DYNAMICS OF MARINE MICROBIAL METABOLISM AND PHYSIOLOGY AT STATION ALOHA

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI´I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OCEANOGRAPHY

MAY 2017

By
John R. Casey

Thesis Committee:
David Karl, Chairperson
Robert Bidigare
Matthew Church
Paul Falkowski
Craig Nelson
Thomas Hemscheidt

We certify that we have read this dissertation and that, in our opinion, it is satisfactory in scope and quality as a dissertation for the degree of Doctor of Philosophy in Oceanography.

DISSERTATION COMMITTEE

_____
Chairperson

_____

_____

_____

_____

ii

Acknowledgements

I would like to convey my deepest gratitude to my advisor David Karl for his guidance and encouragement throughout my graduate studies. Dave is a trailblazer in the most fundamental topics of marine biogeochemistry and microbial oceanography, and his vision and tireless commitment has fostered many broad and large-scale transdisciplinary initiatives. His Herculean efforts have proven truly transformative to both the field and to his collaborators and students, including myself. I was very fortunate to have a supportive and engaging committee, and would like to thank Robert Bidigare, Matthew Church, Paul Falkowski, Thomas Hemscheidt, and Craig Nelson for their efforts and their encouragement. Additionally, I had the pleasure of joining the Systems and Synthetic Biology group at Chalmers University of Technology for 6 months during 2014. This work was supported by an award within the National Science Foundation's Graduate Research Fellowship Program, Graduate Research Opportunities Worldwide, and by the Swedish Research Council, to which I am grateful. My gracious host, Jens Nielsen, provided me with invaluable expertise both from himself and from his group. Especially, I would like to thank Boyang Ji, Adil Mardinoglu, Partho Sarathi Sen, and Saeed Shaoie for sharing their time and expertise. I would also like to acknowledge Benedetto Barone, Karin Björkman, Michael Cooney, Ed DeLong, Sara Ferrón, Lance Fujieki, Eric Grabowski, Sandra Martinez-García, Brian Popp, Mike Rappé, Ty Samo, Chris Schvarcz, Grieg Steward, and Sam Wilson for their many and varied contributions. Finally, and most importantly, I would like to thank my loving family, friends, and partner Alice who have been with me every step of the way.

Table of Contents

Abstract

Marine microbial communities influence global biogeochemical cycles by coupling
the transduction of free energy to the transformation of Earth's essential bio-
elements: H, C, N, O, P, and S. The web of interactions between these processes is
extraordinarily complex, though fundamental physical and thermodynamic
principles should describe its dynamics. In this collection of 5 studies, aspects of the
complexity of marine microbial metabolism and physiology were investigated as
they interact with biogeochemical cycles and direct the flow of energy within the
Station ALOHA surface layer microbial community. In Chapter 1, and at the broadest
level of complexity discussed, a method to relate cell size to metabolic activity was
developed to evaluate allometric power laws at fine scales within picoplankton
populations. Although size was predictive of metabolic rates, within-population
power laws deviated from the broader size spectrum, suggesting metabolic diversity
as a key determinant of microbial activity. In Chapter 2, a set of guidelines was
proposed by which organic substrates are selected and utilized by the heterotrophic
community based on their nitrogen content, carbon content, and energy content. A
hierarchical experimental design suggested that the heterotrophic microbial
community prefers high nitrogen content but low energy density substrates, while
carbon content was not important. In Chapter 3, a closer look at the light-dependent
dynamics of growth on a single organic substrate, glycolate, suggested that growth
yields were improved by photoheterotrophy. The remaining chapters were based
on the development of a genome-scale metabolic network reconstruction of the
cyanobacterium *Prochlorococcus* to probe its metabolic capabilities and quantify

metabolic fluxes. Findings described in Chapter 4 pointed to evolution of the
Prochlorococcus metabolic network to optimize growth at low phosphate
concentrations. Finally, in Chapter 5 and at the finest scale of complexity, a method
was developed to predict hourly changes in both physiology and metabolic fluxes in
Prochlorococcus by incorporating gene expression time-series data within the
metabolic network model. Growth rates predicted by this method more closely
matched experimental data, and diel changes in elemental composition and the
energy content of biomass were predicted. Collectively, these studies identify and
quantify the potential impact of variations in metabolic and physiological traits on
the mêlée of microbial community interactions.

List of Tables

List of Figures

Background and motivation

Microbial control of the Earth system

Earth teems with microbial life. From the hot basement fluids deep within

the oceanic crust (Jungbluth *et al.,* 2013) to the frigid air currents of the upper

stratosphere (Shivaji *et al.,* 2006), microbes thrive in almost any space, no matter

how seemingly inhospitable, between the mantle and the mesosphere. Wherever a

redox potential is found, microbes have designed an energy transduction 'machine'

to flourish. Of course, construction of such a machine requires access to basic

elemental resources – H, C, N, O, P, S, as well as a selection of transition metals. As

with any machine, its construction also requires instructions – the genetic code –

and assembly and repair – the thousands of intermediate biochemical reactions

collectively known as metabolism. Driven by the second law of thermodynamics,

microbial dissipation of these redox potentials in turn drives Earth's biogeochemical

cycles for these bio-elements. Indeed, microbes are both the architects and the

engineers of the chemistry of Earth's crust.

Earth is approximately a thermodynamically closed system; that is, its

boundaries exchange heat and energy but not mass (to any appreciable extent in the

modern geological past). As such, the redox gradients and free energy which sustain

life are ultimately fueled by absorption of solar radiative energy (although 'boutique

metabolisms' reducing radionuclides exist; Lovely *et al.,* 1991), maintaining the

Earth system away from the cataclysmic evenness of thermodynamic equilibrium.

Within the Earth system, however, exists an interconnected web of open systems;

the land, atmosphere, and oceans exchange both mass and energy, and as such their

1

microbial inhabitants grow proximally on free energy sources and bio-elements generated both locally and from system boundaries. In the global oceans, organic matter, gasses, minerals and nutrients are deposited from the atmosphere, transported from coastlines, and ventilated from the seabed. This material, given the availability of a suitable redox pairing, can supply the chemical potential energy to support microbial growth, as is observed at the coastal margins, in hydrothermal vent systems, and near the sediment water interface. In the vast central regions of the major oceanic basins, however, the absorption of photons by light-harvesting microbes within the upper 2% of the water column supplies the overwhelming majority of free energy for the microbial community, and the rest of the food web.

Microbial transduction of radiative energy to chemical potential energy functions by charge separation across a membrane. Either through the water reducing reactions of oxidative photosynthesis or through chromophore photoreduction and subsequent ion translocation, an electrochemical chemiosmotic gradient is established. The resulting kinetic force, often in the form of a proton motive force, is exerted on specialized proteins embedded in intracellular membranes which store a portion of that kinetic energy in the phosphoanhydride bonds of the universal energy carrier, adenosine triphosphate (ATP). ATP is ultimately the energy 'currency' for all cellular activities, distributing chemical potential energy for the synthesis of macromolecular structures, or to activate non-spontaneous biochemical reactions. Some of this material can then be oxidized by other microbes to synthesize ATP, thus feeding the ocean food web, though the

ultimate fate of the chemical potential energy is heat radiated back to the emptiness of space.

Biogeochemical cycles in the global oceans

All six of the major bio-elements, H, C, N, O, P, and S, undergo microbially-mediated but thermodynamically constrained redox reactions, though P was only recently recognized to belong to that list (Karl, 2014). Within the global oceans, all six major bio-elements and the transition metals Fe and Co can limit microbial growth rates or yields, though H limitation is uncommon (c.f., Nishihara *et al.,* 2001). Each of the bio-elements is present in a variety of forms in the oceans; as inorganic monomers, complexed with minerals, or bound within the organic molecules of both living cells and in extracellular medium. The dissolved organic matter pool, the extracellular conduit which links microbial community metabolism, is composed of thousands of chemically distinct species (Mopper *et al.,* 2007). These various forms are produced, consumed and transformed by microbes, photochemical processes and other abiotic equilibrium reactions. Furthermore, each form is unevenly distributed in both space and time. Spatially, net biological production and consumption processes, transport by ocean currents, sinking or floating particles, and diffusive processes drive their distribution. Temporally, preferential use and production by microbes and differences in molecular stability and solubility (due to e.g., temperature, light, salinity) result in molecular half-lives spanning the order of minutes to millennia. Thus it is unsurprising that the details of such an interconnected web of elemental cycles are not completely understood.

The central gyres of the major ocean basins, due in part to their isolation from coastal runoff, and in part to the depth of the permanent thermocline, are relatively depleted of the bio-elements N, P, Fe, and Co. In the North Pacific Subtropical Gyre (NPSG), Earth's largest contiguous biome, vertical exchange from the nutrient-rich ocean interior to the nutrient-poor surface layer is largely insulated by thermal stratification resulting from a strong atmospherically forced temperature gradient which prevents winter mixing depths in excess of the top of the thermocline. Thus, the input of "new" boundary supplied bio-elements in the surface layer of the oligotrophic regions is small relative to thermal wind driven upwelling regions, coastal margins, and higher latitudes where seasonal forcing drives convective winter mixing to form a seasonal thermocline (Dutkiewicz *et al.,* 2005; Falkowski *et al.,* 1998). Instead, new nutrients are supplied by episodic entrainment from mesoscale features like eddies (Johnson *et al.,* 2010) and sub-mesoscale features like filaments and fronts (Guidi *et al.,* 2012; Ascani *et al.,* 2013). New nitrogen is also supplied by nitrogen fixation (Dugdale *et al.,* 1961), although this was considered a negligible source (Dugdale and Goering, 1967) until recently (summarized in Karl *et al.,* 2008). Supply of new nutrients by horizontal advection, zooplankton vertical migration, negatively buoyant particles, precipitation and atmospheric deposition are all potentially important sources, though their relative contributions are poorly constrained. Collectively, these boundary-supplied sources of new nutrients support net biological stock size increases, however the majority of microbial growth demands are met by the highly efficient recycling of energy and bio-elements within a microbial loop (Azam *et al.,* 1983), described in more detail in

the next section. Processes which decouple the stoichiometry of the bio-elements, including variations in biomass composition, production rates, and consumption rates, cause imbalance in availability and effect the functioning of the microbial loop.

Undersampling remains a challenge to ocean biogeochemistry and microbial oceanography; while increasingly detailed molecular methods have been adopted, environmental sensors and the proliferation of unmanned sample collection vehicles have not kept pace. As it has been since the Challenger Expedition of the mid 1800's, microbial processes are conducted at sea, thus limiting the scope of observations to Eulerian time-series sites and surveys. Station ALOHA (A Long-term Oligotrophic Habitat Assessment) and the Bermuda Atlantic Time-series Study site were chosen during the J-GOFS era (Joint Global Ocean Flux Study) as approximately representative of broad central regions of the NPSG and North Atlantic Subtropical Gyre, respectively. Station ALOHA, home to the Hawaii Ocean Time-series program, has been visited at approximately monthly intervals for nearly three decades to regularly characterize a suite of 'core' physical, chemical, and biological processes, and to contextualize numerous experiments and observations. Insights gleaned from the BATS and HOT programs have fundamentally transformed understanding of the role of the oceans in the Earth system (Steinberg *et al.,* 2001; Karl and Church, 2014). Perhaps as new sensors and autonomous platforms are developed, resolution of processes occurring at frequencies and spatial resolutions beyond the capabilities of research vessels will become accessible.

Microbial engineers in the NPSG

Fueled by solar energy capture and an often fleeting supply of essential bio-elements, the microbial loop consists of more than 37,000 microbial species (Sunagawa *et al.,* 2015) from all three domains of life – bacteria, archaea, and eukaryota. In the classical view (Azam *et al.,* 1983), small oxygenic photosynthesizers release dissolved organic carbon which is subsequently consumed by oxidative respiration from heterotrophic bacteria, with inorganic ions like ammonium and phosphate as byproducts. These inorganic nutrients in turn support the growth requirements of the photosynthesizers. Population sizes of both groups are tightly regulated by nanoflagellate grazing, effectively reducing the transfer of carbon to higher trophic levels. Additionally, since the particle size spectrum is small, sinking velocities are low and only a small fraction of the primary production is exported to the mesopelagic. Although original microbial loop concept remains approximately valid, the contemporary view is complicated by the prevalence of mixed metabolisms like photoheterotrophs and mixotrophic nanoflagellates. Furthermore, the 'viral shunt' and other processes leading to cell lysis (e.g., programmed cell death), and even the lines of communication within and between microbial populations affect the retention of the bio-elements within the microbial loop. As understanding of the interactions of marine microbiomes, biogeochemical cycles, and the flow of energy evolves, so too must the models which describe them. A move beyond coarse allometric models and NPZ (nutrient, phytoplankton, zooplankton) models is needed, but formalizing the many interactions mathematically is a daunting task.

Motivation

The motivation for my dissertation was to investigate several aspects of the complexity of marine microbial metabolism and physiology as they interact with biogeochemical cycles and direct the flow of energy within the Station ALOHA surface layer microbial community. The order of the chapters was chosen to progressively look deeper into the complexity of metabolism, beginning with the broadest scale – allometry. In Chapter 1, we develop a method coupling high-resolution size fractionation to flow cytometrically derived cell carbon quotas and radioisotope tracer incubations to determine, at fine scale, the relationship between cell size and metabolic activity, within populations of picoplankton at Station ALOHA. Next, in Chapter 2, we develop a set of guidelines by which heterotrophic microbial populations select organic substrates from the dissolved organic matter pool. In Chapter 3, we focus on the dynamics of one substrate, glycolate, and the role of light in its metabolism. In the remaining two chapters we take one final step closer, focusing in on the metabolism of one particular organism, *Prochlorococcus*. In Chapter 4, we explore the evolution of a *Prochlorococcus* metabolic network and its adaptations to growth at low phosphate concentrations. Finally, in Chapter 5, we incorporate gene expression time-series data into our *Prochlorococcus* metabolic model to predict physiological changes and metabolic fluxes at diel time scales.

References

Ascani, F., Richards, K.J., Firing, E., Grant, S., Johnson, K.S., Jia, Y., Lukas, R., Karl, D.M., 2013. Physical and biological controls of nitrate concentrations in the upper subtropical North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography* 1–16.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., 1983. The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series* 10, 257-263.

Dugdale, R., Goering, J., 1967. Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnology and Oceanography* 12, 196–206.

Dugdale, R. C., Menzel, D.W., Ryther, J.H., 1961. Nitrogen fixation in the Sargasso Sea. *Deep Sea Research* 7, 297-300.

Dutkiewicz, S., Follows, M.J., Parekh, P., 2005. Interactions of the iron and phosphorus cycles: A three-dimensional model study. *Global Biogeochemical Cycles* 19, GB1021.

Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* 281, 200–206.

Guidi, L., Calil, P.H.R., Duhamel, S., Björkman, K.M., Doney, S.C., Jackson, G.A., Li, B., Church, M.J., Tozzi, S., Kolber, Z.S., Richards, K.J., Fong, A.A., Letelier, R.M., Gorsky, G., Stemmann, L., Karl, D.M., 2012. Does eddy-eddy interaction control surface phytoplankton distribution and carbon export in the North Pacific Subtropical Gyre? *Journal of Geophysical Research* 117, G02024.

Johnson, K.S., Riser, S.C., Karl, D.M., 2010. Nitrate supply from deep to near-surface waters of the North Pacific subtropical gyre. *Nature* 465, 1062–1065.

Jungbluth, S.P., Grote, J., Lin, H.-T., Cowen, J.P., Rappé, M.S., 2012. Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. *The ISME Journal* 7, 161–172.

Karl, D.M., 2014. Microbially mediated transformations of phosphorus in the sea: New views of an old cycle. *Annual Review of Marine Science* 6, 279–337.

Karl, D.M., Church, M.J., 2014. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nature Reviews Microbiology* 12, 699–713.

Karl, D.M., Bidigare, R.R., Church, M.J., Dore, J.E., Letelier, R.M., Mahaffey, C., Zehr, J.P., 2008. The nitrogen cycle in the North Pacific Trades biome: an evolving paradigm, in: *Nitrogen in the Marine Environment*, Elsevier Inc., 705–769.

Lovley, D.R., Phillips, E.J.P., Gorby, Y.A., Landa, E.R., 1991. Microbial reduction of uranium. *Nature* 350, 413–416.

Mopper, K., Stubbins, A., Ritchie, J.D., Bialk, H.M., Hatcher, P.G., 2007. Advanced instrumental approaches for characterization of marine dissolved organic matter: extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chemical Reviews* 107, 419-442.

Nishihara, H., Miyata, Y., Miyashita, Y., Bernhard, M., Pohlmann, A., Friedrich, B., Takamura, Y., 2001. Analysis of the molecular species of hydrogenase in the cells of an obligately chemolithoautotrophic, marine hydrogen-oxidizing bacterium *Hydrogenovibrio marinus. Bioscience, Biotechnology, and Biochemistry* 65, 2780–2784.

Shivaji, S., Chaturvedi, P., Suresh, K., Reddy, G.S.N., Dutt, C.B.S., Wainwright, M., Narlikar, J.V., Bhargava, P.M., 2006. *Bacillus aerius* sp. nov., *Bacillus aerophilus* sp. nov., *Bacillus stratosphericus* sp. nov. and *Bacillus altitudinis* sp. nov., isolated from cryogenic tubes used for collecting air samples from high altitudes. *International Journal of Systematic and Evolutionary Microbiology* 56, 1465–1473.

Steinberg, D.K., Carlson, C.A., Bates, N.R., Johnson, R., Michaels, A.F., Knap, A.H., 2001. Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-

scale look at ocean biology and biogeochemistry. *Deep Sea Research Part II: Topical Studies in Oceanography* 48, 1405–1447.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., Cornejo-Castillo, F.M., Costea, P.I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J.M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B.T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans coordinators, Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M.B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S.G., Bork, P., Boss, E., Bowler, C., Follows, M., Karp-Boss, L., Krzic, U., Reynaud, E.G., Sardet, C., Sieracki, M.E., Velayoudon, D., 2015. Structure and function of the global ocean microbiome. *Science* 348, 1261359.

Suttle, C.A., 2007. Marine viruses – major players in the global ecosystem. *Nature Reviews Microbiology* 5, 801-812.

Chapter 1

Size dependence of physiology and metabolism within marine picoplankton

populations

John R. Casey, Karin Björkman, Sara Ferrón, David M. Karl

Abstract

Cell size is increasingly applied as a convenient parameterization of ecosystem models, and is widely applicable to constraining the activity of organisms spanning large size intervals. However, the size structure of the majority of the marine picoplankton assemblage is narrow and at the lower limit of the empirical allometric relationships established so far. We applied a fine-resolution size fractionation method to estimate the photophysiological attributes of North Pacific Subtropical Gyre microbial assemblages. Group-specific carbon retained on each filter was quantified by flow cytometric conversion of light scatter to cellular carbon quotas. Median carbon quotas were 31.5, 36.2, and 9.8 fg C cell^{-1} for surface populations of *Prochlorococcus*, high-scatter bacteria, and low-scatter bacteria, respectively. Carbon-specific rates of primary production, using the ^{14}C method, and phosphate transport, using ^{32}P radiotracers, as a function of cell size resulted in negative power scalings (b) within populations of the picocyanobacterium *Prochlorococcus* and non-pigmented bacteria of $b = -1.3$ and $b = -1.1$, respectively. These findings are in contrast to the positive empirical power scaling comprising the broader prokaryote category ($b = 0.3$), and point to intricate and dynamic variability in cell physiology and metabolism within these important microbial groups.

13

1.1 Introduction

The empirical scaling of metabolic rates to body size, in the form $Y = Y_0M^b$, has been carefully studied for nearly two centuries (Sarrus and Rameaux, 1839; Robiquet and Tillaye, circa 1839 in Klieber, 1932). When organisms spanning more than 20 orders of magnitude in body size are compared, an organism-specific scaling exponent b_{org} of 3/4 quite accurately predicts metabolic rates. This allometric scaling 'law' has been explained by the fractal nature of vascular transport abridging surface area-dependent (proportional to $b_{org} = 2/3$) and volume-dependent (proportional to an isometric b_{org}) rates (West *et al.,* 1997). More recently, other scaling exponents which better describe prokaryotes ($b_{org} > 1$) and protists ($b_{org} = 1$) have been added, citing genome size and endosymbiotic surface area as constraints (DeLong *et al.,* 2010). Convenient as it is in reducing the dimensionality of ecosystem models, the allometric scalings are a lynchpin of the so-called Metabolic Theory of Ecology (MTE; Brown *et al.,* 2004a) and are leveraged for maximum intrinsic growth rates and substrate transport rates (Aksnes and Egge, 1991; Litchman *et al.,* 2007; Finkel *et al.,* 2010; Edwards *et al.,* 2012), assimilation numbers (Finkel, 2001), and elemental stoichiometry (Finkel *et al.,* 2010; Marañón *et al.,* 2013) by 'trait-based' ocean ecosystem models (Follows and Dutkiewicz, 2011). An extensive review of allometric relationships in microbial oceanography is found elsewhere (Marañon, 2015).

As Brown and co-authors of the original MTE paper (Brown *et al.,* 2004a) responded to a very active forum on the topic (Brown *et al.,* 2004b), "When body mass differs by only two- or threefold, or temperature varies by only a degree or

two, other factors can assume equal or greater importance." At Station ALOHA, a site quite similar to vast regions of the worlds oceans, microbial particle size distributions within the range 0.2 – 1.0 μm comprise most of the biomass and photosynthesis (Rii *et al.,* 2016), and temperatures vary seasonally by roughly 3 °C. Accordingly, it should be established if these 'other factors,' namely metabolism, physiology, and community interactions, are indeed important.

In the surface layer at Station ALOHA, the majority of cells within the 0.2 – 1.0 μm size interval belong to the cyanobacterial genus *Prochlorococcus* and the alphaproteobacterial orders Pelagibaterales (SAR11 clade) and Rhodobacterales (*Roseobacter* clade; Ottesen *et al.,* 2014). We asked if allometric relationships could be used to predict within-species metabolic rates for *Prochlorococcus* and small bacteria, and whether these relationships are consistent within the broader prokaryote size interval. We used 0.05 μm size fractionation of radioisotope tracer incubations and subsequent flow cytometric analysis of cellular carbon spectra retained on filters in the interval 0.10 – 1.00 μm to calculate carbon-specific metabolic rates as a function of body size. Our results indicate a negative power law relating carbon-specific rates of primary production and phosphate uptake to body size within populations, in contrast to a positive power law found at broader prokaryote body size scales.

1.2 Methods

1.2.1 Study site and sample collection

Field experiments were conducted on several expeditions (Cruise 1 -

September 2013, Cruise 2 - December 2013, Cruise 3 - May 2015) at Station ALOHA,

a site due North of Kahuku Point, O'ahu (22° 45' N, 158° 00' W). Water samples were

collected using a rosette of PVC Bullister bottles mounted on a frame equipped with

dual conductivity sensors, temperature sensors, pressure sensors, oxygen optodes,

transmissometers, and triplet fluorometers (SBE 911plus, Sea-Bird Electronics,

Inc.). Fast repetition rate fluorometry measurements and experiments designed to

compare the effectiveness of different filtration methods were conducted on Cruise

1. ^{14}C-PP and ^{32}P-PO$_4$$^{3-}$ experiments were conducted on Cruise 3.

1.2.2 Optical measurement of photosynthetic parameters

The cross sectional area of photosystem II (σ_{PSII}) and the ratio of variable

fluorescence to maximal fluorescence emission (F_V/F_M) were measured by fast

repetition rate fluorometry (FRRF; Kolber *et al.,* 1998) using a spectrally resolved

bench-top FRRF with excitation at 445±20, 470±25, 503±30, and 530±35 nm bands

(described in detail in Wilson *et al.,* 2012). Cell suspensions of in-line filtration

isolated fractions were dark adapted for 30 minutes prior to analysis. F_V/F_M is

calculated by FRRF as the ratio of the variable multispectral fluorescence (the

difference between maximum and dark adapted fluorescence yields) and the

maximum fluorescence, when all available photosystem II reaction centers are

reduced. σ_{PSII} values are reported for 470/40 nm excitation only.

1.2.3 ^{14}C-primary production measurements

Rates of inorganic carbon assimilation were measured by the $H^{14}CO_3^-$ tracer method according to standard Hawaii Ocean Time-series protocols (Karl and Dore, 2001; http://hahana.soest.hawaii.edu/hot/methods/results.html). 500 ml seawater samples collected in triplicate from 15 m were inoculated with 10 µCi $H^{14}CO_3^-$ and incubated for 12 hours beginning prior to sunrise and terminated after sunset. Deckboard polycarbonate incubators, shaded to approximately the blue light attenuation at 15 m, were flushed with surface seawater to maintain in situ temperatures. Incubations were terminated by filtration (details below). A 100 µl specific activity sample was collected from each incubation and added to a 20 ml scintillation vial containing 2 ml β-phenethylamine, a CO_2 trap. Filters were placed in 20 ml scintillation vials, acidified to 0.1 N hydrochloric acid and allowed to vent for 24 hours prior to liquid scintillation counting. This measurement approximates a rate between net and gross primary production (Marra, 2002).

1.2.4 ^{32}P-PO$_4^{3-}$ uptake measurements

Phosphate assimilation rates were measured using the radioisotope tracer ^{32}P-PO$_4^{3-}$ (Bjorkman and Karl, 2003). 500 ml seawater samples were collected in triplicate from 15 m and inoculated with 10 µCi carrier-free ^{32}P-PO$_4^{3-}$, which was approximately 0.002 % of the ambient soluble reactive phosphorus concentration. Incubations were conducted for 4.5 h as described above, from 0900-1330 local

time. Incubations were again terminated by filtration (details below) and processed similarly to ^{14}C-PP filters.

1.2.5 Culture H^{14}CO$_3^-$ assimilation experiment

Prochlorococcus strain MIT9301 was acclimated for several generations to growth in PRO99 medium (Moore *et al.,* 2007) at 24 °C in a sinusoidal 14 h light : 10 h dark cycle reaching a maximum irradiance of 20 µmol quanta m^{-2} s^{-1} blue light. 1 ml aliquots of culture suspension were inoculated with 10 µCi H^{14}CO$_3^-$ for 1 h and terminated by 0.5 % paraformaldehyde fixation in the dark at 4 °C for 30 minutes. Fixed samples were diluted to 100 ml 0.2 µm filtered seawater prior to filtration (details below). Filters were processed identically to the above description for natural assemblages. This shorter interval incubation more closely approximates gross primary production, and therefore cannot be compared directly to 12 h ^{14}C-PP field measurements.

1.2.6 Custom membranes

A series of "Plastics-Irradiated-Etched" (PIE; Karl, 2007) 47 mm diameter polycarbonate membrane filters spanning the pore-size range 0.10 – 1.00 µm at 0.05 µm intervals were custom fabricated (Sterlitech Corp.) to finely size fractionate particles. To validate manufacturer stated pore sizes, filter surfaces (shiny side) were imaged by scanning electron microscopy (SEM; Hitachi S-4800 field emission scanning electron microscope; Hitachi) fitted with a silicon drift detector (Oxford INCA X-Act energy dispersive spectroscope; Oxford). Images were analyzed by

automated detection of pore edges and measuring diameter of at least 20 pores per filter using embedded scale bars (Fiji, Schindelin *et al.,* 2012).

1.2.7 Filtration

Several filtration methods were employed, as follows:

In-line filtration – This approach was designed not for quantitative sample recovery, but rather to concentrate filter fractions with as narrow a size spectrum as possible. A series of 10 47 mm diameter polycarbonate filter holders (Nalgene, Thermo Fisher Scientific Inc.), equipped with Luer bleed ports, were plumbed to the spigot of a 10 L carboy containing the seawater sample at the inlet end and dispensed by peristaltic pumping to the mouth of a 10 L carboy receptacle at the outlet end (**Figure 1.1**). Filter holders were housed in a custom-built rack mounted to lab Unistrut. Filter pore sizes were arranged from largest (1.00 μm) to smallest (0.10 μm), again at 0.05 μm intervals. The entire set of 19 pore sizes could not practically be joined together, so two intervals (1.00 μm to 0.55 μm, 0.55 μm to 0.10 μm) were filtered separately. A technique to improve the retention of only particles with diameters within each pore size interval was as follows: following sample filtration, particles were resuspended by repeated purging of the hold-up volume of the filter holder using a syringe attached to the bleed ports. Filters were then back-flushed by reversing the flow direction and drawing from a 10 L carboy of 0.2 μm filtered seawater. This procedure was repeated three times before sampling.

Sequential filtration – This approach was designed for quantitative sample recovery, for analyses sensitive enough to detect filtered fractions with very little retained

biomass. A 500 ml sample was filtered by gentle vacuum (< 70 mBar differential) using a self-contained filter holder and receiver. The collected filtrate was then filtered with the next smaller pore size filter using a separate filter holder, and this process repeated until all fractions were collected.

Parallel filtration – This approach was designed for less sensitive analyses, but provides only semi-quantitative recovery. Although the filtration setup is identical to the sequential filtration method, unfiltered sample is instead split into 500 ml volumes and filtered separately.

1.2.8 Flow cytometry

Samples for flow cytometry were collected from whole seawater samples and from filtrate fractions. For the in-line filtration method, samples were collected from the filter holder dead volume using a syringe attached to the bleed port after resuspension. In sequential or parallel filtration, samples were collected by pipeting from the receiver. 1 ml samples were fixed with paraformaldehyde (0.2 μm syringe filtered; 0.5 % final concentration) in the dark at 4 °C for 30 minutes prior to flash freezing in liquid nitrogen and storage at -80 °C. Samples were analyzed using a jet-in-air Influx Cell Sorter (BD Biosciences, Inc.) equipped with a small particle detector (100X magnification objective in the forward-angle light scatter path) and 200 mW 457 nm and 488 nm lasers (Coherent, Inc.) aligned through the same pinhole. Analog pulse height signals from 530/20 nm, 580/30 nm, and 692/40 nm bandpass filters were log amplified from photomultiplier tube detectors (Hamamatsu, Inc.) and data were collected in listmode files using the operating

software Spigot (BD Biosciences, Inc.). Particles were quantified and discriminated based on their autofluorescence and light scattering properties (Casey *et al.,* 2013), as well as their fluorescence of the nucleic acid stain SYBR Green I (Invitrogen). 0.53 µm polystyrene calibration particles impregnated with Nile Red dye (Spherotech, Inc) were used for optical and fluidic alignment and as an internal scattering standard. Approximately 300 µl of unstained and SYBR Green I stained sample was acquired to ensure a sufficient number of events were recorded. Concentrations were determined using the gravimetric method (Casey *et al.,* 2013).

1.2.9 Data analysis

A method for quantifying population-specific carbon retained on each filter was developed and is available as a Matlab toolbox (Supplementary File 1). The method converts forward-angle light scatter (FSC) signals to cellular carbon content using a previously determined calibration curve (Casey *et al.,* 2013). FSC data of gated populations and internal standard particles are used to generate normalized FSC histograms. To minimize spurious events at the distribution extremes (which can be important for large, rare particles), distributions were smoothed using a normal kernel function with bin widths scaled to the coefficient of variation of the internal standard particles, and the number of bins set equal to the number of FSC channels spanning 95% of the distribution. After applying the carbon-FSC calibration, the carbon transformed scale is then used to generate cellular carbon spectra for each filtrate sample. The carbon spectrum of each filtrate is then subtracted from the next

largest filtrate spectrum and the residual spectrum is integrated to calculate total

particulate carbon caught on each filter.

1.3 Results

1.3.1 Comparison of filtration approaches

SEM imaging was used to quantify the distribution of pore sizes of a

representative filter from each batch of each pore size filter used in this study. Filter

pore sizes were on average within 12 – 16% of manufacturer specifications in the

0.10 – 0.35 μm interval and 3 – 7 % in the 0.40 – 1.00 μm interval. Model I least

squares regression of median pore sizes as a function of manufacturer specified

pore sizes was not significantly different from unity (p = 0.81; **Figure 1.2**) and no

residuals trend was apparent. Three filtration techniques – in-line, parallel, and

sequential (see Methods for a description of these designs) – were tested for

separation effectiveness and recovery. The in-line filtration approach, especially

with resuspension of particles from the filter and recursive backflushing, was

necessary to adequately isolate size fractions, with the third moment of FSC (0.41 ±

0.19) significantly lower than parallel (1.8 ± 0.9) or sequential filtration (1.2 ± 0.6)

techniques (Mann-Whitney Test; p = $7e^{-5}$ and p = $2e^{-4}$, respectively). The positive

(right) skewness of the sequential filtration technique is important, and indicates

that particles smaller than the filter pore size are caught on polycarbonate

membranes (**Figure 1.3**). Membrane filtration is instead best suited to capturing

particles larger than the pore size, though overlapping pores were frequent (2 – 8 %

of total pores imaged) in our electron micrographs. These particles are contaminants in any size fractionation experiment using this filtration technique.

1.3.2 Separation of picoplankton populations

A coarse resolution depth profile (surface, 150 m, and 500 m) collected during Cruise 1 was size fractionated to test whether the major picoplankton populations at Station ALOHA (*Prochlorococcus* and non-autofluorescent bacteria and archaea; for brevity we will refer to this population as "bacteria" herein) could be discretely separated by size. At the surface, these populations overlapped in size spectra almost identically, with a small fraction (3 %) of bacteria present in the 0.30 – 0.35 μm fraction where *Prochlorococcus* was absent (**Figure 1.4**). The bimodal size spectra of *Prochlorococcus* and bacteria is a striking feature, with peaks at 0.55 μm and 0.40 μm in both populations. For *Prochlorococcus*, this distribution is likely due to the cell cycle stage distribution, (this interval corresponds to a 1.8 – 2.6 fold change in cell volume), and *Prochlorococcus* is known to exhibit synchronous division just after sunset (Zinser *et al.,* 2009; Waldbauer *et al.,* 2012; Ribalet *et al.,* 2015), however physiological differences between strains have not been excluded. Synchronicity in natural non-pigmented bacterial population cell division has not previously been reported. At 150 m, just beneath the deep chlorophyll maximum depth (approximately 125 m during Cruise 1), *Prochlorococcus* and bacteria were clearly separated unimodal populations with only 10 % of *Prochlorococcus* within the 0.25 – 0.50 μm bacteria fraction and 7 % of bacteria within the 0.50 – 0.85 μm *Prochlorococcus* fraction, indicating the utility of this method for e.g., down-stream

isotopic or physiological measurements at this depth. At 500 m, no *Prochlorococcus* cells were detected, but a bimodal population of bacteria (and likely archaea; Karner *et al.,* 2001) was detected.

1.3.3 Size spectrum of Prochlorococcus photophysiology

Fast repetition rate fluorometry of in-line filtration isolated picoplankton size fractions revealed a bimodal distribution of maximum photochemical efficiency (F_V/F_M), coincident with the distribution of *Prochlorococcus* cells, despite a constant effective cross sectional area of photosystem II (σ_{PSII}; **Figure 1.5**). σ_{PSII} values in the *Prochlorococcus* size interval 0.35 – 0.85 µm (503 ± 27 A^2). F_V/F_M values spanned the range 0.12 – 0.49 over the constant σ_{PSII} interval of 0.35 – 0.85 µm, which accounted for 88 % of the *Prochlorococcus* population. Both σ_{PSII} and F_V/F_M values were consistent with previously reported results from nitrogen replete and nitrogen deprived *Prochlorococcus* isolates (σ_{PSII} range 430 – 600 A^2; F_V/F_M range 0.2 – 0.6; Steglich *et al.,* 2001).

1.3.4 Size spectrum of picoplankton carbon

Based on samples from 25 m during Cruise 3, cellular carbon content comprised the range 4 – 125 fg C cell^{-1} over the 0.10 – 1.00 µm pore size interval. The resulting carbon density of cells in this range was 240 fg C µm^3, similar to previously reported values (references compiled in Casey *et al.,* 2013) which range from 190 – 470 fg C µm^3 with a median value of 228 fg C µm^3 based on 12 studies. Cell carbon quotas were log-normally distributed, and centered about the median

values 31.5, 36.2, and 9.8 fg C cell^{-1} for *Prochlorococcus*, high-scatter bacteria, and low-scatter bacteria, respectively (**Table 1.1**). These two bacterial populations are often referred to as high and low nucleic acid bacteria, however many configurations are possible, and we unambiguously refer to the populations by their scattering amplitude. The sum of carbon retained on filters, as calculated by the histogram subtraction method (described in Methods; **Figure 1.6** and **Figure 1.7**), recovered 107 ± 8 % of the initial picoplankton carbon. The discrepancy in summed fractions from the total is due to the necessary smoothing function applied to forward-angle light scatter distributions. *Prochlorococcus* contributed 68 ± 2 % (748 ± 18 nmol C L^{-1}), while bacteria and *Synechococcus* contributed the remainder (23 ± 1 % and 8 ± 0 %, respectively) of the total picoplankton carbon (1101 ± 21 nmol C L^{-1}). Exponentially growing *Prochlorococcus* MIT9301 cells were markedly larger (0.65 – 0.80 μm) than the natural population, with an average carbon quota of 57 fg C cell^{-1}.

1.3.5 Size spectrum of picoplankton metabolic rates

Again based on samples from 25m during Cruise 3, ^{14}C-PP varied by roughly an order of magnitude over the *Prochlorococcus* size interval (0.35 – 0.65 μm), with maxima coinciding with cell number and carbon retained on the 0.55 μm filter (**Figure 1.8**). Radioactivity was detectable at low levels in the 0.15 – 0.35 μm interval, where few (< 0.1 %) or no *Prochlorococcus* cells could be detected, likely an indication of bacterial uptake of recently fixed photosynthate or anapleurotic carbon fixation. Low levels of ^{14}C-PP at larger size intervals were due to low

concentrations of small *Synechococcus* cells as well as *Prochlorococcus* cells caught

on larger pore size filters (20 % of total). $^{32}PO_4^{3-}$ assimilation also varied by roughly

an order of magnitude over the bacteria and *Prochlorococcus* size interval (0.20 –

0.65 µm), with maxima coinciding with the bacterial cell number maximum at 0.40

µm, but not the carbon maximum at 0.55 µm (**Figure 1.9**).

1.3.6 Allometric relationships

      Prochlorococcus carbon specific ^{14}C-PP rates (nmol C L^{-1} h^{-1} [nmol C L^{-1}]$^{-1}$)

and picoplankton carbon specific $^{32}PO_4^{3-}$ assimilation rates varied as a power law

function of pore size converted to cell volume ($Y_C = Y_{C,0}M^b$, where C is the cellular

carbon content); the calculated exponent *b* was -1.3 ± 0.4 (p = 2e^{-7}) and -1.1 ± 0.4 (p

= 1e^{-5}), respectively (**Figure 1.10** and **Figure 1.11**, respectively). By weighting

Prochlorococcus carbon specific ^{14}C-PP rates to the *Prochlorococcus* carbon size

spectrum, the mean population growth rate was 0.58 ± 0.15 d^{-1}, similar to maximal

growth rates of high-light adapted ecotypes in culture (0.63 ± 0.06 d^{-1}; Moore *et al.,*

1995), though not as high as recent estimates from surface waters at Station ALOHA

(0.7 – 0.9 d^{-1}; Ribalet *et al.,* 2015). The calculated *b* for *Prochlorococcus* MIT9301

carbon specific $H^{14}CO_3^-$ assimilation rates was -2.2 (p = 0.04; **Figure 1.12**). Growth

rates estimated from a linear scaling of $H^{14}CO_3^-$ assimilation rates to the daily

incubation light profile were 0.31 d^{-1}.

