not present in either model. My analysis also detected some potential differences between *P. tricornutum* and *T. pseudonana* that have not been previously discussed in the literature. Reactions present only in *P. tricornutum* include the metabolism of xylose via xylose isomerase feeding into the pentose phosphate pathway, the transamination of 4-aminobutyrate, the conversion of O-acetyl-L-homoserine to L-homocysteine via O-acetyl-L-homoserine acetate-lyase, L-tryptophan deamination, the conversion of agmatine

to putrescine, the conversion of formamide to formate and ammonia, and the degradation of guanine to xanthine. These reactions were eliminated from *i*Tps1426.

Metabolic pathways and reactions present in *T. pseudonana*, but not *P. tricornutum*, include chitin biosynthesis and degradation, and the cleavage of pyruvate into acetyl-CoA and formate. *i*Tps1426 was modified to include this information. Additionally, lipid metabolism was re-configured in *i*Tps1426 to reflect known differences in lipid composition [29, 30]. An important storage molecule in diatoms is chrysolaminarin, a polymer of 1,3-glucan. Production and respiration of 1,3-glucan and chitin was added to *i*Tps1426. Pathways for the biosynthesis of 2,3-dihydroxypropane-1-sulfonate (DHPS) [6, 15] from serine were added to the model. I included in *i*Tps1426 additional transport reactions for biotin, D-lactate, cyanocobalamin, aquacobalamin, DMSP, DHPS, glycine betaine, N-acetyltaurine, formamide, formate, uracil, acetate, choline, xanthine, UDP-N-acetyl-alpha-D-glucosamine, L-glutamate, L-aspartate, L-isoleucine, L-leucine, L-valine, L-asparagine, L-glutamine, L-alanine, L-histidine, L-serine, L-threonine, and glycine based on information from the literature, gene annotations, and information from other algal models [31, 32].

2.3.2 Constraints on light absorption and photosynthesis

An important addition to *i*Tps1426 was inclusion of a mechanistic model of photon absorption and electron transfer by fucoxanthin chlorophyll *a/c* binding proteins (FCPs), using a similar approach as in the cyanobacterial model constructed by ref. [31]. Photon exchange reactions were added for each 20-nm wavelength in the photosynthetically active radiation (PAR) range (400-700 nm). Photon absorption reactions for each 20-nm bin were included to calculate the amount of photons absorbed by each pigment (chlorophyll a, c_1, c_2, fucoxanthin, diatoxanthin, β-carotene) at each wavelength. The diatom light-harvesting complex is composed of fucoxanthin, chlorophyll c_1, c_2, and chlorophyll a, three pigments that efficiently transfer energy to the reaction centers of PSII and PSI [33]. I included reactions that describe energy transfer efficiency from excited pigments to chlorophyll a in PSII and PSI. I also included pigment de-excitation reactions for diatoxanthin and β-carotene and

Property	*Thalassiosira pseudonana* CCMP 1335	*Phaeodactylum tricornutum* CCAP 1055/1
Model name	*i*Tps1426	*i*LB1027_lipid
Genes		
Total	11,849 (Ambrust *et al.*, 2004)/ 13,344 (Gruber *et al.*, 2015)	10,402 (Bowler *et al.*, 2008)
Included in models	1,426 (12.03% / 10.69%)	1,032 (9.92%)
Complexed[a]	182	172
Reactions		
Reversible	650	423
Irreversible	5,459	4,033
Gene associated	5,605	4,150
Non-gene associated	504	306
Metabolic	5,569	4,093
Transport	446	308
Demand[b]	10	13
Sink[c]	3	1
Exchange[d]	67	30
Biomass	10	8
ATP maintenance	4	3
Extracellular	116	51
Cytoplasm	4,384	3,078
Chloroplast	911	657
Thylakoid lumen	12	7
Mitochondria	540	525
Peroxisome	146	138
Total	6,109	4,456
Unique	5,649	4,130
Blocked[e]	260	381
Orphaned[f]	0	4
Metabolites		
Extracellular	68	30
Cytoplasm	1,509	1,130
Chloroplast	570	428
Thylakoid lumen	15	9
Mitochondria	475	443
Peroxisome	141	132
Total	2,778	2,172
Unique	1,996	1,583
Dead-ends[g]	255	340

[a] Complexed genes are those that together encode multiple subunits of an enzyme. They are denoted as complexes in the gene reaction rules using the word 'and', while genes performing the same reaction are connected by the word 'or'.

Table 2.1: Comparison of attributes of *i*Tps1426 and *i*LB1027_lipid

[b] Demand reactions are unbalanced reactions (have substrates but no products). These reactions deal with metabolites that are known to be produced (and not consumed) but have no degradation pathway, are not substrates in the biomass reaction, and are not known to be transported out of the cell, for example loss of photons (as heat or fluorescence).

[c] Sink reactions are similar to demand reactions but are reversible. These reactions are a source and sink for metabolites that are required by the model but are not part of the extracellular environment, for example chitin.

[d] Exchange reactions are unbalanced extracellular reactions that are used to control the supply or removal of metabolites in the media.

[e] Blocked reactions are reactions that cannot carry flux due to missing reactions in the network.

[f] Orphaned reactions are blocked reactions that are disconnected from the entire network.

[g] Dead-end metabolites are metabolites that are only linked to blocked reactions.

Reaction ID	Name	Reaction
METS1_B_{12}_c	Methionine synthase (B_{12}-dependent), part 1	5mthf_c + cbl1_c + h_c −− > mcbl3_c + thf_c
METS2_B_{12}_c	Methionine synthase (B_{12}-dependent), part 2	hcys__L_c + mcbl3_c −− > cbl1_c + h_c + met__L_c
MTRR_c	Methionine synthase reductase	2.0 amet_c + 2.0 cbl2_c + nadph_c −− > 2.0 ahcys_c + h_c + 2.0 mcbl3_c + nadp_c
MMM1_B_{12}_m	Methylmalonyl-CoA mutase (B_{12}-dependent), part 1	adoch2_m + mmcoa__R_m <=> adoch3_m + mmsucr_m
MMM2_B_{12}_m	Methylmalonyl-CoA mutase (B_{12}-dependent), part 2	adoch3_m + mmsucr_m <=> adoch2_m + succoa_m
COB_{12}_m	Coenzyme B_{12} loading	adocbl_m <=> adoch2_m + cbl2_m

5mthf_c: 5-methyltetrahydrofolate

cbl1_c: Cob(I)alamin

h_c: H^+

mcbl3_c: Methylcob(III)alamin

thf_c: 5,6,7,8-tetrahydrofolate

hcys__L_c: L-Homocysteine

met__L_c: L-Methionine

amet_c: S-adenosyl-L-methionine

cbl2_c: Cob(II)alamin

nadph_c: NADPH

ahcys_c: S-adenosyl-L-homocysteine

nadp_c: $NADP^+$

adoch2_m: 5-deoxyadenosyl radical

mmcoa__R_m: (R)-Methylmalonyl-CoA

adoch3_m: 5-deoxyadenosil radical

mmsucr_m: Methylmalonyl-CoA / Succinyl-CoA radical

succoa_m: Succinyl-CoA

adocbl_m: Adenosylcobalamin

Table 2.2: B_{12}-dependent reactions modified for *i*Tps1426

non-photochemical quenching reactions to dissipate excess energy as heat or fluorescence. The calculations for photon absorption and electron transfer are included in Supplementary Data Set B.3.

Photodamage of the D1 subunit was included as a component of the PSII reaction and a non-growth associated ATP maintenance reaction ('NGAM_D1_h') was added to account for the metabolic cost of D1 protein degradation and biosynthesis. From data in ref. [34], I calculated the number of D1 protein inactivation events per molecule of O_2 evolved from PSII over a range of light intensities. The metabolic cost of D1 degradation included the ATP-cost of phosphorylation and activation of the FtsH protease [35], while the cost of D1 repair was represented by the ATP- and GTP-costs of biosynthesizing a D1 peptide (Supplementary Data Set B.3).

The photosynthesis-irradiance curves for net O_2 production ('EX_o2_e') and gross carbon uptake ('EX_co2_e + EX_hco3_e') were acquired from ref. [36] for *T. pseudonana* cells acclimated to continuous light at 60 μmol photons m^{-2} s^{-1}. An equation to represent the net oxygen production curve was derived with Platt fitting [37] including dark respiration (0.086 mmol O_2 (mg Chl a)$^{-1}$ h^{-1}).

$$EX\_o2\_e(t) = \left(0.279 \cdot (1 - e^{\frac{-0.0044 \cdot I}{0.279}}) - 0.086\right) \cdot \left(\frac{mg\ chl\ a/cell}{gDW/cell}\right) \tag{2.1}$$

The compensation light level (at which net O_2 production = 0) was calculated to be 23.3 μmol photons m^{-2} s^{-1}. Another equation was derived to represent carbon uptake.

$$EX\_co2\_e + EX\_hco3\_e(t) = 0.243 \cdot \left(1 - e^{\frac{-0.0021 \cdot I}{0.243}}\right) \cdot \left(\frac{mg\ chl\ a/cell}{gDW/cell}\right) \tag{2.2}$$

The ATP maintenance reaction lower and upper bounds (LB, UB) were calculated as,

$$LB_{ATPM\_c} = UB_{ATPM\_c} = Ia_{chl\,a}r\phi e \tag{2.3}$$

(equation 2 from ref. [38], Supplementary Data Set B.3), using a compensation light level of 23.3 μmol photons m^{-2} s^{-1} as the light intensity at which the rate of photosynthesis equals the rate of cellular respiration (I). The term $a_{chl\,a}$ is the chlorophyll a specific absorption

coefficient (m^2 mg Chl a^{-1}, [39]), r is the ratio of chlorophyll a to gram dry weight per cell (mg Chl a gDW^{-1}), ϕ is the quantum efficiency of photosynthesis (8 mol O_2 mol $photon^{-1}$), and e is the amount of ATP generated per oxygen (10 mol ATP mol O_2^{-1}). The results of this equation were converted into the flux units for *i*Tps1426 (mmol gDW^{-1} h^{-1}). See Supplementary Data Set B.3 for calculations and references.)

2.3.3 Constraints on alternative electron flow

Non-respiratory reactions that consume reducing equivalents or generate ATP can be important for balancing redox reactions, particularly if they carry large fluxes. Cyclic electron flow ('CEF_h') around PSI is an important mechanism of generating ATP in Viridiplantae ([40], see ❶ in Figure 2.1). But it was found to represent only a small fraction of total electron flow in *T. pseudonana* and other diatoms [12]. Additionally, cyclic electron flow appears to be independent of light intensity, so the bounds of this reaction were constrained to a narrow range of low flux values.

$$0.5 \leq CEF\_h \leq 1\ mmol\ gDW^{-1}h^{-1} \tag{2.4}$$

Nitrate (Figure 2.1, ❷) and sulfate (Figure 2.1, ❸) assimilation reactions are also localized in the plastid, consume significant reducing equivalents, and maintain relatively high flux because of cellular nutrient requirements. After cyclic electron flow and respiration, much of the remaining ATP/NADPH regulation was assumed to involve associated assimilation reactions. The lower bound of nutrient uptake rates for CO_2, HCO_3^-, NO_3^-, NH_4^+, urea, PO_4^{3-}, and SO_4^{2-} were constrained with V_{max} and K_m values from the literature for *T. pseudonana* or related organisms.

$$V = -\frac{V_{max}(mmol\ gDW^{-1}h^{-1}) \cdot [S](mM)}{K_m(mM) + [S](mM)} \tag{2.5}$$

$$LB_{EX\_no3\_e} = -\frac{1.16 \cdot 0.1}{0.004 + 0.1} = -1.11\ mmol\ gDW^{-1}h^{-1} \tag{2.6}$$

$$LB_{EX\_so4\_e} = -\frac{0.10 \cdot 28.8}{0.0014 + 28.8} = -0.097\ mmol\ gDW^{-1}h^{-1} \tag{2.7}$$

where [S] represents substrate concentrations for NO_3 and SO_4 in synthetic seawater media.

2.3.4 Constraints on light-dependent respiration

To further enhance my ability to simulate growth of *T. pseudonana*, I explored appropriate constraints on light dependent respiration, including the specificity of RuBisCO for CO_2, chlororespiration, the Mehler reaction, and energetic coupling between plastids and mitochondria. The respiration reactions include ribulose-1,5-bisphosphate oxygenase ('RUBISO_h'), plastid terminal oxidase ('PTOX_h'), the Mehler reaction (as a component of 'PSI_u'), alternative oxidase ('AOX_m'), and cytochrome *c* oxidase complex IV ('CYOO_m'). Together these reactions regulate the ATP/NADPH ratio in photosynthetic organisms (Figure 2.1, ❹-❾).

Photorespiration results from the oxygenase activity of RuBisCO that generates a 2-phosphoglycolate that cannot enter the Calvin Cycle (Figure 2.1, ❹). In diatoms, photorespiration is truncated and 2-phosphoglycolate is not recycled back to ribulose-1,5-bisphosphate [41]. The absence of a serine-pyruvate aminotransferase and a hydroxypyruvate reductase prevent the conversion of serine to glycerate. Instead, glycolate can be metabolized to malate in the peroxisome, or to serine, NH_4 and CO_2 in the mitochondria; serine can enter phosphoglycerolipid metabolism, cysteine and methionine metabolism, pyruvate metabolism, or possibly DHPS biosynthesis [15]. To activate photorespiration in my model, a constraint was added to reflect the oxygenase activity of RuBisCO. The constraint was estimated as follows. Flux through the oxygenase reaction is a fraction of the carboxylase reaction in RuBisCO:

$$RUBISO\_h = X \cdot RUBISC\_h \tag{2.8}$$

where

RUBISO_h = oxygenase activity of RuBisCO,

RUBISC_h = carboxylase activity of RuBisCO,

and X = the specificity factor (SF) of RuBisCO for CO_2 over O_2. I used a specificity factor (SF) of 79 for CO_2, as determined for the related diatom *T. weissfloggii* [42] because of predicted peptide level similarities between the RuBisCO enzymes from the two diatoms (rbcS is 97% identical, rbcL is 98% identical). The activity of carbon concentrating mech-

anisms in *T. pseudonana* has been modeled and the concentration of CO_2 in the pyrenoid, where most RuBisCO is located, is estimated [43] as 100 μM. Because the pyrenoid is an impermanent proteinaceous body, oxygen is predicted to interact with RuBisCO. However, the concentration of O_2 in the pyrenoid is unknown and difficult to measure [44]. I therefore used the ambient concentration of O_2 in seawater at equilibrium with the atmosphere (200 μM). I estimated X as,

$$X \sim \frac{200\,\mu M}{100\,\mu M} \div 79 = 0.025 \tag{2.9}$$

In the Mehler reaction, oxygen reacts with reduced ferredoxin emerging from photosystem I and forms a superoxide anion. Superoxide dismutase neutralizes two reactive anions into oxygen and hydrogen peroxide and ascorbate peroxidase converts hydrogen peroxide into water. NAD(P)H is consumed through the glutathione-ascorbate cycle (Figure 2.1, ❺). Ref. [45] measured light-dependent oxygen uptake over a range of light intensities by blocking PSII with the inhibitor DCMU in *Cylindrotheca fusiformis*. The DCMU inhibitor blocks the Mehler reaction and photorespiration (but not chlororespiration or mitochondrial respiration). Because photorespiration makes up a small fraction of total respiration, I estimated that oxygen uptake was approximately 20% of gross oxygen production and insensitive to changes in light intensity. I adjusted the rates of superoxide formation at photosystem I ('PSI_u') from ref. [31] to match the range of expected values.

Ref. [12] proposed that reductants from the plastid flow to the mitochondria based on their observations that impairment of mitochondrial respiration with SHAM and antimycin A or myxothiazol inhibitors suppressed flux through PSII. Ref. [13] incorporated these observations into their *P. tricornutum* model simulations and proposed that energetic coupling between the plastid and mitochondria (Figure 2.1, ❻) could be simulated with the constraint that a fraction (C) of photosystem I flux ('PSI_u') goes to NADH:ubiquinone oxidoreductase ('NADHOR_m') in the mitochondria,

$$NADHOR\_m = C \cdot PSI\_u \tag{2.10}$$

In ref. [13], chlororespiration was represented by a constraint, where a fraction (H) of cy-

tochrome b_6/f ('CBFC2_u') electron flow goes through plastid terminal oxidase ('PTOX_h', Figure 2.1, ❼).

$$PTOX\_h = H \cdot CBFC2\_u \tag{2.11}$$

I estimated the relative contribution of mitochondrial respiration and chlororespiration (given a fixed value of photorespiration, the Mehler reaction, and maintenance mitochondrial respiration: $LB_{sink\_13glucan\_c} = -1.02 \times 10^{-5}$ mmol gDW^{-1} h^{-1} [36]) by using *i*Tps1426 to simulate a range of energetic coupling ($C = 0 - 1$) and chlororespiration ($H = 0 - 1$) at 30 μmol photons m^{-2} s^{-1} (the light intensity at which the experiment in ref. [12] was conducted). I found that when 11% of photosystem I flux (PSI_u) is directed towards NADH:ubiquinone oxidoreductase in the mitochondria ($C = 0.11$), PSII_u flux matched measured rates of gross oxygen production, as expected. Therefore, C of 0.11 is assumed to be the correct parameter for energetic coupling between the plastid and mitochondria at low light intensities. When 35% of cytochrome b_6/f flux was diverted to the plastid terminal oxidase reaction ($H = 0.35$), there was a good fit between the relative impairment of mitochondrial respiration (Figure 2.1, AOX_m ❽, CYOO_m ❾) and suppression of PSII flux. Figure 2.2 shows experimental data for *P. tricornutum* and *T. pseudonana* from ref. [12] on the relative impact of flux through PSII when respiration is impaired with different concentrations of SHAM and other inhibitors. I show that when $H = 0.35$, any inhibition of mitochondrial respiration (AOX_m + CYOO_m) due to decreases in C, also impacts flux through PSII_u. Different levels of chlororespiration alter the relationship between mitochondrial respiration and PSII (Figure 2.2), where lower values decrease the y-intercept and produce a steeper slope over a range of C. At this light level, photorespiration makes up 0.6% of respiration, chlororespiration makes up 61.4%, the Mehler reaction makes up 17.5%, and mitochondrial respiration makes up 19.3%. Diatom respiration increases linearly with oxygen production across a range of light intensities [12, 36, 46]. In *T. pseudonana*, estimates of light-dependent respiration ranges from 4 to 12% of gross oxygen production for cells acclimated to different light levels [36]. In ref. [36], light dependent respiration was estimated to be 4% of gross photosynthesis in cells acclimated to 60 μmol photons m^{-2} s^{-1}. But the sensitivity of the relative amounts

of chlororespiration and energetic coupling to changes in light intensity remains unknown given that this experiment was only performed at 30 μmol photons m^{-2} s^{-1}.

2.3.5 Simulations of growth in batch culture

Dynamic flux balance analysis was used to simulate *T. pseudonana* growth in continuous light at 100 μmol photons m^{-2} s^{-1} in the synthetic seawater medium Aquil. The cultures were initialized with biomass concentrations of 5×10^{-5} gDW L^{-1}, or 3,846 cells mL^{-1} using the cell dry weight conversion from ref. [47] for cells grown at 100 μmol photons m^{-2} s^{-1} on continuous light harvested during logarithmic phase (13 pg $cell^{-1}$, Supplementary Data Set B.2). To mimic availability of chrysolaminarin and chitin as cellular energy sources, the media were initialized with 35.5 nM 1,3-glucan and 36.9 nM chitin which represents 23% and 15% [48, 49, 50] of the cellular biomass (gDW L^{-1}), respectively. High density cultures can become light-limited by self-shading when photon absorption capacity exceeds photon delivery [31]. I included self-shading as part of the simulation by dividing the culture into ten concentric slices. The light available in the first slice was used to calculate the resulting net oxygen production, carbon uptake rate, and photosystem II D1 damage rate in that slice. The light that was not absorbed was available for the next slice where it was used to recalculate those rates.

The constraints previously determined for photorespiration and the Mehler reaction were utilized in these simulations. Chlororespiration was not included as there is evidence that chlororespiration is less active at the simulated light conditions [51]. The value for energetic coupling was adjusted to $NADHOR\_m \geq 0.03\, PSI\_u$. At this value, the initial PSII flux was similar to the expected value for that light level [36]. In this simulation, photorespiration contributes 3.6-5.8% of oxygen consumption, the Mehler reaction contributes 68-72%, mitochondrial respiration contributes 21-22%, and chlororespiration contributes 0-2.9% but only when light limitation due to self-shading becomes important (Supplementary Figure B.1). All of the mitochondrial respiration is due to alternative oxidase activity and not cytochrome *c* oxidase complex IV, except for a very small fraction when self-shading is important. Low

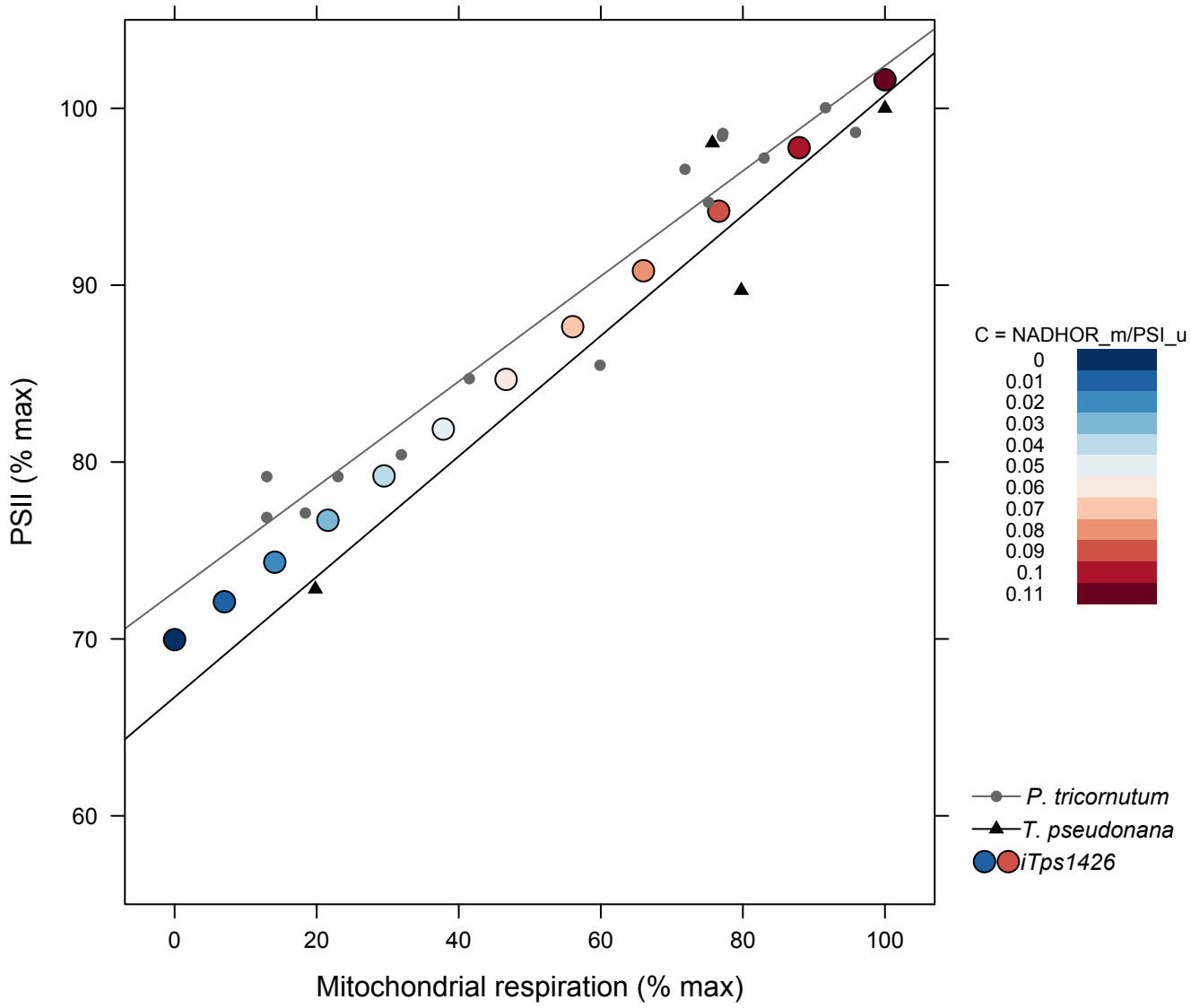


Figure 2.2: Relationship between mitochondrial respiration and photosynthetic activity at 30 μmol photons m^{-2} s^{-1}. Effect of SHAM and antimycin A or myxothiazol inhibitors on dark respiration and photosynthetic activity from ref. [12] in *P. tricornutum* (grey circles) and *T. pseudonana* (black triangles). Comparison of experimental data to simulated effect of inhibiting mitochondrial respiration by modifying the strength of energetic coupling between the plastid and the mitochondria when $H = 0.35$ (chlororespiration) and $X = 0.025$ (photorespiration). Red: higher levels of coupling, blue: lower level of coupling.

flux through cytochrome *c* oxidase may be attributed to the use of parsimonious flux balance analysis and the lack of constraints on different types of mitochondrial respiration. In *T. pseudonana*, the Mehler reaction should contribute to 60% of total oxygen uptake at higher light intensities [52]. Interestingly the contribution of the Mehler reaction is highly variable among different diatom species, and it is higher in *T. pseudonana* compared to other marine phytoplankton. Although the strength of energetic coupling was decreased for this light level so that the Mehler reaction and PSII flux could match expected values from the literature, the relative contribution of mitochondrial respiration remained the same as for 30 μmol photons m^{-2} s^{-1}.

Four simulations were performed with these respiratory constraints in continuous light in Aquil medium to test the effects of different sources of nitrogen as well as N- and P-limitation. Two simulations were performed in nutrient-replete Aquil with either NO_3 (100 μM NO_3, 10 μM PO_4) or NH_4 (100 μM NH_4, 10 μM PO_4) as the nitrogen source; under either condition, nitrogen depletion coupled with light limitation from self-shading ultimately limits cell division. In a third simulation, added NO_3 was reduced four-fold (25 μM NO_3, 10 μM PO_4) such that only NO_3 availability, and not self-shading, limited cell division. In the last simulation, added PO_4 was reduced ten-fold (100 μM NO_3, 1 μM PO_4) such that only PO_4 availability limited cell division.

In the first simulation, self-shading began to impact growth at $t = 64\,h$ (Figure 2.3A). During growth on NH_4 self-shading began even sooner and impacted growth at $t = 48\,h$ (Figure 2.3B). In the NO_3-replete simulation, growth was ultimately N-limited; NO_3 was drawn down from 100 μM to 0.003 nM, and PO_4 was drawn down from 10 μM to 4.7 μM by $t = 304\,h$ (Figure 2.3A). During growth on NH_4, NH_4 was also drawn down from 100 μM to 0.003 nM and PO_4 dropped from 10 μM to 4.7 μM by $t = 304\,h$ (Figure 2.3B). In NO_3-limited media, N-limitation began at $t = 72\,h$ when NO_3 was drawn down from 25 μM to zero; PO_4 remained high at 8.7 μM (Figure 2.3C). In PO_4-limited media, P-limitation began at $t = 64\,h$ when PO_4 was drawn down from 1 μM to zero. Under these conditions, NO_3 continued to be drawn down to zero despite the lack of cell division (Figure 2.3D).

2.3.6 Macronutrient uptake rates

In NO_3-replete media, the nitrate uptake rate gradually declined over the course of the simulation as NO_3 was drawn down (Figure 2.3A). In NH_4-replete media, sulfate uptake rates were slightly higher than in NO_3-replete media (2.3 μmol gDW^{-1} h^{-1} vs 1.8 μmol gDW^{-1} h^{-1} at $t = 0$). During growth on NH_4, the GS-GOGAT cycle played a role in balancing redox reactions, but there was no flux through nitrate and nitrite reductase reactions and so sulfate uptake played a greater role in dissipating reductants. In NO_3-limited media, the nitrate uptake rate declined rapidly early on, completely limiting biomass production (Figure 2.3C). There was a small dip in sulfate uptake for all conditions as nitrate was depleted and *i*Tps1426 began to dissipate excess reducing power by storing chrysolaminarin (as the monomer 1,3-glucan). As the upper limit for chrysolaminarin storage is reached due to limits on biomass production, there was a boost in sulfate uptake rates (Figure 2.3A, 2.3B, 2.3C). In PO_4-limited media, the onset of P-limitation resulted in an increase in nitrate uptake, followed by an increase in sulfate uptake after nitrate was depleted (Figure 2.3D). Together, these results suggest that *T. pseudonana* uses inorganic nutrients to help correct cellular redox imbalances due to energy inputs from light and growth on more reduced sources of nitrogen such as NH_4. The different assimilation pathways have varying ATP requirements and NADPH dissipation capacities. CO_2 uptake requires 3 ATP to oxidize 2 NADPH. NO_3 uptake requires 1 ATP to consume 6 reducing equivalents. The energy requirements of SO_4 are more flexible; SO_4 uptake to cysteine (H_2S) requires 2 ATP to consume 7 reducing equivalents, while SO_4 uptake to cysteate (PAPS) requires 2 ATP to consume 1 NADPH, and SO_4 uptake to UDP-sulfoquinovose (SO_3) require 2 ATP to consume 1 reducing equivalent. Utilization of CO_2, NO_3, or SO_4 to dissipate reductants depends on both ATP availability and maximum possible rate of nutrient uptake. NO_3 uptake is the preferred mechanism of dissipating reductants as it is less energy intensive. After NO_3 is depleted or during growth on more reduced forms of nitrogen, the role of SO_4 is amplified.