1.4 Discussion

When cautiously interpreted, size fractionation is an effective tool for probing the variability of physiology and metabolism within natural food web components and, at the resolution utilized in this study, within microbial populations. Whatever the mechanism, be it flow impaction or pore clogging, particles smaller than the filter pore size are retained; depending on the application, this fraction may significantly bias size fractionation results (**Figure 1.3**). Accounting for this effect, we coupled size fractionation of radioisotope incubations to flow cytometric determinations of cellular carbon and observed significant and non-monotonic (size-independent) variability in photophysiology and metabolic rates within natural picoplankton populations.

Photophysiology variability within the natural *Prochlorococcus* population was similar to the variability observed between nitrogen stressed and nitrogen replete laboratory isolates, suggesting effects on photochemical efficiency are due to distributions of cell-cycle stages or rather due to photophysiological differences between individual strains or 'ecotypes', although the latter cannot be rejected. Constant σ_{PSII} indicates a uniform physiological status of PSII reaction centers across the size spectrum, and is inversely related to F_V/F_M as a function of nutrient supply, within algal classes, and between algal classes (summarized by Suggett *et al.,* 2009). Thus it is intriguing that peaks in F_V/F_M within the *Prochlorococcus* population were associated with two cell abundance maxima, coincidentally separated by a roughly 2-fold difference in cell volume. The late afternoon sampling time coincides with the initiation of cell division in natural *Prochlorococcus* populations (Ribalet *et al.,* 2015)

27

is suggestive of cell-cycle associated changes in photophysiology, though cytometric cell cycle analysis was not performed on these samples.

The relationship between carbon-specific primary productivity or phosphate assimilation rates and cell size differed in our study from those predicted by the broader prokaryote allometric relationship. In natural *Prochorococcus* populations, bacterial populations, and a *Prochlorococcus* isolate MIT9301, the exponents were all negative (b = -1.3, b = -1.1, and b = -2.2, respectively), rather than positive (b = 0.30; DeLong *et al.,* 2010), suggesting that within-population size dependence is independent of between-population size dependence. In both natural populations and a laboratory isolate, a highly negative b indicates that the smallest cells are disproportionately active. In the case of *Prochlorococcus*, which is phase locked to the solar cycle for initiation of cell division, it is possible that larger cells are arrested in cell cycle stage. Remnants of the previous day's cycle, perhaps these cells require very little additional resources to satisfy the remaining carbon and energy quota, and are simply awaiting the signaling cascade to initiate genome replication and division. Although such a synchronicity has not yet been reported for other bacterioplankton, diel rhythms in gene expression are widespread (Ottesen *et al.,* 2014; Aylward *et al.,* 2015) and may include replication.

The empirical allometric power law is a useful approximation of ecosystem function when the size distribution of organisms spans several orders of magnitude. However, the universal b_{org} = ¾ is a victim of the Lumberjack Problem (emergent trends when considering too many logs), and has been challenged by closer inspection of b_{org} within and between microbial groups (DeLong *et al.,* 2010;

reviewed in Marañon, 2015). Transformation of the newly accepted group-specific power functions by normalizing rates to biomass would indicate a positive b ($b = 0.30$; Maranon *et al.,* 2013) within the small prokaryotes, a near zero b within the protists, and $b = -1/4$ within the metazoans. Is it then reasonable to approximate the metabolic rates of prokaryotic life by size alone? Put differently, is the evolution of metabolic diversity, the vast physiological differences between microbes, and the give-and-take of microbial community interactions irrelevant?

At the smallest scales (< 1 μm), the size dependence of microbial metabolism had not until recently (García *et al.,* 2015) been investigated in natural marine bacterioplankton populations, and thus the lower extent of the size range which applies to the prokaryotes $b = 0.30$. Using flow cytometric calibration of cell size and the spectrophotometric determination of respiration by 2-para(iodo-phenyl)-3(nitrophenyl)-5(phenyl) tetrazolium chloride (INT chloride) reduction, García *et al*. (2015) found a cell-specific respiration exponent $b = 1.67 \pm 0.19$, quite close to Delong et al. (2010) estimates ($b = 1.72 \pm 0.07$), and a curiously calculated biomass-specific respiration exponent $b = -1.43$ (García *et al*., 2015). The relationship between small angle light scatter from a cuvette-based flow cytometer and cell size is problematic (Becker *et al.,* 2002; Casey *et al.,* 2013), and rates were normalized to a seemingly arbitrary biovolume of 0.068 μm^3 for cells captured on 0.2, 0.4, and 0.6 μm pore size filters. For these and other reasons, we have not directly compared results from this study, though it is nonetheless interesting that a negative slope was found.

While our results do not necessarily refute previously reported allometric relationships within the prokaryotes, we do demonstrate that within the size spectrum of the smallest prokaryotes exists complexity in photophysiology and metabolic rates which do not scale with broader cell size comparisons. Perhaps it will be some time yet before allometry is applicable to marine microbial communities.

1.5 References

Aksnes, D.L., Egge, J.K., 1991. A theoretical model for nutrient uptake in phytoplankton. *Marine Ecology Progress Series* 70, 65-72.

Aylward, F.O., Eppley, J.M., Smith, J.M., Chavez, F.P., Scholin, C.A., Delong, E.F., 2015. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proceedings of the National Academy of Sciences of the United States of America* 112, 5443-5448.

Becker, A., Meister, A., Wilhelm, C., 2002. Flow cytometric discrimination of various phycobilin-containing phytoplankton groups in a hypertrophic reservoir. *Cytometry* 48, 45–57.

Bjorkman, K.M., Karl, D.M., 2003. Bioavailability of dissolved organic phosphorus in the euphotic zone at Station ALOHA, North Pacific Subtropical Gyre. *Limnology and Oceanography* 48, 1049–1057.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., West, G.B., 2004. Toward a metabolic theory of ecology. *Ecology* 85, 1771-1789.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., West, G.B., 2004b. Response to forum commentary on "Toward a metabolic theory of ecology". *Ecology* 85, 1818–1821.

Casey, J., Aucan, J.P., Goldberg, S.R., Lomas, M.W., 2013. Changes in partitioning of carbon amongst photosynthetic pico- and nano-plankton groups in the Sargasso Sea in response to changes in the North Atlantic Oscillation. *Deep Sea Research Part II: Topical Studies in Oceanography* 93, 58–70.

DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M., Brown, J.H., 2010. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences of the United States of America* 107, 12941-12945.

Edwards, K.F., Thomas, M.K., Klausmeier, A., Litchman, E., 2012. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. *Limnology and Oceanography* 57, 554-566.

Finkel, Z.V., 2001. Light absorption and size scaling of light-limited metabolism in marine diatoms. *Limnology and Oceanography* 46, 86-94.

Finkel Z.V., Beardall J., Flynn K.J., Quigg A., Rees T.A.V., Raven J.A., 2010. Phytoplankton in a changing world: cell size and elemental stoichiometry. *Journal of Plankton Research* 32, 119–37

Follows, M.J., Dutkiewicz, S., 2011. Modeling diverse communities of marine microbes. *Annual Review of Marine Science* 3, 427–451.

García, F.C., García-Martín, E.E., Taboada, F.G., Sal, S., Serret, P., López-Urrutia, Á., 2015. The allometry of the smallest: superlinear scaling of microbial metabolic rates in the Atlantic Ocean. *The ISME Journal* 10, 1029-1036.

Karl, D.M., Dore, J.E., 2001. Microbial ecology at sea: sampling, subsampling and incubation considerations. *Methods in Microbiology* 30, 13–39.

Karner, M., Delong, E.F., Karl, D.M., 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409, 507–510.

Kleiber, M., 1932. Body size and metabolism. *Journal of Agricultural Science* 6, 315-353.

Kolber, Z.S., Prášil, O., Falkowski, P.G., 1998. Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. *Biochimica et Biophysica Acta* 1367, 88-106.

Litchman, E., Klausmeier, C.A., Schofield, O.M., Falkowski, P.G., 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. *Ecology* 10, 1170-1181.

Marañon, E., Cermeño, P., López Sandoval, D.C., Rodríguez Ramos, T., Sobrino, C., Huete Ortega, M., Blanco, J.M., Rodríguez, J., 2013. Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. *Ecology Letters* 16, 371–379.

Marañon, E., 2015. Cell size as a key determinant of phytoplankton metabolism and community structure. *Annual Review of Marine Science* 7, 241–264.

Marra, J., 2002. Approaches to the measurement of plankton production, in: Phytoplankton productivity: carbon assimilation in marine and freshwater ecosystems. P.J. leB. Williams, D.N. Thomas, C.S. Reynolds. Cambridge, Blackwells, 78-108.

Moore, L.R., Coe, A., Zinser, E.R., Saito, M., Sullivan, M.B., Lindell, D., Frois-Moniz, K., Waterbury, J., Chisholm, S.W., 2007. Culturing the marine cyanobacterium *Prochlorococcus. Limnology and Oceanography: Methods* 5, 353–362.

Ottesen, E.A., Young, C.R., Gifford, S.M., Eppley, J.M., Marin, R., Schuster, S.C., Scholin, C.A., Delong, E.F., 2014. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* 345, 207–212.

Ribalet, F., Swalwell, J., Clayton, S., Jiménez, V., Sudek, S., Lin, Y., Johnson, Z.I., Worden, A.Z., Armbrust, E.V., 2015. Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proceedings of the National Academy of Sciences of the United States of America* 112, 8008-8012.

Rii, Y.M., Karl, D.M., Church, M.J., 2016. Temporal and vertical variability in picophytoplankton primary productivity in the North Pacific Subtropical Gyre. *Marine Ecology Progress Series* 562, 1–18.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9, 676–682.

Sarrus, P.; Rameaux, J.-F., 1839. Application des sciences accessoires et principalement des mathématiques àla physiologie générale. *Bulletin de l'Académie Royale de Médecine* 3, 1094–1100.

Steglich, C., Behrenfeld, M.J., Koblizek, M., Claustre, H., Penno, S., Prasil, O., Partensky, F., Hess, W.R., 2001. Nitrogen deprivation strongly affects Photosystem II but not phycoerythrin level in the divinyl-chlorophyll b-containing cyanobacterium *Prochlorococcus marinus. Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1503, 341–349.

Suggett, D.J., Moore, C.M., Hickman, A.E., Geider, R.J., 2009. Interpretation of fast repetition rate (FRR) fluorescence: signatures of phytoplankton community structure versus physiological state. *Marine Ecology Progress Series* 376, 1–19.

Waldbauer, J.R., Rodrigue, S., Coleman, M.L., Chisholm, S.W., 2012. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS ONE* 7, e43432.

West, G.B., 1997. A general model for the origin of allometric scaling laws in biology. *Science* 276, 122–126.

Wilson, S.T., Kolber, Z.S., Tozzi, S., Zehr, J.P., Karl, D.M., 2012. Nitrogen fixation, hydrogen cycling, and electron transport kinetics in *Trichodesmium erythraeum* (cyanobacteria) strain IMS101. *Journal of Phycology* 48, 595-606.

Zinser, E.R., Lindell, D., Johnson, Z.I., Futschik, M.E., Steglich, C., Coleman, M.L., Wright, M.A., Rector, T., Steen, R., Mcnulty, N., Thompson, L.R., Chisholm, S.W., 2009. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS ONE* 4, e5135.

Table 1.1 – Summary statistics of cell carbon quotas for natural populations of *Prochlorococcus* and bacteria. High-scatter (HS) bacteria and low-scatter (LS) bacteria were discriminated by their Sybr Green I fluorescence and side angle light scatter spectra. Concentrations are reported as 10^5 cells ml^{-1}; all other values are reported as fg C cell^{-1}. All values were computed as the mean ± standard deviation of three unfiltered seawater samples collected from 25 m on Cruise 3.

	Prochlorococcus	HS Bacteria	LS Bacteria
Concentration	2.2 ± 0.11	0.4 ± 0.1	4.1 ± 0.1
Variance	234.6 ± 19.1	116.0 ± 0.8	12.5 ± 0.2
Mean	34.4 ± 1.2	34.7 ± 0.3	10.0 ± 0.0
Median	31.5 ± 1.1	32.5 ± 0.4	9.8 ± 0.0
Mode	26.9 ± 6.2	35.4 ± 8.9	9.8 ± 1.2
Standard deviation	15.3 ± 0.6	10.7 ± 0.0	3.5 ± 0.0
Skewness	1.0 ± 0.1	0.6 ± 0.0	0.8 ± 0.0

Figure 1.1– In-line filtration apparatus. Carboys containing unfiltered and pre-filtered seawater are not pictured, but during normal operation are connected to the inlet and outlet, respectively.

Figure 1.2 – Comparison of manufacturer stated pore sizes (median diameter) and SEM calibrated pore sizes. Data are median values for each filter and error bars represent one standard deviation of the mean. Also shown are the line of unity (1:1; black line), a least squares regression (red line), and the 5% and 95% confidence intervals (red shaded region).
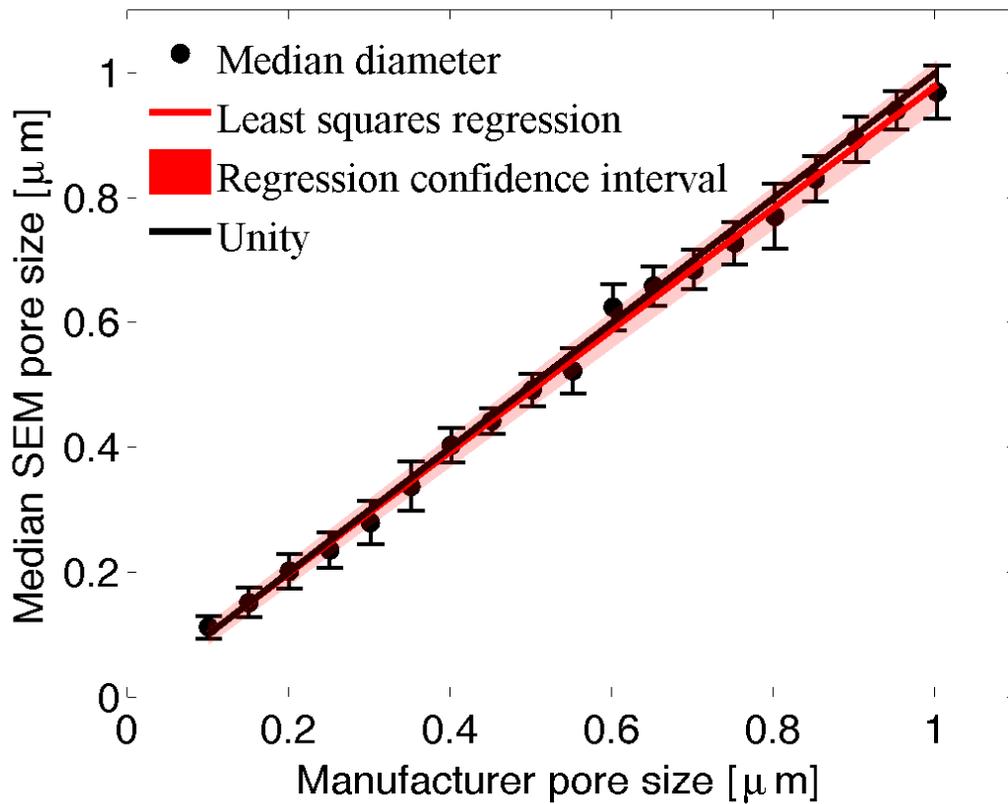
Figure 1.3 – Cell count normalized carbon spectrum of *Prochlorococcus* MIT9301 retained on each filter within the interval comprising 91% of the initial population. Vertical bars represent expected carbon quota of cells with equivalent spherical diameter matching the respective pore size. Note the proportion of each distribution with carbon quotas lower than the pore size derived carbon quota.
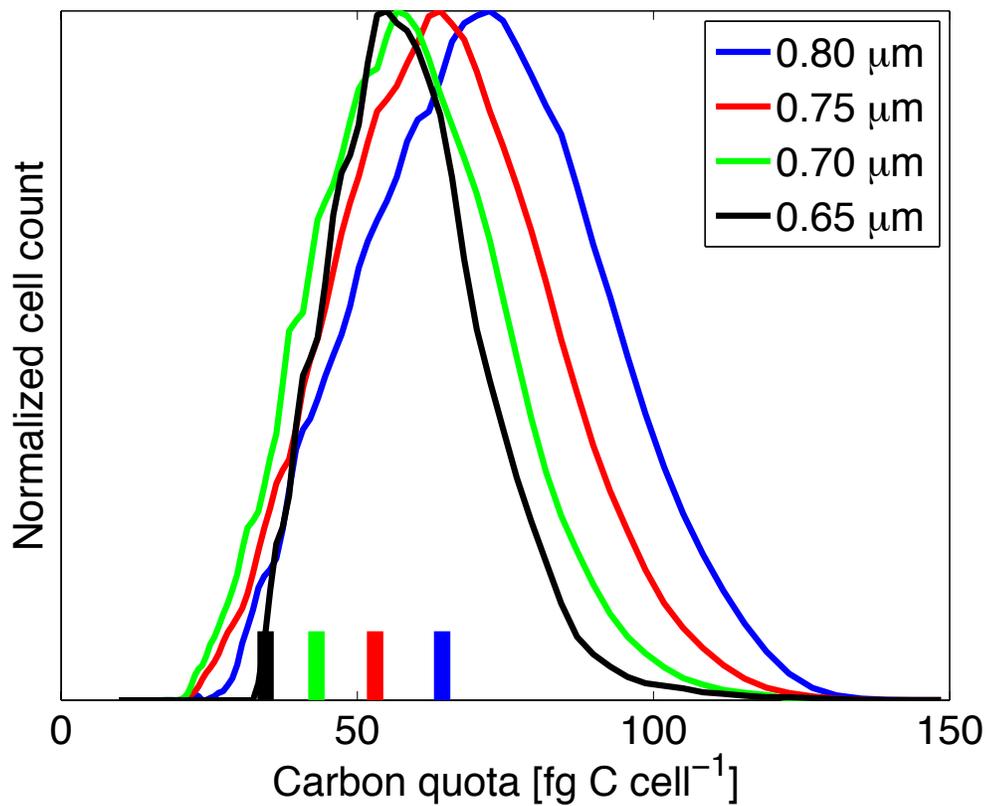
Figure 1.4 – Size spectrum of bacteria (top panel) and *Prochlorococcus* (bottom panel) counts retained on each filter collected at the surface, 150m, and 500m. "Bacteria" refer to non-pigmented microbes, including the Archaea. Note different scales for each group are indicated in the legend.
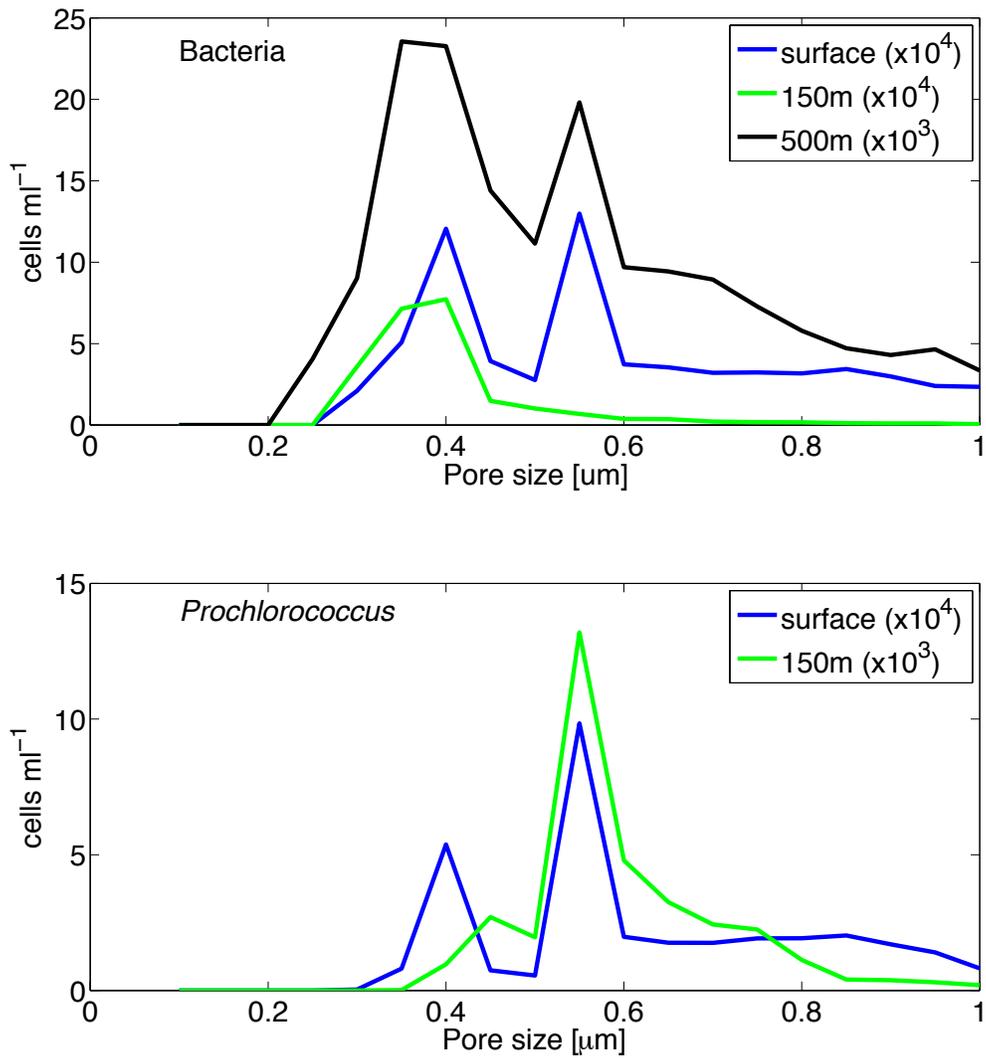
Figure 1.5 – Size spectrum of photophysiology. Variable fluorescence (F_V/F_M) and the effective cross sectional absorption area of photosystem II (σ_{PSII}) values are shown. Note multiple vertical axes scales.
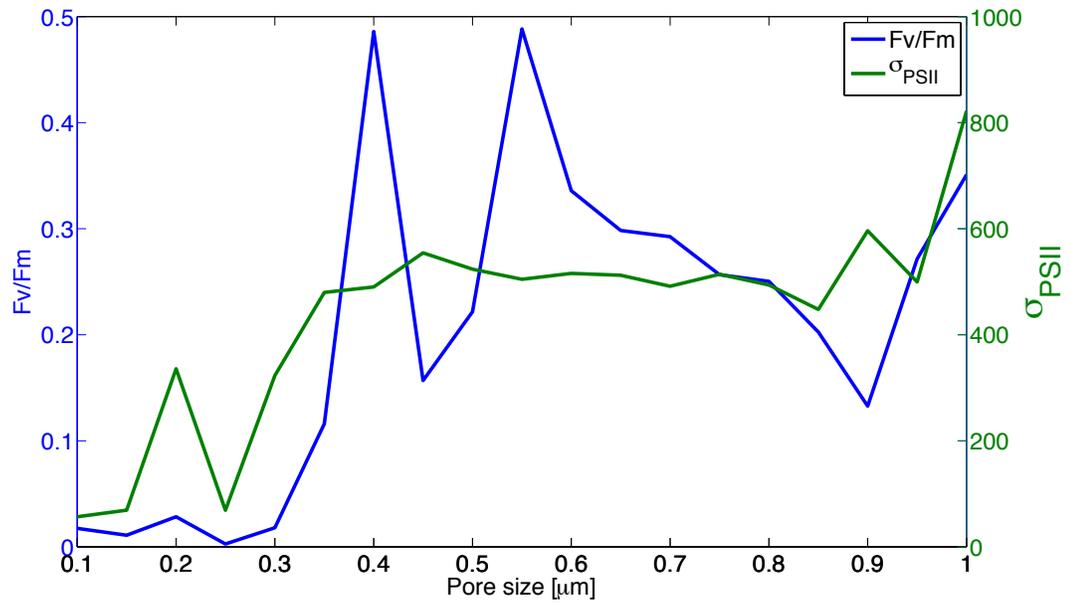
Figure 1.6 – Histogram of *Prochlorococcus* cell carbon quotas retained on each filter. Horizontal scale extends from 4 to 100 fg C cell^{-1}. Each column is a replicate sample, and all vertical scales are equal.
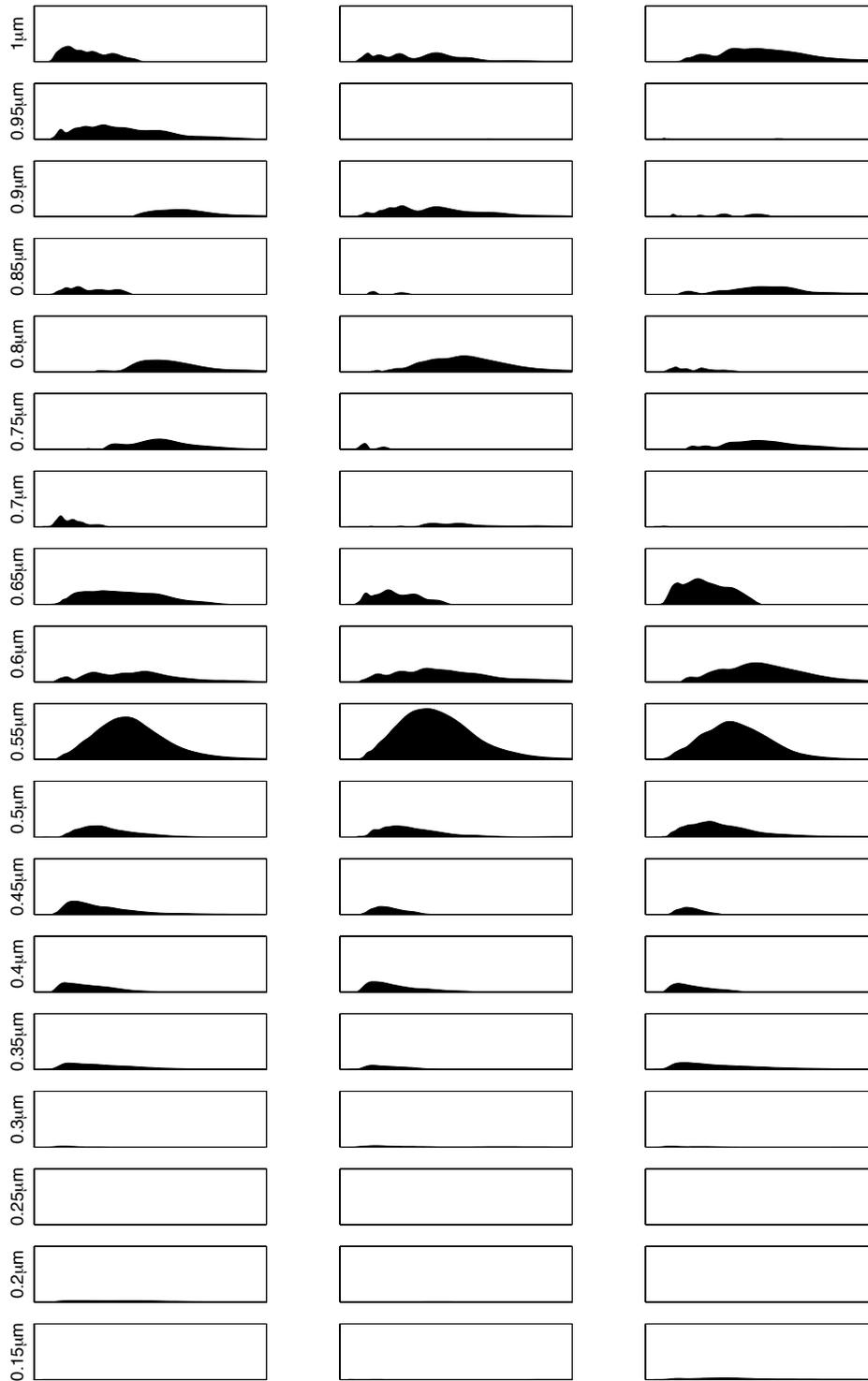
Figure 1.7 – Histogram of bacterial cell carbon quotas retained on each filter. Horizontal scale extends from 4 to 100 fg C cell^{-1}. Each column is a replicate sample, and all vertical scales are equal.
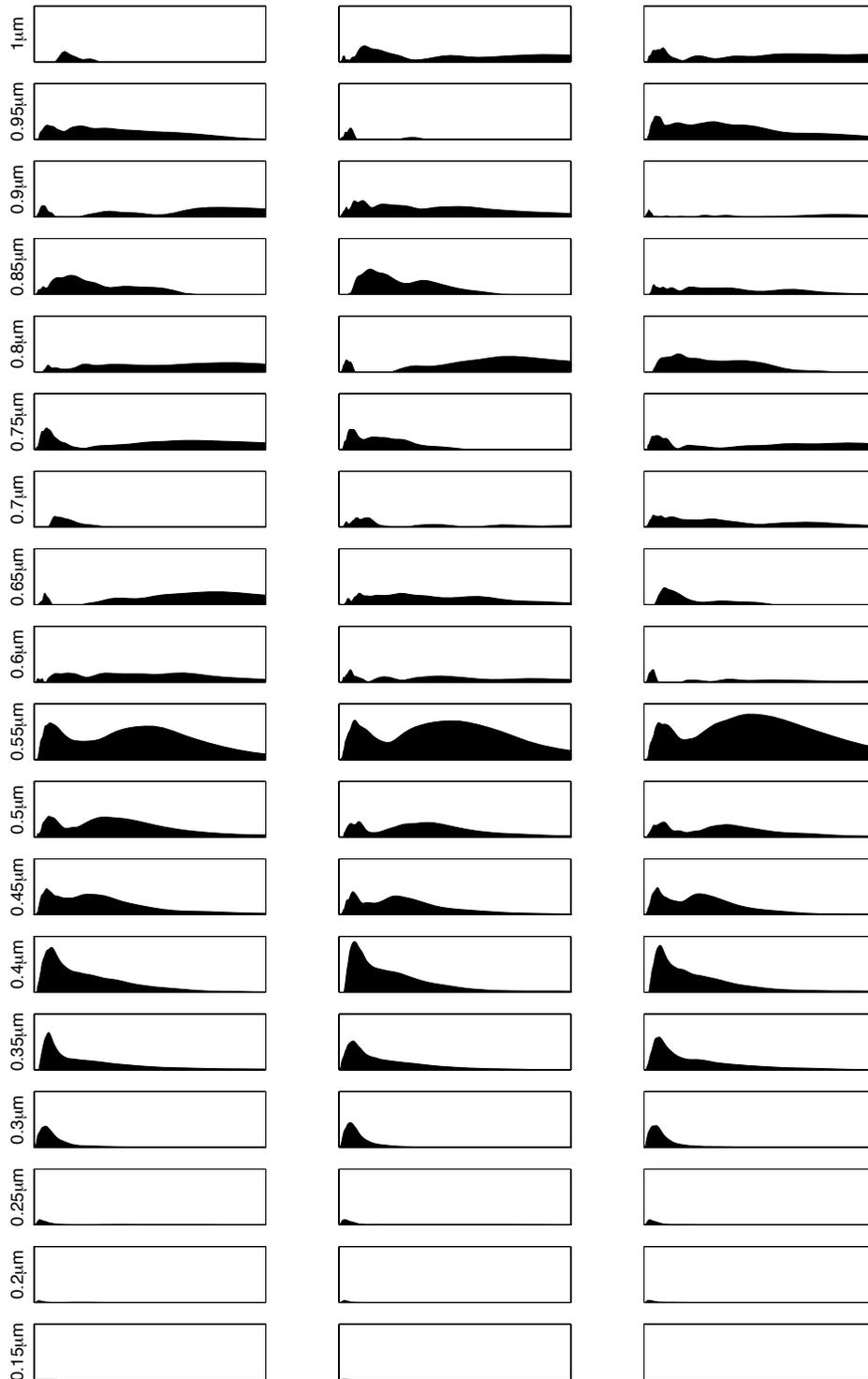
Figure 1.8 – Size spectrum of the number of *Prochlorococcus* cells (top panel),

Prochlorococcus carbon (middle panel), and ^{14}C-PP (bottom panel). Mean values are

shown with error bars representing one standard deviation. ^{14}C-PP data from the

1μm filter fraction are excluded.
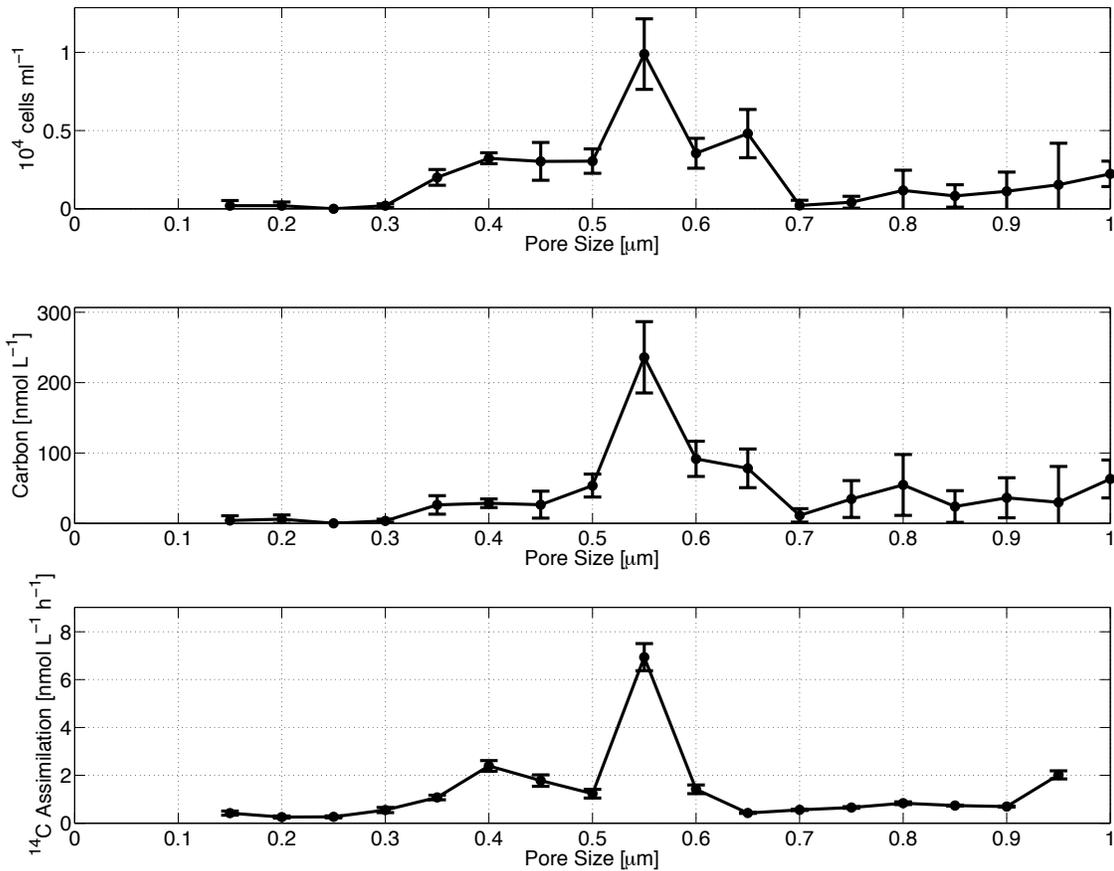
Figure 1.9 – Size spectrum of the number of *Prochlorococcus* and bacterial cells (top panel), *Prochlorococcus* and bacterial carbon (middle panel), and $^{32}PO_4^{3-}$ assimilation rates (bottom panel). Mean values are shown with error bars representing one standard deviation. $^{32}PO_4^{3-}$ assimilation data from the 1μm filter fraction are excluded.
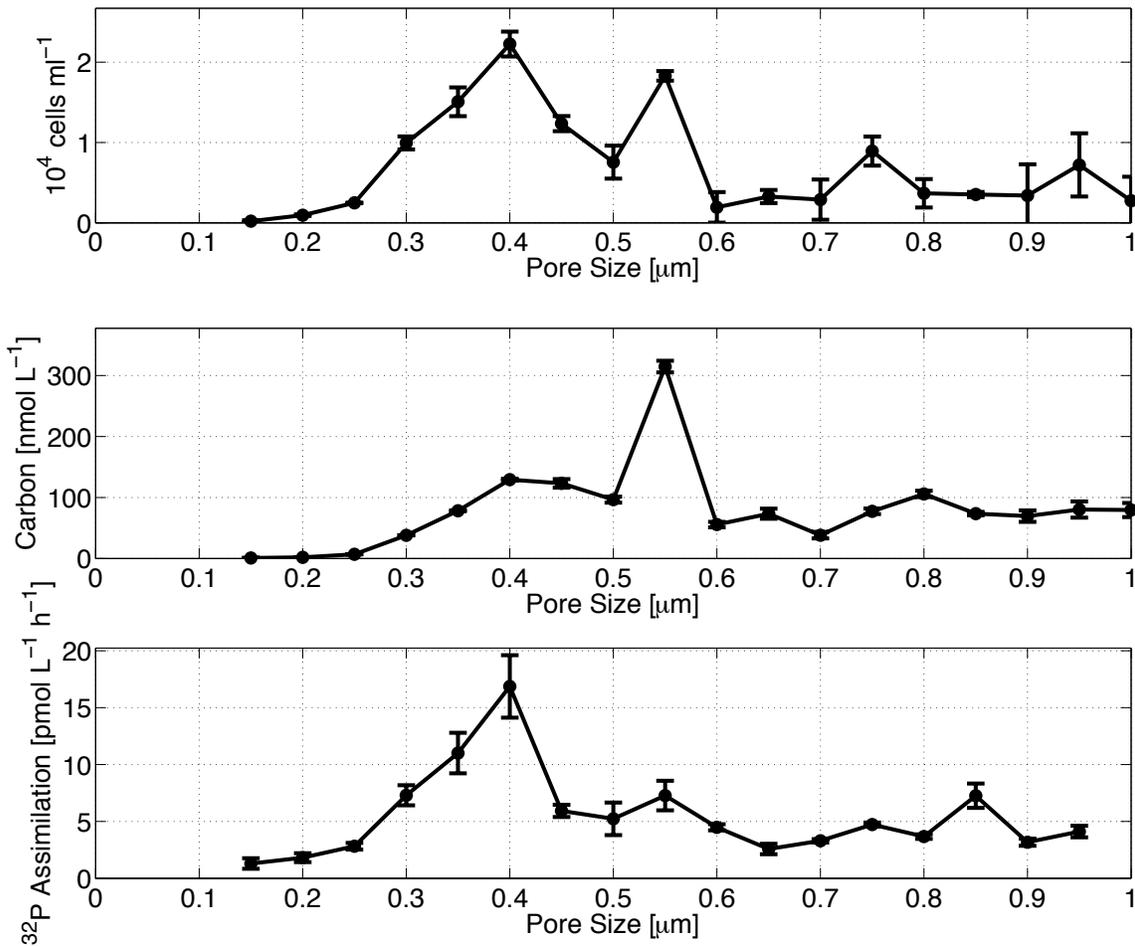
Figure 1.10 – *Prochlorococcus* carbon-specific ^{14}C-PP plotted as a function of equivalent spherical cell volume (based on pore size). Mean values are plotted and error bars represent one standard deviation. Non-linear least squares fit of the power law regression is also shown.
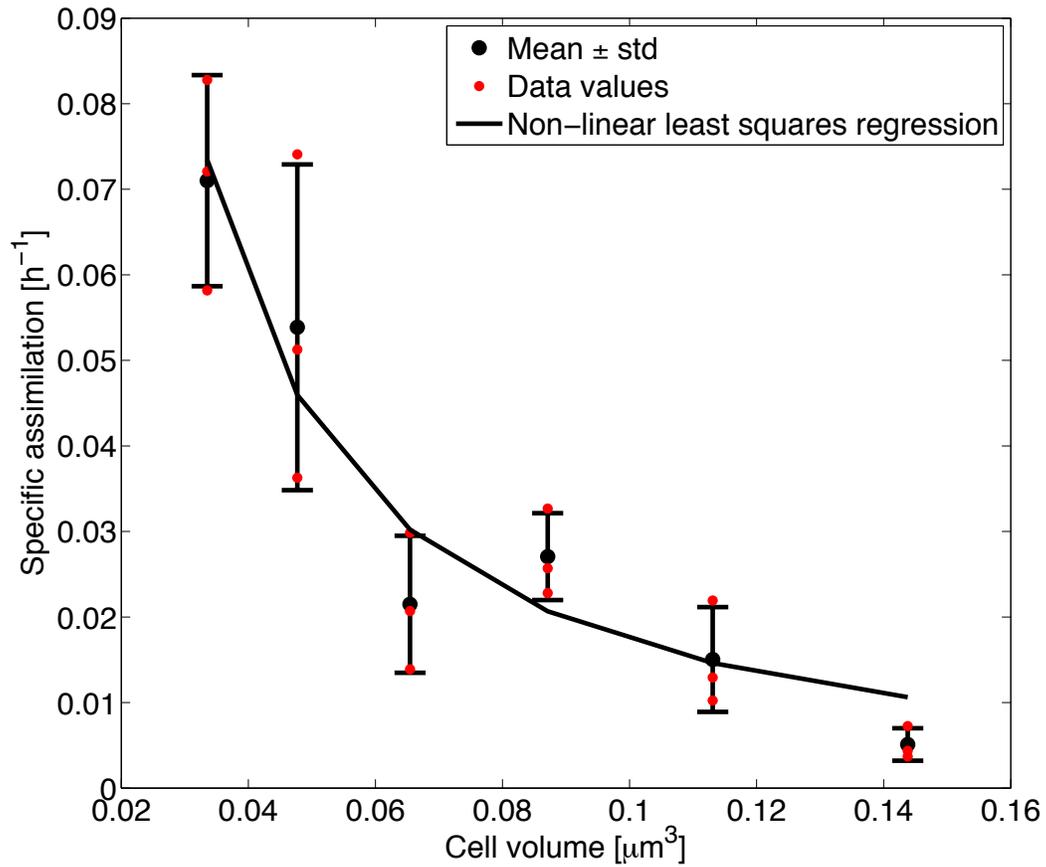
Figure 1.11 – Bacterial carbon-specific $^{32}PO_4^{3-}$ assimilation rates plotted as a function of equivalent spherical cell volume (based on pore size). Mean values are plotted and error bars represent one standard deviation. Non-linear least squares fit of the power law regression is also shown.
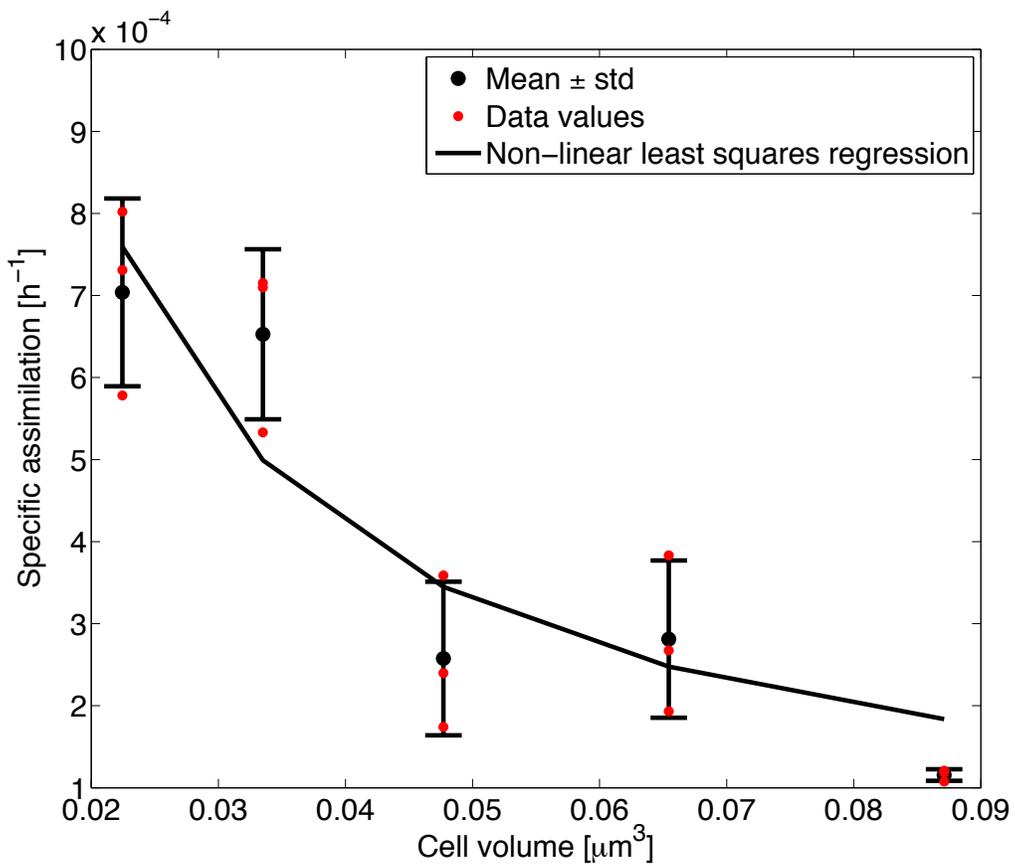
Figure 1.12 –*Prochlorococcus* MIT3901 carbon-specific ^{14}C-bicarbonate

assimilation rates plotted as a function of equivalent spherical cell volume (based on

pore size). Non-linear least squares fit of the power law regression is also shown.
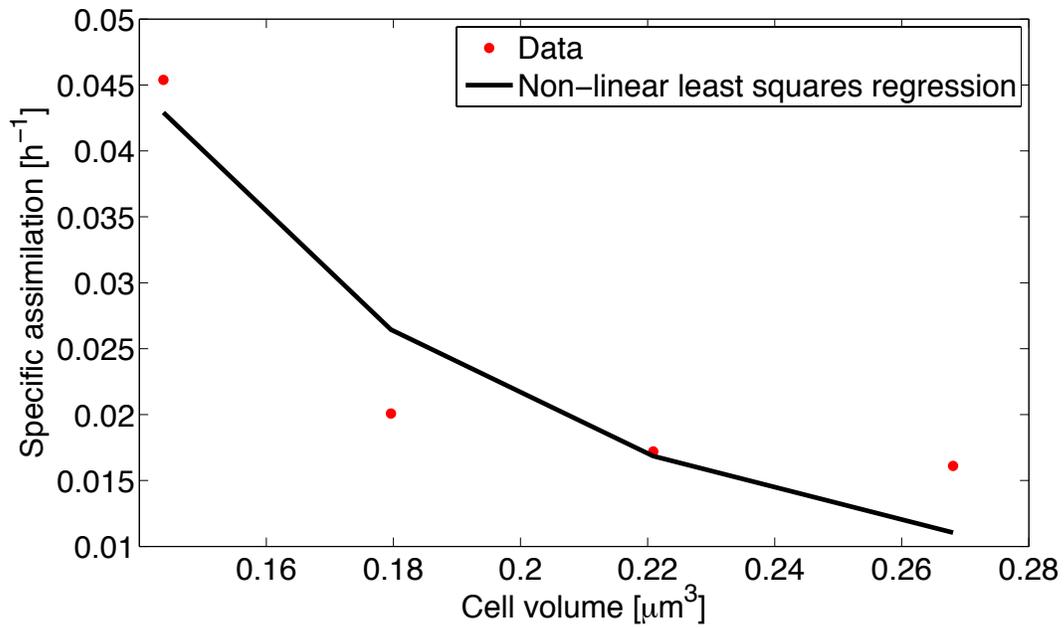
Substrate selection for heterotrophic bacterial growth in the sea

John R. Casey, Paul G. Falkowski, David M. Karl

Abstract

   Growth of heterotrophic microbes requires the extraction of energy, electrons, carbon, and nutrients from a complex and dynamic reservoir of potential substrates. We employed a matrix of selected organic substrates with varying characteristics, and experimentally followed the kinetics of assimilation and respiration to explore the basic principles that govern selection and preferential use based on carbon, nitrogen, and energy content. We further competed these substrates in a combinatorial fashion to evaluate preferential substrate utilization in natural microbial assemblages. Several substrates displayed biphasic kinetic responses and variable respiration : assimilation ratios. Amino acids had the shortest turnover times and were taken up preferentially at ambient concentrations. We also identified a linear relationship between substrate uptake rates and affinity, suggesting the microbial community optimizes the relative abundances of membrane transporters according to substrate demand. When competed against one another at saturating concentrations, substrate assimilation and respiration rates were enhanced or inhibited by up to two orders of magnitude, compared to competitor-free controls. Further, we describe an unexpected trend between the substrate energy density and turnover times, with more energetic, reduced carbon substrates turning over more slowly than more oxidized substrates.

2.1 Introduction

Microbial communities in the oligotrophic gyres are profoundly complex, both in terms of taxonomic diversity and metabolic potential, and interpretations of monoclonal laboratory cultures cannot easily be extended to the natural environment. Consequently, a qualitative, let alone quantitative, understanding of microbial community substrate preference is currently lacking. In the contest for resources, specialists and generalists inevitably compete with unique metabolic strategies. We propose that individual microbes within the community optimize the utilization of growth-limiting resources, but that the overall outcome is based on "rules" within the social order of the microbial community. In the chronically inorganic nutrient depleted surface layer of the North Pacific Subtropical Gyre (NPSG), it might logically follow that dissolved organic nitrogen and/or phosphorus sources should be preferentially utilized by natural heterotrophic microbial communities (e.g., . In this paper, we ask, "What are the basic principles that govern the selection of extracellular molecules for their growth?" We examine the hypothesis that heterotrophic communities in the surface layer of a typical oligotrophic ocean gyre have evolved to preferentially utilize dissolved organic matter (DOM) substrates based in accordance to the hierarchy of energy density (KJ mol C^{-1}), nitrogen mole fraction (mol N mol^{-1}), or carbon mole fraction (mol C mol^{-1}).

Substrates were chosen to represent a broad sampling of the bio-available low molecular weight (LMW) organic substrate spectrum; four 6-carbon compounds with different energy densities (in terms of standard enthalpy of combustion; ΔH_C'), three amino acids with different numbers of nitrogen atoms, and

three carbohydrates with equal energy densities but different numbers of carbon atoms were selected as representatives of each class (**Table 2.1**). Gene products associated with the metabolism of all selected substrates were identified in Station ALOHA metagenome datasets (NCBI Accession: PRJNA13694, PRJNA29033, PRJNA16339) and in genomic DNA sequences of diverse bacteria and eukaryotes isolated from the NPSG surface waters, including alcohol and aldehyde dehydrogenases that are required for the metabolism of 1-hexanol. Other selected compounds (L-leucine, L-lysine, L-arginine, glyceraldehyde, D-ribose, D-mannose, citric acid, and hexanoic acid) are associated with central carbon metabolism, amino acid biosynthesis, and fatty acid synthesis/degradation pathways and therefore are readily metabolized by those organisms capable of transporting them. Individual substrate utilization and kinetics have been extensively investigated for numerous amino acid (*e.g.*, Ayo *et al.*, 2001; Kirchman and Hodson, 1986), sugar (*e.g.*, Azam and Hodson, 1981; Nissen *et al.*, 1984) , nucleotide (*e.g.*, Karl and Bailiff, 1989), and organic acid (*e.g.*, Kieber *et al.*, 1989; Wright and Hobbie, 1966) compounds in the marine microbial populations, however a comparative approach with multiple substrates is quite rare (*c.f.*, Mopper and Kieber, 1991).

To evaluate substrate preference, the assimilation and respiration kinetics of each substrate were measured by radioisotope incubations of natural surface seawater at Station ALOHA. We further hypothesized that substrate uptake (as the sum of assimilation plus respiration) and subsequent metabolism (as the ratio of respiration to uptake) are influenced by the presence of competing substrates. To evaluate substrate competition, each radioisotope tracer was incubated in the

presence of a saturating concentration of each of the other substrates. Apart from a thorough competition of amino acids against different peptides (Kirchman *et al.*, 1984), we are unaware of a competition experiment involving multiple substrate classes in natural marine microbial assemblages.

Collectively, our field measurements indicate that natural heterotrophic microbial communities select substrates based on their nitrogen content over their carbon or energy content. Our results also lead to a somewhat surprising conclusion that higher energy density substrates (1-hexanol and hexanoic acid) turn over more slowly than lower energy density substrates (D-mannose, glyceraldehyde, and citric acid).

2.2 Methods

2.2.1 Sample Collection

The experiments were carried out onboard the *R/V Kilo Moana* at Station ALOHA (22°45'N, 158°W) in March 2014. Seawater samples from 25 m were collected in the dark from 12 l Niskin® bottles each night following a 2100 cast, approximately 3 h after sunset. Samples were collected over the course of 11 days alongside numerous biological, chemical, and hydrographic measurements, including measurements of community respiration based on diel cycles of O_2/Ar (Ferrón *et al.*, *in review*) and *in vivo* electron transport chain activity (Martínez-García *et al.*, 2009). Sample collections, preparations, and incubations were conducted in the dark or under low intensity red light and all glassware was cleaned following recommendations by Fitzwater *et al.* (1982) to minimize trace metals

contamination (pre-combusted, acid-washed, 0.22 μm filtered distilled deionized water-rinsed), which have been shown to affect amino acid uptake rate determinations in natural seawater samples (Ferguson and Sunda, 1984). Incubations were conducted at night to avoid extracellular production of newly synthesized photosynthate and to encourage constant uptake rates (*e.g.*, Carlucci *et al.*, 1984).

2.2.2 Reagents and Supplies

We evaluated 9 compounds: [1,5-^{14}C]citric acid, [1-^{14}C]hexanoic acid, [1-^{14}C]hexanol, [U-^{14}C]D-mannose, [U-^{14}C]L-leucine, [U-^{14}C]L-lysine, [U-^{14}C]L-arginine, [U-^{14}C]glyceraldehyde, and [1-^{14}C]D-ribose (MP Biomedicals, LLC; American Radiolabeled Chemicals, Inc.; **Table 2.1**). Uniformly labeled substrates were chosen, where available, to minimize ambiguity in interpretations of respiration and assimilation (ref), but this is inconsequential to uptake rates determinations. Unlabeled compounds and the reagents β-phenylethylamine (PEA herein), sulfuric acid, formaldehyde (37 wt %) were purchased from Sigma Aldrich Co. Dilute stocks of labeled substrates were made up to 92.5 MBq l^{-1} and specific activities were adjusted to 1.85 GBq mmol^{-1} by the addition of unlabeled carrier, with the exception of 1-hexanol which was prepared to 18.5 MBq l^{-1} at 0.37 GBq mmol^{-1}. 1-Hexanol solutions were always less than 577 μM, the solubility limit for 1-hexanol in water at 20°C.

2.2.3 Sample Preparation and Incubation Conditions

Samples (70 ml) for kinetic and competition incubations were transferred to either 70 or 125 ml glass serum bottles for assimilation and respiration measurements, respectively. Kinetic experiments were initiated by the addition of 9 concentrations of a labeled substrate (from 1 nM to 365 nM, spaced logarithmically), in triplicate (n = 486). Competition experiments were initiated by the addition of 32 nM of a labeled substrate in the presence of 2 μM of each unlabeled substrate competitor, in triplicate (n = 486). A single killed control (fixed at 4 °C for 1 h with 2% w/v formaldehyde) was used for each treatment (n = 162). Following addition of the radiolabeled substrate, assimilation bottles were crimped sealed without a headspace with Teflon stoppers using aluminum caps. The respiration bottles were fitted with "sleeve-style" rubber stoppers pierced with center well cups containing a dry piece of fluted cellulose filter paper (Whatman®) in the headspace. All bottles were incubated upright in the dark for 8 to 10 h, submerged in a bath of circulating surface seawater to approximately maintain the 25 m collection depth temperature (24.0 ± 0.1 °C).

2.2.4 Assimilation

Incubations were terminated by gentle vacuum filtration (<70 mBar) onto nominal 0.3 μm pore size glass fiber filters (GF75, Sterlitech). Bottles, caps, and filter funnels were thoroughly rinsed with 0.2 μm filtered seawater (three rinses of approximately 20 ml each). Filters were transferred to 20 ml glass scintillation vials followed by a 10 ml addition of scintillation cocktail (UltimaGold LLT, Perkin Elmer). [14]C activity was counted on a Perkin-Elmer 2910TR TriCarb liquid scintillation

counter using the spectral index of an external standard for quench correction. Substrate retained on the glass fiber filters or adsorbed to cells was estimated with killed-controls, which were prepared for each concentration of each substrate. Killed controls were incubated alongside live samples and processed identically.

2.2.5 Respiration

Incubations were terminated by soaking the filter paper in the cup with 150 µl PEA and then acidifying the sample by adding 4 ml of 4.5 N sulfuric acid through the gas-tight stopper. The acidified solution was allowed to passively distill for 48 h before removing the stoppers. This procedure changes the dissolved CO_2 solubility and leads to a flux of $^{14}CO_2$ first into the headspace, then into center well cups containing the PEA. Center well cups containing the filter paper were transferred to scintillation vials. 500 µl aliquots of the remaining volume of each sample were also transferred to scintillation vials (n = 486). ^{14}C activity was quantified identically to the filters (see *Assimilation*). Activity other than $^{14}CO_2$ that was retained by the cup was estimated by equilibrating each substrate in 0.2 µm filtered seawater and processing identically to live samples. $^{14}CO_2$ trapping efficiency was estimated by processing a known quantity of $H^{14}CO_3^-$ added to 0.2 µm filtered seawater.

2.2.6 Data analysis

Filter samples (i.e., "assimilation") were corrected by subtracting killed control activities and cup samples (respiration) were corrected for cell-free incubation activities. Transient isotope effects affect apparent uptake rates due to

mass-dependent kinetic isotope fractionation under isotopic disequilibrium; specific transporter fractionation factors differ widely between different enzyme-substrate systems, and the respiration to assimilation ratio influences the magnitude of this effect. However, substrate specific activities and added concentrations were sufficiently high so as to reduce this effect to well below the analytical standard error of the measurment (<5% standard error). Kinetics parameters (**Table 2**) were determined for assimilation (v_A), respiration (v_R), and uptake (v_U) of each substrate by non-linear least squares optimizations for each of four potential kinetics models:

Monophasic:	$v = \frac{V_{max}[S_A]}{K_M+[S_A]}$	(1)
Monophasic/Diffusive:	$v = \frac{V_{max}^{Fast}[S_A]}{K_M^{Fast}+[S_A]} + K_D[S_A]$	(2)
Biphasic:	$v = \frac{V_{max}^{Fast}[S_A]}{K_M^{Fast}+[S_A]} + \frac{V_{max}^{Slow}[S_A]}{K_M^{Slow}+[S_A]}$	(3)
Multiphasic:	$v = \begin{cases} V_{max}^{Fast}\frac{S_A}{K_M^{Fast}+S_A} & \text{if } S_A \leq b \\ V_{max}^{Fast}\frac{b}{K_M^{Fast}+b} + V_{max}^{slow}\frac{S_A-b}{K_M^{Slow}+S_A-b} & \text{if } S_A > b, \end{cases}$	(4)

where K_m [nM] and V_{max} [nM h^{-1}] are the Michaelis-Menten-like half-saturation constant and maximal rate, K_D is a diffusive parameter, and v is the kinetic velocity. To account for systematic analytical variance, which varied as a function of S_A added, we weighted the cost function using standard errors. A sensitivity analysis of the initialization parameters resulted in derived kinetics parameters that were stable to within <1% over a ten-fold range of expected values (data not shown). Models were selected using the minimal Akaike Information Criterion (AIC) using the log-likelihood of each NLS fit; the AIC model selection procedure rewards the goodness-of-fit of each model while penalizing overly complex models. Standard errors for K_m and V_{max} values were propagated by the root sum of squares deviation. Specific

affinities (not biomass normalized) were calculated as the initial slope of the fast

(a_F^o) and slow (a_S^o) subsets. Competition experiments were processed similarly,

except that rates were treated as a fold-change relative to an unamended control.

Community respiration estimates corresponding to the same casts based on

ΔO_2:Ar and electron transport activity measurements varied by 28% and 19% over

the course of the sampling period, respectively (Ferrón *et al.*, *in review*). This day-to-

day natural variability was not taken into account in our error analysis and

potentially influenced V_{max} estimates since our experiments were conducted over an

11 day period, however, both community respiration estimates could explain <1%

of the variance in V_{max} (p = 0.936).

Standard enthalpies of combustion were calculated for each substrate

according to a simplification of the Kharash equation (Kharasch, 1929):

$$\Delta H_C' = 200a_1 + 280a_2 + 220a_3 + 105a_4 + 60a_5 + 40a_6 - 20a_7$$

Where the coefficients are the average bond energies (KJ mol^{-1}) corresponding to

the number of π bonds in aromatic compounds (a_1), non-aromatic double bonds

(a_2), σ and C-H single bonds (a_3), C-N and N-H bonds (a_4), carbonyl groups (a_5),

hydroxy (a_6), and carboxyl groups (a_7).

2.3 Results and Discussion

2.3.1 Method Evaluation

We evaluated the ^{14}C-respirometry method for determining compound-

specific assimilation and respiration rates in natural marine seawater incubations.

We consistently found quantitative recovery 99.85 ± 1.98 % (calculated as the sum

of filter, cup, and remaining activity as a percentage of activity added) and low sample-sample variability (8.3 ± 1.9 %; **Figure 2.1**). Killed control activities were consistently <0.1% of activity added and consistently <15 % of assimilation with the exception of [1-^{14}C]D-ribose (45.5 ± 17.9 % of assimilation), indicating low cell surface and glass fiber filter adsorption over the entire S_A range. The trapping efficiency of PEA soaked filter paper after acidifying a known quantity of H^{14}CO$_3^-$ was 98.99 ± 4.28 %. When the same procedure was followed for 32 nM additions of each labeled substrate in 0.2 μm filtered seawater, the absorption was 0.08 ± 0.04 % of total radioactivity added, with the exception of [1-^{14}C]hexanol which was 10.4 %.

2.3.2 Substrate preference and kinetics

Assimilation, respiration, and uptake kinetics were biphasic or multiphasic within the range 1 nM to 364.5 nM of S_A for 7 of the 9 substrates tested with "breakpoints" (the transition between two transporter systems operating at different concentration ranges) between 20 nM and 90 nM of the S$_A$ (**Figure 2.2**, **Table 3**). No differences in breakpoint concentrations were found between assimilation, respiration, or uptake kinetics curves for a given substrate. Multiphasic kinetics in natural marine bacterial populations have been reported previously for the metabolism of glucose and mixed amino acids (Azam and Hodson, 1981; Nissen *et al.,* 1984) and for ectoenzyme kinetics (Unanue *et al.,* 1999). Artifacts associated with S_A and S_n concentration dependencies of K_m and multiphasic kinetics have been proposed to result from a diffusive model with a sub-unity slope resulting from different membrane permeabilities, intracellular substrate concentrations, and

kinetic fractionation of the radiolabeled substrate uptake between species (Logan and Fleury, 1993). However it is highly unlikely that marine bacteria rely on concentration gradients alone for substrate uptake, especially since intracellular metabolite concentrations are frequently in the μM to mM concentration range (e.g., Bennett *et al.*, 2009); rather, it is much more likely a combination of multiple transporters (Button, 1993). Although the multiphasic model has been theoretically and experimentally described in a multitude of biological systems (Nissen and Nissen, 1983), we cannot discount the possibility that breakpoints were an artifact of *de novo* transporter synthesis during the course of the incubation. The remaining two substrates were not well described by the multiphasic or biphasic models, rather a monophasic model was selected for L-arginine kinetics and an additional diffusive term best fit the kinetics of L-lysine, according to our AIC model selection procedure.

v_R / v_U ratios, which are upper constraints on bacterial growth efficiencies (BGE) due to isotopic disequilibrium of intracellular metabolite pools during short incubations (King and Berman, 1984), were not constant over the concentration range of substrates tested (**Figure 2.3**). Coefficients of variation (CV) of substrate v_R / v_U ratios ranged from 14 to 45 % from glyceraldehyde to L-arginine, respectively. S_A dependencies of BGE are ignored in most bacterial production assays, including those based on [^{3}H]L-leucine assimilation (CV = 28 %), and will introduce ambiguity to their interpretations since ambient substrate concentrations are rarely measured prior to incubations and substrate is usually added at saturating concentrations (> 20 nM; e.g. Church et al., 2004).

Because ambient substrate concentrations (S_n) were not measured, "tracer" substrate concentration additions, ideally <10% of S_n, were not necessarily achieved. We therefore approximate ambient substrate uptake rates as the initial velocity (v_o) at 1 nM S_A. With this proviso, amino acids were taken up at rates higher than other substrates tested. Another metric for substrate affinity, V_{max}/K_m (h^{-1}), was also found to be highest for the amino acids compared with other substrates. We found a linear relationship (R^2 = 0.70, slope = 0.85 ± 0.1, p = 0.0001; **Figure 2.4**) between the ambient substrate uptake rate and the substrate affinity, as expected for Michaelis-Menten kinetics at $S_A + S_n \ll K_m$. Agreement between ambient uptake rates and substrate affinities supports the hypothesis that microbial communities collectively optimize the scavenging of resources according to substrate availability and demand. Since K_m is an intrinsic property of transporter-substrate affinity, both the cell surface density of transporters and the distribution of transporters among different microbial community taxa are optimized for the exploitation of specific substrates. For example, amino acid transporters are likely common to a majority of microbial taxa, and may represent a larger fraction of outer membrane transporter proteins than e.g., fatty acid transporters, as was the case for coastal bacterioplankton (Poretsky *et al.*, 2010).

Over the concentration range tested, we found that 1-hexanol was assimilated at rates comparable to other C_6 substrates. Marine microbes are known to use methanol as both a carbon and energy source, and several alcohol dehydrogenases are common to methylotrophic bacteria including the numerically dominant *Pelagibacter ubique* SAR11 (Sun *et al.,* 2011). 1-Hexanol is

dehydrogenated to hexanaldehyde and further to hexanoic acid and subsequently directed toward fatty acid synthesis or degradation pathways to regenerate acetyl-CoA. Perhaps since 1-hexanol is uniquely connected to hexanoic acid, ambient substrate uptake rates and substrate affinities were nearly identical for 1-hexanol and hexanoic acid, despite almost certainly depending on separate transporters due to differences in charge distribution. We estimated an upper limit on the concentration of 1-hexanol to be 7.0 ± 1.9 nM (**Table 2.4)**, however there are no direct determinations for the ambient concentration of any primary alcohols other than methanol in seawater to compare. K_m and V_{max} values (**Table 2.3**) were comparable to methanol (K_m = 9.3 to 86 nM; V_{max} = 1.0 to 1.2 nM h^{-1} in Dixon *et al.,* 2011). 1-Hexanol respiration could not be determined due to a high sample blank, perhaps due to volatility and subsequent scavenging by PEA, though we are unaware of this reaction at room temperature. We suggest that future respirometry determinations of volatile and semi-volatile substrates use another trapping solvent (e.g., hyamine, a quaternary ammonium hydroxide).

D-Ribose was taken up at unexpectedly low rates and respiration was near the limit of quantitation (3 times the standard deviation of v_R for each S_A concentration) resulting in poor confidence in kinetic constant estimates (3 out of 9 S_A concentrations with $p > 0.05$). This study is the first to report D-ribose uptake rates in natural seawater and the few direct concentration determinations available for the North Pacific suggest D-ribose is typically depleted (<5 mol %) in monomeric (MW < 200 Da), oligomeric (MW 200 - 4000 Da), and polymeric (MW > 4000 Da) fractions of total hydrolysable dissolved carbohydrates (Mopper *et al.,* 1980;

Sakugawa and Handa, 1983). Rather, D-ribose is likely present mostly as the sugar moiety of RNA and free or polymeric nucleotides, which are abundant both in cells and dissolved in seawater (Karl and Bailiff, 1989; Bjorkman and Karl, 2005). Pentosyltransferases, such as purine and pyrimidine nucleoside transferases are probably not active exoenzymes, otherwise we would expect an accumulation of dissolved D-ribose since nucleotide turnover times are relatively short (1-2 d; Björkman and Karl, 2005). Therefore, RNA and nucleotides must be transported prior to dissimilation (e.g., xanthine degradation pathway, purine catabolic pathways, pyrimidine catabolic pathways) or polymerization.

In contrast to D-ribose, the remaining carbon substrates (citric acid, glyceraldehyde, and D-mannose) were rapidly utilized. Of these, only D-mannose uptake has been previously reported, albeit in a mesotrophic lake, with roughly comparable turnover times (20 to 170 h; Bunte and Simon, 1999). Citric acid v_R / v_U ratios (0.65 ± 0.10) were significantly higher ($p < e^{-6}$) than the other carbohydrates glyceraldehyde (0.35 ± 0.05) and D-mannose (0.23 ± 0.05). Glyceraldehyde and D-mannose v_R / v_U ratios were comparable to D-glucose in other marine environments (Williams, 1970; Williams and Yentsch, 1976). Although we expected high catabolism of citric acid due to its role as a tricarboxylic acid (TCA) cycle intermediate, high anabolic assimilation of carbohydrates suggests an unexpected prevalence of carbon storage by gluconeogenesis in marine bacteria inhabiting the NPSG. The disparity in the v_R / v_U ratio for glyceraldehyde and D-mannose further supports this hypothesis. Intermediates at the top of glycolysis (β-D-fructose-6-phosphate derived from D-mannose by hexokinase, M6P isomerase) are expected to

show higher anabolism relative to bottom of glycolysis intermediates (2-phospho-D-glycerate or 3-phospho-D-glycerate derived from glyceraldehyde by glyceraldehyde dehydrogenase or triose kinase, respectively) under conditions where the oxidative pentose phosphate pathway flux is low.

Ambient uptake rates were similar ($p > 0.05$) between amino acids, however differences were observed in the substrate affinity, turnover, kinetic constants, and efficiencies of nitrogen assimilation. For example, the efficiency of nitrogen assimilation, expressed as the nitrogen assimilated as a molar fraction of the nitrogen added, increased in the order of nitrogen atoms per molecule (L-leucine (1) < L-lysine (2) < L-arginine (4)). Turnover times and K_m constants increased in the same order, while ambient substrate uptake rates were similar, suggesting transporter abundance (proportional to V_{max}) was predominantly responsible for the high uptake rates.

Among the carbohydrates, the number of carbon atoms was not a determinant in v_U, v_A, v_R, K_m, V_{max}, a_F^o, a_S^o, or turnover time. However, the turnover time was exponentially dependent on energy density (KJ mol C^{-1}; **Figure 2.5**). It is worth noting that the carbon substrates exhibiting the most rapid turnover (citric acid, glyceraldehyde, and D-mannose) have potential energies, average degrees of reduction, and respiration quotients (RQ) closely matching those of microbial cell biomass (based on an average of *Escherichia coli* and *Saccharomyces cerevisiae*). Few compound-specific turnover time determinations are available for oligotrophic marine environments, still fewer with substrates representing a broad energy density range, however the turnover dependence on substrate energy density

appears to be consistent. For example, in an assessment of substrate-specific turnover times of 10 substrates in the Black Sea water column, L-ornithine was both the highest energy density substrate surveyed (626 KJ mol C^{-1}) and had the longest turnover time (172 d; Mopper and Kieber, 1991). Long turnover times have been reported for other high energy density substrates; acetone (607 KJ mol C^{-1}) turnover was estimated at 41-55 days in the oligotrophic gyres (Dixon *et al.*, 2013), methanol (700 KJ mol C^{-1}) turnover times were 7-33 days in the northeast Atlantic (Dixon *et al*,. 2011), comparable to 1-hexanol turnover times (41 days). Methane (880 KJ mol C^{-1}), the most reduced of all organic substrates, has been estimated to turn over between a few years and hundreds of years in the western North Pacific (Watanabe *et al.*, 1995) and Sargasso Sea (Jones *et al.*, 1991). The number of substrates surveyed in this and other studies are too few to formalize such a pattern, however if this relationship holds for other labile LMW DOM substrates, it may be applied to the existing theoretical framework of ecological thermodynamics since both bacterial growth yield (mol C mol C^{-1} utilized) and bacterial C specific growth rate (mol C mol C^{-1} utilized d^{-1}) tend to increase as a linear and squared function, respectively, of the degree of reduction of substrate (Vallino *et al.,* 1996). We observed no relationship between degree of reduction or energy density and v_R / v_U so the catabolic : anabolic ratio of individual substrates did not align with theoretical predictions of growth yield in our study site. Conversely, for a particular substrate as the sole carbon source for anoxygenic microbes, the thermodynamic efficiency for growth, defined as the product of the Gibbs energy of assimilation : chemical potential energy ratio and the growth rate : uptake rate ratio, is

theoretically lower for reduced substrates (degree of reduction > 4) than oxidized

substrates (Westerhoff, 1993). Correspondingly, the theoretical growth yield and

the ratio of free enthalpy of catabolism to anabolism increases with degree of

reduction of the substrate (Westerhoff et al., 1983). The degree of reduction of

intermediate metabolites in the central carbon pathways of metabolism are

reminiscent of this, since oxidized intermediates are associated with high flux

central carbon metabolic pathways (TCA cycle, glycolysis, and the oxidative pentose

phosphate pathway), while reduced metabolites (long-chain fatty acids and

alcohols) are associated with linear biosynthetic and degradation pathways (e.g.,

fatty acid biosynthesis, β-oxidation) which turn over less rapidly. It is therefore

plausible that the degree of reduction produces an imbalance in the production and

consumption of LMW DOM substrates, resulting in the observed trend with

turnover time. Even for specialists, a single carbon source is unlikely to support

growth in the marine environment, so these generalizations are complicated by

multiple and varying carbon sources.

2.3.3 Substrate competition

DOM composition is dynamic in the near-surface environment on both

spatial and temporal scales; diel oscillations in photochemical production of organic

matter, especially in the surface microlayer (e.g., Carlucci *et al.,* 1984; Kieber *et al.*,

1989; Zhou and Mopper, 1997), and diel oscillations in phytoplankton extracellular

DOM production interact with diurnal mixing cycles to alter the DOM age and

composition throughout the upper water column. The intention of the competition

experiment was to broadly evaluate changes in substrate uptake rates in the presence of saturating concentrations of substrates that could aid in our interpretation of pairwise comparisons of kinetics experiments, in a combinatorial manner. Substrate utilization is constrained by one of two proteins: the membrane transporter (or porin), and a rate-limiting metabolic reaction. For marine bacterial genomes and metagenomes, adequate functional characterizations of membrane transporters are unavailable; transporters are often annotated by homology to no better than the family level. With neither the adequate identification nor the abundance of transporters known, the distinction between transporter rate-limitation and metabolic rate-limitation is unclear. Considering that a typical oligotrophic bacterium (0.4 μm diameter) may have space for only 10^3 to 10^4 individual membrane transporter proteins (based on approximate protein sizes, spacing, and surface area), and given the diversity of potential substrates and minimal genome sizes, it is likely that many transporters are of broad specificity and therefore subject to competitive inhibition. Ultimately, the competitor may act in a great many ways to affect a change in apparent uptake rates of the labeled substrates.

We observed complex interactions between many labeled substrates and their competitors, affecting significant changes in uptake rates and the fraction of uptake respired, spanning more than a factor of 20 and 7, respectively (**Figure 2.6**). In general, competing substrates with similar chemical properties (e.g., L-lysine and L-arginine) and/or mutual metabolic pathways (e.g., hexanoic acid and 1-hexanol) were inhibitors of one another, while substrates of varied structure from disparate

branches of the metabolic network affected assimilation and respiration changes in more complex ways. We highlight several of these interactions below, however, we stress that in all cases, inhibitive or enhanced rates and ratios are exaggerated relative to natural conditions due to the high concentration (2 μM) of competitor added with each labeled substrate.

As might be expected, amino acids were potent inhibitors of other amino acids, with 50.0 ± 0.2 %, 72.7 ± 1.2 %, and 88.1 ± 6.3 % inhibition of uptake for L-leucine, L-lysine, and L-arginine, respectively, regardless of which saturating amino acid competitor was used. This result, together with similarity in the breakpoint of high and low affinity transport systems between these three amino acids, suggests that a high-affinity but broad substrate-specificity amino acid transporter may be prevalent and rate-limiting. Alternatively, metabolic regulation of protein synthesis may be rate-limiting; Kirchman and Hodson (1984) used oligopeptides as surrogates to alter the intracellular amino acid pool size, based on the assumption that free and polymeric amino acids do not share transport systems. They attributed the oligopeptide inhibition of extracellular amino acid uptake to 'buildup' of intracellular amino acid pools and suggested the feedback indicates direct coupling between amino acid uptake and protein synthesis. In our competition experiment, amino acid competitors similarly affected the fraction of substrate uptake that was subsequently respired; L-leucine, L-lysine, and L-arginine v_R / v_U ratios decreased by 53.4 ± 1.5 %, 53.8 ± 14.1 %, and 58.7 ± 11.9 %, respectively. Presumably, decreased v_R / v_U ratios are an indication of connectivity of amino acid pools; degradation of the competitor amino acid to ammonia and subsequent *de novo* protein synthesis

would result in enhanced assimilation of the labeled amino acid. L-leucine also inhibited the uptake of all other substrates, with the curious exception of an increase in both hexanoic acid and 1-hexanol respiration, perhaps suggesting that the flux through fatty acid degradation pathways is increased when nitrogen growth limitation is relieved. L-leucine, the only acidic aliphatic amino acid investigated, responded to carbon substrate competitor additions differently from the basic non-polar amino acids L-lysine and L-arginine. For example, L-leucine uptake was enhanced in the presence of all three carbohydrates and hexanoic acid.

2.3.4 Conclusions

Pairwise comparisons of uptake kinetics and competition experiments in natural seawater incubations are challenging to interpret, however, by comparing substrates spanning a wide range of thermodynamic and chemical properties, we identified three emergent patterns of substrate selection in surface microbial communities. First, amino acids were preferentially utilized at ambient concentrations compared with all other substrates, and the number of nitrogen atoms was important to amino acid uptake kinetics. Amino acids inhibited the uptake of other other amino acids, suggesting a broad specificity transporter system. Second, a linear relationship between v_o and a_F^o was identified, suggesting that interactions between transporter binding efficiency and translation among complex communities have evolved to optimize the consumption of specific substrates. Third, among the non-amino acid substrates, oxidized substrates turn over faster than reduced substrates, consistent with theoretical assimilation

efficiencies. The turnover times of organic carbon substrates should therefore proceed in the order of oxidation state (alkanes > alcohols, ethers, alkenes > ketones, aldehydes, epoxides > carboxylic acids, hydroxy acids). Our results reveal that microbes are indeed picky eaters and that the surface microbial community prefers a "high protein, low calorie" diet.

2.4 Acknowledgements

2.5 References

Ayo, B., Unanue, M., Azua, I., Gorsky, G., Turley, C., Iriberri, J., 2001. Kinetics of glucose and amino acid uptake by attached and free-living marine bacteria in oligotrophic waters. Marine Biology 138: 1071–1076.

Azam, F., Hodson, R.E., 1981. Multiphasic kinetics for D-glucose uptake by assemblages of natural marine bacteria. Marine Ecology Progress Series 6: 13–222.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D.,

2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nature Methods 5: 593–599.

Björkman, K.M., Karl, D., 2005. Presence of dissolved nucleotides in the North Pacific Subtropical Gyre and their role in cycling of dissolved organic phosphorus. Aquatic Microbial Ecology 39: 193–203.

Bunte, C., Simon, M., 1999. Bacterioplankton turnover of dissolved free monosaccharides in a mesotrophic lake. Limnology and Oceanography 44: 1862-1870.

Button, D.K., 1993. Nutrient-limited microbial growth kinetics: overview and recent advances. Antonie van Leeuwenhoek 63: 225–235.