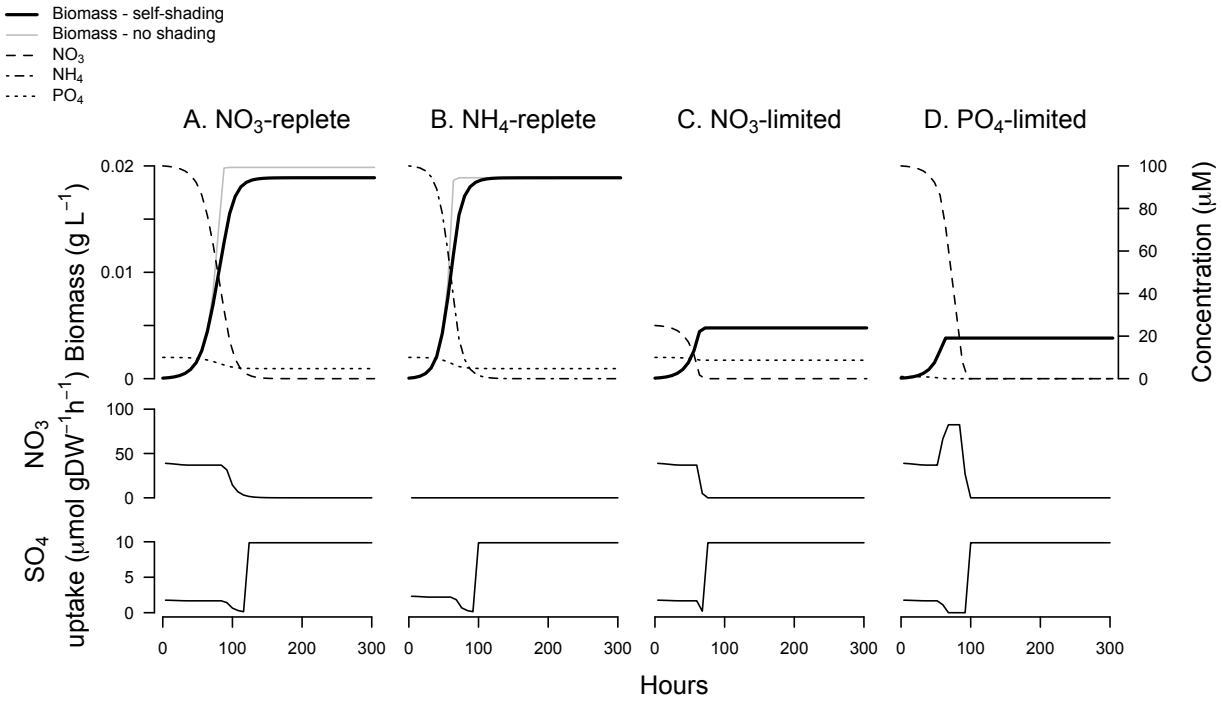


Figure 2.3: Simulated growth curves of *T. pseudonana* in batch culture at 100 μmol photons m^{-2} s^{-1} in nutrient-replete Aquil (A), nutrient-replete modified Aquil where NO_3 was replaced with NH_4 (B), NO_3-limited Aquil (C), and PO_4-limited Aquil (D). Biomass concentration that includes the effects of light-limitation due to self-shading is plotted with a thick black line, while biomass concentration that does not include self-shading is plotted with a thinner grey line. Change in concentration of NO_3, NH_4, and PO_4 (dashed, dot-dashed, and dotted lines) were plotted alongside the growth curves in μM. NO_3 and SO_4 uptake rates for the first (unshaded) slice of the cultures were plotted in μmol gDW^{-1} h^{-1}.

2.3.7 Metabolite secretion

The simulations predicted different patterns of metabolite excretion as nutrient conditions changed. An inherent feature of these simulations is that all assimilated nutrients contribute to either biomass production or metabolite secretion, while nitrate or sulfate uptake serve the additional role of balancing reductants through oxidation of reducing equivalents in the plastid (Figure 2.1). Biomass composition is invariable in the model, so any deviation from the stoichiometry of biomass results in metabolite secretion. I illustrate this point by comparing metabolic flux solutions from the NO_3-replete simulation (Figure 2.3A) at two different time points which represent exponential phase ($t = 0\,h$, Figure 2.4A) and early stationary phase ($t = 88\,h$, Figure 2.4B). Fluxes through the glycolysis (see ① in Figure 2.4), the TCA cycle ②, nitrate assimilation ③, sulfate assimilation ④, and the glycine cleavage system ⑤ were higher during exponential growth (Figure 2.4A). In the early stationary phase, there was relatively more flux through gluconeogenesis, the urea cycle ⑥, and carbon fixation ⑦ (Figure 2.4B). There was less available nutrients during early stationary, so biomass production was lower. Consequentially, *i*Tps1426 balanced reductants by fixing more CO_2; excess carbon was secreted as urea ⑥, glycolate ⑧ and formate ⑨ or stored as chrysolaminarin ⑩.

The overall patterns of metabolite secretion are a response to nutrient availability and redox imbalance in the cell (Figure 2.5). Two metabolites, cyanide and thiocyanate, were secreted at low flux throughout all four simulations. Cyanide ions were secreted as a byproduct of cyanocobalamin utilization to meet cellular vitamin B_{12} requirements. At the onset of nutrient limitation, cyanide is taken back up by *i*Tps1426 and transformed into thiocyanate via the enzyme 3-mercaptopyruvate sulfurtransferase, which is known to play a role in cyanide detoxification. Thiocyanate was subsequently released by cells rather than cyanide. Cyanocobalamin is frequently used as a source of B_{12} in culture work, though other forms of B_{12} predominate in nature [53].

Carbohydrates were secreted or stored as a mechanism of dealing with the CO_2 or HCO_3 fixed through the Calvin cycle when growth was limited by nutrient availability and excess

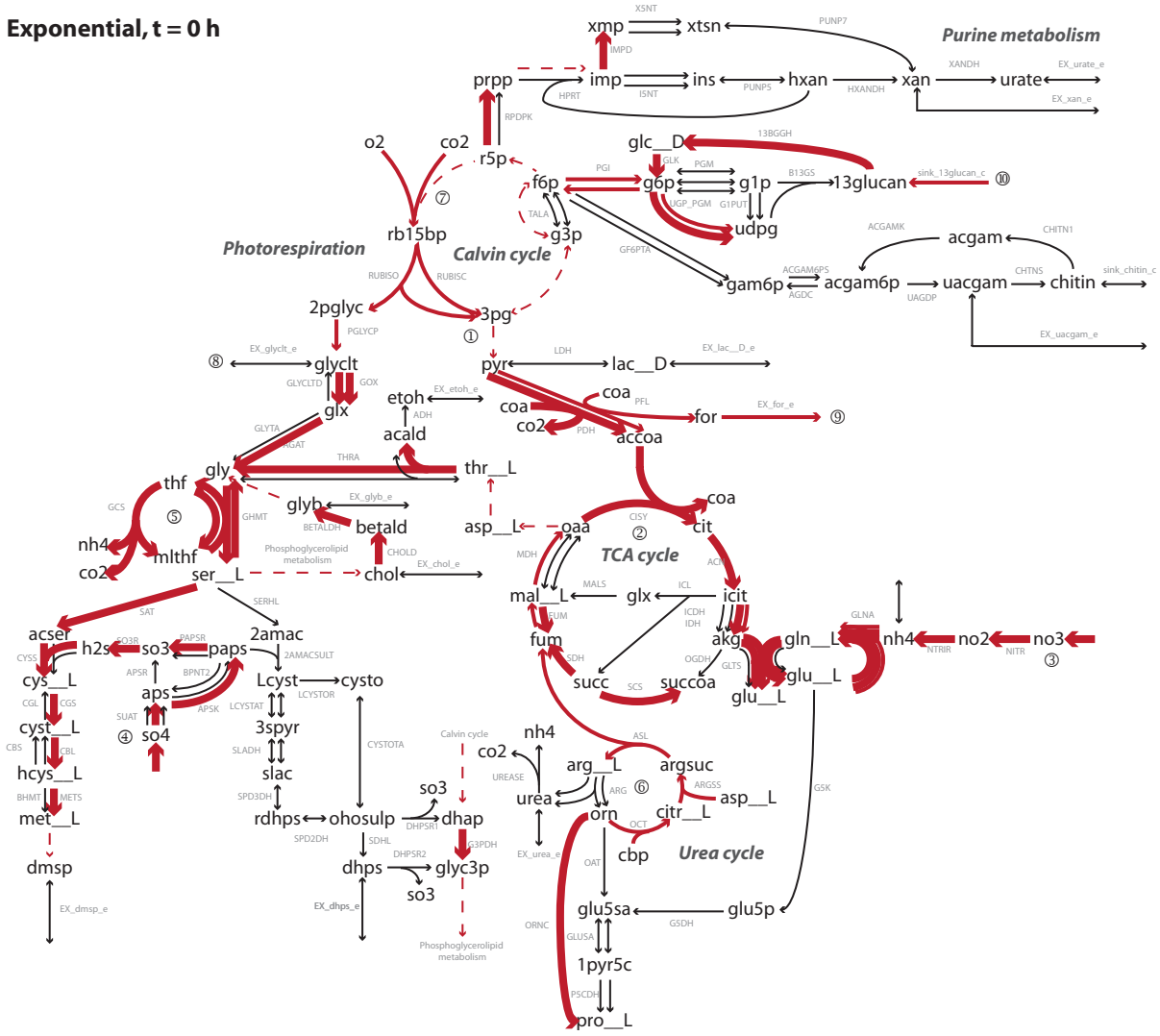

Exponential, t = 0 h
Purine metabolism
Photorespiration
Calvin cycle
TCA cycle
Urea cycle
xmp
xtsn
prpp
imp
ins
hxan
xan
urate
o2
co2
r5p
glc__D
f6p
g6p
g1p
13glucan
udpg
rb15bp
g3p
gam6p
acgam6p
acgam
uacgam
chitin
2pglyc
3pg
glyclt
pyr
lac__D
etoh
acald
coa
for
accoa
glx
gly
thr__L
thf
glyb
betald
asp__L
oaa
cit
nh4
co2
mlthf
ser__L
chol
mal__L
icit
akg
fum
succ
succoa
gln__L
glu__L
nh4
no2
no3
acser
h2s
so3
paps
2amac
cys__L
aps
so4
Lcyst
cysto
cyst__L
3spyr
hcys__L
slac
met__L
rdhps
ohosulp
dhap
dmsp
dhps
glyc3p
so3
nh4
co2
urea
arg__L
argsuc
asp__L
orn
citr__L
cbp
glu5sa
glu5p
1pyr5c
pro__L

Figure 2.4

Early stationary, t = 88 h

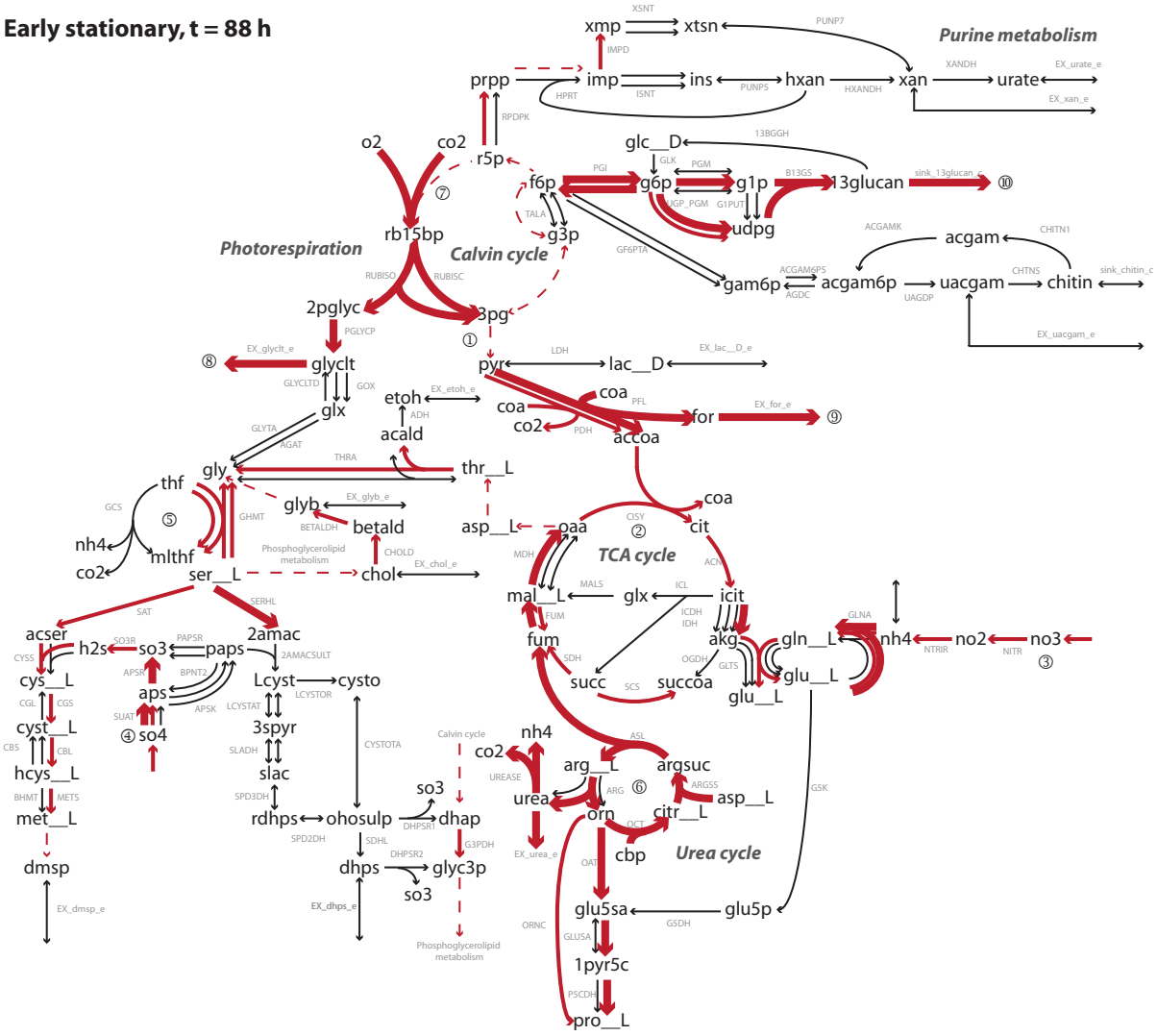


Figure 2.4: (Previous page.) Metabolic maps comparing the flux distributions in NO_3-replete Aquil at 100 μmol photons m^{-2} s^{-1} during exponential phase ($t = 0\,h$) and early stationary phase ($t = 88\,h$). These flux distributions show the results of a single FBA slice at each time point that is not subject to self-shading. Dashed arrows indicate that a metabolic pathway has been abbreviated, black arrows indicate that the reaction carries no flux, red arrows indicate that the reaction carries flux, the thicker arrows carry a higher flux, while the thinner arrows carry a lower flux when comparing the time points. The corresponding metabolite or reaction names for each abbreviation can be found in Supplementary Data Set B.1 or at http://bigg.ucsd.edu. Symbols refer to metabolic pathways and reactions referenced in the text: glycolysis/gluconeogenesis ①, TCA cycle ②, nitrate assimilation ③, sulfate assimilation ④, glycine cleavage system ⑤, urea cycle ⑥, carbon fixation ⑦, glycolate excretion ⑧, formate excretion ⑨, chrysolaminarin storage ⑩.

ATP was available. The chrysolaminarin monomer 1,3-glucan began to be stored at the onset of N-limitation in all four simulations (Figure 2.5). Chrysolaminarin contributes to mitochondrial maintenance respiration [36] during the early growth phase when it is hydrolyzed to glucose and enters glycolysis. 1,3-Glucan can be re-synthesized through gluconeogenesis as a mechanism of storing excess reducing power, but there is an upper limit on chrysolaminarin storage as biomass production stalls.

Formate was secreted at the beginning of all simulations where *i*Tps1426 grew on nitrate (Figure 2.5A, 2.5C, 2.5D). It is a byproduct of the methionine salvage pathway or DMSP biosynthesis, and it is a byproduct of pyruvate formate lyase, an alternative reaction to pyruvate dehydrogenase. Formate is involved in many different pathways including: folate, thiamine, riboflavin, and steroid biosynthesis. Glycolate, a more reduced carbohydrate (Table 2.3), was secreted at the beginning of the NH_4-replete simulation (Figure 2.5B). Glycolate is produced due to the inherent oxygenase activity of RuBisCO. Glycolate can be utilized by the cell through conversion to glyoxylate and then conversion to malate by malate synthase or conversion to glycine by glycine transaminase or alanine-glyoxylate aminotrans-

ferase. Both the glycine cleavage system and the TCA cycle are fueled by the availability of NAD^+; the glycine cleavage system is additionally fueled by the availability of nitrogen to transform glyoxylate into glycine. There was a much higher increase in secretion of the one- and two-carbon molecules, formate and glycolate, as biomass production began to stall, and their production was further amplified by nitrate depletion. As illustrated in my simulations, glycolate was released into the media during NO_3-limiting conditions (including the NO_3-depletion that occurs after P-limitation). I hypothesize that glycolate secretion occurs when flux through the TCA cycle or the glycine cleavage system is insufficient to keep up with the pace of RuBisCO.

Three other carbohydrates were secreted after chrysolaminarin storage stalled. Metabolism of carbohydrates and amino acids results in production and release of different two- and three-carbon molecules under growth-limiting conditions. Acetaldehyde is a byproduct of glycine, serine and threonine metabolism due to the activity of threonine aldolase. The resulting acetaldehyde is subsequently transformed into ethanol by alcohol dehydrogenase or acetate by aldehyde dehydrogenase. The secretion of ethanol is therefore related to glycine production from threonine rather than, or in addition to, glyoxylate. Acetate is involved in acetyl-CoA synthesis, is a byproduct of cysteine synthase, and is involved in amino sugar metabolism. Lactate is a fermentation product of pyruvate produced by the enzyme lactate dehydrogenase when there is insufficient oxygen in the mitochondria for complete glycolysis.

The volume and character of secreted organic nitrogen depends on cause of growth limitation. In NO_3, NH_4-replete, and NO_3-limited conditions there are small fluxes of urea secreted as light-limitation due to self-shading inhibits growth and excess nitrogen is released into the media. As those cultures becomes NO_3 or NH_4 depleted, urea is re-assimilated resulting in the production of L-valine, which itself re-assimilated at the next time point resulting in the production of L-aspartate, folate, and chitin. In the NO_3-limited simulation those fluxes halt once biomass production is inhibited. In the PO_4-limited simulation, higher fluxes of different forms of organic nitrogen are secreted as a result of boosted nitrate uptake rates after growth is limited (Figure 2.3D). Initially, there is secretion of urea, xanthine, folate,

L-asparagine, and L-aspartate. Re-uptake of urea, L-asparagine, and L-aspartate results in secretion of L-valine, L-isoleucine, and chitin synthesis (Figure 2.5D). Urea is the most reduced form of organic nitrogen produced by the model, followed by xanthine, L-asparagine, L-valine, L-isoleucine, folate, chitin, and L-aspartate (Table 2.3). Re-uptake of more highly reduced nitrogen compounds steepens the redox imbalance in the cell. As urea is re-absorbed and as nitrate becomes limiting, other forms of organic nitrogen are secreted due to the high degree of reduction in urea, creating a cascade of nitrogenous product secretion with varying degrees of reduction over the course of the growth curve. When biomass production instead becomes limited by lack of phosphate, the role of nitrate uptake becomes more crucial to balance redox reactions in the plastid, and excess nitrogen is excreted as organic nitrogen to balance these increased rates of nitrate assimilation (Figure 2.5D).

As *i*Tps1426 consumes either organic nitrogen with high degrees of reduction, or as nitrate becomes limiting, sulfate uptake replaces nitrate assimilation as a primary means of consuming reducing equivalents in the plastid. Sulfate assimilation has a higher ATP cost than nitrate uptake. But there is ATP to spare as biomass production becomes inhibited through nutrient limitation. The organic sulfur compounds DMSP and DHPS are secreted in all conditions once nitrate is depleted and chrysolaminarin storage reaches its maximum. DMSP is a byproduct of the amino acid methionine which otherwise would contribute to biomass production. Ref. [15] postulated that there are two possible routes for DHPS biosynthesis, via cysteinolic acid or 3-sulfopyruvate. In the simulations, DHPS was synthesized via cysteinolic acid, likely due to the use of parsimonious flux balance analysis which minimizes the absolute sum of fluxes. DHPS was respired to glycerol-3-phosphate in self-shaded slices below the compensation level. The organic sulfur compound DMSP has a higher degree of reduction than DHPS (Table 2.3) and is secreted to higher concentrations. Most of the organic sulfur is produced in culture slices that are not subject to light-limitation due to self-shading in NO_3- and NH_4-replete simulations, while it is produced equally in all slices in the nutrient limited cultures which are more exposed to light throughout (Supplementary Figure B.2).

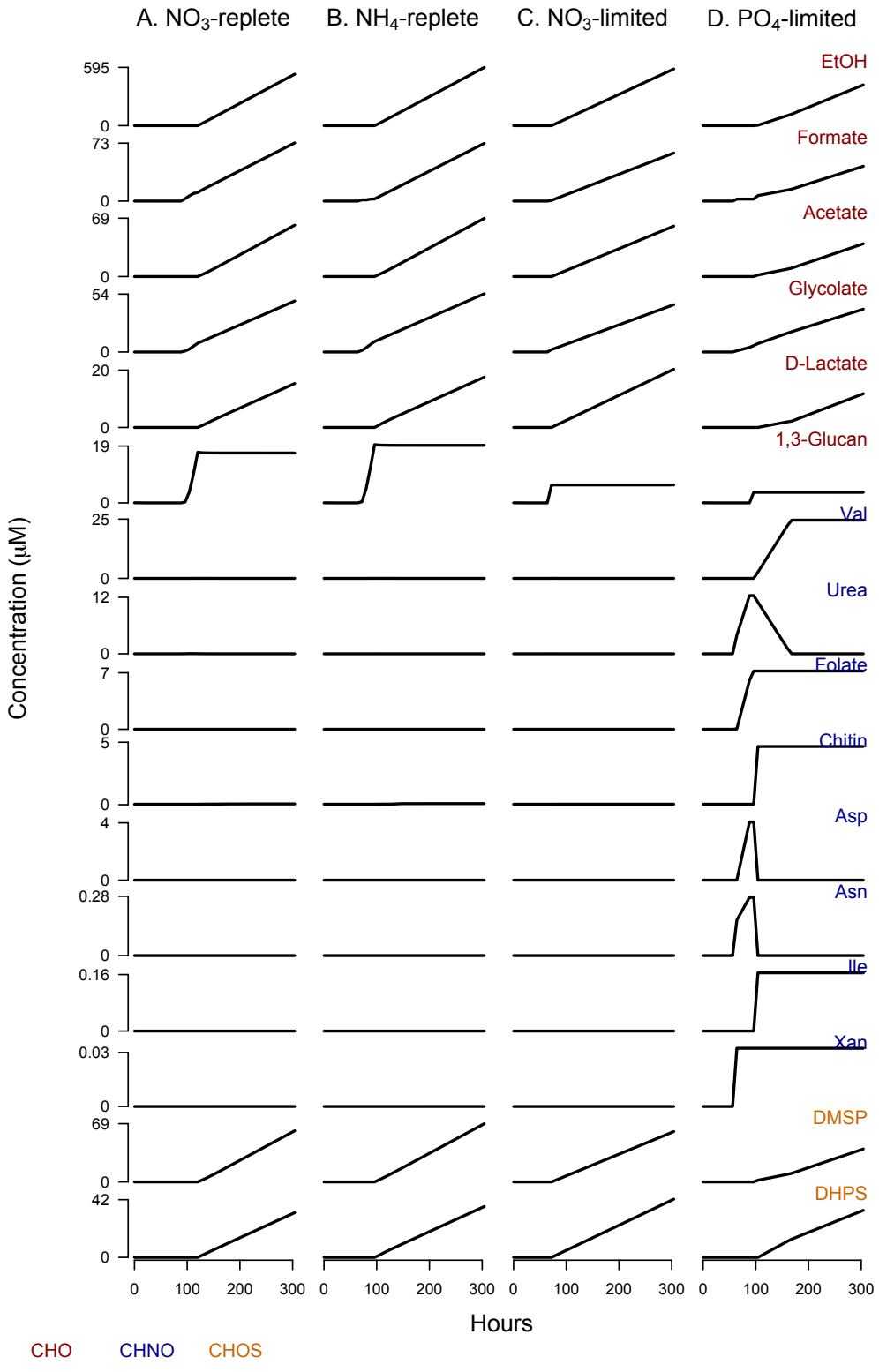

A. NO3-replete
B. NH4-replete
C. NO3-limited
D. PO4-limited
EtOH
Formate
Acetate
Glycolate
D-Lactate
1,3-Glucan
Val
Urea
Folate
Chitin
Asp
Asn
Ile
Xan
DMSP
DHPS
595
73
69
54
20
19
25
12
7
5
4
0.28
0.16
0.03
42
0
Concentration (μM)
0
100
200
300
Hours
CHO
CHNO
CHOS

Figure 2.5

Figure 2.5: Simulated concentrations of metabolites produced over the course of growth experiments at 100 μmol photons m^{-2} s^{-1} in nutrient-replete Aquil (A), nutrient-replete modified Aquil where NO_3 was replaced with NH_4 (B), NO_3-limited Aquil (C), and PO_4-limited Aquil (D). Organic carbon compounds are labeled in red, organic nitrogen compounds are labeled in blue, and organic sulfur compounds are labeled in orange. Secreted metabolites whose change in concentration was less than 1 nM were not plotted.

2.4 Discussion

Thalassiosira pseudonana CCMP 1335 was isolated from Moriches Bay, New York, in 1958, and the whole genome was sequenced in 2004 [54]. Since the genome became available, *T. pseudonana* has been studied from a systems-wide perspective using transcriptomics, proteomics, and metabolomics (eg. [55, 56, 57, 58, 59]). The genome-scale metabolic model of *T. pseudonana* created here incorporates currently available physiological and genomic data and will serve as a powerful tool to generate hypotheses about diatom metabolism and interpret molecular and physiological data.

In this study, I identified and incorporated appropriate constraints on photosynthesis, respiration, cyclic electron flow, non-growth associated ATP maintenance, and nutrient uptake based on a synthesis of available data. It was assumed that photosynthesis and respiration rates are impacted by light intensity only, rather than nutrient limitation or nutrient source as there is little available data to constrain nutrient impacts. Under these assumptions, nitrate and sulfate assimilation were shown to play important roles in balancing redox reactions in the plastid. Nitrate uptake is more effective at dissipating reductants than sulfate uptake because of a higher V_{max} and reduced consumption of ATP relative to NADPH equivalents. Importantly, nitrate is typically limiting both in coastal regions [60] and in the subtropical gyres [61] whereas sulfate is consistently available at a high concentration (28 mM, [62]) throughout the ocean. In my simulations, I used the kinetics of sulfate uptake from maize

root due to the lack of available data on diatoms. Marine phytoplankton may benefit from a higher V_{max} for sulfate than maize roots given the nitrate-poor and a sulfate-rich ocean environment. A higher V_{max} for sulfate in marine phytoplankton could further enhance the relative role of sulfate assimilation in redox balance. Although my focus was on growth limitation by either nitrogen or phosphorus, other types of nutrient-limitation, or photoinhibitory light levels may similarly impact redox balance and subsequent release of organic compounds.

Rates of oxygen production and respiration are typically measured in nutrient-replete steady state conditions across a range of light levels [36]. Yet, phytoplankton may also adjust respiration rates under different nutrient conditions. For example, the freshwater green alga *Selenastrum minutum* decreases oxygen production and respiration rates when NH_4 is added during N-limited conditions [63]. Additionally, alternative oxidase expression is upregulated in the diatom *P. tricornutum* under stressful conditions including high light intensity, iron or nitrate limitation, or supraoptimal temperatures, indicating increased rates of mitochondrial respiration [64]. If the cell adjusts respiration rates in response to changing redox pressures, then nitrate and sulfate uptake may contribute little to redox balance. But nitrogen-replete diatoms have also been observed to secrete ammonium during rapid increases in irradiance [65], indicating a role for nitrate uptake and secretion of nitrogenous compounds in response to fluctuating redox pressures.

A range of one- to three-carbon molecules were predicted to be secreted by *i*Tps1426 under different nutrient conditions. Glycolate production by diatoms has been documented as a by-product of photorespiration [66, 67]; *i*Tps1426 accurately predicted secretion of this compound once the photorespiration constraint was added. Secretion of ethanol, acetate, formate, and D-lactate by diatoms is suggestive of some level of fermentation, which has been less well documented. Anaerobic bacterial cultures are known to accumulate ethanol, acetate, and formate, and metabolic models of *E. coli* quantitively predict their secretion [17]. In *E. coli*, ethanol is typically produced by alcohol dehydrogenase that detoxifies acetaldehyde produced by pyruvate decarboxylase during fermentation. *T. pseudonana* lacks

the gene for pyruvate decarboxylase, and instead produces acetaldehyde as a by-product of threonine aldolase. In *i*Tps1426, acetaldehyde is either converted into ethanol and excreted or converted into acetate which may rejoin the TCA cycle as acetyl-CoA. Similar secretion of these fermentation products has been observed in cyanobacteria during dark and anaerobic conditions. Strains that produced large quantities of glycolate and formate typically did not produce ethanol, lactate, and acetate and vice versa [68]. Interestingly, I observed that ethanol, acetate, and lactate production tended to occur only after chrysolaminarin storage stalled. D-Lactate accumulation has been observed in *Chlamydomonas reinhardtii* subjected to anaerobic conditions, although only trace amounts were excreted – most of the lactate was later re-oxidized [69]. Such accumulation of D-lactate has been linked to inhibition of mitochondrial respiration [70]. I stimulated mitochondrial respiration with the constraint $NADHOR\_m \geq 0.03\, PSI\_u$ but the ratio of these fluxes did not increase beyond 0.03 as nitrate was depleted, causing D-lactate production. In parsimonious flux balance analysis, the secondary objective is to minimize the absolute sum of fluxes; D-lactate secretion or accumulation could be a more optimal solution than an increase in energetic coupling to maintain PSII flux. These simulations appear to predict anaerobic metabolism as nitrate is depleted because there is no dynamic adjustment of respiratory constraints. I hypothesize that secretion of fermentative products is more likely in nature when respiration rates do not increase rapidly enough to match increased redox pressure, perhaps during fluctuating light conditions. Given that flux balance analysis cannot differentiate between molecule storage and molecule excretion, measures of these compounds in culture would indicate whether *T. pseudonana* re-oxides any produced D-lactate in a manner similar to *C. reinhardtii* or whether these compounds are secreted and available for bacterial consumption.

I accurately predicted chrysolaminarin accumulation both in conditions of N-limitation and extreme P-limitation, as shown experimentally by ref. [71]. I imposed an upper limit on chrysolaminarin storage (23% of biomass, [48]). Once the storage capacity was reached, sulfate uptake and excretion of organosulfur compounds became more important. The storage or excretion of high molecular weight carbohydrates has the potential to dissipate large

amounts of energy and reductants. Enhanced secretion of exopolysaccharides (EPS) has been observed during periods of increased carbon assimilation, growth on NH_4, nutrient depletion, especially phosphorus, and growth inhibition [72, 73, 74]. The upper limit on EPS secretion is likely higher than chrysolaminarin storage as it is not constrained by cell volume and could be an important regulator of redox balance. A pathway for EPS synthesis and secretion was beyond the scope of the model but would be interesting to study in future iterations of *i*Tps1426.