Carlucci, A.F., Craven, D.B., Henrichs, S.M., 1984. Diel production and microheterotrophic utilization of dissolved free amino acids in waters off Southern California. Applied and Environmental Microbiology 48: 165-170.

Church, M.J., Ducklow, H.W., Karl, D., 2004. Light dependence of [3H]Leucine incorporation in the oligotrophic North Pacific Ocean. Applied and Environmental Microbiology 70: 4079-4087.

Dixon, J.L., Beale, R., Nightingale, P.D., 2011. Microbial methanol uptake in northeast

Atlantic waters. ISME Journal 5: 704–716.

Dixon, J.L., Beale, R., Nightingale, P.D., 2013. Production of methanol, acetaldehyde, and acetone in the Atlantic Ocean. Geophysical Research Letters 40: 4700–4705.

Ferguson, R.L., Sunda, W.G., 1984. Utilization of amino acids by planktonic marine bacteria: Importance of clean technique and low substrate additions. Limnology and Oceanography 29: 258–274.

Ferrón, S., Wilson, S. T., Martínez-García, S., Quay, P. D., and Karl, D. M. *in review.* Metabolic balance in the mixed layer of the oligotrophic North Pacific Ocean from diel changes in O_2/Ar saturation ratios. Geophysical Research Letters.

Jones, R.D., 1991. Carbon monoxide and methane distribution and consumption in the photic zone of the Sargasso Sea. Deep Sea Research Part A 38: 625-635.

Karl, D., Bailiff, M., 1989. The measurement and distribution of dissolved nucleic acids in aquatic environments. Limnology and Oceanography 34: 543–558.

Kharasch, M.S., 1929. Heats of combustion of organic compounds. Bureau of Standards Journal of Research 2: 359–430.

King, G.M., Berman, T., 1984. Potential effects of isotopic dilution on apparent

respiration in ^{14}C heterotrophy experiments. Marine Ecology Progress Series 19: 175–180.

Kirchman, D., Hodson, R., 1984. Inhibition by peptides of amino acid uptake by bacterial populations in natural waters: implications for the regulation of amino acid transport and incorporation. Applied and Environmental Microbiology 47: 624–631.

Kirchman, D.L., Hodson, R.E., 1986. Metabolic regulation of amino acid uptake in marine waters. Limnology and Oceanography 31: 339–350.

Martínez-García, S., Fernández, E., Aranguren-Gassis, M., Teira, E., 2009. In vivo electron transport system activity: a method to estimate respiration in natural marine microbial planktonic communities. Limnology and Oceanography Methods 7: 459–469.

Mopper, K., Dawson, R., Liebezeit, G., Ittekkot, V., 1980. The monosaccharide spectra of natural waters. Marine Chemistry 10: 55–66.

Mopper, K., Kieber, D.J., 1991. Distribution and biological turnover of dissolved organic compounds in the water column of the Black Sea. Deep Sea Research Part A. Oceanographic Research Papers 38: 1021–1047.

Nissen, H., Nissen, P., Faroor, A., 1984. Multiphasic uptake of D-glucose by an oligotrophic marine bacterium. Marine Ecology Progress Series 16: 155–160.

Poretsky, R.S., Sun, S., Mou, X., Moran, M.A., 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environmental Microbiology 12: 616–627.

Sakugawa, H., Handa, N., 1983. Chemical studies of dissolved carbohydrates in seawater. Journal of the Oceanographical Society of Japan 39: 279–288.

Sun, J., Steindler, L., Thrash, J.C., Halsey, K.H., Smith, D.P., Carter, A.E., Landry, Z.C., Giovannoni, S.J., 2011. One carbon metabolism in SAR11 pelagic marine bacteria. PLoS ONE 6: e23973

Unanue, M., Ayo, B., Agis, M., Slezak, D., Herndl, G.J., Iriberri, J., 1999. Ectoenzymatic activity and uptake of monomers in marine bacterioplankton described by a biphasic kinetic model. Microbial Ecology 37: 36–48.

Vallino, J., Hopkinson, C.S., Hobbie, J.E., 1996. Modeling bacterial utilization of dissolved organic matter: optimization replaces Monod growth kinetics. Limnology and Oceanography 41: 1591-1609.

Watanabe, S., Higashitani, N., Tsurushima, N., Tsunogai, S., 1995. Methane in the

western North Pacific. Journal of Oceanography 51: 39–60.

Williams, P.J. leB., 1970. Heterotrophic utilization of dissolved organic compounds in the sea I. Size distribution of population and relationship between respiration and incorporation of growth substrates. Journal of the Marine Biological Association of the United Kingdom 50: 859–870.

Williams, P.J. leB., Yentsch, C.S., 1976. An examination of photosynthetic production, excretion of photosynthetic products, and heterotrophic utilization of dissolved organic compounds with reference to results from a coastal subtropical sea. Marine Biology 35: 31–40.

Wright, R.R., Hobbie, J.E., 1966. Use of glucose and acetate by bacteria and algae in aquatic ecosystems. Ecology 47: 447-464.

Zhou, X., Mopper, K., 1997. Photochemical production of low-molecular-weight carbonyl compounds in seawater and surface microlayer and their air-sea exchange. Marine Chemistry 56: 201–213.

Table 2.1 - Comparison of selected substrates. Carbon enthalpies of combustion ($\Delta H_C'$) were calculated according to the Kharasch Equation (see Methods). Respiration quotients (RQ) were derived from balanced respiration equations.

Compound	$\Delta H_C'$ [KJ mol C^{-1}]	RQ [mol CO$_2$: mol O$_2$]	Nitrogen [N atoms]	Carbon [C atoms]
citric acid	377	1	0	6
hexanoic acid	600	1.33	0	6
1-hexanol	667	2	0	6
glyceraldehyde	486	1	0	3
D-ribose	473	1	0	5
D-mannose	483	1	0	6
L-leucine	615	0.8	1	6
L-lysine	633	0.86	2	6
L-arginine	652	1.1	4	6

Table 2.2 - Parameters used in this study.

Parameter	Description	Units
$a_F{}^o$	substrate affinity of fast transport system (K_m/V_{max})	h^{-1}
$a_S{}^o$	substrate affinity of slow transport system (K_m/V_{max})	h^{-1}
K_m	Michaelis-Menten half-saturation constant	nM
K_D	rate of diffusion	h^{-1}
S_A	substrate added concentration	nM
S_n	natural (ambient) substrate concentration	nM
v	transport velocity	nM h^{-1}
v_o	initial velocity	nM h^{-1}
v_R	respiration rate	nM h^{-1}
$\underline{v}_A$	assimilation rate	nM h^{-1}
v_U	uptake rate ($v_R + v_A$)	nM h^{-1}
V_{max}	maximum velocity	nM h^{-1}

Table 2.3 - Summary of uptake kinetics for each labeled substrate. Standard deviation reported in parentheses. *ND*- not determined.

Substrate	Model	Breakpoint [nM]	V_{max} (Fast) [nM h^{-1}]	K_m (Fast) [nM]	V_{max} (Slow) [nM h^{-1}]	K_m (Slow) [nM]	R^2
citric acid	Multiphasic	52.3	0.094 (0.003)	6.2 (0.6)	0.195 (0.003)	54.2 (3.2)	1.00
hexanoic acid	Multiphasic	45.5	0.050 (0.004)	16.6 (2.4)	0.712 (0.150)	912.1 (109.1)	1.00
1-hexanol	Multiphasic	89.1	0.037 (0.003)	6.9 (2.8)	0.218 (0.004)	243.8 (19.7)	1.00
glyceraldehyde	Multiphasic	22.9	0.096 (0.010)	11.4 (1.5)	0.410 (0.021)	80.3 (8.7)	1.00
D-ribose	Biphasic	ND	0.001 (7.1e^{-5})	0.3 (0.1)	0.004 (2.2e^{-4})	97.7 (11.5)	0.97
D-mannose	Multiphasic	20.7	0.024 (0.001)	2.6 (0.4)	0.151 (0.001)	157.1 (15.1)	1.00
L-leucine	Biphasic	ND	0.035 (0.001)	0.6 (0.2)	0.203 (0.015)	368.3 (57.1)	0.97
L-lysine	Diffusive	ND	0.096 (0.003)	3.7 (0.6)	ND	$K_D = 0.0003$	0.97
L-arginine	Monophasic	ND	2.028 (0.090)	255.1 (16.1)	ND	ND	0.99

Table 2.4 - Summary of turnover times and K_m(Fast)+S_n. Standard deviation reported in parentheses. D-Ribose not reported (see text).

CompoundName	Turnover Time [h]	$K_m + S_n$ [nM]
citric acid	75.3 (9.2)	7.2 (0.4)
hexanoic acid	277.2 (29.6)	12.9 (1.0)
1-hexanol	981.3 (262.0)	7.0 (0.4)
glyceraldehyde	109.0 (10.9)	9.7 (1.2)
D-mannose	112.2 (13.4)	2.6 (0.1)
L-leucine	44.1 (20.7)	2.4 (0.2)
L-lysine	54.4 (15.1)	7.0 (0.4)
L-arginine	74.6 (16.4)	38.1 (10.0)

Figure 2.1 - Summary of performance of assimilation and respiration methods. A: "Box-and-Whiskers" plot of loss of radioisotope for each labeled substrate. Calculated for all samples (n = 972) as the percent of isotope added minus the sum of activities caught on the filter, trapped in the cup, and left over in solution. B: Stacked bar plot of percent recovery of respirometry for each labeled substrate and $Na_2H^{14}CO_3$ (see Methods: Respiration; n = 12). Solid bars represent percent of added activity trapped in the cups; open bars represent percent of added activity remaining in solution. C: "Box-and-Whiskers" plot of killed control percent of assimilation for all kinetics samples (n = 486).
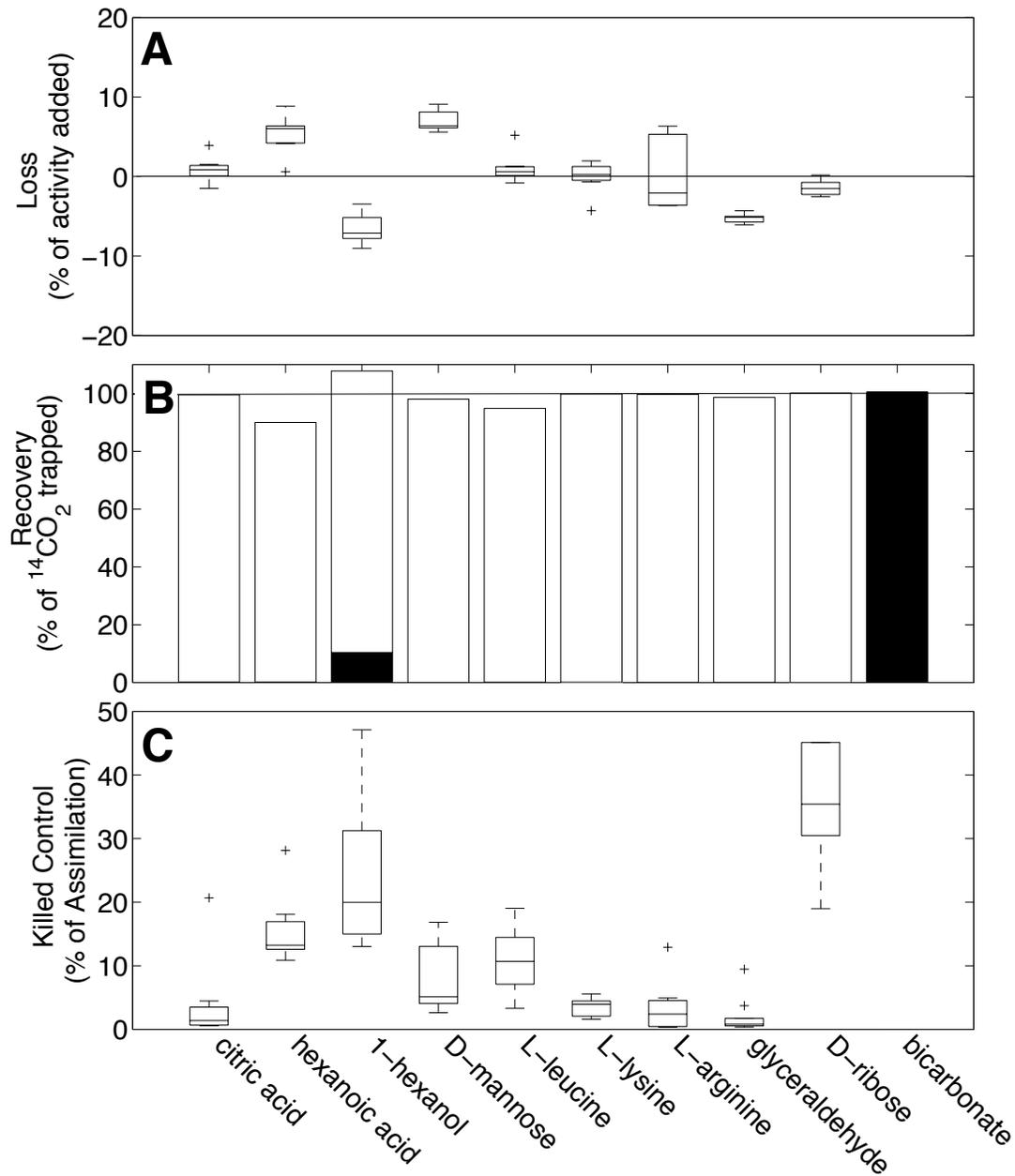
Figure 2.2 - Uptake kinetics ($v_U = v_A + v_R$) for all substrates. Mean values (Data from triplicate samples (open circles) are shown with corresponding NLS regressions (lines) are shown for AIC selected models. Values for all transport systems are given in Table 3.
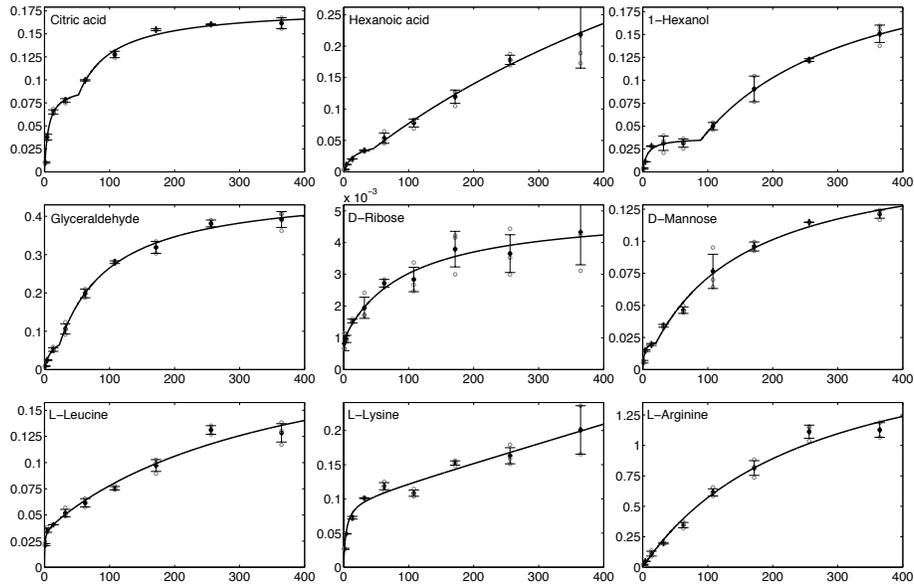
Figure 2.3 - Respiration : uptake (v_R / v_U) ratio for A: carbohydrates, B: carboxylic acids and C: amino acids over the S_A range (log scale). 1-Hexanol is not shown (see text).
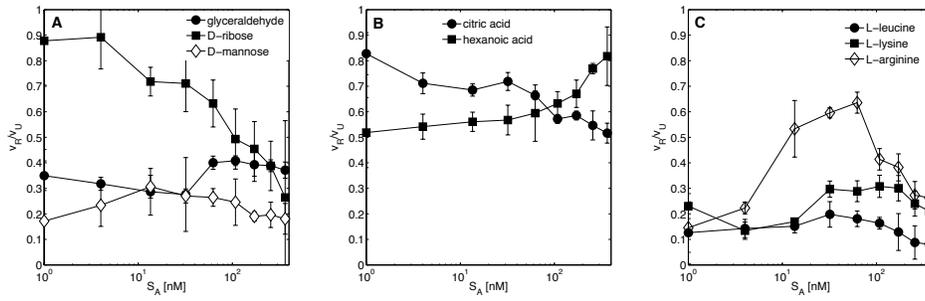
Figure 2.4 - "Ambient" substrate uptake rates (v_o) as a function of substrate affinity (a_F^o, a_S^o). Closed symbols correspond to fast uptake kinetics and open symbols correspond to slow uptake kinetics. Error bars represent one standard deviation for both variables. Values are plotted with the line of unity for reference.
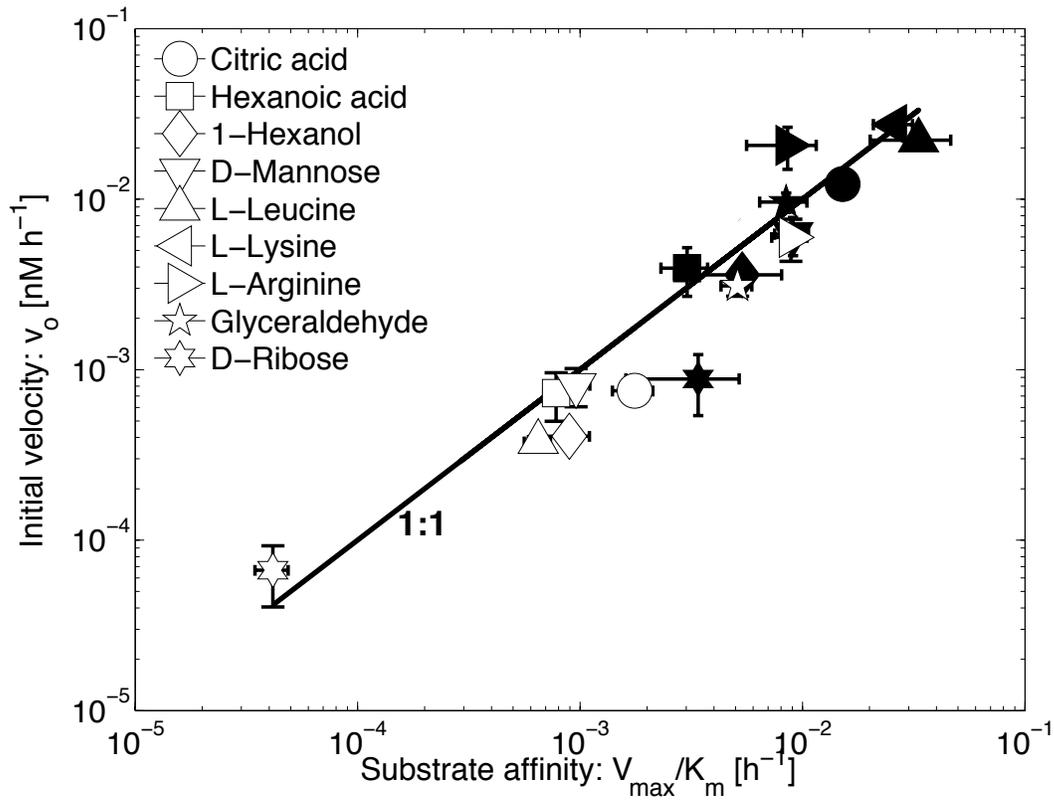
Figure 2.5 - Turnover time of carbon substrates as a function of energy density. Error bars represent one standard deviation of the mean.
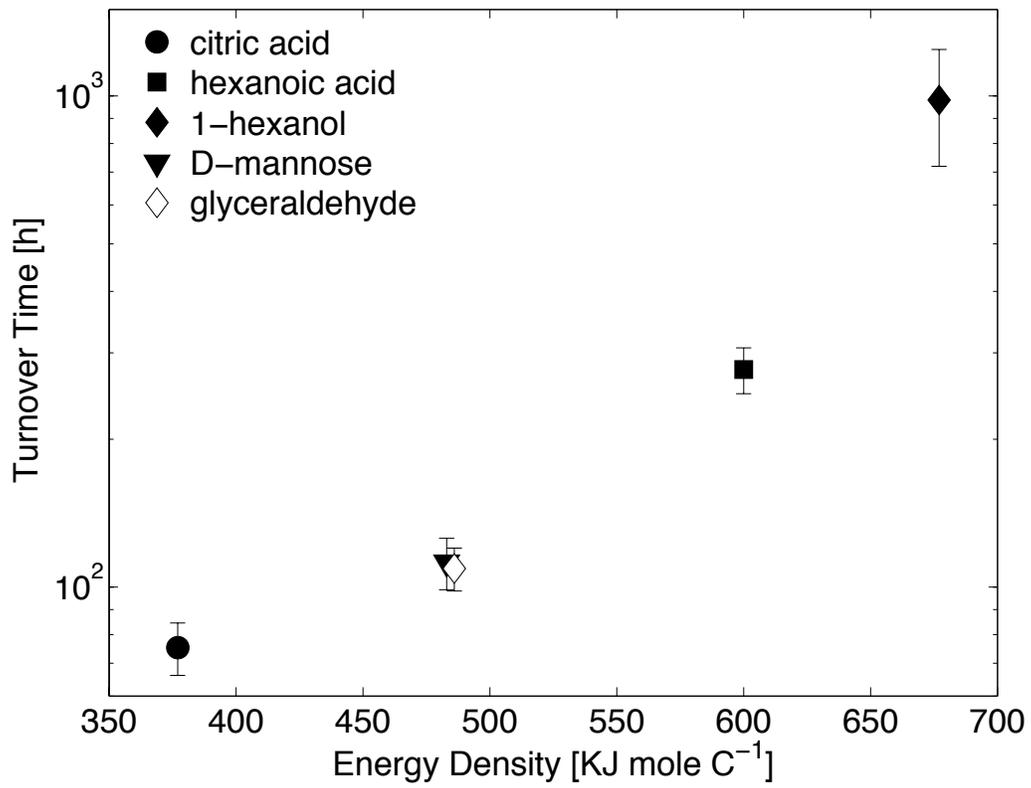
Figure 2.6 - Summary of competition experiments. Colors represent relative % change compared to control (the diagonal is defined as unity, 100%). Symbols represent significant (*p*<0.05; Student's t-test) Substrate names are abbreviated for clarity: *cit* – citric acid, *hxoic* – hexanoic acid, *hxol* – 1-hexanol, *leu* – L-leucine, *lys* – L-lysine, *arg* – L-arginine, *glyc* – glyceraldehyde, *ribose* – D-ribose, *mann* – D-mannose.
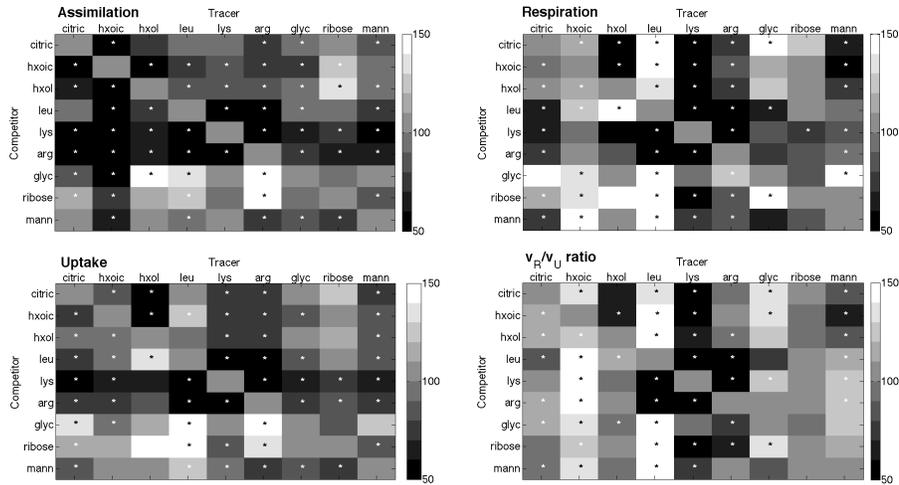
Photoheterotrophy enhances microbial growth efficiency

John R. Casey, Sara Ferrón, David M. Karl

Abstract

Molecular evidence for proteorhodopsin and bacteriochlorophyll based photoheterotrophy is widespread in oligotrophic marine microbial community metagenomes, and has been implicated in light-enhanced growth rates, substrate uptake rates, and anapleurotic carbon fixation, thus complicating the web of interactions within the 'microbial loop'. We quantified photoheterotrophic metabolism of the oxidized organic acid, glycolate, a fast-turnover (0.63 d^{-1}) and exclusively phytoplankton-derived substrate at an oligotrophic site in the subtropical North Pacific Ocean. As expected, concentration-dependent changes in uptake rates were observed over the diel cycle, with maxima occurring at midday. Although no light-enhanced substrate uptake rates were observed, samples exposed to light altered assimilation and respiration rates, resulting in an approximately 4-fold increase in glycolate-specific growth efficiency. Energy demand for such a metabolic change was linearly related to light, strongly suggesting phototrophy.

3.1 Introduction

The web of microbially mediated transformations of carbon and energy in the oceans is intricate and dynamic. The conduit through which a great majority of oceanic respiration is channeled is the marine dissolved organic matter (DOM) reservoir. DOM is composed of thousands of unique molecules in widely varying concentrations (Mopper *et al.,* 2007), and the spectrum of turnover spans from minutes (Fuhrman and Ferguson, 1986) to millennia (Ziolkowski and Druffel, 2010). Among other physicochemical attributes, the thermodynamic properties of organic substrates governs their turnover, with high enthalpy substrates supporting sub-optimal microbial growth rates and therefore turning over more slowly than low enthalpy substrates (Casey *et al.,* 2015). However, since the discovery of two unique light-harvesting systems widespread in marine bacteria and archaea, aerobic anoxygenic photoautotrophy (AAP; Shiba *et al.,* 1979) and proteorhodopsin (PR) photoautotrophy (PRP; Béjà *et al.,* 2000), the traditional view of a primary producer-DOM-secondary producer microbial loop (Azam *et al.,* 1983) may need revision (Karl, 2014). Collectively, PRP and AAP bacteria and archaea comprise most of the total heterotrophic microbial community in oligotrophic marine ecosystems (Rusch *et al.,* 2007), and PR's have been found in diverse bacterial phyla (McCarren and DeLong, 2007), including the numerically dominant alpha-proteobacterium *Pelagibacter ubique* SAR11, and in marine archaea (Frigaard *et al.,* 2006). While nutrient and ion transport have been associated with rhodopsins (Feng *et al.,* 2013; Kwon *et al.,* 2013; Inoue *et al.,* 2013; Yoshizawa *et al.,* 2014; Dimroth, 1990; Chan *et al.,* 1981), both PR and the bacteriochlorophyll (BChl) complex are capable of

generating a proton motive force (pmf) to supplement the ATP demands of biosynthetic and maintenance functions. The PR pmf generated has been shown in monoclonal cultures to markedly stimulate growth rates (Gomez-Consarnau *et al.,* 2007; Kimura *et al.,* 2011; Palovaara *et al.,* 2014), anapleurotic carbon fixation rates (Palovaara *et al.,* 2014), substrate uptake rates (Alonso-Saez *et al.,* 2006; Michelou *et al.,* 2007; Mary *et al.,* 2008), and to resuscitate carbon-starved cells (Gomez-Consarnau *et al.,* 2010; Steindler *et al.,* 2011). Whether the high abundance and diversity of AAP and PRP in the marine environment indicates a physiological cost-benefit solution to energy limitation of heterotrophic microbial growth on thermodynamically sub-optimal substrates remains unclear.

The PRP and AAP pmf may provide microbes with a readily available energy source to supplement, or perhaps to partly relieve oxidative phosphorylation demands. Indeed, Koblížek and co-authors (Koblížek *et al.,* 2010) measured a 70% reduction in respiration rates of an AAP *Roseobacter* isolate when grown on glutamate as a sole carbon source in the presence of light. Accordingly, the 'shaft work' provided to facultative photoorganoheterotrophs by photochemical energy transduction should decouple substrate chemical energy potential from anabolic yields of obligate chemoorganoheterotrophs (von Stockar *et al.,* 2006). We hypothesized that light inhibition of photoorganoheterotrophic respiration alters substrate-specific growth yields within natural microbial assemblages, especially for substrates more oxidized than their anabolic end-products.

A suitable substrate to evaluate light-enhanced growth yield is the oxidized hydroxy acid, glycolate. Glycolate is secreted as a result of photorespiration from

photoautotrophic microbes like high-light adapted *Prochlorococcus* strains

(Bertilsson *et al.,* 2005), and some of which lack a complete salvage pathway (Casey

et al., 2016). Since photorespiration is likely the sole extracellular source of

glycolate, and since glycolate permease transporters and glycolate oxidases and

dehydrogenases are present in SAR11, glycolate represents a direct transfer of

oxidized, newly fixed photosynthate to support heterotrophic carbon and energy

demands. In this study, radiorespirometry experiments were conducted to

determine the concentration-dependent kinetics of glycolate uptake, the diel cycle of

glycolate uptake, and the effect of light within that cycle and as a function of depth

within the euphotic zone (5 – 100 m).

3.2 Methods

3.2.1 Station locations and sample collection

Stations sampled were on two separate expeditions (Cruise 1 – September,

2013 at 22° 75' N, 158° 00' W and Cruise 2 – July-August 2015 at 24° 25' N, 156° 45'

W) in the North Pacific Subtropical Gyre north of the island of O'ahu. The kinetics

experiment was carried out during Cruise 1. Cruise 2 followed a semi-Lagrangian

track near the center of an anticyclonic mode-water eddy feature, facilitated by

maintaining ship's position with World Ocean Circulation Experiment Surface

Velocity Profile drifters with 15 m depth drogues. Sampling for glycolate tracer

incubation time-series was conducted at 4 h intervals, uninterrupted over the

course of two days. Additionally, surface tethered drifter arrays, designed to

accommodate sample bottles suspended at 5, 25, 50, 75, and 100 m were deployed

before, and again after the time-series experiments. Water samples were collected

using a rosette of PVC Bullister bottles mounted on a frame equipped with dual

conductivity sensors, temperature sensors, pressure sensors, oxygen optodes,

transmissometers, and triplet fluorometers (SBE 911plus, Sea-Bird Electronics,

Inc.). Photosynthetically active radiation (PAR; 400 – 700 nm wavelength band) was

measured both in time-series experiments by shipboard quantum cosine sensor (LI-

190R; LI-COR Inc.) with data logger (LI-1000; LI-COR Inc.), and also in depth profiles

(8 – 190 m) by a free-falling profiling hyperspectral radiometer (Hyperpro, Satlantic

LP.). Incubation time-integrated PAR was calculated by scaling shipboard PAR to

incubator transmittance.

3.2.2 Community stocks, production, and respiration data

Alongside glycolate incubations, samples were collected for particulate

carbon (PC), chlorophyll a (Chl a), primary productivity (PP) rates, and microbial

community respiration (MCR) rates. PC, Chl a, and PP measurements were

conducted according to Hawaii Ocean Time-series standard protocols (Karl *et al.,*

2001; http://hahana.soest.hawaii.edu/hot/methods/results.html). Briefly, for PC, 4

liter samples were pre-filtered through 202 μm screen mesh and pressure filtered

onto pre-combusted 25 mm glass fiber filters (Whatman GF/F). Filters were then

stored at -20°C until analysis by high-temperature combustion elemental analyzer

(CE-440, Exeter Analytical, Inc.). For Chl a, 2 liter samples were filtered as above and

stored in acetone at -20°C until fluorometric analysis. For PP, 500 ml samples were

incubated *in situ* on a surface-tethered array deployed before sunrise and recovered

after sunset. Bottles were spiked with $H^{14}CO_3^-$ to yield a final radioactivity of approximately 2 MBq L^{-1}. After a 14 hour incubation, samples were filtered onto GF/F filters, acidified in glass scintillation vials with 1 ml 2M HCl and allowed to vent for 24 hours prior to the addition of 10 ml Ultima Gold LLN cocktail and liquid scintillation counting. MCR measurements were conducted according to the difference between gross oxygen production and net oxygen change (Ferrón *et al.,* 2016). Briefly, samples were collected in 125 ml Pyrex glass bottles, enriched to 2300 ‰ $H_2^{18}O$ (97.2% ^{18}O) and incubated *in situ* along with the PP array. Mass-to-charge (*m/z*) ratios of 32 ($^{16}O^{16}O$) and 34 ($^{18}O^{18}O$) were quantified and normalized to the biologically inert gas Ar (*m/z* 40) by a silicone membrane inlet quadrupole mass spectrometer (MIMS gas analyzer, Bay Instruments; HiCube 80 Eco, Pfeiffer Vacuum) equipped with a Faraday collector and secondary electron multiplier.

3.2.3 Glycolate kinetics and uptake experiments

Incubations for glycolate kinetics and diel uptake rates were conducted in semi-transparent acrylic (shaded to approximately match the 15 m depth of sampling) or darkened deckboard incubators flushed with circulating surface seawater to maintain *in situ* temperatures. Incubator bath temperatures were monitored by Hobo Pendant Data Loggers (Onset Computer Corp.).

Experimental procedures for glycolate incubations were described in Casey et al. (2015). Briefly, 60 ml samples were spiked with [U-^{14}C]-glycolic acid calcium salt (^{14}C-glycolate herein; American Radiolabeled Chemicals, Inc.) at a specific activity of 1.48 TBq mol^{-1}. For the kinetics experiment, nine spike concentrations

were added, ranging from 1 nM to 348 nM, spaced logarithmically. For all other

incubations, spike concentrations were 10 nM. Uptake time series samples were

collected at 4 h intervals for two days. All samples were incubated in triplicate and a

500 μL total activity aliquot was collected from each sample prior to incubation. For

assimilation rates, samples were filtered under gentle vacuum (<70 mBar) directly

after incubation onto 25 mm glass fiber filters (nominal pore size 0.3 μm; GF75,

Sterlitech Corp.) and rinsed with three volumes of 20 ml 0.2 μm filtered seawater.

Filters were transferred to 20 ml glass scintillation vials and submerged in 10 ml

scintillation cocktail (UltimaGold LLT, Perkin Elmer). To account for ^{14}C-glycolate

adsorbed to cells or glass fiber filters, a "killed-control" replicate sample poisoned

with 2% final concentration paraformaldehde was included prior to each

incubation. Killed-controls were incubated alongside live samples and processed

identically. Assimilation depth profiles were conducted alongside PP and MCR *in situ*

arrays.

For respiration rates, 125 ml glass serum bottles were fitted with rubber

sleeve stoppers pierced with center well cups containing a dry piece of fluted

cellulose paper (Whatman #2) suspended in the headspace. Respiration incubations

were terminated by first soaking the filter paper with 150 μL phenethylamine and

then acidifying the sample with 4 ml 4.5 N sulfuric acid through the gas-tight

stopper. The acidified sample was allowed to react for at least 48 hours before

removing the stoppers. This procedure is designed to completely capture the

respired $^{14}CO_2$. Center well cups were transferred to 20 ml glass scintillation vials

and submerged in 10 ml scintillation cocktail. Quantitative recovery (100 ± 4%) was

determined in the kinetics experiment, and recovery was independent of substrate concentration.

3.3 Results

3.3.1 Glycolate kinetics experiment

Two distinct kinetics profiles were observed for dark glycolate assimilation (v_a) and respiration (v_R; **Figure 3.1**). Assimilation followed a monophasic Michaelis-Menten function with a resulting half-saturation concentration ($K_{m,a}$) of 118 nM and a maximum velocity ($V_{max,a}$) of 28 nM h^{-1}. Respiration rates did not appear to completely saturate over the concentration range tested, therefore $K_{m,r}$ and $V_{max,r}$ could not be determined. The resulting uptake parameters $K_{m,u}$ and $V_{max,u}$ were calculated to be 195 nM and 152 nM h^{-1}. Glycolate-specific growth efficiency ($100*v_a/v_a+v_R$) varied as a logistic function of substrate concentration added, with highest efficiencies (29.3 ± 0.9 %) corresponding to $S_A < 57$ nM. At saturating concentrations, the growth efficiency approached 12%.

3.3.2 Glycolate diel time-series experiment

Uptake rates varied by roughly 3-fold (2.8 ± 0.7 fold) over the diel cycle, in phase with the solar cycle, and no difference between shaded and unshaded uptake rates were observed (two-sample t-test with unequal variance; $p = 0.39$; **Figure 3.2**). Assimilation rates also followed a diel cycle (3.1 ± 1.1 fold), but with maximal rates occurring in unshaded samples in the early morning (0600 – 1000 local time). Assimilation rates in shaded samples were $35 \pm 7\%$ lower than unshaded samples

(p = 0.007) during daylight hours, but were similar to nighttime samples (p = 0.42). In contrast, respiration rates in shaded samples were 120 ± 14% higher than shaded samples (p = 9e^{-6}) during daylight hours, and unshaded samples were similar over the entire daylength (p = 0.71). The discrepancy between daytime light-dependent responses in assimilation and respiration rates resulted in large variations (3.5 ± 0.9 fold; p = 9e^{-7}) in glycolate-specific growth efficiencies, ranging from 24 ± 6% in shaded daytime samples to 82 ± 8% in unshaded daytime samples.

3.3.3 Glycolate assimilation depth profile experiment

Assimilation rates decreased exponentially with depth (**Figure 3.3**), and were more closely matched to PAR (Model II geometric mean least squares fit; r^2 = 0.99) than to PP (r^2 = 0.79). In consideration of incubation lengths, array assimilation rates at 25 m were similar to time-series assimilation rates averaged over the array deployment interval (p = 0.71).

3.4 Discussion

Glycolate, a low molecular weight (76 Da) organic acid, may constitute an important flux of both carbon and energy within the marine microbial community metabolism. Dark glycolate uptake kinetics indicated an upper bound (Laws, 1983) ambient concentration of 30 ± 6 nM, based on turnover time as a function of substrate added (Dietz *et al.,* 1977). This estimate, though an upper bound, is roughly half the nighttime concentration measured by HPLC at an oligotrophic (60-70 ng Chl a L^{-1}) site in the tropical eastern North Atlantic (66 nM; Leboulanger *et al.,*

1997), and is well below the $K_{m,u}$ (119 nM). During the dark kinetics incubation, the glycolate pool turned over approximately daily (1.1 ± 0.1 d); however, this estimate is likely an underestimate due to disequilibrium with respect to production (which is exclusively during daylight hours).

Time-series incubations showed a characteristic diel cycle in uptake rates in phase with the solar cycle, independent of whether the sample was exposed to light. Taken together with our understanding of a photorespiratory source of glycolate, these results indicate ambient concentration-dependent rates, rather than light-enhanced uptake by heterotrophs. Although the time series experiment and kinetics experiment were conducted on separate expeditions with somewhat different phytoplankton community stocks and rates (Cruise 1: Chl a = 80 ng L^{-1}, PP = 8.1 mg C L^{-1} d^{-1}; Cruise 2: Chl a = 137 ng L^{-1}, PP = 9.9 mg C L^{-1} d^{-1}), nighttime uptake rates (10 nM spike) collected during the time-series experiment closely matched the corresponding values from the kinetics curve (p = 0.81). With this caution, we estimate a 2-fold change in ambient glycolate concentrations over the course of the diel cycle, which is consistent with estimates from the North Atlantic (2.4 ± 1.2 fold; Leboulanger *et al.,* 1998). Glycolate-specific respiration rates accounted for approximately 3% of total community oxygen respiration rates, and considering the glycolate respiration quotient, 6% of total community respired CO_2 (assuming a total community RQ of 1.0; del Georgio *et al.,* 2006; c.f., Berggren *et al.,* 2011). On a carbon basis, glycolate production rates accounted for less than 4% of gross oxygen production in the time-series incubations, however this is can also be interpreted as an underestimate of gross photorespiration since salvage pathways are present in

some photoautotrophs. Due to methodological challenges, photorespiration rates have not been directly quantified in the oceans, and may be an important but largely ignored flux of carbon.