Amino acid secretion has been observed in a variety of different diatom species in nitrate-replete conditions [75]. Approximately 5% of the extracellular carbon excreted by *Skeletonema costatum* was as amino acids and the relative composition of amino acids changed from dawn to dusk, and exponential to stationary phase [21]. In *i*Tps1426, only transport reactions for amino acids which are known to be abundantly excreted by diatoms were included [76]. If putative transporter genes could be identified for those reactions, then transport reactions for all chemically similar amino acids which were transported by those genes were also added. This resulted in secretion of L-valine, L-isoleucine, L-aspartate, L-asparagine by *i*Tps1426. A study on the extracellular metabolome of *T. pseudonana* found that small peptides were an important component of secreted metabolites – possibly as a by-product of protein turnover [59].

I predicted that the organic sulfur compounds dimethylsulfoniopropionate (DMSP) and 2,3-dihydroxypropane-1-sulfonate (DHPS) were excreted during conditions of prolonged nitrogen limitation including growth on more reduced forms of nitrogen. It has been demonstrated in *T. pseudonana* that nitrate limitation causes the greatest increase in DMSP production and phosphate causes a smaller increase in intracellular DMSP [77]. It has been suggested that DMSP may replace stores of proline, another osmolyte, in nitrate-limiting conditions [77]. I observed that urea and amino acid consumption in the PO_4-limited simulation further stimulated DHPS excretion. DHPS is a newly discovered organosulfur compound that has a lower degree of reduction than DMSP (Table 2.3). I included a hypothetical DHPS biosynthesis pathway in the *i*Tps1426 using information from ref. [15]. In this model, DHPS

could be produced from a serine precursor and respired to glycerol-3-phosphate. DMSP and DHPS have been measured at high concentrations in the cytosol [15], and there may be turnover in the relative composition of different osmolytes under different conditions rather than, or in addition to, excretion.

Many of the compounds produced and excreted by *i*Tps1426 are known compatible solutes. In part, this is because osmolytes are a class of compounds that are known to be transported in and out of the cell – and there is an otherwise limited knowledge of which compounds diatoms secrete [59]. In live organisms, osmolyte excretion is influenced by osmotic pressure. The simulation results could be demonstrating changes in the relative composition of osmolytes inside the cell rather than or in addition to excretion patterns. Future research could attempt to model more dynamic changes in biomass composition, particularly changes in the relative abundance of different osmolytes.

*i*Tps1426 could be extended in a variety of directions. The biomass objective function could be extended to include trace metal requirements and the details of metal-requiring protein biosynthesis, as trace metal limitation is an important nutrient condition in the ocean. A model of silica metabolism has not yet been developed for any metabolic model and would greatly improve *i*Tps1426 as well as *i*LB1027_lipid. Silica metabolic modeling requires information on the specific long-chain polyamine composition of the *T. pseudonana* frustule, their associated biosynthesis reactions, the amount of silica precipitated per polyamine, and ideally identification of the genes associated with each reaction. A model of EPS synthesis and secretion could also be an important feature to develop in the study of redox metabolism. An effort to better characterize transporters between organelles and the cytosol as well as extracellular transporters would significantly improve predictions of metabolite secretion and mechanisms of energetic coupling between the plastid and the mitochondria.

This model of overflow metabolism as a means of dissipating reductants simulated production of ecologically important metabolites and generated hypotheses for when these compounds may be excreted. Notably, many of the metabolites secreted by *i*Tps1426 can be consumed by marine bacteria. For example, a subset of marine bacteria can utilize glycolate

Metabolite	Formula	Charge	Degree of reduction (wrt. to NO_3^-)
O_2	O_2	0	-4
Pi	HO_4P	-2	0
SO_4^{2-}	O_4S	-2	0
HCO_3^-	CHO_3	-1	0
NO_3^-	NO_3	-1	0
CO_2	CO_2	0	0
Magnesium	Mg	2	0
Formate	CHO_2	-1	2
Glycolate	$C_2H_3O_3$	-1	3
Acetate	$C_2H_3O_2$	-1	4
D-Lactate	$C_3H_5O_3$	-1	4
1,3-Glucan	$C_{12}H_{20}O_{10}$	0	4
L-Aspartate	$C_4H_6NO_4$	-1	5
Chitin	$C_8H_{13}NO_5$	0	5
DHPS	$C_3H_7O_5S$	-1	5.33
Biomass	$C_{30.5}H_{48.5}N_{5.3}O_{11.0}P_{0.28}S_{0.24}Mg_{0.04}$	-0.49	5.85
Ethanol	C_2H_6O	0	6
Cyanocobalamin	$C_{63}H_{88}N_{14}O_{14}PCo$	0	6.19
Folate	$C_{19}H_{17}N_7O_6$	-2	6.2
L-Isoleucine	$C_6H_{13}NO_2$	0	6.33
L-Valine	$C_5H_{11}NO_2$	0	6.4
DMSP	$C_5H_{10}O_2S$	0	6.4
Biotin	$C_{10}H_{15}N_2O_3S$	-1	6.6
L-Asparagine	$C_4H_8N_2O_3$	0	7
Thiamine	$C_{12}H_{17}N_4OS$	1	7.5
Xanthine	$C_5H_4N_4O_2$	0	8
NH_4^+	H_4N	1	8
Cyanide	CN	-1	10
Thiocyanate	CNS	-1	16
Urea	CH_4N_2O	0	16

Table 2.3: Degree of reduction of exchanged metabolites

as a sole carbon source [78], and transcripts for glycolate oxidase were found to vary on a diel cycle during a phytoplankton bloom [79]. Some marine bacteria from the Roseobacter clade can utilize acetate as a sole carbon source [80]. Many roseobacteria are, however, unable to reduce nitrate or nitrite and some cannot even oxidize sulfate [81]. Instead they rely on organic nitrogen and sulfur compounds produced by phytoplankton or other bacteria. Alphaproteobacteria and Gammaproteobacteria are known to degrade DMSP into methanethiol (CH_4S) or dimethyl sulfide (DMS) [82]. Ref. [6] found that *T. pseudonana* secretes DHPS that supports growth of the alphaproteobacterium *Ruegeria pomeroyi* during an interaction where the bacteria provided vitamin B_{12} to the diatom. This work demonstrates that environmental conditions impact the physiology of diatoms, which in turn can impact the types of metabolites secreted. Small differences in the metabolic networks of different types of phytoplankton, as well as differences in respiration, and how different species react to fluctuations in environmental conditions and redox imbalances can impact the character and quantity of metabolites secreted. These are likely major factors that structure the bacterial community associated with phytoplankton and control bacterial succession over the course of a bloom [83].

2.5 Materials and methods

2.5.1 Network reconstruction and curation

A genome-scale metabolic model of *Thalassiosira pseudonana* CCMP 1335 was generated using *i*LB1027_lipid (the model of *Phaeodactylum tricornutum* CCAP 1055/1, [13]) as a starting point, based on similarities between the two organisms. The *T. pseudonana* proteome was acquired from a dataset produced by ref. [84], in which open reading frames (ORFs) were improved by ensuring that each gene starts with 'ATG', encodes an uninterrupted reading frame that ends in a stop codon, is less than 10 kb in length, and has EST support. The ORFs used in a *T. pseudonana* network reconstruction for BioCyc were retrieved as well [85]; these proteins were re-annotated in 2012 using the JGI annotation pipeline for

eukaryotes [86]. The *P. tricornutum* chromosomal proteome was downloaded from EnsemblProtists (ASM15095v2), the plastid and mitochondrial proteomes were downloaded from NCBI (acc no.: NC_008588.1, HQ840789.1). A gene ID conversion table was provided by the authors of *i*LB1027_lipid and was used to update the gene IDs in the published model. OrthoMCL [26] was used to identify gene orthologs of *P. tricornutum* and the two sets of *T. pseudonana* ORFs. A network of *T. pseudonana* reactions was generated by retaining reactions in *i*LB1027_lipid that contained gene orthologs from *T. pseudonana* and deleting reactions with no gene ortholog. Spontaneous reactions in *i*LB1027_lipid were preserved in the *T. pseudonana* model. Protein localization predictions were performed on both *T. pseudonana* ORF sets (see Subcellular Protein Localization, below). BiGG IDs [87] were used for reactions and metabolites in the *T. pseudonana* network, and these were assigned to six different compartments: cytosol ('c'), mitochondria ('m'), peroxisome ('x'), chloroplast ('h'), thylakoid lumen ('u'), endoplasmic reticulum ('r'), or extracellular ('e').

The guidelines from an established genome-scale reconstruction protocol [88] were implemented to refine the *T. pseudonana* model. All genes from ref. [84] with orthologs in the reconstruction, all genes assigned to a reaction in BioCyc [85], and all *T. pseudonana* genes without an ortholog in *P. tricornutum* were annotated with InterProScan [89] and those annotations were used to verify gene-protein-reaction associations, and to detect missing genes, reactions, and pathways in the model. KEGG [90] and BioCyc [85] databases were used as an aid in model curation to make comparisons between organisms. The literature on *T. pseudonana* was examined for experimental evidence on the existence of different reactions and for protein localization data (see references and notes in Supplementary Data Set B.1). When experimental protein localization data was available, it superseded the subcellular localization prediction. For each reaction, mass and charge balance were verified and links to external databases were added for each reaction and metabolite.

Dead-end metabolites were identified as metabolites present only in blocked reactions; blocked reactions cannot carry flux due to reactions missing in the network and dead-end metabolites cannot be produced by the model as they are only part of blocked reactions.

These reactions were identified with the flux analysis module in COBRApy and the gapfilling module was used to identify gaps in the network [91]. Gaps were filled if a gene for the missing reaction could be identified, if there is physiological evidence that the reaction exists, if the majority of the pathway is otherwise present in the model, or if the reaction was required to produce biomass. In *T. pseudonana*, I first checked whether the reaction was present in another compartment and if there was any evidence that the subcellular localization prediction was too stringent (e.g., a low confidence prediction may be more likely based on the localization of other reactions in the pathway), if there is the possibility of dual-targeting, or if a different gene with the correct localization for the reaction could be identified. Occasionally, the JGI ORFs (rather than the ORFs from ref. [84]) provided the missing gene for a reaction, or a more likely subcellular localization prediction. If there was no evidence that a protein in the model was incorrectly targeted, then transport reactions between compartments were added to connect the network. Demand reactions were added for metabolites that are known to be produced, but their degradation pathways or their fractional abundance in biomass are unknown [88]. Sink reactions were added for metabolites are variable biomass components that are known to be dynamically produced and consumed. *i*Tps1426 is available in SBML format as Supplementary File B.1 and will be deposited in the BioModels database.

2.5.2 Subcellular protein localization

The protein localization pipeline developed by ref. [13] was updated for this analysis. For chloroplast targeting predictions, TargetP [92] was replaced with ASAFind [84], a plastid proteome prediction tool developed for diatoms and other algae with plastids derived from a secondary endosymbiosis. All *T. pseudonana* proteins were used as input for SignalP 4.1 [93], TargetP 1.1 [92], HECTAR 1.3 [94], Mitoprot II 1.101 [95], ASAFind 1.1.7 [84], predictNLS 1.3 [96], and scanProsite [97]. PredictNLS, a tool for predicting nucleus targeted proteins, was run in batch mode using a script that re-implements predictNLS 1.3 in Python (https://github.com/peterjc/pico_galaxy/tree/master/tools/predictnls). ScanProsite was run to search for two peroxisomal targeting signals "[SAC]-[KRH]-[LM]>" and "S-S-L>" [98]

and the PROSITE pattern PS00342 describing microbody C-terminal targeting signals, as well as the endoplasmic reticulum (ER) targeting signal "[KD]-[DE]-E-L>" [92, 99] and the PROSITE pattern PS00014 describing other endoplasmic reticulum targeting sequences. All other programs were run with default settings. ER-targeted proteins were defined as 'Not plastid, SignalP positive' or 'Plastid, low confidence' by ASAFind and contained an ER-targeting signal identified by scanProsite. Plastid-targeted proteins include those identified by ASAFind as 'Plastid, high confidence' or 'Plastid, low confidence' with no recognized ER targeting signal. Mitochondria targeted proteins are SignalP negative and predictNLS negative and match one of the following criteria: (A) have a Mitoprot II score > 0.9, (B) have a Mitoprot II score > 0.8 and are mitochondria targeted according to HECTAR or have a mitochondrial targeting peptide according to TargetP, or (C) are predicted to be mitochondria targeted by HECTAR and have a mitochondrial targeting peptide according to TargetP. Peroxisome targeted proteins are SignalP negative, predictNLS negative, not mitochondria targeted, and contain a peroxisome targeting signal according to scanProsite. Proteins in the plastid and mitochondrial genomes were assigned to reactions in the plastid and mitochondria, respectively. All remaining proteins were assigned to the cytosol. 408 sequences from the optimized gene catalog could not be run through this subcellular localization pipeline despite curation by ref. [84] because they have internal stop codons, or were either too long or short for some of the programs.

2.5.3 Biomass objective function

Biomass reactions convert metabolites in the model into cellular components that comprise the millimolar contribution to 1 g dry weight (gDW) of cell mass under specific growth conditions [88]. A biomass reaction was developed for *T. pseudonana* cells acclimated to $\sim$100 μmol photons m^{-2} s^{-1} using biomass composition measurements from the literature and from genome sequence data. The biomass reaction constitutes DNA, RNA, protein, free amino acids, pigments, carbohydrates, lipids (phospholipids, sulfolipids, galactolipids, glycerolipids), and a soluble pool of osmolytes, polyamines, and cofactors (see Supplementary

Data Set B.2 for calculations and references). Measurements derived from the literature were typically performed on samples harvested from batch cultures during the exponential growth phase, at temperatures ranging from 15-21^{o}C (optimal growth temperature is 21^{o}C).

2.5.4 Mechanistic model of light-harvesting

Using ref. [31] as an guideline, stoichiometric reactions were generated to represent light harvesting in *T. pseudonana*. The pigment weight and composition [100, 101] was used in combination with the weight-specific absorption spectra for each pigment [39, 102] to calculate the relative absorption of each pigment within 20 nm bins in the photosynthetically active radiation range (PAR range; 400-700 nm). Excitation energy transfer reactions were generated to account for energy loss in the transfer of excitation energy from different pigments in the FCPs to chlorophyll *a* in the reaction centers. Photon absorption was constrained for each wavelength using the absorption spectrum of *T. pseudonana* cells and the light intensity spectrum of a cool white fluorescent bulb according to the methodology in ref. [31]. PSI and PSII reactions were modified to include charge separation and recombination reactions. The rate of photodamage to the PSII D1 subunit was included as a proportion of the flux through PSII [31, 34]. A non-growth associated ATP maintenance reaction was generated to account for the costs of D1 subunit degradation and biosynthesis [31, 103]. See Supplementary Data Set B.3 for calculations and references.

2.5.5 Flux Balance Analysis

FBA simulates the flow of metabolites through a network of reactions using the mass balance equation,

$$\frac{dx}{dt} = S \cdot v = 0 \tag{2.12}$$

given that,

$$LB \leq v \leq UB \tag{2.13}$$

where the change in metabolite concentration (dx) over time (dt) is equal to the stoichiometric matrix (S) describing a reaction network times a vector of fluxes (v). Intracellular metabolites are assumed to be at steady state, which allows v to be solved while maximizing an objective function within certain bounds (LB, UB). The GLPK solver was used to optimize all FBA problems.

Growth simulations of *i*Tps1426 were performed by dynamic FBA. I modified Matlab scripts provided by E. Borenstein to implement the method published in ref. [104] in Python with the COBRApy package [91] for a photosynthetic organism. The growth period was divided into thirty-eight 8-hour intervals, and steady state was assumed for each time interval so that FBA could be used to solve for the growth rate and reaction flux distributions every 8 hours. I determined the uptake limit for the metabolite j at each time point t, such that the lower bound of each exchange reaction is equal to either the flux predicted by the Michaelis-Menten equation or the concentration of each metabolite x_j per gram dry weight (gDW) biomass ($bio(t)$) per time period Δt, whichever is closer to zero.

$$LB_j(t) = -min\left(\frac{V_{max} \cdot x_j}{k_m + x_j}, \frac{x_j}{bio(t) \cdot \Delta t}\right) \tag{2.14}$$

Michaelis-Menten parameters were calculated from the literature on *T. pseudonana* for HCO_3, CO_2 [105], NO_3 [106], PO_4 [107], NH_4 [108], and urea [109] (Supplementary Data Set B.3). SO_4 uptake rates were obtained from maize roots [110]. The concentrations of HCO_3 and CO_2 in Aquil were calculated for seawater concentrations at 390 ppm pCO_2 [34]. As in ref. [104], V_{max} was set to 20 and K_m was set to 0.05 for all other uptake reactions, with the exception of H_2O and H^+ which were assumed to diffuse freely and V_{max} was set to 1000. The biomass-dependent lower bound assumes that *i*Tps1426 exists in a well-mixed environment where it can sense nutrient-availability and adjust its uptake rate accordingly [104]. The lower bound for CO_2 and HCO_3 uptake were always constrained with Michaelis-Menten parameters. O_2 production fluxes were constrained during the light period using experimental values from the photosynthesis-irradiance (I) curve for cells acclimated to 60 μmol photons m^{-2} s^{-1} published in ref. [36] using Platt fitting [37] parameters and dark

respiration (0.086 mmol O_2 (mg Chl a)$^{-1}$ h^{-1}).

$$NPP = \left(0.279 \cdot \left(1 - e^{\frac{-0.0044 \cdot I}{0.279}}\right) - 0.086\right) \cdot \left(\frac{mg\ chl\ a/cell}{gDW/cell}\right) \tag{2.15}$$

The compensation light level (at which net O_2 production (NPP) = 0) was calculated to be 23.3 μmol photons m^{-2} s^{-1}. If I > 23.3, then the bounds on the oxygen exchange reaction ('EX_o2_e') were set as

$$0.99 \cdot \leq EX\_o2\_e(t) \leq 1 \cdot NPP \tag{2.16}$$

Otherwise, NPP was negative the bounds on EX_o2_e were set as

$$1 \cdot NPP \leq EX\_o2\_e(t) \leq 0 \tag{2.17}$$

The lower bound of the sum of CO_2 and HCO_3 uptake was constrained with gross primary productivity values for cells acclimated to 60 μmol photons m^{-2} s^{-1} using Platt fitting [37, 36].

$$-1 \cdot 0.243 \cdot \left(1 - e^{\frac{-0.0021 \cdot I}{0.243}}\right) \cdot \left(\frac{mg\ chl\ a/cell}{gDW/cell}\right) \leq EX\_co2\_e + EX\_hco3\_e \tag{2.18}$$

After each time point, the concentration of O_2, HCO_3, and CO_2 was re-equilibrated with the atmosphere so that concentrations remained constant (well-mixed). Photon uptake was constrained for each 20 nm in the PAR spectral range using the method published in ref. [31],

$$LB_{\lambda_{20}}(t) = -min\left(r_{\lambda_{20}} \cdot I \cdot a_{\lambda_{20}}, \frac{r_{\lambda_{20}} \cdot I \cdot A}{bio(t)} + \sum v_{\lambda_{20} i*}\right) \tag{2.19}$$

$$UB_{\lambda_{20}}(t) = -0.9999 \cdot min(PFA, PFD) \tag{2.20}$$

, where photon uptake is constrained by the photon absorption flux (PFA) or the photon flux density (PFD), whichever is closer to zero. The term $r_{\lambda_{20}}$ is the fraction of the light source's photon flux at a particular wavelength integrated within a 20 nm bin, $a_{\lambda_{20}}$ is the absorption of light at that wavelength (m^2 gDW^{-1}) [111], and A is set to the surface area of the flask (a 300 mL tissue culture flask = $\sim$0.0225 m^2). The culture volume was divided into 10 slices and $v_{\lambda_{20} i*}$ is the sum of photon flux density in all previous slices to account for self-shading

in culture. The light intensity (I) in each slice was back-calculated from the photon flux values and used to calculate the rate of D1 damage for the PSII reaction ('PSII_u'), the rate of O_2 evolution ('EX_o2_e'), and the lower bound of carbon uptake. The medium was initialized with a concentration of 1,3-glucan and chitin representing 23% [48] and 15% [49, 50] of initial biomass, respectively. The upper bounds of 1,3-glucan and chitin sink reactions only unblocked when their concentration was less than 23% and 15% of biomass at that time point. iTps1426 was constrained with the photorespiration and mitochondrial respiration constraints developed above. I used parsimonious flux balance analysis (pFBA) to generate single solutions for each time point. pFBA optimizes the model objective and then minimizes the total sum of flux to obtain single solution that contains no thermodynamically infeasible loops. The second objective imitates a possible cellular objective of minimizing protein biosynthesis. The fraction of optimum was set to 0.99 so that biomass production must be at least 99% of optimum.

The resulting flux distribution was used to calculate the biomass of iTps1426 and the concentration of metabolite x_j in the next time interval.

$$bio(t_2) = bio(t_1) \cdot e^{\mu \Delta t} \tag{2.21}$$

$$x_j(t_2) = x_j(t_1) + \sum \frac{v_j}{\mu}[bio(t_1)(e^{\mu \Delta t} - 1)] \tag{2.22}$$

, where μ is the growth rate of iTps1426 and v_j is the flux of each metabolite. If $\mu = 0$, then it was rounded up to 1e-6 (the error tolerance). All fluxes were all converted from mmol gDW^{-1} h^{-1} to μmol gDW^{-1} h^{-1} to avoid numerical precision issues associated with small fluxes.

2.6 Acknowledgements

I would like to thank Jennifer Levering for providing a list of gene IDs for the iLB1027_lipid model, Elhanan Borenstein for providing Matlab scripts for dynamic Flux Balance Analysis, and his student Colin McNally for his help with the scripts.

References

[1] PG Falkowski. "The evolution of modern eukaryotic phytoplankton." In: *Science* 305.5682 (July 2004), pp. 354–60. ISSN: 0036-8075. DOI: 10.1126/science.1095964.

[2] PA Sims, DG Mann, and LK Medlin. "Evolution of the diatoms: insights from fossil, biological and molecular data." In: *Phycologia* 45.4 (July 2006), pp. 361–402. ISSN: 0031-8884. DOI: 10.2216/05-22.1.

[3] AE Allen, A Vardi, and C Bowler. "An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms." In: *Current Opinion in Plant Biology* 9.3 (June 2006), pp. 264–73. ISSN: 13695266. DOI: 10.1016/j.pbi.2006.03.013.

[4] AE Allen et al. "Evolution and metabolic significance of the urea cycle in photosynthetic diatoms." In: *Nature* 473.7346 (May 2011), pp. 203–7. ISSN: 0028-0836. DOI: 10.1038/nature10074.

[5] C Bowler et al. "The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes." In: *Nature* 456.7219 (Nov. 2008), pp. 239–44. ISSN: 0028-0836. DOI: 10.1038/nature07410.

[6] BP Durham et al. "Cryptic carbon and sulfur cycling between surface ocean plankton." In: *Proceedings of the National Academy of Sciences* 112.2 (Jan. 2015), pp. 453–7. ISSN: 0027-8424. DOI: 10.1073/pnas.1413137112.

[7] JF Allen. "Photosynthesis of ATP-electrons, proton pumps, rotors, and poise." In: *Cell* 110.3 (Aug. 2002), pp. 273–6. ISSN: 00928674. DOI: 10.1016/S0092-8674(02)00870-X.

[8] DJ Kyle, I Ohad, and CJ Arntzen. "Membrane protein damage and repair: Selective loss of a quinoneprotein function in chloroplast membranes." In: *Proceedings of the National Academy of Sciences* 81.13 (July 1984), pp. 4070–4. ISSN: 0027-8424. DOI: 10.1073/pnas.81.13.4070.

[9] K Asada et al. “The water-water cycle as alternative photon and electron sinks.” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 355.1402 (Oct. 2000), pp. 1419–31. ISSN: 09628436. DOI: 10.1098/rstb.2000.0703.

[10] Y Munekage et al. “Cyclic electron flow around Photosystem I is essential for photosynthesis.” In: *Nature* 429.6991 (June 2004), pp. 579–82. ISSN: 00280836. DOI: 10.1038/nature02598.

[11] JF Allen. “Oxygen reduction and optimum production of ATP in photosynthesis.” In: *Nature* 256.5518 (Aug. 1975), pp. 599–600. ISSN: 0028-0836. DOI: 10.1038/256599a0.

[12] B Bailleul et al. “Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms.” In: *Nature* 524.7565 (Aug. 2015), pp. 366–9. ISSN: 0028-0836. DOI: 10.1038/nature14599.

[13] J Levering et al. “Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom.” In: *PLoS ONE* 11.5 (May 2016), e0155038. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0155038.

[14] J Kim et al. “Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*.” In: *The Plant Journal* 85.1 (Jan. 2016), pp. 161–76. ISSN: 09607412. DOI: 10.1111/tpj.13081.

[15] BP Durham et al. “Disentangling microbial networks in the sea: sulfonate-based trophic interactions between ocean plankton.” 2019.

[16] ST Dyhrman et al. “A microbial source of phosphonates in oligotrophic marine systems.” In: *Nature Geoscience* 2.10 (Oct. 2009), pp. 696–9. ISSN: 1752-0894. DOI: 10.1038/ngeo639.

[17] A Varma and BO Palsson. “Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.” In: *Applied and environmental microbiology* 60.10 (Oct. 1994), pp. 3724–31. ISSN: 0099-2240. DOI: PMC201879.

[18] I Zlotnik and Z Dubinsky. "The effect of light and temperature on DOC excretion by phytoplankton." In: *Limnology and Oceanography* 34.5 (July 1989), pp. 831–9. ISSN: 19395590. DOI: 10.4319/lo.1989.34.5.0831.

[19] MW Lomas and PM Gilbert. "Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms." In: *Limnology and Oceanography* 44.3 (May 1999), pp. 556–72. ISSN: 00243590. DOI: 10.4319/lo.1999.44.3.0556.

[20] SM Myklestad. "Release of extracellular products by phytoplankton with special emphasis on polysaccharides." In: *Science of The Total Environment* 165.1-3 (Apr. 1995), pp. 155–64. ISSN: 00489697. DOI: 10.1016/0048-9697(95)04549-G.

[21] E Granum, S Kirkvold, and SM Myklestad. "Cellular and extracellular production of carbohydrates and amino acids by the marine diatom *Skeletonema costatum*: Diel variations and effects of N depletion." In: *Marine Ecology Progress Series* 242 (Oct. 2002), pp. 83–94. ISSN: 01718630. DOI: 10.3354/meps242083.

[22] KE Helliwell et al. "Insights into the evolution of vitamin B_{12} auxotrophy from sequenced algal genomes." In: *Molecular Biology and Evolution* 28.10 (Oct. 2011), pp. 2921–33. ISSN: 0737-4038. DOI: 10.1093/molbev/msr124.

[23] KR Heal et al. "Two distinct pools of B_{12} analogs reveal community interdependencies in the ocean." In: *Proceedings of the National Academy of Sciences* 114.2 (Jan. 2017), pp. 364–9. ISSN: 0027-8424. DOI: 10.1073/pnas.1608462114.

[24] G Peers and NM Price. "Copper-containing plastocyanin used for electron transport by an oceanic diatom." In: *Nature* 441.7091 (May 2006), pp. 341–4. ISSN: 1476-4687. DOI: 10.1038/nature04630.

[25] I Grouneva, A Rokka, and E Aro. "The thylakoid membrane proteome of two marine diatoms outlines both diatom-specific and species-specific features of the photosyn-

thetic machinery." In: *Journal of Proteome Research* 10.12 (Oct. 2011), pp. 5338–53. ISSN: 15353893. DOI: 10.1021/pr200600f.

[26] L Li, CJ Stoeckert, and DS Roos. "OrthoMCL: Identification of ortholog groups for eukaryotic genomes." In: *Genome Research* 13.9 (Sept. 2003), pp. 2178–89. ISSN: 1088-9051. DOI: 10.1101/gr.1224503.

[27] A Montsant. "Comparative genomics of the pennate diatom *Phaeodactylum tricornutum*." In: *Plant Physiology* 137.2 (Feb. 2005), pp. 500–13. ISSN: 0032-0889. DOI: 10.1104/pp.104.052829.

[28] M Samukawa et al. "Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*." In: *Photosynthesis Research* 121.2-3 (Sept. 2014), pp. 235–49. ISSN: 0166-8595. DOI: 10.1007/s11120-014-9967-x.

[29] JE Hunter. "Phytoplankton lipidomics : Lipid dynamics in response to microalgal stressors." PhD thesis. University of Southampton, Oct. 2015, pp. 1–230.

[30] ET Yu et al. "Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Baccilariophyceae) during starvation." In: *Journal of Applied Phycology* 21.6 (Dec. 2009), pp. 669–81. ISSN: 0921-8971. DOI: 10.1007/s10811-008-9400-y.

[31] JT Broddrick et al. "Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis." In: *Proceedings of the National Academy of Sciences* 113.51 (Dec. 2016), E8344–53. ISSN: 0027-8424. DOI: 10.1073/pnas.1613446113.

[32] RL Chang et al. "Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism." In: *Molecular Systems Biology* 7.518 (2011), pp. 1–13. ISSN: 17444292. DOI: 10.1038/msb.2011.52.

[33] E Papagiannakis et al. "Spectroscopic characterization of the excitation energy transfer in the fucoxanthin-chlorophyll protein of diatoms." In: *Photosynthesis Research* 86.1-2 (Nov. 2005), pp. 241–50. ISSN: 0166-8595. DOI: 10.1007/s11120-005-1003-8.

[34] G Li and DA Campbell. "Rising CO_2 interacts with growth light and growth rate to alter Photosystem II photoinactivation of the coastal diatom *Thalassiosira pseudonana*." In: *PLoS ONE* 8.1 (Jan. 2013), e55562. ISSN: 19326203. DOI: 10.1371/journal.pone.0055562.

[35] DA Campbell et al. "Photosystem II protein clearance and FtsH function in the diatom *Thalassiosira pseudonana*." In: *Photosynthesis Research* 115.1 (Mar. 2013), pp. 43–54. ISSN: 01668595. DOI: 10.1007/s11120-013-9809-2.

[36] NL Fisher and KH Halsey. "Mechanisms that increase the growth efficiency of diatoms in low light." In: *Photosynthesis Research* 129.2 (Aug. 2016), pp. 183–97. ISSN: 15735079. DOI: 10.1007/s11120-016-0282-6.