High-light adapted 'ecotypes' (eHL) of the marine cyanobacterium *Prochlococcus*, the most abundant photoautotroph at Station ALOHA, lacks glycolate oxidase or glycolate dehydrogenase, an essential step in the salvage pathway for photorespiratory glycolate regeneration of 3-phosphoglycerate. Because the precursor 2-phosphoglycolate is toxic to central carbon fixation pathways, *Prochloroccus* actively excretes glycolate via an ATP-binding cassette efflux transporter. In cultures of two eHL *Prochlorococcus* strains, glycolate excretion was 3% of carbon fixation (Bertilsson *et al.*, 2005), remarkably close to our upper bound estimate. It should be noted that the diazotrophic cyanobacterium *Crocosphaera*, which was relatively abundant during the time-series expedition (100-700 cells ml^{-1}; 3% of GOP; Wilson *et al.*, in review), does have a complete photorespiratory salvage pathway. Therefore we cannot eliminate the possibility that *Crocosphaera* could take up glycolate, though we could not find mention of any photoautotroph utilizing glycolate. Notwithstanding, *Prochlorococcus* was the dominant primary producer during the time-series expedition, and was likely the major glycolate producer. Interestingly, a major consumer of glycolate is likely the numerically dominant heterotroph at Station ALOHA, *Pelagibacter ubique* SAR11 (SAR11 hereafter), a small alphaproteobacterium with an absolute growth requirement for pyruvate and either glycolate or one of several photorespiratory salvage pathway intermediate metabolites. SAR11 has both a glycolate transporter and glycolate

oxidase which yields glyoxylate and hydrogen peroxide. In addition to the apparent

co-evolution of these two dominant oligotrophs (Braakman *et al., in review*), SAR11

and much of the heterotrophic microbial community at Station ALOHA (Rusch *et al.,*

2007), have genes for proteorhodopsin-based phototrophy, prompting our

investigation into the light-dependent metabolism of glycolate. With a respiration

quotient (mol CO_2 : mol O_2) of 2, a carbon redox number of 3, and a standard carbon

molar enthalpy of combustion (ΔH°_c) of 340 KJ $[C\text{-mol}]^{-1}$, glycolate is a relatively

poor energy substrate. Accordingly, heterotrophic growth on oxidized carboxylic

and hydroxy acids like acetate and fatty acids typically requires the operation of the

glyoxylate shunt (Kornberg, 1966), a bypass of two CO_2 evolving steps of the

oxidative tricarboxylic acid pathway by way of isocitrate lyase and malate synthase

which allows the net accumulation of carbon through acetyl-CoA. However, at least

one alternative pathway utilizing glyoxylate is present in SAR11 and many other

heterotrophs, consisting of a heterotrophic analogue of the photorespiratory

salvage pathway which can supply precursors for gluconeogenesis (by way of 3-

phosphoglycerate) or a number of amino acid synthesis pathways (by way of L-

glycine). These anabolic pathways cannot be sustained without a supplemental

energy source, since the ATP/NAD(P)H ratio and yield of the glyoxylate shunt using

glycolate as a substrate does not satisfy the demands of e.g., protein synthesis

(calculation based on *i*AF1260, a metabolic model of *Escherichia coli* K-12 MG1655;

Feist *et al.,* 2007). The central finding of this study, that exposure to light enhances

the glycolate-specific growth efficiency, points to the possibility that the pmf

generated by proteorhopsin phototrophy or by the BChla-complex yields sufficient

energy to divert glycolate flux from the mostly catabolic glyoxylate shunt to the mostly anabolic pathways. Considering the ΔH°_c and the maximum chemical potential energy yield of glycolate respiration, and assuming a quantitative energy transduction by PR or BChla-complex, the energy yield was closely correlated with PAR irradiance integrated over each incubation ($r^2 = 0.94$; $p = 4e^{-6}$; **Figure 3.4**). Unfortunately, it is not possible to 'scale' glycolate-specific phototrophic energy yields to total photoheterotrophy, since the composition of the myriad additional growth substrates and their respective uptake rates and light-dependent growth efficiencies is not known. Even if during their co-evolution (Braakman *et al., in review*) *Prochlorococcus* has adapted to 'growing SAR11' on a thermodynamically optimal substrate (glycolate and pyruvate; thermodynamic efficiency = 20-24%; Westerhoff *et al.,* 1983), SAR11 could be fighting back by enhancing growth yields with the help of PR.

3.5 References

Alonso-Saez, L., Gasol, J. M., Lefort, T., Hofer, J., Sommaruga, R., 2006. Effect of natural sunlight on bacterial activity and differential sensitivity of natural bacterioplankton groups in northwestern Mediterranean coastal waters. *Applied and Environmental Microbiology* 72**:** 5806–5813.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., 1983. The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series* 10: 257-263.

Béjà, O., 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* 289, 1902–1906.

Berggren, M., Lapierre, J.F., del Giorgio, P.A., 2011. Magnitude and regulation of bacterioplankton respiratory quotient across freshwaters environmental gradients. *ISME Journal* 6, 984-993.

Bertilsson, S., Berglund, O., Pullin, M., Chisholm, S.W., 2005. Release of dissolved organic matter by *Prochlorococcus*. *Vie et Milieu* 55, 225–232.

Braakman, R., Follows, M.J., Chisholm, S.W., in review. Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*.

Chan, W.Y., Mosca, P., Rennert, O.M., 1981. Lithium nephrotoxicity: a review. *Annals of Clinical Laboratory Science* 11, 343-349.

Casey, J., Falkowski, P.G., Karl, D.M., 2015. Substrate selection for heterotrophic bacterial growth in the sea. *Marine Chemistry* 177: 349-356.

del Giorgio, P.A., Pace, M., Fischer, D., 2006. Relationship of bacterial growth efficiency to spatial variation in bacterial activity in the Hudson River. *Aquatic Microbial Ecology* 45, 55–67.

Dietz, A.S., Albright, L.J., Tuominen, T., 1977. Alternative model and approach for determining microbial heterotrophic activities in aquatic systems. *Applied and Environmental Microbiology* 33: 817-823.

Dimroth, P., 1990. Mechanisms of sodium transport in bacteria. *Philosophical Transactions of the Royal Society of London B* 326: 465-477.

Feng, S., Powell, S.M., Wilson, R., Bowman, J.P., 2013. Light-stimulated growth of proteorhodopsin-bearing sea-ice psychrophile *Psychroflexus torquis* is salinity dependent. *ISME Journal* 7, 2206–2213.

Ferrón, S., del Valle, D.A., Bjorkman, K.M., Quay, P.D., Church, M.J., Karl, D.M., 2016. Application of membrane inlet mass spectrometry to measure aquatic gross primary production by the ^{18}O *in vitro* method. *Limnology and Oceanography: Methods* 14, 610–622.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction *for Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* 3: 121.

Frigaard, N.-U., Martinez, A., Mincer, T.J., Delong, E.F., 2006. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439, 847–850.

Fuhrman, J.A., Ferguson, R.L., 1986. Nanomolar concentrations and rapid turnover of dissolved free amino-acids in seawater-agreement between chemical and microbiological measurements. *Marine Ecology Progress Series* 33, 237–242.

Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., Pedrós-Alió, C., Pinhassi, J., 2007. Light stimulates growth of proteorhodopsin-containing marine flavobacteria. *Nature* 445, 210–213.

Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L., González, J.M., Pinhassi, J., 2010. Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biology* 8, e1000358.

Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., 2013. A light-driven sodium ion pump in marine bacteria. *Nature Communications* 4: 1678.

Karl, D.M., Dore, J.E., 2001. Microbial ecology at sea: sampling, subsampling and incubation considerations. *Methods in Microbiology* 30, 13–39.

Karl, D.M., 2014. Solar energy capture and transformation in the sea. *Elementa: Science of the Anthropocene* 2, 000021.

Kimura, H., Young, C.R., Martinez, A., Delong, E.F., 2011. Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. *ISME Journal* 5, 1641–1651.

Koblížek, M., Mlčoušková, J., Kolber, Z., Kopecký, J., 2010. On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Archives of Microbiology* 192, 41–49.

Kornberg, H.L., 1966. The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochemistry Journal* 99: 1-11.

Kwon, S.K., Kim, B.K., Song, J.Y., Kwak, M.J., Lee, C.H., Yoon, J.H., Oh, T.K., Kim, J.F., 2013. Genomic makeup of the marine flavobacterium *Nonlabens* (*Donghaeana*) *dokdonensis* and identification of a novel class of rhodopsins. *Genome Biology and Evolution* 5, 187–199.

Laws, E.A., 1983. Plots of turnover times versus added substrate concentrations provide only upper bounds to *in situ* substrate concentrations. *Journal of Theoretical Biology* 101: 147-150.

Leboulanger, C., Oriol, L., Jupin, H., Desolas-gros, C., 1997. Diel variability of glycolate in the eastern tropical Atlantic Ocean. *Deep-Sea Research Part I: Oceanographic Research Papers* 44, 2131–2139.

Leboulanger, C., Serve, L., Comellas, L., Jupin, H., 1998. Determination of glycolic acid released from marine phytoplankton by post-derivatization gas chromatography-mass spectrometry. *Phytochemical Analysis* 9, 5–9.

Mary, I., Tarran, G.A., Warwick, P.E., Terry, M.J., Scanlan, D.J., Burkill, P.H., Zubkov, M.V., 2008. Light enhanced amino acid uptake by dominant bacterioplankton groups in surface waters of the Atlantic Ocean. *FEMS Microbiology Ecology* 63, 36–45.

McCarren, J., Delong, E.F., 2007. Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environmental Microbiology* 9, 846–858.

Michelou, V.K., Cottrell, M.T., Kirchman, D.L., 2007. Light-stimulated bacterial production and amino acid assimilation by cyanobacteria and other microbes in the North Atlantic Ocean. *Applied and Environmental Microbiology* 73, 5539–5546.

Mopper, K., Stubbins, A., Ritchie, J.D., Bialk, H.M., Hatcher, P.G., 2007. Advanced instrumental approaches for characterization of marine dissolved organic matter:

Extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chemical Reviews* 107, 419-442.

Palovaara, J., Akram, N., Baltar, F., Bunse, C., Forsberg, J., Pedrós-Alió, C., González, J.M., Pinhassi, J., 2014. Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 111: E3650-E3658.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.-H., Falcón, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., Venter, J.C., 2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* 5, e77.

Shiba, T., Simidu, U., Taga, N., 1979. Distribution of aerobic bacteria which contain bacteriochlorophyll a. *Applied and Environmental Microbiology* 38: 43-45.

Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., 2011. PLOS ONE: Energy starved *Candidatus pelagibacter ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One* 6: e19725.

von Stockar, U., Maskow, T., Liu, J., Marison, I.W., 2006. Thermodynamics of microbial growth and metabolism: an analysis of the current situation. *Journal of Biotechnology* 121, 517–533.

Westerhoff, H.V., Hellingwerf, K.J., Van Dam, K., 1983. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proceedings of the National Academy of Sciences of the United States of America* 80, 305–309.

Wilson, S.T., Aylward, F.O., Ribalet, F., Barone, B., Casey, J.R., Connell, P.E., Eppley, J.A., Ferron, S., Romano A.E., Turk-Kubo, K.A., Vislova, A., Armbrust, E.V., Caron, D.A., Church, M.J., Zehr, J.P., Karl, D.M., DeLong, E.F., in review. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocosphaera*. *Nature Microbiology*.

Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W., Delong, E.F., Kogure, K., 2014. Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 111, 6732–6737.

Ziolkowski, L.A., Druffel, E.R.M., 2010. Aged black carbon identified in marine dissolved organic carbon. *Geophysical Research Letters* 37, L16601.

Figure 3.1 – Kinetics experiment Glycolate assimilation and respiration as a function of substrate added. Error bars represent one standard deviation of the mean rates at each concentration. Michaelis-Menten non linear least squares regression line is shown for assimilation data.
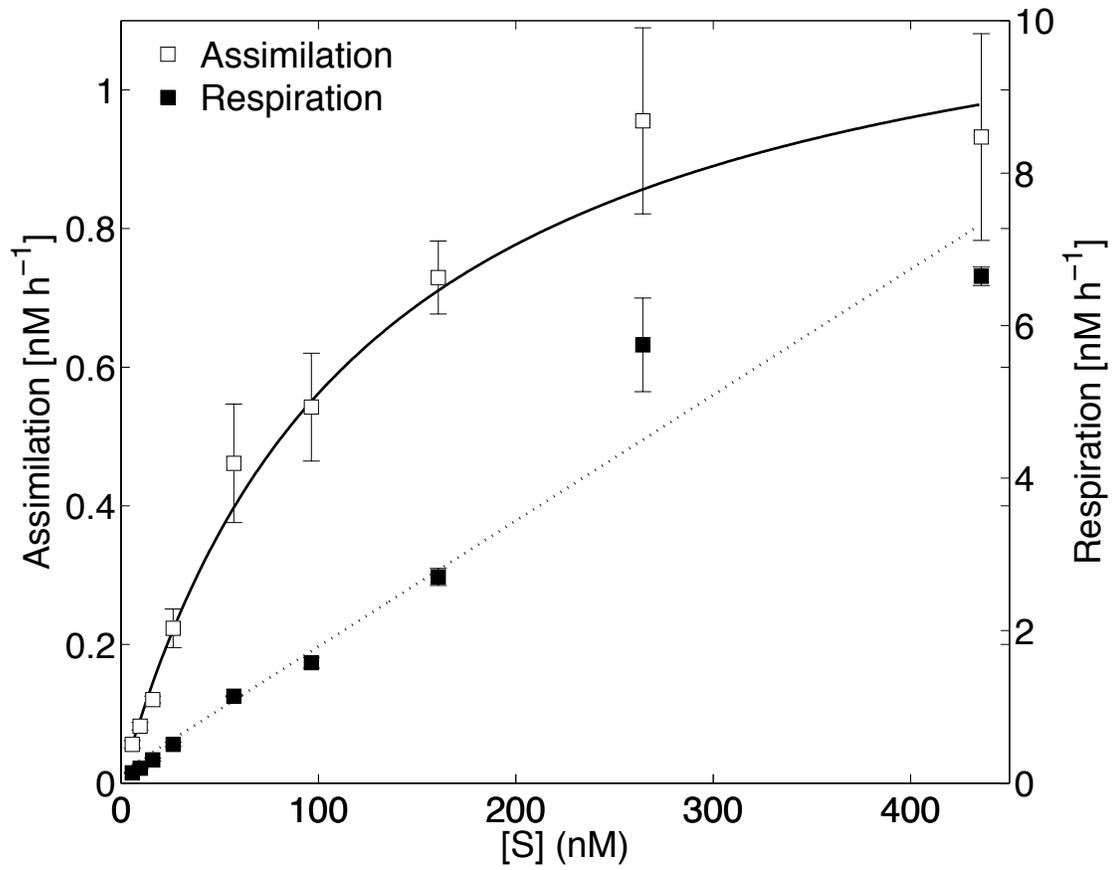
Figure 3.2 – Time-series experiment Glycolate uptake (top panel), assimilation (second panel), respiration (third panel), and specific growth efficiency (bottom panel) for shaded and unshaded incubations over the course of the diel time-series experiment. Mean (large symbols) and individual data points (small symbols) are shown for clarity. Photosynthetically available radiation (PAR) data is overlaid (green line) in each panel.
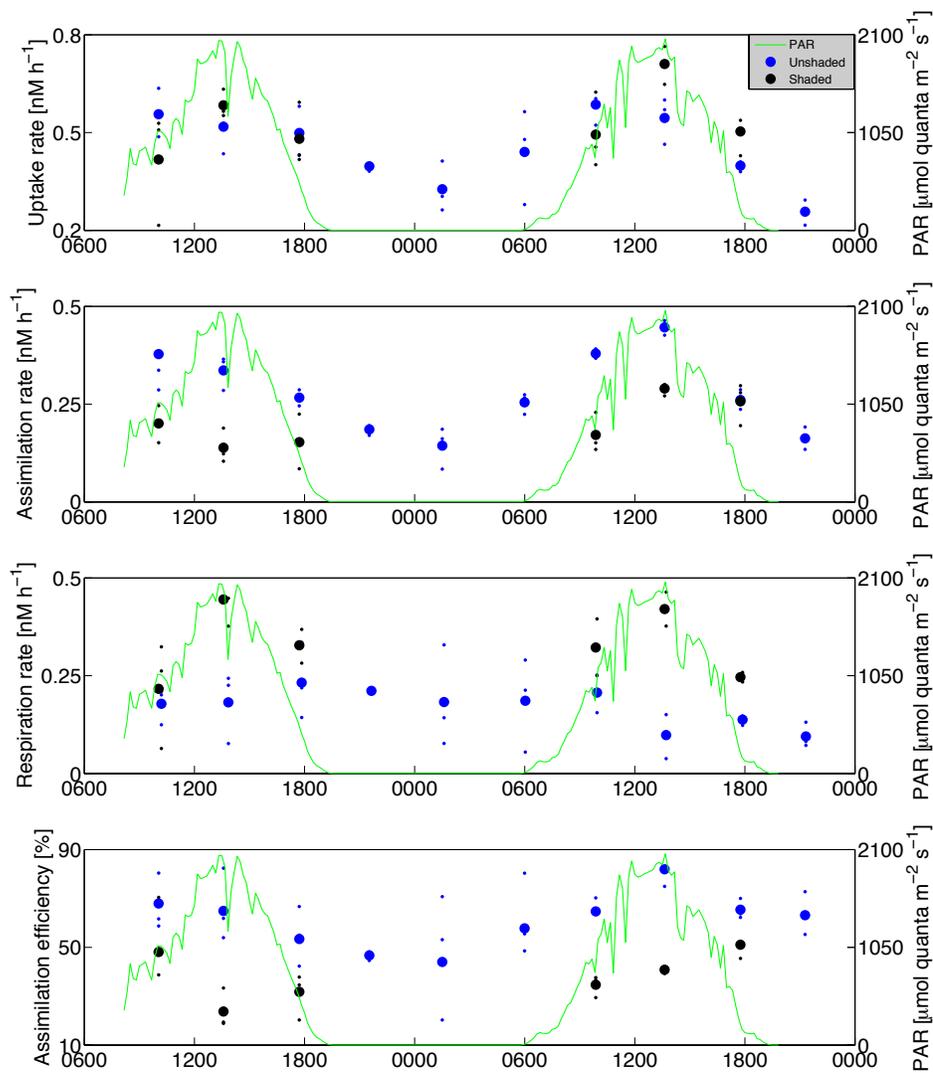
Figure 3.3 – Glycolate assimilation array Depth profile of glycolate assimilation

rates. Error bars represent one standard deviation of the mean rates at each depth.
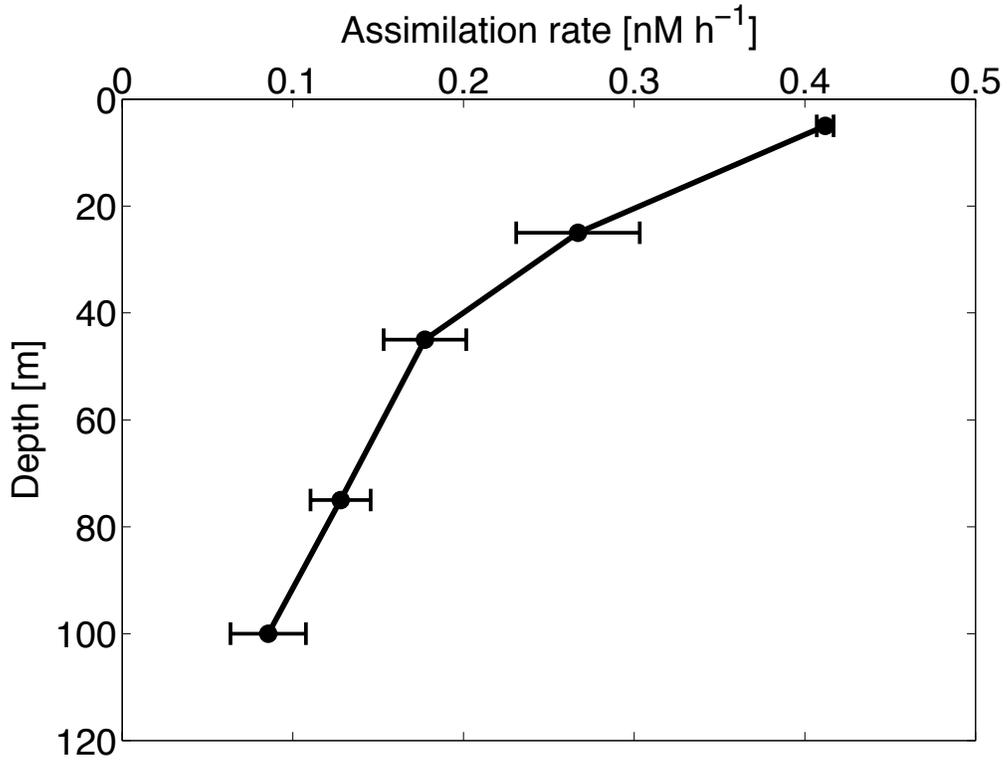
Figure 3.4 – Energy transduction Calculated glycolate-specific energy transduction (see main text for details) over the diel time-series experiment (top panel) and as a function of incubation time integrated PAR (bottom panel). Model II geometric mean least squares regression line is shown.
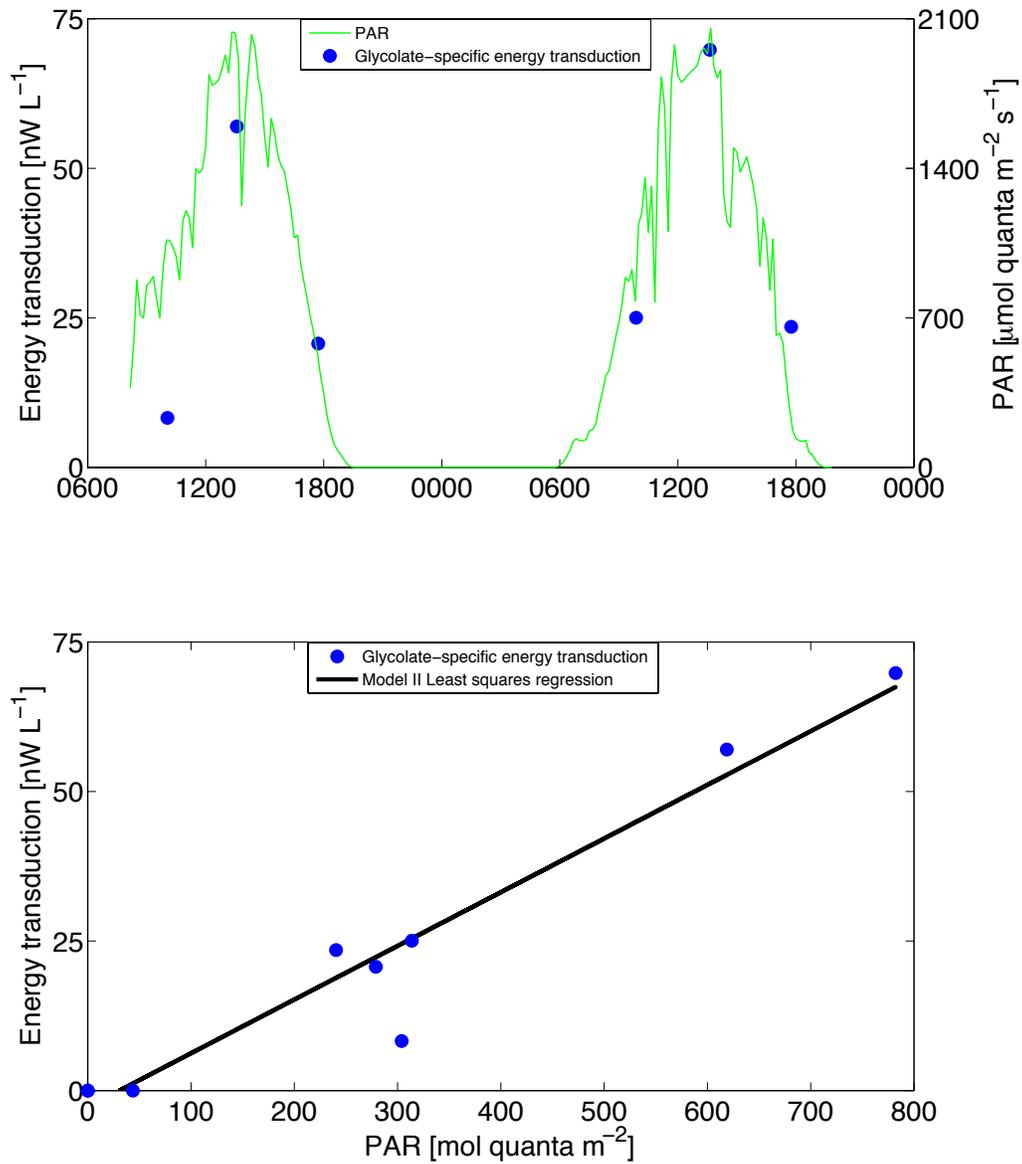
Adaptive evolution of phosphorus metabolism in Prochlorococcus

John R. Casey, Adil Mardinoglu, Jens Nielsen, David M. Karl

Abstract

Inorganic phosphorus is scarce in the Eastern Mediterranean Sea, where the high-light adapted ecotype HLI of the marine picocyanobacterium *Prochlorococcus marinus* thrives. Physiological and regulatory control of phosphorus acquisition and partitioning has been observed in HLI both in culture and in the field, however the optimization of phosphorus metabolism and associated gains for its phosphorus limited growth (PLG) phenotype have not been studied. Here we reconstructed a genome-scale metabolic network of the HLI axenic strain MED4 (*i*JC568), consisting of 568 metabolic genes related with 794 reactions involving 680 metabolites distributed in 6 sub-cellular locations. *i*JC568 was used to quantify metabolic fluxes under PLG conditions and we observed a close correspondence between experimental and computed fluxes. We found that MED4 has minimized its dependence on intracellular phosphate, not only through drastic depletion of phosphorus containing biomass components, but also through network-wide reductions in phosphate-reaction participation and the loss of a key enzyme, succinate dehydrogenase. These alterations occur despite the stringency of having relatively few pathway redundancies and an extremely high proportion of essential metabolic genes (47%; defined as the percentage of lethal *in silico* gene knockouts). These strategies are examples of nutrient-controlled adaptive evolution and confer a dramatic growth rate advantage to MED4 in phosphorus limited regions.

Importance

Microbes are known to employ three basic strategies to compete for limiting elemental resources: (*i*) cell quotas may be adjusted by alterations to cell physiology

114

or by substitution for a more plentiful resource, (*ii*) stressed cells may synthesize high-affinity transporters, and (*iii*) cells may access more costly sources from internal stores, by degradation or by petitioning other microbes. In the case of phosphorus, a limiting resource in vast oceanic regions, the cosmopolitan cyanobacterium *Prochlorococcus marinus* thrives by adopting all three strategies and a fourth, previously unknown strategy. By generating a detailed model of its metabolism, we found that strain MED4 has evolved a way to reduce its dependence on phosphate by minimizing the number of enzymes involved in phosphate transformations, despite the stringency that nearly half of its metabolic genes are essential for survival. Relieving phosphorus limitation, both physiologically and throughout intermediate metabolism, substantially improves phosphorus-specific growth rates.

4.1 Introduction

The picocyanobacterium *Prochlorococcus marinus* is the numerically dominant photoautotroph in the vast oligotrophic gyres, where it often contributes a majority of carbon fixation [1]. Ecotypes of the *Prochlorococcus* lineage occupy a broad ecological niche space, and its success has been attributed to its small size, a highly streamlined and nearly minimal genome, and physiological adaptations to low nutrient environments (2). Natural populations and laboratory isolates adjust their elemental quotas widely in response to nutrient supply by a variety of intriguing mechanisms. Among other notable adaptations, they are capable of utilizing organic substrates to supplement inorganic nutrient deficits (3, 4, 5), replacing phospholipids for sulfolipids and glycolipids under phosphate-deplete conditions (6), and coordinating proteome-wide control of nitrogen allocation under nitrogen stress (7). The first axenic strain cultivated of the *e*MED4 high-light ecotype lineage, MED4 (and synonymous genotypes PCC9511, CCMP1378, and CCMP1986 [8]), has been the subject of numerous physiological studies, and was originally isolated from the chronically phosphate-deplete surface waters of the Eastern Mediterranean Sea, where it is numerically dominant (9, 10). The phosphorus-stress response has been described in detail for MED4 regulation of gene expression (11, 12), and for physiological responses of the cell cycle (13), elemental composition (14), phosphorus substrate uptake rates (15), alkaline phosphatase activity (3), and lipid composition (6). The regulatory circuit includes upregulation of the sensor kinase complex *phoBR* and subsequent upregulation of the high-affinity phosphate transporter system *pstABCS* and *pho* operon including

alkaline phosphatase. We hypothesized that, in addition to these regulatory and physiological responses, MED4 has optimized, through adaptive gene loss (16), its metabolic network to cope with low phosphate availability. To this end, we sought a quantitative method to predict metabolic capabilities of the MED4 genotype and its phosphorus limited growth (PLG) phenotype.

Genome scale metabolic (GEM) network reconstructions represent a cornerstone of systems biology, serving as both a 'knowledgebase' for contextualizing physiological and multi-omics data types, and as a framework for computational approaches such as constraint-based modeling (17, 18). GEMs are available for a broad spectrum of microbes and model organisms, ranging widely in network size, complexity, and "quality" (based on the scoring criteria of Thiele *et al.* [19]). Quantitative metabolic flux predictions using constraint-based flux balance analysis (FBA) have been validated experimentally for a range of different organisms. Furthermore, FBA and related approaches have proven valuable for strain engineering, natural product yield optimization, identification of inhibition targets for drug therapies, and numerous other industrial and medical applications (20). Despite these routine applications, to our knowledge no ecological applications have been reported. Indeed, GEMs of ecologically relevant microbes could complement trait-based and cellular resource allocation models, which benefit from broader taxonomic coverage, and perhaps be nested in global biogeochemical models.

Here we reconstructed a GEM for MED4 and used the model to simulate growth under a variety of defined media to describe its metabolic capabilities and quantify fluxes associated with phosphate, carbon, and light limited growth (PLG, CLG and LLG, respectively) conditions. The imprint of adaptive evolution in phosphate-deplete conditions was found throughout the MED4 metabolic network, with implications for global cellular elemental turnover and energy metabolism. To explore MED4 metabolism, we first describe fundamental properties of the metabolic network in relation to a diverse selection of microbial GEMs. We then used FBA to compute its metabolic capabilities, and compared model simulations with experimental data from both the lab and the field. Finally, we discuss genome wide alterations to phosphorus metabolism and implications for the PLG phenotype.

4.2 Methods

4.2.1 Network reconstruction

A metabolic network of MED4 was created following the reconstruction process detailed previously (19, 48). Briefly, an initial draft reconstruction was created by identifying protein homology with the KEGG Orthology (KO) database supplied through the BioMet Toolbox (http://www.sysbio.se/BioMet). Hidden Markov models (HMM) of protein sequences for each KO were retrieved and queried against the MED4 reference genome (NCBI GenBank: BX548174.1). Metabolic genes which were excluded from HMM hits were individually examined using different resources (NCBI, UniProt, IMG, BioCyc, ProPortal). General and unbalanced reactions were excluded and extensive manual curation was performed for gap-filling and balancing process, due in part to the conservative reaction assignment criteria, as well as the incomplete genome annotation (30% of ORFs assigned to putative or unknown function), typical of marine cyanobacteria (e.g., 48% of ORFs assigned to putative or unknown function for Synechocystis sp. PCC6803). Draft model reactions were checked for elemental and charge balance, for known substrate and cofactor specificity, and for directionality. Reaction directionality was determined by thermodynamic favorability (49) followed by manual inspection and elimination of futile cycles and according to guidelines described in (19). Cofactor specificity, especially with regard to preference for NAD(H) and NADP(H), was often unknown; however, 4 genes (PMM1127, PMM1145, PMM1146, PMM1147) encoding the alpha and beta subunits of the reversible membrane-bound nicotinamide nucleotide transhydrogenase (EC 1.6.1.1;

R00112) effectively eliminated the need to differentiate these important cofactors.

Proteins were assigned to one of six sub-cellular locations: cytoplasmic membrane, periplasm, thylakoid membrane, thylakoid lumen, cytoplasm, or carboxysome. Protein localization was based on amino acid sequences using the PSORTb algorithm for bacteria (50) and the ExPASy tool DAS-TMfilter for transmembrane domain prediction (51). Proteins associated with the thylakoid membranes and carboxysomes are not predicted by PSORTb, and were instead inferred from homology to a detailed photophysiological model for *Synechocystis* sp. PCC6803 (52).

Gaps were identified by iteratively examining dead-end metabolites and associated blocked reactions, and returning to the literature for evidence of synthesis, degradation, secretion or uptake of associated metabolites. Conserved domains from the resulting orphan reactions were then queried against the MED4 genome by protein homology using BLASTp. This process identified several ORFs with high sequence identity. The resulting well-connected network was then queried for futile cycles and transport and exchange reactions were added. Several exchange reactions were added for protein complexes (e.g., acyl carrier protein, lipoylprotein) which are not explicitly synthesized by the network, though these carry no flux and are included only for modeling purposes. Fake exchange reactions were also added for dead-end metabolites not included in the biomass objective for which no transporters are annotated (e.g., glycolaldehyde, 7-aminomethyl-7-carbaguanine, methanol). These reactions can carry flux and are considered analogous to diffusive transport. A tunable ATP sink was introduced, also for

modeling purposes, to account for costs associated with photo-damage above an experimentally determined irradiance (549 µmol photons m^{-2} s^{-1} [53]), though this reaction is constrained to zero unless explicitly stated herein. Transporter proteins are particularly poorly annotated in the MED4 genome, so physiological evidence alone was required for transporter presence in some cases. Because transport may, variously, be chemiosmotic (symporter/antiporter ion pumps) or mediated by ATP hydrolysis, it is likely that *i*JC568 is not accurately charge balanced with respect to major ions (e.g., K$^+$, Ca^{2+}). In all cases, reaction presence/absence was scored for evidence according to (19).

The process of building an *in silico* metabolic reconstruction is, historically, a series of iterative improvements whereby the model grows in size and complexity, often with detail added to specific pathways as experimental data become available. Open code and computational design is essential to this process, and we have made efforts to enable community contributions. The model is fully MIRIAM compliant and is available in standard formats (SBML for RAVEN and BioOpt at http://biomet-toolbox.org; Excel format in **Supplementary File 4.1**). Since naming conventions and database link identifiers differ widely, the Excel file contains additional fields to identify reactions (SBO Terms, KEGG Orthology, EC Codes), metabolites (molecular formula, molecular weight, charge, IUPAC Names, InChI, InChIKey, PubChem CID, and KEGG Compound), and genes (KEGG Gene, NCBI Accession, UniProt ID) which are intended to aid in formatting conversions for ease of sharing. Simulation results and the BOF are available as tabs in a separate Excel file (**Supplementary File 4.2**).

4.2.2 Constraint-based modeling

FBA and several related approaches were employed in this manuscript. In the dynamic state, FBA seeks to maximize or minimize a metabolic function, such as biomass growth or ATP dissipation, subject to constraints on fluxes:

Minimize:
$$Z = \mathbf{-c}^T \mathbf{v}$$
Subject to: (1)
$$\mathbf{S} \cdot \mathbf{v} \mathbf{-b} = \frac{d\mathbf{X}}{dt},$$
$$v_j^{LB} \leq v_j \leq v_j^{UB}$$

where *c* is a vector of coefficients of length *n* identifying the objective reaction in the flux vector *v* of length *n*. *S* is the stoichiometric matrix of metabolites and reactions of dimension *m* x *n*, *b* is a vector of exchange fluxes of length *n*, and *X* is a vector of metabolite concentrations of length *m*. In the steady state, the problem is restated by:

$$\frac{d\mathbf{X}}{dt} = 0,$$
(2)

implying that there is no net accumulation or depletion of any metabolite pools. The optimization package Mosek (Mosek ApS, Denmark) was used to find the primal solution of the linear programming (LP) problem. Elemental flux sums (Φ_i) were calculated using the elemental matrix *E*, constructed for hydrogen, carbon, nitrogen, oxygen, phosphorus, and sulfur from metabolite molecular formulas:

$$\Phi_i = \frac{1}{2} \sum_j |E_{ij} v_j|$$

(3)

4.2.3 Shadow Prices

Sensitivity analyses were based on so-called Shadow Prices of the dual

solution to the LP problem according to:

$$\sum_{i=1}^{m} \lambda_i = \frac{dZ}{dX_i}$$

(4)

where dual variables $\boldsymbol{\lambda}$ of length m are assigned to steady state constraints, and

variables $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ are assigned to the flux constraints $\boldsymbol{v}^{LB}$ and $\boldsymbol{v}^{UB}$, respectively:

Minimize:
$$\text{-}\mathbf{q}_1 \mathbf{v}^{LB} - \mathbf{q}_2 \mathbf{v}^{UB}$$
Subject to:
$$\mathbf{c}^T = \boldsymbol{\lambda}^T \mathbf{S} + \mathbf{q}_1^T + \mathbf{q}_2^T,$$
$$\mathbf{q}_1 \leq 0, \mathbf{q}_2 \geq 0$$

(5)

4.2.4 Reporter metabolites and reporter subnetworks

A hypothesis-based method to identify key biological features around which

transcriptional changes occur was implemented to interpret the phosphorus stress

response, using the algorithms for reporter metabolites (54) and reporter

subnetworks (55). Both algorithms map the p-values and fold-changes from a

differential expression dataset (11) to the metabolic network using gene-protein-

reaction associations. The reporter metabolites algorithm ranks metabolite nodes

based on the normalized transcriptional response of its neighboring protein nodes

according to Z-scores assigned to each edge. The reporter subnetworks algorithm

expands on this concept by randomly sampling aggregates of reporter nodes, and again ranking each aggregate according to its Z-score.

4.2.5 Biomass objective function

A detailed biomass objective function (BOF) is essential to any high-quality GEM. *i*JC568 includes detailed biomass composition data collated from the MED4 literature under similar growth conditions (PRO99 medium, 14:10 hour light:dark cycle with peak intensities of 40-80 µmol photons m^{-2} s^{-1}), where available. Our BOF includes the protein amino acid composition, lipid profiles, pigment content, cell wall composition, carbohydrate content, DNA nucleotide fraction, RNA nucleotide fraction, and mineral and trace element composition, for a total of 121 compounds. However, detailed biochemical composition data are lacking for intracellular metabolite concentrations in MED4, and our BOF lacked information on free nucleotides, free amino acids, and the soluble pool (BioPool) concentrations, which were instead taken from the more completely characterized cyanobacterium Synechocystis sp. PCC6803. Cumulatively, these three pools make up less than 5% of ash-free dry weight (DW) and correspond to 4% of the variance of the growth rate under optimal growth, mostly (59%) due to spermidine and nicotinamide dinucleotides (**Supplementary File 4.2**). Growth associated maintenance (GAM) and non-growth associated maintenance (NGAM) ATP requirements were calculated according to the method described by Feist *et al.* (56). The sensitivity of growth rate to alterations in the biomass composition (Ψ_k) was evaluated by brute-force, analogous to the calculation of shadow prices:

$$\Psi_k = \frac{dZ}{d\boldsymbol{X}_k^{BIO}},$$
$$\boldsymbol{X}_k^{BIO} = \boldsymbol{S}_{BIO} + \boldsymbol{a}_k^T \boldsymbol{S}_{BIO} \tag{6}$$

where $\boldsymbol{X}_k{}^{BIO}$ is the biomass equation $\boldsymbol{S}_{BIO}$ with variable composition k. This is

accomplished by varying either a pool of biomass precursors, where BIO is the index

of the biomass reaction and $\boldsymbol{a}_k$ is a vector of ones with an element of variable

magnitude corresponding to a crude fraction (e.g., protein) or a specific compound

within that crude fraction (e.g., L-lysine). These targeted elements for each biomass

precursor pool or compound were varied by an arbitrarily small interval (-1 ppm $\leq$

$\Delta \boldsymbol{a}_k \leq$ 1 ppm) and FBA was then performed to quantify the resulting change in

growth rate ($\boldsymbol{\Delta Z}$).