[37] T Platt, CL Gallegos, and WG Harrison. "Photoinhibition of photosynthesis in natural assemblages of marine-phytoplankton." In: *Journal of Marine Research* 38.4 (1980), pp. 687–701.

[38] JR Geider, BA Osborne, and JA Raven. "Light dependence of growth and photosynthesis in *Phaeodactylum tricornutum* (Bacillariophyceae)." In: *Journal of Phycology* 21.4 (Oct. 1985), pp. 609–19. ISSN: 00223646. DOI: 10.1111/j.0022-3646.1985.00609.x.

[39] E Devred et al. "Future retrievals of water column bio-optical properties using the Hyperspectral Infrared Imager (HyspIRI)". In: *Remote Sensing* 5.12 (Dec. 2013), pp. 6812–37. ISSN: 20724292. DOI: 10.3390/rs5126812.

[40] T Shikanai. "Cyclic electron transport around Photosystem I: Genetic approaches." In: *Annual Review of Plant Biology* 58.1 (June 2007), pp. 199–217. ISSN: 1543-5008. DOI: 10.1146/annurev.arplant.58.091406.110525.

[41] A Davis et al. “Clarification of photorespiratory processes and the role of malic enzyme in diatoms.” In: *Protist* 168.1 (Feb. 2017), pp. 134–53. ISSN: 16180941. DOI: 10.1016/j.protis.2016.10.005.

[42] JN Young et al. “Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms.” In: *Journal of Experimental Botany* 67.11 (May 2016), pp. 3445–56. ISSN: 14602431. DOI: 10.1093/jxb/erw163.

[43] JN Young and BM Hopkinson. “The potential for co-evolution of CO_2-concentrating mechanisms and Rubisco in diatoms.” In: *Journal of experimental botany* 68.14 (June 2017), pp. 3751–62. ISSN: 14602431. DOI: 10.1093/jxb/erx130.

[44] AMC Heureux et al. “The role of Rubisco kinetics and pyrenoid morphology in shaping the CCM of haptophyte microalgae.” In: *Journal of experimental botany* 68.14 (Sept. 2017), pp. 3959–69. ISSN: 14602431. DOI: 10.1093/jxb/erx179.

[45] P Claquin, JC Kromkamp, and V Martin-Jezequel. “Relationship between photosynthetic metabolism and cell cycle in a synchronized culture of the marine alga *Cylindrotheca fusiformis* (Bacillariophyceae).” In: *European Journal of Phycology* 39.1 (Feb. 2004), pp. 33–41. ISSN: 09670262. DOI: 10.1080/0967026032000157165.

[46] J Waring et al. “Light-induced responses of oxygen photoreduction, reactive oxygen species production and scavenging in two diatom species.” In: *Journal of Phycology* 46.6 (Nov. 2010), pp. 1206–17. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2010.00919.x.

[47] MR Brown et al. “Effects of harvest stage and light on the biochemical composition of the diatom *Thalassiosira pseudonana*.” In: *Journal of Phycology* 32.1 (Feb. 1996), pp. 64–73. ISSN: 0022-3646. DOI: 10.1111/j.0022-3646.1996.00064.x.

[48] Mark Hildebrand, Kalpana Manandhar-Shrestha, and Raffaela Abbriano. “Effects of chrysolaminarin synthase knockdown in the diatom *Thalassiosira pseudonana*: Impli-

cations of reduced carbohydrate storage relative to green algae." In: *Algal Research* 23 (Apr. 2017), pp. 66–77. ISSN: 22119264. DOI: 10.1016/j.algal.2017.01.010.

[49] A Ozkan and GL Rorrer. "Effects of CO_2 delivery on fatty acid and chitin nanofiber production during photobioreactor cultivation of the marine diatom *Cyclotella* sp." In: *Algal Research* 26 (July 2017), pp. 422–30. ISSN: 22119264. DOI: 10.1016/j.algal.2017.07.003.

[50] A Ozkan and GL Rorrer. "Lipid and chitin nanofiber production during cultivation of the marine diatom *Cyclotella* sp. to high cell density with multistage addition of silicon and nitrate." In: *Journal of Applied Phycology* 29.4 (Aug. 2017), pp. 1811–8. ISSN: 15735176. DOI: 10.1007/s10811-017-1113-7.

[51] R Goss and T Jakob. "Regulation and function of xanthophyll cycle-dependent photoprotection in algae." In: *Photosynthesis Research* 106.1-2 (Nov. 2010), pp. 103–22. ISSN: 01668595. DOI: 10.1007/s11120-010-9536-x.

[52] JC Kromkamp and J Peen. "Oxygen consumption in the light by unicellular algae." In: 4 (May 1992), pp. 1–6.

[53] KR Heal et al. "Determination of four forms of vitamin B_{12} and other B vitamins in seawater by liquid chromatography/tandem mass spectrometry." In: *Rapid communications in mass spectrometry : RCM* 28.22 (Oct. 2014), pp. 2398–404. ISSN: 10970231. DOI: 10.1002/rcm.7040.

[54] EV Armbrust et al. "The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism." In: *Science* 306.5693 (Oct. 2004), pp. 79–86. ISSN: 0036-8075. DOI: 10.1126/science.1101156.

[55] J Ashworth et al. "Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*." In: *Proceedings of the National Academy of Sciences* 110.18 (Apr. 2013), pp. 7518–23. ISSN: 0027-8424. DOI: 10.1073/pnas.1300962110.

[56] JJ Valenzuela et al. "Ocean acidification conditions increase resilience of marine diatoms". In: *Nature Communications* 9.1 (June 2018). ISSN: 20411723. DOI: 10.1038/s41467-018-04742-3.

[57] SJ Bender et al. "Transcriptional responses of three model diatoms to nitrate limitation of growth." In: *Frontiers in Marine Science* 1 (Mar. 2014), pp. 1–15. ISSN: 2296-7745. DOI: 10.3389/fmars.2014.00003.

[58] ST Dyhrman et al. "The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response." In: *PLoS ONE* 7.3 (Mar. 2012). ISSN: 19326203. DOI: 10.1371/journal.pone.0033768.

[59] K Longnecker, MC Kido Soule, and EB Kujawinski. "Dissolved organic matter produced by *Thalassiosira pseudonana*." In: *Marine Chemistry* 168 (Jan. 2015), pp. 114–23. ISSN: 0304-4203. DOI: 10.1016/j.marchem.2014.11.003.

[60] JA Downing. "Marine nitrogen: Phosphorus stoichiometry and the global N:P cycle." In: *Biogeochemistry* 37.3 (1997), pp. 237–52.

[61] DM Karl and RM Letelier. "Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes." In: *Marine Ecology Progress Series* 364 (July 2008), pp. 257–68. ISSN: 0171-8630. DOI: 10.3354/meps07547.

[62] S Emerson and J Hedges. "Oceanography background: dissolved chemicals, circulation and biology in the sea." In: *Chemical Oceanography and the Marine Carbon Cycle*. Cambridge: Cambridge University Press, pp. 3–32. DOI: 10.1017/CBO9780511793202.002.

[63] DH Turpin et al. "Interactions between photosynthesis, respiration, and nitrogen assimilation in microalgae." In: *Canadian Journal of Botany* 66.10 (1988), pp. 2083–97. ISSN: 0008-4026. DOI: 10.1139/b88-286.

[64] O Murik et al. “Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom.” In: *New Phytologist* 221 (2019), pp. 1303–16. ISSN: 0028646X. DOI: 10.1111/nph.15479.

[65] MW Lomas. “Ammonium release by nitrogen sufficient diatoms in response to rapid increases in irradiance.” In: *Journal of Plankton Research* 22.12 (May 2000), pp. 2351–66. ISSN: 14643774. DOI: 10.1093/plankt/22.12.2351.

[66] MS Parker et al. “Induction of photorespiration by light in the centric diatom *Thalassiosira weissflogii* (Bacillariophyceae): Molecular characterization and physiological consequences.” In: *Journal of Phycology* 40 (May 2004), pp. 557–67. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2004.03184.x.

[67] MS Parker and EV Armbrust. “Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae).” In: *Journal of Phycology* 41.6 (Nov. 2005), pp. 1142–53. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2005.00139.x.

[68] H Heyer and WE Krumbein. “Excretion of fermentation products in dark and anaerobically incubated cyanobacteria.” In: *Archives of Microbiology* 155.3 (Feb. 1991), pp. 284–7. ISSN: 03028933. DOI: 10.1007/BF00252213.

[69] DW Husic and NE Tolbert. “Anaerobic formation of D-lactate and partial purification and characterization of a pyruvate reductase from *Chlamydomonas reinhardtii*.” In: *Plant Physiology* 78.2 (June 1985), pp. 277–84. ISSN: 0032-0889. DOI: 10.1104/pp.78.2.277.

[70] DW Husic and NE Tolbert. “Inhibition of glycolate and D-lactate metabolism in a *Chlamydomonas reinhardtii* mutant deficient in mitochondrial respiration.” In: *Proceedings of the National Academy of Sciences* 84.6 (Mar. 1987), pp. 1555–9. ISSN: 0027-8424. DOI: 10.1073/pnas.84.6.1555.

[71] S Myklestad. "Production of carbohydrates by marine planktonic diatoms. II. Influence of the ratio in the growth medium on the assimilation ratio, growth rate, and production of cellular and extracellular carbohydrates by *Chaetoceros affinis* var. *willei* (Gran) Hustedt and *Skeletonema costatum* (Grev.) Cleve". In: *Journal of Experimental Marine Biology and Ecology* 29.2 (Sept. 1977), pp. 161–79. ISSN: 00220981. DOI: 10.1016/0022-0981(77)90046-6.

[72] N Staats, LJ Stal, and LR Mur. "Exopolysaccharide production by the epipelic diatom *Cylindrotheca closterium*: Effects of nutrient conditions." In: *Journal of Experimental Marine Biology and Ecology* 249.1 (June 2000), pp. 13–27. ISSN: 00220981. DOI: 10.1016/S0022-0981(00)00166-0.

[73] T Alcoverro, E Conte, and L Mazzella. "Production of mucilage by the adriatic epipelic diatom *Cylindrotheca closterium* (Bacillariophyceae) under nutrient limitation." In: *Journal of Phycology* 36.6 (Dec. 2000), pp. 1087–95. ISSN: 0022-3646. DOI: 10.1046/j.1529-8817.2000.99193.x.

[74] R Urbani et al. "Extracellular carbohydrates released by the marine diatoms *Cylindrotheca closterium*, *Thalassiosira pseudonana* and *Skeletonema costatum*: Effect of P-depletion and growth status." In: *Science of the Total Environment* 353.1-3 (Dec. 2005), pp. 300–6. ISSN: 00489697. DOI: 10.1016/j.scitotenv.2005.09.026.

[75] W Admiraal, H Peletier, and RWPM Laane. "Nitrogen metabolism of marine planktonic diatoms; excretion, assimilation and cellular pools of free amino acids in seven species with different cell size." In: *Journal of Experimental Marine Biology and Ecology* 98.3 (July 1986), pp. 241–63. ISSN: 00220981. DOI: 10.1016/0022-0981(86)90216-9.

[76] H Sarmento et al. "Phytoplankton species-specific release of dissolved free amino acids and their selective consumption by bacteria." In: *Limnology and Oceanography* 58.3 (May 2013), pp. 1123–35. ISSN: 00243590. DOI: 10.4319/lo.2013.58.3.1123.

[77] E Bucciarelli and WG Sunda. "Influence of CO_2, nitrate, phosphate, and silicate limitation on intracellular dimethylsulfoniopropionate in batch cultures of the coastal diatom *Thalassiosira pseudonana*." In: *Limnology and Oceanography* 48.6 (Nov. 2003), pp. 2256–65. ISSN: 00243590. DOI: 10.4319/lo.2003.48.6.2256.

[78] WWY Lau and EV Armbrust. "Detection of glycolate oxidase gene glcD diversity among cultured and environmental marine bacteria." In: *Environmental Microbiology* 8.10 (July 2006), pp. 1688–702. ISSN: 14622912. DOI: 10.1111/j.1462-2920.2006.01092.x.

[79] WWY Lau, RG Keil, and EV Armbrust. "Succession and diel transcriptional response of the glycolate-utilizing component of the bacterial community during a spring phytoplankton bloom." In: *Applied and Environmental Microbiology* 73.8 (Apr. 2007), pp. 2440–50. ISSN: 00992240. DOI: 10.1128/AEM.01965-06.

[80] JM González et al. "*Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments." In: *International Journal of Systematic and Evolutionary Microbiology* 53.5 (Sept. 2003), pp. 1261–9. ISSN: 1466-5026. DOI: 10.1099/ijs.0.02491-0.

[81] RJ Newton et al. "Genome characteristics of a generalist marine bacterial lineage." In: *The ISME Journal* 4.6 (July 2010), pp. 784–98. ISSN: 1751-7370. DOI: 10.1038/ismej.2009.150.

[82] EC Howard et al. "Abundant and diverse bacteria involved in DMSP degradation in marine surface waters." In: *Environmental microbiology* 10.9 (Sept. 2008), pp. 2397–410. ISSN: 1462-2920. DOI: 10.1111/j.1462-2920.2008.01665.x.

[83] H Teeling et al. "Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom." In: *Science* 336.6081 (May 2012), pp. 608–11. ISSN: 0036-8075. DOI: 10.1126/science.1218344.

[84] A Gruber et al. “Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage.” In: *The Plant Journal* 81.3 (Feb. 2015), pp. 519–28. ISSN: 09607412. DOI: 10.1111/tpj.12734. URL: http://doi.wiley.com/10.1111/tpj.12734.

[85] R Caspi et al. “The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.” In: *Nucleic Acids Research* 42.D1 (Jan. 2014), pp. 471–80. ISSN: 03051048. DOI: 10.1093/nar/gkt1103.

[86] K Parker. “Metabolic network construction based on the genome of the marine diatom *Thalassiosira pseudonana* and the analysis of genome-wide transcriptome data to investigate triacylglyceride accumulation.” PhD thesis. San Jose State University, Dec. 2013, pp. 1–93. ISBN: 9781303715341.

[87] ZA King et al. “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models.” In: *Nucleic Acids Research* 44.D1 (Oct. 2016), pp. D515–22. ISSN: 1362-4962. DOI: 10.1093/nar/gkv1049.

[88] I Thiele and BO Palsson. “A protocol for generating a high-quality genome-scale metabolic reconstruction.” In: *Nature protocols* 5.1 (Jan. 2010), pp. 93–121. ISSN: 1750-2799. DOI: 10.1038/nprot.2009.203. arXiv: NIHMS150003.

[89] P Jones et al. “InterProScan 5: genome-scale protein function classification.” In: *Bioinformatics* 30.9 (May 2014), pp. 1236–40. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu031.

[90] M Kanehisa et al. “KEGG: New perspectives on genomes, pathways, diseases and drugs.” In: *Nucleic Acids Research* 45.D1 (Nov. 2017), pp. D353–61. ISSN: 13624962. DOI: 10.1093/nar/gkw1092.

[91] A Ebrahim et al. “COBRApy: COnstraints-Based Reconstruction and Analysis for Python.” In: *BMC Systems Biology* 7.1 (Aug. 2013), p. 74. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74.

[92] O Emanuelsson et al. "Locating proteins in the cell using TargetP, SignalP and related tools." In: *Nature protocols* 2.4 (Jan. 2007), pp. 953–71. ISSN: 1754-2189. DOI: 10.1038/nprot.2007.131.

[93] TN Petersen et al. "SignalP 4.0: discriminating signal peptides from transmembrane regions." In: *Nature Methods* 8.10 (Oct. 2011), pp. 785–6. ISSN: 1548-7091. DOI: 10.1038/nmeth.1701.

[94] B Gschloessl, Y Guermeur, and JM Cock. "HECTAR: A method to predict subcellular targeting in heterokonts." In: *BMC Bioinformatics* 9.1 (Sept. 2008), p. 393. ISSN: 1471-2105. DOI: 10.1186/1471-2105-9-393.

[95] MG Claros and P Vincens. "Computational method to predict mitochondrially imported proteins and their targeting sequences." In: *European journal of biochemistry / FEBS* 241.3 (Nov. 1996), pp. 779–86. ISSN: 0014-2956. DOI: 10.1111/j.1432-1033.1996.00779.x.

[96] M Cokol, R Nair, and B Rost. "Finding nuclear localization signals." In: *EMBO reports* 1.5 (2000), pp. 411–5. ISSN: 1469-3178. DOI: 10.1093/embo-reports/kvd092.

[97] A Gattiker, E Gasteiger, and A Bairoch. "ScanProsite: a reference implementation of a PROSITE scanning tool." In: *Applied Bioinformatics* 1.2 (2002), pp. 107–8. ISSN: 1175-5636.

[98] NH Gonzalez et al. "A single peroxisomal targeting signal mediates matrix protein import in diatoms." In: *PLoS ONE* 6.9 (Sept. 2011). ISSN: 19326203. DOI: 10.1371/journal.pone.0025316.

[99] PG Kroth et al. "A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis." In: *PLoS One* 3.1 (Jan. 2008), e1426. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0001426.

[100] A McCarthy et al. “Elevated carbon dioxide differentially alters the photophysiology of *Thalassiosira pseudonana* (Bacillariophyceae) and *Emiliania huxleyi* (Haptophyta).” In: *Journal of Phycology* 48.3 (Apr. 2012), pp. 635–46. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2012.01171.x.

[101] S Mas et al. “Diel variations in optical properties of *Imantonia rotunda* (Haptophyceae) and *Thalassiosira pseudonana* (Bacillariophyceae) exposed to different irradiance levels.” In: *Journal of Phycology* 44.3 (May 2008), pp. 551–63. ISSN: 00223646. DOI: 10.1111/j.1529-8817.2008.00501.x.

[102] T Fujiki and S Taguchi. “Relationship between light absorption and the xanthophyll-cycle pigments in marine diatoms.” In: *Cell* 48.2 (2001), pp. 96–103. ISSN: 13430874.

[103] AM Feist and BO Palsson. “The biomass objective function”. In: *Current Opinion in Microbiology* 13.3 (Apr. 2010), pp. 344–9. ISSN: 13695274. DOI: 10.1016/j.mib.2010.03.003.

[104] H Chiu, R Levy, and E Borenstein. “Emergent biosynthetic capacity in simple microbial communities.” In: *PLoS Computational Biology* 10.7 (July 2014), e1003695. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003695.

[105] S Trimborn et al. “The effect of pCO_2 on carbon acquisition and intracellular assimilation in four marine diatoms.” In: *Journal of Experimental Marine Biology and Ecology* 376.1 (July 2009), pp. 26–36. ISSN: 00220981. DOI: 10.1016/j.jembe.2009.05.017.

[106] P Cermeño et al. “Competitive dynamics in two species of marine phytoplankton under non-equilibrium conditions.” In: *Marine Ecology Progress Series* 429 (May 2011), pp. 19–28. ISSN: 01718630. DOI: 10.3354/meps09088.

[107] MJ Perry. “Phosphate utilization by an oceanic diatom in phosphorus-limited chemostat culture and in the oligotrophic waters of the central North Pacific.” In: *Limnology and Oceanography* 21.1 (Jan. 1976), pp. 88–107. ISSN: 00243590. DOI: 10.4319/lo.1976.21.1.0088.

[108] K Yin, PJ Harrison, and Q Dortch. "Lack of ammonium inhibition of nitrate uptake for a diatom grown under low light conditions." In: *Journal of Experimental Marine Biology and Ecology* 228.1 (Oct. 1998), pp. 151–65. ISSN: 00220981. DOI: 10.1016/S0022-0981(98)00025-2.

[109] JJ McCarthy. "The uptake of urea by marine phytoplankton." In: *Journal of Phycology* 8.3 (Sept. 1972), pp. 216–22. ISSN: 15298817. DOI: 10.1111/j.1529-8817.1972.tb04031.x.

[110] FF Nocito. "Heavy metal stress and sulfate uptake in maize roots." In: *Plant Physiology* 141.3 (May 2006), pp. 1138–48. ISSN: 0032-0889. DOI: 10.1104/pp.105.076240.

[111] D Stramski, A Sciandra, and H Claustre. "Effects of temperature, nitrogen, and light limitation on the optical properties of the marine diatom *Thalassiosira pseudonana*." In: *Limnology and Oceanography* 47.2 (Mar. 2002), pp. 392–403. ISSN: 00243590. DOI: 10.4319/lo.2002.47.2.0392.

Chapter 3

GENOME-SCALE METABOLIC MODELING OF INTERACTION BETWEEN THE B12-AUXOTROPH *THALASSIOSIRA PSEUDONANA* AND THE B12-PRODUCING MARINE BACTERIUM *RUEGERIA POMEROYI*

3.1 Abstract

Vitamin B_{12} auxotrophy among diatoms is considered an important mechanism governing interactions with bacteria in the Roseobacter clade, which are principle producers of B_{12} (cobalamin) in the ocean. The cobalamin-producing bacterium *Ruegeria pomeroyi* DSS-3 supports growth of the diatom *Thalassiosira pseudonana* CCMP 1335 in the absence of external cobalamin additions to the media in exchange for organic matter, particularly the compound 2,3-dihydroxypropane-1-sulfonate (DHPS). I reconstructed a genome-scale metabolic network of *R. pomeroyi* that integrates genomic and physiological data to study the systemic basis of interaction with *T. pseudonana*. The model of DSS-3 (*i*Rpo1302) is a network of 1,695 reactions and 919 metabolites that cover the enzymatic activities of 1,302 genes (∼30% of the bacterial genome). The network topology of *R. pomeroyi* and a previous reconstruction of *T. pseudonana* (*i*Tps1426) were used to re-analyze transcriptomic data from each organism in co-culture versus each organism in mono-culture collected after 8 hours. Metabolites around which there was differential transcriptional regulation were identified in the reconstructed networks when the two treatments were compared. A signature of B_{12}-limitation for *T. pseudonana* and a signature of diatom-metabolite utilization for *R. pomeroyi* were identified. Flux Balance Analysis, a constraint-based modeling method, was used to simulate the exchange of metabolites between the two metabolic models and demonstrate that B_{12}-limited growth of *T. pseudonana* could explain the observed patterns of metabolite

secretion. Cobalamin is a cofactor for two different enzymes in diatoms, methionine synthase and methylmalonyl-CoA mutase. Yet, the lack of cobalamin has far-reaching effects on diatom metabolism due to the involvement of methionine synthase in the folate cycle and methyltransferase reactions, through the generation of S-adenosyl-methionine. Furthermore, genome-scale metabolic modeling of photosynthetic organisms has demonstrated that growth-limiting conditions can be a boon for opportunistic marine bacteria as the assimilation of inorganic nutrients and secretion of excess metabolites are important mechanisms for balancing redox reactions and dissipating energy, particularly in conditions of high light intensity.

3.1.1 Importance

Diatoms are a group of eukaryotic phytoplankton that are important primary producers in coastal regions. Dissolved organic matter secreted by diatoms fuels secondary production by heterotrophic bacterioplankton, resulting in rapid cycling of organic matter in the surface ocean. Transcriptomics and metabolomics have revealed that the physiology of diatoms is affected by the presence of different bacteria in their environment, and that marine bacteria interact with diatoms using various strategies. However, metabolic transfer between trophic levels remain difficult to describe because those fluxes are likely ephemeral, involve a diverse assortment of molecules, and occur over microscopic distances. Constraint-based metabolic modeling can be used to rapidly simulate nutrient and metabolic fluxes in sequenced marine organisms. There is a need in the field of Oceanography to connect genes to ecosystems. Genome-scale metabolic modeling gives us the ability to integrate a wealth of molecular information to effectively develop and test systems-level hypotheses.

3.2 Introduction

Interaction between phytoplankton and heterotrophic bacteria is the most basic unit of the marine food web, linking primary and secondary production [1, 2]. In the surface ocean, phytoplankton transform inorganic nutrients and CO_2 into algal biomass and metabolites,

up to 60% of which is rapidly consumed by bacterioplankton [3]. The trophic transfer of organic matter from primary to secondary producers is not a passive unidirectional process, where bacteria scavenged nutrients from a relatively homogeneous dilute pool. Instead, organic matter distribution is concentrated and patchy at the microscale [4, 5]. Molecular approaches have unveiled a realm where marine microbes exhibit unique metabolic and regulatory responses to each other's presence [6, 7, 8]. In this paradigm, organic matter characteristics and residence time are determined by the dynamics of phytoplankton-bacteria interactions [2].

Diatoms and bacteria from the Roseobacter clade co-occur in coastal regions [9], and these associations persist in culture [10]. During a diatom bloom in the North Sea, roseobacteria showed high levels of expression for transporters of low molecular weight compounds [11], indicating that these bacteria are partly responsible for rapid dissolved organic matter cycling in the surface ocean. Roseobacteria often have large genomes ($>$ 4 Mb) and broad metabolic diversity [12]. They are known to consume many of the metabolites produced by phytoplankton and also have the capacity to synthesize vitamins and secondary metabolites.

Early culturing work demonstrated that many diatoms require vitamin B_{12} (cobalamin) for growth, and that cobalamin-producing marine bacteria could support their growth in culture [13]. It is now clear that B vitamin auxotrophy is widespread among eukaryotic phytoplankton; many require these cofactors for growth but are unable to synthesize them. About half of eukaryotic phytoplankton in culture have an absolute requirement for vitamin B_{12} (cobalamin), while a quarter require vitamin B_1 (thiamine) and 8% require B_7 (biotin) [14]. In contrast, 86% of roseobacteria are capable of *de novo* B_{12} biosynthesis, while approximately half are auxotrophic for vitamin B_1 or B_7 [14]. Cobalamin-limitation can have profound impacts on mitochondrial metabolism, one-carbon metabolism (folate cycling), and methylation. Methionine synthase converts homocysteine to methionine in a two-step reaction involving two methyl transfers, from methylcob(III)alamin to homocysteine, and from 5-methyltetrahydrofolate to cob(I)alamin. Methionine synthase plays an important role in the folate cycle as the only enzyme capable of regenerating tetrahydrofolate (THF) from 5-

methyltetrahydrofolate. Methionine is also the precursor to S-adenosyl-methionine (SAM), a co-substrate in thiamine, and biotin biosynthesis reactions, as well as a methyl donor in methyltransferase reactions. Methylmalonyl-CoA mutase is the final step in the degradation of valine, isoleucine, and odd-chain fatty acids.

The extension of presumed vitamin-based mutualistic interactions to the global scale [15] has been controversial, as many phytoplankton may scavenge enough cobalamin from nutrient upwelling [16] and decaying cells in the surrounding seawater to support growth [17]. Yet, certain regions of the ocean such as the Ross Sea [18] and the Antarctic ice edge [19] have been shown to be co-limited for iron and B_{12} under conditions when the growth of B_{12}-producing marine bacteria is limited by iron availability. B_{12} co-deficiency with iron and nitrogen has also been detected at the eastern boundary of the South Atlantic Gyre [20]. Depletion of multiple vitamins has been measured along the California coast [21]. *Thaumarchaeota*, one of the principle B_{12} producers in the ocean, are more reliant on copper than iron availability [22], but are more abundant at and below the deep chlorophyll maximum [23].

Recent co-culture studies also support the possibility of multiple phytoplankton/bacteria vitamin-based mutualisms. The green picoeukaryote *Ostreococcus tauri* and the roseobacterium *Dinoroseobacter shibae* can form a mutualistic interaction where *O. tauri* growth is supported by cobalamin and thiamine produced by *D. shibae*, while *D. shibae* growth is supported by niacin, biotin, and *p*-aminobenzoic acid (a folate precursor) from the alga [24]. The diatom *Thalassiosira pseudonana* is auxotrophic for cobalamin; it has two B_{12}-dependent enzymes, methymalonyl-CoA mutase (MMM) and methionine synthase (METH), and three accessory proteins involved in supplying the cofactor to the enzymes. NADH:cob(II)alamin oxidoreductase (CBLA) transports cobalamin into the mitochondria, cob(I)alamin adenosyltransferase (CBLB) transforms cob(I)alamin into adenosylcobalamin, the cofactor for MMM, and methionine synthase reductase (MTRR) converts cob(II)alamin into methylcob(III)alamin, the cofactor for METH [25]. *Ruegeria pomeroyi* is prototrophic for cobalamin; it has B_{12}-dependent enzymes as well as the capacity to synthesize vitamin B_{12} *de*

novo. In addition to the B_{12}-dependent enzymes mentioned previously, *R. pomeroyi* has type II ribonucleotide reductase that utilizes the adenosylcobalamin cofactor.

Ref. [7] demonstrated that unexpected nutrient substitutions are possible when *T. pseudonana* was grown in the absence of B_{12} supplements, but in the presence of the cobalamin-producing bacterium *R. pomeroyi*. Rather than utilize sulfate which is abundant in seawater, *R. pomeroyi* utilized organic sulfur produced by the diatom in the forms of 2,3-dihydroxypropane-1-sulfonate (DHPS) and N-acetyltaurine. In addition, *R. pomeroyi* likely received organic nitrogen or NH_4^+ from *T. pseudonana* as the bacterium cannot assimilate nitrate, the only source of nitrogen in the media. In Chapter 2, I demonstrated that *T. pseudonana* may secrete organic nitrogen or sulfur as a mechanism to balance redox reactions under growth limiting conditions.

Here constraint-based metabolic modeling was used to study the interaction between *T. pseudonana* and *R. pomeroyi* from a systems-level perspective. A new genome-scale metabolic model was developed for *R. pomeroyi* for use with the previously developed genome-scale metabolic model of *T. pseudonana* (Chapter 2). Dynamic Flux Balance Analysis was used to simulate the growth of each organism in mono-culture and in co-culture. Each organism impacts its own extracellular space which feeds into the medium, impacting the nutrient conditions of the other organism in the next time point. The models were constrained with the transcriptomics data from ref. [7] to determine which aspects of the interaction might be transcriptionally controlled. Cobalamin-limitation played an important role in the interaction. DHPS production, in particular, was caused by the inhibition of methionine synthase by B_{12} deficiency. In addition, the growth inhibition caused by B_{12}-limitation prompted *T. pseudonana* to release more organic matter as a mechanism for balancing redox reactions and dissipating energy. *R. pomeroyi* benefited from the increased flux and more highly reduced organic matter, while some redox stress was alleviated for *T. pseudonana* as *R. pomeroyi* oxidized the diatom byproducts.