An exhaustive search algorithm was implemented to quantify the change in growth

rate as a function of varying biomass precursor pool compositions that satisfied a

particular carbon:phosphorus molar ratio. In this way equation 1 is additionally

subject to:

$$e_m^{BIO} \in \{Q_{min,m}^{C:P} \leq Q_m^{C:P} \leq Q_{max,m}^{C:P}\} \tag{7}$$
$$Q_m^{C:P} = \frac{\sum_{l=1}^{L} \sum_{k=1}^{K} a_{kl} Q^C}{\sum_{l=1}^{L} \sum_{k=1}^{K} a_{kl} Q^P}$$

where Q^C is the number of carbon atoms and Q^P is the number of phosphorus atoms

of each compound (k) in each biomass precursor pool (l), and e_m^{BIO} refers to the mth

target C:P composition ($Q_m^{C:P}$), derived from the elemental matrix $\boldsymbol{E}$, within an

interval 10% below $Q_{min,m}^{C:P}$ and above $Q_{max,m}^{C:P}$. **Table 4.1** summarizes the BOF pool

composition and sensitivity (detailed composition and sensitivity under carbon,

light, phosphorus, and nitrogen limited growth conditions provided in **Supplementary File 4.2**).

4.2.6 Culture conditions and analytical procedures

Axenic *Prochlorococcus marinus* strain MED4 (courtesy of S.W. Chisholm) was grown in 30 ml batch in 70 ml borosilicate glass tubes in modified PRO99 low nutrient seawater-based media (31). P-limited growth was achieved after three transfers into 2 µM $H_2PO_4^-$ with a resulting N:P of 200. Cells were grown at 24ºC under cool white fluorescent light programmed to a parabolic 14 h : 10 h light : dark cycle reaching a peak irradiance of 45 µmol photons m^{-2} s^{-1}. Cell growth and contamination were monitored daily by flow cytometry (57). Cells were harvested by centrifugation (14,000 g) and the pellets resuspended in 100 µl 0.2 µm filtered seawater containing 0.2% paraformaldehyde. Aliquots of harvested cells were allowed to fix in the dark at 4ºC for 30 minutes prior to analysis of lipopolysaccharide by the *Limulus* amebocyte lysate spectrophotometric method (58).

4.3 Results

4.3.1 Reconstruction of iJC568 and its computing performance

We reconstructed the GEM for *Prochlorococcus marinus* str. MED4, termed *i*JC568, which consists of 568 metabolic genes encoding 794 reactions with 680 metabolites distributed among 6 sub-cellular locations (cytoplasmic membrane, periplasm, thylakoid membrane, thylakoid lumen, cytoplasm, and carboxysome;

Supplementary File 4.1). A summary of *i*JC568 network properties, and a comparison with the ensemble is given in **Table 4.2**.

To verify the *i*JC568 biomass objective function (BOF; see Methods) composition mass and energy budgets, elemental stoichiometry and standard enthalpies were calculated and compared with reported experimental data. Elemental stoichiometry of the BOF composition was within the error of reported values for carbon, nitrogen, and phosphorus ratios under balanced growth conditions (**Supplementary File 4.2** [14]). We calculated the heats of combustion (21) for each of the 121 compounds comprising the BOF (**Supplementary File 4.2**). By comparing these values to their energy cofactor demands (calculated as the sum of nucleotide triphosphate, nicotinamide dinucleotide, and flavin adenine nucleotide standard enthalpies), a slope of 29.5 KJ [mol ATP]$^{-1}$ was found, which is quite similar to the theoretical standard enthalpy of ATP hydrolysis (30 KJ [mol ATP]$^{-1}$ [22]). The resulting aggregate energy density of MED4 was 28 KJ g DW^{-1}, comparable with *Escherichia coli* (23 KJ g DW^{-1}) and *Saccharomyces cerevisiae* (21 KJ g DW^{-1} [23]).

We verified FBA results by comparing simulated growth rates, exchange fluxes, and internal fluxes with experimental data by simulating experimental conditions. Growth rates were compared with a fairly extensive set of culture experiments grown on a broad selection of defined media compositions and light profiles. The most commonly reported growth condition was a 14 h:10 h light:dark cycle at 20-24°C, reaching a peak irradiance ranging from 10 to 56 μmol photons m^{-2} s^{-1} blue light (24-27) in PRO99 medium. Zinser *et al.*, (26) provided the most comprehensive dataset relating carbon fixation rates and photophysiology

parameters to growth rates, with sampling intervals (2 hourly) most relevant to our instantaneous flux distributions. By simulating their growth conditions over a diel cycle, we calculated an optimal growth rate of 0.62 d^{-1} while experimental growth rates were 0.62 ± 0.04 d^{-1}. Short-term [^{14}C]-bicarbonate primary production measurements fell between model net and gross primary production for most of the light cycle (**Figure 4.1**). Further comparisons of *i*JC568 photosynthetic parameters (ATP/NADPH yields, quantum yields, photosynthetic quotient, optimal growth irradiance, net and gross primary production), growth yields, exchange fluxes (protons, CO^2, bicarbonate, nutrients), and central carbon metabolism metrics (phosphoglycerate kinase/phosphoglycerate mutase flux, anapleurotic CO^2 fixation) were in close agreement with those reported for strain MED4, where available, and *Synechocystis* str. PCC 6803 (**Supplementary File 4.2**). However, the tricarboxylic acid (TCA) cycle and photosynthetic electron flow pathways differed considerably from *Synechocystis* str. PCC 6803 and are discussed below.

4.3.2 Metabolic genes and their essentiality

The proportion of MED4 genes encoding metabolic reactions relative to the total number of genes (30%) is significantly higher than the GEM ensemble (19 ± 6%, p = 1e^{-7}), consistent with loss of many regulatory functions (28). Nearly all (99.3%) of the metabolic genes were expressed under conditions tested by Wang *et al.*, (29). The 4 non-expressed metabolic genes included, surprisingly, malate dehydrogenase (PMM1023), two lipid biosynthetic genes diacylglyceride kinase (PMM0183) and diacylglycerol phosphatidyltransferase (PMM0798), and

128

unsurprisingly, arsenate reductase (PMM0512). However, it should be noted that these 4 genes showed low, but detectable, expression levels in natural samples from the North Pacific Subtropical Gyre (30). We compiled essential metabolic genesets based on photolithoautotrophic growth on minimal medium (31) and on a supplemental medium (including the 39 carbon substrates, 34 nitrogen substrates, and 95 phosphorus substrates predicted to support growth if suitable transporters were present). Simulated single-gene knockouts were performed for each metabolic gene, and we required that enzyme complexes be complete for the corresponding reaction to proceed. The photolithoautotrophic essential geneset consists of 266 genes, or 47% of the metabolic genes. Although most (88%) of the metabolic genes in *i*JC568 belong to the 'core' of the *Prochlorococcus* pan-genome (compared with 65% of the whole genome), non-lethal genes were enriched (17%) in 'flexible' genes when compared with essential genes (8%). A similar pattern was seen for gene essentiality for mixotrophic growth in supplemental medium (**Supplementary File 4.2**), though a further 196 genes (34% of metabolic genes) produced lethal mutants only under specific conditions (termed 'variable-essential'). As in the photolithoautotrophic case, non-essential genes were more frequently part of the 'flexible' pan-genome (18%) than variable-essential genes (15%) or essential genes (8%). A genetic system remains elusive for *Prochlorococcus marinus*, so individual knockouts are yet to validate these results, however this is likely a conservative estimate since false negatives are likely using an *in silico* approach. Where available, the essential geneset of the ensemble (**Supplementary File 4.2**) ranged from 12% (*Pseudomonas putida* str. KT2440) to 38% (*Synechocystis* str. PCC6803) of metabolic

129

genes, reinforcing the adaptive gene loss hypothesis for *Prochlorococcus* (16).

Examples of bacteria with exceptionally high gene essentiality include the obligate

parasites *Mycoplasma genitalium* str. G37 (79% of whole genome [32]) and

Haemophilus influenzae str. Rd KW20 (47% of whole genome [33]).

4.3.3 The role of phosphate in MED4 metabolism

We examined the role of phosphate in MED4 by quantifying its connectivity,

dynamic coupling, and turnover within the *i*JC568 metabolic network. A

fundamental attribute of the stoichiometric matrix **S** (see Methods) is the

connectivity of the column and row space, defined here as the degree distribution of

the undirected bipartite graph. Metabolite participation (i.e., the number of

reactions associated with a particular metabolite) was assessed for *i*JC568 and the

ensemble by normalizing the degree distribution to the number of non- exchange

and transport reactions of each network. For example, the obligate anaerobes

Thermotoga maritima and *Methanosarcina barkeri* str. Fusaro had oxygen

metabolite participations near zero. Patterns of metabolite participation generally

clustered together according to taxonomic group (**Figure 4.2**), however *i*JC568

deviated from other cyanobacteria for orthophosphate, with the lowest

participation among all ensemble models.

The fixed matrix **S** is itself a transformation of the flux vector **v** to the vector

of concentration time derivatives d**X**/dt. Therefore, studying the topology of the

matrix **S** and its four fundamental subspaces (the row and null space comprising the

flux vector, and the column and left null space comprising the vector of

concentration time derivatives) is useful for interrogating network properties and for comparing the structural features and dynamic coupling of networks. Among the numerous factorization methods for analyzing the unconstrained solution space of the stoichiometric matrix, the most intuitive is the singular value decomposition (SVD), given by $S = U\Sigma V^T$. The m x r column-space (U) and r x n row-space (V) matrices contain the linearly independent orthonormal eigenvectors of the four fundamental subspaces of the m x n matrix S, and the singular values σ = diag(Σ) that define the set $\{\sigma^1, \ldots, \sigma^r\}$, where r is the rank of S, are measures of the distortion induced by each linear transformation. In this way, the first mode (σ^1) is the weight given to the first systems reaction, a linearly dependent set of reactions forming a basis pathway that is effectively a spanning set of S. For iJC568, and typical of other networks, the first systems reaction maps to a set of reactions involving proton translocation, due to the central role of the proton motive force in the electron transport chain and photosystems. The first systems reactions correspond quite closely to the metabolites with the highest flux sums (see Methods), with the exception of the photon flux, which participates in the 3rd and 6th modes in iJC568. In other organisms, subsequent modes vary in composition according to the metabolic capabilities of the organism (34). Further analysis of the systems reactions indicated that phosphate metabolism is a less dominant feature of iJC568. While the first three modes (ATP hydrolysis, NAD(P) redox, and the proton motive force) were similar for iJC568, phosphate transformations were associated with the 8th mode, compared with the 4th mode of *Escherichia coli*, *Haemophilus influenzae*, and *Helicobacter pylori* (34) or the 4th or 5th mode of each of the

131

ensemble models. Accordingly, the fractional singular value (e.g., the percent weight of a particular mode) associated with phosphate transformations was lower in *i*JC568 (0.9%) compared with other phyla (range 1.9% to 6.3%; **Supplementary Figure 4.1**).

The implications of a diminished role for phosphate in MED4 were investigated by comparing the elemental turnover of intermediate metabolism based on flux sums and quotas for hydrogen, carbon, nitrogen, oxygen, phosphorus, and sulfur with the high-quality reconstruction (*i*TO977 [35]) for *Saccharomyces cerevisiae*. Internal flux sums and turnover were normalized for the differences in optimal growth rate, transport flux of each element, and cellular elemental quotas between iTO977 and *i*JC568. Phosphorus turnover was approximately an order of magnitude higher than the other elements in both organisms, but *i*TO977 turnover was nearly 3-fold higher than *i*JC568 (**Supplementary Figure 4.2**). Surprisingly little of this turnover (40%) was due to ATP hydrolysis and ADP phosphorylation, with dissolved P_i demand for ATP synthase primarily recycled from the Calvin-Benson-Bassham (CBB) cycle reactions D-glyceraldehyde-3-phosphate:NAD+ oxidoreductase and sedoheptulose 1,7-bisphosphate 1-phosphohydrolase. The majority of the remaining 60% of the P turnover was shared between reactions with phosphorylated central carbon metabolites, nucleic acid intermediates, and dinucleotide energy carriers, implying that much of the difference in turnover is due to phosphate participation.

4.3.4 Physiological response to low phosphate

132

Prochlorococcus is known to have an extremely flexible elemental stoichiometry, perhaps a key to coping with variable supply of nutrients and extended periods of nutrient starvation typically encountered in the oligotrophic surface waters. Populations of *Prochlorococcus* in the periodically P-limited Sargasso Sea exhibited wide particulate C:P ratios (120:1 to 350:1), varying latitudinally (36). When grown in batch culture under P-limitation (molar $NH_4^+:H_2PO_4^- = 800:1$), MED4 particulate C:P increased to 464:1 ± 28:1, compared with 121:1 ± 17:1 under balanced growth (molar molar $NH_4^+:H_2PO_4^- = 16:1$ [14]). The partitioning of P in crude fractions of MED4 biomass, calculated by the elemental composition of the BOF, is predominantly bound in RNA (45%), DNA (23%), cell wall (di-trans-poly-cis-undecaprenyl diphosphate and lipid A disaccharide; 15%), and the soluble pool (especially inorganic P, nucleotides, folate cofactors, and several vitamins; 14%). The remaining P quota (2%) is found in lipids and in protein fractions. Since the discovery that P-limited MED4 and other *Prochlorococcus* strains have virtually eliminated phospholipids (2% of total lipid) in favor of sulfolipids and glycolipids (66% and 32% of total lipid, respectively [6]), the majority remains in cell wall and nucleotide fractions. DNA-P is static throughout the cell cycle phase G_1, and the whole proteome phosphorylation state is unlikely to vary significantly, so it follows that the ability to modulate C:P ratios to such extremes (~464:1) requires that all of these fractions must be capable of drastic reductions. Accounting for the 33% increase of the C-quota under P-limitation (14), the cumulative P quota in non-DNA pools (lipid, protein, RNA, cell wall, and the soluble pool) must be reduced by 85% to achieve a C:P of 464:1, and

the additional constraint of genome replication exacerbates this problem. MED4 must therefore regulate C:P ratios beyond those reached by lipid head-group substitution alone; such a reduction undoubtedly has profound impacts on cellular metabolism and physiology. An exhaustive search (see Methods) was implemented to quantify the growth rate advantage imparted on the PLG and CLG phenotypes by varying crude fractions of biomass to meet a range of feasible cellular C:P ratios (**Figure 4.3**). Over the allowable range of C:P ratios (120:1 to 528:1), the change in growth rates for the CLG phenotype (14 ± 7%) was identical (p = 0.71) to the coefficients of variation (CV) within any particular biomass composition (12 to 14%). In contrast, growth rates increased 370 ± 12% over the allowable C:P range for the PLG phenotype, with smaller compositional variations (CV = 2 to 9%; two-sample F-test $p \leq 1e^{-6}$). To identify which biomass components would yield the highest growth rate gains, we performed a brute-force sensitivity analysis (Ψ_k; see Methods; **Supplementary File 4.2**). Positive Ψ_k values imply an increase in growth rate resulting from a unit decrease in a particular biomass precursor pool k or an individual compound within a specified biomass precursor pool k. Among the crude biomass fractions, DNA, RNA, lipid, and cell wall were responsible for 96% of the growth rate sensitivity. Since DNA content is considered static in G_1, the highest growth sensitivity crude fractions were RNA (Ψ_k =0.45), cell wall (Ψ_k =0.15), and the soluble pool (Ψ_k =0.13). Within the cell wall crude fraction, di-trans-poly-cis-undecaprenyl diphosphate and Lipid A disaccharide were responsible for 74% and 26% of the sensitivity, respectively. Within the soluble pool, most (70%) of the sensitivity was due to nicotinamide dinucleotides.

Differential gene expression may be used to infer changes in metabolism using the reporter metabolites and reporter subnetwork algorithms (see Methods). By ranking Z-scores for each reporter metabolite, we identified a set of metabolites which were associated with up- or down-regulated proteins from differential expression (11) and changes in protein abundance (37) of phosphorus limited MED4 cultures. The top reporter subnetworks included lipopolysaccharide (LPS) synthesis, tRNA synthesis, cell wall synthesis, and a large subnetwork associated with the CBB cycle branches for carbon fixation, lower glycolysis, and the reductive PPP (**Supplementary Figure 4.3**). Similarly, shadow prices (λ; see Methods) can be used to infer the degree to which production of certain metabolites are limiting growth. By constraining a sub-optimal upper bound on the P_i transport rate and calculating the shadow prices, a set of 28 metabolites were negative, or limiting growth. These include phosphate esters, nucleotides, cell wall precursors, NADH, and the phosphorylated central carbon metabolites 3-phosphoglycerate and 2-phosphoglycolate. A comparison of reporter metabolites and shadow prices showed agreement between Z-scores and the relative magnitude $|-\lambda|$ (**Supplementary File 4.2**). These qualitative (gene expression) and quantitative (shadow prices) predictions are complementary but independent methods and were partially validated in our laboratory comparison of LPS in culture (see Methods), where P-limited cells showed a 55 ± 4 % reduction compared with P-replete cells.

In culture, P-stress induced changes not only in acquisition mechanisms and biosynthetic pathways but also in central carbon metabolic pathways and the photosynthetic apparatus. Following initiation of P-stress, photosystem II (PSII) was subsequently degraded with a concomitant decrease in carbon fixation proteins, though photosystem I (PSI) and subunits of the ATP synthase complex remained intact (37). This finding was also apparent for the *in silico* PLG phenotype in *i*JC568, with an increase in the PSI/PSII photon absorption ratio at I_{max} as a non-linear function of the P_i transporter flux, converging on a new steady state for ATP and reductant for the PLG phenotype. We compared growth and key photosynthetic fluxes between the LLG and PLG conditions by phenotype phase plane (PhPP) analysis varying light and P_i uptake rates (**Figure 4.4** and **Supplementary Figure 4.4**). In *i*JC568, the linear electron flow (LEF) pathway begins with PSII, cytochrome b_6f (Cyt b_6f), PSI, and ends with ferredoxin-NADP$^+$ reductase (FdR). LEF is linked by the plastocyanin ($Cu^{2+}PC/Cu^+PC$), ferredoxin (Fd_{ox}/Fd_{red}), and plastoquinone (PQ/PQH_2) pools. A set of alternative electron flow (AEF) pathays include cyclic electron flow (CEF) around PSI via the NADPH dehydrogenase type 1 (NDH) or via ferredoxin:quinone oxidoreductase (FQR), pseudocyclic electron flow (PCEF) around PSII via cytochrome oxidase bd (COX; MED4 apparently lacks the aa3-type cytochrome c oxidase), the Mehler reaction, and photorespiration. The activity of AEF pathways affect a number of fundamental fluxes, including the ATP/NADPH ratio, photosynthetic efficiency, quantum yield, and the photosynthetic quotient. Under optimal growth conditions (along the line of optimality; L.O.), the ratio of PSI to PSII absorption was 2.3, with the entirety of the PSI flux split between CEF

136

around PSI via NDH to prevent over-reduction of the PQ pool, and to NADPH via FdR

to maintain the optimal ATP/NADPH ratio of 1.30. Under LLG conditions (above the

L.O.), the ratio of PSI to PSII absorption increased to 2.4, with the PSI flux mostly

diverted to $NADP^+$, at the expense of CEF via NDH. Under PLG conditions (below the

L.O.), ATP deficits resulted in a PSI to PSII absorption ratio of 0.8, with excess

reductant diverted to PCEF around PSII via COX and LEF to NADPH from PSI. CEF

around PSI was diverted to FQR from NDH under PLG conditions.

It is plausible that persistent changes to the optimal path of electron flow

under P-limited conditions have resulted in a restructuring of the MED4

photosynthetic apparatus. MED4 lacks the genes encoding succinate dehydrogenase

(SDH), which catalyzes the succinate-fumarate couple and directly links the

oxidation of TCA cycle intermediates to the reduction of the PQ pool in the

photosynthetic electron chain. Succinate oxidation forms an AEF pathway which

supplies reductant to $Cu^{2+}PC$ at the cost of PQH_2 via Cyt b_6f. Furthermore, all

cyanobacteria have a branched TCA cycle, lacking 2-oxoglutarate dehydrogenase,

however *Prochlorococcus marinus* and marine *Synechococcus* spp. also lack the

recently-discovered analogous enzymes 2-oxoglutarate decarboxylase (2OGDC) and

succinic semialdehyde dehydrogenase (SSADH), which are necessary to regenerate

succinate from 2-oxoglutarate (38) and have long been implicated in obligate

photolithoautotrophy (39). To quantify the effect of SDH and the branched TCA

cycle on the energy budget of *i*JC568, we compared the NAD(P)H costs of four *in*

silico variants grown photolithoautotrophically: wild type (WT), +2OGDC+SSADH,

+SDH, and +2OGDC+SSADH+SDH. NAD(P)H costs were calculated as the change in

flux sums for each variant after forcing a net accumulation of succinate ($dX_i/dt > 0$; where i corresponds to succinate). For both the WT and the +2OGDC+SSADH mutant, the NAD(P)H cost of succinate synthesis was primarily due to the CEF enzyme NDH and the Calvin-Benson-Bassham pathway enzyme triosephosphate dehydrogenase. For both the +SDH and +2OGDC+SSADH+SDH mutants, electron flow largely bypasses NDH, reducing the cost of succinate synthesis. However, with the +SDH mutant (lacking the cyanobacterial-type TCA cycle), additional NAD(P)H costs were required for precursor sythesis via malate dehydrogenase. In summary, and in detail in **Supplementary File 4.2**, the NAD(P)H cost of regenerating succinate *de novo* for mutants with the cyanobacterial TCA cycle increases from 14 to 16 molecules NAD(P)H in the absence of SDH, while the reverse is true for MED4 WT, which decreases from 21 to 14 molecules NAD(P)H in the absence of SDH. These findings are qualitatively in agreement with wild-type and –SDH, –2OGDC, and –SSADH mutants of *Synechococcus* sp. PCC7002 (38). As a consequence of the MED4 (WT) TCA pathway, under PLG conditions the absence of SDH reduces the PSI flux by 30%, resulting in a 16% to 37% increase in quantum yield (mol CO_2 reduced [mol photons]$^{-1}$) over the range of PLG phenotypes.

4.4 Discussion

We explored metabolic strategies of the *Prochlorococcus* PLG phenotype in an effort to characterize its adaptation to low phosphate marine environments. In addition to sulfolipid substitution and utilization of organophosphorus to satisfy P-demand, we uncovered three additional strategies that reduce the influence of phosphorus control over optimal growth flux topology: (a) choreographed reductions in enzymes reliant on orthophosphate as a substrate across the entire metabolic network, (b) reductions in phosphorus-rich biomass constituents, and (c) alterations to photosynthetic and respiratory electron flow. Phosphorus, the 'staff of life' (40), plays myriad roles in the structural, regulatory, and energetic functions of all cellular life. The phosphate residue provides hydrophilicity and increased water solubility of the parent chain, provides a charge to prevent membrane permeation, and provides a nucleophile repellent to resist hydrolysis. Phosphoester bonds modulate post-translational regulation of protein function and phosphodiester bonds form the connective tissue of the DNA and RNA backbone. The phosphoanhydride bonds of nucleotide polyphosphates and polyphosphates contain the ultimate source of chemical energy required for all metabolism and, once hydrolyzed, the free monomeric metaphosphate ion is a strong electrophile and phosphorylating agent, capable of phosphorylating even aromatic amine rings, attacking carbonyl groups and ketones to their enol phosphates. It is, therefore, unsurprising that hydrolysis, esterification, and isomerization of phosphorylated metabolites are ubiquitous functions in metabolic networks. Since typical intermediate metabolite pool concentrations meet or exceed the associated enzyme

139

half-saturation constant for growth in rich media (41), a phosphorus limited

metabolic state would, presumably, influence substrate binding kinetics widely. At

the level of pathways, phosphate availability would distribute rate limitation

broadly, in keeping with metabolic control analysis (42), which finds a non-zero

elasticity coefficient for almost all enzymes, resulting in flux control shared among

all participating reactions in a pathway.

A strategy to alleviate metabolic control of a single, persistent rate limiting

co-substrate might be to selectively reduce its degree ('participation') in the

network. In a comparison of *i*JC568 and the ensemble, phosphate participation was

lowest for *i*JC568. Reductions in phosphate participation results in a greatly

diminished "role" (fractional contribution of the phosphate transformation systems

reaction to the singular value spectrum) for phosphate from a network perspective,

suggesting that low phosphate availability may have guided gene loss during the

evolution of MED4, streamlining the loss of genes associated with phosphate

transformations. Low phosphate reaction participation in *i*JC568 contributed to

decreased total elemental phosphorus fluxes in a comparison with *i*TO977, resulting

in slower turnover and possibly lowering soluble phosphate concentration

requirements to maintain optimal fluxes throughout the network. This prediction is

supported by experimental determinations of intracellular phosphate

concentrations and their responses to phosphate availability (**Supplementary File

4.2**). Under P-replete conditions, intracellular phosphate concentrations were 7-fold

lower in MED4 than in another oligotrophic marine picocyanobacterium

Synechococcus WH7803 (43). Importantly, when grown in P-deplete medium, MED4

intracellular phosphate decreased by only 22 ± 3% compared with 69 ± 4% for

Synechococcus WH7803. A similar contrast might be made between MED4 and

Saccharomyces cerevisiae (44), although the experimental conditions could not be

directly compared. To our knowledge, this is the first example of nutrient control of

metabolic network evolution.

Beyond the inherent architecture of the metabolic network, MED4 is known

to respond physiologically to low phosphate availability by widely altering its

elemental C:P ratio. The range of the flexible elemental stoichiometry of MED4 in

culture and in the field presents a problem if carbon quotas increase by only 20%

and C:P ratios approach the theoretical limit of 528:1 under severe phosphorus

limitation (C:P = 464:1): how can genome replication be feasible when a fully

replicated genome alone translates to a C:P of 264:1? Based on comparison of the

phosphatidylglycerol content of P-replete and P-limited cultures (45), phospholipid

for sulfo- and glycolipid head group substitution accounts for 2% of the required P

quota reduction, so we sought *in silico* methods to identify where the additional

98% of P quota reductions were to be found. Sensitivity analysis of the BOF

composition suggested that reductions in RNA and the cell wall were likely

candidates. The selective reduction in phospholipid, cell wall and RNA synthesis was

also observed by the method of reporter subNetworks from differential gene

expression of balanced versus phosphate limited growth conditions. Our exhaustive

search algorithm predicted a set of feasible biomass compositions for a range of C:P

ratios, with optimal growth corresponding to a phosphorus composition of DNA :

RNA: Lipid: BioPool: CellWall of 1: 2.8±0.5: 4.7±0.9: 1.6±0.4: 3.0±0.5. This optimal

biomass composition was partially validated in our culture experiments with a shift in the DNA : CellWall ratio from 1.51 under CLG to 2.75 under PLG conditions, assuming LPS remains proportional to cell wall content. It is unclear what physiological effects such a dramatic reduction in cell wall might have; even under rich medium growth MED4 has a reduced cell wall thickness (19 nm), compared with 34 nm thickness in a strain isolated from deeper in the euphotic zone where phosphate limitation is less prevalent (46).

Consistent with the theme of low phosphate guided gene loss, the conspicuous absence of SDH in MED4 and other *e*HL-I strains prompted us to investigate the role of this otherwise ubiquitous enzyme under a variety of growth conditions. The reversible succinate-fumarate couple and its catalyst, SDH, are found in all three domains of life, the last universal ancestor, and were probably conserved throughout organismal evolution (47). Additionally, SDH represents a unique connection between the TCA cycle and respiratory and photosynthetic electron flow, and it is thus under considerable evolutionary pressure that *e*HL-I shed SDH. The loss of SDH creates an unexpected link between photosynthetic quantum yield and phosphorus limited growth, and it is at least suggestive that the gene coding for the A subunit of SDH, *sdhA*, is the one annotated gene that differentiates the high light ecotypes eHLI, which dominates the P-limited Mediterranean and eHLII, which dominates the predominantly N-limited Atlantic and Pacific.

4.4.1 Concluding remarks

Nearly three decades after its isolation, MED4 has very likely undergone considerable laboratory evolution, however its metabolic potential remains imprinted with the signature of the chronically phosphorus depleted surface waters of the Eastern Mediterranean Sea. Metabolic network reconstruction and constraint-based modeling revealed previously unknown evolutionary strategies for organisms perpetually coping with low phosphate availability. These strategies include a re-design of the metabolic network to alleviate metabolic control of a single substrate, global control of phosphorus partitioning in biomass components, and optimization of photosynthetic electron flow.

4.5 Acknowledgements

4.6 References

1. Liu, H., Nolla, H. A., and Campbell, L. 1997. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. Aquatic Microbial Ecology 12:39–47. http://dx.doi.org/10.3354/ame012039

2. Partensky, F. and Garczarek, L. 2010. *Prochlorococcus*: advantages and limits of minimalism. Annual Review of Marine Science 2:305–331. http://dx.doi.org/10.1146/annurev-marine-120308-081034

3. Moore, L., Ostrowski, M., Scanlan, D., Feren, K., and Sweetsir, T. 2005. Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. Aquatic Microbial Ecology 39:257–269. http://dx.doi.org/10.3354/ame039257

4. Casey, J., Lomas, M. W., Michelou, V., Dyhrman, S., Orchard, E., Ammerman, J. W., and Sylvan, J. 2009. Phytoplankton taxon-specific orthophosphate (Pi) and ATP utilization in the western subtropical North Atlantic. Aquatic Microbial Ecology 58:31–44. http://dx.doi.org/10.3354/ame01348

5. Bjorkman, K., Duhamel, S., and Karl, D. M. 2012. Microbial group specific uptake kinetics of inorganic phosphate and adenosine-5-triphosphate (ATP) in the North Pacific Subtropical Gyre. Frontiers in Microbiology 3:1–17. http://dx.doi.org/10.3389/fmicb.2012.00189

6. Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T., and Devol, A. H. 2006. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. Proceedings of the

National Academy of Sciences of the United States of America 103:8607–

8612. http://dx.doi.org/10.1073/pnas.0600540103

7. Gilbert, J. D. J. and Fagan, W. F. 2010. Contrasting mechanisms of proteomic

    nitrogen thrift in *Prochlorococcus.* Molecular Ecology 20:92–104.

    http://dx.doi.org/10.1111/j.1365-294X.2010.04914.x

8. Partensky, F., Hoepffner, N., Li, W., Ulloa, O., and Vaulot, D. 1993.

    Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated

    from the North Atlantic and the Mediterranean Sea. Plant Physiology

    101:285–296.

9. Thingstad, T. F., Krom, M. D., Mantoura, R., Flaten, G., Groom, S., Herut, B.,

    Kress, N., Law, C. S., Pasternak, A., Pitta, P., Psarra, S., Rassoulzadegan, F.,

    Tanaka, T., Tselepides, A., Wassmann, P., Woodward, E. M. S., Wexels Riser, C.,

    Zodiatis, G., and Zohary, T. 2005. Nature of phosphorus limitation in the

    ultraoligotrophic eastern Mediterranean. Science 309: 1068–1071.

    http://dx.doi.org/10.1126/science.1112632

10. Garczarek, L., Dufresne, A., Rousvoal, S., West, N. J., Mazard, S., Marie, D.,

    Claustre, H., Raimbault, P., Post, A. F., Scanlan, D. J., and Partensky, F. 2007.

    High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the

    Mediterranean Sea in summer. FEMS Microbiology Ecology 60:189–206.

    http://dx.doi.org/10.1111/j.1574-6941.2007.00297.x

11. Martiny, A. C., Coleman, M. L., and Chisholm, S. W. 2006. Phosphate

    acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide

    adaptation. Proceedings of the National Academy of Science of the United

States of America 103:12552–12557.

http://dx.doi.org/10.1073/pnas.0601301103

12. Reistetter, E. N., Krumhardt, K., Callnan, K., Roache-Johnson, K., Saunders, J. K., Moore, L. R., and Rocap, G. 2013. Effects of phosphorus starvation versus limitation on the marine cyanobacterium *Prochlorococcus* MED4 II: gene expression. Environmental Microbiology 15: 2129–2143.

http://dx.doi.org/10.1111/1462-2920.12129

13. Parpais, J., Marie, D., Partensky, F., Morin, P., and Vaulot, D. 1996. Effect of phosphorus starvation on the cell cycle of the photosynthetic prokaryote *Prochlorococcus* spp. Marine Ecology Progress Series 132:265–274.

http://dx.doi.org/10.3354/meps132265

14. Bertilsson, S., Berglund, O., Karl, D. M., and Chisholm, S. W. 2003. Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. Limnology and Oceanography 48:1721–1731. 10.4319/lo.2003.48.5.1721

15. Krumhardt, K. M., Callnan, K., Roache-Johnson, K., Swett, T., Robinson, D., Reistetter, E. N., Saunders, J. K., Rocap, G., and Moore, L. R. 2013. Effects of phosphorus starvation versus limitation on the marine cyanobacterium *Prochlorococcus* MED4 I: uptake physiology. Environmental Microbiology 15:2114–2128. http://dx.doi.org/10.1111/1462-2920.12079

16. Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., and Johnson, J. 2007. Patterns and

implications of gene gain and loss in the evolution of *Prochlorococcus.* PLoS Genetics 3:e231. http://dx.doi.org/10.1371/journal.pgen.0030231

17. Palsson, B. Ø. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge University Press. 2015.

18. Nielsen, J. 2003. It is all about metabolic fluxes. Journal of Bacteriology 185:7031–7035. http://dx.doi.org/10.1128/JB.185.24.7031-7035.2003

19. Thiele, I. and Palsson, B. Ø. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5:93–121. http://dx.doi.org/10.1038/nprot.2009.203

20. Mardinoglu, A. and Nielsen, J. 2015. New paradigms for metabolic modeling of human cells. Current Opinion in Biotechnology 34:91–97. http://dx.doi.org/10.1016/j.copbio.2014.12.013

21. Kharasch, M. S. 1929. Heats of combustion of organic compounds. Bureau of Standards Journal of Research 2:359–430.

22. Nelson, D. and Cox, M. Lehninger Principles of Biochemistry. W.H. Freeman and Company. 2005.

23. Cordier, J., Butch, B., Birou, B., and von Stockar, U. 1987. The relationship between elemental composition and heat of combustion of microbial biomass. Applied Microbiology and Biotechnology 25:305–312. http://dx.doi.org/10.1007/BF00252538

24. Moore, L. and Goericke, R. 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments,

fluorescence and absorptive properties. Marine Ecology Progress Series 116:259–275. http://dx.doi.org/10.3354/meps116259

25. Bruyant, F., Babin, M., Sciandra, A., Marie, D., Genty, B., Claustre, H., Blanchot, J., Bricaud, A., Rippka, R., Boulben, S., Louis, F., and Partensky, F. 2001. An axenic cyclostat of *Prochlorococcus* PCC 9511 with a simulator of natural light regimes. Journal of Applied Phycology 13:135–142. http://dx.doi.org/10.1023/A:1011144310988

26. Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., Wright, M. A., Rector, T., Steen, R., Mcnulty, N., Thompson, L. R., and Chisholm, S. W. 2009. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. PLoS ONE 4:e5135. http://dx.doi.org/10.1371/journal.pone.0005135

27. Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. 2012. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. PLoS ONE 7:e43432. http://dx.doi.org/10.1371/journal.pone.0043432

28. Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W., and Hess, W. R. 2008. The challenge of regulation in a minimal photoautotroph: Non-coding RNAs in *Prochlorococcus*. PLoS Genetics 4:e1000173. http://dx.doi.org/10.1371/journal.pgen.1000173

29. Wang, B., Lu, L., Lv, H., Jiang, H., Qu, G., Tian, C., and Ma, Y. 2014. The transcriptome landscape of *Prochlorococcus* MED4 and the factors for

stabilizing the core genome. BMC Microbiology 14:11.

http://dx.doi.org/10.1186/1471-2180-14-11

30. Ottesen, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin, R., Schuster, S. C., Scholin, C. A., and DeLong, E. F. 2014. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science 345:207–212. http://dx.doi.org/10.1126/science.1252476

31. Moore, L., Coe, A., Zinser, E., Saito, M., Sullivan, M. B., Lindell, D., Frois-Moniz, K., Waterbury, J., and Chisholm, S. W. 2007. Culturing the marine cyanobacterium *Prochlorococcus*. Limnology and Oceanography: Methods 5:353–362. http://dx.doi.org/10.4319/lom.2007.5.353

32. Akerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N., and Mekalanos, J. J. 2002. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proceedings of the National Academy of Sciences of the United States of America 99:966–971. http://dx.doi.org/10.1073/pnas.012602299

33. Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Hutchinson, C. A., III, Smith, H. O., and Venter, J. C. 2006. Essential genes of a minimal bacterium. Proceedings of the National Academy of Sciences of the United States of America 103:425–430. http://dx.doi.org/10.1073/pnas.0510013103

34. Famili, I. and Palsson, B. Ø. 2003. Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices.

Journal of Theoretical Biology 224: 87–96.

http://dx.doi.org/10.1016/S0022-5193(03)00146-2

35. Österlund, T., Nookaew, I., Bordel, S., and Nielsen, J. 2013. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Systems Biology 7: 36. http://dx.doi.org/10.1186/1752-0509-7-36

36. Martiny, A. C., Pham, C. T. A., Primeau, F. W., Vrugt, J. A., Moore, J. K., Levin, S. A., and Lomas, M. W. 2013. Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. Nature Geoscience 6:279–283. http://dx.doi.org/10.1038/ngeo1757

37. Fuszard, M. A., Wright, P. C., and Biggs, C. A. 2010. Cellular acclimation strategies of a minimal picocyanobacterium to phosphate stress. FEMS Microbiology Letters 306:127–134. http://dx.doi.org/10.1111/j.1574-6968.2010.01942.x

38. Zhang, S. and Bryant, D. A. 2011. The tricarboxylic acid cycle in cyanobacteria. Science 334: 1551–1553. http://dx.doi.org/10.1126/science.1210858

39. Wood, A. P., Aurikko, J. P., and Kelly, D. P. 2004. A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy? FEMS Microbiology Reviews 28:335–352. http://dx.doi.org/10.1016/j.femsre.2003.12.001

40. Karl, D.M. 2000. Phosphorus, the staff of life. Nature 406:31–33. http://dx.doi.org/10.1038/35017686

41. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nature Methods 5:593–599. http://dx.doi.org/10.1038/nchembio.186

42. Hofmeyr, J. H. S. and Cornish-Bowden, A. 1991. Quantitative assessment of regulation in metabolic systems. European Journal of Biochemistry 200:223–236. http://dx.doi.org/10.1111/j.1432-1033.1991.tb21071.x

43. Mine, A., Coleman, M., and Colman, A. Enzyme mediated nutrient regeneration following lysis of *Synechococcus* WH7803. In Ocean Sciences Meeting, New Orleans, LA. 2016.