3.3 Results

3.3.1 Reconstruction of the R. pomeroyi metabolic model

A genome-scale metabolic model of *Ruegeria pomeroyi* DSS-3 (*i*Rpo1302) was reconstructed using available literature information and a previous reconstruction of a related alphaproteobacterium, *Rhodobacter sphaeroides* 2.4.1. (*i*Rsp1140, [26]). Both *R. sphaeroides* and *R. pomeroyi* belong to the *Rhodobacteraceae* family, and 50% of the genes in *R. pomeroyi* are shared with *R. sphaeroides*. Several major pathways were removed from *i*Rsp1140 to create *i*Rpo1302 including photosynthetic carbon fixation, carotenoid biosynthesis, biosynthesis of photosynthesis proteins, chlorophyll metabolism, starch and sucrose metabolism, and polyketide sugar unit biosynthesis. Several pathways were added to *i*Rpo1302, including dimethylsulfoniopropionate (DMSP) catabolism [27], 2,3-dihydroxypropane-1-sulfonate (DHPS) catabolism [28], N-acetyltaurine catabolism [28], isethionate catabolism [29], trimethylamine catabolism [30], choline-O-sulfate catabolism [31], a reductive acetyl-CoA pathway for carbon monoxide utilization [32], phenylacetate catabolism, ectoine catabolism, phosphonate and phosphinate metabolism, indole-3-acetic acid biosynthesis, catabolism of xanthine to glyoxylate and ammonia [33], and molybdenum cofactor biosynthesis. An alternative isoleucine biosynthesis pathway was also added to *i*Rpo1302 [27, 34]. Transport reactions were added for indole-3-acetic acid, acrylate, citrate, DMSP, DHPS, N-acetyltaurine, acyl-homoserine lactones, methylamine, dimethylamine, trimethylamine, trimethylamine N-oxide, (R)-3-hydroxybutanoate, choline-O-sulfate, ectoine, isethionate, methylphosphonate, carbon monoxide, cyanocob(III)alamin, aquacob(III)alamin, cyanide, allophanate, UDP-N-acetyl-D-glucosamine, formamide, carnitine, sarcosine, dimethylglycine, spermine, alpha-ribazole, 3-dehydroshikimate, shikimate, chorismate, phenylpyruvate, 3-methylthiopropionate, and 4-aminobenzoate. These transport reactions were added based on literature data showing utilization of these compounds, detection of these metabolites in the growth medium of *R. pomeroyi* [35], or transcriptomic evidence that these compounds may be utilized or secreted during microbial interactions [7, 8]. *i*Rpo1302 simulates the activity of ∼30% of the genes in

the *R. pomeroyi* genome, creating a model of metabolic activity of 919 unique metabolites by 1,695 reactions (Table 3.1, Supplementary Data Set C.1).

3.3.2 Metabolic network analysis of transcriptomic data

Ref. [7] analyzed transcriptomic data from 8 h and 56 h after cultures were initiated to identify several metabolites that were likely exchanged between *R. pomeroyi* and *T. pseudonana*; ref. [28] confirmed the exchange of DHPS. Metabolic pathways activated by these nutritional exchanges were also inferred from the relative change in concentration of intracellular metabolites [7]. The *i*Tps1426 and the *i*Rpo1302 network were used to re-analyze transcriptomic data from refs. [28, 7] and to identify metabolites associated with genes differentially transcribed in co-culture [37]. For each gene in the metabolic network, p-values were converted into z-scores where z-scores were calculated for each metabolite as the sum of z-scores for those genes required to catalyze a reaction involving a given metabolite (neighboring gene) divided by the square root of the number of neighbouring genes. These z-scores were then corrected for the background distribution of z-scores for each level of connectivity in the metabolic network (Supplementary Figure C.1). This method allows z-score assignment to each metabolite in both genome-scale metabolic models and data collection time points and evaluation of how transcriptional regulation may affect measured changes in metabolite concentration.

In this analysis of *R. pomeroyi* after 8 h in co-culture versus mono-culture, corrected z-scores ranged from 7.2 for DHPS to -7.1 for AMP, and the median z-score was 0.13, signifying that DHPS was a metabolite that was differentially regulated while AMP was not. Although there is no score cut-off for what constitutes differential regulation [38], the higher scoring metabolites are considered to be indicators of the metabolic differences that arise as a result of transcriptional regulation (or that prompt changes in transcription). Ref. [7] predicted based on transcriptomic analyses that *R. pomeroyi* would import N-acetyltaurine, glycine betaine or proline, glutamate or aspartate, branched-chain amino acids, an amide or urea, altronate or galactarate, acetate, DHPS, and export bisulfite in

Feature	*Ruegeria pomeroyi* DSS-3
Model name	*i*Rpo1302
Genes	
Genome	4,370 [36]
Model	1,302 (29.8%)
Complexed	539
Reactions	
Reversible	656
Irreversible	1,039
Gene associated	1,244
Non-gene associated	451
Metabolic	961
Transport	484
Exchange	238
Demand	6
Biomass	6
ATP maintenance	1
Extracellular	239
Periplasm	245
Cytoplasm	1,211
Total	1,695
Blocked	188 (11.1%)
Orphaned	2
Metabolites	
Extracellular	240
Periplasm	244
Cytoplasm	915
Total	1,399
Unique	919
Dead-ends	202 (14.4%)

Table 3.1: Summary of *i*Rpo1302 model properties

co-culture based on upregulated transport genes and downstream metabolic genes that were differentially regulated. The z-score calculations indicated that these were some of the highest scoring metabolites in the network (Supplementary Data Set C.2). Ref. [7] directly measured a higher concentration of some metabolites when *R. pomeroyi* was in co-culture including, L-cysteate, sulfolactate, sn-glycerol-3-phosphate, UDP-glucose, L-phenylalanine, guanosine-3-phosphate, dGMP, D-gluconic acid, 2-dehydro-D-gluconate, UMP, CMP. All of these metabolites had high z-scores with the exceptions of UDP-glucose, D-gluconic acid, 2-dehydro-D-gluconate. Myo-inositol, and D-sedoheptulose-1,7-phosphate were also detected in the co-culture study; these metabolites are not part of the *i*Rpo1302 model because there is limited genetic evidence connecting inositol phosphate metabolism to the rest of the model and no genetic evidence for the production of D-sedoheptulose-1,7-phosphate in *R. pomeroyi*. These analyses also identified additional metabolites with high z-scores that are predicted to either accumulate within, or be excreted by, *R. pomeroyi* based on differential transcription patterns. These metabolites include N-acetyltaurine, 2-oxo-3-hydroxy-propane-1-sulfonate, bisulfite, allophanate, D-altronate, branched chain amino acids, UDP-N-acetyl-glucosamine and other peptidoglycan or lipopolysaccharide biosynthesis intermediates, purine metabolism intermediates, adenosylcobyrinate-*a,c*-diamide, and adenosylcobalamin (Figure 3.1A, Supplementary Data Set C.2). Some, but not all, of these metabolites were targeted for measurement in ref. [7]; those that were measured did not have a significantly different concentration between treatments. Some metabolites are predominantly associated with genes down-regulated in co-culture (Figure 3.1A), including carbon monoxide, CO_2, sulfite, glycolate, and glycerol. Carbon monoxide and possibly glycerol (from bacterial glycerol stocks) may be the main sources of carbon for *R. pomeroyi* in mono-culture. In the absence of organic sulfur from *T. pseudonana*, *R. pomeroyi* is likely utilizing more oxidized forms of sulfur.

Ref. [7] predicts that in co-culture, *T. pseudonana* upregulates vitamin B_{12}, lipid, and amino acid uptake; upregulates chitin utilization, resulting in production of UDP-N-acetyl-glucosamine; and downregulates sulfate uptake. Ref. [7] only measured metabolites in *T.*

pseudonana at the 56-h time point when more genes were differentially regulated. Z-scores were generated with transcription data for *T. pseudonana* collected at the 8-h and 56-h time points. After 8 h, corrected z-scores ranged from 3.0 for extracellular cyanocobalamin to -4.9 for N-acetyl-D-glucosamine with a median score of -0.04. Reporter metabolites in the 90^{th} percentile include cyanocob(III)alamin and aquacob(III)alamin, choline, urea, biotin, sulfate, plastidial bicarbonate and other intermediates in the Calvin-Benson cycle, amino acids, some precursors for L-methionine biosynthesis including L-serine, L-homocysteine, a putative product of DHPS degradation glycerol-3-phosphate [7], a carbohydrate metabolism precursor UDP-alpha-D-galactose, intermediates in branched-chain amino acid degradation, S-adenosyl-homocysteine, intermediates in ubiquinone biosynthesis, some lipids including triacylglycerols, 1,2-diacyl-sn-glycerols, and fatty acids (Figure 3.1B, Supplementary Data Set C.2).

After 56 h in co-culture, corrected z-scores for *T. pseudonana* ranged from 3.7 for plastid-localized H^+ to -5.3 for mitochondrial succinyl-CoA with a median score of -1.2. The transcriptional response of *T. pseudonana* to co-culture conditions is much stronger after 56 h compared to the 8-h time point. Only the B_{12} acquisition gene CBA1 was significantly upregulated at both time points; after 56 h, there was downregulation of sulfo/glycerolipid and fatty acid biosynthesis, downregulation of carbomoyl phosphate synthase, and upregulation of chitinase and the UDP-N-acetyl-glucosamine transporter [7]. Vitamin B_{12} was not among the top-scoring metabolites after 56h, though amino acids, lipids, sulfate, and UDP-N-acetyl-glucosamine were. Ref. [7] found L-homocysteine and D-gluconic acid concentrations to be significantly lower in co-culture, while L-homocysteine had a low z-score and D-gluconic acid was not present in *i*Tps1426 due to lack of genetic evidence for the biosynthesis of this metabolite. Other metabolites with high z-scores in this analysis include oxygen and hydrogen peroxide, plastidial oxygen, CO_2, PSII-associated metabolites, pigments including chlorophyll *a*, terpenoids, and carotenoids, carbamoyl phosphate, S-adenosyl-methionine, S-adenosyl-homocysteine and ubiquinone biosynthesis intermediates, pentose phosphate intermediates, branched chain amino acids and their degradation products, lipids, UMP, GTP,

and TCA cycle intermediates (Figure 3.1C, Supplementary Data Set C.2).

3.3.3 Simulation of interaction between T. pseudonana and R. pomeroyi

The interaction between *T. pseudonana* and *R. pomeroyi* was simulated using dynamic Flux Balance Analysis (FBA). Genome-scale metabolic models of *T. pseudonana* and *R. pomeroyi* (*i*Tps1426 and *i*Rpo1302) represent metabolism by connecting a network of metabolites with all reactions known to occur in an organism. The existence of a reaction is supported by gene annotations and physiological evidence where available; genome-scale models can integrate a wealth of information on the metabolism of an organism in a way that is mathematically solvable. Flux distributions through a network can be calculated using Flux Balance Analysis by adding constraints on nutrient exchange reactions, and selecting an objective function (for example, maximize biomass production). This method assumes steady state conditions, optimal growth temperature, and fixed biomass composition. In dynamic FBA, the growth curve is split up into steady state chunks and the resulting fluxes determine nutrient uptake and metabolite secretion into the media. Condition-specific models were produced with the GIM^{3}E algorithm [39]. GIM^{3}E creates irreversible Mixed Integer Linear Programming (MILP) models and associates penalty coefficients with each reaction. Penalties denoting the level of transcription were mapped to each reaction using the associated genes in the metabolic model, ensuring that pathways with higher levels of transcription (or lower penalties) were active. Co-culture conditions consisted of growth on a 16-h light/8-h dark cycle at 160 μmol photons m^{-2} s^{-1} in $f/2$ media, in the absence of vitamin B_{12}. As a control, a mono-culture of *T. pseudonana* was simulated in the same conditions with the addition of external cyanocob(III)alamin. The growth curve was divided into seventy 4-h steady state intervals and the uptake rate of nutrients or metabolites for each model was calculated. Flux Balance Analysis was performed for each model at each time point, and the resulting specific growth rate and metabolite flux predictions were used to update biomass and metabolite concentrations in the medium.

Cell concentration from ref. [7] was converted to estimates of biomass concentration using

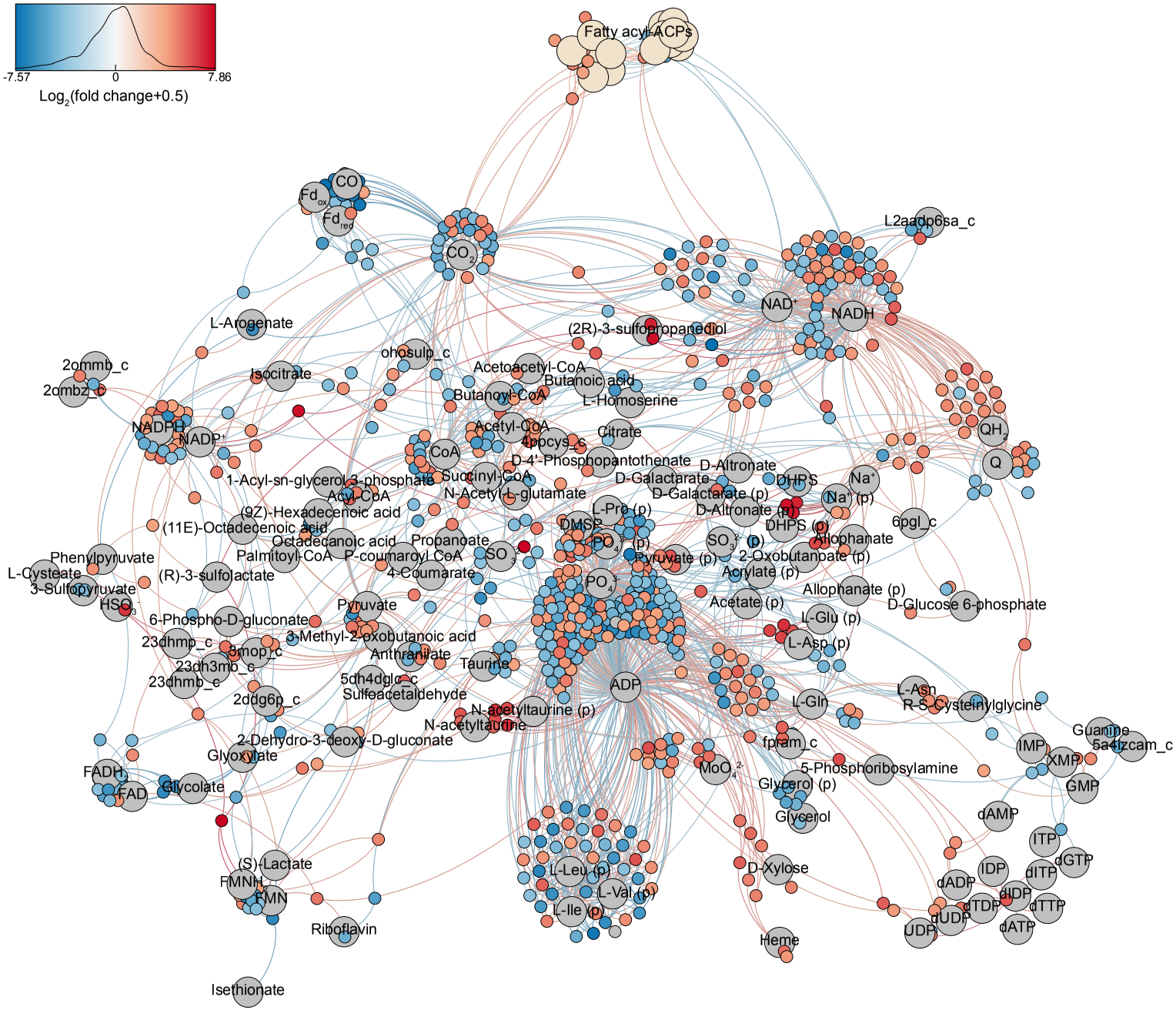
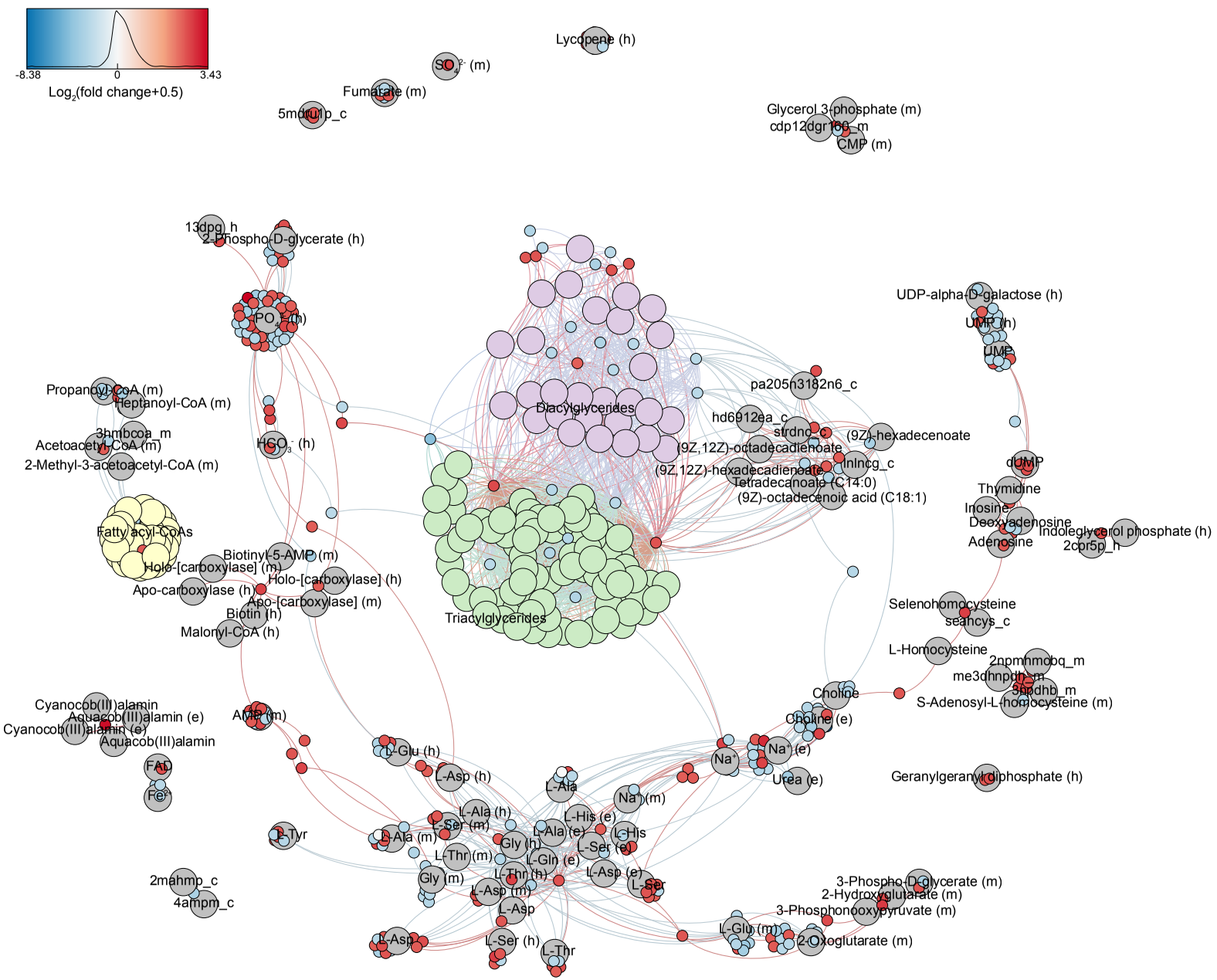

-7.57
0
7.86
Log2(fold change+0.5)
Fatty acyl-ACP$
CO
Fdox
Fdred
CO2
L2aadp6sa_c
NAD+
NADH
L-Arogenate
(2R)-3-sulfopropanediol
ohosulp_c
2ommb_c
2ombz_c
Isocitrate
Acetoacetyl-CoA
Butanoic acid
Butanoyl-CoA
L-Homoserine
NADPH
NADP+
Acetyl-CoA
Citrate
4ppcys_c
CoA
D-4'-Phosphopantothenate
QH2
Q
1-Acyl-sn-glycerol 3-phosphate
Succinyl-CoA
D-Altronate
D-Galactarate
DHPS
Na+
Acyl-CoA
N-Acetyl-L-glutamate
D-Galactarate (p)
Na+ (p)
(9Z)-Hexadecenoic acid
L-Pro (p)
D-Altronate (p)
6pgl_c
(11E)-Octadecenoic acid
DMSP
DHPS (p)
Octadecanoic acid
Propanoate
PO4 3- (p)
SO3 2- (p)
Allophanate
Phenylpyruvate
Palmitoyl-CoA
P-coumaroyl CoA
SO3 2-
Pyruvate (p)
2-Oxobutanoate (p)
L-Cysteate
(R)-3-sulfolactate
4-Coumarate
Acrylate (p)
3-Sulfopyruvate
PO4 3-
Allophanate (p)
HSO3-
Acetate (p)
D-Glucose 6-phosphate
Pyruvate
6-Phospho-D-gluconate
L-Glu (p)
3-Methyl-2-oxobutanoic acid
23dhmp_c
L-Asp (p)
3mop_c
Anthranilate
Taurine
23dh3mb_c
23dhmb_c
5dh4dglc_c
ADP
L-Asn
Sulfoacetaldehyde
L-Gln
R-S-Cysteinylglycine
2ddg6p_c
N-acetyltaurine (p)
N-acetyltaurine
Guanine
5a4izcam_c
2-Dehydro-3-deoxy-D-gluconate
fpram_c
IMP
Glyoxylate
MoO4 2-
5-Phosphoribosylamine
XMP
FADH2
Glycerol (p)
GMP
FAD
Glycolate
Glycerol
dAMP
ITP
dGTP
(S)-Lactate
L-Leu (p)
D-Xylose
IDP
dADP
dITP
FMNH2
L-Val (p)
dIDP
FMN
L-Ile (p)
dTDP
dTTP
dUDP
Riboflavin
Heme
dUTP
dATP
Isethionate

Figure 3.1


-8.38
0
3.43
Log2(fold change+0.5)
Lycopene (h)
SO4 2- (m)
Fumarate (m)
5mdu1p_c
Glycerol 3-phosphate (m)
cdp12dgr160_m
CMP (m)
13dpg_h
2-Phospho-D-glycerate (h)
PO4 3- (h)
Propanoyl-CoA (m)
Heptanoyl-CoA (m)
3hmbcoa_m
Acetoacetyl-CoA (m)
2-Methyl-3-acetoacetyl-CoA (m)
HSO3- (h)
Fatty acyl-CoAs
Biotinyl-5-AMP (m)
Holo-[carboxylase] (m)
Holo-[carboxylase] (h)
Apo-carboxylase (h)
Apo-[carboxylase] (m)
Biotin (h)
Malonyl-CoA (h)
Diacylglycerides
Triacylglycerides
pa205n3182n6_c
hd6912ea_c
strdnc_c
(9Z)-hexadecenoate
(9Z,12Z)-octadecadienoate
(9Z,12Z)-hexadecadienoate
lnlncg_c
Tetradecanoate (C14:0)
(9Z)-octadecenoic acid (C18:1)
UDP-alpha-D-galactose (h)
UMP (h)
UME
dUMP
Thymidine
Inosine
Deoxyadenosine
Adenosine
Indoleglycerol phosphate (h)
2cpr5p_h
Selenohomocysteine
seahcys_c
L-Homocysteine
2npmhmobq_m
me3dhnpdh_m
3mpdhb_m
S-Adenosyl-L-homocysteine (m)
Choline
Choline (e)
Na+ (e)
Urea (e)
Geranylgeranyl diphosphate (h)
Cyanocob(III)alamin
Aquacob(III)alamin (e)
Cyanocob(III)alamin (e)
Aquacob(III)alamin
FAD
Fe2+
AMP (m)
L-Glu (h)
L-Asp (h)
L-Ala
Na+ (m)
L-Ala (h)
L-Ser (m)
L-His (e)
L-Ala (e)
L-His
L-Ala (m)
Gly (h)
L-Ser (e)
L-Thr (m)
L-Gln (e)
L-Asp (e)
Gly (m)
L-Thr (h)
L-Ser
L-Asp (m)
L-Asp
L-Asp
L-Ser (h)
L-Thr
L-Tyr
2mahmp_c
4ahmpm_c
3-Phospho-D-glycerate (m)
2-Hydroxyglutarate (m)
3-Phosphonooxypyruvate (m)
L-Glu (m)
2-Oxoglutarate (m)

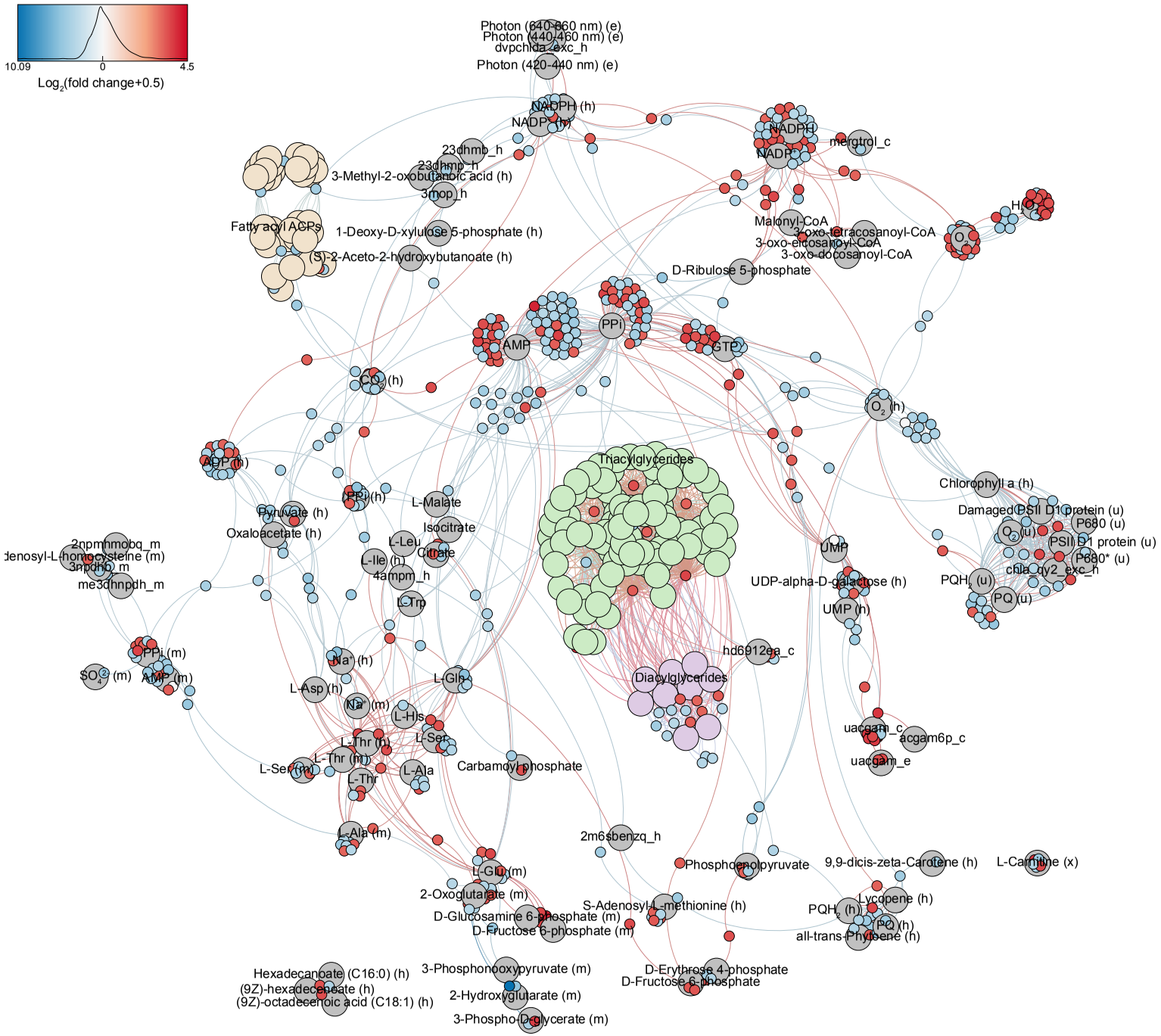

10.09
0
4.5
Log2(fold change+0.5)
Photon (640-660 nm) (e)
Photon (440-460 nm) (e)
dvpchlca_exc_h
Photon (420-440 nm) (e)
NADPH (h)
NADP+ (h)
NADPH
NADP+
mergtrol_c
23dhmb_h
23dhmp_h
3-Methyl-2-oxobutanoic acid (h)
3mop_h
Fatty acyl ACPs
1-Deoxy-D-xylulose 5-phosphate (h)
(S)-2-Aceto-2-hydroxybutanoate (h)
Malonyl-CoA
3-oxo-tetracosanoyl-CoA
3-oxo-eicosanoyl-CoA
3-oxo-docosanoyl-CoA
H2O2
O2
D-Ribulose 5-phosphate
PPi
AMP
GTP
CO2 (h)
O2 (h)
ADP (h)
Triacylglycerides
Chlorophyll a (h)
Damaged PSII D1 protein (u)
P680 (u)
O2 (u)
PSII D1 protein (u)
P680* (u)
chla_qy2_exc_h
PQH2 (u)
PQ (u)
UMP
UDP-alpha-D-galactose (h)
UMP (h)
(PPi) (h)
L-Malate
Isocitrate
Pyruvate (h)
Oxaloacetate (h)
L-Leu
Citrate
L-Ile (h)
4ampm_h
L-Trp
2ppmhmobq_m
denosyl-L-homocysteine (m)
3mpdhb_m
me3dhnpdh_m
PPi (m)
SO4 (m)
AMP (m)
Na+ (h)
L-Asp (h)
L-Gln
Na+ (m)
L-His
L-Ser
L-Thr (m)
L-Thr (m)
L-Ser (m)
L-Ala
L-Thr
Carbamoyl phosphate
L-Ala (m)
hd6912ea_c
Diacylglycerides
uacgam_c
acgam6p_c
uacgam_e
2m6sbenzq_h
L-Glu (m)
2-Oxoglutarate (m)
D-Glucosamine 6-phosphate (m)
D-Fructose 6-phosphate (m)
Phosphoenolpyruvate
S-Adenosyl-L-methionine (h)
9,9-dicis-zeta-Carotene (h)
L-Carnitine (x)
Lycopene (h)
PQH2 (h)
PQ (h)
all-trans-Phytoene (h)
Hexadecanoate (C16:0) (h)
(9Z)-hexadecenoate (h)
(9Z)-octadecenoic acid (C18:1) (h)
3-Phosphonooxypyruvate (m)
2-Hydroxyglutarate (m)
3-Phospho-D-glycerate (m)
D-Erythrose 4-phosphate
D-Fructose 6-phosphate

Figure 3.1: (Previous page.) Network of reporter metabolites with z-scores in the 90^{th} percentile and genes associated with each metabolite for (a) *R. pomeroyi* after $8h$, (b) *T. pseudonana* after $8h$, and (c) *T. pseudonana* after $56\,h$ in co-culture versus mono-culture. Small circles indicate neighboring genes and are coloured according $\log_2$(fold-change + 0.5); the color bar indicates the range of values as well as the density distribution of genes. Red indicates increased transcript abundance in co-culture and blue indicate decreased transcript abundance in co-culture. Metabolites are indicated by larger circles, those located in compartments other than the cytosol labeled with abbreviations in parentheses for each organelle. Metabolites associated with fatty acid/lipid metabolism are colored according to class: fatty acyl ACPs (brown), fatty acyl CoAs (yellow), diacylglycerols (pink), and triacylglycerols (green). Hydrogen and water are not shown. The corresponding metabolite names for each abbreviation can be found at http://bigg.ucsd.edu or in Supplementary Data Set B.1 or C.1. Metabolites located in compartments/organelles other than the cytosol are indicated with the following abbreviations: periplasm ('p'), extracellular ('e'), mitochondria ('m'), chloroplast ('h'), thylakoid lumen ('u'), peroxisome ('x').

cell dry weight measurements from the literature for both *T. pseudonana* and *R. pomeroyi*. *T. pseudonana* was starved for vitamin B_{12} over three sequential transfers prior to the start of the experiment; the experimental growth curves of *T. pseudonana* upon addition of B_{12} in mono-culture did not demonstrate exponential growth and had relatively low cell yields. Recuperation from growth-limiting conditions is not steady state (violating one of the FBA assumptions) and likely involves shifts in biomass composition. Initial simulations predicted a much higher growth rate (represented by the 'DM_biomass_c' reaction) for *T. pseudonana* than the experiment predicted. To try and improve the accuracy of the simulation, the upper bound of *i*Tps1426 biomass production was set to the maximum observed growth rate in the experiment. Similarly, the bounds of *i*Rpo1302 biomass production were set to the minimum and maximum observed growth rates in the experiment. Nevertheless, *i*Tps1426 reached a higher biomass yield in mono-culture than experimental cell counts of

T. pseudonana would predict. But because the steady state assumption was violated, the cell dry weight measurement likely fluctuates over the course of the experiment, so the accuracy of the diatom biomass estimate is not known. In co-culture, growth of *i*Tps1426 is suppressed by vitamin B_{12} limitation and never reaches a plateau, while the experimental data suggest that *T. pseudonana* should have a similar growth rate in co- and mono-culture but reach a lower final yield in co-culture. The *R. pomeroyi* growth curve shows high initial growth rates until the 32-*h* time point and slower linear growth thereafter. The growth rate of *i*Rpo1302 is slow throughout the simulation and does not reach a plateau (Figure 3.2).