44. Auesukaree, C., Homma, T., Tochio, H., Shirakawa, M., Kaneko, Y., and Harashima, S. 2004. Intracellular phosphate serves as a signal for the regulation of the PHO pathway in *Saccharomyces cerevisiae*. Journal of Biological Chemistry 279:17289–17294. 10.1074/jbc.M312202200

45. Van Mooy, B., Fredricks, H. F., Pedler, B. E., and Dyhrman, S. T. 2009. Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature 458:69–72. http://dx.doi.org/10.1038/nature07659

46. Ting, C. S., Hsieh, C., Sundararaman, S., Mannella, C., and Marko, M. 2007. Cryo-electron tomography reveals the comparative three-dimensional architecture of *Prochlorococcus*, a globally important marine cyanobacterium. Journal of Bacteriology 189:4485–4493. http://dx.doi.org/10.1128/JB.01948-06

47. Wächtershäuser, G. 1990. Evolution of the first metabolic cycles. Proceedings of the National Academy of Sciences of the United States of America 87:200–204.

48. Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., and Nielsen, J. 2013. The RAVEN Toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. PLoS Computational Biology 9:e1002980. http://dx.doi.org/10.1371/journal.pcbi.1002980

49. Jankowski, M. D., Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. Biophysical Journal 95:1487–1499. http://dx.doi.org/10.1529/biophysj.107.124784

50. Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., and Brinkman, F. S. L. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. http://dx.doi.org/10.1093/bioinformatics/btq249

51. Cserzo, M., Eisenhaber, F., Eisenhaber, B., and Simon, I. 2003. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. Bioinformatics 20:136–137. http://dx.doi.org/10.1093/bioinformatics/btg394

52. Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., and Thiele, I. 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proceedings of the National Academy of Sciences of the

United States of America 109:2678–2683.

http://dx.doi.org/10.1073/pnas.1117907109

53. Partensky, F., Hoepffner, N., Li, W., Ulloa, O., and Vaulot, D. 1993. Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and the Mediterranean Sea. Plant Physiology 101:285–296. http://dx.doi.org/10.1104/pp.101.1.285

54. Patil, K. R. and Nielsen, J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proceedings of the National Academy of Sciences of the United States of America 102:2685–2689. http://dx.doi.org/10.1073/pnas.0406811102

55. Oliveira, A., Patil, K., and Nielsen, J. 2008. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. BMC Systems Biology 2: 17. http://dx.doi.org/10.1186/1752-0509-2-17

56. Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. Ø. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology 3:121. http://dx.doi.org/10.1038/msb4100155

57. Casey, J., Aucan, J. P., Goldberg, S. R., and Lomas, M. W. 2013. Changes in partitioning of carbon amongst photosynthetic pico- and nano-plankton groups in the Sargasso Sea in response to changes in the North Atlantic

Oscillation. Deep Sea Research Part II: Topical Studies in Oceanography

93:58–70. http://dx.doi.org/10.1016/j.dsr2.2013.02.002

58. Watson, S. W., Novitsky, T. J., Quinby, H. L., and Valois, F. W. 1977.

Determination of bacterial number and biomass in the marine environment.

Applied and Environmental Microbiology 33:940–946.

Table 4.1 - Crude biomass composition and growth sensitivity (Ψ) of *i*JC568. DW – ash free dry weight.

Component	Composition (% of Total DW)	Ψ (% of Total)
DNA	1.2	<1
RNA	4.7	2
Protein	58.1	41
Lipid	11.5	35
Pigments	3.8	5
Cell Wall	5.0	5
Carbohydrate	2.9	7
Free nucleic acids	<0.1	<1
Free amino acids	2.1	<1
BioPool	2.9	3
Mineral and Trace metals	2.4	<1

Table 4.2 - Summary of *i*JC568 properties. Complexed - subunit encoding genes; Blocked - reactions associated with dead-end metabolites; Orphaned - reactions not connected to the network; Gap filled - metabolic reactions with no annotated gene; Transport - including diffusive reactions and porins; Exchange - boundary transport used for modeling.

Genes		568	
	Complexed	302	53%
Reactions		794	
	Blocked	23	3%
	Orphaned	3	<1%
	Gap filled	60	8%
	Reversible	329	42%
	Transport	63	8%
	Exchange	79	10%
Metabolites			
	Unique	597	88%

Figure 4.1 - Diel simulation. Comparison of calculated net and gross primary production against short-term [14C]-bicarbonate primary production measurements reported by (26). The light profile followed a gradual increase from darkness to a peak irradiance of 232 μmol photons m^{-2} s^{-1} which was held constant for four hours, followed by a gradual decrease to darkness.
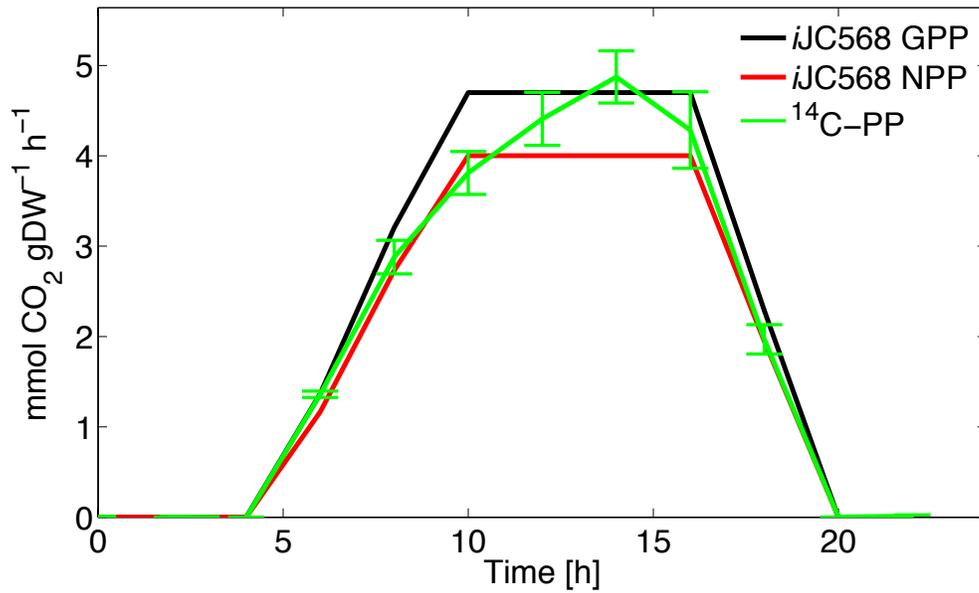
Figure 4.2 - Metabolite participation. Comparison of the 10 highest degree metabolites (excluding H_2O and H^+) between the Ensemble, grouped by phylum, and *i*JC568 (black square markers). Abbreviations: Pi - orthophosphate; PPi - diphosphate; AMP - adenosine monophosphate; L-Glut - L-glutamate.
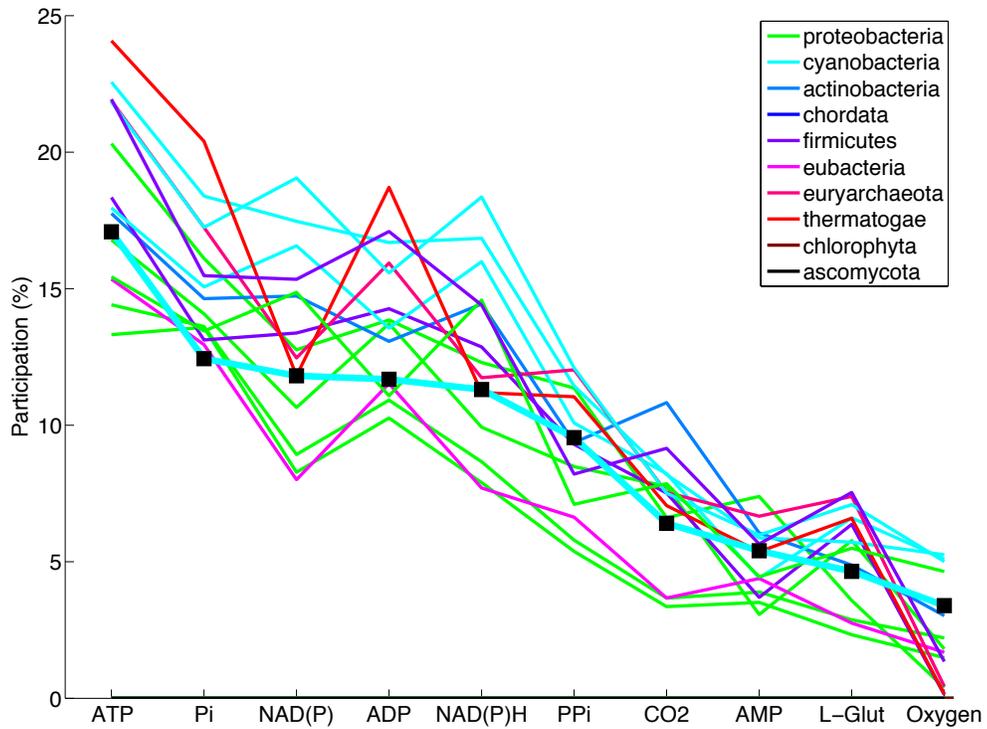
Figure 4.3 - Simulated growth rate as a function of altered biomass compositions. Values represent the calculated growth rates associated with a composition of DNA, RNA, lipid, cell wall, and soluble pool which correspond to each interval of the C:P ratio range. Growth rates were compared by constraining the orthophosphate transporter flux (red) or the carbon fixation flux (black) to sub-optimal rates. The number of biomass compositions at each C:P ratio is indicated (n).
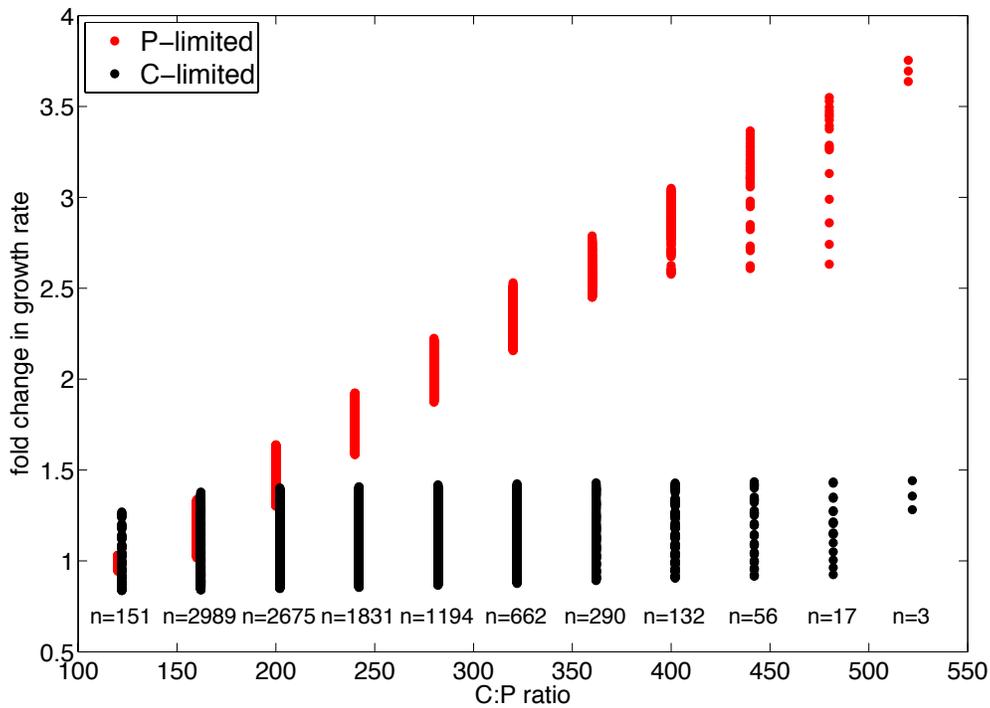
Figure 4.4 - Phenotype phase-planes of light and phosphate uptake for key photosynthetic fluxes. In each panel, the white line of optimality (L.O.) indicates optimal growth and delineates LLG (above) and PLG (below) phenotypes. Abbreviations: FdR - ferredoxin-NADP$^+$ reductase; FqR - ferredoxin:quinone oxidoreductase; COX - cytochrome oxidase bd; Cyt$_{b6f}$ - cytochrome b$_6$f; NDH - NADPH dehydrogenase type 1.
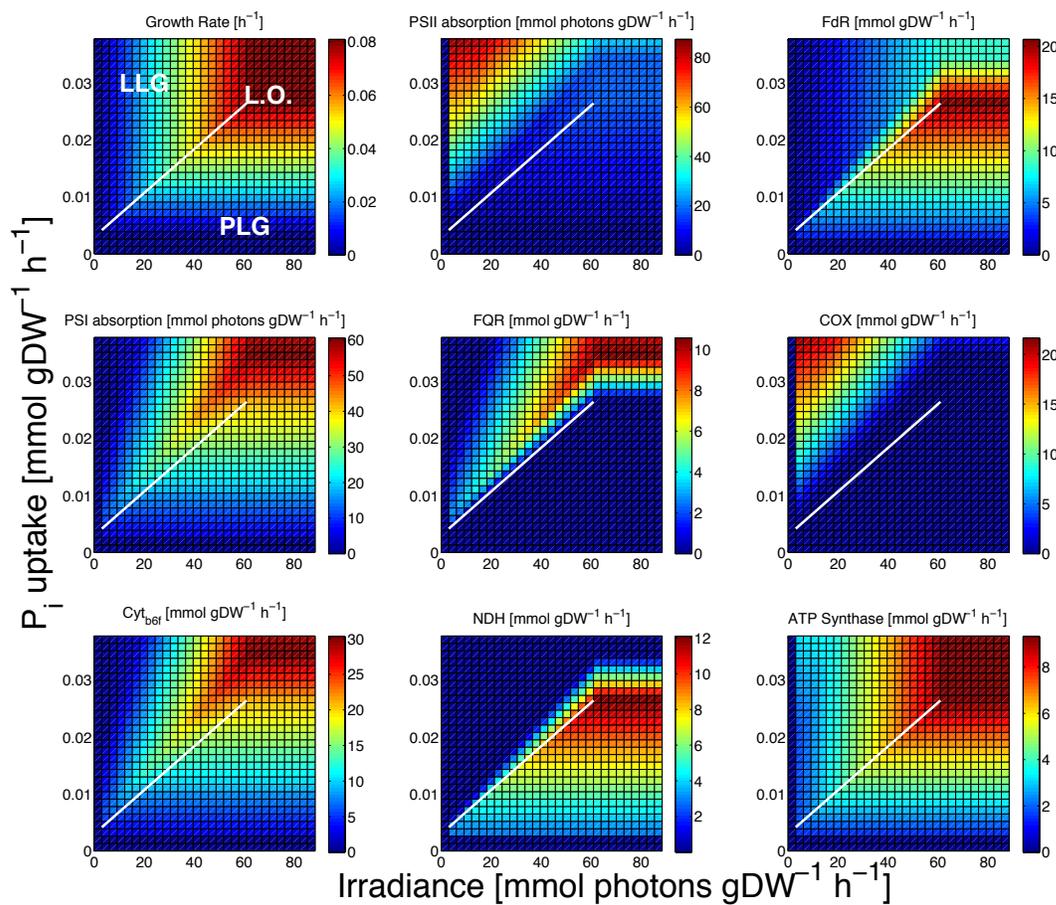
4.7 Supplementary

Supplementary Figure 4.1 - Fractional singular values of the phosphate

transformation systems reaction for *i*JC568 and the ensemble. References for each

ensemble model can be found in **Supplementary File 4.2**. Figure S1.eps
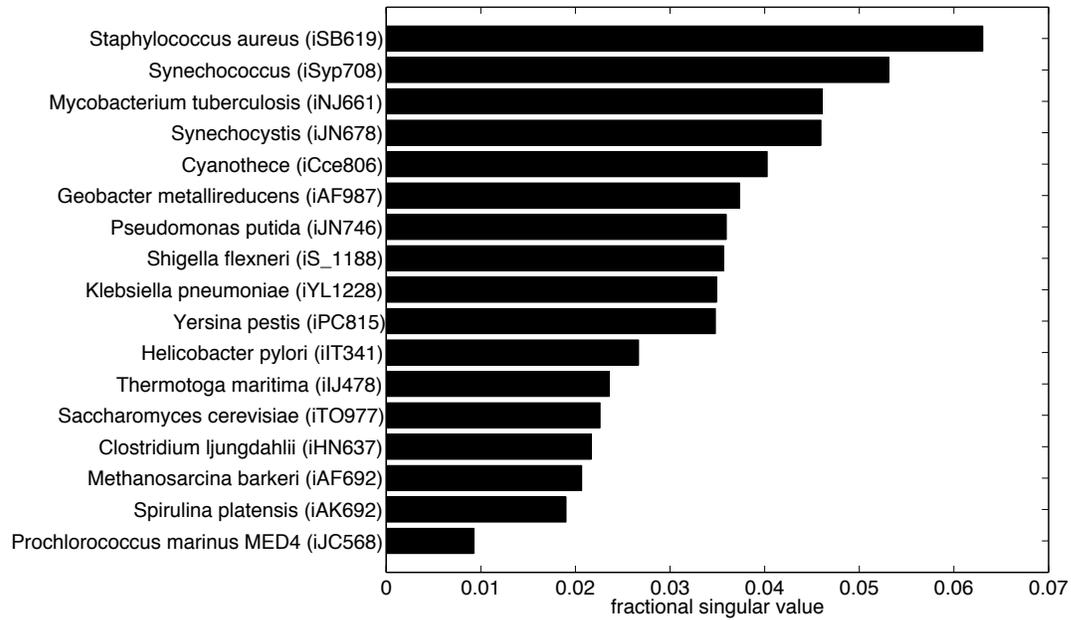
Supplementary Figure 4.2 - Elemental flux sums and turnover comparison of

*i*JC568 and *i*TO977. Figure S2.eps
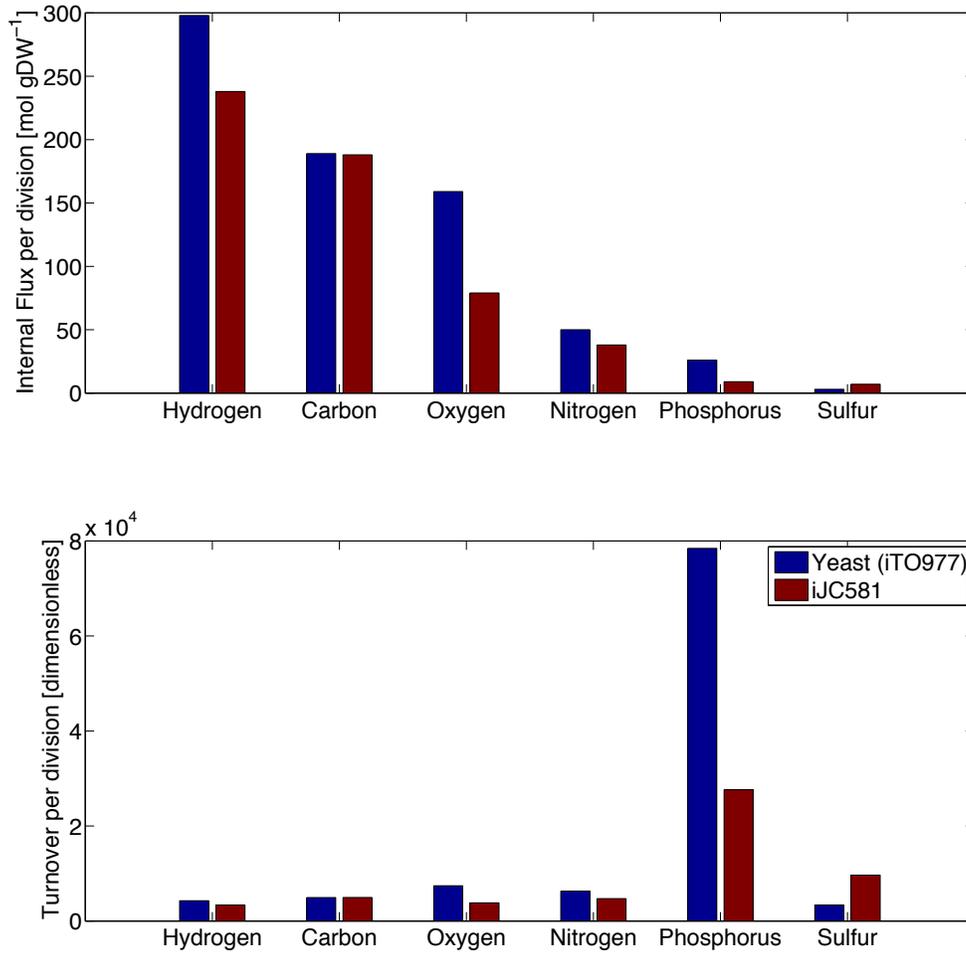
Supplementary Figure 4.3 - Reporter subNetworks identified from differential

expression in P-replete and P-limited media. Data from (11). <u>Figure S3.eps</u>
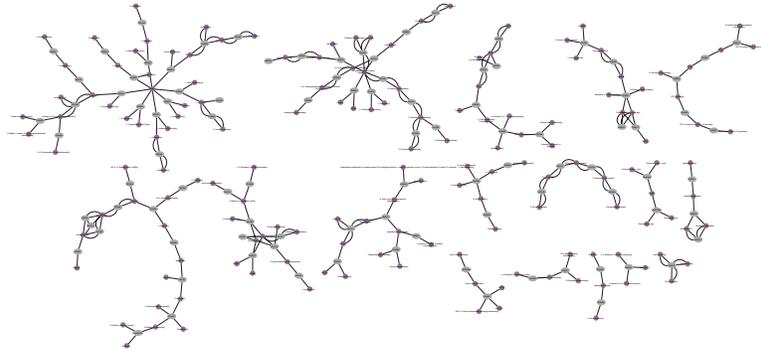
Supplementary Figure 4.4 - Illustration of changes to photosynthetic electron flow under optimal growth conditions, light-limited growth conditions, phosphorus limited growth conditions, and phosphorus limited growth conditions for the *in silico* SDH knock-in mutant. The center panel is a detailed view of the *i*JC568 photosystem, including the transport of protons across the thylakoid membrane (orange text), cofactors associated with each reaction (black arrows and numbers), and the stoichiometry of metabolites associated with each reaction (blue arrows and numbers). Reactions belonging to the LEF (pink), CEF (yellow), PCEF (orange), and the succinate dehydrogenase knock-in (black) include photosystem II (PSII), ferredoxin-NADP$^+$ reductase (FdR), photosystem I (PSI), ferredoxin:quinone oxidoreductase (FQR), cytochrome oxidase bd (COX), cytochrome b$_6$f, NADPH dehydrogenase type 1 (NDH), ATP synthase, and succinate dehydrogenase (SDH). Reactions catalyze oxidations (towards the left) and reductions (towards the right) of ferredoxin (Fd$_{ox}$/Fd$_{red}$), plastoquinone (PQ/PQH$_2$), NADP$^+$/NADPH, and plastocyanin (Cu^{2+}/Cu$^+$). Arrows in the condition-specific panels (top left, top right, bottom left, bottom right) are scaled by the flux of electrons, based on the individual fluxes, the stoichiometry of each metabolite, and the number of valence electrons

exchanged, normalized to the incident number of photons absorbed. Figure S4.tif
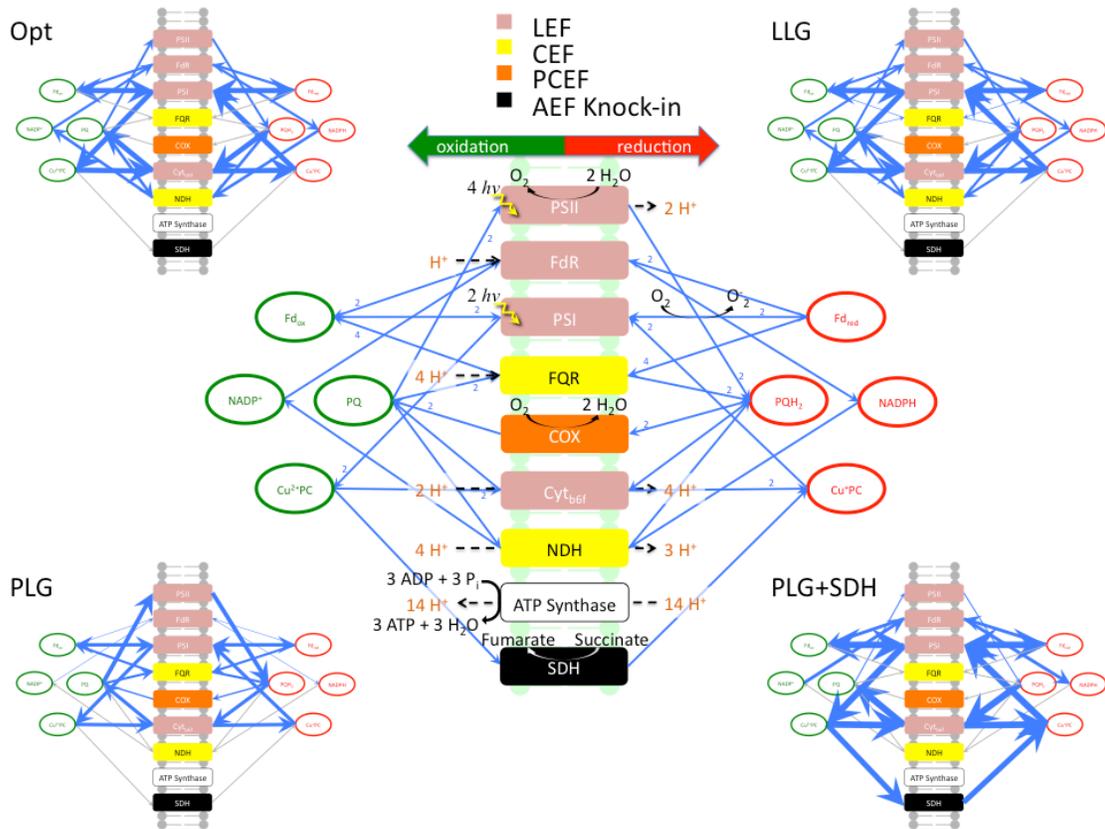
Supplementary File 4.1 - *i*JC568 model in Excel format. BioOpt format and RAVEN

SBML format are available for download at http://biomet-toolbox.org/. SI_File_1.xls

Supplementary File 4.2 - Accompanying datasets used and produced in this

manuscript. SI_File_2.xls

readme - Description of each dataset as well as hyperlinks to navigate.

BOF - Molar and mass based composition of crude biomass fractions and their

individual components.

Elemental stoichiometry - Comparison of BOF elemental stoichiometry with

experimental data.

Enthalpy of combustion - Heats of combustion for each of the biomass components

on a molar and carbon molar basis.

CLG Sensitivity - Carbon limited growth biomass sensitivity.

LLG Sensitivity - Light limited growth biomass sensitivity.

PLG Sensitivity - Phosphorus limited growth biomass sensitivity.

NLG Sensitivity - Nitrogen limited growth biomass sensitivity.

repMets vs Shadow - Table of the top 10 most positive and negative shadow prices

for PLG conditions and their corresponding Z-scores from the reporterMetabolites

algorithm.

Rate validations - Comparison of growth rates, photosynthetic parameters, and

internal fluxes for *i*JC568 and experimental data.

Gene Information - Annotations and identifiers for each metabolic gene included in

*i*JC568, gene length, strand sense, whether the gene belongs to the core or flexible

Prochlorococcus pan-genome, the expression level from Wang *et al.,* (29), and the gene product molecular weight and isoelectric point.

Gene Essentiality - Results from *in silico* gene knockouts. This worksheet includes essential and non-essential genes from the autotrophic and mixotrophic growth simulations. Mixotrophic growth includes a third classification for 'variable' essential genes which were lethal deletions only under certain media compositions.

Ensemble Models - Table summarizing ensemble models (number of metabolites, reactions, metabolic genes, total genes, essential genes).

Intracellular P_i - Table of intracellular phosphate concentrations in P-replete and P-limited media for *Prochlorococcus marinus* MED4, *Synechococcus* WH7803, *Escherichia coli* MG1655, and *Saccharomyces cerevisiae*.

Succinate costs in mutants – Table of NAD(P)H costs associated with *de novo* succinate synthesis for each of the strain variants (WT, +SDH, +2OGDC+SSADH, and +2OGDC+SSADH+SDH). Values for NAD(P)H consuming reactions represent the difference between fluxes in the steady-state solution and fluxes in the forced accumulation solution.

A data-driven model of the diel cycle of Prochlorococcus *physiology and metabolism*

John R. Casey

Abstract

Predicting the dynamics of cellular phenotype from a static genotype poses a profoundly complex challenge, with broad applications. Towards this goal, physiology and metabolism constitute a partial description of cellular phenotype, and can be modeled within a stoichiometric framework. Integrating transcriptome time-series data with a genome-scale metabolic network reconstruction of *Prochlorococcus* metabolism, we describe an approach for predicting time-of-day dependent biomass compositions. By discretizing the resulting biomass objective function in time, diel flux balance analysis simulations provide a time-series of context-specific flux distributions and growth rates. Integrated growth rates determined with this approach differ from those of a static objective function, and compare favorably with experimental results. Predicted changes in the elemental ratios and enthalpy of biomass over the diel cycle may broadly impact microbial food web processes.

5.1 Introduction

Through their broad applications in bioprocess engineering and systems medicine, genome-scale stoichiometric flux models of metabolism (GEM's) have become a routine approach to the quantification of metabolic fluxes and yields (Varma and Palsson, 1994) and subsequently the design of efficient process strains (Nielsen and Keasling, 2016) and the elucidation of new drug targets for metabolic diseases (Mardinoglu and Nielsen, 2015). Optimization approaches with GEM's are poised to provide powerful predictions of complex and dynamic interactions between environmentally relevant organisms, capitalizing on their gene-protein-reaction (GPR) associations and the explosion of 'omics data from environmental samples as metabolic controls. Accordingly, an approach to incorporating environmental transcriptomic or proteomic data into the GEM framework is needed, without the requirement of synoptic and detailed physiological and biochemical contextual information, which are not available for most environmental samples.

A common approach to interpreting gene expression data is to infer metabolic changes and physiological alterations from genes associated with a particular function or pathway. However, pathways and certainly individual enzymes cannot be observed in isolation; emergent from the circuitry of a highly connected metabolic network topology is a system of thermodynamically controlled logic gates which govern metabolism (Chubukov *et al.,* 2014). To this end, several methods exist for interfacing GEM's with gene expression data, from widely different angles. Differential expression data have been used to identify 'reporter' metabolites, and their annealed sub-networks, associated with significant changes

in expression (Patil and Nielsen, 2005), to alter flux constraints in a binary fashion (Covert and Palsson, 2002; Shlomi *et al.,* 2007; Covert *et al.,* 2008; Colijn *et al.,* 2009), to probabilistically alter flux constraints (Chandrasekaran and Price, 2010) or to match flux changes to differential expression changes (Jensen and Papin, 2011). A separate approach minimizes a cost function for metabolic adjustment relative to differential expression (Becker and Palsson, 2008; Collins *et al.,* 2012). Common to all these approaches is a key assumption about a proportionality relationship between gene expression levels and fluxes, an assumption that has been empirically challenged (Chubukov *et al*., 2013; Korchanowski *et al.,* 2013; Gerosa *et al.,* 2013). We implemented a subtly different approach, interpreting the relative expression of metabolic genes as a synoptic view of metabolic capabilities, rather than as binary or proportional control variables on fluxes. Our approach is heavily based on a combination of GIMME (Becker and Palsson, 2008), TEAM (Collins *et al.,* 2012), and flexFBA (Birch *et al.,* 2014). By allowing for flexibility in the objective function, we generate condition-dependent or time-dependent objective functions reflecting the 'ability' of the entire network to produce a particular component of biomass by constraints imposed from relative expression levels, while penalizing disproportionate biomass compositions. Fluxes can then be optimized in subsequent simulations replicating experimental conditions to each discretized objective function.

Much of the variance in the frequency domain of marine microbial metabolism is found at the diel scale, with synchronous effects radiating throughout the marine food web. Although it is interesting to consider that the coordination of

171

periodic changes in physiology and metabolism imparts some advantage to microbial communities, it is unclear whether this attribute is important at longer time-scales. However, physiological alterations would imply changes to both the elemental stoichiometry and the energy density of biomass. A dominant primary producer in much of the nutrient-deplete subtropical oceans is the tiny cyanobacterium *Prochlorococcus* (Liu *et al.,* 1997), serving to some extent as a pacemaker for community interactions and governing the balance between community production and respiration at broad scales. Time-series studies at 12 d^{-1} sampling frequencies of a high-light adapted isolate, MED4, have revealed widespread coordination of gene expression (Zinser *et al.,* 2009; Waldbauer *et al.,* 2012), suggesting diel rhythms to both physiology and metabolism. In addition to the great diel vertical migration (DVM) of large zooplankton to feed near the surface at night, is has recently been suggested that periodic mortality in *Prochlorococcus* (Ribalet *et al.,* 2015), perhaps mediated by phagocytotic protists (Caron *et al.,* 1991), is in phase with the DVM. Relative displacement of the grazing cycle from the *Prochlorococcus* physiology cycle would result in the trophic transfer of biomass chemically altered from time-invariant grazing or time-invariant biomass composition. In addition, relative displacement of the grazing cycle from the *Prochlorococcus* metabolic cycle may alter complementary heterotrophic microbial metabolism, growth and thus the daily photosynthesis-respiration balance.

In the current study, we consider only laboratory-based (isolate) time-series datasets for which extensive experimental data are avialable (Zinser *et al.,* 2009; Waldbauer *et al.,* 2012) to evaluate whether an extension to metatranscriptomic

datasets from field surveys is warranted. These transcriptomic time-series datasets were integrated with a MED4 GEM to predict patterns in biochemical composition, elemental ratios, growth, and metabolic fluxes over the diel cycle.

5.2 Methods

5.2.1 GEM reconstruction

A GEM for *Prochlorococcus* MED4 (*i*JC568) consisting of 568 metabolic genes encoding 794 reactions with 680 metabolites distributed in 6 subcellular locations (cytoplasmic membrane, periplasm, thylakoid membrane, thylakoid lumen, cytoplasm, and carboxysome) was performed according to standard procedures (Thiele and Palsson, 2010; Agren *et al*., 2013) and described elsewhere (Casey *et al.,* 2016). Briefly, an initial draft reconstruction was manually curated for reaction directionality, substrate and co-factor specificity, gap-filling, mass and charge balance, elimination of futile cycles, and protein localization. Metabolic capabilities and detailed biomass compositions were acquired from the literature and from experimental data (described in Casey *et al.,* 2016) and growth and non-growth associated maintenance ATP requirements were calculated according to the method of ATP utilization costs and estimated polymerization costs (Feist *et al.,* 2007).

5.2.2 Incorporation of gene expression time-series as constraints

Gene expression time-series datasets were acquired from two nearly identical *Prochlorococcus* MED4 isolate experiments which we will refer to as Dataset 1 (Zinser *et al.,* 2009) and Dataset 2 (Waldbauer *et al.,* 2012). In both

experiments, axenic isolates were grown on replete medium (PRO99; Moore *et al.,* 2007) under a simulated natural 14 h : 10 h light:dark regime, reaching a peak irradiance of 232 μmol photons m^{-2} s^{-1} at an incubation temperature of 24 °C. Transcriptome samples were collected at 2 h intervals, over the course of two days. To normalize expression levels over the diel cycle, a cumulative distribution function (CDF) was generated for each transcript.

Transcripts and their mean relative expression diel time-series (e^{met}) were detrended and the residuals were collapsed to a single 24 h diel cycle by averaging. 2 h binned residual means were mapped to *i*JC568 metabolic genes and to their corresponding enzyme catalyzed reaction or reactions. A threshold (ξ) was applied to the CDF of relative expression level time-series of each reaction to identify time points when a reaction was assumed to be inactive. If a reaction was catalyzed by an enzyme coded by a multiple genes, we assumed that reaction activity was present only if all subunits of the complex satisfied the condition $e^{met} \geq \xi$. In this fashion, reactions were constrained to zero flux, $v_j^{lb} = v_j = v_j^{ub} = 0$ (see below). This approach differs from TEAM and GIMME by removing the cost function associated with fluxes which differ from gene expression level changes, and does not require or benefit from extensive *a priori* knowledge of changing media compositions and biomass data. Instead, we consider that gene expression is itself a reflection of the full complexity of the physiological and physicochemical state of the system, the lags associated with translation, and the inertia of 'system memory' pre-programmed into circadian rhythms.

Modeling Simulations

A stoichiometric matrix, S ($m \times n$), was constructed for *i*JC568, where m is the number of metabolites and n is the number of reactions. Each element S_{ij} corresponds to the stoichiometric coefficient of the ith metabolite participating in the jth reaction. The product of S and a vector of reaction fluxes, v (of length n; in units of mmol gDW^{-1} h^{-1}) is set equal to the accumulation of each metabolite pool dX/dt, where X is a vector of metabolite concentrations of length m. Within a time step of 2 h, we assumed there was no net accumulation (or depletion) of any metabolite pool, and the system is considered only at steady state, $S \bullet v = 0$.

$$\begin{bmatrix} S_{11} & S_{12} & \dots & S_{1n} \\ S_{21} & S_{22} & \dots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \dots & S_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \frac{dX_1}{dt} \\ \frac{dX_2}{dt} \\ \vdots \\ \frac{dX_m}{dt} \end{bmatrix} = 0$$

A 'biomass reaction' vector S_{BIO} of length m and appended to S, is the molar fraction of each metabolite comprising the dry weight (DW) of cellular material in aggregate (in units of mmol gDW^{-1}). The complete S_{BIO} vector is available as a supplementary file elsewhere (Casey *et al.,* 2016). Correspondingly, an element v_{BIO} was appended to v and the resulting matrices are $\hat{S}$ and $\hat{v}$.

$$\left[\begin{array}{cccc|c} S_{11} & S_{12} & \dots & S_{1n} & S_{1BIO} \\ S_{21} & S_{22} & \dots & S_{2n} & S_{2BIO} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ S_{m1} & S_{m2} & \dots & S_{mn} & S_{mBIO} \end{array} \right] \left[\begin{array}{c} v_1 \\ v_2 \\ \vdots \\ v_n \\ \hline v_{BIO} \end{array} \right] = 0$$

The linear programming (LP) problem for flux balance analysis was then solved by optimizing the 'biomass objective function' (BOF). The primal solution

given by the LP solver (Mosek ApS, Denmark) is a set of fluxes $\hat{\boldsymbol{v}}$ such that biomass

growth v_{BIO} was maximized.

maximize
$$Z = c^T \hat{v}$$

subject to
$$\hat{\boldsymbol{S}} \hat{\boldsymbol{v}} = 0$$
$$v_j^{lb} \leq vj \leq v_j^{ub}$$

where c is a binary vector with an element of one at the index corresponding to v_{BIO}

(in units of h^{-1}). A set of lower and upper boundary constraints on each flux (v_j^{lb} and

v_j^{ub}, respectively) was determined by reaction directionality in most cases, with an

arbitrarily high flux (±1000 mmol gDW^{-1} h^{-1}) as boundaries on reversible reactions,

and zero as a lower boundary on irreversible reactions in the forward direction. In

all simulations, a constraint on the carboxylase flux of ribulose bisphosphate

carboxylase/oxygenase was implemented based on the analytically determined

maximum photosynthetic rate of *Prochlorococcus* MED4 (Zinser *et al.,* 2009). As

noted above, relative gene expression values lower than a threshold $e^{met}_j \leq \xi$ were

iteratively included as additional constraints, using a flexible objective function

(described below).