Under the transcriptome-constrained conditions using the GIM^{3}E algorithm, *i*Rpo1302 did not excrete any form of B_{12} despite taking up Co^{2+} from the media to synthesize vitamin B_{12} for its own biomass production. Results from the GIM^{3}E algorithm are known to be qualitatively but not quantitatively accurate [40], because the objective of minimizing the penalty reaction results in a flux distribution where only the most transcriptionally upregulated reactions are used, eliminating flux through alternative pathways with less of a transcription signal. Therefore, maximal aquacob(III)alamin production ('EX_aqcobal_e') was fixed as a constraint and where iRpo1302 was forced to produce 0.0001 of the optimal value.

In these simulated interactions, the growth of *i*Rpo1302 is dependent on organic nitrogen and NH_4 as well as other sources of organic carbon and sulfur produced by *i*Tps1426. Changes in concentration of important *T. pseudonana* biomass components were also monitored in the simulation. Growth of *i*Tps1426 is dependent on the biosynthesis and secretion of aquacob(III)alamin by *i*Rpo1302. In each simulation, *i*Tps1426 secretes a variety of different forms of reduced and organic nitrogen: NH_4, urea, L-asparagine, L-aspartate, L-histidine, L-alanine, L-glutamate, L-glutamine, glycine, L-leucine, L-serine, L-threonine, formamide, urate, xanthine, choline, glycine betaine, UDP-N-acetyl-glucosamine, and folate (Figure 3.3). Over the course of the simulation, *i*Rpo1302 takes up all of these metabolites except NH_4, glycine, L-threonine, formamide, and folate. *i*Rpo1302 does not include transporters for L-threonine or folate, although it does include a transporter for the folate precur-

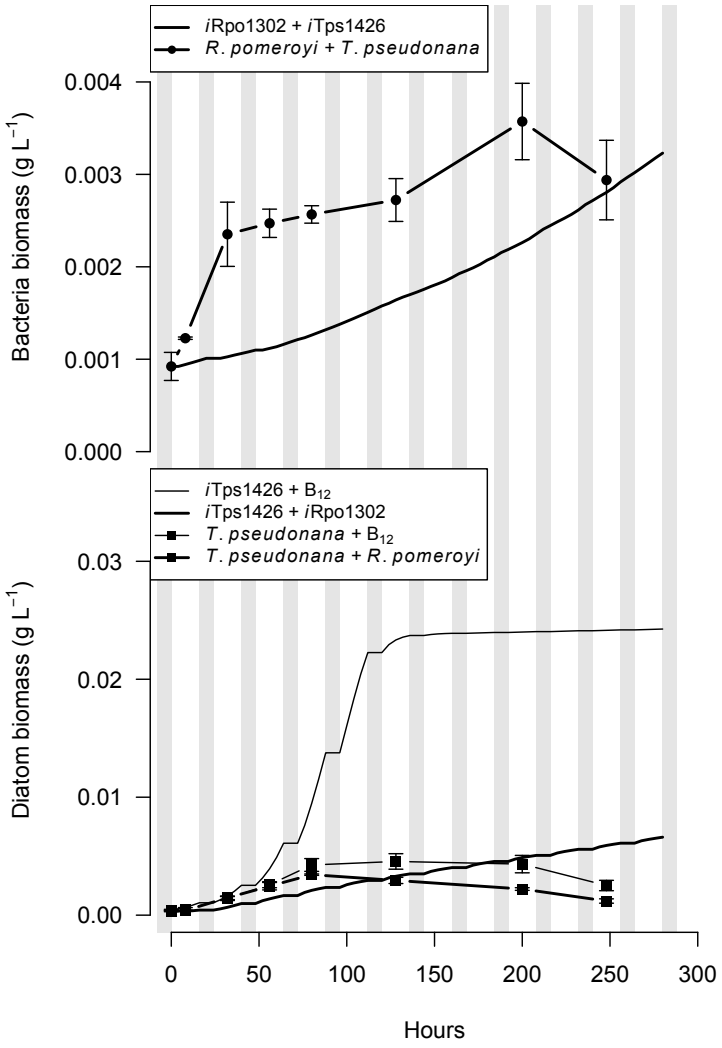


Figure 3.2: Simulated growth curves of *T. pseudonana* in co-culture with *R. pomeroyi* in the absence of vitamin B_{12} and in mono-culture with vitamin B_{12} over a 16:8-hour light/dark cycle at 160 μmol photons m^{-2} s^{-1} in $f/2$ media. These simulations were compared to experimental data under the same growth conditions. Experimental data was converted from cell concentration to biomass using cellular dry weight measurements from ref. [41] and [42]. Experimental data points are plotted with circles for *R. pomeroyi* and squares for *T. pseudonana*. Simulated and experimental data for cells in co-culture are connected with thicker lines, and data for cells in mono-culture with thinner lines.

sors *p*-aminobenzoate and chorismate. In co-culture, *i*Tps1426 produces the organic sulfur compounds DHPS and DMSP. *i*Rpo1302 consumes both DHPS and DMSP and secretes the metabolic byproducts of the DMSP cleavage pathway, CH_4S, as well as an intermediate of DHPS catabolism, 3-sulfopyruvate. Tetrathionate is also secreted by *i*Rpo1302. *i*Tps1426 produces a variety of carbohydrates in co-culture including formate, glycolate, ethanol, and acetate which is consumed by *i*Rpo1302. In co-culture, there is also biosynthesis of triacylglycerols (TAGs) by *i*Tps1426.

In mono-culture there was increased nitrate and phosphate utilization, likely due to the lack of growth-limiting conditions. 1,3-Glucan, a component of the carbohydrate storage molecule chrysolaminarin, and chitin were produced more in mono-culture than in co-culture. There is some utilization of membrane lipids and triacylglycerols in mono-culture. Many secreted metabolites reached a higher concentration in mono-culture, likely because they are not consumed by *i*Rpo1302, but there is also differential production of metabolites due to transcriptional constraints and nutrient conditions. In mono-culture there are higher concentrations of L-asparagine, L-aspartate, L-histidine, L-alanine, L-glutamate, and L-leucine, while in co-culture there are higher concentrations of L-glutamine, glycine, L-serine, and L-threonine in the media. Additionally, there are higher concentrations of urea, UDP-N-acetyl-glucosamine, folate, and DHPS in co-culture, but higher concentrations of formamide, urate, xanthine, choline, and DMSP in mono-culture (Figure 3.3).

Many of the metabolites secreted in either condition are associated with differentially regulated genes and had high z-scores in the previous analysis. After $8\,h$ in culture, triacylglycerols and lipids have high z-scores in the *T. pseudonana* network and are predominantly associated with genes that are downregulated in co-culture; these metabolites are connected to choline. Phosphatidylcholines are also relatively high scoring (though not in the 90^{th} percentile) and choline has a high z-score and appears to be associated with downregulated genes in co-culture (Figure 3.1B). In the mono-culture simulation, there is degradation of lipids and TAGs resulting in choline excretion (Figure 3.3). Homocysteine, S-adenosylhomocysteine (SAH), and adenosine, have a high z-score in the *T. pseudonana* network

(Figure 3.1B). Homocysteine is a precursor to methionine and DMSP biosynthesis; SAH and adenosine are connected to methyltransferase reactions and the regeneration of S-adenosyl-methionine (Figure 3.4). If B_{12}-depletion is blocking methionine synthase and homocysteine utilization, then DHPS may be secreted as an alternative to DMSP (Figure 3.3). DHPS and its degradation products are some of the highest scoring metabolites in the *R. pomeroyi* network (Figure 3.1A) and in metabolite measurements [7]. Similarly, the availability of adenosine could result in an abundance of purines, which could be utilized in the synthesis and excretion of UDP-N-acetyl-glucosamine, another high scoring metabolite (Figure 3.1B, 3.1C, Figure 3.4). The excretion of folate in co-culture could also arise as a response to B_{12} limitation (Figure 3.3, Figure 3.4). Finally, many of the amino acids secreted in either condition have high z-scores in all of the networks.

3.4 Discussion

In Chapter 2, I presented the genome-scale metabolic model of *Thalassiosira pseudonana* CCMP 1335, a well-studied diatom that has been important in the field of marine microbiology for the last 60 years. Here, I reconstructed the genome-scale model of *Ruegeria pomeroyi* DSS-3, a heterotrophic marine bacterium isolated from US coastal waters in 1998 using DMSP as a substrate. DSS-3 was sequenced in 2004 [43]; it is a member of the Roseobacter clade and was found to grow on a variety of algal-derived metabolites as well as inorganic compounds in seawater. Experimental work by ref. [7] has demonstrated that *R. pomeroyi* can support the growth of *T. pseudonana* deprived of cobalamin. In return, *T. pseudonana* assimilates nitrate and provides a source of organic nitrogen that *R. pomeroyi* can utilize for growth. Transcriptomics data showing the effects of interaction, coupled with targeted measurement of metabolites, have demonstrated that the diatom produces an organic sulfur molecule, 2,3-dihydroxypropane-1-sulfonate (DHPS), which is catabolized by *R. pomeroyi* [28]. Growth of the bacterium is likely supported by a variety of compounds including sulfonates, sugar derivatives, and organic nitrogen [7]. *T. pseudonana* also exhibits a transcriptional response in co-culture that may be analogous to plant recognition signal

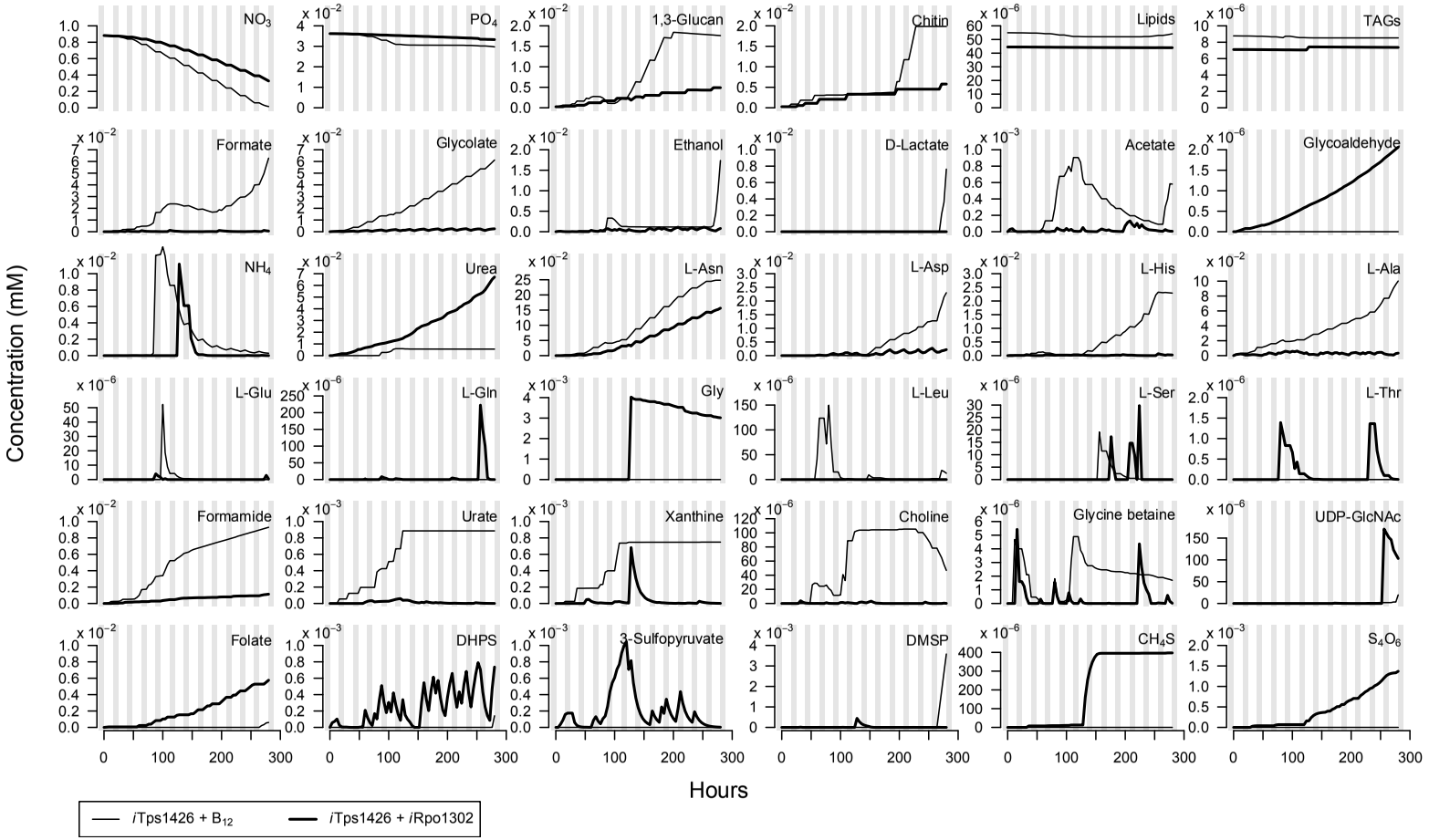


Figure 3.3: Dynamic Flux Balance Analysis simulation of cellular or extracellular metabolite concentration in the media for *i*Tps1426 grown in co-culture with *i*Rpo1302 (thicker lines) and *i*Tps1426 grown in mono-culture with vitamin B_{12} added (thinner lines). The concentrations of different metabolites vary by several orders of magnitude; note the magnification of the y-axes for different metabolites.

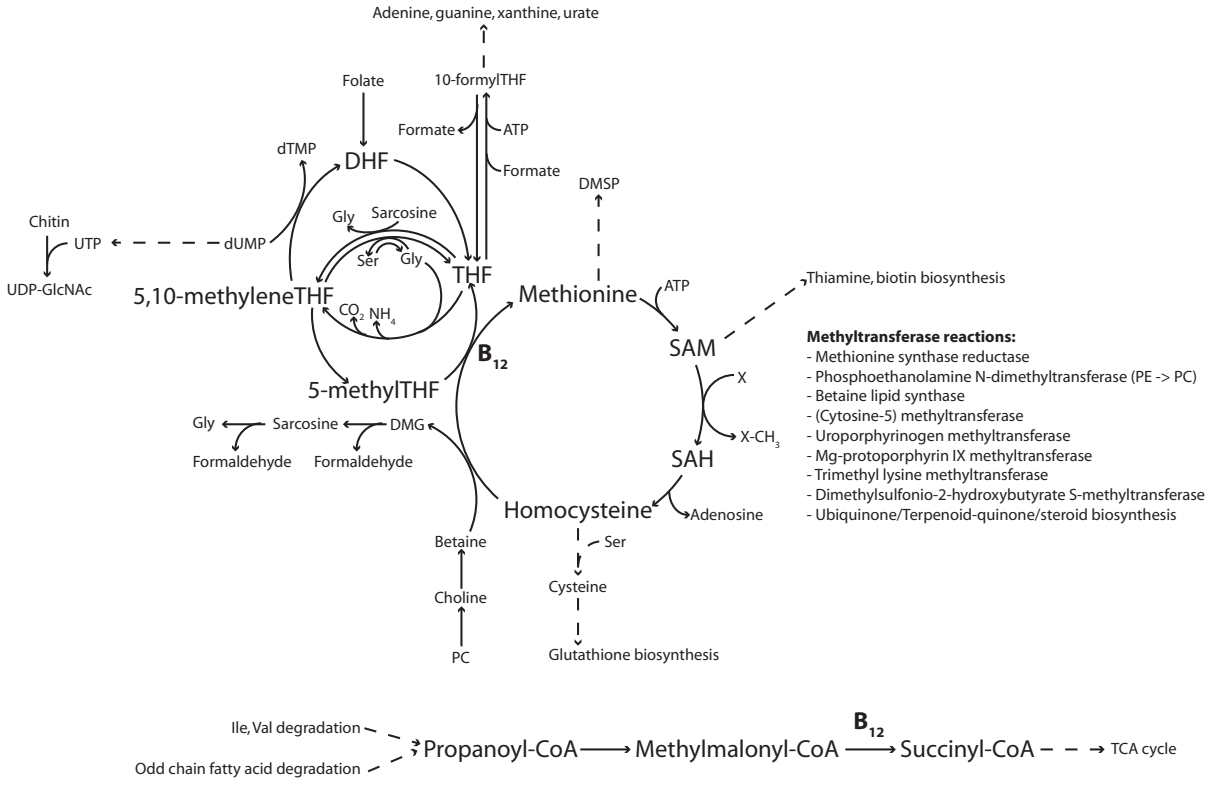


Figure 3.4: Diagram demonstrating the impacts of B_{12}-dependent methionine synthase and methylmalonyl-CoA mutase on diatom metabolism.

cascades [7]. This trophic nutrient exchange may be emblematic of diatom-bacteria interactions in parts of the ocean where diatoms are dependent on bacteria for cobalamin [19]. The combined usage of transcriptomics and metabolomics is powerful and allows us to infer which metabolites may be exchanged between organisms. The addition of network information and genome-scale metabolic modeling simulations allows us to predict the changes in flux over time and to understand the causes of metabolite excretion.

The metabolic network topology of *i*Tps1426 and *i*Rpo1302 was used to re-analyze transcriptomics data from ref. [7]. Reporter metabolites are defined as those involved in reactions associated with genes that are differentially regulated in the two treatment conditions. After $8\,h$ in co-culture, reporter metabolites in the *T. pseudonana* network carry the signature of cobalamin limitation. In humans, vitamin B_{12} limitation is identified by high blood levels of homocysteine and methylmalonic acid; it can also imitate the effects of folate deficiency due to the role of methionine synthase in the folate cycle [44]. Some of the top reporter metabolites (cobalamin, L-homocysteine, S-adenosyl-L-homocysteine) are part of the methionine cycle, while other metabolites are by-products of the cycle or are involved in reactions that require the cofactor S-adenosyl-L-methionine (Figure 3.4). Adenosine is a byproduct of homocysteine formation from SAH, choline (a breakdown product of phosphatidylcholine) is a precursor of glycine betaine which together with homocysteine forms dimethylglycine and methionine. The compounds 30-nonaprenyl-4,5-dihydroxybenzoate ('3npdhb_m'), 3-nonaprenyl-4-hydroxy-5-methoxybenzoate ('me3dhnpdh_m'), 2-decaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinone ('2npmhmobq_m'), and 2-nonaprenyl-3-methyl-6-methoxy-1,4-benzoquinone ('me2np6mobq_m') are metabolic intermediates of methyltransferase reactions for ubiquinone biosynthesis catalyzed by enzymes with SAM as a co-substrate. 4-amino-2-methyl-5-phosphomethylpyrimidine ('4ampm_c') is a metabolic intermediate in the biosynthesis of thiamine from SAM. Other reporter metabolites are involved in reactions with folate cofactors including glycine, L-serine, and dUMP. There are also reporter metabolites associated with dysfunction of methylmalonyl-CoA mutase, including propanoyl-CoA, the odd chain fatty acid heptanoyl-CoA, the valine degradation product 2-methyl-3-oxopropanoate,

and the isoleucine degradation product 2-methyl-3-acetoacetyl-CoA. Many of the differentially regulated genes of *T. pseudonana* in co-culture correspond to those that are regulated by B_{12} [45].

For the *R. pomeroyi* metabolic network, reporter metabolites after 8 h in culture are compounds identified in ref. [7] that *T. pseudonana* has excreted into the media and their downstream catabolic products, sugar derivatives that could have been exopolysaccharide components from the diatom, as well as metabolites that are more likely to be sources of carbon for *R. pomeroyi* in mono-culture (CO, glycerol). Many of the reporter metabolites that appear to transfer from the diatom to *R. pomeroyi* could be the by-products of cobalamin limitation. When methylmalonyl-CoA mutase is inhibited, isoleucine and valine degradation cannot proceed perhaps causing the diatom to excrete these amino acids. Although N-acetyltaurine is not included in *i*Tps1426 due to limited knowledge about its biosynthesis, its precursor is likely cysteine, which could be available in excess in the diatom due to the inhibition of methionine synthase. DHPS and its downstream catabolism products are some of the metabolites with the highest z-scores in *R. pomeroyi*. DMSP biosynthesis may be inhibited in *T. pseudonana* due to limited methionine availability, and DHPS excretion may replace DMSP as a mechanism to rid the cell of excess sulfur (Chapter 2). UDP-N-acetyl-glucosamine, a breakdown product of chitin, is one of the top differentially regulated metabolites in *R. pomeroyi*. UDP-N-acetyl-glucosamine could be a source of nucleotides for the bacterium, of which GTP is heavily involved in cobalamin biosynthesis.

After 56h in culture, the *T. pseudonana* reporter metabolites have less of a cobalamin limitation signature, though SAM and SAH and some of their co-substrates are still high scoring. Other high scoring metabolites include TCA cycle and branched chain amino acid intermediates. I hypothesize that this transcriptional pattern is indicative of recovery from cobalamin limitation as metabolites that had accumulated previously are now being processed. Ref. [7] found that many differentially regulated genes at this time point are analogous to the plant signal recognition pathways and oxidative burst metabolism that is activated during a pathogen attack. The metabolic signature of this response includes metabolites such as

NADPH, O_2, H_2O_2, and GTP. NADPH oxidase consumes O_2 and produces superoxide radicals that act as signals or can kill invading pathogens. Guanylate cyclase synthesizes the secondary messenger cGMP from GTP in response to changes in intracellular calcium levels.

Some of the expected metabolic exchanges between *T. pseudonana* and *R. pomeroyi* could be simulated using a combination of B_{12}-limiting conditions and transcriptional constraints. The excretion of DHPS and folate can be attributed to B_{12} limitation in *i*Tps1426. Ref. [7] postulated that the degradation of chitin and subsequent secretion of UDP-N-acetylglucosamine could be a defensive response against bacterial attachment to the diatom frustule. An alternative explanation emerging from this work is that B_{12}-limitation resulted in a buildup of dUMP, for which the production of UDP-N-acetylglucosamine is a sink (Figure 3.4).

Traditional models of marine biogeochemistry are low resolution when it comes to biology. Only recently have we begun to distinguish between the roles of different phytoplankton [46]. The development of genome-scale models for representative marine microorganisms may be key to increasing the biological resolution of ecosystem models down to the molecular level. To date, four genome-scale models of marine microorganisms have been developed, including *Phaeodactylum tricornutum* CCAP 1055/1 [47], *Prochlorococcus marinus* MED4 [48], *Dinoroseobacter shibae* DFL12T [49], and *Pseudoalteromonas haloplanktis* TAC125 [50]. To this pool I have added two more models, *Thalassiosira pseudonana* CCMP 1335 (Chapter 2), a well-studied diatom that is auxotrophic for vitamin B_{12}, and *Ruegeria pomeroyi* DSS-3, a B_{12}-producing member of the Roseobacter clade that is known to consume a broad range of algal-derived metabolites. Modeling the interactions of key players in different marine biomes at the genomic level could provide more accurate predictions of dissolved organic matter characteristics and residence times in the ocean.

3.5 Materials and methods

3.5.1 Reconstruction and curation of an R. pomeroyi model

A model (*i*Rpo1302) of *Ruegeria pomeroyi* DSS-3 was generated using a previously reconstructed model of *Rhodobacter sphaeroides* 2.4.1 [26] as a template. The *R. pomeroyi* proteome [43] and updated annotations [36] were downloaded from Roseobase (http://www.roseobase.org). The *R. sphaeroides* proteome was downloaded from NCBI (acc no.: NC_007493.2). The metabolite and reaction IDs in *i*Rsp1140 were converted into BiGG IDs [51], and OrthoMCL [52] was used to identify orthologous gene clusters for these two organisms. Reactions in *i*Rsp1140 that contained gene orthologs in *R. pomeroyi* were retained, and reactions with no gene orthologs were deleted. In keeping with BiGG convention [51], reactions and metabolites in *i*Rpo1302 were assigned to three different compartments: cytosol ('c'), periplasm ('p'), or extracellular ('e'). These assignments were retained from *i*Rsp1140 where gene orthologs were detected. The guidelines from an established genome-scale reconstruction protocol were used to implement and refine *i*Rpo1302 [53]. The gene annotations from Roseobase were used to verify gene-protein-reaction associations and detect missing genes, reactions, and pathways. The annotations were also used to verify or edit cofactor usage in each *R. pomeroyi* reaction, and the BiGG database (http://bigg.ucsd.edu) was used to determine the likely reaction direction from other bacterial models. When cofactor-binding annotations were ambiguous, BRENDA and BiGG were used to look for the most common cofactor usage or the *i*Rsp1140 reaction was retained.

A biomass objective function was developed for *i*Rpo1302 using biomass composition measurements from the literature and from data derived from the genome sequence. The biomass reaction constitutes DNA, RNA, protein, phospholipids, and a soluble pool of polyamines and cofactors (Supplementary Data Set C.3). Measurements derived from the literature were typically taken from continuous culture conditions where *R. pomeroyi* was carbon-limited.

Dead-end metabolites were identified as metabolites that were only present in blocked

reactions; blocked reactions were identified from the flux analysis module in COBRApy and the gapfilling module was used to identify gaps in the network [54]. Gaps were filled if a gene for the missing reaction could be identified, if there was physiological evidence in the literature that the reaction exists, if the majority of the pathway was otherwise present in the model, or if the reaction was required to produce biomass. Demand reactions from *i*Rsp1140 were carried over into *i*Rpo1302 if synthesis of those metabolites were maintained. The non-growth associated ATP maintenance (NGAM) cost calculated for *R. sphaeroides* in low light conditions [55] were used in *i*Rpo1302.

The *i*Rpo1302 model was tested for growth on the substrates studied in refs. [56, 57, 30, 31, 32, 33, 58, 28]. If *i*Rpo1302 did not grow, the reaction network was examined and corrected. *i*Rpo1302 was tested for Type III pathways [53] with ExPa [59], using a Python script modified from ref. [60]. Extreme pathways were resolved either by investigating the evidence for the existence of each reaction, or reaction direction. *i*Rpo1302 is available in SBML format in the supplementary material (Supplementary File C.1) and will be deposited in the BioModels database.

3.5.2 Transcriptional regulation of metabolism

RNA-seq data from ref. [7] was downloaded from NCBI (acc no.: PRJNA261079). Transcripts from each sample were quantified with kallisto, a program that pseudoaligns reads to reference genes [61], and sleuth [62] was used to calculate p-values, $\log_2(\mathrm{FC} + 0.5)$, and median transcripts per million for each gene. P-values for each gene were converted to z-scores using the inverse normal cumulative distribution of 1-p. The network topology of *i*Rpo1302 and *i*Tps1426 was used to identify highly regulated (aka reporter) metabolites [37]. Highly regulated metabolites are defined as compounds around which transcriptional changes are most likely to occur.

To calculate the z-score of each metabolite, the network was represented as a metabolic graph where metabolites and genes are nodes, and the edges represent the link between enzymes catalyzing a reaction and the metabolites involved. The z-score of each metabolite

Z_{met} was calculated as:

$$Z_{met} = \frac{1}{\sqrt{k}} \sum Z_{ni} \tag{3.1}$$

where k is the number of neighbouring enzymes and Z_{ni} is the z-score for neighbouring enzymes of a metabolite. These scores were corrected for the background distribution of the network by randomly choosing 100 sets of k enzymes from the metabolic graph and calculating the mean and standard deviation of k enzymes. The resulting noisy distributions were smoothed with the UnivariateSpline function from the SciPy package (similar to ref. [63], Supplementary Figure C.1) and each z-score was corrected by subtracting the mean and dividing by the standard deviation for k enzymes. Networks of metabolites with scores in the top 90^{th} percentile were plotted with their associated genes.