The traditional FBA formulation requires that the network is capable of

synthesizing, simultaneously, all components of the BOF in the static ratios defined

in the biomass reaction. If, for example, a simulated knock-out of the enzyme

polyprenyl synthetase would prevent synthesis of an important component of the

cell wall and result in an infeasible LP solution. To allow cellular physiology to vary

over the diel cycle, a flexible objective approach (flexFBA, adapted from Birch *et al.,* 2014), was applied. Leaving S_{iBIO} intact in $\hat{\boldsymbol{S}}$, individual elements of S_{iBIO} were parsed to a diagonalized square matrix (of dimensions $m \times m$) and appended to the reaction space of $\hat{\boldsymbol{S}}$. Accordingly, a vector $\boldsymbol{f}$ of length m representing the fractional fulfillment of each element of S_{iBIO} was appended to $\hat{\boldsymbol{v}}$. We refer to these matrices as $\tilde{\boldsymbol{S}}$ and $\tilde{\boldsymbol{v}}$:

$$
\left[
\begin{array}{cccc|c|cccc}
S_{11} & S_{12} & \ldots & S_{1n} & S_{1BIO} & S_{1BIO} & 0 & \ldots & 0 \\
S_{21} & S_{22} & \ldots & S_{2n} & S_{2BIO} & 0 & S_{2BIO} & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ldots & \ddots & \vdots \\
S_{m1} & S_{m2} & \ldots & S_{mn} & S_{mBIO} & 0 & \ldots & \ldots & S_{mBIO}
\end{array}
\right]
\left[
\begin{array}{c}
v_1 \\
v_2 \\
\vdots \\
v_n \\
\hline
v_{BIO} \\
\hline
f_1 \\
f_2 \\
\vdots \\
f_m
\end{array}
\right] = 0
$$

To encourage a complete biomass composition (i.e., to avoid the scenario that growth is allowed without e.g., a particular deoxyribonucleotide), an L1-norm penalty (γ) is applied to fractional fulfillments less than proportional to the fractional fulfillment of ATP. The steady-state is maintained, and thus the optimization becomes:

maximize

$$
Z = f_{ATP} - \gamma \sum_{i=1}^{n} (f_{ATP} - f_i) + \beta v_{BIO}
$$

subject to

$$
\tilde{\boldsymbol{S}} \tilde{\boldsymbol{v}} = 0
$$
$$
v_j^{lb} \le v_j \le v_j^{ub}
$$
$$
0 \le v_{BIO} \le f_i
$$
$$
f_i \le f_{ATP}
$$

where β is a weight applied to the BOF to incentivize maximal growth rate. The additional constraints require that the growth rate of the BOF is less than the

fractional fulfillment of any component of the BOF, and that fractional fulfillment of all BOF components are less than the ATP fractional fulfillment. A comprehensive description of the flexFBA formulation is available elsewhere (Birch *et al.,* 2014).

Solving the flexFBA problem iteratively using constraints imposed by relative gene expression on each reaction yields a matrix $\boldsymbol{F}$ (of dimensions $m \times n \times k$). Accordingly, each element F_{ijk} corresponds to the ith fractional fulfillment resulting from the jth reaction constraint for each kth time point. The cumulative effect of reaction constraints on each component, f_i^{total} was evaluated at each time point t according to:

$$f_i^{total}(t) = \frac{\sum_{j=1}^{n} f_{ij}(t)}{\sum_{j=1}^{n} f_{BIO,j}(t)}, \; \forall t \{t \in 1 \dots k\} \text{ in } F_{ijk}$$

where $f_{BIO,j}$ is the fractional fulfillment of the consolidated biomass component. The resulting vector f^{total} was then normalized to the biomass reaction S_{BIO} and evaluated at each time point:

$$S_{iBIO}(t) = S_{iBIO} \frac{f_i^{total}(t) - min(f_i^{total}(t))}{max(f_i^{total}(t)) - min(f_i^{total}(t))}$$

The final set of vectors $S_{BIO}(t)$ were implemented as time-dependent BOF equations (BOF$_t$) in diel simulations and further used to calculate the elemental composition of biomass synthesized at each time point. Briefly, growth irradiance values from each study were converted to photon fluxes by the wavelength specific absorption efficiency and optical cross sections for MED4 (Morel *et al.,* 1993). These values were included as upper boundaries on the photon flux at each time point. Nighttime respiration was assumed to be satisfied by respiration of glycogen stores,

which were calculated as the daytime time-integrated glycogen flux. Other nighttime

energy sources are feasible for iJC568, including beta-oxidation of lipids or

polyphosphate substrate-level phosphorylation, however these were not

considered. Using these time-dependent constraints and the BOF$_t$, FBA was used to

solve for growth rates and fluxes in discretized time. Elemental growth, the

elemental composition of biomass accumulated over the time-series (E_k; mmol gDW$^-$

^{1}), was calculated according to:

$$E_k = E^0 + \sum_{k=1}^{t} \sum_{i=1}^{m} \mu_k \left(S_{iBIOk} \; E_{iBIOk} \right) \Delta t$$

where E^0 is the initial elemental composition of biomass, μ_k is the growth rate at

time k, and E_{iBIOk} is the elemental composition of the ith component of the biomass

equation in S_{iBIOk}, determined from molecular formulas. Elemental growth time-

series were calculated for C, H, O, N, P, and S, and their ratios. Additionally, the

carbon redox number and the enthalpy of combusion of biomass was calculated at

each time point. Standard enthalpies of combustion of each molecule ($\Delta H^0_{C,i}$) were

calculated according to a modification of the Kharash equation (Kharash, 1929).

We stress that degradation processes, including the respiration of glycogen

stores and the many protein repair and replacement activities, are not included in

fractional fulfillment or elemental budgets. Modeling degradation processes in the

stoichiometric framework would require knowledge of the physicochemical

stability, thermodynamic properties, structural properties, kinetics, and abundance

of all proteins represented in the network. Lacking this information, our results

indicate only synthetic activities and are to be considered gross rather than net budgets. In the physiological context, repair of proteins with short half-lives (*i.e.*, those that are rapidly degraded in vivo, *e.g.*, the D2 protein of photosystem II) could be misinterpreted as net synthesis using this method, and extension of this approach would likely fail to provide meaningful results in spores, slow-growing vegetative cells and tissues.

5.3 Results

5.3.1 Determination of biosynthetic activities over the diel cycle

A common approach to interpreting metabolic activities of microbial metabolism over *e.g.*, a diel cycle is to plot relative expression levels of the enzyme which catalyzes the synthesis of a molecule of interest. We will refer to this approach as the 'nearest neighbor' method. Few quantitative compound-specific time-series datasets are available for *Prochlorococcus* MED4 or ecotype MED4 at hourly intervals, however diel datasets for DNA replication are available from both isolate studies incorporated in this work. The timing of DNA replication in *Prochlorococcus* is tightly regulated, both in culture (Bruyant *et al.,* 2005) and in field samples (Liu *et al.,* 1997), and most replicating cells are in S-phase within an hour before or after local 'sunset'. Thus it is expected that all four deoxyribonucleotides be phase locked to the diel cycle and exhibit a similar profile between all three datasets, given that no metabolic function apart from DNA synthesis is known for dATP, dGTP, dCTP, and dTTP in MED4. The timing of DNA nucleotide synthesis $S_{iBIO}(t)$ from our flexible objective simulations was closely

correlated with flow cytometry cell cycle analysis based determinations of the timing of the peak in S-phase cells in all three datasets, whereas relative gene expression of the nine 5'-phosphotransferase enzymes responsible for synthesizing the four deoxyribonucleotides in MED4 was not (**Figure 5.1**).

Although the special case of DNA does not necessarily hold for other biomass components, coherence between datasets and phase locking to the diel cycle are expected, as >86% of *Prochlorococcus* transcripts corresponding to metabolic genes in *i*JC568 were significantly periodic (false discovery rate < 0.05) in the two experimental studies. To evaluate consistency of results between datasets, we compared the distribution of p-values of correlation coefficients by pairwise Pearson's linear correlations, with the hypothesis that the correlation is greater than zero (**Figure 5.2**). The proportion of significantly similar flexFBA results between datasets was 55%, compared to 35% for nearest neighbor results. To evaluate the periodicity of flexFBA results against nearest neighbor results, diel time series were fit to sines, and plotted as the distribution of coefficients of determination (**Figure 5.3**). While non-parametric tests for periodicity are available, and may better describe more complex periodic functions of gene expression regulation beyond a simple harmonic oscillator, the sine fit is still instructive. The proportion of R^2 values greater than 0.8 was 51% for flexFBA, compared with 31% for the nearest neighbor method.

5.3.2 BOF and BOF$_t$ simulated growth rates

The growth rates reported in the original studies, based on changes in cell number at 24 h intervals, were 0.66 ± 0.04 d^{-1} (Dataset 1; Zinser *et al.,* 2009) and 0.56 ± 0.02 d^{-1} (Dataset 2; Waldbauer *et al.,* 2012). Simulations with flexible objectives (BOF$_t$) using the irradiance data from each study resulted in growth rates of 0.60 d^{-1} (Dataset 1) and 0.58 d^{-1} (Dataset 2), whereas the growth rate from the fixed objective (BOF) simulation using the irradiance data from Dataset 1 was 0.81 d^{-1} (**Figure 5.4**). Large variations in instantaneous growth rates were observed between datasets for BOFt simulations at the first sampling time-point following 'sunrise' and preceeding 'sunset'. Closer inspection revealed that these differences were primarily due to an asymmetry in the relative expression values of key photosystem I genes (*psaAB*) which were similar to the CDF threshold value ($e^{met}_j \approx \xi = 0.2$) at these time-points. Because of the absence of succinate dehydrogenase, constraining the flux through photosystem I to zero dramatically reduces *i*JC568 photochemical efficiency. Thus, our approach for the determination of objective functions is sensitive to significantly periodic metabolic genes with small amplitude oscillations relative to high-frequency noise throughout a time-series, and is likely a bias introduced by sequencing of low relative abundance transcripts. Presmoothing of time-series could be accomplished with a Gaussian filter, based on a recent proposal that an autoregressive model with Gaussian noise explains constitutive gene expression dynamics at the single-cell level in *Escherichia coli* strains (Tanouchi *et al.,* 2015).

5.3.3 Diel cycle of crude fractions of biomass

The BOF (or BOF_t) can be aggregated into crude fractions (e.g., protein, DNA, etc.) which tend to co-vary *in vivo*. This grouping introduces some noise originating from variability within a particular fraction; as an interesting example, results from both datasets indicate a difference in the timing of lipid head-group synthesis with digalactosyldiacylglycerol (DGDG) and phospatidylglycerol (PG) maxima occurring between 0400 and 0800 local time, and sulfoquinovodiacylglycerol (SQDG) and monogalactosyldiacylglycerol (MGDG) maxima occurring between 1000 and 1400 local time. This discrepancy may be due to differing lipid compositions of thylakoid and cell membranes, since peak synthesis times for DGDG and PG coincided with pigment peak synthesis times, or may perhaps indicate an energy storage pool composed of SQDG and MGDG lipids not included in our simulations. Nevertheless, the timing of the fractional fulfillments of each crude fraction (**Figure 5.5**) were consistent between datasets ($p < 0.05$ for all fractions except 'minerals' where $p = 0.84$). The 'mineral' fraction was noisy due to few genes (15 in total) associated only with transporters governing the appropriate flux constraints, and while no gene is annotated in MED4 for Mo transport, this element is included in S_{BIO}. In this case, individual ions or minerals (e.g., Zn) should be treated individually; similarly, the 'dissolved pool' is a composite of many metabolites with varying functions (e.g., osmolytes, energy co-factors, electrolytes). Despite these caveats, the general sequence of synthesis begins pre-dawn with pigments, followed by lipids and glycogen at midday, genome replication in mid-afternoon, followed by free nucleic acids, RNA, protein and cell-wall synthesis following cell division after sunset. When plotted as mass fractions over the diel cycle (**Figure 5.6**), differences in the timing

183

of peak glycogen and lipid synthesis are especially apparent between Dataset 1 and
Dataset 2, while other crude fractions follow a generally similar pattern.

5.3.4 Diel cycle of biomass elemental ratios

Solutions to BOF_t simulations yield both instantaneous and cumulative
elemental compositions of biomass (E_{BIOk}). Mass and elemental balances, which are
expected for the steady state condition, were validated by comparing the biomass
time-integrated fluxes at each BOF_t composition to time-integrated system
boundary fluxes. All balances were within 0.1% of an ideal closed system, with
leakage in mass attributed to rounding errors in molecular weights.

"Instantaneous" and "cumulative" elemental compositions of biomass in
discrete time were quite similar between Dataset 1 and Dataset 2 (**Figure 5.7**).
"Instantaneous" is defined as the composition of synthesis *de novo* at each 2 h
interval, while "cumulative" refers to the aggregate composition weighted by growth
rates and integrated in discrete 2 h intervals over the 24 h time-series, i.e., what one
might measure by harvesting cells at each timepoint, neglecting dedgradation.
Driven primarily by protein relative to glycogen and lipid, instantaneous C:N
synthesis ratios varied from approximately 4 mol C [mol N]$^{-1}$ after cell division to a
peak of 6 to 8 mol C [mol N]$^{-1}$ at midday. A more dramatic variation (approximately
4-fold change) was observed in C:P ratios, driven by the cumulative effect of minima
in phospholipid and RNA levels in the 2 to 4 h following sunrise, an effect that is
echoed in the N:P ratio. C:S ratios were driven primarily by the dominant lipid
SQDG, and to a lesser extent the amino acids L-cysteine and L-methionine, with

184

resulting instantaneous synthesis ratios spanning the range 40 – 140 mol C [mol S]$^{-1}$. The instantaneous C:H ratio was effectively static over the diel cycle, and C:O ratios varied by 30 to 60% with minima during daytime as expected from the reduction potential of photosynthesis. While the instantaneous elemental composition of biomass synthesized at each timepoint was highly variable, the cumulative elemental composition was dampened. Due to relatively low instantaneous growth rates at night, differences in BOF_t were most sensitive to variations in crude fractions during the daylight hours, and the accumulated elemental ratios varied by less than 15% over the diel cycle (**Figure 5.8**).

Using the instantaneous elemental composition of synthesized biomass, a calculation of the average carbon redox number as a function of time is trivial. While the C:O ratio was variable over the diel cycle, reflecting the photochemical reduction power of photosynthesis, the average carbon redox number was relatively static (**Figure 5.9**). It seems that the timing of protein synthesis offsets any net reduction of biomass, effectively shuttling electrons from newly fixed carbon atoms in sugars to carbon atoms in proteins. Instead, it is the total enthalpy of combustion of biomass synthesized (instantaneously) over the diel cycle that is minimal during daylight hours. Since the change in entropy of biomass over time could be assumed to be fairly static, the transduction of exergy to Gibbs energy from low carbon enthalpy stores (e.g., $\Delta H^0_{C,glycogen}$ = 466 KJ [mol C]$^{-1}$) during the daytime to high carbon enthalpy structural components like pigments ($\Delta H^0_{C,pigments}$ = 542 ± 11 KJ [mol C]$^{-1}$) and protein ($\Delta H^0_{C,protein}$ = 536 ± 63 KJ [mol C]$^{-1}$) at nighttime must be

accompanied by increased heat dissipation or by entropy production by metabolite excretion during the daylight hours.

5.4 Discussion

Spurious feedback effects resulting from the dynamics of complex networks and high-frequency noise in gene expression levels are obstacles to the inference of the timing of physiological and metabolic activities. For example, a comparison of time-series from the two datasets considered in this study of relative gene expression levels for those enzymes directly catalyzing the synthesis of deoxyribonucleic acids were poorly correlated with the known patterns of DNA synthesis for the MED4 isolate *in vivo*. By incorporating gene expression of all metabolic genes with *i*JC568, accurate predictions could be made for the timing of synthesis of all four deoxyribonucleotides. Although the predictions of the timing of synthesis of each crude fraction of biomass using the flexible objective approach were consistent between datasets, not sensitive to the arbitrary expression level cutoff parameter ξ within the range (less than 14 % variation in the fractional fulfillment of crude fractions over time in the interval $0.15 \leq \xi \leq 0.35$), and relatively well described by simple harmonics, experimental data beyond DNA synthesis are lacking to validate our predictions.

Extension of the complete set of biosynthetic predictions to generate conditional biomass objective functions resulted in improved growth rate predictions relative to observations than with a fixed steady state objective function in diel simulations. While time-series data were used in this study, the flexible

objectives method is by no means limited to this format and could easily be adapted to differential expression data from stress response experiments or adaptive evolution studies. However, the current system is not a realistic model of *in vivo* physiology, due to the dynamic nature of storage reservoirs which are not represented in a linear stoichiometric framework and were instead parameterized. Despite this obvious limitation, gene-expression-based estimates of time resolved biomass objectives allow for the calculation of the accumulation of storage pools like glycogen, which we assume to be the major source of respiratory energy and carbon for nighttime activity. The accumulated glycogen in the isolate studies during the daytime was 185 and 171 mg gDW^{-1} and nighttime biosynthetic ATP requirements resulted in a glycogen carbon-specific growth efficiency of 22 % and 29 % for Dataset 1 and Dataset 2, respectively. Validation of the magnitude of these stores will enable the calculation of light-dependent respiration rates, an analytically challenging measurement.

The elemental composition of instantaneous biomass production computed from BOF$_t$ showed consistent patterns in C:N:P:S ratios, and to a lesser extent C:H:O ratios over the diel cycle. When computed for accumulated biomass, the variability in these ratios was suppressed, but not completely. Since *Prochlorococcus* mortality appears to also be periodic (Ribalet *et al.,* 2015), even subtle changes in the elemental stoichiometry of biomass consumed by grazers could have a profound impact on the composition of the sinking flux, the stoichiometry of upward eddy diffusion supplied nutrients, and thus the productivity of the surface layer. We also noted a dramatic change in $\Delta H^0_{C,biomass}$ synthesized over the diel cycle. Relative

displacement of the grazing cycle and the $\Delta H^0_{C,biomass}$ cycle would bias the trophic energy transfer efficiency. In addition, this prediction has implications for the time-integrated Gibbs energy balance and suggests the need for either enhanced heat flow during the daytime, which we do not predict from daytime respiration rates, or more likely the excretion of 'overflow' metabolites, further fueling the microbial food web.

5.5 References

Becker, S.A., Palsson, B.Ø., 2008. Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology* 4, e1000082.

Birch, E.W., Udell, M., Covert, M.W., 2014. Incorporation of flexible objectives and time-linked simulation with flux balance analysis. *Journal of theoretical biology* 345, 12–21.

Bruyant, F., Babin, M., Genty, B., Prášil, O., Behrenfeld, M.J., Claustre, H., Bricaud, A., Garczarek, L., Holtzendorff, J., Koblizek, M., Dousova, H., Partensky, F., 2005. Diel variations in the photosynthetic parameters of *Prochlorococcus* strain PCC 9511: Combined effects of light and cell cycle. *Limnology and Oceanography* 50, 850-863

Caron, D.A., Lim, E.L., Miceli, G., Waterbury, J.B., Valois, F.W., 1991. Grazing and utilization of chroococcoid cyanobacteria and heterotrophic bacteria by protozoa in

laboratory cultures and a coastal plankton community. *Marine Ecology Progress Series* 76, 205–217.

Casey, J., Mardinoglu, A., Nielsen, J., Karl, D.M., 2016. Adaptive evolution of phosphorus metabolism in *Prochlorococcus. mSystems* 1, e00065–16.

Chandrasekaran, S., Price, N.D., 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America* 107, 17845–17850.

Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R.J., Jules, M., Link, H., Aymerich, S., Stelling, J.O.R., Sauer, U., 2013. Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis. Molecular Systems Biology* 9, 1–13.

Chubukov, V., Gerosa, L., Kochanowski, K., Sauer, U., 2014. Coordination of microbial metabolism. *Nature Reviews Microbiology* 12, 327–340.

Colijn, C., Brandes, A., Zucker, J.D., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.-Y., Moody, D.B., Murray, M., Galagan, J.E., 2009. Interpreting expression data with metabolic flux models: Predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Computational Biology* 5, e1000489.

Collins, S.B., Reznik, E., Segre, D., 2012. Temporal expression-based analysis of metabolism. *PLoS Computational Biology* 8, e1002781.

Covert, M.W., Palsson, B.Ø., 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli. Journal of Biological Chemistry* 277, 28058–28064.

Covert, M.W., Xiao, N., Chen, T.J., Karr, J.R., 2008. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics* 24, 2044–2050.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* 3, 121.

Gerosa, L., Kochanowski, K., Heinemann, M., Sauer, U., 2013. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Molecular Systems Biology* 9, 1–11.

Jensen, P.A., Papin, J.A., 2011. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* 27, 541–547.

Kochanowski, K., Sauer, U., Chubukov, V., 2013. Somewhat in control—the role of transcription in regulating microbial metabolic fluxes. *Current Opinion in Biotechnology* 24, 987-993.

Kharash, M.S., 1929. Heats of combustion of organic compounds. *Bureau of Standards Journal of Research* 2, 359-430.

Liu, H., Nolla, H.A., Campbell, L., 1997. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquatic Microbial Ecology* 12, 39–47.

Mardinoglu, A., Nielsen, J., 2015. New paradigms for metabolic modeling of human cells. *Current Opinion in Biotechnology* 34, 91–97.

Moore, L.R., Coe, A., Zinser, E.R., Saito, M., Sullivan, M.B., Lindell, D., Frois-Moniz, K., Waterbury, J., Chisholm, S.W., 2007. Culturing the marine cyanobacterium *Prochlorococcus. Limnology and Oceanography: Methods* 5, 353–362.

Morel, A., Ahn, Y., Partensky, F., Vaulot, D., Claustre, H., 1993. *Prochlorococcus* and *Synechococcus*: A comparative study of their optical properties in relation to their size and pigmentation. *Journal of Marine Research* 51, 617–649.

Nielsen, J., Keasling, J.D., 2016. Engineering cellular metabolism. *Cell* 164, 1185–1197.

Patil, K.R., Nielsen, J., 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America* 102, 2685–2689.

Ribalet, F., Swalwell, J., Clayton, S., Jiménez, V., Sudek, S., Lin, Y., Johnson, Z.I., Worden, A.Z., Armbrust, E.V., 2015. Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proceedings of the National Academy of Sciences of the United States of America* 112, 8008-8012.

Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E., 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology* 3, 101.

Tanouchi, Y., Pai, A., Park, H., Huang, S., Stamatov, R., Buchler, N.E., You, L., 2015. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* 523, 357–360.

Varma, A., Palsson, B.Ø., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology* 60, 3724-3731.

Waldbauer, J.R., Rodrigue, S., Coleman, M.L., Chisholm, S.W., 2012. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS ONE* 7, e43432.

Zinser, E.R., Lindell, D., Johnson, Z.I., Futschik, M.E., Steglich, C., Coleman, M.L., Wright, M.A., Rector, T., Steen, R., Mcnulty, N., Thompson, L.R., Chisholm, S.W., 2009. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus. PLoS ONE* 4, e5135.

Figure 5.1 – Comparison of the timing of deoxyribonucleotide synthesis between relative gene expression levels (e^{met}; top panels) using the 'nearest neighbor' method and fractional fulfillments calculated with the flexible objectives flexFBA method (f_i; bottom panels) for both Dataset 1 (blue lines and markers) and Dataset 2 (green lines and markers). The fraction of cells in S-phase from flow-cytometry cell cycle analysis data (Zinser *et al.,* 2009) are included in each panel (black lines).
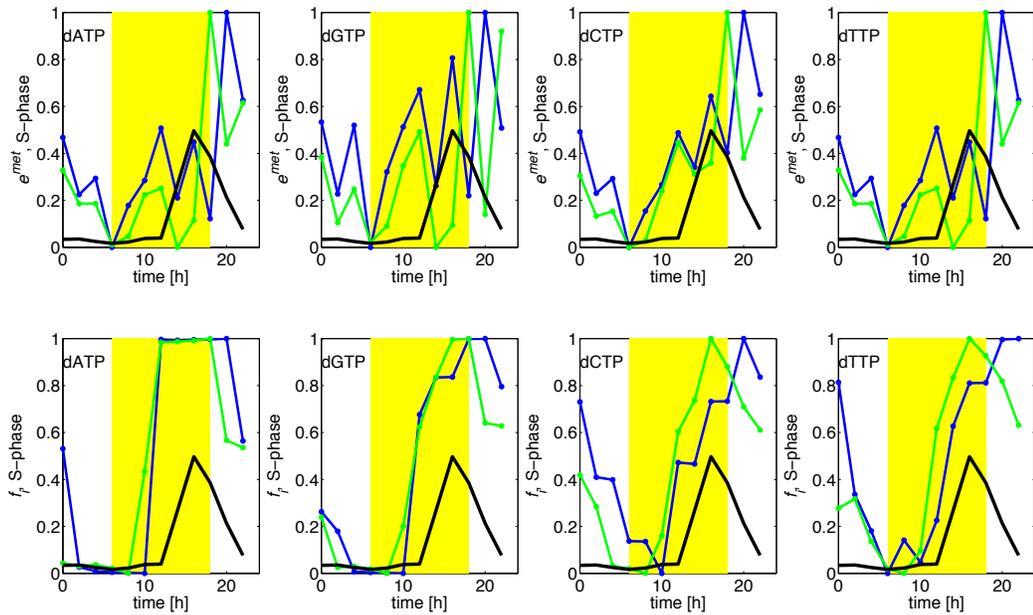
Figure 5.2 – Summary comparison of correlation p-values between Dataset 1 and

Dataset 2 using the nearest neighbor gene expression method (top panel) and

fractional fulfillments calculated with the flexible objectives flexFBA method

(bottom panel).
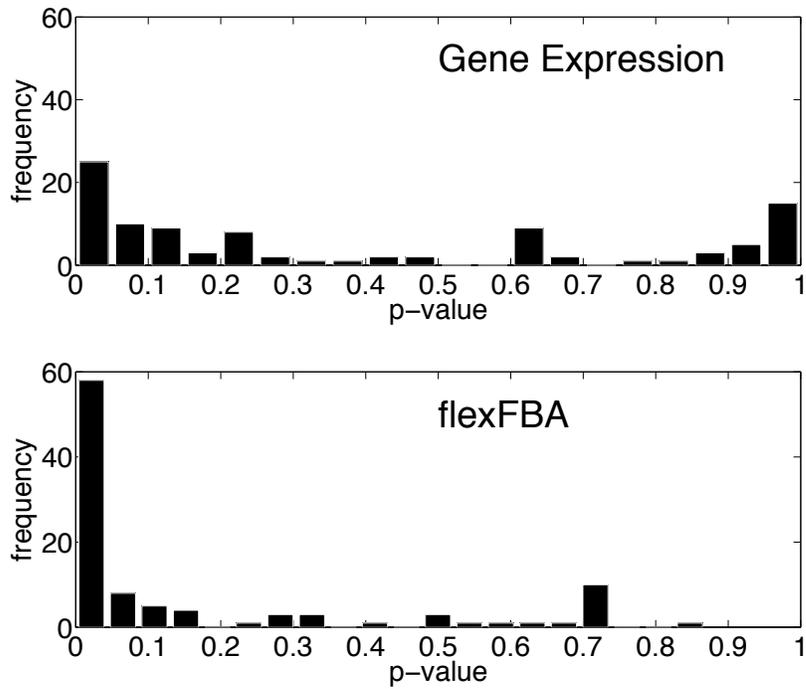
Figure 5.3 – Summary comparison of coefficients of determination resulting from harmonic regression using the nearest neighbor gene expression method (top panel) and fractional fulfillments calculated with the flexible objectives flexFBA method (bottom panel) for Dataset 1 (blue bars) and Dataset 2 (green bars).
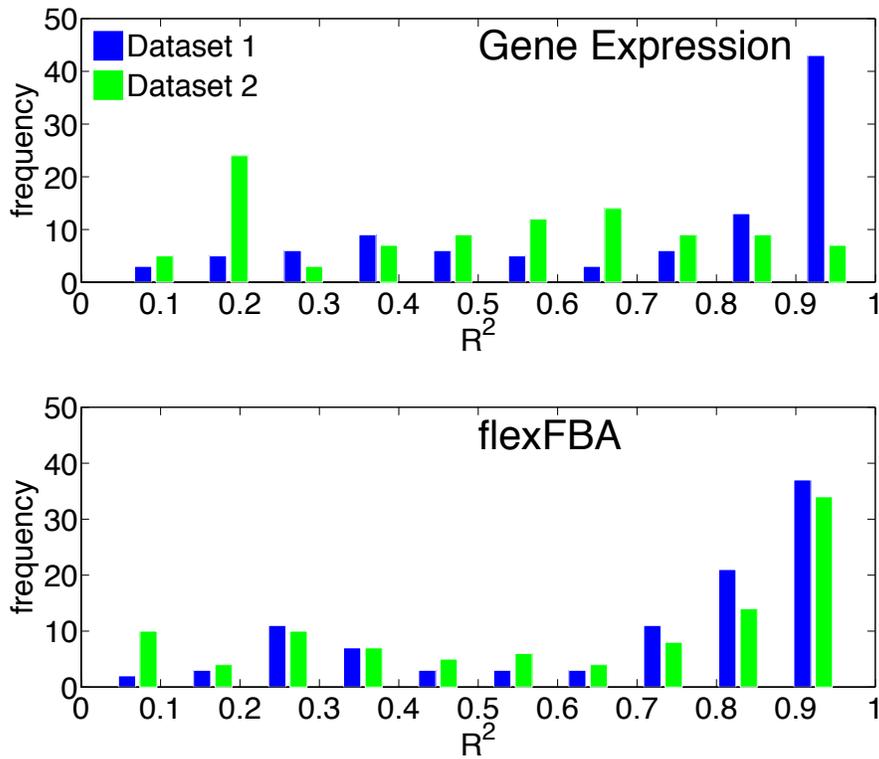
Figure 5.4 – Comparison of instantaneous growth rates during diel simulations using BOF_t calculated from Dataset 1 (blue lines and markers), Dataset 2 (green lines and markers) and using a fixed BOF (black lines and markers).
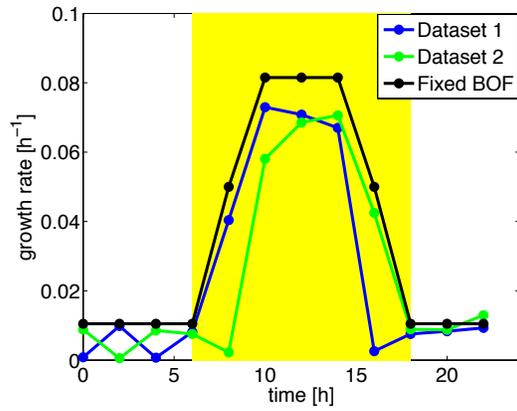
Figure 5.5 – Diel cycle of fractional fulfillments of crude fractions of biomass (f_{crude}) used to compute BOF_t from Dataset 1 (blue lines and markers) and Dataset 2 (green lines and markers). The daylight period is indicated in yellow. f_{crude} was calculated as the average of individual fractional fulfillments (f_i).
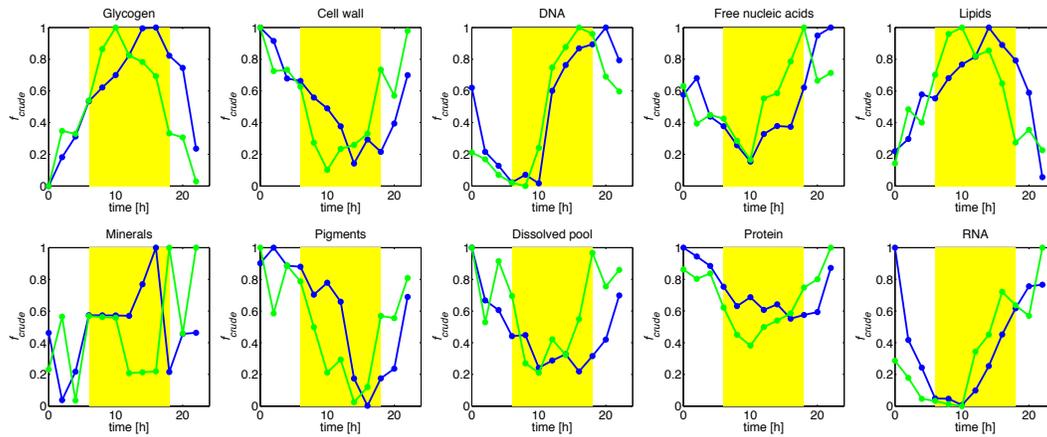
Figure 5.6 – Comparison between Dataset 1 (left panel) and Dataset 2 (right panel) of the diel cycle of crude fractions of biomass synthesized at each time point. The proportions should be interpreted as gross production and do not include degradation, including the respiration of accumulated glycogen.
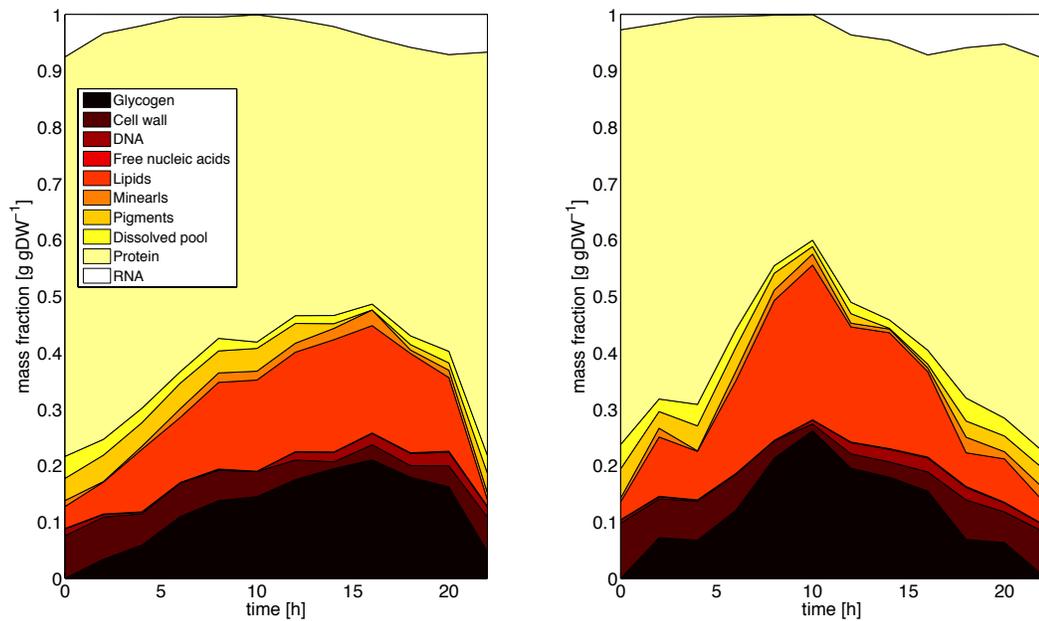
Figure 5.7 – Instantaneous elemental ratios of BOF$_t$ for Dataset 1 (blue lines and markers) and Dataset 2 (green lines and markers).
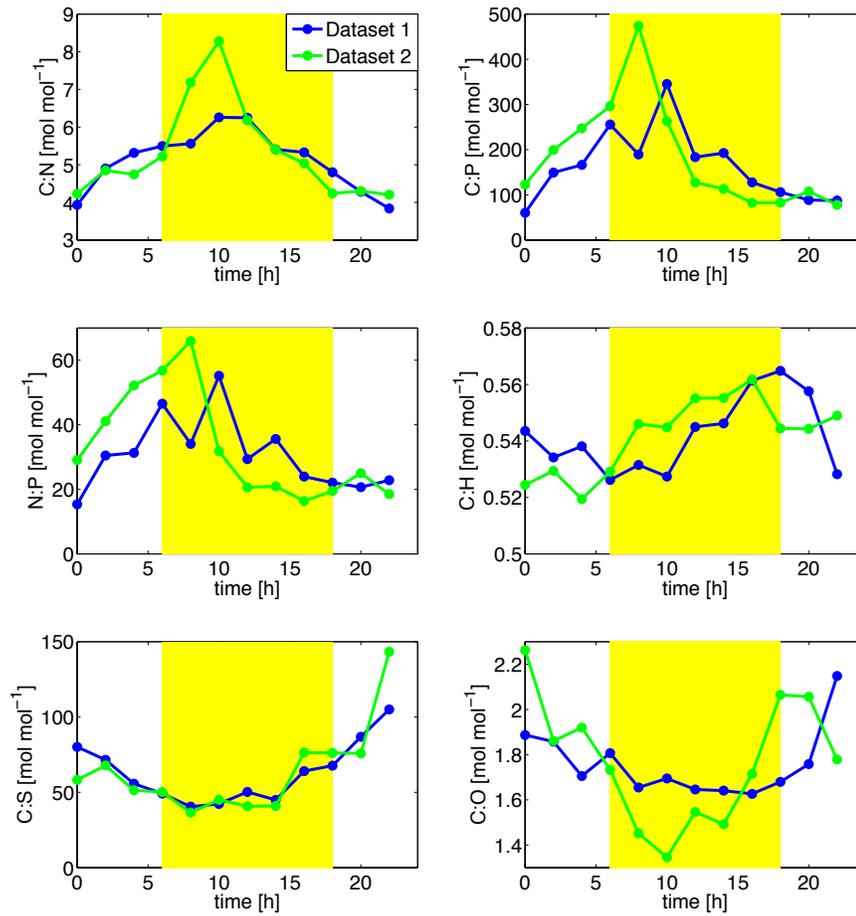
Figure 5.8 – Cummulative elemental ratios of synthesized biomass for Dataset 1 (blue lines and markers) and Dataset 2 (green lines and markers).
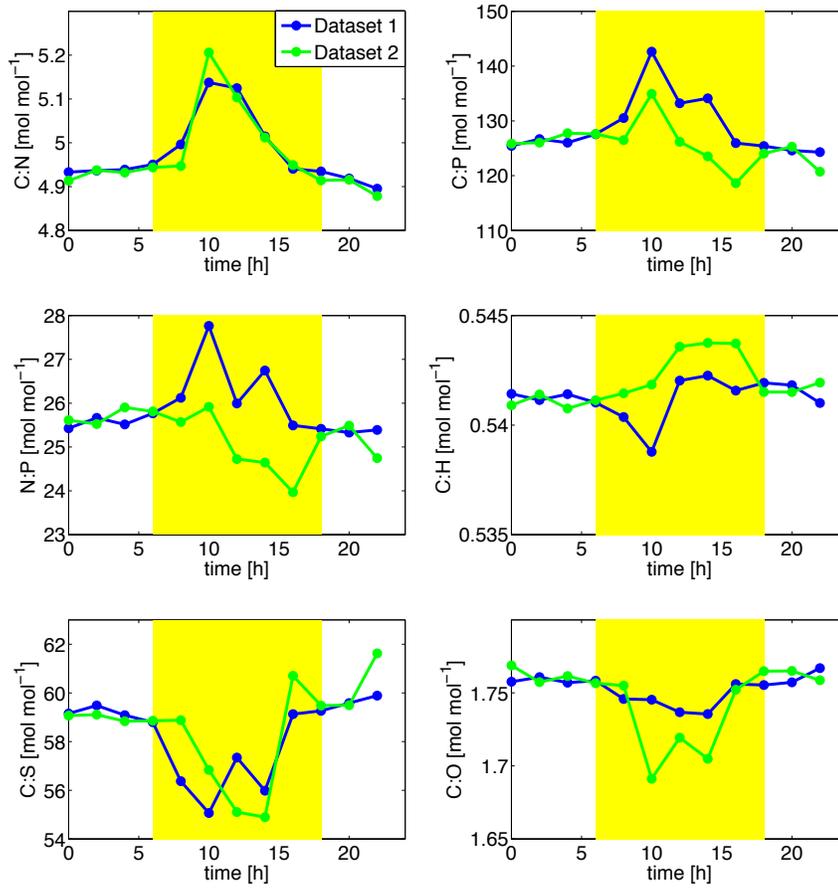
Figure 5.9 – Average carbon redox number and total enthalpy of combustion (ΔH^0_C) of BOF$_t$ instantaneous compositions for Dataset 1 (blue lines and markers) and Dataset 2 (green lines and markers).