3.5.3 Transcriptomics constraints on metabolism

Condition-specific models constrained by the available transcriptomic data were created [7]. The *T. pseudonana* model (*i*Tps1426) discussed in Chapter 2 and the *R. pomeroyi* model (*i*Rpo1302) developed in this study were constrained with transcriptomics data for the 8-*h* and 56-*h* time points to create separate models for growth in co-culture or in diatom mono-culture with the addition of B_{12}. The GIM^{3}E (Gene Inactivation Moderated by Metabolism, Metabolomics and Expression) algorithm [39] was used to add penalty coefficients to each reaction in proportion with the level of transcription associated with the genes responsible and to create a mixed integer linear programming problem (MILP) that would make the models irreversible. The GIM^{3}E module was built for an outdated version of COBRApy and had to be updated to work with COBRApy 0.13.4 (Supplementary File C.1). In the original GIM^{3}E algorithm, reactions without an associated gene did not receive a penalty coefficient and as a result condition-specific models preferentially utilized reactions that were unannotated. The algorithm was modified by adding a penalty coefficient to reactions without an associated gene by calculating the mean penalty coefficient for the model and then adding this number to those reactions, unless they were known to be spontaneous,

exchange, demand, sink, or biomass production reactions. All transport reactions were given the mean penalty coefficient because our confidence in their associated gene annotation is quite low.

3.5.4 Dynamic Flux Balance Analysis

The dynamic Flux Balance Analysis method discussed in Chapter 2 was used to simulate interaction between GIM^{3}E constrained models or to simulate GIM^{3}E constrained growth of *i*Tps1426 in mono-culture. Growth simulations of *i*Rpo1302 and *i*Tps1426 were performed under a 16:8-hour light/dark cycle at 160 μmol photons m^{-2} s^{-1} in $f/2$ medium in the presence or absence of cyanocob(III)alamin (vitamin B_{12}). The growth period was divided into seventy 4-h intervals, and steady state was assumed for each time interval so that Flux Balance Analysis (FBA) could be used to solve for the growth rate and reaction flux distributions in each organism for every 4-h. The medium was initialized with a concentration of 1,3-glucan representing 23% of initial biomass [64], a concentration of chitin representing 15% of initial biomass [65, 66], and a concentration of lipids and TAGs representing 13% and 2% of initial biomass [42]. Each growth simulation was initialized with the cell concentrations of *T. pseudonana* and *R. pomeroyi* measured in ref. [7]. The growth simulation of *i*Tps1426 + B_{12} was initialized with a biomass concentration of $4.35 \cdot 10^{-4}$ gDW L^{-1}, or 33,477 diatoms mL^{-1} using the cell dry weight conversion from ref. [42] for cells grown at 100 μmol photons m^{-2} s^{-1} on continuous light harvested during logarithmic phase (13 pg cell^{-1}). The growth simulation of *i*Tps1426 + *i*Rpo1302 in co-culture was initialized with a biomass concentration of $3.53 \cdot 10^{-4}$ gDW L^{-1} for *i*Tps1426, or 27,116 diatoms mL^{-1}, and $9.22 \cdot 10^{-4}$ gDW L^{-1} for *i*Rpo1302, or 2,030,908 bacteria mL^{-1} using the cell dry weight conversion from ref. [41] for cells grown at steady state in C-limited media (0.454 pg cell^{-1}).

Light limitation due to self-shading was implemented using the method discussed in Chapter 2. The 300-mL tissue culture flask used in refs. [28, 7] had a surface area that was approximately 0.0225 m^2. O_2 production fluxes for *i*Tps1426 were constrained for each slice of culture volume during the light period using experimental values from the photosynthesis-

irradiance curve for cells acclimated to 200 μmol photons m^{-2} s^{-1} published in ref. [67] using Platt fitting parameters [68].

$$NPP = \left(0.775 \cdot \left(1 - e^{(\frac{-0.0035 \cdot I}{0.775})}\right) - 0.13\right) \cdot \left(\frac{mg\ chl\ a/cell}{gDW/cell}\right) \tag{3.2}$$

The compensation light level (at which net O_2 production (NPP) = 0) was calculated to be 40.25 μmol photons m^{-2} s^{-1}. The lower bound of the sum of CO_2 and HCO_3 uptake was constrained with gross primary productivity values for cells acclimated to 200 μmol photons m^{-2} s^{-1} [67].

$$-1 * 0.605 \cdot \left(1 - e^{\frac{-0.0028 \cdot I}{0.605}}\right) \cdot \left(\frac{mg\ chl\ a/cell}{gDW/cell}\right) \leq EX\_co2\_e + EX\_hco3\_e \tag{3.3}$$

After each time point, the concentration of O_2, HCO_3, and CO_2 was re-equilibrated with the atmosphere so that concentrations remained constant (well-mixed).

During the light phase, *i*Tps1426 was constrained with the photorespiration constraint determined in Chapter 2. The upper bound of the biomass demand reaction ('DM_biomass_c') was constrained with the maximum measured growth rate from ref. [7]. The bounds of the *i*Rpo1302 biomass demand reaction ('DM_biomass_c') were constrained with the minimum and maximum measured growth rates from ref. [7] and aquacobalamin production ('EX_aqcobal_e') was set to 0.0001 of optimum. The penalty reaction is a demand reaction for all the penalty coefficients and ϕ is the minimized penalty reaction value [39] given the model constraints. The lower bound of the penalty reaction was constrained to ϕ/1.1 and the upper bound was constrained to ϕ*1.1. In the last step the biomass demand reaction was maximized. During the dark phase, the chlororespiration constraint was included (see Chapter 2), the lower bound of cyclic electron flow ('CEF_h') was set to zero, and the ATP maintenance reaction ('ATPM_c') was optimized in *i*Tps1426. The *i*Tps1426 penalties were not constrained in the dark. Parsimonious Flux Balance Analysis (pFBA) was used to minimize the absolute sum of fluxes and achieve a single solution. The *i*Rpo1302 model was treated the same as during the light period. Prior to the 36-*h* time point, the 8-*h* transcriptomic data were used to constrain *i*Tps1426; while the 56-*h* transcriptomic data was used

after 32 h. For iRpo1302, the 8-h transcriptomic data were used for the entire simulation. The GLPK solver was used to solve each mixed integer linear programming problem with an error tolerance of $1e-6$.

3.6 Acknowledgements

I would like to thank Bryndan Durham for providing data from her papers and for answering questions about experimental details.

References

[1] LR Pomeroy. "The ocean's food web, a changing paradigm." In: *Bioscience* 24.9 (Sept. 1974), pp. 499–504. DOI: 10.2307/1296885.

[2] F Azam and F Malfatti. "Microbial structuring of marine ecosystems." In: *Nature Reviews* 5.10 (Oct. 2007), pp. 782–91. ISSN: 1740-1534. DOI: 10.1038/nrmicro1747.

[3] PJLB Williams. "The balance of plankton respiration and photosynthesis in the open oceans." In: *Nature* 394.6688 (July 1998), pp. 55–7. ISSN: 00280836. DOI: 10.1038/27878.

[4] R Stocker and JR Seymour. "Ecology and physics of bacterial chemotaxis in the ocean." In: *Microbiology and Molecular Biology Reviews* 76.4 (Nov. 2012), pp. 792–812. ISSN: 1092-2172. DOI: 10.1128/MMBR.00029-12.

[5] T Fenchel. "Marine Plankton Food Chains." In: *Annual Review of Ecology and Systematics* 19.1 (Nov. 1988), pp. 19–38. ISSN: 0066-4162. DOI: 10.1146/annurev.es.19.110188.000315.

[6] SA Amin et al. "Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria." In: *Nature* 522.7554 (May 2015), pp. 98–101. ISSN: 0028-0836. DOI: 10.1038/nature14488.

[7] BP Durham et al. "Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system." In: *Environmental Microbiology* 19.9 (Sept. 2017), pp. 3500–13. ISSN: 14622912. DOI: 10.1111/1462-2920.13834.

[8] M Landa et al. "Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom." In: *The ISME Journal* 11 (July 2017), pp. 2677–90. ISSN: 1751-7362. DOI: 10.1038/ismej.2017.117.

[9] A Buchan, MA Moran, and JM Gonzalez. "Overview of the marine *Roseobacter* lineage." In: *Applied and environmental microbiology* 71.10 (Oct. 2005), pp. 5665–77. ISSN: 00224456. DOI: 10.1128/AEM.71.10.5665-5677.2005.

[10] SA Amin, MS Parker, and EV Armbrust. "Interactions between diatoms and bacteria." In: *Microbiology and Molecular Biology Reviews* 76.3 (Sept. 2012), pp. 667–84. ISSN: 1092-2172. DOI: 10.1128/MMBR.00007-12.

[11] H Teeling et al. "Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom." In: *Science* 336.6081 (May 2012), pp. 608–11. ISSN: 0036-8075. DOI: 10.1126/science.1218344.

[12] RJ Newton et al. "Genome characteristics of a generalist marine bacterial lineage." In: *The ISME Journal* 4.6 (July 2010), pp. 784–98. ISSN: 1751-7370. DOI: 10.1038/ismej.2009.150.

[13] KC Haines and RRL Guillard. "Growth of vitamin B_{12}-requiring marine diatoms in mixed laboratory cultures with vitamin B_{12}-producing marine bacteria." In: *Journal of Phycology* 10.3 (Sept. 1974), pp. 245–52. ISSN: 15298817. DOI: 10.1111/j.1529-8817.1974.tb02709.x.

[14] SA Sañudo-Wilhelmy et al. "The role of B vitamins in marine biogeochemistry." In: *Annual Review of Marine Science* 6.1 (Jan. 2014), pp. 339–67. ISSN: 1941-1405. DOI: 10.1146/annurev-marine-120710-100912.

[15] MT Croft et al. "Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria." In: *Nature* 438.7064 (Nov. 2005), pp. 90–3. ISSN: 0028-0836. DOI: 10.1038/nature04056.

[16] DM Karl. "Nutrient dynamics in the deep blue sea." In: *Trends in Microbiology* 10.9 (Sept. 2002), pp. 410–8. ISSN: 0966842X. DOI: 10.1016/S0966-842X(02)02430-7.

[17] MR Droop. "Vitamins, phytoplankton and bacteria: symbiosis or scavenging?" In: *Journal of Plankton Research* 29.2 (Jan. 2007), pp. 107–13. ISSN: 0142-7873. DOI: 10.1093/plankt/fbm009.

[18] EM Bertrand et al. "Iron limitation of a springtime bacterial and phytoplankton community in the Ross sea: Implications for vitamin B_{12} nutrition." In: *Frontiers in Microbiology* 2 (Aug. 2011), pp. 1–12. ISSN: 1664-302X. DOI: 10.3389/fmicb.2011.00160.

[19] EM Bertrand et al. "Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge." In: *Proceedings of the National Academy of Sciences* 112.32 (Aug. 2015), pp. 9938–43. ISSN: 0027-8424. DOI: 10.1073/pnas.1501615112.

[20] TJ Browning et al. "Nutrient co-limitation at the boundary of an oceanic gyre." In: *Nature* 551.7679 (Nov. 2017), pp. 242–6. ISSN: 0028-0836. DOI: 10.1038/nature24063.

[21] SA Sanudo-Wilhelmy et al. "Multiple B-vitamin depletion in large areas of the coastal ocean." In: *Proceedings of the National Academy of Sciences* 109.35 (Aug. 2012), pp. 14041–5. ISSN: 0027-8424. DOI: 10.1073/pnas.1208755109.

[22] W Qin et al. "Stress response of a marine ammonia-oxidizing archaeon informs physiological status of environmental populations". In: *ISME Journal* 12.2 (2018), pp. 508–19. ISSN: 17517370. DOI: 10.1038/ismej.2017.186.

[23] KR Heal et al. "Two distinct pools of B_{12} analogs reveal community interdependencies in the ocean." In: *Proceedings of the National Academy of Sciences* 114.2 (Jan. 2017), pp. 364–9. ISSN: 0027-8424. DOI: 10.1073/pnas.1608462114.

[24] MB Cooper et al. "Cross-exchange of B-vitamins underpins a mutualistic interaction between *Ostreococcus tauri* and *Dinoroseobacter shibae*." In: *The ISME Journal* 13 (Sept. 2018), pp. 334–45. ISSN: 17517370. DOI: 10.1038/s41396-018-0274-y.

[25] KE Helliwell et al. "Insights into the evolution of vitamin B_{12} auxotrophy from sequenced algal genomes." In: *Molecular Biology and Evolution* 28.10 (Oct. 2011), pp. 2921–33. ISSN: 0737-4038. DOI: 10.1093/molbev/msr124.

[26] S Imam, DR Noguera, and TJ Donohue. "Global insights into energetic and metabolic networks in *Rhodobacter sphaeroides*." In: *BMC Systems Biology* 7 (Sept. 2013), p. 89. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-89.

[27] CR Reisch et al. "Metabolism of dimethylsulphoniopropionate by *Ruegeria pomeroyi* DSS-3." In: *Molecular Microbiology* 89.4 (Aug. 2013), pp. 774–91. ISSN: 0950382X. DOI: 10.1111/mmi.12314.

[28] BP Durham et al. "Cryptic carbon and sulfur cycling between surface ocean plankton." In: *Proceedings of the National Academy of Sciences* 112.2 (Jan. 2015), pp. 453–7. ISSN: 0027-8424. DOI: 10.1073/pnas.1413137112.

[29] S Weinitschke et al. "Gene clusters involved in isethionate degradation by terrestrial and marine bacteria". In: *Applied and Environmental Microbiology* 76.2 (Jan. 2010), pp. 618–21. ISSN: 0099-2240. DOI: 10.1128/AEM.01818-09.

[30] I Lidbury, JC Murrell, and Y Chen. "Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: implications for marine carbon and nitrogen cycling." In: *The ISME Journal* 9.3 (Aug. 2015), pp. 760–9. ISSN: 1751-7362. DOI: 10.1038/ismej.2014.149.

[31] I Lidbury et al. "Comparative genomics and mutagenesis analyses of choline metabolism in the marine *Roseobacter* clade." In: *Environmental Microbiology* 17.12 (Dec. 2015), pp. 5048–62. ISSN: 14622920. DOI: 10.1111/1462-2920.12943.

[32] M Cunliffe. "Physiological and metabolic effects of carbon monoxide oxidation in the model marine bacterioplankton *Ruegeria pomeroyi* DSS-3." In: *Applied and Environmental Microbiology* 79.2 (Jan. 2013), pp. 738–40. ISSN: 00992240. DOI: 10.1128/AEM.02466-12.

[33] M Cunliffe. "Purine catabolic pathway revealed by transcriptomics in the model marine bacterium *Ruegeria pomeroyi* DSS-3." In: *FEMS Microbiology Ecology* 92.1 (Nov. 2016), pp. 1–6. ISSN: 15746941. DOI: 10.1093/femsec/fiv150.

[34] C Risso et al. "Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*." In: *Journal of Bacteriology* 190.7 (Apr. 2008), pp. 2266–74. ISSN: 00219193. DOI: 10.1128/JB.01841-07.

[35] WM Johnson, MC Kido Soule, and EB Kujawinski. "Evidence for quorum sensing and differential metabolite production by the marine heterotroph, *Ruegeria pomeroyi*, in response to DMSP." In: *The ISME Journal* 10 (Feb. 2016), pp. 2304–16. ISSN: 1751-7362. DOI: 10.1038/ismej.2016.6.

[36] AR Rivers, CB Smith, and MA Moran. "An updated genome annotation for the model marine bacterium *Ruegeria pomeroyi* DSS-3." In: *Standards in Genomic Sciences* 9.1 (Dec. 2014), p. 11. ISSN: 1944-3277. DOI: 10.1186/1944-3277-9-11.

[37] KR Patil and J Nielsen. "Uncovering transcriptional regulation of metabolism by using metabolic network topology." In: *Proceedings of the National Academy of Sciences* 102.8 (Feb. 2005), pp. 2685–9. ISSN: 0027-8424 (Print) 0027-8424 (Linking). DOI: 10.1073/pnas.0406811102.

[38] AP Oliveira, KR Patil, and J Nielsen. "Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks." In: *BMC*

Systems Biology 2.1 (Feb. 2008), p. 17. ISSN: 1752-0509. DOI: 10.1186/1752-0509-2-17.

[39] BJ Schmidt et al. "GIM^{3}E: condition-specific models of cellular metabolism developed from metabolomics and expression data." In: *Bioinformatics* 29.22 (Nov. 2013), pp. 2900–8. ISSN: 1460-2059. DOI: 10.1093/bioinformatics/btt493.

[40] D Machado and M Herrgård. "Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism." In: *PLoS Computational Biology* 10.4 (Apr. 2014). ISSN: 15537358. DOI: 10.1371/journal.pcbi.1003580.

[41] L Chan et al. "Transcriptional changes underlying elemental stoichiometry shifts in a marine heterotrophic bacterium". In: *Frontiers in Microbiology* 3.may (2012), pp. 1–24. ISSN: 1664-302X. DOI: 10.3389/fmicb.2012.00159.

[42] MR Brown et al. "Effects of harvest stage and light on the biochemical composition of the diatom *Thalassiosira pseudonana*." In: *Journal of Phycology* 32.1 (Feb. 1996), pp. 64–73. ISSN: 0022-3646. DOI: 10.1111/j.0022-3646.1996.00064.x.

[43] MA Moran et al. "Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment." In: *Nature* 432.7019 (Dec. 2004), pp. 910–3. ISSN: 0028-0836. DOI: 10.1038/nature03170.

[44] R Green et al. "Vitamin B_{12} deficiency." In: *Nature Reviews Disease Primers* 3.5 (June 2017), p. 17040. ISSN: 2056-676X. DOI: 10.1038/nrdp.2017.40.

[45] MF Romine et al. "Elucidation of roles for vitamin B_{12} in regulation of folate, ubiquinone, and methionine metabolism." In: *Proceedings of the National Academy of Sciences* 114.7 (Feb. 2017), E1205–14. ISSN: 0027-8424. DOI: 10.1073/pnas.1612360114.

[46] MJ Follows et al. "Emergent biogeography of microbial communities in a model ocean." In: *Science* 315.5820 (Mar. 2007), pp. 1843–6. ISSN: 1095-9203. DOI: 10.1126/science.1138544.

[47] J Levering et al. "Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom." In: *PLoS ONE* 11.5 (May 2016), e0155038. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0155038.

[48] JR Casey et al. "Adaptive evolution of phosphorus metabolism in *Prochlorococcus*." In: *mSystems* 1.6 (Dec. 2016), e00065–16. ISSN: 2379-5077. DOI: 10.1128/mSystems.00065-16.

[49] R Rex et al. "Swimming in light: A large-scale computational analysis of the metabolism of *Dinoroseobacter shibae*." In: *PLoS Computational Biology* 9.10 (Oct. 2013). ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1003224.

[50] M Fondi et al. "Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125." In: *Environmental Microbiology* 17.3 (Mar. 2015), pp. 751–66. ISSN: 14622912. DOI: 10.1111/1462-2920.12513.

[51] ZA King et al. "BiGG Models: A platform for integrating, standardizing and sharing genome-scale models." In: *Nucleic Acids Research* 44.D1 (Oct. 2016), pp. D515–22. ISSN: 1362-4962. DOI: 10.1093/nar/gkv1049.

[52] L Li, CJ Stoeckert, and DS Roos. "OrthoMCL: Identification of ortholog groups for eukaryotic genomes." In: *Genome Research* 13.9 (Sept. 2003), pp. 2178–89. ISSN: 1088-9051. DOI: 10.1101/gr.1224503.

[53] I Thiele and BO Palsson. "A protocol for generating a high-quality genome-scale metabolic reconstruction." In: *Nature protocols* 5.1 (Jan. 2010), pp. 93–121. ISSN: 1750-2799. DOI: 10.1038/nprot.2009.203. arXiv: NIHMS150003.

[54] A Ebrahim et al. "COBRApy: COnstraints-Based Reconstruction and Analysis for Python." In: *BMC Systems Biology* 7.1 (Aug. 2013), p. 74. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74.

[55] S Imam et al. "Quantifying the effects of light intensity on bioproduction and maintenance energy during photosynthetic growth of *Rhodobacter sphaeroides*." In: *Photosynthesis Research* 123.2 (Feb. 2015), pp. 167–82. ISSN: 01668595. DOI: 10.1007/s11120-014-0061-1.

[56] JM González et al. "*Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments." In: *International Journal of Systematic and Evolutionary Microbiology* 53.5 (Sept. 2003), pp. 1261–9. ISSN: 1466-5026. DOI: 10.1099/ijs.0.02491-0.

[57] I Lidbury, JC Murrell, and Y Chen. "Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria." In: *Proceedings of the National Academy of Sciences* 111.7 (Feb. 2014), pp. 2710–5. ISSN: 1091-6490. DOI: 10.1073/pnas.1317834111.

[58] X Mou et al. "Genes for transport and metabolism of spermidine in *Ruegeria pomeroyi* DSS-3 and other marine bacteria." In: *Aquatic Microbial Ecology* 58.3 (Feb. 2010), pp. 311–21. ISSN: 0948-3055. DOI: 10.3354/ame01367.

[59] SL Bell and B Palsson. "expa: A program for calculating extreme pathways in biochemical reaction networks." In: *Bioinformatics* 21.8 (Dec. 2005), pp. 1739–40. ISSN: 13674803. DOI: 10.1093/bioinformatics/bti228.

[60] JO Hay et al. "Integration of a constraint-based metabolic model of *Brassica napus* developing seeds with ^{13}C-metabolic flux analysis." In: *Frontiers in Plant Science* 5 (Dec. 2014), pp. 1–18. ISSN: 1664-462X. DOI: 10.3389/fpls.2014.00724.

[61] NL Bray et al. "Near-optimal probabilistic RNA-seq quantification." In: *Nature Biotechnology* 34.5 (Apr. 2016), pp. 525–7. ISSN: 1087-0156. DOI: 10.1038/nbt.3519.

[62] H Pimentel et al. "Differential analysis of RNA-seq incorporating quantification uncertainty." In: *Nature Methods* 14.7 (June 2017), pp. 687–90. ISSN: 15487105. DOI: 10.1038/nmeth.4324.

[63] K Sameith et al. "Functional modules integrating essential cellular functions are predictive of the response of leukaemia cells to DNA damage." In: *Bioinformatics* 24.22 (Nov. 2008), pp. 2602–7. ISSN: 13674803. DOI: 10.1093/bioinformatics/btn489.

[64] Mark Hildebrand, Kalpana Manandhar-Shrestha, and Raffaela Abbriano. "Effects of chrysolaminarin synthase knockdown in the diatom *Thalassiosira pseudonana*: Implications of reduced carbohydrate storage relative to green algae." In: *Algal Research* 23 (Apr. 2017), pp. 66–77. ISSN: 22119264. DOI: 10.1016/j.algal.2017.01.010.

[65] A Ozkan and GL Rorrer. "Effects of CO_2 delivery on fatty acid and chitin nanofiber production during photobioreactor cultivation of the marine diatom *Cyclotella* sp." In: *Algal Research* 26 (July 2017), pp. 422–30. ISSN: 22119264. DOI: 10.1016/j.algal.2017.07.003.

[66] A Ozkan and GL Rorrer. "Lipid and chitin nanofiber production during cultivation of the marine diatom *Cyclotella* sp. to high cell density with multistage addition of silicon and nitrate." In: *Journal of Applied Phycology* 29.4 (Aug. 2017), pp. 1811–8. ISSN: 15735176. DOI: 10.1007/s10811-017-1113-7.

[67] NL Fisher and KH Halsey. "Mechanisms that increase the growth efficiency of diatoms in low light." In: *Photosynthesis Research* 129.2 (Aug. 2016), pp. 183–97. ISSN: 15735079. DOI: 10.1007/s11120-016-0282-6.

[68] T Platt, CL Gallegos, and WG Harrison. "Photoinhibition of photosynthesis in natural assemblages of marine-phytoplankton." In: *Journal of Marine Research* 38.4 (1980), pp. 687–701.

CONCLUSION

Planetary elemental cycling is driven by geological acid-base chemistry and biological reduction-oxidation reactions [1]. The end products of one metabolic pathway becomes the substrate for another; in this way, synergy between oxidative and reductive pathways is the basis of biochemical cycling. In fact, oxidoreductase genes are able to differentiate communities of different microbes much more effectively than taxonomic markers or any other enzyme category [2]. When segregated in different organisms, redox reactions are the foundation of microbial interactions, and are in some ways analogous to energetic coupling between the plastid and mitochondria.

Photosynthesis evolved ∼3 billion years ago and oxidized the planet [3]. It is the only energy-generating pathway that does not depend on pre-formed chemical bonds [1]. In oxygenic photosynthesis, light energy oxidizes water, releasing electrons and protons that serve to reduce inorganic C, N, P, and S. Photosynthetic cells can complete the redox cycle by respiring reduced compounds, or excrete them into the environment, providing food for heterotrophs.

Genome-scale metabolic modeling demonstrates the importance of redox balance, particularly for photosynthetic organisms dealing with high energy inputs from light. The effects of photosynthetic metabolism likely impact the diversity of compounds produced by diatoms and consumed by marine heterotrophic microbes. In turn, nutrient-limitation and nutrient source can potentially be controlled by marine bacteria, impacting the physiology of diatoms. Flux Balance Analysis has the potential to link molecular data to ecosystem processes – allowing for greater biological complexity and more realistic outcomes from ecosystem models. My hope is that the genome-scale metabolic models developed in my work can be further refined and improved over time, and that models for other representative marine microbes

may be developed in the future.

References

[1] PG Falkowski, T Fenchel, and EF Delong. "The microbial engines that drive Earth's biogeochemical cycles." In: *Science* 320.5879 (May 2008), pp. 1034–9. ISSN: 1095-9203. DOI: 10.1126/science.1153213.

[2] S Ramírez-Flandes, B González, and O Ulloa. "Redox traits characterize the organization of global microbial communities". In: *Proceedings of the National Academy of Sciences* 116.9 (Feb. 2019), pp. 3630–5. ISSN: 0027-8424. DOI: 10.1073/pnas.1817554116.

[3] DE Canfield. "The early history of atmospheric oxygen: Homage to Robert M. Garrels". In: *Annual Review of Earth and Planetary Sciences* 33.1 (2005), pp. 1–36. ISSN: 0084-6597. DOI: 10.1146/annurev.earth.33.092203.122711.

Appendix A

CHAPTER 1

Name	HTCC2559*	SA60	Pm1	HV2
Cultured?	Y	Y	N	Y
Source	250-m water, Sargasso Sea	*P. multiseries* CLNN-17, CLN-25 × CLN-48 (2007)	*P. multiseries* CLN-47, CL-143 × CL-147 (2001)	*P. multiseries* GGA2, Puget Sound (2010)
Year of isolation (or detection)	2001	2010	2006	2012
Status	Complete	Draft	Draft	Not sequenced
Number of bases	2,952,962	2,878,364	3,010,226	-
%GC	33.9	33.66	33.78	-
Protein coding genes	2,678	2,637	2,769	-
RNA coding genes	49	31	49	-

* This is the type strain for *Croceibacter atlanticus*. It was isolated and characterized by Cho & Giovannoni (2003). The genome was published by Oh et al. (2010).

Table A.1: Summary of *Croceibacter atlanticus* genome and isolate information

Metabolism	*T. pseudonana* pathway	*Croceibacter* pathway
Amino acids		
Beta-alanine	-	beta-alanine biosynthesis III
Alanine	alanine biosynthesis III	-
Isoleucine	isoleucine biosynthesis I (from threonine)	-
Valine	valine biosynthesis	-
Leucine	leucine biosynthesis	-
Glutamate	glutamate biosynthesis I, II, III, IV, V	glutamate biosynthesis I, II, III, IV, V, arginine degradation I
Glutamine	glutamine biosynthesis I, II (tRNA-dependent), III	glutamine biosynthesis I
Aspartate	aspartate biosynthesis	aspartate biosynthesis
Asparagine	asparagine biosynthesis I, II, III (tRNA-dependent)	asparagine biosynthesis I, II
Arginine	arginine biosynthesis I, II (acetyl cycle)	-
Citrulline	citrulline biosynthesis	-
Ornithine	ornithine biosynthesis	-
Homocysteine	-	homocysteine biosynthesis
Homoserine	homoserine biosynthesis	homoserine biosynthesis
Serine	serine biosynthesis	serine biosynthesis (partial)
Cysteine	cysteine biosynthesis I	cysteine biosynthesis I (partial)
Selenocysteine	selenocysteine biosynthesis II (archaea and eukaryotes) (partial)	-
Methionine	methionine biosynthesis I (partial), II (partial)	methionine biosynthesis II (partial)
Glycine	glycine biosynthesis I, II, III, IV	glycine biosynthesis I, II
Threonine	threonine biosynthesis from homoserine	threonine biosynthesis from homoserine
Lysine	lysine biosynthesis IV	lysine biosynthesis IV (partial)
Phenylalanine	phenylalanine bioysynthesis	phenylalanine bioysynthesis
Tyrosine	tyrosine biosynthesis II, III, IV	tyrosine biosynthesis IV
Tryptophan	tryptophan biosynthesis (partial)	tryptophan biosynthesis
Proline	proline biosynthesis I, ornithine degradation I,	
arginine degradation VI (arginase 2 pathway)	proline biosynthesis I, arginine degradation VI (arginase 2 pathway)	
Histidine	histidine biosynthesis (partial)	histidine biosynthesis
Vitamins		
Thiamine (vit. B1)	thiamin diphosphate biosynthesis IV (eukaryotes)	thiamin diphosphate biosynthesis I/II
Riboflavin (vit. B2)	flavin biosynthesis I (bacteria and plants)	flavin biosynthesis I (bacteria and plants)
Nicotinic acid (vit. B3)	NAD biosynthesis I (from aspartate) (partial)	NAD biosynthesis I (from aspartate) (partial)
Pantothenic acid (vit. B5)	phosphopantothenate biosynthesis I	phosphopantothenate biosynthesis I (partial)
Pyridoxine (vit. B6)	pyridoxal 5'-phosphate biosynthesis I (partial)	pyridoxal 5'-phosphate biosynthesis I
Biotin (vit. B7)	biotin biosynthesis from 8-amino-7-oxononanoate I (partial)	biotin biosynthesis from 8-amino-7-oxononanoate I, II
Folic acid (vit. B9)	tetrahydrofolate biosynthesis	tetrahydrofolate biosynthesis
Cyanocobalamin (vit. B_{12})	-	-
p-Aminobenzoic acid	p-aminobenzoate biosynthesis (partial)	p-aminobenzoate biosynthesis
Inorganic nutrients		
Ammonia	ammonia assimilation I, II, III	ammonia assimilation I, II, III
Nitrate	nitrate reduction II (assimilatory)	-
Phosphate	phosphate acquisition	phosphate acquisition
Sulfate	sulfate reduction I, II, III (assimilatory)	sulfate reduction I, II, III (assimilatory) (partial)

Table A.2: Complementary and competing metabolisms in *Croceibacter atlanticus* and *Thalassiosira pseudonana*

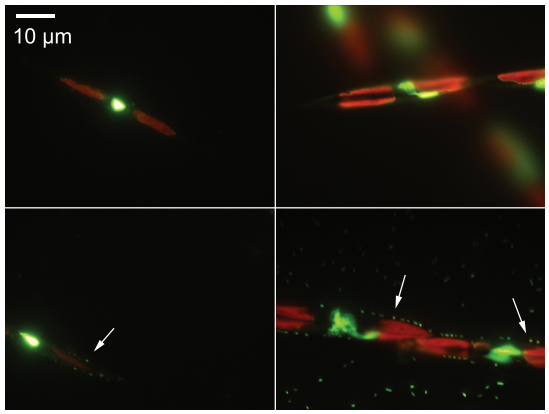


Figure A.1: Epifluorescence micrographs of *P. multiseries* GGA2 and *P. fradulenta* OC1 (left/right) in the absence (top row) or presence (bottom row) of *C. atlanticus* HV2. Diatom plastids autofluoresce red, SYBR Green-stained DNA from the diatom nucleus and the bacteria fluoresce green; diatom-attached bacteria are indicated with arrows.

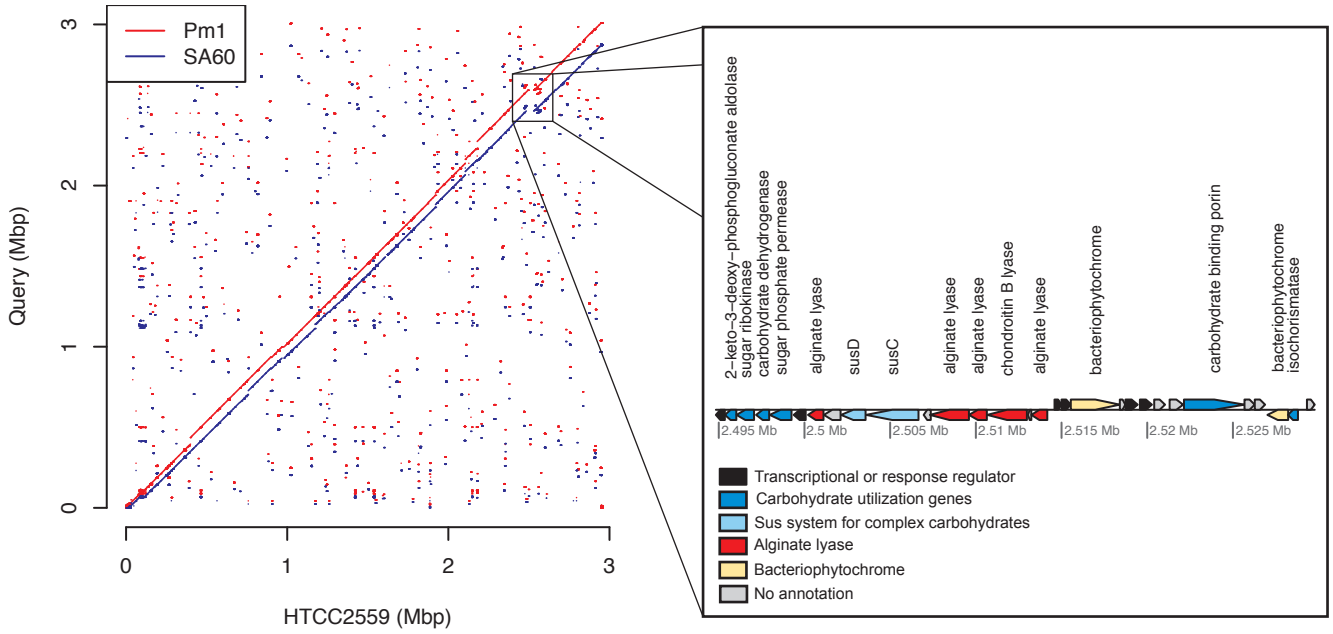


Figure A.2: Left: Nucleotide dot plot alignments of *C. atlanticus* genomes Pm1 and SA60 to the type strain HTCC 2559 (Mbp = mega base pairs). Pm1 and SA60 genomes are $\geq 98\%$ identical to HTCC2559. Breaks from the 1:1 line are genome segments that do not align to HTCC 2559. Points in the background are additional sequences that align to a particular location on HTCC 1559. Right: Pm1 and SA60 both lack genes putatively involved in alginate degradation and two bacteriophytochromes. Gene arrangements and annotations from HTCC2559 are highlighted.

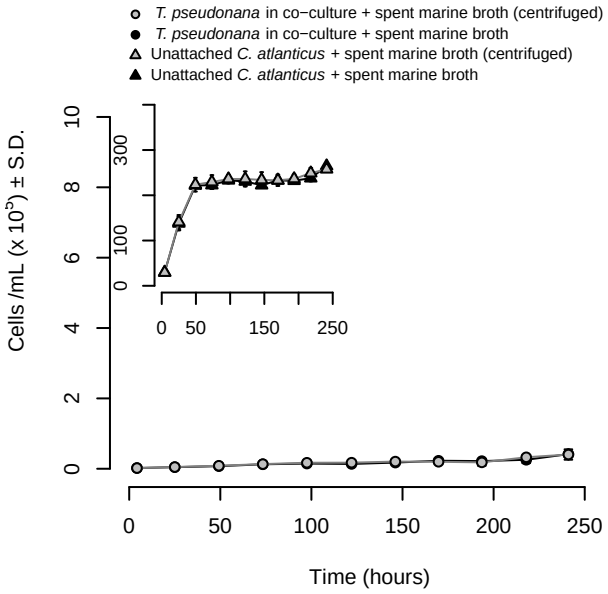


Figure A.3: There is no effect of the centrifugation washing process on the growth of *C. atlanticus* and its impact on *T. pseudonana*. Symbols mark the mean of diatom and unattached bacterium cell concentration (circles, triangles) in co-culture. *C. atlanticus* cells were either centrifuged twice at 3,220×g for five minutes at 10°C or added directly to *T. pseudonana* cultures (grey, black; $n = 3$ biological replicates ± 1 standard deviation; note that many s.d. bars are smaller than the symbol size).

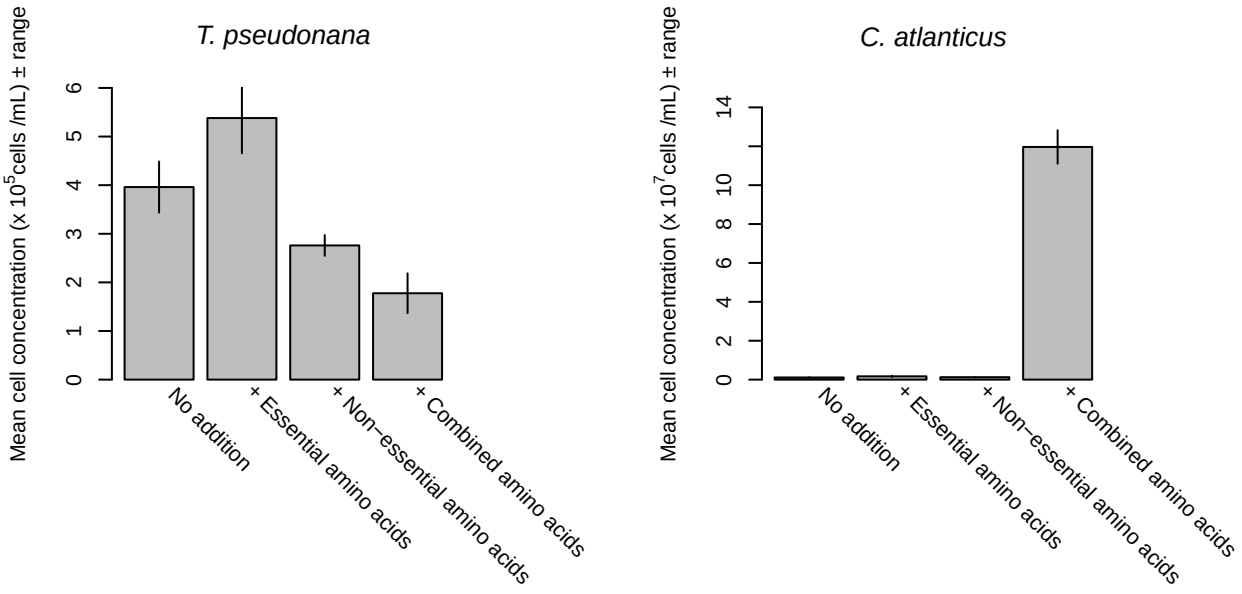


Figure A.4: Cell concentrations of *T. pseudonana* and Aquil-washed *C. atlanticus* after five days in co-culture ($n = 2$). Co-cultures were supplemented with essential amino acids, non-essential amino acids, a mix of both, or nothing.

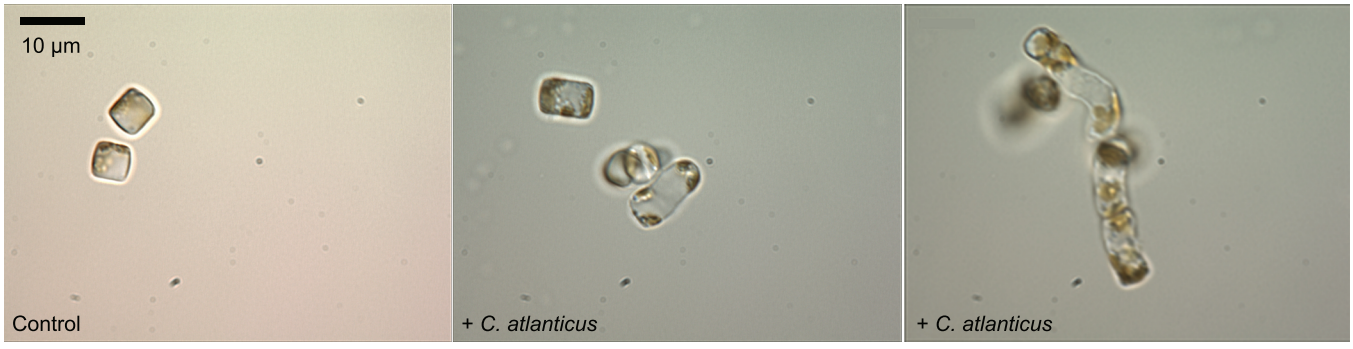
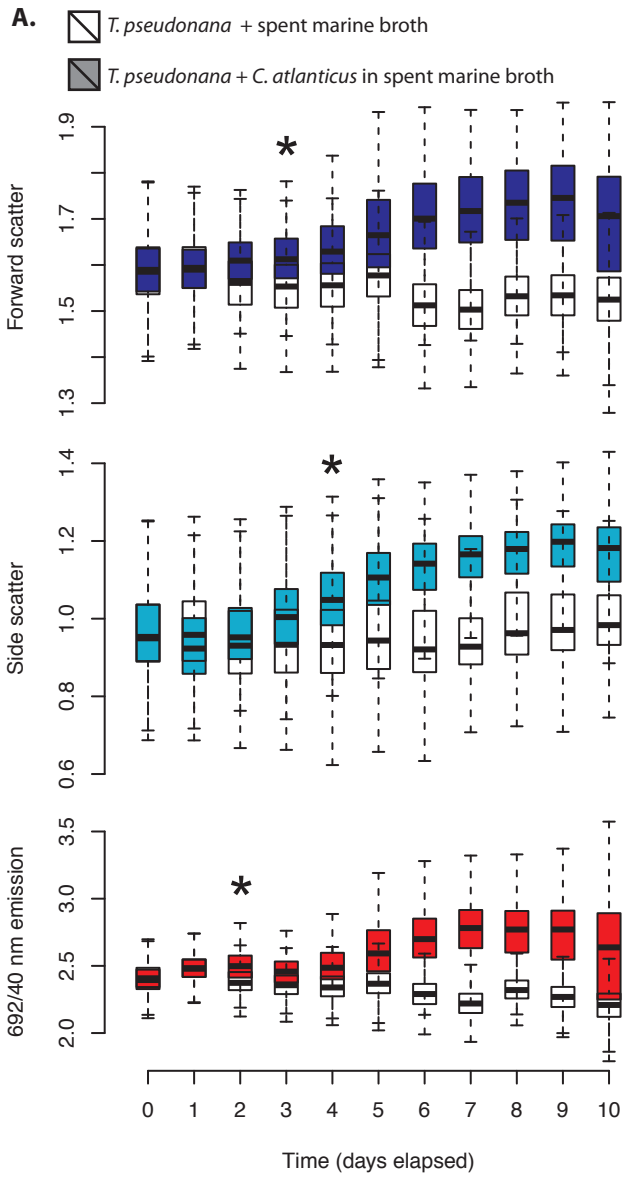
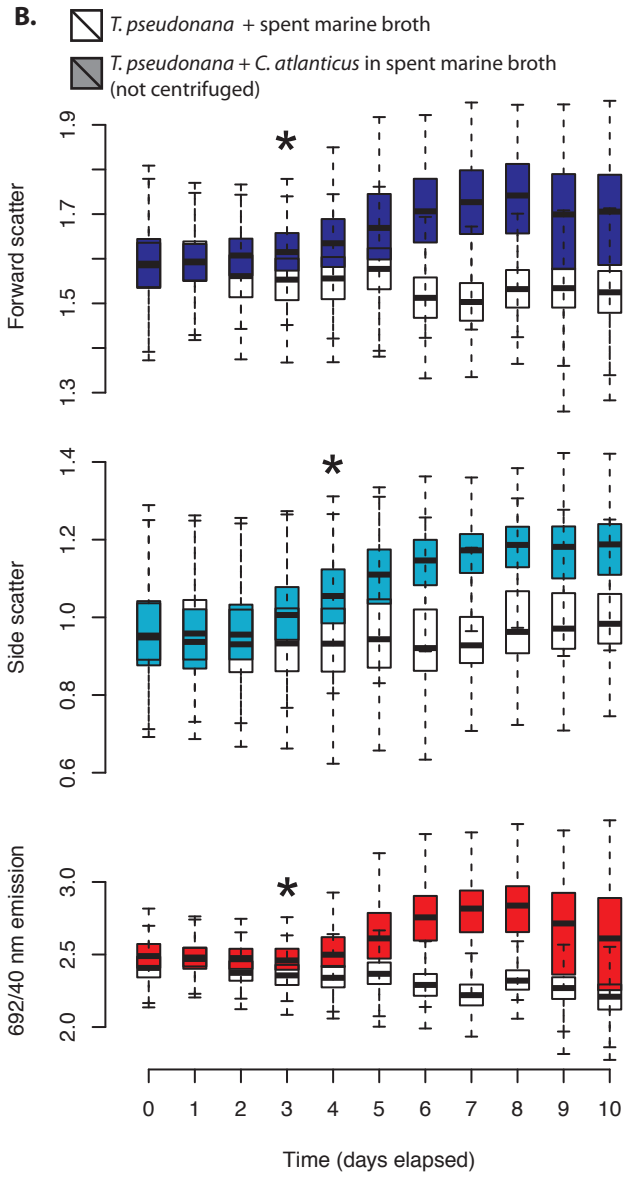


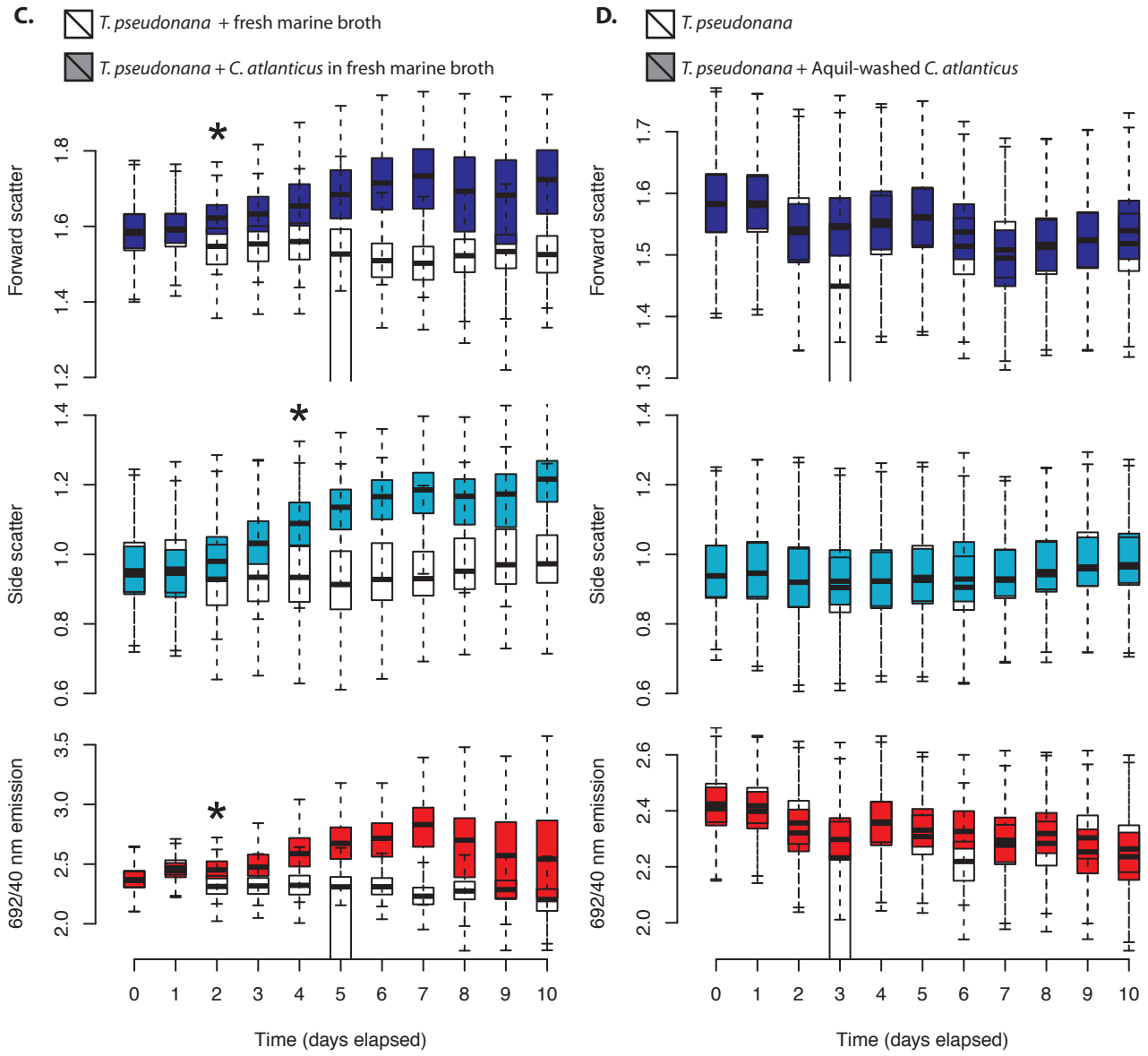
Figure A.5: Light microscopy images of *T. pseudonana* demonstrating the impact of *C. atlanticus* on diatom morphology. Fields were arbitrarily selected from a single sample of each treatment at one time point.



(a) *T. pseudonana* + spent marine broth versus *T. pseudonana* + *C. atlanticus* in spent marine broth.



(b) *T. pseudonana* + spent marine broth versus *T. pseudonana* + *C. atlanticus* in spent marine broth (not centrifuged).



(c) *T. pseudonana* + fresh marine broth versus *T. pseudonana* + *C. atlanticus* in fresh marine broth.

(d) *T. pseudonana* versus *T. pseudonana* + Aquil-washed *C. atlanticus*.

Figure A.6: Box-and-whiskers plots showing change in *T. pseudonana* cell forward scatter, side scatter, and red fluorescence distributions over time. Coloured boxes represent co-culture treatments and white boxes represent control treatments ($n = 3$). Asterisks mark the first day where the Wilcoxon Rank Sum test detected a significant difference between populations. Outliers noted in the Material & Methods section were retained in these plots to show their effect, but were removed from statistical analysis.

Appendix B

CHAPTER 2

Data Set B.1: Table of reactions and metabolites and references in *i*Tps1426

Data Set B.2: Biomass composition calculations and references for *Thalassiosira pseudonana*

Data Set B.3: Constraint calculations and references

File B.1: *i*Tps1426 in SBML format

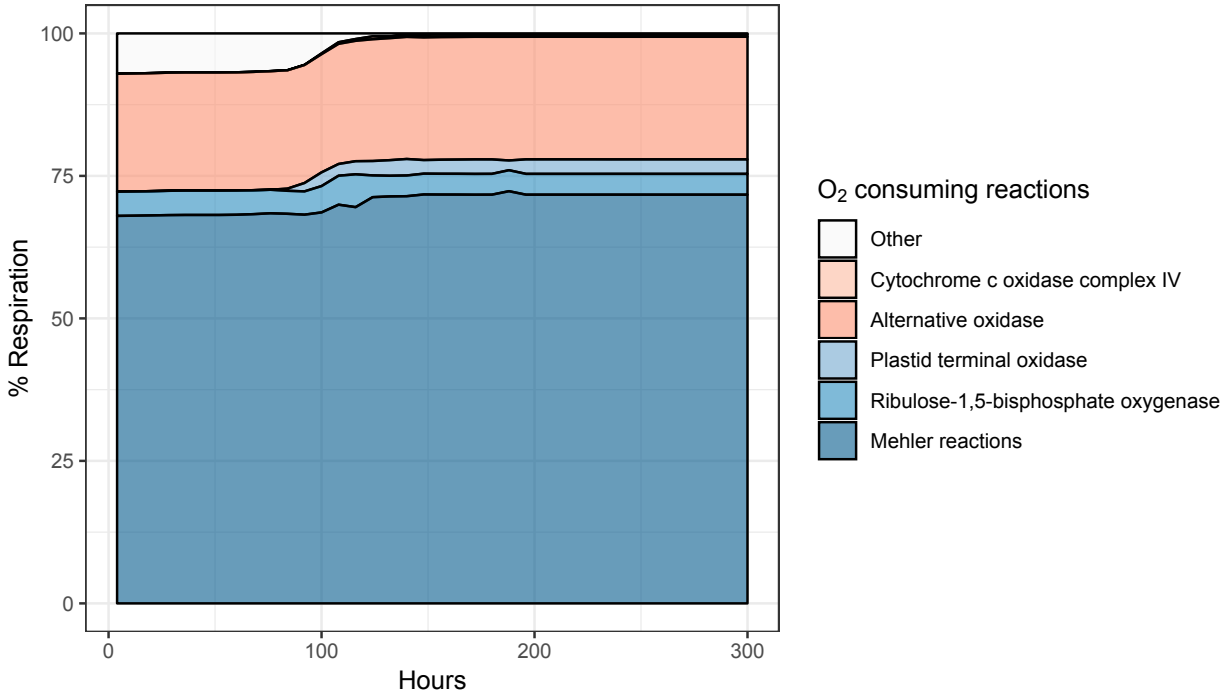


Figure B.1: The relative contribution of different oxygen-consuming reactions to respiration during simulated growth of *i*Tps1426 in Aquil at 100 μmol photons m^{-2} s^{-1}. Respiration reactions in the plastid are blue, reactions in the mitochondria are red, and all other oxygen-consuming reactions are white.

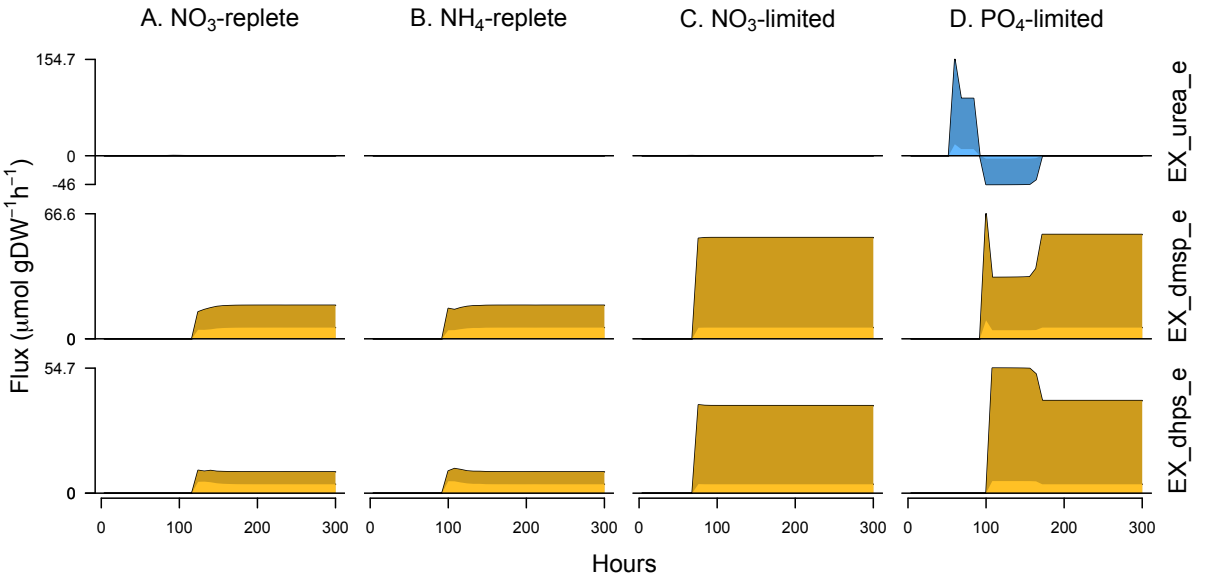


Figure B.2: Simulated fluxes of urea ('EX_urea_e', blue), DMSP ('EX_dmsp_e', yellow), and DHPS ('EX_dhps_e', yellow) over the course of the simulated growth experiments at 100 μmol photons m^{-2} s^{-1} in nutrient-replete Aquil (A), nutrient-replete modified Aquil where NO_3 was replaced with NH_4 (B), NO_3-limited Aquil (C), and PO_4-limited Aquil (D). The lighter colour shows the fluxes for the first slice of the culture, which is never subjected to self-shading, while the darker colour shows the sum of all fluxes for that compound at that time point.

Appendix C

CHAPTER 3

Data Set C.1: Table of reactions, metabolites, and references for *i*Rpo1302

Data Set C.2: File showing metabolite z-scores, $\log_2$(fold-change), and p-values from metabolite measurements in ref. [1]

Data Set C.3: Biomass composition calculations for *i*Rpo1302

File C.1: *i*Rpo1302 in SBML format

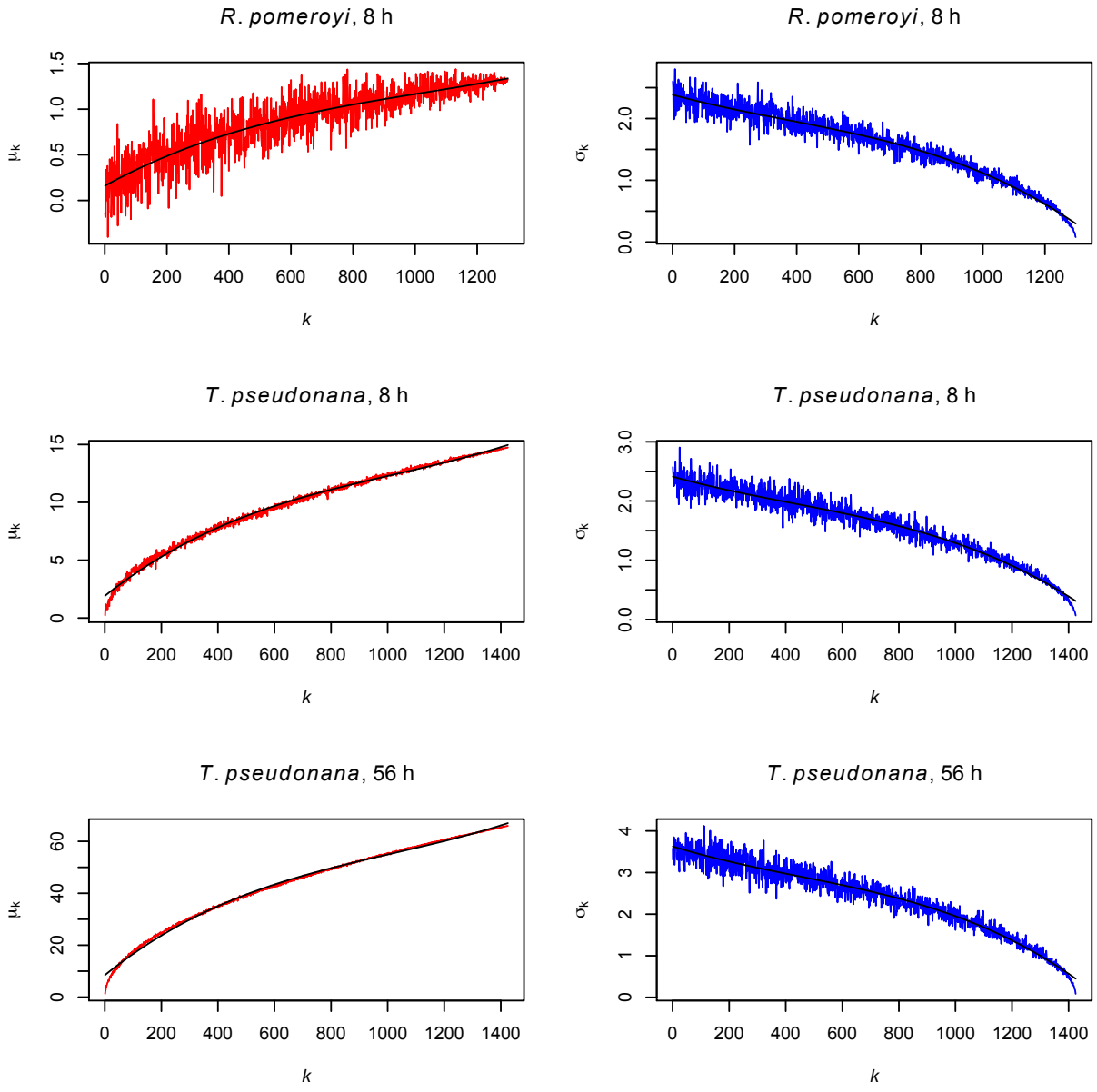


Figure C.1: Mean (red) and standard deviation (blue) of z-scores for randomly sampled sets of enzymes with k neighbours, smoothed with a univariate spline (black line). Z-scores were calculated for each enzyme in the genome-scale networks of *R. pomeroyi* and *T. pseudonana* using transcriptomics data comparing growth in co-culture versus mono-culture after 8 h and 56 h.